Studies in Computational Intelligence 583

Van-Nam Huynh
Vladik Kreinovich
Songsak Sriboonchitta
Komsan Suriya *Editors*

Econometrics of Risk

Springer

Studies in Computational Intelligence

Volume 583

Series editor

Janusz Kacprzyk, Polish Academy of Sciences, Warsaw, Poland
e-mail: kacprzyk@ibspan.waw.pl

About this Series

The series "Studies in Computational Intelligence" (SCI) publishes new developments and advances in the various areas of computational intelligence—quickly and with a high quality. The intent is to cover the theory, applications, and design methods of computational intelligence, as embedded in the fields of engineering, computer science, physics and life sciences, as well as the methodologies behind them. The series contains monographs, lecture notes and edited volumes in computational intelligence spanning the areas of neural networks, connectionist systems, genetic algorithms, evolutionary computation, artificial intelligence, cellular automata, self-organizing systems, soft computing, fuzzy systems, and hybrid intelligent systems. Of particular value to both the contributors and the readership are the short publication timeframe and the world-wide distribution, which enable both wide and rapid dissemination of research output.

More information about this series at http://www.springer.com/series/7092

Van-Nam Huynh · Vladik Kreinovich
Songsak Sriboonchitta · Komsan Suriya
Editors

Econometrics of Risk

Springer

Editors
Van-Nam Huynh
Japan Advanced Institute of Science and Technology
Nomi
Japan

Vladik Kreinovich
Department of Computer Science
University of Texas at El Paso
El Paso, TX
USA

Songsak Sriboonchitta
Faculty of Economics
Chiang Mai University
Chiang Mai
Thailand

Komsan Suriya
Faculty of Economics
Chiang Mai University
Chiang Mai
Thailand

ISSN 1860-949X ISSN 1860-9503 (electronic)
Studies in Computational Intelligence
ISBN 978-3-319-13448-2 ISBN 978-3-319-13449-9 (eBook)
DOI 10.1007/978-3-319-13449-9

Library of Congress Control Number: 2014956205

Springer Cham Heidelberg New York Dordrecht London

Printed on acid-free paper

Springer International Publishing AG Switzerland is part of Springer Science+Business Media (www.springer.com)

Preface

Econometrics is the application of mathematical, statistical, and computational methods to economic data. Econometrics adds empirical content to economic theory, allowing theories to be tested and used for forecasting and policy evaluation.

One of the most important aspects of economics—and one of the most difficult tasks in analyzing economic data—is how to properly take into account economic risk. Proper accounting of risks is vitally important for keeping the economy stable and prosperous.

The economic crises of the 1990s has shown that the traditional methods of risk analysis, methods based on simplified Gaussian statistical descriptions of economic phenomena and corresponding risks, are often not sufficient to adequately describe economic risks. Because of this insufficiency, new methods have been developed, in particular, methods using non-Gaussian heavy-tailed distributions, methods using non-Gaussian copulas to properly take into account dependence between different quantities, methods taking into account imprecise ("fuzzy") expert knowledge, and many other innovative techniques.

This volume contains several state-of-the-art papers devoted to econometrics of risk. Some of these papers provide further theoretical analysis of the corresponding mathematical, statistical, computational, and economical models. Several other papers describe applications of the novel risk-related econometric techniques to real-life economic situations.

We hope that this versatile volume will help practitioners to learn how to apply new techniques of econometrics of risk, and help researchers to further improve the existing models and to come up with new ideas on how to best take into account economic risks.

We want to thank all the authors for their contributions and all anonymous referees for their thorough analysis and helpful comments.

The publication of this volume is partly supported by the Chiang Mai School of Economics (CMSE), Thailand. Our thanks to Dean Pisit Leeahtam and CMSE for providing crucial support. Our special thanks to Prof. Hung T. Nguyen for his valuable advice and constant support.

We would also like to thank Prof. Janusz Kacprzyk (Series Editor) and Dr. Thomas Ditzinger (Senior Editor, Engineering/Applied Sciences) for their support and cooperation in this publication.

Nomi, Japan, January 2015 — Van-Nam Huynh
El Paso, TX, USA — Vladik Kreinovich
Chiang Mai, Thailand — Songsak Sriboonchitta
Komsan Suriya

Contents

Part I Fundamental Theory

Challenges for Panel Financial Analysis 3
Cheng Hsiao

Noncausal Autoregressive Model in Application to Bitcoin/USD Exchange Rates 17
Andrew Hencic and Christian Gouriéroux

An Overview of the Black-Scholes-Merton Model After the 2008 Credit Crisis 41
Chadd B. Hunzinger and Coenraad C.A. Labuschagne

What if We Only Have Approximate Stochastic Dominance? 53
Vladik Kreinovich, Hung T. Nguyen and Songsak Sriboonchitta

From Mean and Median Income to the Most Adequate Way of Taking Inequality into Account 63
Vladik Kreinovich, Hung T. Nguyen and Rujira Ouncharoen

Belief Aggregation in Financial Markets and the Nature of Price Fluctuations 75
Daniel Schoch

The Dynamics of Hedge Fund Performance 85
Serge Darolles, Christian Gouriéroux and Jérome Teiletche

The Joint Belief Function and Shapley Value for the Joint Cooperative Game 115
Zheng Wei, Tonghui Wang, Baokun Li and Phuong Anh Nguyen

Distortion Risk Measures Under Skew Normal Settings 135
Weizhong Tian, Tonghui Wang, Liangjian Hu and Hien D. Tran

Towards Generalizing Bayesian Statistics: A Random Fuzzy Set Approach 149
Hien D. Tran and Phuong Anh Nguyen

Local Kendall's Tau 161
P. Buthkhunthong, A. Junchuay, I. Ongeera, T. Santiwipanont and S. Sumetkijakan

Estimation and Prediction Using Belief Functions: Application to Stochastic Frontier Analysis 171
Orakanya Kanjanatarakul, Nachatchapong Kaewsompong, Songsak Sriboonchitta and Thierry Denœux

The Classifier Chain Generalized Maximum Entropy Model for Multi-label Choice Problems 185
Supanika Leurcharusmee, Jirakom Sirisrisakulchai, Songsak Sriboonchitta and Thierry Denœux

Part II Applications

Asymmetric Volatility of Local Gold Prices in Malaysia 203
Mohd Fahmi Ghazali and Hooi Hooi Lean

Quantile Regression Under Asymmetric Laplace Distribution in Capital Asset Pricing Model 219
Kittawit Autchariyapanitkul, Somsak Chanaim and Songsak Sriboonchitta

Evaluation of Portfolio Returns in Fama-French Model Using Quantile Regression Under Asymmetric Laplace Distribution 233
Kittawit Autchariyapanitkul, Somsak Chanaim and Songsak Sriboonchitta

Analysis of Branching Ratio of Telecommunication Stocks in Thailand Using Hawkes Process 245
Niwattisaiwong Seksiri and Napat Harnpornchai

Forecasting Risk and Returns: CAPM Model with Belief Functions . . . 259
Sutthiporn Piamsuwannakit and Songsak Sriboonchitta

Correlation Evaluation with Fuzzy Data and its Application in the Management Science 273
Berlin Wu, Wei-Shun Sha and Juei-Chao Chen

Empirical Evidence Linking Futures Price Movements of Biofuel Crops and Conventional Energy Fuel 287
Jianxu Liu, Songsak Sriboonchitta, Roland-Holst David, Zilberman David and Aree Wiboonpongse

Optimal Portfolio Selection Using Maximum Entropy Estimation Accounting for the Firm Specific Characteristics 305
Xue Gong and Songsak Sriboonchitta

Risk, Return and International Portfolio Analysis: Entropy and Linear Belief Functions 319
Apiwat Ayusuk and Songsak Sriboonchitta

Forecasting Inbound Tourism Demand to China Using Time Series Models and Belief Functions 329
Jiechen Tang, Songsak Sriboonchitta and Xinyu Yuan

Forecasting Tourist Arrivals to Thailand Using Belief Functions 343
Nyo Min, Jirakom Sirisrisakulchai and Songsak Sriboonchitta

Copula Based Polychotomous Choice Selectivity Model: Application to Occupational Choice and Wage Determination of Older Workers 359
Anyarat Wichian, Jirakom Sirisrisakulchai and Songsak Sriboonchitta

Estimating Oil Price Value at Risk Using Belief Functions 377
Panisara Phochanachan, Jirakom Sirisrisakulchai and Songsak Sriboonchitta

Broad Monetary Condition Index: An Indicator for Short-Run Monetary Management in Vietnam 391
Pham Thi Tuyet Trinh and Nguyen Thien Kim

Analysis of International Tourism Demand for Cambodia 415
Chantha Hor and Nalitra Thaiprasert

Modeling the Impact of Internet Broadband on e-Government Service Using Structural Equation Model 427
Sumate Pruekruedee, Komsan Suriya and Niwattisaiwong Seksiri

Assessing Sectoral Risk Through Skew-Error Capital Asset Pricing Model: Empirical Evidence from Thai Stock Market 435
Nuttanan Wichitaksorn and S.T. Boris Choy

Strategic Path to Enhance the Impact of Internet Broadband on the Creative Economy in Thailand: An Analysis with Structural Equation Model 449
Sumate Pruekruedee and Komsan Suriya

Impact of Mobile Broadband on Non-life Insurance Industry in Thailand and Singapore 457
Niwattisaiwong Seksiri and Komsan Suriya

Using Conditional Copula to Estimate Value-at-Risk in Vietnam's Foreign Exchange Market 471
Vu-Linh Nguyen and Van-Nam Huynh

The Effects of Foreign Direct Investment and Economic Development on Carbon Dioxide Emissions 483
Shu-Chen Chang and Wan-Tran Huang

Author Index 497

Part I
Fundamental Theory

Challenges for Panel Financial Analysis

Cheng Hsiao

Abstract We consider panel financial analysis from a statistical perspective. We discuss some main findings and challenges in the area of (i) estimating standard errors; (ii) joint dependence; (iii) to pool or not to pool; (iv) aggregation and predictions; (v) modeling cross-sectional dependence; and (vi) multiple-dimensional statistics.

1 Introduction

Panel data contain more degrees of freedom and more sample variability than cross-sectional or time series data. It not only provides the possibility of obtaining more accurate statistical inference, but also provides the possibility of constructing and testing more realistic behavioral hypotheses; see, e.g., [29, 30]. However, panel data also raise many methodological challenges. This paper considers some statistical issues of using panel data in finance research. We consider (i) estimation of standard errors; (ii) multiple equations modeling; (iii) to pool or not to pool; (iv) aggregation and predictions; (v) cross-sectional dependence and (vi) multi-dimensional statistics.

2 Estimation of Panel Standard Errors

Consider a single equation model often used in corporate finance or asset pricing models for N cross-sectional units observed over T time periods,

$$y_{it} = \underset{\sim}{x}_{it}'\underset{\sim}{\beta} + v_{it}, \quad i = 1, \ldots, N, \quad t = 1, \ldots, T, \tag{2.1}$$

C. Hsiao (✉)
Department of Economics, University of Southern California, Los Angeles, USA
e-mail: chowderlad@gmail.com

C. Hsiao
WISE, Xiamen University, Xiamen, China

C. Hsiao
Faculty of Economics, Chiang Mai University, Chiang Mai, Thailand

V.-N. Huynh et al. (eds.), *Econometrics of Risk*, Studies in Computational Intelligence 583, DOI 10.1007/978-3-319-13449-9_1

where

$$v_{it} = \alpha_i + \lambda_t + u_{it}, \tag{2.2}$$

α_i denotes the individual-specific effects that vary across i, but stay constant over time, λ_t denotes the time-specific effects that are individual-invariant but time-varying, and u_{it} denotes the impact of those omitted variables that vary across i and over time. t. Covariance transformation is often used to remove the impacts of α_i and λ_t; see, e.g., [29], Chap. 3. The covariance estimator of $\underset{\sim}{\beta}$ is defined as

$$\hat{\underset{\sim}{\beta}}_{cv} = \left(\sum_{i=1}^{N}\sum_{t=1}^{T}\tilde{\underset{\sim}{x}}_{it}\tilde{\underset{\sim}{x}}'_{it}\right)^{-1}\left(\sum_{i=1}^{N}\sum_{t=1}^{T}\tilde{\underset{\sim}{x}}_{it}\tilde{y}_{it}\right), \tag{2.3}$$

where

$$\tilde{\underset{\sim}{x}}_{it} = (\underset{\sim}{x}_{it} - \bar{\underset{\sim}{x}}_i - \bar{\underset{\sim}{x}}_t + \bar{\underset{\sim}{x}}), \quad y_{it} = (y_{it} - \bar{y}_i - \bar{y}_t + \bar{y}),$$

$$\bar{\underset{\sim}{x}}_i = \frac{1}{T}\sum_{t=1}^{T}\underset{\sim}{x}_{it}, \quad \bar{\underset{\sim}{x}}_t = \frac{1}{N}\sum_{i=1}^{N}\underset{\sim}{x}_{it}, \quad \bar{\underset{\sim}{x}} = \frac{1}{N}\sum_{i=1}^{N}\bar{\underset{\sim}{x}}_i = \frac{1}{T}\sum_{t}\bar{\underset{\sim}{x}}_t.$$

Statistical inference on $\underset{\sim}{\beta}$ depends on the property of u_{it}. "Although the literature has used an assortment of methods to estimate standard errors in panel data sets, the chosen method is often incorrect and the literature provides little guidance to researchers as to which method should be used. In addition, some of the advice in the literature is simply wrong." ([54], p. 436).

Vogelsang [64] showed that the covariance matrix estimate proposed in [22] based on the Newey-West [48] heteroscedastic autocorrelation (HAC) covariance matrix estimator of cross-section averages,

$$T\left(\sum_{i=1}^{N}\sum_{t=1}^{T}\tilde{\underset{\sim}{x}}_{it}\tilde{\underset{\sim}{x}}'_{it}\right)^{-1}\hat{\tilde{\Omega}}\left(\sum_{i=1}^{N}\sum_{t=1}^{T}\tilde{\underset{\sim}{x}}_{it}\tilde{\underset{\sim}{x}}'_{it}\right)^{-1}, \tag{2.4}$$

is robust to heterosceasticity, autocorrelation and spatial dependence, where

$$\hat{\tilde{\Omega}} = \frac{1}{T}\left\{\sum_{t=1}^{T}\hat{\tilde{\underset{\sim}{\nu}}}^{*}_{t}\hat{\tilde{\underset{\sim}{\nu}}}^{*'}_{t} + \sum_{j=1}^{T-1}k\left(\frac{j}{m}\right)\left[\sum_{t=j+1}^{T}\hat{\tilde{\underset{\sim}{\nu}}}^{*}_{t}\hat{\tilde{\underset{\sim}{\nu}}}^{*'}_{t-j} + \sum_{t=j+1}^{T}\hat{\tilde{\underset{\sim}{\nu}}}^{*}_{t-j}\hat{\tilde{\underset{\sim}{\nu}}}^{*'}_{t}\right]\right\},$$

$$\hat{\underset{\sim}{\nu}}^{*}_{it} = \tilde{\underset{\sim}{x}}_{it}(\tilde{y}_{it} - \tilde{\underset{\sim}{x}}'_{it}\hat{\underset{\sim}{\beta}}_{cv}), \quad \hat{\tilde{\underset{\sim}{\nu}}}^{*}_{t} = \frac{1}{N}\sum_{i=1}^{N}\tilde{\underset{\sim}{x}}_{it}(\tilde{y}_{it} - \bar{\underset{\sim}{x}}_{it}\hat{\underset{\sim}{\beta}}_{cv}) = \frac{1}{N}\sum_{i=1}^{N}\hat{\tilde{\underset{\sim}{\nu}}}^{*}_{it},$$

$k\left(\frac{j}{m}\right) = 1 - \frac{j}{m}$ if $\left|\frac{j}{m}\right| < 1$ and $k\left(\frac{j}{m}\right) = 0$ if $\left|\frac{j}{m}\right| > 1$ m an a priori chosen positive constant less than or equal to T. The choice of m depends on how strongly an investigator thinks about the serial correlation of the error u_{it}.

The Vogelsang 2012 estimator [64] of the covariance matrix of the covariance estimator, $\underset{\sim}{\beta}_{cv}$ (2.4), is consistent when errors are autocorrelated and heterocedastic provided $\underset{\sim}{x}_{it}$ is strictly exogenous. As noted by Nerlove [47], "all interesting economic behavior is inherently dynamic, dynamic panel models are the only relevant models; what might superficially appear to be a static model only conceals underlying dynamics, since any state variable presumed to influence present behavior is likely to depend in some way on past behavior." When lagged dependent variables appear in the explanatory variables to capture the inertia in human behavior, strict exogeneity of $\underset{\sim}{x}_{it}$ is violated. Not only the covariance estimator is biased if the time series dimension T is finite, no matter how large the cross-sectional dimension N is (e.g., [29], Chap. 3 [5, 24, 35]), so is the Vogelsang 2012 estimator [64] of the covariance matrix. General formulae for the estimator and its covariance matrix when the errors are autocorrelated and heterocedastic for dynamic panel model remain to be developed.

3 Multiple Equations Modeling

One of the prominent features of econometrics analysis is the incorporation of economic theory into the analysis of numerical and institutional data. Economists, from León Walras onwards, perceive the economy as a coherent system. The interdependence of sectors of an economy is represented by a set of functional relations, each representing an aspect of the economy by a group of individuals, firms, or authorities. The variables entering into these relations consist of a set of *endogenous* (or *joint dependent*) variables, whose formations are conditioning on a set of exogenous variables which the economic theory regards as given; see, e.g., [57]. Combining the joint dependence and dynamic dependence, a Cowles Commission structural equation model could be specified as,

$$\underset{\sim}{B}\underset{\sim}{y}_{it} + \underset{\sim}{\Gamma}\underset{\sim}{y}_{i,t-1} + \underset{\sim}{C}\underset{\sim}{x}_{it} = \underset{\sim}{\eta}_i + \underset{\sim}{u}_{it}, \quad i = 1, \ldots, N; \quad t = 1, \ldots, T, \tag{3.1}$$

where $\underset{\sim}{y}_{it} = (y_{1,it}, y_{2,it}, \ldots, y_{G,it})'$, $\underset{\sim}{y}_{i,t-1} = (y_{1i,t-1}, y_{2i,t-1}, \ldots, y_{Gi,t-1})'$ are $G \times 1$ contemporaneous and lagged joint dependent variables, $\underset{\sim}{x}_{it}$ is a $k \times 1$ vector of strictly exogenous variables, $\underset{\sim}{\eta}_i$ is a $G \times 1$ vector of time-invariant individual-specific effects and $\underset{\sim}{u}_{it}$ are assumed to be independently, identically distributed over i and t with zero mean and nonsingular covariance matrix Ω_u. We assume that $\underset{\sim}{y}_{i,0}$ are observed.

The distinct feature of panel dynamic simultaneous equations models are the joint dependence of $\underset{\sim}{y}_{it}$ and the presence of time persistent effects $\underset{\sim}{\eta}_i$ in the ith individual's time series observations. The joint dependence of $\underset{\sim}{y}_{it}$ makes $B \neq I_G$ and $|B| \neq 0$.

Premultiplying B^{-1} to (3.1) yields the reduced form specification

$$\underset{\sim}{y}_{it} = H_1\underset{\sim}{y}_{i,t-1} + H_2\underset{\sim}{x}_{it} + \underset{\sim}{\alpha}_i + \underset{\sim}{v}_{it}, \tag{3.2}$$

where $H_1 = -B^{-1}\Gamma$, $H_2 = -B^{-1}C$, $\underset{\sim}{\alpha}_i = B^{-1}\underset{\sim}{\eta}_i$ and $\underset{\sim}{v}_{it} = B^{-1}\underset{\sim}{u}_{it}$.

Statistical inference can only be made in terms of observed data. The joint dependence of observed variables raises the possibility that many observational equivalent

structures could generate the same observed phenomena; see, e.g., [26]. Moreover, the presence of time-invariant individual-specific effects ($\underset{\sim}{\eta}_i$ or $\underset{\sim}{\alpha}_i$) creates correlations of all current and past realized endogenous variables even $\underset{\sim}{u}_{it}$ is independently, identically distributed across i and over t with nonsingular covariance matrix Ω_u. Hsiao and Zhou [36] show that the standard Cowles Commission rank and order conditions (e.g., [28]) for the identification of (2.1) still holds provided the roots of $|B - \lambda\Gamma| = 0$ lie outside the unit circle.

When the process is stationary, both the likelihood approach and the generalized method of moments (GMM) approach can be used to make inference on (3.1) or (3.2); see, e.g., [17, 36]. The advantages of the GMM approach are that there is no need to specify the probability density function of the random variable or to worry about how to treat the initial values, y_{i0}. The disadvantages are that in many cases the GMM approach does not guarantee a global minimum and there could be huge number of moment conditions to consider, for instance, the number of moment conditions for the Arellano-Bond [10] type GMM is of order T^2. Moreover, Akashi and Kunitomo [2, 3] show that the GMM approach of estimating the structural form (3.1) is inconsistent if $\frac{T}{N} \to c \neq 0 < \infty$ as both N and T are large. Even though the GMM approach can yield consistent estimator for the reduced form model (3.2), following the approach of [2, 3, 5], Hsiao and Zhang show [35] that it is asymptotically biased of order $\sqrt{\frac{T}{N}}$ when both N and T are large. The limited Monte Carlo studies conducted by Hsiao and Zhang [35] show that whether an estimator is asymptotically biased or not plays a pivotal role in statistical inference. The distortion of the size of the test for the Arellano-Bond [10] type GMM test could be 100 % for a nominal size of 5 % if $\frac{T}{N} \to \neq 0 < \infty$.

The advantages of the likelihood approach are that a likelihood function is a natural objective function to maximize and the number of moment conditions is fixed independent of N and T. The quasi maximum likelihood estimator (QMLE) is asymptotically unbiased independent of the way N or T or both tend to infinity; see, e.g., [35, 36]. The disadvantages are that specific assumptions about the initial values $\underset{\sim}{y}_{i0}$ need to be made and specific assumptions of the data generating process of $\underset{\sim}{x}_{it}$ need to be imposed to get around the issue of incidental parameters; see, e.g., [36, 38]. When the initial distributions of $\underset{\sim}{y}_{i0}$ are misspecified, the QMLE is consistent and asymptotically unbiased only if N is fixed and $T \to \infty$. When $\frac{N}{T} \to c \neq 0 < \infty$ as $N, T \to \infty$, the QMLE is asymptotically biased of order $\sqrt{\frac{N}{T}}$.

4 To Pool or Not to Pool

Panel data, by nature, focus on individual outcomes. Factors affecting individual outcomes are numerous. Yet a model is not a mirror image of the reality, but a simplification of reality. A good model wishes to capture the essentials that affect

the outcomes while allowing the existence of unobserved heterogeneity. When a variable of interest, say y, is modeled as a function of some important factors, say the $K+m$ variables, $\underset{\sim}{w} = (\underset{\sim}{x}', \underset{\sim}{z}')$, where $\underset{\sim}{x}$ and $\underset{\sim}{z}$ are of dimension K and m, respectively,

$$y_{it} = \underset{\sim}{x}_{it}'\underset{\sim}{\beta}_i + \underset{\sim}{z}_{it}'\underset{\sim}{\gamma}_i + u_{it}, \quad i = 1, \ldots, N, \quad t = 1, \ldots, T. \tag{4.1}$$

One way to justify pooling is to test if $\underset{\sim}{\beta}_i = \bar{\underset{\sim}{\beta}}$ and $\underset{\sim}{\gamma}_i = \bar{\underset{\sim}{\gamma}}_i$ for all i. However, the homogeneity assumption is often rejected by empirical investigation; see, e.g., [40]. When $\underset{\sim}{\beta}_i$ and $\underset{\sim}{\gamma}_i$ are treated as fixed and different for each i, the only advantage of pooling is to put the model (4.1) in Zellner's [65] seemingly unrelated regression framework to improve the efficiency of the estimates of the individual behavioral equation.

One way to accommodate heterogeneity across individuals in pooling is to use a mixed random and fixed coefficient framework proposed by Hsiao, Appelbe and Dineen [37],

$$\underset{\sim}{y} = X\underset{\sim}{\beta} + Z\underset{\sim}{\gamma} + \underset{\sim}{u}, \tag{4.2}$$

where

$$\underset{NT \times 1}{\underset{\sim}{y}} = (\underset{\sim}{y}_1', \underset{\sim}{y}_2', \ldots, \underset{\sim}{y}_N'), \quad \underset{\sim}{y}_i = (y_{i1}, \ldots, y_{iT})', \quad i = 1, \ldots, N,$$

$$\underset{NT \times NK}{X} = \begin{pmatrix} X_1 & 0 & . & . & 0 \\ 0 & X_2 & 0 & . & . \\ . & . & . & . & . \\ . & . & . & . & . \\ 0 & . & . & 0 & X_N \end{pmatrix},$$

$$\underset{NT \times Nm}{Z} = \begin{pmatrix} Z_1 & 0 & . & . & 0 \\ 0 & Z_2 & . & . & . \\ . & . & . & . & . \\ . & . & . & . & . \\ 0 & . & . & . & Z_N \end{pmatrix},$$

$$\underset{T \times K}{X_i} = (\underset{\sim}{x}_{it}'), \quad \underset{T \times m}{Z_i} = (\underset{\sim}{z}_{it}')$$

$$\underset{NT \times 1}{\underset{\sim}{u}} = (\underset{\sim}{u}_1', \ldots, \underset{\sim}{u}_N')',$$

$$\underset{NK \times 1}{\underset{\sim}{\beta}} = (\underset{\sim}{\beta}_1', \ldots, \underset{\sim}{\beta}_N')', \text{ and } \underset{Nm \times 1}{\underset{\sim}{\gamma}'} = (\underset{\sim}{\gamma}_1', \ldots, \underset{\sim}{\gamma}_N')'.$$

The coefficients $\underset{\sim}{\gamma}$ are assumed fixed and different. The coefficients $\underset{\sim}{\beta}$ is assumed to be subject to stochastic constraints of the form

$$\underset{\sim}{\beta} = A\bar{\underset{\sim}{\beta}} + \underset{\sim}{\alpha}, \tag{4.3}$$

where A is an $NK \times L$ matrix with known elements, $\underset{\sim}{\bar{\beta}}$ is an $L \times 1$ vector of constants, and $\underset{\sim}{\alpha}$ is assumed to be randomly distributed with mean 0 and nonsingular covariance matrix. When $A = \underset{\sim}{\iota}_N \otimes I_K$, where $\underset{\sim}{\iota}_N$ is an $N \times 1$ vector of 1's, $\underset{\sim}{\beta}_i = \underset{\sim}{\bar{\beta}} + \underset{\sim}{\alpha}_i$, i.e., individual $\underset{\sim}{\beta}_i$ is randomly distributed with mean $\underset{\sim}{\bar{\beta}}$. The justification for (4.3) is that conditional on $\underset{\sim}{Z}'_{it}\underset{\sim}{\gamma}_i$, individual's responses towards changes in $\underset{\sim}{x}$ are similar. The difference across i is due to chance mechanism, i.e., satisfying de Finetti's [20] exchangeability criteria. Hsiao et al. [37] propose a Bayesian solution to obtain best predictors of $\underset{\sim}{\beta}$ and $\underset{\sim}{\gamma}$.

The advantage of Bayesian framework over the sampling framework to consider the issue of poolability is that all sampling tests essentially exploit the implications of a certain formulation in a specific framework; see, e.g., [15]. They are indirect in nature. The distribution of a test statistic is derived under a specific null, but the alternative is composite. The rejection of a null hypothesis does not automatically imply the acceptance of a specific alternative. It would appear more appropriate to treat the pooling issues as a model selection issue. Hsiao and Sun [32] propose to classify the conditional variables $\underset{\sim}{w}$ into $\underset{\sim}{x}$ and $\underset{\sim}{z}$ (i.e. the dimension of K and m) using some well known model selection criterion such as Akaike's information criterion [1] or Schwarz's Bayesian information criteria [58]. If $m = 0$ simple pooling is fine. If $m \neq 0$ then one can consider pooling conditioning on $Z_i\underset{\sim}{\gamma}_i$. Their limited Monte Carlo studies appear to show that combining the Bayesian framework with some model selection criterion works well in answering the question of to pool or not to pool.

5 Aggregation and Predictions

One of the tools for reducing the real world detail is through "suitable" aggregation. However, for aggregation not to distort the fundamental behavioral relations among economic agents, certain "homogeneity" conditions must hold between the micro units. Many economists have shown that if micro units are heterogeneous, aggregation can lead to very different relations among macro variables from those of the micro relations; see, e.g., [43, 44, 49, 59, 61, 63].

For instance, consider the simple dynamic equation,

$$y_{it} = \gamma_i y_{i,t-1} + \underset{\sim}{x}'_{it}\underset{\sim}{\beta}_i + \alpha_i + u_{it}, \quad |\gamma_i| < 1, \quad i = 1, \ldots, N, \tag{5.1}$$

where the error u_{it} is covariance stationary. Equation (5.1) implies a long-run relation between y_{it} and $\underset{\sim}{x}_{it}$,

$$y_{it} - \underset{\sim}{x}'_{it}\underset{\sim}{b}_i - \eta_i = v_{it} \tag{5.2}$$

where $\underset{\sim}{b}_i = (1-\gamma_i)^{-1}\underset{\sim}{\beta}_i$, $\eta_i = (1-\gamma_i)^{-1}\alpha_i$, $v_{it} = (1-\gamma_i)^{-1}u_{it}$.

Let $y_t = \sum_{i=1}^{N} y_{it}$ and $\underset{\sim}{x}_t = \sum_{i=1}^{N} \underset{\sim}{x}_{it}$, then a similar long-run relation between $\underset{\sim}{y}_t$ and $\underset{\sim}{x}_t$,

$$y_t - \underset{\sim}{x}'_t\underset{\sim}{b} - c = v_t, \tag{5.3}$$

holds for a stationary v_t if and only if either of the following conditions hold [39]:

(i) $\frac{1}{1-\gamma_i}\underset{\sim}{\beta}_i = \frac{1}{1-\gamma_j}\underset{\sim}{\beta}_j$ for all i and j; or

(ii) if $\frac{1}{1-\gamma_i}\underset{\sim}{\beta}_i \neq \frac{1}{1-\gamma_j}\underset{\sim}{\beta}_j$, then $\underset{\sim}{x}'_t = (\underset{\sim}{x}'_{1t}, \ldots, \underset{\sim}{x}'_{Nt})$ must lie on the null space of D for all t, where $D' = \left(\frac{1}{1-\gamma_1}\underset{\sim}{\beta}'_1 - \underset{\sim}{b}', \ldots, \frac{1}{1-\gamma_N}\underset{\sim}{\beta}'_N - \underset{\sim}{b}'\right)$.

These conditions are fairly restrictive. If "heterogeneity" is indeed present in micro units, then shall we predict the aggregate outcome based on the summation of estimated micro relations or shall we predict the aggregate outcomes based on the estimated aggregate relations? Unfortunately, there is not much work on this specific issue. In choosing between whether to predict aggregate variables using aggregate (H_a) or disaggregate equations (H_d), Grunfeld and Griliches [23] suggest using the criterion of:

$$\text{Choose } H_d \text{ if } \underset{\sim}{e}'_e\underset{\sim}{e}_d < \underset{\sim}{e}'_a\underset{\sim}{e}_a, \text{ otherwise choose } H_a \tag{5.4}$$

where $\underset{\sim}{e}_d$ and $\underset{\sim}{e}_a$ are the estimates of the errors in predicting aggregate outcomes under H_d and H_a, respectively. The Grunfeld and Griliches criterion is equivalent to using simple average of micro-unit prediction to generate aggregate prediction if (5.4) holds. As discussed by Hsiao and Wan [34] that if cross-sectional units are not independent, there are many other combination approaches that could yield better aggregate forecasts, such as Bates and Granger regression approach [16], Buckland et al. Bayesian averaging [18], Hsiao and Wan eigenvector approach [34], Swanson and Zeng information combination [60], etc. (for a survey of forecast combinations, see [62]). However, if a model is only a local approximation, then frequent structural breaks could occur from a model's perspective even there is no break in the underlying structure. In this situation, it is not clear there exists an optimal combination of micro forecasts. Perhaps, "robustness" is a more relevant criterion than "optimality"; see, e.g., [53].

6 Cross-Sectional Dependence

Most panel inference procedures assume that apart from the possible presence of individual invariant but period varying time-specific effects, the effects of omitted variables are independently distributed across cross-sectional units. Often economic theory predicts that agents take actions that lead to interdependence among themselves. For example, the prediction that risk averse agents will make insurance contracts allowing them to smooth idiosyncratic shocks implies dependence in consumption across individuals. Contagion of views could also lead to herding or imitating behavior; see, e.g., [4]. Cross-sectional units could also be affected by common omitted factors. The presence of cross-sectional dependence can substantially complicate statistical inference for a panel data model.

Ignoring cross-sectional dependence in panel data could lead to seriously misleading inference; see, e.g., [33, 56]. However, modeling cross-sectional dependence is a lot more complicated than modeling serial dependence. There is a natural order of how a variable evolves over time. Cross-sectional index is arbitrary. There is no natural ordering. Three popular approaches for taking account the cross-sectional dependence are: spatial approach; see, e.g., [8, 9, 41, 42], factor approach (e.g., [11, 12]), and cross-sectional mean augment approach (e.g., [50, 53]). The spatial approach assumes that there exists a known $N \times N$ spatial weight matrix W, where the i, jth element of W, w_{ij}, gives the strength of the interaction between the ith and jth cross-sectional units. The conventional specification assumes that the diagonal elements, $w_{ii} = 0$ and $\sum_{j=1}^{N} w_{ij} = 1$ through the row normalization. The only term unknown is the absolute strength, ρ. However, to ensure the interaction between the ith and jth unit has a "decaying" effect among cross-sectional units where the "distance" between them increases, ρ is assumed to have absolute value less than 1. Apart from the fact that it is difficult to have prior information to specify w_{ij}, it also raises the issue of relations between observed sample and the population. If N is not the population size, the restriction that $\sum_{j=1}^{N} w_{ij} = 1$ and $|\rho| < 1$ implies that as N increases, each element of $w_{ij} \to 0$.

Another approach to model cross-sectional dependence is to assume that the variable or error follows a linear factor model,

$$v_{it} = \sum_{j=1}^{r} b_{ij} f_{jt} + u_{it} = \underset{\sim}{b}_i' \underset{\sim}{f}_t + u_{it}, \tag{6.1}$$

where $\underset{\sim}{f}_t = (f_{1t}, \ldots, f_{rt})'$ is a $r \times 1$ vector of random factors with mean zero, $\underset{\sim}{b}_i = (b_{i1}, \ldots, b_{ir})'$, is a $r \times 1$ nonrandom factor loading coefficients, u_{it} represents the effects of idiosyncratic shocks which is independent of $\underset{\sim}{f}_t$ and is independently distributed across i with diagonal covariance matrix D.

An advantage of factor model over the spatial approach is that there is no need to prespecify the strength of correlations between units i and j. The disadvantage is that when no restriction is imposed on the factor loading matrix $B = (\underset{\sim}{b}_{ij}')$, it implies strong cross-sectional dependence [19]. Unless B is known, there is no way to find transformation to control the impact of cross-sectional dependence on statistical inference. Bai [11] has proposed methods to estimate a model with factor structure error term. However, most financial data contains large number of cross-sectional units. When N is large, the estimation of the factor loading matrix, B, is not computational feasible.

Instead of estimating $\underset{\sim}{f}_t$ and $\underset{\sim}{b}$, Pesaran suggests [50] a simple approach to filter out the cross-sectional dependence by augmenting the cross-sectional mean of observed data to a model. For instance, Pesaran, Schuermann and Weiner propose [53] a global vector autoregressive model (VAR) for an $m \times 1$ dimensional random variables, $\underset{\sim}{w}_{it}$ to accommodate dynamic cross-dependence by considering

$$\Phi_i(L)(\underset{\sim}{w}_{it} - \Gamma_i \underset{\sim}{w}_{it}^*) = \underset{\sim}{\varepsilon}_{it}, \quad i = 1, 2, \ldots, N, \tag{6.2}$$

where $\Phi_i(L) = I + \Phi_{1i}L + \cdots + \Phi_{pi}L^{p_i}$, and L denotes the lag operator,

$$\underset{\sim}{w}^*_{it} = \sum_{j=1}^{N} r_{ij}\underset{\sim}{w}_{jt}, \tag{6.3}$$

$$r_{ii} = 0, \quad \sum_{j=1}^{N} r_{ij} = 1, \text{ and } \sum_{j=1}^{N} r_{ij}^2 \to 0 \text{ as } N \to \infty. \tag{6.4}$$

The weight r_{ij} could be $\dfrac{1}{N-1}$ for $i \neq j$, or constructed from trade value or other measures of some economic distance and could be time-varying. The global average $\underset{\sim}{w}^*_{it}$ is inserted into the individual i's VAR model,

$$\Phi_i(L)\underset{\sim}{w}_{it} = \underset{\sim}{y}_{it} \tag{6.5}$$

to take account of the cross-sectional dependence. When $\underset{\sim}{w}^*_{i,t-j}$ can be treated as weakly exogenous (predetermined), the estimation of each i can proceed using standard time series estimation techniques; see, e.g., [54]. Pesaran et al. [53] show that the weak exogeneity assumption of $\underset{\sim}{w}^*_{it}$ hold for all countries except for the U.S. because of U.S.'s dominate position in the world. They also show that (6.2) yields better results than (6.5) when cross-sectional units are correlated.

The advantage of Pesaran's 2006 cross-sectional mean-augmented approach [50] to take account the cross-sectional dependence is its simplicity. However, there are restrictions on its application. The method works when $\underset{\sim}{b}_i'\underset{\sim}{f}_t = c_i\bar{\underset{\sim}{b}}'\underset{\sim}{f}_t$ for all t or if $\underset{\sim}{f}_t$ can be considered as a linear combinations of $\bar{y}_t$ and $\bar{\underset{\sim}{x}}_t$. It is hard to ensure $\underset{\sim}{b}_i'\underset{\sim}{f}_t$ if $r > 1$. For instance, consider the case that $r = 2, \underset{\sim}{b}_i' = (1, 1), \bar{\underset{\sim}{b}}' = (2, 0)$, $\underset{\sim}{f}_t' = (1, 1)$, then $\underset{\sim}{b}_i'\underset{\sim}{f}_t - \bar{\underset{\sim}{b}}'\underset{\sim}{f}_t = 2$. However, if $\underset{\sim}{f}_s' = (2, 0)$, then $\underset{\sim}{b}_i'\underset{\sim}{f}_s = 2$ while $\bar{\underset{\sim}{b}}'\underset{\sim}{f}_s = 4$. If $\underset{\sim}{b}_i'\underset{\sim}{f}_t = c_{it}\bar{\underset{\sim}{b}}'\underset{\sim}{f}_t$, cross-sectional mean $\underset{\sim}{c}_i'\bar{\underset{\sim}{w}}_t$ does not approximate (6.1). Additional conditions are need to approximate $\underset{\sim}{b}_i'\underset{\sim}{f}_t$.

7 Multi-dimensional Statistics

Panel data is multi-dimensional. Phillips and Moon [55] have shown that the multi-dimensional asymptotics is a lot more complicated than one-dimensional asymptotics. Financial data typically have cross-sectional and time-series dimension increase at the same rate or some arbitrary rate. Moreover, computing speed and storage capability have enabled researchers to collect, store and analyze data sets of very high dimensions. Multi-dimensional panel will become more available. Classical asymptotic theorems under the assumption that the dimension of data is fixed (e.g., [7]) appear to be inadequate to analyze issues arising from finite sample of very high dimensional data; see, e.g., [14]. For example, Bai and Saranadasa [13] proved that

when testing the difference of means of two high dimensional populations, Dempster's 1958 non-exact test [21] is more powerful than Hotelling's 1931 T^2-test [27] even though the latter is well defined. Another example is in the regression analysis economists sometimes consider optimal ways to combine a set of explanatory variables to capture their essential variations as a dimension reduction method when the degrees of freedom are limited (e.g., [6]) or to combine a number of independent forecasts to generate a more accurate forecast; see, e.g., [62]. The former leads to principal component analysis that chooses the combination weights as the eigenvectors corresponding to the largest eigenvalues of the covariance matrix of the set of variables in question. The latter leads to choosing the combination weights proportional to the eigenvector corresponding to the smallest eigenvalue of the prediction mean square error matrix of the set of independent forecasts [31]. However, the true covariance matrix is unknown. Economists have to use the finite sample estimated covariance matrix (or mean square error matrix) in lieu of the true one. Unfortunately, when the dimension of the matrix (p) relative to the available sample (n) is large, $\frac{p}{n} = c \neq 0$, the sample estimates can be very different from the true ones and whose eigenvectors may point in a random direction [46]; for an example, see [31]. Many interesting and important issues providing insight to finite and large sample issues for high dimensional data analysis remain to be worked out and can be very useful to economists and/or social scientists; see, e.g., [14].

8 Concluding Remarks

Panel data contain many advantages (but they also raise many methodological issues; see, e.g., [29, 30]. This paper attempts to provide a selective summary of what have been achieved and challenging issues confronting panel financial analysis. In choosing an appropriate statistical method to analyze the panel financial data on hand, it is helpful to keep several factors in mind. First, what advantages do panel data offer us in adapting economic theory for empirical investigation over data sets consisting of a single cross-section or time series? Second, what are the limitations of panel data and the econometric methods that have been proposed for analyzing such data. Third, the usefulness of panel data in providing particular answers to certain issues depends critically on the compatibility between the assumptions underlying the statistical inference procedures and the data generating process. Fourth, when using panel data, how can we increase the efficiency of parameter estimates? "Analyzing economic data (or financial data) requires skills of synthesis, interpretation and empirical imagination. Command of statistical methods is only a part, and sometimes a very small part, of what is required to do a first-class empirical research" [25]. Panel data are no panacea. Nevertheless, if "panel data are only a little window that opens upon a great world, they are nevertheless the best window in econometrics" [45].

Acknowledgments This work is supported in part by the China National Science Foundation grant #71131008. I would like to thank a referee for helpful comments.

References

1. Akaike, H.: Information theory and an extension of the maximum likelihood principle. In: Petrov, B.N., Csaki, F. (eds.) Proceedings of the 2nd International Symposium Information Theory, pp. 267–281. Akademiai Kiado, Budapest (1973)
2. Akashi, K., Kunitomo, N.: Some properties of the LIML estimator in a dynamic panel structural equation. J. Econom. **166**, 167–183 (2012)
3. Akashi, K., Kunitomo, N.: The Limited Information Maximum Likelihood Approach to Dynamic Structural Equation Models. Annals of the Institute of Statistical Mathematics (forthcoming) (2014)
4. Akerlof, G.A., Shiller, R.J.: Animal Spirits: How Human Psychology Drives the Economy, and Why It Matters for Global Capitalism. Princeton University Press, Princeton (2009)
5. Alvarez, J., Arellano, M.: The time series and cross-section asymptotics of dynamic panel data estimators. Econometrica **71**, 1121–1159 (2003)
6. Amemiya, T.: On the use of principal components of independent variables in two-stage least-squares estimation. Int. Econ. Rev. **7**, 283–303 (1966)
7. Anderson, T.W.: An Introduction to Multivariate Analysis, 2nd edn. Wiley, New York (1985)
8. Anselin, L.: Spatial Econometrics: Methods and Models. Kluwer Academic, Dordrecht (1988)
9. Anselin, L., Le Gallo, J., Jayet, H.: Spatial panel econometrics. In: Mátyás, L., Sevestre, P. (eds.) The Econometrics of Panel Data, 3rd edn, pp. 625–660. Springer, Berlin (2008)
10. Arellano, M., Bond, S.: Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations. Rev. Econ. Stud. **58**, 277–297 (1991)
11. Bai, J.: Panel data models with interactive fixed effects. Econometrica **77**, 1229–1279 (2009)
12. Bai, J., Ng, S.: Determining the number of factors in approximate factor models. Econometrica **70**, 191–221 (2002)
13. Bai, Z.D., Saranadasa, H.: Effect of high dimension: by an example of a two sample problem. Stat. Sin. **6**, 311–329 (1996)
14. Bai, Z.D., Silverstein, J.W.: CLT of linear spectral statistics of large-dimensional sample covariance matrices. Ann. Probab. **32**(1A), 553–605 (2004)
15. Baltagi, B.H., Bresson, G., Pirotte, A.: To Pool or Not to Pool? In: Matyas, L., Sevestre, P. (eds.) The Econometrics of Panel Data, pp. 517–546. Springer, Berlin (2008)
16. Bates, J.M., Granger, C.M.W.: The combination of forecasts. Oper. Res. Q. **20**, 451–468 (1969)
17. Binder, M., Hsiao, C., Pesaran, M.H.: Estimation and inference in short panel vector autoregressions with unit roots and cointegration. Econom. Theory **21**(4), 795–837 (2005)
18. Buckland, S.T., Burnhamn, K.P., Augustin, N.H.: Model selection: an integral part of inference. Biometrics **53**, 603–618 (1997)
19. Chudik, A., Pesaran, M.H., Tosetti, E.: Weak and strong cross-section dependence and estimation of large panels. Econom. J. **14**(1), C45C90 (2011)
20. de Finetti, B.: Foresight: its logical laws, its subjective sources. In: Kyberg Jr, H.E., Smokler, H.E. (eds.) Studies in Subjective Probability, pp. 93–158. Wiley, New York (1964)
21. Dempster, A.P.: A high dimensional two sample significance test. Ann. Math. Stat. **29**(4), 995–1010 (1958)
22. Driscoll, J.C., Kraay, A.C.: Consistent covariance matrix estimation with spatially dependent panel data. Rev. Econ. Stat. **80**(4), 549–560 (1998)
23. Grunfeld, Y., Griliches, Z.: Is aggregation necessarily bad? Rev. Econ. Stat. **42**, 1–13 (1960)
24. Hahn, J., Kuersteiner, G.: Asymptotically unbiased inference for a dynamic panel model with fixed effects when both N and T are large. Econometrica **70**(4), 1639–1657 (2002)
25. Heckman, J.J.: Econometrics and empirical economics. J. Econom. **100**, 3–6 (2001)
26. Hood, W.C., Koopmans, T.C. (eds.): Studies in Econometric Method. Wiley, New York (1953)
27. Hotelling, H.: The economics of exhaustible resources. J. Polit. Econ. **39**(2), 137–175 (1931)
28. Hsiao, C.: Autoregressive modeling and causal ordering of economic variables. J. Econ. Dyn. Control **4**(1), 243–259 (1982)
29. Hsiao, C.: Panel Data Analysis, 2nd edn. Cambridge University Press, Cambridge (2003)

30. Hsiao, C.: Panel data analysis-advantages and challenges. TEST **16**, 1–22 (2007)
31. Hsiao, C.: The creative tension between statistics and econometrics. Singap. Econ. Rev. **57**, 125007-1–125007-11 (2012)
32. Hsiao, C., Sun, B.H.: To pool or not to pool panel data. In: Krishnakumar, J., Ronchetti, E. (eds.) Panel Data Econometrics: Future Directions, Papers in Honor of Professor Pietro Balestra, pp. 181–198. North Holland, Amsterdam (2000)
33. Hsiao, C., Tahmiscioglu, A.K.: Estimation of dynamic panel data models with both individual and time specific effects. J. Stat. Plan. Inference **138**, 2698–2721 (2009)
34. Hsiao, C., Wan, S.K.: Is there an optimal forecast combination? J. Econom. **178**, 294–309 (2014)
35. Hsiao, C., Zhang, J.: IV, GMM or MLE to Estimate Dynamic Panel Data Models when both N and T are Large. Mimeo, University of Southern California (2013)
36. Hsiao, C., Zhou, Q.: Statistical inference for panel dynamic simultaneous equations models. J. Econom. (forthcoming) (2014)
37. Hsiao, C., Appelbe, T.W., Dineen, C.R.: A general framework for panel data analysis—with an application to Canadian customer dialed long distance service. J. Econom. **59**, 63–86 (1993)
38. Hsiao, C., Pesaran, M.H., Tahmiscioglu, A.K.: Maximum likelihood estimation of fixed effects dynamic panel data models covering short time periods. J. Econom. **109**, 107–150 (2002)
39. Hsiao, C., Shen, Y., Fujiki, H.: Aggregate versus disaggregate data analysis—a paradox in the estimation of money demand function of Japan under the low interest rate policy. J. Appl. Econom. **20**, 579–601 (2005)
40. Kuh, E.: Capital Stock Growth: A Micro-Econometric Approach. North-Holland, Amsterdam (1963)
41. Lee, L.F., Yu, J.: Estimation of spatial autoregressive panel data models with fixed effects. J. Econom. **154**, 165–185 (2010)
42. Lee, L.F., Yu, J.: Some recent developments in spatial panel data models. Reg. Sci. Urban Econ **40**, 255–271 (2010)
43. Lewbel, A.: Aggregation and with log linear models. Rev. Econ. Stud. **59**, 535–554 (1992)
44. Lewbel, A.: Aggregation and simple dynamics. Am. Econ. Rev. **84**, 905–918 (1994)
45. Mairesse, J.: Comment on panel data analysis—advantages and challenges. TEST **16**, 37–41 (2007)
46. Nadler, B.: Finite sample approximation results for principal component analysis, a matrix perturbation approach. Ann. Stat. **36**, 2791–2817 (2008)
47. Nerlove, M.: An Essay on the History of Panel Data Econometrics, paper presented at 2000 Panel Data Conference in Geneva (2000)
48. Newey, W.K., West, K.D.: A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. Econometrica **55**(3), 703–708 (1987)
49. Pesaran, M.H.: Estimation and Inference in Large Heterogeneous Panels with Cross Section Dependence, Cambridge Working Papers in Economics 0305, Faculty of Economics, University of Cambridge (2003)
50. Pesaran, M.H.: Estimation and inference in large heterogeneous panels with cross-section dependence. Econometrica **74**, 967–1012 (2006)
51. Pesaran, M.H., Smith, R.: Estimation of long-run relationships from dynamic heterogeneous panels. J. Econom. **68**, 79–114 (1995)
52. Pesaran, M.H., Tosetti, E.: Large panels with common factors and spatial correlations. J. Econom. **161**, 182–202 (2010)
53. Pesaran, M.H., Schuermann, T., Weiner, S.M.: Modelling regional interdependencies using a global error-correction macroeconometrics model. J. Bus. Econ. Stat. **22**, 129–162 (2004)
54. Peterson, M.A.: Estimating standard errors in finance panel data sets: comparing approaches. Rev. Financ. Stud. **22**(1), 435–480 (2009)
55. Phillips, P.C.B., Moon, H.E.: Linear regression limit theory for nonstationary panel data. Econometica **67**(5), 1057–1111 (1999)
56. Phillips, P.C.B., Sul, D.: Bias in dynamic panel estimation with cross-sectional dependence. Econom. Theory (2007)

57. Roberts, M., Whited, T.M.: Endogeneity in Empirical Corporate Finance. Simon School working paper no. FR11-29 (2012)
58. Schwarz, G.E.: Estimating the dimension of a model. Ann. Stat. **6**(2), 461464 (1978)
59. Stoker, T.M.: Empirical approaches to the problem of aggregation over individuals. J. Econ. Lit. **31**, 1827–1874 (1993)
60. Swanson, N.R., Zeng, T.: Choosing among competing econometric forecasts: regression-based forecast combination using model selection. J. Forecast. **20**, 425–440 (2001)
61. Theil, H.: Linear Aggregation of Economic Relations. North-Holland, Amsterdam (1954)
62. Timmermann, A.: Forecast combinations. In: Elliott, G., Granger, C.W.J., Timmermann, A. (eds.) Handbook of Economic Forecasting, vol. 1. Elsevier, Amsterdam (2006)
63. Trivedi, P.K.: Distributed lags, aggregation and compounding: some econometric implications. Rev. Econ. Stud. **52**, 19–35 (1985)
64. Vogelsang, T.: Heteroscedasticity, autocorrelation and spatial correlation robust inference in linear panel models with fixed effects. J. Econom. **166**, 303–319 (2012)
65. Zellner, A.: An efficient method of estimating seemingly unrelated regressions and tests for aggregation bias. J. Am. Stat. Assoc. **57**, 348–368 (1962)

Noncausal Autoregressive Model in Application to Bitcoin/USD Exchange Rates

Andrew Hencic and Christian Gouriéroux

Abstract This paper introduces a noncausal autoregressive process with Cauchy errors in application to the exchange rates of the Bitcoin electronic currency against the US Dollar. The dynamics of the daily Bitcoin/USD exchange rate series displays episodes of local trends, which can be modelled and interpreted as speculative bubbles. The bubbles may result from the speculative component in the on-line trading. The Bitcoin/USD exchange rates are modelled and predicted.

JEL number: C14 · G32 · G23

1 Introduction

In recent months digital currencies (sometimes referred to as crypto-currencies) and their standard bearer, Bitcoin, have been garnering more public attention (see [36]). This can likely be attributed to two factors. Public adoption of the digital currency is beginning to become more commonplace (see [26]) and its more nefarious uses are slowly being exposed (see [19]).

A prime example of the first point is the University of Nicosia in Cyprus. The University is the largest private university in Cyprus and is beginning to accept bitcoins as tuition payment. The university's reasoning is that they wish to be at the forefront of global commerce, but there may be other reasons at play. More recently Cyprus has gone through significant financial stress and many of the country's depositors will likely face significant losses (see [35]). Mistrust of the established financial system may lead institutions to begin accepting alternative means of payment.

A. Hencic
York University, Department of Economics, 4700 Keele St, Toronto, ON M3J 1P3, Canada
e-mail: ahencic@econ.yorku.ca

C. Gourieroux (✉)
University of Toronto and CREST, Max Gluskin House, 150 St. George Street,
Toronto, ON M5S 3G7, Canada
e-mail: gouriero@ensae.fr

V.-N. Huynh et al. (eds.), *Econometrics of Risk*, Studies in Computational Intelligence 583, DOI 10.1007/978-3-319-13449-9_2

As for the nefarious uses of bitcoins, the most recent story about the raid on the website The Silk Road can speak to the dark side of digital and anonymous currency. In October of 2013 the FBI shut down The Silk Road for allegedly selling illegal drugs and charged its owner with a whole host of offenses. Critics of digital currencies say that the anonymity provided to their users is dangerous and should be further regulated. The government of the United States has responded to these concerns by implementing rules to attempt to curb the use of digital currencies in money laundering (see [34]). With the market capitalization of bitcoin surpassing $12 Billion USD (see [4]), and its ever increasing adoption, further study of the uses, threats and mechanisms that govern digital currencies is needed.

The objective of this paper is to examine the dynamics of the Bitcoin/USD exchange rate and to predict its future evolution. The dynamics of the series are characterized by the presence of local trends and short-lived episodes of soaring Bitcoin/USD rates, followed by sudden almost vertical declines. These patterns are referred to as bubbles. In economics, bubbles in asset prices have been introduced in the context of the rational expectation hypothesis in the seventies and as a result of the speculative behavior of traders. The bubbles in the Bitcoin/USD rate may originate from (a) the fact that the bitcoin market is still an emerging market with a lot of speculative trading, (b) the asymmetric information and crowd phenomena (see e.g. [12] for the analogous on Nasdaq), (c) the lack of a centralized management and control of exchange rate volatility, (d) the deterministic supply of bitcoins and the evolution of the volume over time. As the volume of bitcoins available on the market is exogenously determined, this enhances the bitcoin price and exchange rate volatility.

Because of the presence of local explosive trends, depicted as bubbles, the Bitcoin/USD exchange rate cannot be modelled by any traditional ARIMA or ARCH models (see e.g. [1]). In this paper, we use the mixed causal-noncausal autoregressive process with Cauchy errors [21, 22] to estimate and predict the Bitcoin/USD exchange rate.

The structure of the paper is as follows. In Sect. 2, we describe the bitcoin as an electronic currency, and we explain the mechanisms of bitcoin trading and storage. Next, we describe the data and the period of interest that includes a bubble burst and crash. A speculative bubble is a nonlinear dynamic feature that can be accommodated by the aforementioned noncausal autoregressive process. In Sect. 3, we review the properties of noncausal processes and introduce the associated inference and prediction methods. The application to the Bitcoin/ US Dollar exchange rates recorded on the Mt. Gox[1] exchange market is presented in Sect. 4. The noncausal model is used to predict the occurrence of the bubble in the Bitcoin/USD exchange rate. Section 5 concludes the paper.

[1] Formerly magic: the gathering online exchange.

2 The Bitcoin/USD Exchange Rate

2.1 Bitcoin Currency

Bitcoin (BTC) is an electronic currency originally created by a developer under the pseudonym Satoshi Nakamoto in 2009 (see [14]). The electronic currency is distributed on a peer-to-peer network anonymously between any two accounts. There is no formal denomination or name for units of the currency other than 1.00 BTC being referred to as a bitcoin and the smallest possible denomination, 10^{-8} BTC, being a "satoshi".

The bitcoin can be purchased on a virtual exchange market, such as mtgox.com against the US Dollar or other currencies.[2] Users of the currency store it on a private digital "wallet". This wallet has no personal identification with an individual and is comprised of three components: an address, a private key, and a public key. There is nothing that connects a wallet to an individual. This level of anonymity has been one of the driving forces behind the currency's popularity. The bitcoin can be used to purchase a number of goods and services that are listed on the Bitcoin website.

Three types of wallets exist: the software wallet, the mobile wallet and the web wallet. Software wallets are installed directly on a computer and allow the user complete control over the wallet. Mobile wallets are installed on mobile devices and operate the same way. Web wallets host an individual's bitcoins online. All of these wallets can be accessed with just the private key assigned to the address. Again, there is nothing to associate a physical human being with a Bitcoin address other than if the person owns the hardware on which the wallet is installed.

As of December 2, 2013 the total market capitalization of bitcoin is approximately $12 billion USD (see [9]). Bitcoin is traded 24 h a day on various exchanges, the largest of which include Mt. Gox (based in Japan)[3] and BTC China (recently the world's largest BTC exchange [28]). The former is a real time exchange whereas BTC China is a fixed rate exchange (see [8]). Bitcoins are denominated in USD on Mt. Gox and in Renminbi on BTC China. After a clarification by the People's Bank of China on Bitcoin's status at the beginning of December 2013, the exchanges on BTC China can only be done in Chinese Yuan, and the users have to now provide their identity using, for example, a passport number. Trading and use of Bitcoin is forbidden in Thailand.

The trading volume of bitcoin on Mt. Gox has slowly increased over time as adoption of the currency has increased. On its first day of trading on Mt. Gox, the total volume of bitcoin traded was 20 units. Obviously this is a very small number in comparison to the 3,436,900 bitcoins in circulation at that time. However, trading volume has gradually increased since then as Bitcoin has become more generally accepted and garnered more attention. Trading volume reached an all-time high on

[2] The transactions on this market have been suspended as of February 25, 2014. The reason is yet to be revealed, but an attack by hackers has been declared.

[3] It represented 12 % of the trades before it collapsed.

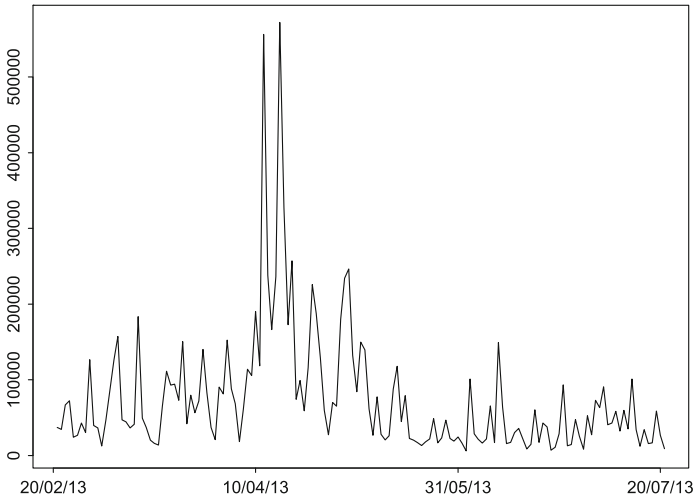


Fig. 1 Bitcoin volume, Feb–July 2013

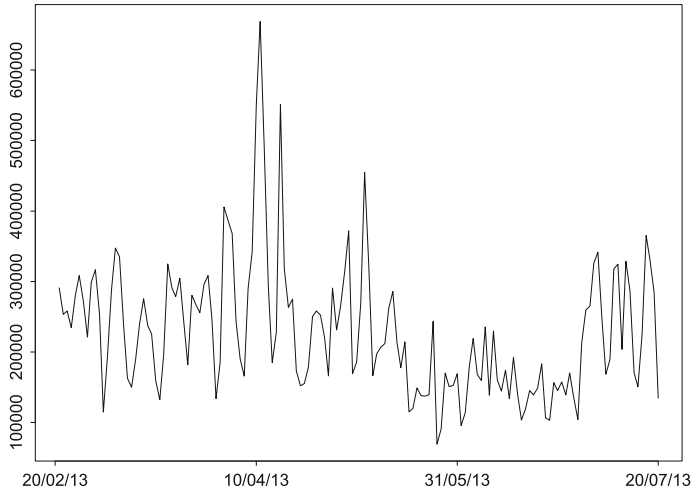


Fig. 2 Bitcoin transactions, Feb–July 2013

April 15, 2013 with 572,185.7 bitcoins changing hands on Mt. Gox. At the time there were approximately 11,027,700 units in existence, meaning that on this day approximately 5 % of all bitcoins in circulation were traded on Mt. Gox.

The long term supply of BTC will never exceed 21,000,000 units. However, the daily volume traded on the platforms can be much smaller. The traded volume was 31,800 BTC on Mt. Gox on December 8, 2013. The evolution of the traded volume of bitcoins between February and July 2013 is displayed in Figs. 1 and 2.

Figure 1 provides the daily volume exchanged against USD while Fig. 2 provides the daily volume of bitcoins used for real transactions that is for the sale and purchase of goods and services offered in bitcoin. The daily volumes are small compared to the capitalization of bitcoin, showing that this emerging market may encounter liquidity problems.

Bitcoins are produced in such a way that the volume of new bitcoins produced will be halved every four years until the volume of new coins produced decays to zero. At this point the final supply of bitcoins will be fixed (the exact amount of units varies depending on rounding, but it will be less than 21 million units) (see [7]). Bitcoins are produced in a process referred to as "mining". Computers on the Bitcoin network solve complex mathematical problems and are rewarded for their work with a predetermined amount of bitcoins, referred to as a "block reward", and a transaction fee. The current block reward is 25 bitcoins (see [6]). In order to control the supply of bitcoins being produced the difficulty of these problems is automatically adjusted so that the time between solutions averages 10 min.

2.2 Bitcoin Transactions

To ensure the security of transactions, the Bitcoin system uses public key cryptography. Each individual has one or more addresses with an associated private and public key. The system is totally anonymous and balances are only associated with an address and its keys. Only the user with the private key can sign a transfer of bitcoins to another party, whereas anybody in the network can validate the signature and transaction using the user's public key (see [6]). When a transaction occurs, one user sends another an amount of bitcoins and signs the transaction with their private key. The user who sends the bitcoins announces a public key and it falls on the network to verify the signature. The user then broadcasts the transaction on the Bitcoin network. In order to prevent double spending the details about a transaction are sent to as many other computers on the network as possible in a block. Each computer on this network has a registry of these blocks called a "block chain". In order for the newest block to be accepted into the chain, it must be valid and must include proof of work (the solution to the aforementioned math problem). When a block is announced the miners work to verify the transaction by solving the math problem. When a solution is reached it is verified by the rest of the network. This allows for the tracking of the life of every individual bitcoin produced.

Thus for any individual to double spend their Bitcoins, their computing power would have to exceed the combined computing power of all other Bitcoin computers.

Alternatives to Bitcoin have already begun to spring up. The largest competitor is Litecoin, which as of December 2, 2013 has a market capitalization of $695,376,891 USD [8]. Litecoin seeks to be an improvement over Bitcoin by attempting to overcome some of the more technical issues facing Bitcoin.

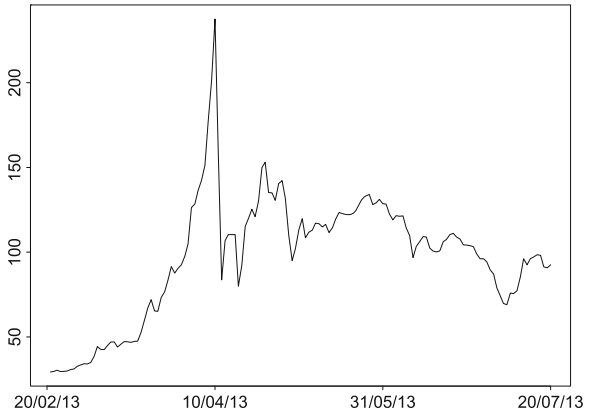


Fig. 3 Bitcoin/USD exchange rate, Feb–July 2013

2.3 The Data

In our empirical study, we consider the Bitcoin/USD exchange rate from the first part of year 2013 that includes a bubble, which bursted on April 10, 2013.

More specifically, the sample consists of 150 observations on the daily closing values of the Bitcoin/USD exchange rate over the period February 20–July 20, 2013. The dynamics of the data is displayed in Fig. 3. We observe a nonlinear trend as well as the bubble that peaked at the virtual time t = 50. The sample median, interquartile range and total range[4] are 103.27, 46.69 and 208.21, respectively. For comparison,

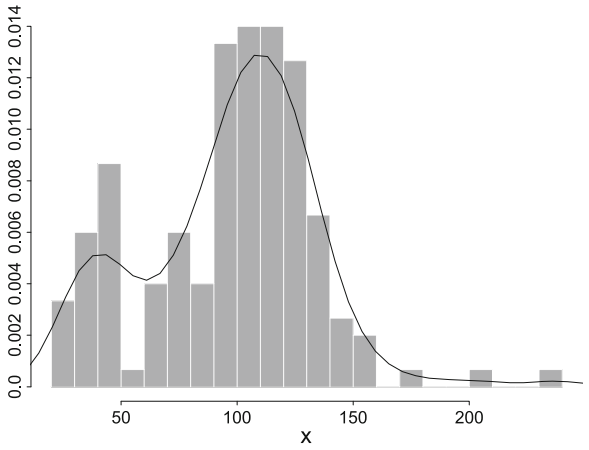


Fig. 4 Bitcoin histogram

[4] The difference between the sample max and min.

the sample mean and variance are 96.98 and 1327.63, respectively. These standard summary statistics can be misleading, since their usual interpretation assumes the stationarity of the process. This assumption is clearly not satisfied for the Bitcoin/USD exchange rate. Figure 4 shows the histogram and a kernel-based density estimate of the sample marginal density.

Both estimates display fat tails, as suggested by the fact that the total range is five times greater than the interquartile range. Also, the histogram indicates a discontinuity in the left tail, which shows as an almost bimodal pattern in the kernel-smoothed density estimate.

3 The Model

This section presents the mixed causal-noncausal autoregressive process, explains how this process accommodates the bubble effects observed in the Bitcoin/USD exchange rate series and discusses the estimation and inference.

3.1 The Noncausal and Mixed Autoregressive Process

A mixed (causal-noncausal) autoregressive process is a stochastic process $\{y_t;\ t = 0, \pm 1, \pm 2, \dots\}$, defined by:

$$\Psi(L^{-1})\Phi(L)y_t = e_t, \tag{1}$$

where $\Psi(L^{-1})$ and $\Phi(L)$ are polynomials in the negative (resp. positive) powers of the lag operator L, such that $\Psi(L^{-1}) = 1 - \psi_1 L^{-1} - \cdots - \psi_s L^{-s}$ and $\Phi(L) = 1 - \phi_1 L - \cdots - \phi_r L^r$. The roots of both polynomials are assumed to lie outside the unit circle, and error terms e_t are identically and independently distributed. When $\phi_1 = \cdots = \phi_r = 0$, model (1) defines a pure noncausal autoregressive process of order s, while for $\psi_1 = \cdots = \psi_s = 0$, the process y_t is the traditional pure causal AR(r) process. When some of the coefficients of both polynomials are non-zero, we obtain a mixed process that contains both the lags and leads of y_t. Under the above assumptions, there exists a unique stationary solution to Eq. (1). This solution admits a strong, two-sided moving average representation:

$$y_t = \sum_{j=-\infty}^{\infty} \xi_j e_{t-j},$$

where the ξ_j's are the coefficients of an infinite order polynomial in positive and negative powers of the lag operator L and such that: $\Xi(z) = \sum_{j=-\infty}^{\infty} \xi_j z^j = [\Psi(z^{-1})]^{-1}[\Phi(z)]^{-1}$.

When errors e_t are normally distributed, the causal and noncausal components of the dynamics cannot be distinguished, and model (1) is not identifiable. However, the causal and noncausal autoregressive coefficients are identifiable when the process (e_t) is not Gaussian.[5] For example, [24] consider t-Student distributed errors, while [21, 22] discuss the properties of the purely noncausal autoregressive process ($r = 0$, $s = 1$) with Cauchy distributed errors. In particular, the density of a Cauchy distributed random variable X with location μ and scale γ is:

$$f(e_t) = \frac{1}{\pi}\left[\frac{\gamma}{(x-\mu)^2+\gamma^2}\right]$$

In Sect. 3, we will assume that $e_t \sim Cauchy(0, \gamma)$. A particular feature of the Cauchy distribution is that the expected value as well as all population moments of any higher order do not exist.

3.2 The Bubble Effect

The trajectory of the Bitcoin/USD exchange rate displays repetitive episodes of upward trends, followed by instantaneous drops, which are called bubbles. In general, a bubble has two phases: (1) a phase of fast upward (or downward) departure from the stationary path that resembles an explosive pattern and displays an exponential rate of growth, followed by (2) a phase of sudden almost vertical drop (or upspring) back to the underlying fundamental path. There exist several definitions of a bubble in the economic literature. The first definition was introduced by Blanchard [4] in the framework of rational expectation models. The formal definition by Blanchard as well as the later definitions by Blanchard and Watson [5], Evans [18] all assume a nonlinear dynamic models of x_t (say) with two components, one of which depicts the fundamental path of x_t, while the second one represents the bubble effect. The economic explanation of this phenomenon is as follows: a bubble results from the departure of a price of an asset from its fundamental value. In the context of the Bitcoin/USD exchange rate, the bubbles may result from the speculative trading that makes the rate deviate quickly above its trend, although it is hard to say if the trend is representative of the fundamental value of the bitcoin. Indeed, the bitcoin is a virtual currency, which is backed neither on a real asset, nor on the performance of a firm or a national economy.

So far, the bubbles were considered in the time series literature as nonstationary phenomena and treated similarly to the explosive, stochastic trends due to unit roots. In fact, the existing tests for the presence of a bubble are essentially tests of a

[5] See e.g. [13], [33, Theorem 1.3.1.], for errors with finite variance, Breidt [10] for errors with finite expectation and infinite variance, [22] for errors without finite expectation, as the Cauchy errors considered in the application.

breakpoint in the general explosive stochastic trend of a nonstationary process (see e.g. [31, 32]).

Gourieroux and Zakoian [21] propose a different approach and assume that the bubbles are rather short-lived explosive patterns caused by extreme valued shocks in a noncausal, stationary process. Formally, that process is a noncausal AR(1) model with Cauchy distributed errors. The approach in reverse time, based on a noncausal model allows for accommodating the asymmetric pattern of the bubble. The merit of the Cauchy distributed errors is in replicating the sudden spike in the reverse time trajectory that is observed as a bubble burst from the calendar time perspective. Such a noncausal or mixed process has to be examined conditionally on the information of the current and past rates. It is known that a noncausal, linear autoregressive process also has a nonlinear causal autoregressive dynamics, except in the Gaussian case. This is the special nonlinear feature, which makes it suitable for modelling the bubbles in Bitcoin/USD exchange rate. Moreover the noncausal autoregressive model allows for forecasting the occurrence of a future bubble and the time of bubble burst. The methodology of forecasting is discussed in Sect. 3.4 and illustrated in the application in Sect. 4.

3.3 Estimation and Inference

The traditional approach to the estimation of causal time series models relies on the Box-Jenkins methodology that consists of three steps: identification, estimation and diagnostics. In application to noncausal and mixed processes, most of the traditional Box-Jenkins tools of analysis need to be interpreted with caution. The reason is that most of the traditional estimators are based on the first- and second-order sample moments of the process and rely on the Gaussian approximation of its density, while the noncausal processes need to be non-Gaussian to solve the aforementioned identification purpose and may have infinite moments of order one and/or two.

(a) *Identification*

The autocorrelation function (ACF) is the basic tool for detecting temporal dependence. By construction, the ACF estimators rely on a implicit normality assumption, as they are computed from the sample moments up to order two. Due to the aforementioned nonidentifiability problem, the ACF cannot reveal whether a time series is causal or not, as it yields identical results in either case. It remains however a valid tool for detecting serial dependence in variables with infinite variances (see [2]). In particular Andrews and Davis show that the total autoregressive order $p = r + s$ can be inferred from the autocorrelation function, while r and s need to be inferred from the estimated models by comparing their fit criteria, computed from the sample.

For variables with infinite variance, [15] established the asymptotic properties of the sample autocorrelation $\hat{\rho}$ at lag l defined as:

$$\hat{\rho}(l) = \frac{\hat{\gamma}(l)}{\hat{\gamma}(0)}, \ \hat{\gamma}(l) = \frac{1}{T}\sum_{t=1}^{T-l}(y_t - \bar{y})(y_{t+l} - \bar{y}), \ l > 0,$$

where T is the sample size and $\bar{y} = \frac{1}{T}\sum_{t=1}^{T} y_t$.

In the presence of Cauchy errors, the standard confidence intervals of the ACF are no longer valid as the sample ACF is no longer asymptotically normally distributed and has a nonstandard speed of convergence. By using the results of [15, 21] (Proposition 6) show that the sample autocorrelations of a noncausal AR(1) process with Cauchy errors and autoregressive coefficient ρ have a limiting stable distribution and a rate of convergence that is different from the standard $\sqrt{T}$ rate. More specifically, let us denote the vector of sample autocorrelations up to lag M: by $\hat{\rho}_T = (\hat{\rho}_T(1), \ldots, \hat{\rho}_T(M))'$ and consider the true values $\rho = (\rho, \ldots, \rho^M)'$. Then,

$$\frac{T}{\ln T}(\hat{\rho}_T - \rho) \xrightarrow{d} Z = (Z_1, \ldots, Z_M),'$$

where for $l = 1, \ldots, M$, $Z_l = \sum_{j=1}^{\infty}[\rho^{j+l} - \rho^{|j-l|}]S_j/S_0$, and $S_1, S_2 \ldots$ is an i.i.d. sequence of symmetric 1-stable random variables independent of the positive 1/2 stable random variable S_0. The limiting true values can be interpreted as pseudo-autocorrelations, as the autocorrelations themselves do not exist in a process with infinite variance.

(b) *Estimation*

The standard Gaussian quasi-maximum likelihood approach can no longer be used to estimate the autoregressive parameters due to the Gaussian-specific identification problem. However, when the distribution of the errors is non-Gaussian, the estimation of the parameters in noncausal and mixed processes can be based on the maximum likelihood estimator, which preserves its speed of convergence and asymptotic normality (see [24]). The maximum likelihood method differs slightly from that used in causal processes. It is called the "approximate maximum likelihood" for the reason that the sample used in the approximate likelihood is reduced to $T - (r + s)$ observations.[6] Indeed, the first error to be included in the likelihood function that can be written without a value of y_t prior to the sample is e_{r+1}. To see that, assume $\psi_1 = \cdots = \psi_s = 0$ and write:

$$e_{r+1} = y_t - \phi_1 y_{t-1} - \cdots - \phi_r y_{t-r}.$$

Suppose now that $\phi_1 = \cdots = \phi_r = 0$ The last error in the sample to be included in the likelihood function that can be written without the values of y_t posterior to the sample is

$$e_{T-s-1} = y_T - \psi_1 y_{T+1} - \cdots - \psi_s y_{T+s}$$

[6] The approximate likelihood disregards the first r state variables that summarize the effect of shocks before time r and the last s state variables that summarize the effect of shocks after time $T - s$ [21, 22] and is therefore constructed from shocks $e_{r+1}, \ldots e_{T-s-1}$ only.

The Approximate Maximum Likelihood (AML) is defined as:

$$(\hat{\Psi},\ \hat{\Phi},\ \hat{\theta}) = Argmax_{\Psi,\Phi,\theta} \sum_{t=r+1}^{T-s} \ln\ g[\Psi(L^{-1})\Phi(L)y_t; \theta], \tag{2}$$

where $g[.; \theta]$ denotes the probability density function of e_t.

Lanne and Saikkonen [24] show that the traditional Wald tests and other inference methods, like the AIC and SBC fit criteria based on the approximated likelihood remain valid. The fit criteria are essentially used for determining the autoregressive orders r and s of the process. The model with the r and s that minimizes one of the fit criteria is selected at this "ex post" identification stage, analogously to the choice of orders p and q in an ARMA(p, q) process.

(c) *Diagnostics*

The diagnostic checking consists of testing if the estimated shocks $\hat{e}_t = \hat{\Psi}(L^{-1})\hat{\Phi}(L)y_t$ of the model are strong white noise. The asymptotic distribution of the sample autocorrelation of the residuals is different from the standard one derived for processes with finite variance. For instance, for a noncausal Cauchy autoregressive process of order 1, the limiting distribution of the residual autocorrelation estimator at lag 1 is:

$$\frac{T}{\ln T}\hat{r}_T(1) \xrightarrow{d} \rho^*(1 + 2\rho^*)S_1/S_0,$$

where ρ^* is the noncausal autoregressive coefficient of process Y. Contrary to the standard process with finite variance, the limiting distribution depends on ρ^*.

3.4 Forecasting

Due to the different dynamics of non-Gaussian processes in the calendar and the reverse times, the "backcasting" algorithm in the spirit of Newbold [30] is no longer valid. Nevertheless, it is possible to extend the concept of the Kalman filter and make it applicable to noncausal and mixed processes. The approach consists of three steps (see e.g. [20]):

Step 1 Filter shocks e_t for $t = 1, \ldots, T$ and the causal and noncausal state components of the process.

Step 2 Estimate the predictive distribution of the noncausal component of the process by a look-ahead estimator.

Step 3 Simulate the future noncausal components of the process by a sampling importance resampling (SIR) algorithm and infer the simulated future values of the process.

This methodology is used in the next section to derive the prediction intervals.

4 Application

4.1 ACF Analysis

The traditional Box-Jenkins approach starts from the analysis of the sample autocorrelation function (ACF). The ACF provides information on the possible linear serial dependence in the series, but its interpretation can be rather misleading in the case of extreme events.

The standard confidence interval for testing the statistical significance of the autocorrelations is based on the approximate limiting standard normal distribution of the autocorrelation estimator at a given lag, under the null hypothesis that the true value of that autocorrelation is zero. Hence, with the sample size of 150, the statistically significant autocorrelations exceed 0.16 in absolute value.

In order to establish the confidence interval for Cauchy distributed errors, we approximate the limiting distribution of the pseudo-autocorrelation estimator given in Sect. 3.3 by simulations. We draw independent standard normals e_1, e_2, e_3 and build the ratio

$$Z = \frac{S_1}{S_0} = \frac{e_1 e_3^2}{e_2}$$

where $S_1 = \frac{e_1}{e_2}$ is a symmetric 1-stable random variable and $S_0 = \frac{1}{e_3^2}$ is a symmetric 0.5 stable random variable. From the 25th and 975th order statistics from a sample of 1,000 values of Z multiplied by $\frac{\ln(150)}{150}$, we obtain the confidence interval $[-0.36, 0.36]$. Under the null hypothesis of zero pseudo-autocorrelation at lag l, the statistically significant autocorrelation at lag l is less than 0.36 in absolute value with the asymptotic probability of 95 %.

In Fig. 5, we plot the ACF of the data with the standard confidence interval and the interval adjusted for infinite variance.

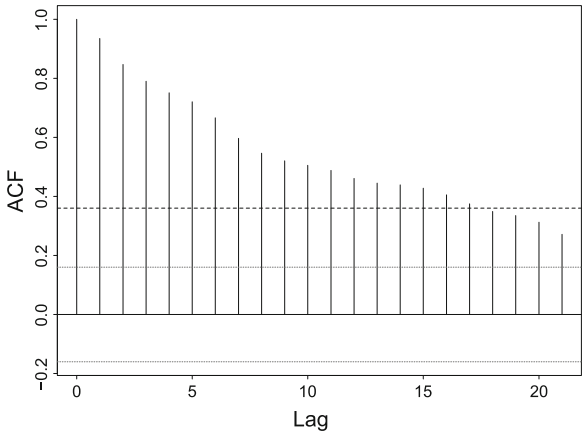


Fig. 5 ACF, Bitcoin

The ACF displays slow, linear decay, which resembles the patterns observed in unit root processes. Moreover, the Dickey-Fuller and the Augmented Dickey-Fuller ADF(4) tests accept the null hypothesis of a unit root in the data with p-values 0.4 and 0.6, respectively. However, it is easily checked that the standard procedure of transforming the data into first differences and estimating a stationary ARMA cannot accommodate the nonlinear features of the series.

4.2 Global and Local Trends

In the Bitcoin/USD exchange rate series, it is important to disentangle the fundamental and the bubble components. The fundamental component is modelled as a nonlinear deterministic trend[7] and the bubble component as a noncausal autoregressive process with Cauchy errors. Accordingly, we define the Bitcoin/USD rate as:

$$rate_t = trend_t + y_t,$$

(a) *Estimation of the trend and detrended series*

In order to remove the trend, we fit a nonlinear function of time by regressing the data on a 3rd degree polynomial in time. The detrended series, obtained as the following series of residuals:

$$y_t = rate_t + 3.045 - 3.854t + 3.499t^2 - 0.866t^3$$

is calculated and plotted in Fig. 6. The marginal density of y_t is shown in Fig. 7.

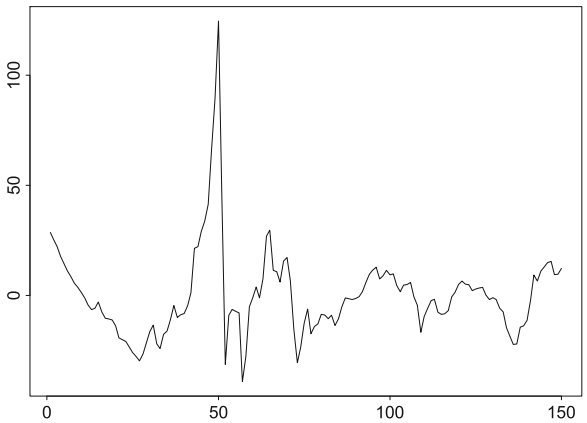


Fig. 6 Detrended series

[7] Alternatively, it can be represented by a model with a stochastic trend, assumed independent of the shocks that create the speculative bubble.

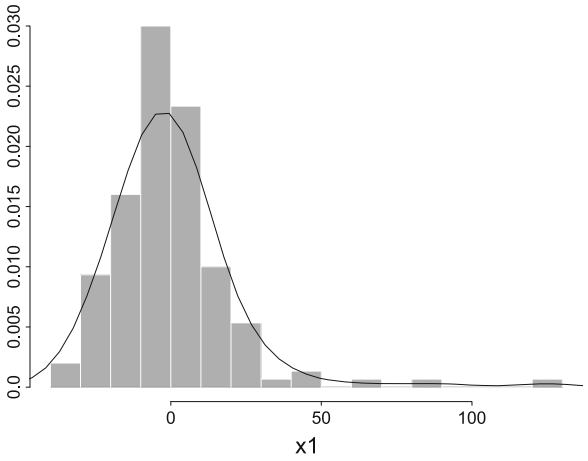


Fig. 7 Detrended series, histogram

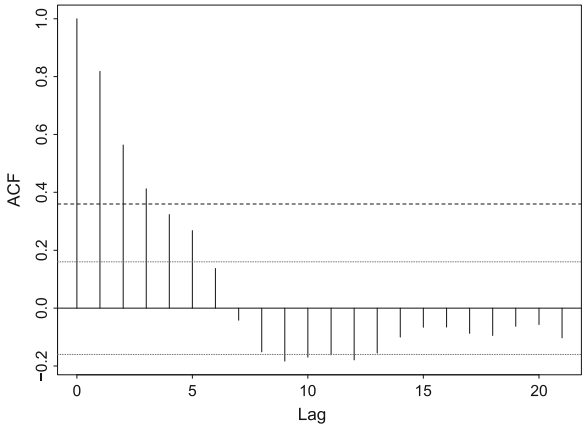


Fig. 8 Detrended series, ACF

We observe that the detrended series no longer displays the bimodal pattern, while it preserves the peaked and long-tailed shape of the density of the Bitcoin/USD rate. The ACF function of the detrended series, given in Fig. 8, shows considerably less persistence than the original series and indicates short linear memory.

(b) *Noncausal analysis of the detrended series*

Next, the detrended series is modelled as a noncausal autoregressive process. Let us first consider a noncausal Cauchy AR(1) process:

$$y_t = \psi y_{t+1} + e_t, \tag{3}$$

Table 1 AR(1) Parameter estimates

	Parameter	Standard error	t-ratio
ψ	0.9122	0.024	37.025
γ	2.734	0.113	8.833
$-\ln L$	496.165	–	–

where e_t are independent and Cauchy distributed with location 0 and scale γ, $e_t \sim Cauchy(0, \gamma)$. At this point, it is interesting to compare the trajectory of y_t with the simulated path of a noncausal AR(1) with the autoregressive coefficient 0.9, as displayed in [21, Fig. 4]. It is clear that the dynamics of the transformed Bitcoin/USD rate and of the simulated series resemble one another. The model is estimated by maximizing the approximated log-likelihood function, based on the Cauchy density function:

$$\ln L(\psi, \gamma) = (T-1)[-\ln(\pi) + \ln(\gamma)] - \sum_{t=1}^{T-1}[\ln((y_t - \psi y_{t+1})^2 + \gamma^2)]. \quad (4)$$

The parameter estimates and the estimated standard errors are given in Table 1:

The residuals are plotted in Fig. 9.

The model has not only removed the serial correlation, as shown in Fig. 10, but has also removed the asymmetry due to the bubble (see Fig. 7).

The speculative subperiod is just characterized by a rather standard volatility clustering.

An additional autoregressive term can be introduced. We consider a noncausal AR(2) model:

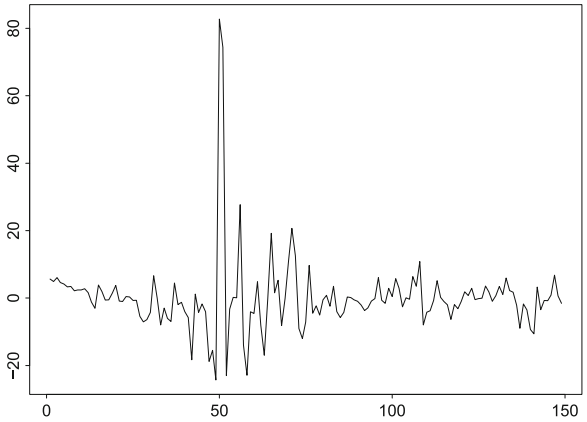


Fig. 9 Residuals, noncausal AR(1)

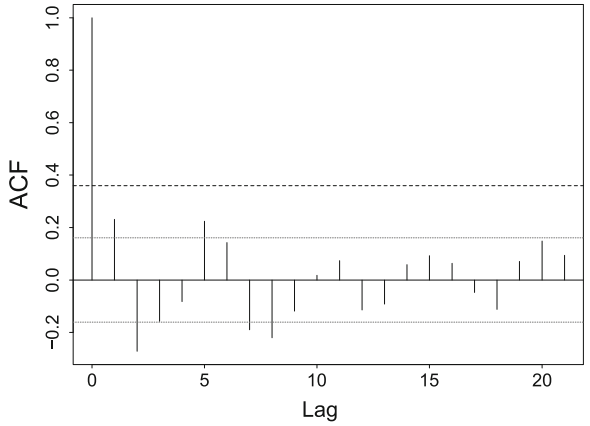


Fig. 10 ACF residuals, noncausal AR(1)

Table 2 AR(2) Parameter estimates

	Parameter	Standard error	t-ratio
ψ_1	1.316	0.077	17.0465
ψ_2	−0.401	0.065	−6.166
γ	2.433	0.112	7.894
$-\ln L$	478.709	–	–

$$y_t = \psi_1 y_{t+1} + \psi_2 y_{t+2} + e_t, \tag{5}$$

and estimate it by the approximated maximum likelihood.

The roots of the noncausal polynomial are 1.194 and 2.084 and the noncausal AR(2) is stationary. The residuals of the noncausal AR(2) satisfy also white noise features, as their autocorrelations are not statistically significant. The noncausal AR(2) provides good fit to the data and the value of its log likelihood function at the maximum is very close to that of the next model considered with the same number of parameters (Table 2).

The next specification considered is a mixed autoregressive model MAR(1,1) with both causal and noncausal orders equal to 1. The estimated model is:

$$(1 - \phi L)(1 - \psi L^{-1}) y_t = e_t. \tag{6}$$

The parameter estimates are provided in Table 3. Both roots of the polynomials lie outside the unit circle.

In order to accommmodate remaining residual autocorrelation, we estimate the model MAR(2, 2) (see Table 4):

Table 3 MAR(1, 1) Parameter estimates

	Parameter	Standard error	t-ratio
ψ	0.678	0.028	23.864
ϕ	0.717	0.023	30.507
γ	2.559	0.109	8.552
$-\ln L$	479.402	–	–

Table 4 MAR(2, 2) Parameter estimates

	Parameter	Standard error	t-ratio
ψ_1	0.739	0.025	29.495
ψ_2	0.032	0.023	1.367
ϕ_1	0.501	0.063	7.912
ϕ_2	0.114	0.027	4.084
γ	2.413	0.112	7.842
$-\ln L$	471.507	–	–

Table 5 MAR(2, 1) Parameter estimates

	Parameter	Standard error	t-ratio
ψ_1	0.632	0.046	13.479
ϕ_1	0.664	0.037	17.715
ϕ_2	0.157	0.033	4.679
γ	2.481	0.114	7.911
$-\ln L$	470.658	–	–

$$(1 - \phi_1 L - \phi_2)(1 - \psi_1 L^{-1} - \psi_2 L^{-2})y_t = e_t. \tag{7}$$

Both polynomials have real-valued roots outside the unit circle. We observe that the parameter ψ_2 is not significant. Therefore, below, we estimate the MAR(2, 1) model (see Table 5).
The autoregressive polynomial in past $y's$ has real-valued roots outside the unit circle.

The noncausal AR(2) process and the MAR(1, 1) process are not equivalent from the modeling point of view. Noncausal parameters ψ_j (resp. causal parameters ϕ_j) have a significant impact on the rate of increase of the bubble (resp. decrease of the bubble). The noncausal AR(2) model is able to fit bubbles with two possible rates of increase, corresponding to ψ_1 and ψ_2, but with sharp decrease due to the absence of causal autoregressive parameter. The mixed MAR(1, 1) model is flexible enough to fit any asymmetric bubbles. For illustration purpose, we focus below on the mixed MAR(1, 1) model.

4.3 The Causal and Noncausal Components

The MAR(1, 1) process can be decomposed into a "causal" and a "noncausal" components (see e.g. [20]):

$$y_t = \frac{1}{1-\phi\psi}(u_t + \phi v_{t-1}), \tag{8}$$

where the noncausal component is defined by

$$u_t - \psi u_{t+1} = e_t, \tag{9}$$

and the causal component by

$$v_t - \phi v_{t-1} = e_t. \tag{10}$$

The causal component v_t (resp. the noncausal component u_t) is a combination of current and lagged values (resp. of the current and future values) of the noise e_t. This explains the terminology used above.

In Fig. 11, we provide the filtered noise and the filtered causal and noncausal components of the detrended series.

The series of filtered e_t is close to a series of independent random variables, as suggests the ACF given in Fig. 12. From the series of filtered e_t, we can see the dates of extreme shocks.

The component series determine jointly the bubble patterns. In particular, the component u with parameter ψ determines the growth phase of the bubble, while the v component and parameter ϕ determine the bubble burst.

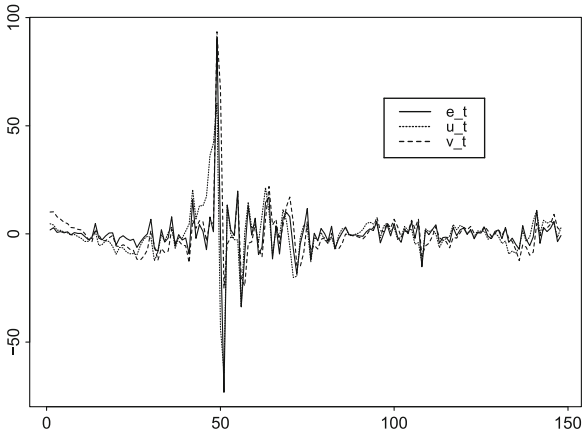


Fig. 11 Components series of the MAR(1, 1)

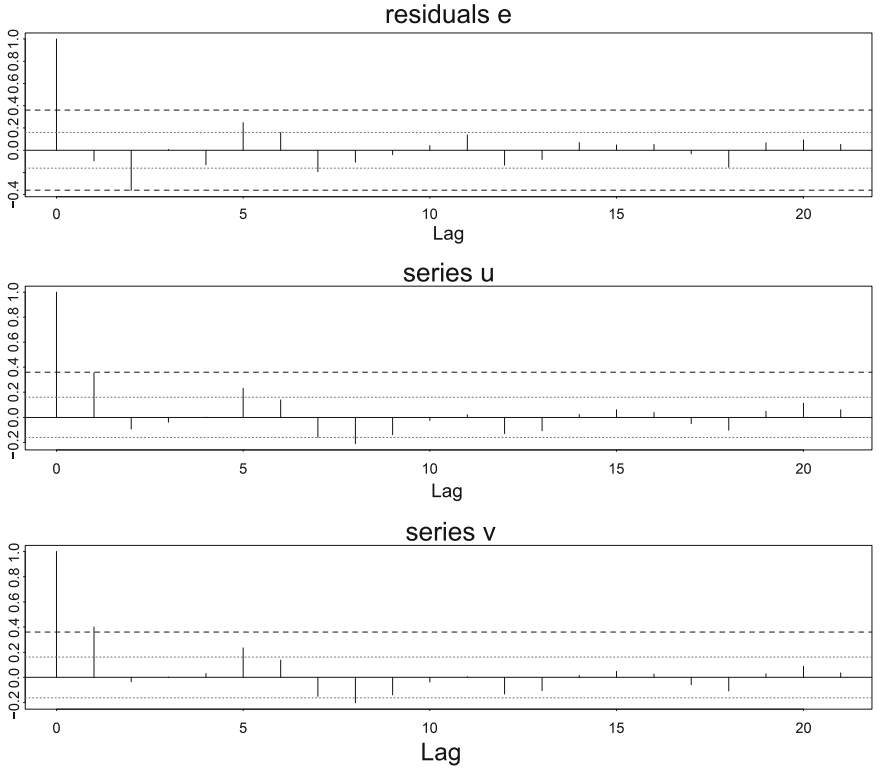


Fig. 12 ACF of the MAR(1, 1) components

4.4 Prediction

Let us now consider the prediction performance of the MAR(1, 1) with nonlinear deterministic trend. In particular, we will study its ability of predicting the future bubbles.

For this purpose we need to predict the future path of detrended process y at some horizon H, that is $y_{T+1}, \ldots, y_{T+H}$, given the available information $y_1, \ldots, y_T$. We are interested in nonlinear prediction of this path represented by the predictive density of $y_{T+1}, \ldots, y_{T+H}$. This analysis depends on horizon H. It becomes more complex when H increases, but also more informative concerning the possible future downturns. There exist consistent approximations of this joint predictive density, based on a look-ahead estimator, which admit closed form expressions. They can be used for small H to display the predictive densities and in general to compute any moment of the type:

$$E(a(y_{T+1}, \ldots, y_{T+H})|y_1, \ldots, y_T),$$

by simulation or numerical integration.

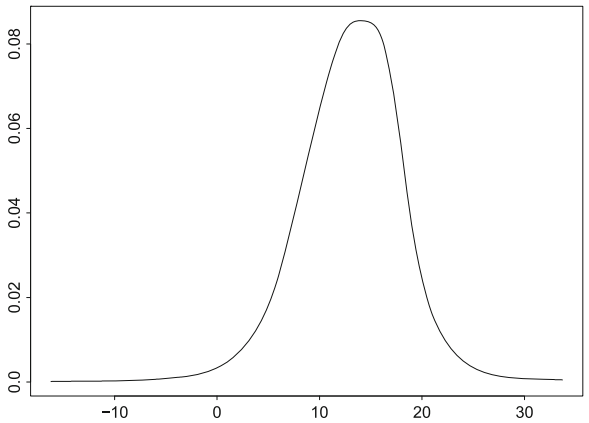


Fig. 13 Predictive density at horizon 1

For $s = 1$, the predictive distribution at horizon H is defined as [20]:

$$\hat{\Pi}(u_{T+1}, \ldots, u_{T+H}|\hat{u}_T) = \left\{ \hat{g}(\hat{u}_T - \hat{\psi} u_{T+1}) \hat{g}(u_{T+1} - \hat{\psi} u_{T+2}) \right.$$
$$\left. \ldots \hat{g}(u_{T+H-1} - \hat{\psi} u_{T+H}) \sum_{t=1}^{T-1} \hat{g}(u_{T+H} - \hat{\psi} \hat{u}_t) \right\}$$
$$\left\{ \sum_{t=1}^{T-1} \hat{g}(\hat{u}_T - \hat{\psi} \hat{u}_t) \right\}^{-1} \tag{11}$$

We display in Fig. 13 the predictive density at horizon 1, in Fig. 14 the joint preditive density at horizon 2, and its contour plot in Fig. 15.

We observe some asymmetry in the predictive density of y_{T+1}, which would not have been detected with a standard Gaussian ARMA model.

The joint predictive density at horizon 2 is much more informative. For the MAR(1, 1) process, all the predictive densities depend on the information by means of two state variables only, which are y_T, u_T, or equivalently by y_T, y_{T-1}. The values of the state variables for the Bitcoin data are: $u_T = 2.87$, $y_{T-1} = 9.64$, $y_T = 12.27$. Thus at the end of the observation period we are in an increasing phase of the detrended series. From the knowledge of the joint predictive density, we can infer the type of pattern that will follow that increasing phase. The series can continue to increase (the upper North-East semi-orthant in Fig. 15 from the top point with coordinates of about $(y_T, y_T) = (12.27, 12.27)$), increase and then slowly decrease (the bottom North-East semi-orthant), or immediately decrease sharply (the bottom South-West semi-orthant) and so on.

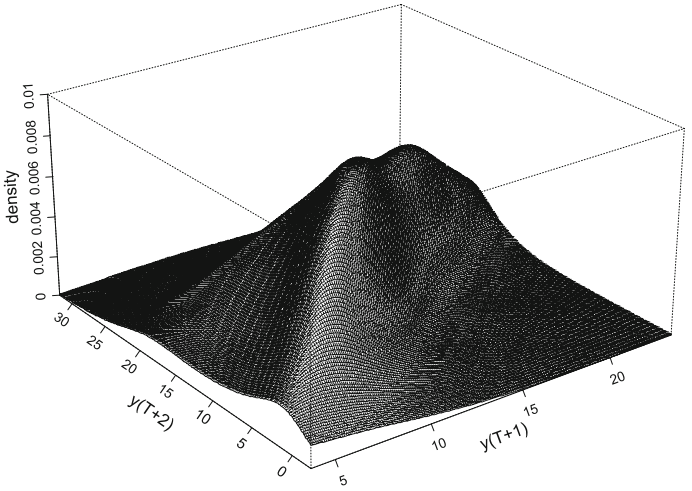


Fig. 14 Joint predictive density at horizons 1 and 2

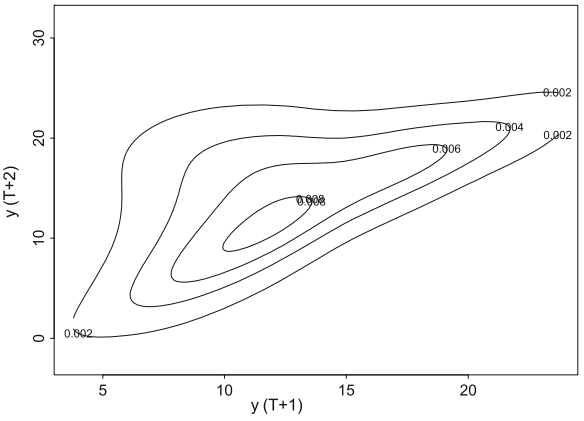


Fig. 15 Contour plot of the joint predictive density at horizon 2

The joint predictive density has a complicated pattern, far from Gaussian, with a strong dependence in extreme future risks in some directions. Its associated copula is close to an extreme value copula (see e.g. [3] for examples of extreme value copulas). By considering Fig. 15, we see that the probability of a continuing increase of y (North-East orthant) is rather high, but so is the probability of a sharp downturn at date $T + 1$ (South-West orthant). However, the probability of a downturn at date $T + 2$ (South-East orthant) is small. Thus the joint predictive density can be used to recognize the future pattern of y by comparing the likelihood of the different scenarios, in particular to evaluate the probability of the downturn at dates $T + 1$, $T + 2$, etc. The above discussion based on the graphical representation is limited

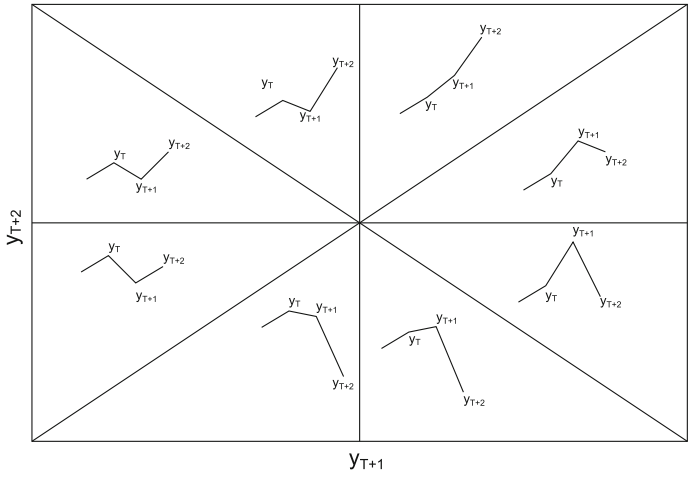


Fig. 16 Predicted dynamics

to horizon 2 (see Fig. 16). However, the probabilities of the different types of future patterns can be evaluated numerically for larger H.

5 Conclusion

The causal-noncausal autoregressive models have been proposed as nonlinear dynamic models that are able to fit speculative bubbles. We applied thismethodology to analyse the Bitcoin/USD exchange rates over the period February–July 2013. Indeed, speculative bubbles appeared in that period and could be used to calibrate the parameters of the model. We have considered a mixed model with both causal and noncausal orders equal to 1, estimated the parameters by the Approximated Maximum Likelihood, filtered the underlying components of the process to better understand the type of existing bubbles. Next, we built the joint predictive density of the future path. This joint density was used to predict the future patterns of the process, a kind of model-based chartist approach, in particular to evaluate the likelihood of the future dates of downturn.

The series of Bitcoin/USD exchange rates has been used as a playground for analyzing the relevance of the causal-noncausal modeling to capture bubble phenomena. It was typically an example of highly speculative emerging market. Recently, several exchange platforms have closed, temporarily[8] or definitely. Other platforms were submitted to regulations. There is clearly a need of supervision to better protect the investors in bitcoins against the theft of their bitcoins, but also against the

[8] The French platform Bitcoin-Central has been closed for 5 months in 2013 due to hackers attack. Nevertheless the customers had still the possibility to withdraw their bitcoins.

speculative behavior of large bitcoin holders. This supervision will likely make disappear the previously observed speculative bubbles and perhaps the market for this electronic currency itself.

However, there will still exist a large number of financial markets, not necessarily emerging, with frequently appearing bubbles. Examples are the markets for commodity futures, and the markets with high frequency trading. These are potential applications for the causal-noncausal model presented in this paper.

References

1. Andrews, B., Calder, M., Davis, R.: Maximum likelihood estimation for α-stable autoregressive processes. Ann. Stat. **37**, 1946–1982 (2009)
2. Andrews, B., Davis, R.: Model identification for infinite variance autoregressive processes. J. Econom. **172**, 222–234 (2013)
3. Balkema, G., Embrechts, P., Nolde, N.: The shape of asymptotic dependence. In: Shirayev, A., Varadhan, S., Presman, E. (eds.) Springer Proceedings in Mathematics and Statistics, special volume. Prokhorov and Contemporary Probability Theory, vol. 33, pp. 43–67 (2013)
4. Blanchard, O.: Speculative bubbles: crashes and rational expectations. Econ. Lett. **3**, 387–389 (1979)
5. Blanchard, O., Watson, M.: Bubbles, rational expectations and financial markets. In: Wachtel, P. (ed.) Crisis in the Economic and Financial Structure, pp. 295–315, Lexington (1982)
6. Bitcoin: Introduction. Retrieved 4 December 2013, https://en.bitcoin.it/wiki/Introduction (2013)
7. Bitcoin: Controlled Supply. Retrieved 4 December 2013, https://en.bitcoin.it/wiki/Controlled_Currency_Supply (2013)
8. Bitcoin: Trade. Retrieved 4 December 2013, https://en.bitcoin.it/wiki/Trade (2013)
9. Blockchain. Bitcoin Market Capitalization. Retrieved from Bitcoin Block Explorer, http://blockchain.info/charts/market_cap
10. Breidt, F., Davis, R.: Time reversibility, identifiability and independence of innovations for stationary time series. J. Time Ser. Anal. **13**, 273–390 (1992)
11. Breidt, F., Davis, R., Lii, K.: Maximum likelihood estimation for noncausal autoregressive processes. J. Multivar. Anal. **36**, 175–198 (1991)
12. Brunnermeier, M.: Asset Pricing under Asymmetric Information: Bubbles, Crashes, Technical Analysis and Herding. Oxford University Press, Oxford (2001)
13. Cheng, Q.: On the unique representation of non-Gaussian linear processes. Ann. Stat. **20**, 1143–1145 (1992)
14. Davis, J.: The Crypto-Currency, The New Yorker. Retrieved from http://www.newyorker.com/reporting/2011/10/10/111010fa_fact_davis (2011)
15. Davis, R., Resnick, S.: Limit theory for moving averages of random variables with regularly varying tail probabilities. Ann. Probab. **13**, 179–195 (1985)
16. Davis, R., Resnick, S.: Limit theory for the sample covariance and correlation functions of moving averages. Ann. Stat. **14**, 533–558 (1986)
17. Davis, R., Song, L.: Noncausal Vector AR Processes with Application to Economic Time Series, DP Columbia University (2012)
18. Evans, G.: Pitfalls in testing for explosive bubbles in asset prices. Am. Econ. Rev. **81**, 922–930 (1991)
19. Flitter, E.: FBI shuts alleged online drug marketplace, Silk Road, Reuters. Retrieved from http://www.reuters.com/article/2013/10/02/us-crime-silkroad-raid-idUSBRE9910TR20131002 (2013)

20. Gourieroux, C., Jasiak, J.: Filtering, Prediction and Estimation of Noncausal Processes. CREST (2014)
21. Gourieroux, C., Zakoian, J.M.: Explosive Bubble Modelling by Noncausal Cauchy Autoregressive Process. CREST (2013)
22. Gourieroux, C., Zakoian, J.M.: On Uniqueness of Moving Average Representation of Heavy Tailed Stationary Processes. CREST (2013)
23. Lanne, M., Saikkonen, P.: Noncausal autoregressions for economic time series. J. Time Ser. Econom. **3**(3), Article 2 (2011)
24. Lanne, M., Luoto, J., Saikkonen, P.: Optimal forecasting of nonlinear autoregressive time series. Int. J. Forecast. **28**, 623–631 (2010)
25. Lanne, M., Saikkonen, P.: Noncausal vector autoregression. Econom. Theory **29**, 447–481 (2013)
26. Li, S.: Bitcoin now accepted as tuition payment at a Cyprus University, Los Angeles Times. Retrieved from http://www.latimes.com/business/money/la-fi-mo-cyprus-university-bitcoin-20131120,0,3194094.story#axzz2mXKIff7E (2013)
27. Litecoin Block Explorer: Litecoin Block Explorer Charts. Retrieved from 4 December 2013, http://ltc.block-explorer.com/charts (2013)
28. Liu, J.: BTC China the world's largest Bitcoin trading platform, ZD Net. Retrieved from http://www.zdnet.com/btc-china-the-worlds-largest-bitcoin-trading-platform-7000023316/ (2013)
29. Muth, J.: Rational expectations and the theory of price movements. Econometrica **29**, 315–335 (1961)
30. Newbold, P.: The exact likelihood function for a mixed autoregressive-moving average process. Biometrika. **61**, 423– 426 (1974)
31. Phillips, P., Shi, S., Yu, J.: Testing for Multiple Bubbles, DP Cowles Foundation, 1843 (2012)
32. Phillips, P., Wu, Y., Yu, J.: Explosive behavior in the 1990s Nasdaq: when did exuberance escalate asset values? Int. Econ. Rev. **52**, 201–226 (2011)
33. Rosenblatt, M.: Gaussian and Non-Gaussian Linear Time Series and Random Fields. Springer, New York (2000)
34. Sparshott, J.: Web Money Gets Laundering Rule, The Wall Street Journal. Retrieved from http://online.wsj.com/news/articles/SB10001424127887324373204578374611351125202 (2013)
35. Tagaris, K.: Cyprus details heavy losses for major bank customers, Reuters. Retrieved from http://www.reuters.com/article/2013/03/30/us-cyprus-parliament-idUSBRE92G03I20130330 (2013)
36. Velde, F.R.: Bitcoin: A primer, Chicago Fed Letter. Retrieved from http://www.chicagofed.org/digital_assets/publications/chicago_fed_letter/2013/cfldecember2013_317.pdf (2013)

An Overview of the Black-Scholes-Merton Model After the 2008 Credit Crisis

Chadd B. Hunzinger and Coenraad C.A. Labuschagne

Abstract The 2008 credit crisis exposed the over-simplified assumptions of the Black-Scholes-Merton (BSM) model. This paper provides an overview of some of the adjustments forced on the BSM model by the 2008 credit crisis to maintain the relevance of the model. The inclusion of credit value adjustment (CVA), debit value adjustment (DVA), funding value adjustment (FVA) and the posting of collateral in the BSM model are discussed.

1 Introduction

The credit crisis of 2008 was a dramatic event for financial markets. This was the beginning of the financial tsunami that would plague and force changes in global markets for many years to come. The economic meltdown that followed had massive effects on many everyday issues such as house prices, interest rates and inflation. Investment banks were also affected and numerous investment banks either defaulted or were taken over by the U.S. Federal Reserve to avoid default. The impact on financial derivative pricing did not escape the 2008 credit crisis.

Prior to the 2008 credit crisis, pricing the value of a derivative was relatively straightforward. Universally, practitioners and many academics agreed on the pricing method used to price a derivative. The method was well-known: discount future expected cash flows under the risk-neutral measure to the present date using the

C.B. Hunzinger
Rand Merchant Bank, 1 Merchant Place, Cnr Fredman Drive and Rivonia Road,
2196 Sandton, South Africa
e-mail: chaddhunzinger@gmail.com

C.B. Hunzinger
Research Associate, Faculty of Economics and Financial Sciences,
Department of Finance and Investment Management, University of Johannesburg,
P.O. Box 524, 2006 Aucklandpark, South Africa

C.C.A. Labuschagne (✉)
Department of Finance and Investment Management, University of Johannesburg,
P.O. Box 524, 2006 Aucklandpark, South Africa
e-mail: coenraad.labuschagne@gmail.com

V.-N. Huynh et al. (eds.), *Econometrics of Risk*, Studies in Computational Intelligence 583, DOI 10.1007/978-3-319-13449-9_3

risk-free rate. This method was derived from the fundamental theory laid down by Black, Scholes and Merton in the 1970s (see Black and Scholes [3] and Merton [26]).

Although there are many known approaches to option pricing, which includes heavy-tailed distribution techniques, the continuous time Black-Scholes-Merton (BSM) model is considered by many financial practitioners to be adequate for option pricing, irrespective of its over-simplified assumptions. It was and still is, widely used in practice, as it is well understood and yields a framework in which both pricing and hedging is possible. The deep-rooted acceptance of the BSM model is further cemented by the fact that the discrete time Cox, Ross and Rubinstein (CRR) model, which is a discretisation of the BSM model, is very useful and easy to implement in practice (see Cox et al. [10]).

The 2008 credit crisis drove home the fact that what was used in practice prior to the crisis as an approximation (also called a proxy) for the theoretical notion of a risk-free interest rate, as required by the BSM model, is totally inadequate to yield realistic results.

The myth that banks are risk-free was disproved by the 2008 credit crisis. The default of what we used to call *too big to fail* banks, such as Lehman Brothers and Bear Stearns, which defaulted in the 2008 credit crisis, disproved the myth that banks are risk-free (see Gregory [15]).

The 2008 credit crisis also exposed the inadequate management of counterparty credit risk. Counterparty credit risk (also known as default risk) between two parties, is the two-sided risk that one of the counterparties will not pay, as obligated on a trade or a transaction between the two parties.

Changes need to be made to the usual ways in which "business was conducted" prior to the 2008 credit crisis and these changes need to be addressed and incorporated in the models used prior to the 2008 credit crisis.

The aim of this paper is to present the current state of affairs with regard to the BSM model.

For terminology not explained in the paper, the reader is referred to Alexander [1] or Hull [18].

2 Credit Value Adjustment (CVA) and Debit Value Adjustment (DVA)

Over-the-counter (OTC) derivative trading is done directly between two parties without any supervision on an exchange (this, however, is going to change in the future due to stipulations in the Basel III Accord).

Even before the 2008 credit crisis, banks realised that many corporate clients are not risk-free; therefore, in OTC derivative trades, banks charged their clients a credit value adjustment (CVA). CVA is defined as the fair market value of the expected loss of an OTC derivative trade given that the opposite counterparty defaults.

Many of the banks' clients believed that banks were risk-free; therefore, the clients would accept the price that banks offered them and in turn did not charge banks a

CVA on the trade. Reference papers on CVA include Sorensen and Bollier [29], Jarrow and Turnbull [24] and Duffie and Huang [13].

As a result of the 2008 credit crisis, banks are not seen as risk-free anymore. One implication of this is the inclusion of a debit value adjustment (DVA) in the derivative's price. DVA is defined as the fair market value of the expected gain of an OTC derivative given own default. The origins of DVA are found in Duffie and Huang [13]; however, their paper deals mostly with swaps. Gregory [14] and Brigo and Capponi [4] examine bilateral credit risk in general and derive DVA formally. In essence, DVA is the adjustment clients charge the bank for the bank's own credit risk. Therefore, from the client's point of view, the adjustment is known as CVA and from the bank's point of view the adjustment is known as DVA.

One aspect of pricing with CVA and DVA is that it allows two credit-risky counterparties to trade with each other. If two counterparties charged each other a CVA and do not include the offsetting DVA term, then the two counterparties would not agree on the price of the derivative. The inclusion of DVA allows symmetric prices. The concept of symmetric prices means that two counterparties will price the derivative at the same price.

DVA is a hotly debated and controversial quantity. The reason for the controversy is that the DVA amount can only be realised when the bank defaults. If the bank is out-of-the-money on a trade and defaults, then the bank only needs to pay a recovery of the mark-to-market (MTM); therefore, the bank benefits from its own default. It can be compared to buying life insurance. The policy will only be realised after the death of the policy holder. Some practitioners argue that DVA simply cannot be hedged effectively.

Gregory and German [16] describe DVA as a double edged sword. On the one hand, it creates a symmetric world where counterparties can readily trade with one another, even when their underlying default probabilities are high. On the other hand, the nature of DVA and its implications and potential unintended consequences create some additional complexity and potential discomfort. From an accounting point of view, adding DVA to the price makes sense; however, the regulators are not so sure. Risk Magazine on 6th February 2012 reported that

> Accountants want banks to report as profits the impact of widening credit spreads on their liabilities, but regulators are moving in the other direction.

(see Carver [8]). The accounting rules International Financial Reporting Standards (IRFS) 13 and Financial Accounting Standards Board (FASB) 157 require DVA. However, the Basel III committee has decided to ignore any DVA relief in capital calculations.

3 The Risk-Free Rate: The Proxies LIBOR Versus OIS

The BSM model requires that one has to discount future expected cash flows under the risk-neutral measure using the risk-free rate. The risk-free rate is the theoretical

rate of return on an investment with no risk of financial loss. The risk-free rate defines the expected growth rates of market variables in a risk-neutral world.

In practice, the pertinent question is: which interest rate should be used as a proxy for the risk-free rate?

The London Interbank Offered Rate (LIBOR) rate is the rate that banks could freely borrow and lend at. Prior to the 2008 credit crisis, practitioners constructed a curve from LIBOR rates, Eurodollar futures and swap rates, which Hull and White [21] refer to as a LIBOR swap curve. The 3-month LIBOR swap curve was used by practitioners as a proxy for the risk-free rate.

An overnight interest rate swap (OIS) is a swap for which the overnight rate is exchanged for a fixed interest rate for a certain tenor (also known as maturity). An overnight index swap references an overnight rate index, such as the Fed funds rate, as the underlying for its floating leg, while the fixed leg would be set at an assumed rate.

Before the 2008 credit crisis, the LIBOR-OIS spread, which is the difference between the LIBOR rate and the OIS rate, was only a few basis points. It was stable and not significant (see Gregory [15]).

The 2008 credit crisis caused a significant spread between 3-month LIBOR and the OIS rate. The LIBOR-OIS spread spiked to hundreds of basis points in the aftermath of the default of Lehman Brothers in September 2008 and has remained significant ever since. Many practitioners believe that the spread between LIBOR and OIS rates describes the health of the banking industry. As one can see from Fig. 1 the banks in 2008 were not "in good shape" during the crisis. The fact that the LIBOR-OIS spread has remained significant illustrates why banks are not risk-free. These shifts made it apparent that LIBOR incorporates an adjustment for the credit risk of banks and swap

Fig. 1 The spread between 3-month LIBOR and OIS during the 2008 crisis. *Source* http://www.soberlook.com, May 31, 2014

rates correspond to the risk of unsecured short-term loans to financial institutions; therefore, the LIBOR swap curve is an imperfect proxy for the risk-free rate. The OIS rate appears to be the preferred choice as a proxy for the risk-free rate (see Hull and White [21] and Hunzinger and Labuschagne [23]).

4 Collateral and Funding Costs

The 2008 credit crisis emphasised the importance of the managing of counterparty credit risk.

One of the ways to mitigate counterparty credit risk is by posting collateral in a derivative trade. Collateral is a borrower's pledge of specific assets to a lender, to secure repayment of a liability.

Banks required collateral posted from their counterparties on certain trades prior to the 2008 credit crisis. But as it became apparent that banks are not risk-free, clients require that banks now also post collateral on some transactions. For exchange traded derivatives, i.e. stock option, counterparty credit risk is not an issue, because the two counterparties in the trade are required to post margins to the exchange.

The posting of collateral in a derivative trade is regulated by a Credit Support Annex (CSA). A CSA is a contract that documents collateral agreements between counterparties in trading OTC derivative securities. The trade is documented under a standard contract called a Master Agreement, developed by the International Swaps and Derivatives Association (ISDA). The 2010 ISDA margin survey suggests that 70 % of net exposure arising from OTC derivative transactions are collateralised (source: www2.isda.org).

After the 2008 credit crisis many banks have started to use OIS rates for discounting collateralised transactions and LIBOR swap rates for discounting non-collateralised transactions. This can be clarified by considering the fundamental paper of Piterbarg [27] in which he notes fundamental facts regarding derivative pricing when collateral is posted.

Piterbarg [27] notes that when pricing a zero-threshold CSA trade, where the collateral is cash and in the the same currency as the derivative, the cash flows should be discounted using the collateral rate of that particular currency. Collateral posted overnight will earn a rate similar to the index rate referenced in an OIS. Furthermore, when the trade is not collateralised, then the cash flows should be discounted using the funding rate of the bank. He also notes that one may price a derivative trade by always discounting the future expected cash flows using a collateral rate and making a funding value adjustment (FVA). FVA is a correction made to the risk-free price of an OTC derivative to account for the funding cost in a financial institution.

Posting collateral in an OTC trade may mitigate counterparty credit risk and funding costs; however, this depends on the collateral posted in the trade and how often this collateral is readjusted according to market movements. Collateral can be changed daily, weekly or monthly, which will affect the exposures of the two counterparties.

The 2008 credit crisis drove home the realisation that banks are not risk-free. This resulted in banks becoming reluctant to lend to each other and banks became unable to borrow at preferential rates. This resulted in banks charging a FVA on transactions. When managing a trading position, one needs cash to conduct operations such as hedging or posting collateral. This shortfall of cash can be obtained from the treasury of the bank. The funding cost adjustment (FCA) is the cost of lending money at a funding rate which is higher than the risk-free rate. The firm may also receive cash in the form of collateral or a premium. The funding benefit adjustment (FBA) is the benefit earned when excess cash is invested at a higher rate than the risk-free rate. Therefore, the funding value adjustment has two components

$$\text{FVA} = \text{FBA} + \text{FCA},$$

where the FBA and the FCA terms have opposite signs.

Funding value adjustment arises because of two factors. Firstly, because banks cannot borrow at the risk-free rate any more and secondly because of collateralised trades. Figure 2 illustrates how a funding cost adjustment arises in terms of a trading floor set-up. Let us say for example a trader enters into a trade with a corporate client and at this point in time the trader is in-the-money on the trade. At the same time the trader enters into another trade with a hedge counterparty to hedge out the trader's exposure to the client. Because the trader is in-the-money on the client trade, the trader will be out-the-money on the hedge. Let us also assume that the trade with the hedge counterparty is collateralised; therefore, the trader is required to posted collateral to the hedge counterparty. The collateral posted by the trader will earn a collateral rate. If the client trade was not traded with a CSA (no collateral will be posted in the trade) and then the trader needs to fund the collateral requirement from the treasury of the bank. The trader cannot fund a short fall of cash from the treasury

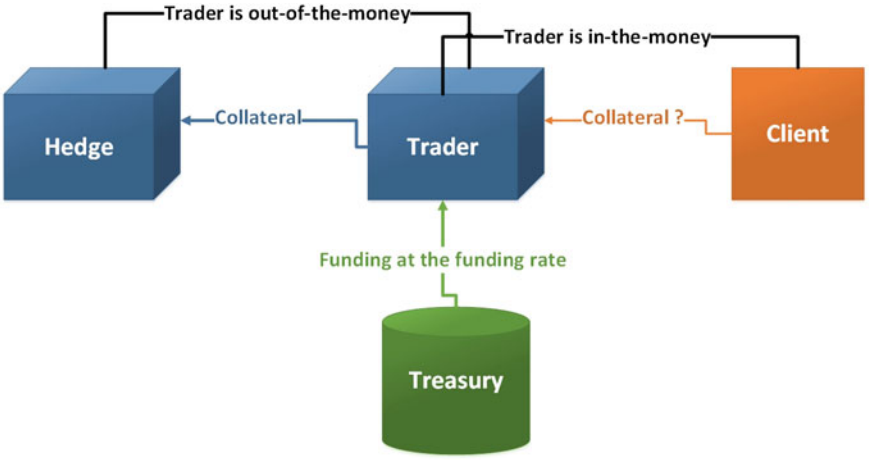


Fig. 2 A graphical illustration of funding cost adjustment

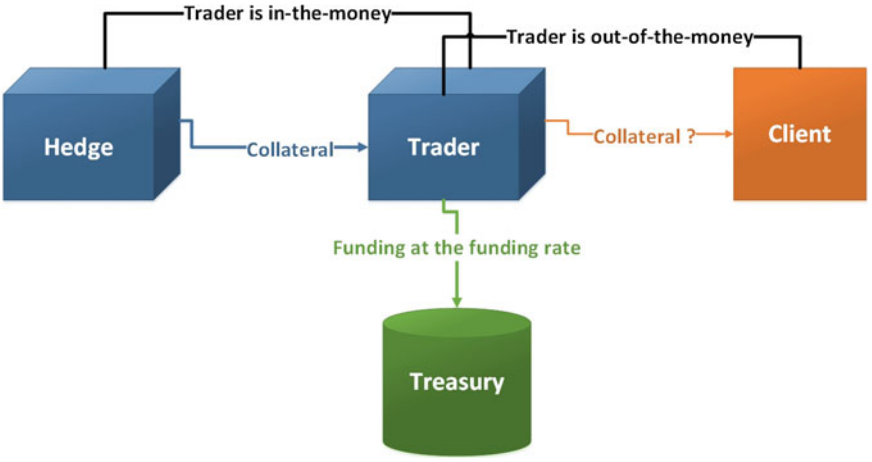


Fig. 3 A graphical ilustration of funding benefit adjustment

at the risk-free rate but at a higher funding rate. This excess cost of funding at the funding rate is exactly a funding cost adjustment. On the other hand, if the client trade is traded with a CSA, the client will be required to post collateral to the trader, and hence the trader can pass this collateral amount to the hedge counterparty. This situation results in no funding costs. The natural question at this point is: how does this scenario now differ from that prior to 2008? Before the crisis, the trader could fund from the treasury at a risk-free rate; hence, if the trader received collateral from the client or required funding from the treasury, this funding is at the risk-free rate. Therefore no funding cost adjustment would occur in this set-up.

Figure 3 illustrates how a funding cost adjustment arises in terms of a trading floor set-up. In this case, the trader is out-of-the-money on the client trade and in-the-money on the hedge. Now the hedge counterparty is required to post collateral with the trader. If no CSA is placed between the client and the trader, then the trader can place these funds with the treasury and earn a rate better than the collateral rate. This extra benefit is known as a funding benefit adjustment. If the trader is required to post collateral to the client, then there is no resulting benefit.

In this example, we assume that rehypothecation is possible. Rehypothecation is the practice by banks and brokers of using, for their own purposes, assets that have been posted as collateral by their clients.

5 The FVA Debate

The inclusion of FVA in pricing financial instruments is a controversial issue. Hull and White [19, 20] argue against it. They argue that the funding costs and benefits realised in a trade, violate the idea of risk-neutral pricing and should not be included

in the pricing of the derivative. Inclusion of FVA in the price of a derivative trade violates the law of one price in the market because the two counterparties may price a trade and obtain a different outcome.

There are views that are different from those of Hull and White. Laughton and Vaisbrot [25] suggest that in practice the market is not complete and the uniqueness of prices and the law of one price will not hold (see Harrison and Pliska [17]). They state that applying the so called FVA to the risk-neutral value, is justified. Banks with a lower funding rate will be more competitive on trades that require funding. This is fully consistent with the current situation in the markets, as theory should aim to be. In summary, they believe that the beautiful and elegant theory of BSM is not applicable and needs to be rethought because of the theory's unrealistic assumptions, especially post the credit crisis.

Castagna [9] also disagrees with Hull and White. In the BSM model there is only one interest rate and that is the risk-free interest rate. Castagna suggests if one considers a framework where more than one interest rate exists, such as a risk-free rate and a funding rate, then one could still produce a replicating portfolio which perfectly replicates the derivative. If a bank can only invest at the risk-free rate and fund at a higher funding rate, then it is well known that this will not impediment the replication of the derivative (see Bergman [2], Rubinstein and Cox [28] and Hunzinger and Labuschagne [23]). This will lead to a different prices for the buy side and the sell side; however, a closed form solution will still exist. Castagna suggests that models need to be amended in order to be more useful to traders. They should remove the assumption of the ability to borrow at the risk-free rate to finance trades.

Inclusion of both FBA and DVA in the price could also lead to double counting: FVA references the firms own funding spread (which is the difference between the funding rate of the bank and the risk-free rate) in both terms, FCA and FBA. The funding spread is based on the credit rating of the firm. The counterparty's credit spread (which is the difference between the yield on a firm's credit risky bond and the yield of a risk-free bond) is referenced in the CVA term and the firm's own credit spread in the DVA term. A change in the credit rating of the bank leads to a change in the price of the derivative. Since the DVA and FBA terms have the same sign, the change in the price is reflected twice if both the DVA and FBA terms are included in the valuation.

6 The BSM Model

In order to discuss extensions to the BSM model which follow from the discussions above, we include a summary of the BSM model and its assumptions for the convenience of the reader.

The interest rate assumptions of the BSM model are:

- The BSM model is a model with a single interest rate.
- This interest rate is the risk-free rate r.

The extensions to BSM model are concerned with amendments to these interest rate assumptions. The extended BSM model assumes multiple interest rates.

Other assumptions of the BSM model include:

- Stock prices follow geometric Brownian motion.
- Short selling is permitted.
- There are no taxes and transaction costs.
- No dividends on the underlying (although the BSM model can be adjusted to include dividends).
- No arbitrage opportunities exist.
- Continous trading of securities.

Given the described assumptions, we present the BSM partial differential equation (PDE). Let T denote the fixed time of maturity of a derivative contract and σ as the volatility of the underlying security, in this case a stock price. The Black-Scholes-Merton PDE is given by

$$\frac{\partial f}{\partial t} + rS\frac{\partial f}{\partial S} + \frac{1}{2}\sigma^2 S^2 \frac{\partial^2 f}{\partial S^2} = rf$$

where f, is the price of a derivative which is contingent on the stock price S_t and time $t \in [0, T]$. For a European call and put option with strike K, the BSM PDE has solution

$$V_t = \alpha \left(S_0 N\left(\alpha d_1\right) - Ke^{-r(T-t)} N\left(\alpha d_2\right) \right)$$

where

$$d_1 = \frac{\ln\left(\frac{S_0}{K}\right) + \left(r + \frac{1}{2}\sigma^2\right)(T-t)}{\sigma\sqrt{(T-t)}}$$

and

$$d_2 = d_1 - \sigma\sqrt{T-t},$$

where $\alpha = 1$ for a call option and $\alpha = -1$ for a put option. $N(x)$ is the cumulative distribution function of the standard normal distribution.

In practice a continuous-time model, such as the BSM model, is difficult to implement and is usually discretised to facilitate implementation. The Cox, Ross and Rubinstein (CRR) discrete-time model is a discretisation of the BSM model. Details of the CRR model can be found in Cox et al. [10] or Steland [30].

6.1 The BSM Model Which Includes Collateral and Funding Costs

Standard pricing theory excludes the intricacies of the collateralisation of the market. The posting of collateral in a derivative trade changes the traditional way in which a derivative is priced.

Piterbarg [27] extends the BSM continuous-time model to include collateral in a derivative trade and shows how the posting of collateral in a derivative trade affects the price.

In the Piterbarg model, the price of a collateralised derivative trade is given by

$$\text{risky price} = \text{risk-free price} + \text{FVA}.$$

The risk-free price is the BSM-price of a derivative that includes no credit risk and funding costs. This price is calculated by discounting all expected cash flows at the risk-free rate. The risky price is defined as the risk-free price plus any adjustments. Piterbag's paper won him the Quant of the Year Award in 2011.

It is possible to extend the CRR model to include dividends and collateral. Moreover, by using ideas along the lines of those in Hunzinger and Labuschagne [23], it can be shown that discretising Piterbarg's model (which is the BSM model that includes collateral and dividends) coincides with the aforementioned model. This is achieved by showing that Piterbarg's PDE, which represents the value of a collateralised derivative trade, can be represented as an expectation via the Feynman-Kac theorem.

6.2 The BSM Model Which Includes CVA, DVA and FCA

Currently, there are three unified frameworks which incorporate funding costs, collateral and credit risk into a derivative trade. These frameworks are proposed by

1. Piterbarg (see [27]), Burgard and Kjaer (see [6, 7]).
2. Brigo et al. (see [5]).
3. Crépey (see [11, 12]).

We take a closer look at the Burgard and Kjaer framework. The model proposed by Burgard and Kjaer [6] gives the price of a derivative trade by

$$\text{risky price} = \text{risk-free price} + \text{CVA} + \text{DVA} + \text{FCA},$$

where the CVA and DVA terms have opposite signs. The risky price is given by the risk-free price plus the adjustments for CVA, DVA and FCA.

It is possible to extend the CRR model to include CVA, DVA and FCA. Moreover, it can be shown that discretising the Burgard and Kjaer model (which is the BSM model that includes CVA, DVA and FCA) coincides with the aforementioned model. This is achieved by showing that Burgard and Kjaer's PDE can be represented as an expectation via the Feynman-Kac theorem. The details may be found in Hunzinger and Labuschagne [23].

Burgard and Kjaer [7] extends these two models discussed in Sects. 4 and 5 to create a general framework to price a credit risky derivative that is collateralised. This general framework, which is in the form of a PDE, reduces to the models presented

in the previous two subsections if certain assumptions are made, the details of which are contained in Burgard and Kjaer [7].

7 Conclusion

The adaptions of the BSM model required post the 2008 credit crisis are hotly debated amongst academics and practitioners. There is an intense controversy in the financial quantitative industry regarding inclusion of FVA when pricing financial instruments, as it could be argued that inclusion of FVA violates the law of one price.

The 2008 crisis has had a massive effect on derivative pricing and has plagued our markets with uncertainty. There is no general consensus about how to price a derivative trade after the 2008 credit crisis. This uncertainty has presented regulators, practitioners and academics with new challenges around financial markets. If all the market participants share ideas, then these challenges could possibly be overcome.

The debate continuous.

Acknowledgments The authors would like to thank Carlous Reinecke for the helpful discussions. The second named author was supported by the NRF (Grant Number 87502).

References

1. Alexander, C.: Quantitative Methods in Finance, Market Risk Analysis, vol. 1. Wiley, Hoboken (2008)
2. Bergman, Y.Z.: Option pricing with differential interest rates. Rev. Financ. Stud. **8**, 475–500 (1995)
3. Black, F., Scholes, M.: The pricing of options and corporate liabilities. J. Polit. Econ. **81**, 637–654 (1973)
4. Brigo, D., Capponi, A.: Bilateral counterparty risk valuation with stochastic dynamical models and application to CDSs, SSRN working paper (2008)
5. Brigo, D., Perini, D., Pallavicini, A.: Funding, collateral and hedging: uncovering the mechanics and the subtleties of funding valuation adjustments, SSRN working paper (2011)
6. Burgard, C., Kjaer, M.: Partial differential equation representations of derivatives with bilateral counterparty risk and funding costs. J. Credit Risk **7**, 75–93 (2011)
7. Burgard, C., Kjaer, M.: Generalised CVA with funding and collateral via semi-replication, SSRN working paper (2012)
8. Carver, L.: Show me the money: banks explore DVA hedging. Risk Mag. **25**, 35–37 (2012)
9. Castagna, A.: Yes, FVA is a cost for derivatives desks, SSRN working paper, IASON Ltd (2012)
10. Cox, J., Ross, S., Rubinstein, M.: Option pricing: a simplified approach. J. Financ. Econ. **7**, 229–263 (1979)
11. Crépey, S.: Bilateral counterparty risk under funding constraints part I: pricing, forthcoming in Mathematical Finance (2012)
12. Crépey, S.: Bilateral counterparty risk under funding constraints Part II: CVA, Forthcoming in Mathematical Finance (2012)
13. Duffie, D., Huang, M.: Swap rates and credit quality. J. Financ. **51**, 921–950 (1996)
14. Gregory, J.: Being two faced over counterparty credit risk. Risk **20**, 86–90 (2009)

15. Gregory, J.: Counterparty Credit Risk and Credit Value Adjustment—A Continuing Challenge for Global Financial Markets, 2nd edn. Wiley, London (2012)
16. Gregory, J., German, I.: Closing out DVA? SSRN working paper, Barclays, London (2012)
17. Harrison, M., Pliska, S.: Martingales and stochastic integrals in the theory of continuous trading, stochastic processes and their applications. Stoch. Process. Appl. **11**, 215–260 (1981)
18. Hull, J.: Options, Futures, and Other Derivatives, Harlow 8th edn. Pearson Education Limited, Upper Saddle River (2012)
19. Hull, J., White, A.: The FVA debate. Risk (25th anniversary edition) (2012)
20. Hull, J., White, A.: The FVA debate continued. Risk **10** (2012)
21. Hull, J., White, A.: LIBOR versus OIS: the derivatives discounting dilemma. J. Invest. Manag. **11**, 14–27 (2013)
22. Hull, J., White, A.: Collateral and credit issues in derivatives pricing, SSRN working paper (2013)
23. Hunzinger, C., Labuschagne, C.C.A.: The Cox, Ross and Rubinstein tree model which includes counterparty credit risk and funding costs. N. Am. J. Econ. Financ. **29**, 200–217 (2014)
24. Jarrow, R., Turnbull, S.: Pricing options on financial securities subject to default risk. J. Financ. **1**, 53–86 (1995)
25. Laughton, S., Vaisbrot, A.: In defense of FVA—a response to Hull and White. Risk **25**, 18–24 (2012)
26. Merton, R.: Theory of rational option pricing. Bell. J. Econ. Manag. Sci. **4**, 141–183 (1973)
27. Piterbarg, V.: Funding beyond discounting: collateral agreements and derivatives pricing, Risk Mag. 97–102 (2010)
28. Rubinstein, M., Cox, J.C.: Options Market, 1st edn. Prentice-Hall, New York (1985)
29. Sorensen, E., Bollier, T.: Pricing swap default risk. Financ. Anal. J. **50**, 23–33 (1994)
30. Steland, A.: Financial Statistics and Mathematical Finance Methods, Models and Applications. Wiley, Singapore (2012)

What if We Only Have Approximate Stochastic Dominance?

Vladik Kreinovich, Hung T. Nguyen and Songsak Sriboonchitta

Abstract In many practical situations, we need to select one of the two alternatives, and we do not know the exact form of the user's utility function—e.g., we only know that it is increasing. In this case, stochastic dominance result says that if the cumulative distribution function (cdf) corresponding to the first alternative is always smaller than or equal to the cdf corresponding to the second alternative, then the first alternative is better. This criterion works well in many practical situations, but often, we have situations when for most points, the first cdf is smaller but at some points, the first cdf is larger. In this paper, we show that in such situations of approximate stochastic dominance, we can also conclude that the first alternative is better—provided that the set of points x at which the first cdf is larger is sufficiently small.

1 Stochastic Dominance: Reminder and Formulation of the Problem

In finance, we need to make decisions under uncertainty. In financial decision making, we need to select one of the possible decisions: e.g., whether we sell or buy a given financial instrument (share, option, etc.). Ideally, we should select a decision which leaves us with the largest monetary value x. However, in practice, we cannot predict exactly the monetary consequences of each action: because of the changing

V. Kreinovich (✉)
Department of Computer Science, University of Texas at El Paso, 500 W. University, El Paso, TX 79968, USA
e-mail: vladik@utep.edu

H.T. Nguyen
Department of Mathematical Sciences, New Mexico State University, Las Cruces, NM 88003, USA
e-mail: hunguyen@nmsu.edu

H.T. Nguyen · S. Sriboonchitta
Faculty of Economics, Chiang Mai University, Chiang Mai, Thailand

S. Sriboonchitta
e-mail: songsak@econ.chiangmai.ac.th

V.-N. Huynh et al. (eds.), *Econometrics of Risk*, Studies in Computational Intelligence 583, DOI 10.1007/978-3-319-13449-9_4

external circumstances, in similar situations the same decision can lead to gains and to losses. Thus, we need to make a decision in a situation when we do not know the exact consequences of each action.

In finance, we usually have probabilistic uncertainty. Numerous financial transactions are occurring every moment. For the past transactions, we know the monetary consequences of different decisions. By analyzing these past transactions, we can estimate, for each decision, the frequencies with which this decision leads to different monetary outcomes x. When the sample size is large—and for financial transactions it is large—the corresponding frequencies become very close to the actual probabilities. Thus, in fact, we can estimate the probabilities of different values x.

Comment. Strictly speaking, this is not always true: we may have new circumstances, we can have a new financial instrument for which we do not have many records of its use—but in most situations, knowledge of the probabilities is a reasonable assumption.

How to describe the corresponding probabilities. As usual, the corresponding probabilities can be described either by the probability density function $f(x)$ or by the cumulative distribution function $F(t) \stackrel{\text{def}}{=} \text{Prob}(x \le t)$.

If we know the probability density function $f(x)$, then we can reconstruct the cumulative distribution function as $F(t) = \int_{-\infty}^{t} f(x)\,dx$. Vice versa, if we know the cumulative distribution function $F(t)$, we can reconstruct the probability density function as its derivative $f(x) = F'(x)$.

How to make decisions under probabilistic uncertainty: a theoretical recommendation. Let us assume that we have several possible decisions whose outcomes are characterized by the probability density functions $f_1(x)$, $f_2(x)$, ... According to the traditional decision making theory (see, e.g., [3, 5–7]), the decisions of a rational person can be characterized by a function $u(x)$ called *utility function* such that this person always selects a decision with the largest value of expected utility $\int f_i(x) \cdot u(x)\,dx$.

A decision corresponding to the probability distribution function $f_1(x)$ is preferable to the decision corresponding to the probability distribution function $f_2(x)$ if

$$\int f_1(x) \cdot u(x)\,dx > \int f_2(x) \cdot u(x)\,dx,$$

i.e., equivalently, if

$$\int \Delta f(x) \cdot u(x)\,dx > 0,$$

where we denoted $\Delta f(x) \stackrel{\text{def}}{=} f_1(x) - f_2(x)$.

Comment. It is usually assumed that small changes in x lead to small changes in utility, i.e., in formal terms, that the function $u(x)$ is differentiable.

From a theoretical recommendation to practical decision. Theoretically, we can determine the utility function of the decision maker. However, since such a determination is very time-consuming, it is rarely done in real financial situations. As a result, in practice, we only have a partial information about the utility function.

One thing we know for sure if that the larger the monetary gain x, the better the resulting situation; in other words, we know that the utility $u(x)$ grows with x, i.e., the utility function $u(x)$ is increasing.

Often, this is the only information that we have about the utility function. How can we make a decision in such a situation?

How to make decisions when we only know that utility function is increasing: analysis of the problem. When is the integral $\int \Delta f(x) \cdot u(x)\,dx$ positive?

To answer this question, let us first note that while theoretically, we have gains and losses which can be arbitrarily large, in reality, both gains and losses are bounded by some value T. In other words, $f_i(x) = 0$ for $x \le -T$ and for $x \ge T$ and thus,

$$F_i(-T) = \text{Prob}_i(x \le -T) = 0$$

and

$$F_i(T) = \text{Prob}_i(x \le T) = 1.$$

In this case,

$$\int \Delta f(x) \cdot u(x)\,dx = \int_{-T}^{T} \Delta f(x) \cdot u(x)\,dx.$$

Let us now take into account that since $\Delta f(x) = f_1(x) - f_2(x)$, $f_1(x) = F_1'(x)$, and $f_2(x) = F_2'(x)$, we can conclude that $\Delta f(x) = \Delta F'(x)$, where

$$\Delta F(x) \stackrel{\text{def}}{=} F_1(x) - F_2(x).$$

We can therefore apply integration by parts

$$\int_{\ell}^{u} a'(x) \cdot b(x)\,dx = a(x) \cdot b(x)|_{\ell}^{u} - \int_{\ell}^{u} a(x) \cdot b'(x)\,dx,$$

with $a(x) = \Delta f(x)$ and $b(x) = u(x)$, to the above integral. As a result, we get the formula

$$\int_{-T}^{T} \Delta f(x) \cdot u(x)\,dx = \Delta F(x) \cdot u(x)|_{-T}^{T} - \int \Delta F(x) \cdot u'(x)\,dx.$$

Since $F_1(-T) = F_2(-T) = 0$, we have

$$\Delta F(-T) = F_1(-T) - F_2(-T) = 0.$$

Similarly, from $F_1(T) = F_2(T) = 1$, we conclude that

$$\Delta F(T) = F_1(T) - F_2(T) = 0.$$

Thus, the first term in the above expression for integration by parts is equal to 0, and we have

$$\int_{-T}^{T} \Delta f(x) \cdot u(x)\,dx = -\int \Delta F(x) \cdot u'(x)\,dx.$$

We know that the utility function is increasing, so $u'(x) \geq 0$ for all x. Thus, if $\Delta F(x) \leq 0$ for all x—i.e., if $F_1(x) \leq F_2(x)$ for all x—then the difference $\int \Delta f(x) \cdot u(x)\,dx$ is always non-negative and thus, the decision corresponding to the probability distribution function $f_1(x)$ is preferable to the decision corresponding to the probability distribution function $f_2(x)$.

This is the main idea behind *stochastic dominance* (see, e.g., [4, 8]):

Stochastic dominance: summary. If $F_1(x) \leq F_2(x)$ for all x and the utility function $u(x)$ is increasing, then the decision corresponding to the probability distribution function $f_1(x)$ is preferable to the decision corresponding to the probability distribution function $f_2(x)$.

Comments.

- The condition $F_1(x) \leq F_2(x)$ for all x is not only sufficient to conclude that the first alternative is better, it is also necessary. Indeed, if $F_1(x_0) > F_2(x_0)$ for some x_0, then, since both cumulative distribution functions $F_i(x)$ are differentiable and thus, continuous, there exists an $\varepsilon > 0$ such that $F_1(x) > F_2(x)$ for all x from the interval $(x_0 - \varepsilon, x_0 + \varepsilon)$.
  We can then take a utility function which:
  - is equal to 0 for $x \leq x_0 - \varepsilon$,
  - is equal to 1 for $x \geq x_0 + \varepsilon$, and
  - is, e.g., linear for x between $x_0 - \varepsilon$ and $x_0 + \varepsilon$.

  For this utility function, we have

$$\int F_1(x) \cdot u'(x)\,dx > \int F_2(x) \cdot u'(x)\,dx,$$

  and thus,

$$\int f_1(x) \cdot u(x)\,dx = -\int F_1(x) \cdot u'(x)\,dx < -\int F_2(x) \cdot u'(x)\,dx = \int f_2(x) \cdot u(x)\,dx,$$

  so the first alternative is worse.

- Sometimes, we have additional information about the utility function. For example, the same amount of additional money h is more valuable for a poor person than for the rich person. This can be interpreted as saying that for every value $x < y$ and, the increase in utility $u(x+h) - u(x)$ is larger than (or equal to) the increase $u(y+h) - u(y)$. If we take the resulting inequality

$$u(x+h) - u(x) \geq u(y+h) - u(y),$$

  divide both sides by h, and tends h to 0, we conclude that $u'(x) \geq u'(y)$ when $x < y$. In other words, it is reasonable to conclude that the derivative $u'(x)$ of the utility function is decreasing with x—and thus, that its second derivative is negative.
  If this property is satisfied, then we can perform one more integration by parts and get a more powerful criterion for decision making—for situations when we do not know the exact utility function.

What if the stochastic dominance condition is satisfied "almost always": formulation of the problem. Let us return to the simple situation when we only know that utility is increasing, i.e., that $u'(x) \geq 0$. In this case, as we have mentioned, if we know that $F_1(x) \leq F_2(x)$ for *all* x, then the first alternative is better. In many cases, we can use this criterion and make a decision.

However, often, in practice, the inequality $F_1(x) \leq F_2(x)$ holds for "almost all" values x—i.e., it is satisfied for most values x except for the values x from some small interval. Unfortunately, in this case, as we have shown, the traditional stochastic dominance approach does now allow to make any conclusion—even when the interval is really small. It would be nice to be able to make decisions even if we have *approximate* stochastic dominance.

What we plan to do in this paper. In this paper, we show that, under reasonable assumptions, we can make definite decisions even under approximate stochastic dominance—provided, of course, that the deviations from stochastic dominance are sufficiently small.

Comment. A similar—but somewhat different—problem is analyzed in [1], where it is shown that under certain assumptions, approximate stochastic dominance implies that the first alternative is not much worse than the second one—i.e., if we select the first alternative instead of the second one, we may experience losses, but these losses are bounded, and the smaller the size of the area where $F_1(x)$ is larger than $F_2(x)$, the smaller this bound.

2 How to Make Decisions Under Approximate Stochastic Dominance: Analysis of the Problem

Additional reasonable assumptions about the utility function $u(x)$. In the previous text, we used the fact that the utility function $u(x)$ increases with x, i.e., that its derivative $u'(x)$ is non-negative. Theoretically, we are thus allowing situations when

this derivative is extremely small—e.g., equal to 10^{-40}—or, vice versa, extremely large—e.g., equal to 10^{40}.

From the economical viewpoint, however, such too small or too large numbers make no sense. If the derivative is too small, this means that for all practical purposes, the person does not care whether he or she gets more money—which may be true for a monk leading a spiritual life, but not for agents who look for profit. Similarly, if the derivative $u'(x)$ is, for some x, too large, this means that, in effect, the utility function is discontinuous at this x, i.e., that adding a very small amount of money leads to a drastic increase in utility—and this is usually not the case.

These examples show that not only the derivative $u'(x)$ should be non-negative, it cannot be too small and it cannot be too large. In other words, there should be some values $0 < s < L$ for which

$$s \le u'(x) \le L$$

for all x.

This additional assumption helps us deal with situation of approximate stochastic dominance. Let us show that the above additional assumption $0 < s \le u'(x) \le L$ enables us to deal with approximate stochastic dominance. Indeed, we want to make sure that

$$\int \Delta F(x) \cdot u'(x)\, dx \le 0.$$

In the case of stochastic dominance, we have $\Delta F(x) \le 0$ for all x, but we consider the case of *approximate* stochastic dominance, when $\Delta F(x) > 0$ for some values x. To deal with this situation, let us represent the desired integral as the sum of the two component integrals:

- an integral over all the values x for which $\Delta F(x) \le 0$, and
- an integral over all the values x for which $\Delta F(x) > 0$:

$$\int \Delta F(x) \cdot u'(x)\, dx = \int\limits_{x:\Delta F(x)\le 0} \Delta F(x) \cdot u'(x)\, dx + \int\limits_{x:\Delta F(x)>0} \Delta F(x) \cdot u'(x)\, dx.$$

We want to prove that the sum of these two component integrals is bounded, from above, by 0. To prove this, let us find the upper bound for both integrals.

For the values x for which $\Delta F(x) \le 0$, the largest possible value of the product $\Delta F(x) \cdot u'(x)$ is attained when the derivative $u'(x)$ is the smallest possible—i.e., when this derivative is equal to s. Thus, we conclude that

$$\Delta F(x) \cdot u'(x) \le s \cdot \Delta F(x).$$

Therefore,

$$\int\limits_{x:\Delta F(x)\le 0} \Delta F(x) \cdot u'(x)\, dx \le s \cdot \int\limits_{x:\Delta F(x)\le 0} \Delta F(x)\, dx.$$

Since $\Delta F(x) \le 0$, we have $\Delta F(x) = -|\Delta F(x)|$ and thus,

$$\int\limits_{x:\Delta F(x)\le 0} \Delta F(x)\cdot u'(x)\,dx \le -s\cdot \int\limits_{x:\Delta F(x)\le 0} |\Delta F(x)|\,dx.$$

For the values x for which $\Delta F(x) > 0$, the largest possible value of the product $\Delta F(x)\cdot u'(x)$ is attained when the derivative $u'(x)$ is the largest possible—i.e., when this derivative is equal to L. Thus, we conclude that

$$\Delta F(x)\cdot u'(x) \le L\cdot \Delta F(x).$$

Therefore,

$$\int\limits_{x:\Delta F(x)>0} \Delta F(x)\cdot u'(x)\,dx \le L\cdot \int\limits_{x:\Delta F(x)>0} \Delta F(x)\,dx.$$

By combining the bounds on the two component integrals, we conclude that

$$\int \Delta F(x)\cdot u'(x)\,dx \le -s\cdot \int\limits_{x:\Delta F(x)\le 0} |\Delta F(x)|\,dx + L\cdot \int\limits_{x:\Delta F(x)>0} \Delta F(x)\,dx.$$

The integral $\int \Delta F(x)\cdot u'(x)\,dx$ is non-positive if the right-hand side bound is non-positive, i.e., if

$$-s\cdot \int\limits_{x:\Delta F(x)\le 0} |\Delta F(x)|\,dx + L\cdot \int\limits_{x:\Delta F(x)>0} \Delta F(x)\,dx \le 0,$$

i.e., equivalently, if

$$L\cdot \int\limits_{x:\Delta F(x)>0} \Delta F(x)\,dx \le s\cdot \int\limits_{x:\Delta F(x)\le 0} |\Delta F(x)|\,dx,$$

or

$$\int\limits_{x:\Delta F(x)>0} \Delta F(x)\,dx \le \frac{s}{L}\cdot \int\limits_{x:\Delta F(x)\le 0} |\Delta F(x)|\,dx.$$

This condition is satisfied when the set of all the values x for which $\Delta F(x) > 0$ is small—in this case the integral over this set is also small and thus, smaller than the right-hand side.

Let us describe the resulting criterion in precise terms.

3 How to Make Decisions Under Approximate Stochastic Dominance: Main Result

Formulation of the problem. We have two alternatives, characterized by the cumulative distribution functions $F_1(x)$ and $F_2(x)$. We need to decide which of these two alternatives is better.

What we know about the utility function $u(x)$. We know that the utility function $u(x)$ describing the agent's attitude to different monetary values x is non-decreasing: $u'(x) \geq 0$. Moreover, we assume that we know two positive numbers $s < L$ such that for every x, we have

$$s \leq u'(x) \leq L.$$

Stochastic dominance: reminder. If $F_1(x) \leq F_2(x)$ for all x, i.e., if $\Delta F(x) \leq 0$ for all x (where we denoted $\Delta F(x) = F_1(x) - F_2(x)$), then the first alternative is better.

New criterion for the case of approximate stochastic dominance. If $\Delta F(x) > 0$ for some values x, but the set of all such x is small, in the sense that

$$\int\limits_{x:\Delta F(x)>0} \Delta F(x)\,dx \leq \frac{s}{L} \cdot \int\limits_{x:\Delta F(x)\leq 0} |\Delta F(x)|\,dx,$$

then the first alternative is still better.

Comments.

- It is interesting that a similar expression appears in another context: namely, in the study of different notions of transitivity of stochastic relations; see, e.g., [2]. Indeed, adding $\int_{x:\Delta F(x)\leq 0} |\Delta F(x)|\,dx$ to both sides of the above inequality, and taking into account that the resulting integral in the left-hand side is simply an integral of $|\Delta F(x)| = |F_2(x) - F_1(x)|$ over all possible x, we conclude that

$$\int |F_2(x) - F_1(x)|\,dx \leq \left(1 + \frac{s}{L}\right) \cdot \int\limits_{x:F_2(x)-F_1(x)>0} (F_2(x) - F_1(x))\,dx.$$

  The right-hand side of the new inequality can be described as the interval, over all possible x, of the function $(F_2(x) - F_1(x))_+$, where, as usual, for any function $f(x)$, its positive part $f_+(x)$ is defined as $f_+(x) \stackrel{\text{def}}{=} \max(f(x), 0)$. Thus, this inequality can be represented as

$$\int |F_2(x) - F_1(x)|\,dx \leq \left(1 + \frac{s}{L}\right) \cdot \int (F_2(x) - F_1(x))_+\,dx,$$

or, equivalently, as

$$\frac{\int (F_2(x) - F_1(x))_+ \, dx}{\int |F_2(x) - F_1(x)| \, dx} \geq \frac{1}{1 + \dfrac{s}{L}}.$$

The left-hand side of this inequality is known as the *Proportional Expected Difference*, it is used in several results about transitivity [2].

- The same idea can extend the stochastic dominance criterion corresponding to $u''(x) \leq 0$ to the case when this criterion is satisfied for "almost all" values x.

Acknowledgments We acknowledge the partial support of the Center of Excellence in Econometrics, Faculty of Economics, Chiang Mai University, Thailand. This work was also supported in part by the National Science Foundation grants HRD-0734825 and HRD-1242122 (Cyber-ShARE Center of Excellence) and DUE-0926721. The authors are thankful to Bernard de Baets for valuable discussions, and to the anonymous referees for valuable suggestions.

References

1. Bruni, R., Cesarone, F., Scozzari, A., Tardella, F.: A new stochastic dominance approach to enhanced index tracking problems. Econ. Bull. **32**(4), 3460–3470 (2012)
2. de Baets, B.: State-of-the-art in reciprocal relations. In: Proceedings of the International Workshop Information, Uncertainty, and Imprecision WIUI'2014, Olomouc, Czech Republic, 7–11 April 2014
3. Fishburn, P.C.: Nonlinear Preference and Utility Theory. John Hopkins Press, Baltimore (1988)
4. Levy, H.: Stochastic Dominance: Investment Decision Making under Uncertainty. Springer, New York (2006)
5. Luce, R.D., Raiffa, R.: Games and Decisions: Introduction and Critical Survey. Dover, New York (1989)
6. Nguyen, H.T., Kreinovich, V., Wu, B., Xiang, G.: Computing Statistics Under Interval and Fuzzy Uncertainty. Springer, Berlin (2012)
7. Raiffa, H.: Decision Analysis. McGraw-Hill, Columbus (1997)
8. Sriboonchitta, S., Wong, W.-K., Dhompongsa, S., Nguyen, H.T.: Stochastic Dominance and Applications to Finance, Risk and Economics Hardcover. CRC Press, Boca Raton (2009)

From Mean and Median Income to the Most Adequate Way of Taking Inequality into Account

Vladik Kreinovich, Hung T. Nguyen and Rujira Ouncharoen

Abstract How can we compare the incomes of two different countries or regions? At first glance, it is sufficient to compare the mean incomes, but this is known to be not a very adequate comparison: according to this criterion, a very poor country with a few super-rich people may appear to be in good economic shape. A more adequate description of economy is the *median* income. However, the median is also not always fully adequate: e.g., raising the income of very poor people clearly improves the overall economy but does not change the median. In this paper, we use known techniques from group decision making—namely, Nash's bargaining solution—to come up with the most adequate measure of "average" income: geometric mean. On several examples, we illustrate how this measure works.

1 Mean Income, Median Income, What Next?

Mean income and its limitations. At first glance, if we want to compare the economies of two countries or two regions, all we need to do is divide, for each country, the total income by the number of people and compare the resulting values of mean income. If the mean income in country A is larger than the mean income

V. Kreinovich (✉)
Department of Computer Science, University of Texas at El Paso 500 W. University, El Paso, TX 79968, USA
e-mail: vladik@utep.edu

H.T. Nguyen
Department of Mathematical Sciences, New Mexico State University, Las Cruces, New Mexico 88003, USA
e-mail: hunguyen@nmsu.edu

H.T. Nguyen
Faculty of Economics, Chiang Mai University, Chiang Mai, Thailand

R. Ouncharoen
Department of Mathematics, Chiang Mai University, Chiang Mai, Thailand
e-mail: rujira.o@cmu.ac.th

V.-N. Huynh et al. (eds.), *Econometrics of Risk*, Studies in Computational Intelligence 583, DOI 10.1007/978-3-319-13449-9_5

in country B, this means that the economy of country A is in better shape than the economy of country B.

In many cases, this conclusion is indeed justified, but not always. The fact that the mean has limitations can be illustrated by a known joke: "What happens when Bill Gates walks into a bar? On average, everyone becomes a millionaire." This is a joke, but this joke reflects a serious problem: if a billionaire moves into a small and very poor country, the mean income in this country would increase but the country would remain very poor, contrary to the increase in the mean.

In other words, when comparing different economies, we need to take into account not only the total income, but also the degree of inequality in income distribution.

Comment. In technical terms, we would like the proper measure of "average" income to not change much if we add of an *outlier* like Bill Gates. In statistics, the corresponding property of statistical estimates is known as *robustness*; see, e.g., [9, 29]. In these terms, the main problem of the mean is that it is not robust.

Medium income: a more adequate measure. To avoid the above problem, economists proposed several alternatives to the mean income. The most widely used alternative is the *median* income, i.e., the income level for which the income of exactly half of the population is above this level—and the income of the remaining half is below this level. For example, this is how the Organization for Economic Cooperation and Development (OECD) compares economies of different countries: by listing both their mean incomes and their median incomes; see, e.g., [23].

Median resolves some of the problems related to mean: for example, when Bill Gates walks into a bar, the mean income of people in the bar changes drastically, but the median does not change much.

Comment. The main problem with the mean, as we have mentioned, is that the mean is not robust. From this viewpoint, median—a known robust alternatives to the mean [9, 29]—seems a reasonable replacement of the mean.

Limitations of the median and remaining practical problem. While the median seems to be a more adequate measure of "average" income than the mean, it is not a perfect measure. For example, if the incomes of all the people in the poorer half increase—but do not exceed the previous median—the median remains the same. This is not a very adequate measure for governments that try to lift people out of poverty. Similarly, if the income of the poorer half drastically decreases, we should expect the adequate measure of "average" income to decrease—but the median remains unchanged.

Comment. After we reformulated the problem with mean in terms of robustness, a reader may be under the impression that robustness is all we seek. Alas, the above limitation shows that the problem of finding an appropriate measure of "average" income goes beyond robustness; namely:

- the main problem of mean is that it is *not robust*—it changes too much when we would like to change it a little bit;

- however, while the median *is* robust, it has another problem—it is "too robust": it changes too little (actually, not at all) when we would like it to change.

This example shows that we cannot solve our problem by simply reducing it to a known statistical problem of designing robust estimates, we do need to solve the original problem of estimating the "average" income.

How this practical problem is resolved now. At present, economists propose different heuristic measures of "average" income which are supposedly more adequate than mean and median. There is no absolutely convincing arguments in favor of this or that measure; as a result, researchers use emotional and ideological arguments; see, e.g., [24].

What we do in this paper. In this paper, we show that under some reasonable conditions, it is possible to find the most adequate way how to take inequality into account when gauging the "average" income.

2 Analysis of the Problem and the Resulting Measure

The problem of gauging "average" income can be viewed as a particular case of a problem of group decision making. For the problem of gauging "average" income—when taken "as is"—there is no immediate solution yet. Let us show, however, that this gauging problem can be reformulated in terms of a problem for which many good solutions have been developed—namely, the problem of group decision making.

To explain this reformulation, let us start with the simplest possible case our main problem: the case when in each of the two compared regions, there is perfect equality: all the people in the first region have the same income x, and all the people in the second region have the same income y. In this case clearly:

- if $x > y$, this means that the first region is in better economic shape, and
- if $x < y$, this means that the second region is in better economic shape.

What if we consider a more realistic case of inequality, when people in the first region have, in general, different incomes $x_1, \ldots, x_n$, and people in the second area also have, in general, different incomes $y_1, \ldots, y_m$? How can we then compare the two regions?

A natural idea is to reduce this comparison to the case when all the incomes are equal. In other words:

- first, we find the value x such that for the group of all the people from the first region, incomes $x_1, \ldots, x_n$ are equivalent—in terms of group decision making—to all of them getting the same income x;
- then, we find the value y such that for the group of all the people from the second region, incomes $y_1, \ldots, y_n$ are equivalent—in terms of group decision making—to all of them getting the same income y;

- finally, we compare the resulting values x and y: if $x > y$, then the first economy is in better shape, otherwise, if $x < y$, the second economy is in better shape.

Comment. Our main idea is to reduce the econometric problem of finding an adequate measure for "average" income" to a game-theoretic problem of cooperative group decision making. This idea is in line with the emerging view that game theory—which was originally invented as a *general* theory of group behavior—should be (and can be) successfully applied not only to situations of conflicting competition, but also to more general problems of economics—in particular, to problems of financial econometrics.

From the idea to the algorithm. To transform the above idea to the algorithm, let us recall a reasonable way to perform group decision making. In group decision making, we need to order situations with different individual incomes. To be more precise, in group decision making, we consider situations with different individual *utility values* $u_1, \ldots, u_n$—since different people value different income levels differently; see, e.g., [6, 14, 22, 26]. In this case, as shown by the Nobelist John Nash, under some reasonable assumptions, the most adequate solution is to select the alternative for which the product of the utilities $\prod_{i=1}^{n} u_i$ is the largest possible; see, e.g., [14, 21, 26].

The utility is usually proportional to a power of the money: $u_i = C_i \cdot x_i^a$ for some $a \approx 0.5$; see, e.g., [11–13]. Substituting these utility values into Nash's formula, we get the product $\prod_{i=1}^{n} C_i \cdot \prod_{i=1}^{n} x_i^a$. In these terms, to find the value x for which the selection $(x_1, \ldots, x_n)$ is equivalent to x, we must find x for which

$$\prod_{i=1}^{n} C_i \cdot \prod_{i=1}^{n} x_i^a = \prod_{i=1}^{n} C_i \cdot \prod_{i=1}^{n} x^a.$$

Dividing both sides of this equality by the constant $\prod_{i=1}^{n} C_i$ and extracting power a from both sides, we conclude that $\prod_{i=1}^{n} x_i = \prod_{i=1}^{n} x = x^n$. Thus, the value x which describes the income distribution $(x_1, \ldots, x_n)$ is equal to $x = \sqrt[n]{x_1 \cdot \ldots \cdot x_n}$—the *geometric mean* of the income values. So, we arrive at the following conclusion.

Resulting measure of "average" income which most adequately described "average" income: geometric mean. Suppose that we need to compare the economies of two regions. Let us denote the incomes in the first region by $x_1, \ldots, x_n$ and the incomes in the second region by $y_1, \ldots, y_m$. To perform this comparison, we compute the geometric averages $x = \sqrt[n]{x_1 \cdot \ldots \cdot x_n}$ and $y = \sqrt[m]{y_1 \cdot \ldots \cdot y_m}$ of the two regions; then:

- if $x > y$, we conclude that the first region is in better economic shape, and
- if $x < y$, we conclude that the second region is in better economic shape.

From the mathematical viewpoint, comparing geometric means $\overline{x}$ and $\overline{y}$ is equivalent to comparing the logarithms of these means. Here,

$$\ln(\overline{x}) = \ln(\sqrt[n]{x_1 \cdot \ldots \cdot x_n}) = \frac{\ln(x_1) + \cdots + \ln(x_n)}{n}.$$

Thus, the logarithm of the geometric mean $\overline{x}$ is equal to mean value $E[\ln(x)]$ of the logarithm of the income—and therefore,

$$\overline{x} = \exp(E[\ln(x)]) = \exp\left(\int \ln(x) \cdot f(x)\, dx\right).$$

So, to compare the economies in two different regions, we need to compare the mean values $E[\ln(x)]$ of the logarithm of the income x in these regions.

Relation between the new measure and the mean income: an observation. It is well known that the geometric mean is always smaller than or equal than the arithmetic mean, and they are equal if and only if all the numbers are equal; see, e.g., [1, 32].

Thus, the new measure of "average" income is always smaller than or equal to the mean income, and it is equal to the mean income if and only if all the individual incomes are the same—i.e., if and only if we have perfect equality.

3 First Example of Using the New Measure of "Average" Income: Case of Low Inequality

Case of low inequality: informal description. Let us first consider the case when inequality is low, i.e., when most people have a reasonable income, and the proportion of very poor and very rich people is not that large.

Towards a formal description. The fact that most incomes are close to one another means that most of these incomes are close to the mean income μ. In mathematical statistics, deviations from the mean are usually described by the standard deviation σ; see, e.g., [30]. In these terms, low inequality means that the standard deviation σ is small. Let us analyze what happens in this case.

Case of low inequality: analysis of the problem. As we have mentioned, the new inequality measure has the form $\overline{x} = \exp(E[\ln(x)])$. Thus, to compare the economies in two different regions, we need to compare the mean values $E[\ln(x)]$ of the logarithm of the income x in these regions.

Since the deviations from the mean $x - \mu$ are relatively small, we have can substitute $x = \mu + (x - \mu)$ into the formula for $E[\ln(x)]$ and ignore higher order terms in the expansion in $x - \mu$. According to the Taylor series for the logarithm, we have:

$$\ln(x) = \ln(\mu + (x - \mu)) = \ln(\mu) + \frac{1}{\mu} \cdot (x - \mu) - \frac{1}{2\mu^2} \cdot (x - \mu)^2 + \cdots$$

By taking the mean value of both sides and taking into account that $E[x - \mu] = \mu - \mu = 0$ and that $E[(x - \mu)^2] = \sigma^2$, we conclude that

$$E[\ln(x)] = \ln(\mu) - \frac{1}{2\mu^2} \cdot \sigma^2 + \cdots$$

Since we assumed that the deviations of x from μ are small, we can preserve only the first terms which shows the dependence on these deviations and ignore higher order terms in this expansion. As a result, we get an approximate formula

$$E[\ln(x)] \approx \ln(\mu) - \frac{\sigma^2}{2\mu^2}.$$

Thus, for $\overline{x} = \exp(E[\ln(x)])$, we get

$$\overline{x} = \exp(E[\ln(x)]) \approx \exp\left(\ln(\mu) - \frac{\sigma^2}{2\mu^2}\right) = \exp(\ln(\mu)) \cdot \exp\left(-\frac{\sigma^2}{2\mu^2}\right).$$

The first factor is equal to μ. To estimate the second factor, we can again use the fact that σ is small; in this case, we can expand the function $\exp(z)$ in Taylor series and keep only the first term depending on σ:

$$\exp\left(-\frac{\sigma^2}{2\mu^2}\right) = 1 - \frac{\sigma^2}{2\mu^2} + \cdots \approx 1 - \frac{\sigma^2}{2\mu^2}.$$

Substituting this expression into the above formula for $\overline{x} = \exp(E[\ln(x)])$, we conclude that

$$\overline{x} = \mu \cdot \left(1 - \frac{\sigma^2}{2\mu^2}\right) = \mu - \frac{\sigma^2}{2\mu}.$$

Thus, we arrive at the following conclusion.

Resulting formula. In the case of low inequality, the "average" income is equal to

$$\overline{x} = \mu - \frac{\sigma^2}{2\mu},$$

where μ is the average income and σ is the standard deviation.

Analysis of this formula. The larger the inequality, the larger the standard deviation σ, and the less preferable is the economy. The above formula provides an exact quantitative description of this natural qualitative idea.

Comments.

- The new measure takes inequality into account, and it avoids the ideological ideas of weighing inequality too much: if an increase in the mean income comes at the expense of an increase in inequality, this is OK, as long as the above combination of means and standard deviation increases.
- This example is one of the cases which shows that the new measure is more adequate than, e.g., the median. For example, if the incomes are normally distributed, then the median simply coincides with the mean, and so, contrary to our intuitive expectations, the increase in inequality does not worsen the median measure of economics. In contrast, the new measure does go down when inequality increases.

4 Second Example of Using the New Measure of "Average" Income: Case of a Heavy-Tailed Distribution

Heavy-tailed (usually, Pareto) distributions are ubiquitous in economics. In the 1960s, Benoit Mandelbrot, the author of fractal theory, empirically studied the price fluctuations and showed [15] that large-scale fluctuations follow the Pareto power-law distribution, with the probability density function $f(x) = A \cdot x^{-\alpha}$ for $x \geq x_0$, for some constants $\alpha \approx 2.7$ and x_0. For this distribution, variance is infinite. The above empirical result, together with similar empirical discovery of heavy-tailed laws in other application areas, has led to the formulation of *fractal theory*; see, e.g., [16, 17].

Since then, similar Pareto distributions have been empirically found in other financial situations [3–5, 7, 18, 20, 25, 28, 31, 33, 34], and in many other application areas [2, 8, 16, 19, 27].

Formulation of the problem. Let us consider the situations when the income distribution follows Pareto law, with probability density $f(x)$ equal to 0 for $x \leq x_0$ and to $A \cdot x^{-\alpha}$ for $x \geq x_0$.

Once we know x_0 and α, we can determine the parameter A from the condition that $\int f(x)\,dx = 1$. For the above expression, this condition leads to $A \cdot \dfrac{x_0^{-(\alpha-1)}}{\alpha - 1} = 1$, hence $A = (\alpha - 1) \cdot x_0^{\alpha-1}$.

For this distribution, we want to compute the mean income, the median income, and the newly defined "average" income.

Mean income. The mean income is equal to $\mu = \int x \cdot f(x)\,dx$, i.e., for the Pareto distribution, to

$$\int_{x_0}^{\infty} A \cdot x^{1-\alpha}\,dx = A \cdot \frac{x^{2-\alpha}}{2-\alpha}\Big|_{x_0}^{\infty} = A \cdot \frac{x_0^{2-\alpha}}{\alpha - 2}.$$

Substituting the above value of A, we conclude that the mean is equal to

$$\mu = \frac{\alpha - 1}{\alpha - 2} \cdot x_0.$$

Median income. The median income m can be determined from the condition that $\int_m^\infty f(x)\,dx = \frac{1}{2}$. For the Pareto distribution, this means

$$\int_m^\infty A \cdot x^{-\alpha}\,dx = A \cdot \frac{m^{-(\alpha-1)}}{\alpha - 1} = \frac{1}{2}.$$

Substituting the above expression for A into this formula, we conclude that $\frac{m^{-(\alpha-1)}}{x_0^{-(\alpha-1)}} = \frac{1}{2}$, hence $\frac{m^{\alpha-1}}{x_0^{\alpha-1}} = 2$, and $m = x_0 \cdot 2^{1/(\alpha-1)}$.

New measure of "average" income. For the new measure of average income $\overline{x}$, its logarithm is equal to the expected value of $\ln(x)$:

$$\ln(\overline{x}) = \int \ln(x) \cdot f(x)\,dx = \int_{x_0}^\infty \ln(x) \cdot A \cdot x^{-\alpha}\,dx.$$

This integral can be computed by integration by part; so, we get

$$\begin{aligned}
\ln(\overline{x}) &= \ln(x) \cdot \frac{A \cdot x^{1-\alpha}}{1-\alpha}\Bigg|_{x_0}^\infty - \int_{x_0}^\infty \frac{1}{x} \cdot \frac{A \cdot x^{1-\alpha}}{1-\alpha}\,dx \\
&= \ln(x_0) \cdot \frac{A \cdot x_0^{-(\alpha-1)}}{\alpha - 1} - \int_{x_0}^\infty \frac{A \cdot x^{-\alpha}}{1-\alpha}\,dx \\
&= \ln(x_0) \cdot \frac{A \cdot x_0^{-(\alpha-1)}}{\alpha - 1} - \frac{A \cdot x^{-(\alpha-1)}}{(1-\alpha)^2}\Bigg|_{x_0}^\infty \\
&= \ln(x_0) \cdot \frac{A \cdot x_0^{-(\alpha-1)}}{\alpha - 1} + \frac{A \cdot x_0^{-(\alpha-1)}}{(1-\alpha)^2}.
\end{aligned}$$

Substituting the expression $A = (\alpha - 1) \cdot x_0^{-(\alpha-1)}$ into this formula, we get

$$\ln(\overline{x}) = \ln(x_0) + \frac{1}{\alpha - 1},$$

hence

$$\overline{x} = \exp(\ln(\overline{x})) = x_0 \cdot \exp\left(\frac{1}{\alpha - 1}\right).$$

Comment. When $\alpha \to \infty$, the distribution tends to be concentrated on a single value x_0—i.e., we have the case of absolute equality. In this case, as expected, all three characteristics—the mean, the median, and the new geometric mean—tends to the same value x_0.

5 Auxiliary Result: The New Measure of "Average" Income May Explain the Power-Law Character of Income Distribution

In the previous section, we analyzed how the new measure of "average" income $\overline{x} = \exp(\int \ln(x) \cdot f(x)\,dx)$ behaves in situations when the income distribution follows a power law.

Interestingly, the power law itself can be derived based on this inequality measure. Indeed, suppose that all we know about the income distribution is the value $\overline{x}$ and the lower bound $\delta > 0$ on possible incomes (this lower bound reflects the fact that a human being needs some minimal income to survive). There are many possible probability distributions $f(x)$ which are consistent with this information. In such situation, out of all such distributions, it is reasonable to select a one for which the entropy $S \stackrel{\text{def}}{=} -\int f(x) \cdot \ln(f(x))\,dx$ is the largest; see, e.g., [10].

To find the distribution that maximizes the entropy S under the constraints $\exp\left(\int \ln(x) \cdot f(x)\,dx\right) = \overline{x}$ and $\int f(x)\,dx = 1$, we can use the Lagrange multiplier technique that reduces this constraint optimization problem to the unconstrained problem of optimizing a functional

$$-\int f(x) \cdot \ln(f(x))\,dx + \lambda_1 \cdot \left(\exp\left(\int \ln(x) \cdot f(x)\,dx\right) - \overline{x}\right) + \lambda_2 \cdot \left(\int f(x)\,dx - 1\right),$$

for appropriate Lagrange multipliers λ_i. Differentiating this expression with respect to $f(x)$ and equating the derivative to 0, we conclude that

$$-\ln(f(x)) - 1 + \lambda_1 \cdot C \cdot \ln(x) + \lambda_2 = 0,$$

where $C \stackrel{\text{def}}{=} \exp(\int \ln(x) \cdot f(x)\,dx)$ and thus $C = \overline{x}$. Thus,

$$\ln(f(x)) = (\lambda_2 - 1) + \lambda_1 \cdot \overline{x} \cdot \ln(x).$$

Applying the function $\exp(z)$ to both sides of this equality, we conclude that $f(x) = A \cdot x^{-\alpha}$, where $A = \exp(\lambda_2 - 1)$ and $\alpha = -\lambda_1 \cdot \overline{x}$. So, we indeed get the empirically observed power law for income distribution.

Acknowledgments This work was supported in part by Chiang Main University, and also by the US National Science Foundation grants HRD-0734825 and HRD-1242122 (Cyber-ShARE Center of Excellence) and DUE-0926721. The authors are thankful to the anonymous referees for valuable suggestions.

References

1. Beckenbach, E.F., Bellman, R.: Introduction to Inequalities. Mathematical Association of America, New York (1975)
2. Beirlant, J., Goegevuer, Y., Teugels, J., Segers, J.: Statistics of Extremes: Theory and Applications. Wiley, Chichester (2004)
3. Chakrabarti, B.K., Chakraborti, A., Chatterjee, A.: Econophysics and Sociophysics: Trends and Perspectives. Wiley-VCH, Berlin (2006)
4. Chatterjee, A., Yarlagadda, S., Chakrabarti, B.K.: Econophysics of Wealth Distributions. Springer, Milan (2005)
5. Farmer, J.D., Lux, T. (eds.): Applications of statistical physics in economics and finance. A special issue of the J. Econ. Dyn. Control, **32**(1), 1–320 (2008)
6. Fishburn, P.C.: Nonlinear Preference and Utility Theory. John Hopkins Press, Baltimore (1988)
7. Gabaix, X., Parameswaran, G., Vasiliki, P., Stanley, H.E.: Understanding the cubic and half-cubic laws of financial fluctuations. Phys. A **324**, 1–5 (2003)
8. Gomez, C.P., Shmoys, D.B.: Approximations and randomization to boost CSP techniques. Ann. Oper. Res. **130**, 117–141 (2004)
9. Huber, P.J.: Robust Statistics. Wiley, Hoboken (2004)
10. Jaynes, E.T.: Probability Theory: The Logic of Science. Cambridge University Press, Cambridge (2003)
11. Kahneman, D.: Thinking: Fast and Slow. Farrar, Straus, and Girroux, New York (2011)
12. Kahneman, D., Tversky, A.: Advances in prospect theory: cumulative representation of uncertainty. J. Risk Uncertainty **5**, 297–324 (1992)
13. Kozielecki, J.: Psychologicl Decision Theory. Kluwer, Dordrecht (1981)
14. Luce, R.D., Raiffa, R.: Games and Decisions: Introduction and Critical Survey. Dover, New York (1989)
15. Mandelbrot, B.: The variation of certain speculative prices. J. Bus. **36**, 394–419 (1963)
16. Mandelbrot, B.: The Fractal Geometry of Nature. Freeman, San Francisco (1983)
17. Mandelbrot, B., Hudson, R.L.: The (Mis)behavior of Markets: A Fractal View of Financial Turbulence. Basic Books, New York (2006)
18. Mantegna, R.N., Stanley, H.E.: An Introduction to Econophysics: Correlations and Complexity in Finance. Cambridge University Press, Cambridge (1999)
19. Markovich, N. (ed.): Nonparametric Analysis of Univariate Heavy-Tailed Data: Research and Practice. Wiley, Chichester (2007)
20. McCauley, J.: Dynamics of Markets, Econophysics and Finance. Cambridge University Press, Cambridge (2004)
21. Nash, J.: The bargaining problem. Econometrica **18**(2), 155–162 (1950)
22. Nguyen, H.T., Kreinovich, V., Wu, B., Xiang, G.: Computing Statistics Under Interval and Fuzzy Uncertainty. Springer, Berlin (2012)
23. Organization for Economic Cooperation and Development (OECD), OECD Statistics (2014) http://stats.oecd.org/
24. Piketty, T.: Capital in the Twenty-First Century. Harvard University Press, Cambridge (2014)
25. Rachev, S.T., Mittnik, S.: Stable Paretian Models in Finance. Wiley, New York (2000)
26. Raiffa, H.: Decision Analysis. McGraw-Hill, Columbus (1997)
27. Resnick, S.I.: Heavy-Tail Phenomena: Probabilistic and Statistical Modeling. Springer, New York (2007)

28. Roehner, B.: Patterns of Speculation—A Study in Observational Econophysics. Cambridge University Press, Cambridge (2002)
29. Rousseeuw, P., Leroy, A.: Robust Regression and Outlier Detection. Wiley, New York (1987)
30. Sheskin, D.J.: Handbook of Parametric and Nonparametric Statistical Procedures. Chapman & Hall/CRC, Boca Raton (2011)
31. Stanley, H.E., Amaral, L.A.N., Gopikrishnan, P., Plerou, V.: Scale invariance and universality of economic fluctuations. Phys. A **283**, 31–41 (2000)
32. Steele, J.M.: The Cauchy-Schwarz Master Class: An Introduction to the Art of Mathematical Inequalities. Cambridge University Press, New York (2004)
33. Stoyanov, S.V., Racheva-Iotova, B., Rachev, S.T., Fabozzi, F.J.: Stochastic models for risk estimation in volatile markets: a survey. Ann. Oper. Res. **176**, 293–309 (2010)
34. Vasiliki, P., Stanley, H.E.: Stock return distributions: tests of scaling and universality from three distinct stock markets. Phys. Rev. E: Stat. Nonlinear Soft Matter Phys. **77**(3), Pt. 2, Publ. 037101 (2008)

Belief Aggregation in Financial Markets and the Nature of Price Fluctuations

Daniel Schoch

Abstract We present a model of financial markets, where the belief of the market, expressed by a normal distribution over asset returns, is formed by aggregating in a dynamically consistent way individual subjective beliefs of the market participants, which are likewise assumed to follow normal distributions. We apply this model to a market of traders with standard CARA preferences with the aim of identifying an intrinsic source of price fluctuations. We find that asset prices depend on both Gaussian parameters mean and variance of the market belief, but argue that the latter changes slower than the former. Consequently, price fluctuations are dominated by the covariance matrix of the market participants' subjective beliefs about expected asset returns.

1 Introduction

Portfolio theory, both modern and post-modern, attempts to model investor decisions solely by the means of preferences and public information. The basis of the decision are statistical information on the past performance of assets, generally given in the form of two parameters, average return, and an indicator of risk. The most simple and classical variant is the Capital Asset Pricing Model. However, it fails to explain stock returns [4]. Results improve when the model is generalized to include time-dependence and considers the correlation with certain prediction variables for stock market returns [6].

All of these rational-expectation models inherit a notion of portfolio dominance in the mean return/risk space [12, 13]. For a given risk-free return rate, the models predict the existence of a unique dominant market portfolio among the risky assets, if the latter form a strictly convex set. Consequently, all investors choose the same risky portfolio, even if their preferences differ. Although this is clearly not a realistic conclusion, the underlying premise of the CAPM that investors have equal expectations is hardly challenged on its own. The main line of criticism goes against the single

D. Schoch (✉)
Nottingham School of Economics, Jalan Broga, 43500 Semenyih, Malaysia
e-mail: Daniel.Schoch@nottingham.edu.my

V.-N. Huynh et al. (eds.), *Econometrics of Risk*, Studies in Computational Intelligence 583, DOI 10.1007/978-3-319-13449-9_6

factor nature of the model. But even in generalized models with multiple factors, the diversity of investment decisions remains unexplained as long as all of these factors stem from the same public indicators.

A different line towards understanding price formation has been outlined in the literature on prediction markets. The asset is a 1$ binary real option on a future event. Investors are distributed according to their subjective belief. An equilibrium condition for the market price can be derived [2]. However, since subjective expected utility maximization is assumed, investors spend their whole budget on the single asset in question and also this model can not account for investment diversity. Moreover, since the model is based on risk-neutral preferences, it is unsuitable for belief aggregation in financial markets, where subjective uncertainty should play a role. It had been noticed that, in contrast to a widespread belief [15], the aggregated probability (average belief) can not be equated with the market price under the assumptions of this model [10]; this only holds if certain forms of risk aversion are introduced [5, 16]. This points towards the significance of including some representation of subjective risk in belief aggregation.

There is a good reason that, when subjective beliefs on risk is included, belief aggregation is indeed more suitable for the financial market than for prediction markets. Real events cause externalities, which make bets on a prediction market suitable as insurance against those events, which influences the prices and causes deviations from the prices they would have if they were mere objects of speculations in itself. Thus prices are more likely to represent beliefs in stock market than in prediction markets.

We present a model of dynamically consistent belief aggregation for stock returns. As a rationality constraint, we introduce dynamic consistency as a no-arbitrage condition. Beliefs about mean returns and volatilities are found to be aggregated in a unique way.

2 Belief Aggregation

Let $\langle \mathscr{A}, X \rangle$ denote a measure space, and m be any (not necessarily probability-) measure. Consider a set P of probability measures with the following properties:

- All $p \in P$ are m-continuous (i.e., can be represented with a measurable density by m).
- P is convex.
- P is closed under conditionalization: If $p \in P$ and $p(A) > 0$, then $p_A \in P$, where p_A denotes the conditional measure limited to the set A.
- For each $x \in X$, the Dirac measure

$$\delta_x(A) = 1_A(x)$$

  for all $A \in \mathscr{A}$ can be approximated by a sequence of measures in P with respect to the weak topology, which is induced by the convergence of expectation values of bounded continuous functions [9].

We call $p^1, \ldots, p^n \in P$ **compatible** if and only if there is an $A \in \mathscr{A}$ such that $p^i(A) > 0$ for all i and $B \in \mathscr{A}$, $B \subseteq A$

$$p^i(B) > 0 \Leftrightarrow p^j(B) > 0$$

for each $i, j \in \{1, \ldots, n\}$. A **probability aggregation function** (of order n) maps each compatible n-tuple $p^1, \ldots, p^n \in P$ to a probability measure $f\left(p^1, \ldots, p^n\right) \in P$.

Definition 1 A probability aggregation function satisfies the **unanimity** condition if and only if

$$f(p, \ldots, p) = p.$$

It satisfies **anonymity** (or **symmetry**) if and only if for each permutation k: $\{1, \ldots, n\} \to \{1, \ldots, n\}$

$$f\left(p^1, \ldots, p^n\right) = f\left(p^{k(1)}, \ldots, p^{k(n)}\right).$$

Although arbitrage possibilities may exist in real markets, we assume that traders are aware of this fact and try to avoid giving others the opportunities to write a Dutch book on their cost due to their diverging beliefs. They have an interest in acting as if there is a unique market belief, such that they can formulate their trading strategy in terms of differences of their own belief relative to the market, avoiding exposing bilateral belief differences to third parties. Moreover, they will assure that new incoming information is being processed in a time-consistent manner ruling out time-dependent Dutch books.

Definition 2 A probability aggregation function f is called **dynamically consistent** if and only if for any event A which is p^i-non-null for $i = 1, \ldots, n$, aggregation and conditionalization commute: if $f\left(p^1, \ldots, p^n\right) = p$, then

$$f\left(p_A^1, \ldots, p_A^n\right) = p_A.$$

Theorem 3 *Let P be the set of Borel probability measures on the real line. A probability aggregation function on P is dynamically consistent if and only if for every compatible set of measures $p^1, \ldots, p^n \in P$ with densities $\rho_1, \ldots, \rho_n$, the density ρ of the aggregated measure $f\left(p^1, \ldots, p^n\right)$ can be written, up to a normalization constant, as a weighted geometrical mixture*

$$\rho \propto (\rho_1)^{\alpha_1} \cdot \cdots \cdot (\rho_n)^{\alpha_n}, \tag{1}$$

where $\alpha_i \geq 0$. The aggregation function satisfies unanimity if and only if $\sum_{i=1}^n \alpha_i = 1$. If, additionally, anonymity holds, there is exactly one probability aggregation function, which is given by the aggregated density by the geometrical average

$$\rho \propto (\rho_1 \cdot \cdots \cdot \rho_n)^{1/n}. \tag{2}$$

The proof of the theorem is given in the appendix. Let us now assume that there are n market participants considering investment in an equity. The participants commonly believe that returns are normally distributed, but disagree on the parameters mean and volatility of the distribution. Just as in the CAPM, investors base their investment decisions solely on mean returns and volatility, but they use both their private subjective beliefs and the aggregated market belief to adjust their portfolio.

From Eq. (1) we obtain immediately that univariate normal distributions $p^1, \ldots, p^n$ with mean μ_i and standard deviation σ_i are aggregated to a normal distribution with mean

$$\bar{\mu} = \sum_{i=1}^{n} \alpha_i \cdot \frac{\bar{\sigma}^2}{\sigma_i^2} \cdot \mu_i \tag{3}$$

and standard deviation $\bar{\sigma}$ satisfying

$$\frac{1}{\bar{\sigma}^2} = \sum_{i=1}^{n} \frac{\alpha_i}{\sigma_i^2}. \tag{4}$$

These could be the common knowledge necessary to circumvent the no-trade theorem [11].
For a symmetric aggregation function, these equations reduce to

$$\bar{\mu} = \frac{1}{n} \sum_{i=1}^{n} \frac{\bar{\sigma}^2}{\sigma_i^2} \cdot \mu_i \tag{5}$$

and

$$\frac{1}{\bar{\sigma}^2} = \frac{1}{n} \sum_{i=1}^{n} \frac{1}{\sigma_i^2}. \tag{6}$$

Equation (3) says that the mean belief of the market about the stock's average return is a weighted average of the mean beliefs of the individuals. The weighting factors

$$\alpha_i \cdot \frac{\bar{\sigma}^2}{\sigma_i^2} \tag{7}$$

are positive and sum up to one according to (4). The first subfactor α_i describes the agent's influence on the market opinion. The second factor $\bar{\sigma}^2/\sigma_i^2$ can be interpreted as the relative subjective certainty of the agent about her belief. The lower person i's subjective uncertainty σ_i^2 is relative to the market uncertainty $\bar{\sigma}^2$, the greater his contribution to the market opinion. If the subjective certainty is greater than one, the subject is more confident in his or her belief μ_i on the return than the average market participant. According to (4), the market certainty equals the weighted average of the individual subjective certainties.

3 The Portfolio

We adopt the most simple framework commonly used for the study of information in financial markets [7, 14]. It consists of a market with two assets, a risky one and a riskless one. We assume trader i to be a Subjective Expected Utility (SEU) maximizer exhibiting constant absolute risk aversion (CARA) utilities of the form

$$U_i(w_i) = -\exp(-a_i \cdot w_i)$$

depending on the wealth w_i earned from the portfolio. The parameter a_i is the agent's degree of risk aversion. Each investor is holding a certain number x_i of the risky asset, which he or she believes to yield normally distributed returns $r_i \sim N(\mu_i, \sigma_i^2)$ during the investment period, as well a certain amount m_i of a risk-free asset with constant return r_f. At the end of the investment period the investor consumes the total return of his portfolio

$$w_i = r_i \cdot x_i + r_f \cdot m_i.$$

The initial budget equality is given by

$$w_{0i} = p \cdot x_i + m_i,$$

with p being the price of the risky asset, in terms of the riskless asset. We rewrite

$$w_i = (r_i - r_f \cdot p) \cdot x_i + r_f \cdot w_{0i}.$$

Maximizing SEU

$$E[U(w_i)] = -\exp\left\{-a_i \cdot \left(E[w_i] - \frac{a_i}{2} \cdot Var[w_i]\right)\right\}$$

we find that the investor's demand of the risky asset x_i is proportional to the perceived expected return in excess of $r \cdot p$ and inversly proportional to the perceived variances of the returns,

$$x_i = \frac{\mu_i - r_f \cdot p}{a_i \cdot \sigma_i^2}. \tag{8}$$

The proportionality factor is a direct function of the degree of risk aversion. The demand of the risky asset turns out to be independent of the initial wealth and the demand m_i of the riskless asset. We allow the demand to become negative indicating short selling.

We now apply this result to our belief aggregation model. The average market demand per investor amounts to

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{1}{n}\sum_{i=1}^{n} \frac{\mu_i}{a_i \cdot \sigma_i^2} - r_f \cdot p \cdot \frac{1}{n}\sum_{i=1}^{n} \frac{1}{a_i \cdot \sigma_i^2}. \tag{9}$$

Setting the agent's opinion weight to

$$\alpha_i = \frac{a}{n \cdot a_i}, \tag{10}$$

by (3) and (4) this can be rewritten as

$$\bar{x} = \frac{\bar{\mu} - r_f \cdot p}{a \cdot \bar{\sigma}^2}. \tag{11}$$

Noting that Eqs. (8) and (11) are identical in form we conclude that a total market of SEU maximizers with CARA utilities and Gaussian beliefs acts as a single CARA agent with Gaussian beliefs. This is a well-known result (see e.g. [7]). What is new here is the insight that this market-representing agent's belief formation as an aggregation of the individual beliefs can be characterized by dynamic consistency. Unanimity requires

$$1 = \sum_{i=1}^{n} \alpha_i \Rightarrow \frac{1}{a} = \frac{1}{n}\sum_{i=1}^{n} \frac{1}{a_i},$$

thus the market's inverse risk aversion turns out to be the average individual inverse risk aversion.

The *i*th trader's demand can now be written as

$$x_i = \frac{a}{a_i} \cdot \frac{\bar{\sigma}^2}{\sigma_i^2} \cdot \bar{x} + \frac{1}{a_i \cdot \sigma_i^2} \cdot (\mu_i - \bar{\mu}) \tag{12}$$

Observe that the only way the demand of the risky asset depends on its price is via the average demand $\bar{x}$. If $\mu_i = \mu$, $\sigma_i = \sigma$, and $a_i = a$, the equation reduces to $x_i = \bar{x}$. Thus, *ceteris paribus*, a trader exhibiting market belief and market risk aversion demands exactly an average share.

Equation (12) can be easily interpreted. The first term represents the decider's choice to buy a certain amount proportional to his or her relative certainty (7) and relative inverse risk aversion $\frac{a}{a_i}$. For simplicity, consider the case where the agent exhibits the market risk aversion $a_i = a$. If she has more confidence in his belief than the market, $\sigma_i < \sigma$, then she acquires a portion greater than the average demand, otherwise she demands a less than average share of the cake. The second term represents a further redistribution of the stock according to the trader's 'excess bullishness' $\mu_i - \bar{\mu}$ and proportional to his absolute certainty $1/\sigma_i^2$ and his inverse risk aversion $1/a_i$. From (3), (4) and (10) it follows that these terms sum up to zero. The trader will buy an additional amount of the stock if he or she believes that the equity is underestimated by the market, $\mu_i > \bar{\mu}$, and will buy less if he considers it overestimated, $\mu_i < \bar{\mu}$.

4 The Nature of Price Fluctuations

In tradition rational expectation models, price varies due to external random sources or demand fluctuations [7, 8]. However useful external random sources are for the study of information, they do not provide any explanation on the nature of price fluctuations. Prices vary according to the changing decisions of the market participants, and any external causes should be mediated through the decider's preferences and beliefs. To gain insight into other potential sources of price fluctuations, we rewrite Eq. (11) to obtain the price of the risky asset,

$$p = \frac{1}{r_f} \cdot \left(\bar{\mu} - a \cdot \bar{\sigma}^2 \cdot \bar{x} \right). \tag{13}$$

The risk-free interest rate r_f and the taste of the investors change only very slowly, and so does the aggregated market risk aversion a, which is a function of the individual risk aversions $a_1, \ldots, a_n$ only. Thus for moderate investment horizonts, r_f and a could be regarded constant.

The aggregated average demand $\bar{x}$ can change only if there is either some elasticity in supply or short selling is allowed. If supply of the risky asset is infinitely inelastic, and short selling is excluded, then market clearance assures that the aggregated supply is constant. This assumption is, for example, used in the Grossman-Stiglitz model on informational inefficiency, where an external source of randomness is directly attached to the returns [7]. But then, in equilibrium, also demand must be constant. We conclude that fluctuations of $\bar{x}$ can in general account for price volatility only as far as it constitutes random deviations from the equilibrium price.

This leaves as potential explananda of volatility only the characteristics of the aggregated market belief, which is given by the three equations

$$\bar{\mu} = \sum_{i=1}^{n} \lambda_i \cdot \mu_i, \tag{14}$$

$$\lambda_i = \frac{1}{n} \cdot \frac{a}{a_i} \cdot \frac{\bar{\sigma}^2}{\sigma_i^2}, \tag{15}$$

$$\sum_{i=1}^{n} \lambda_i = 1. \tag{16}$$

There are three reasons why the fluctuations of $\bar{\mu}$ can be expected to be more influencial than those of $\bar{\sigma}^2$. First, since prices are positive, the first term in (13) must dominate the second. Second, it is psychologically feasible that agents adapt beliefs about future expected returns faster than beliefs about future risks. Beliefs on risks depend on historical volatilities, which need longer time periods to observe than prices themselves. Since the subjective mean returns $\mu_1, \ldots, \mu_n$ enter only equation (14), but not Eqs. (16) and (15), fluctuations of $\bar{\mu}$ should be higher than those of $\bar{\sigma}^2$. Third, changes in beliefs on risks are more likely to follow the market, since

the professional traders have a professional risk management system behind them, while noise traiders, ignorant of historical volatilities, have no reason to change their initial risk assessment. Thus, the dominating term in the volatility of the asset price (13) is the first summand $\bar{\mu}$.

As a first approximation, we can assume that changes of the subjective risks are small and highly correlated, thus λ_i is nearly constant. We then can write

$$\text{Var}\,(p) = \frac{1}{r_f^2}\lambda Q \lambda, \tag{17}$$

where $Q_{ij} = \text{Cov}\left(\mu_i, \mu_j\right)$ is the covariance matrix of the individual expected returns and λ is the vector of weights given by (15).

There are two interesting limited cases. If traders form their beliefs independently, Q reduces to a diagonal matrix, and the standard deviation (volatility) of the price decreases with $1/\sqrt{n}$ with the numbers of traders. A large market of independent noise traders therefore tends to alleviate price fluctuations. The other interesting case is that of perfect correlation due to herding behaviour. The standard deviation of the price turns out a weighted average of the standard deviations of the individual expected returns independent of n. Herding behaviour turns out to be necessary to propagate individual belief changes to market price volatility.

The dominating influence on stock price volatility, however, is its inverse dependency on the risk-free interest rate. Such a negative relation between low-interest policy and high implied volatility has been established in a recent study [3]. The authors found a "strong co-movement between the VIX, the stock market option-based implied volatility, and monetary policy." Risk aversion decreases with a lax monetary policy.

Appendix: Proof of the Main Theorem

Proof We say that $x, y \in X$ are separable if and only if there is an $A \in \mathscr{A}$ with $x \in A$ and $y \notin A$. Clearly, x and y are separable if and only if $\delta_x \neq \delta_y$ on $\mathscr{A}$. Let $x, y, z, w \in X$ be pairwise separable and consider the probability measures

$$p^i = a_i \cdot \delta_x + b_i \cdot \delta_y + c_i \cdot \delta_z + d_i \cdot \delta_w, \quad i = 1, \ldots, n.$$

for $a_i, b_i, c_i, d_i \in [0, 1]$ with $a_i + b_i + c_i + d_i = 1$. Since x, y, and z are pairwise separable and $\mathscr{A}$ is closed under intersection, there are pairwise disjoint events $E_x \ni x$, $E_y \ni y$, and $E_z \ni z$. Let $A = E_x \cup E_y$ and $B = E_x \cup E_z$, then $p^i\,(A) = r^i$, $p^i\,(B) = q^i$. The conditional measures are

$$p_A^i = q^i \cdot \delta_x + \left(1 - q^i\right) \cdot \delta_y,$$
$$p_B^i = r^i \cdot \delta_x + \left(1 - r^i\right) \cdot \delta_z.$$

It is easy to see that A and B are independent, $p^i(A \cap B) = p^i(A) \cdot p^i(B)$, if and only if $d_i = \frac{a_i c_i}{b_i}$.

Since by assumption, P is convex and each Dirac measure can be weakly approximated by a sequence of measures in P, we can extend f uniquely to convex combinations of Dirac measures. Applying dynamic consistency we obtain

$$f\left(p^1, \ldots, p^n\right) = a \cdot \delta_x + b \cdot \delta_y + c \cdot \delta_z + d \cdot \delta_w$$

with $a, b, c, d \in [0, 1]$ with $a+b+c+d = 1$. There exists a function $\varphi : [0, 1]^n \to [0, 1]$ with

$$\varphi\left(\frac{a_1}{b_1}, \ldots, \frac{a_n}{b_n}\right) = \frac{a}{b}$$

for all $a_1, \ldots, a_n, b_1, \ldots, b_n$ with $a_i \leq b_i$ and $1 \geq b_i > 0$ and

$$a = \varphi(a_1, \ldots, a_n),$$
$$b = \varphi(b_1, \ldots, b_n).$$

This implies

$$\varphi\left(\frac{c_1}{b_1}, \ldots, \frac{c_n}{b_n}\right) = \frac{c}{b},$$
$$\varphi\left(\frac{a_1}{c_1}, \ldots, \frac{a_n}{c_n}\right) = \frac{a}{c},$$

and thus

$$\varphi\left(\frac{a_1}{b_1}, \ldots, \frac{a_n}{b_n}\right) = \varphi\left(\frac{a_1}{c_1}, \ldots, \frac{a_n}{c_n}\right) \cdot \varphi\left(\frac{c_1}{b_1}, \ldots, \frac{c_n}{b_n}\right).$$

In other words, for any $x_1, \ldots, x_n \in [0, 1]$ and $y_1, \ldots, y_n \in (0, 1]$ we find

$$\varphi(x_1 \cdot y_1, \ldots, x_n \cdot y_n) = \varphi(x_1, \ldots, x_n) \cdot \varphi(y_1, \ldots, y_n).$$

Thus there exist real numbers $\alpha_1, \ldots, \alpha_n$ such that

$$\varphi(x_1, \ldots, x_n) = x_1^{\alpha_1} \cdot \cdots \cdot x_n^{\alpha_n}.$$

For the probabilities we therefore obtain

$$a = \varphi(a_1, \ldots, a_n) = a_1^{\alpha_1} \cdot \cdots \cdot a_n^{\alpha_n}.$$

Since probabilities are bounded by one, the α_i must be non-negative. This proves that (1) holds for the densities.

Unamity holds if and only if $a = \varphi(a, \ldots, a)$ if and only if $\sum_{i=1}^{n} \alpha_i = 1$. Anonymity hold if and only if φ is symmetric with respect to permutations of the arguments, which holds if and only if all exponents are equal (see [1]).

References

1. Aczel, J.: Lectures on Functional Equations and Their Applications. Mathematics in Science and Engineering, vol. 19. Academic Press, New York (1966)
2. Ali, M.: Probability and utility estimates for racetrack bettors. J. Polit. Econ. **85**, 803–815 (1977)
3. Bekaert, G., Marie, H., Marco Lo, D.: Risk, uncertainty and monetary policy. J. Monetary Econ. **60**(7), 771–788 (2013)
4. Fama, E., French, K.: The CAPM: Theory and Evidence. Working Paper No. 550, Center for Research in Security Prices, University of Chicago (2003)
5. Gjerstad, S.: Risk Aversion, Beliefs, and Prediction Market Equilibrium. Economic Science Laboratory, University of Arizona (2005)
6. Guo, H.: A Rational pricing explanation for the failure of the CAPM. *Federal Reserve Bank of St. Louis Review*, May/June 2004, vol. 86(3), pp. 23–33 (2004)
7. Grossman, S., Joseph, S.: On the impossibility of informationally efficient markets. Am. Econ. Rev. **70**(3), 393–408 (1980)
8. Grossman, S.J.: An introduction to the theory of rational expectations under asymmetric information. Rev. Econ. Stud. **48**(4), 541–559 (1981)
9. Klenke, A.: Probability Theory. Springer, New York (2006)
10. Manski, C.: Interpreting the predictions of prediction markets. Econ. Lett. **91**(3), 425–429 (2006)
11. Milgrom, P., Stokey, N.: Information, trade and common knowledge. J. Econ. Theory **26**, 11–21 (1982)
12. Muth, J.F.: Optimal properties of exponentially weighted forecasts. J. Am. Stat. Assoc. **55**(290), 299–306 (1960)
13. Muth, J.F.: Rational expectations and the theory of price movements. Econometrica **29**, 315–335 (1961)
14. Shleifer, A., et al.: Noise trader risk in financial markets. J. Polit. Econ. **98**(4), 703–738 (1990)
15. Wolfers, J., Zitzewitz, E.: Prediction markets. J. Econ. Perspect. **18**, 107–126 (2004)
16. Wolfers, J., Zitzewitz, E.: Interpreting Prediction Market Prices as Probabilities, NBER Working Papers 12200, National Bureau of Economic Research Inc. (2005)

The Dynamics of Hedge Fund Performance

Serge Darolles, Christian Gouriéroux and Jérome Teiletche

Abstract The ratings of fund managers based on past performances of the funds and the rating dynamics are crucial information for investors. This paper proposes a stochastic migration model to investigate the dynamics of performance-based ratings of funds, for a given risk-adjusted measure of performance. We distinguish the absolute and relative ratings and explain how to identify their idiosyncratic and systematically persistent (resp. amplifying cycles) components. The methodology is illustrated by the analysis of hedge fund returns extracted from the TASS database for the period 1994–2008.

1 Introduction

The journals for investors write lead articles or even make their cover page on the ratings of funds. These ratings are generally based on (quantitative) rankings of fund managers based on past performances of the funds. These rankings and their dynamics are used to define the management fees of fund managers, by introducing money incentives based on the evolution of their performances, such as the high water mark scheme [see e.g. [1, 5]]. They are also used by the fund managers to attract new clients and increase the net asset value of the fund, i.e., is the size of the portfolio to be managed. They are finally used by the investors to construct more robust portfolio management, such as positional portfolio management [see e.g. [14]].

S. Darolles (✉)
Université Paris-Dauphine and CREST, Place du Marchal de Lattre de Tassigny,
75016 Paris, France
e-mail: serge.darolles@dauphine.fr

C. Gouriéroux
CREST and University of Toronto, 15 Boulevard Gabriel Péri,
92245 Malakoff Cedex 1, France
e-mail: christian.gourieroux@ensae.fr

J. Teiletche
Unigestion, Avenue de Champel 8C, 1206 Geneve, Switzerland
e-mail: jteiletche@unigestion.com

V.-N. Huynh et al. (eds.), *Econometrics of Risk*, Studies in Computational Intelligence 583, DOI 10.1007/978-3-319-13449-9_7

Our paper proposes a stochastic migration model to investigate the dynamics of performance-based rankings of funds, for a given risk-adjusted measure of performance. We distinguish the absolute and relative ranking and explain how to identify the idiosyncratic and systematically persistent or amplifying cycles (i.e., procyclical) components. The methodology is illustrated by the analysis of hedge fund returns extracted from the TASS database for the period 1994–2008.

The outline of the paper is as follows. Section 2 introduces a general framework for analyzing the joint dynamics of performances of a set of funds. This analysis is based on rankings derived from a quantitative performance measure. The ranking can be absolute, when the performance level is taken into account, or relative when only the rank of the performance matters. The ranking histories of the different funds are used to construct matrices which provide the transition frequencies from one level to another one. We describe the static and dynamic stochastic transition models, which capture the uncertainty on transition and their dynamics. In Sect. 3, this methodology is applied to a set of hedge funds. Using data from the TASS database, we estimate the absolute and relative transition matrices and analyze their dynamics. In particular, we identify the idiosyncratic and systematic persistence components. Section 4 contains the conclusions.

2 Fund Performance Dynamics

2.1 Performance and Ranking

2.1.1 Performance

There exists many risk-adjusted performance measures.[1] They differ by the way they summarize the notion of risk, by the investment horizon, or by the information taken into account in their computation. For instance, for given horizon and historical performance measures, several authors have proposed the Sharpe ratio [21, 22], which is the ratio of a mean return by its realized volatility, the Sortino ratio [25], which is the ratio of a mean return by its realized semi-standard deviation, or the L-performances [6].

These performance measures can also be computed from conditional distribution to account for either the increase of available information over time, or for the fact that the investment in funds is used to complete an investment in another type of asset, leading to the fitted performance measures [4, 17]. In the following discussions, we consider a given risk-adjusted performance measure.

[1] In the hedge fund literature, the term "performance" is often used for return, and thus is an ex-post notion. In our paper we are interested in ex-ante performance measures, i.e., before observation. These measures are adjusted for the uncertainty on future returns.

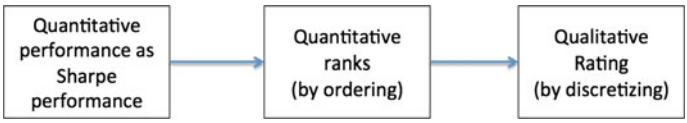
2.1.2 Rating

As in the credit risk or mutual fund industry, the practice is to diffuse information on hedge funds by means of qualitative ratings regularly elaborated by specialized agencies. These qualitative ratings can be derived from ranking based on a selected risk-adjusted performance measure S_{it}, where i denotes the hedge fund and t the date, in different ways.

(i) *Absolute rating*. The idea is to select a set of thresholds $a_1 < \cdots < a_K$, say, and to define the class according to the location of the risk-adjusted performance measure with respect to the thresholds. The best rating is obtained if $a_K < S_{it}$, the second best if $a_{K-1} < S_{it} < a_K$, and so on …
(ii) *Relative rating*. Relative ratings are derived by comparing the performance of a hedge fund to the performance of the other funds. More precisely, the different hedge funds can be ranked according to their performances at date t from the worst one to the best one. Then, we deduce the rank $Rk_{i,t}$ of hedge fund i among the n funds of the population. The relative risk-adjusted performances are these cross-sectional ranks, and the discretization is deduced by thresholds defined on these cross-sectional ranks. The best rating class includes for instance, the 10 % of hedge funds with the best risk-adjusted performances, and so on.

These relative performance rankings depend on the selected population of funds (perimeter) and this perimeter can change over time. Do we compute these ranks for all funds (mutual and hedge funds), or simply for the hedge funds? Do we include the funds of funds? How do we treat the new created funds or correct for the survival bias?

Thus, there exists a large number of qualitative ratings according to the underlying risk-adjusted performance measure, to the absolute or relative approach, to the number and levels of selected thresholds, and to the selected perimeter in the relative rating case. From a mathematical point of view, both absolute and relative ratings are sample counterparts of theoretical performance ranks. The theoretical relative performance ranks are $Rk_{i,t} = F_t(S_{it})$, where F_t is the distribution of performance at time t, whereas the theoretical absolute performance ranks are $Rk^*_{i,t} = F(S_{it})$, where F is the marginal distribution of performance, i.e. $F = \lim_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T} F_t$ (see Appendix 1 for a more detailed discussion). Then these ranks are discretized to get the associated qualitative rating; see Scheme 1.



Scheme 1 From quantitative performance to ratings

2.2 Stochastic Migration and Migration Correlation

Let us now assume a given rating approach. The rating can be computed by fund and date, providing individual rating histories. The rating histories define panel data indexed by both fund and time. The total number of observations depends on the number of funds included in the group (cross-sectional dimension) and the number of observation dates (time dimension). We describe below the basic stochastic migration model and define the notion of migration correlation. Finally, we explain how to estimate the different parameters of such a stochastic migration model [see e.g. [7, 12]].

2.2.1 Homogeneous Population of Funds

Let us denote the (qualitative) rating histories by:

$$Y_{i,t},\ i = 1, \ldots, n,\ t = 1, \ldots, T,$$

where i indicates the fund and t the date when the rating is published (or computed from lagged returns). The variable $Y_{i,t}$ is a qualitative variable with alternatives $k = 1, \ldots, K$. These alternatives are ranked in such a way class $k = 1$ indicates the best performing funds, $k = 2$ the second best performing funds…, up to the last class $k = K$, which corresponds to inactivity, i.e. either fund default, or no reporting. A simple type of model underlies the practice of fund rating agencies or fund of hedge funds managers, when they aggregate individual rating histories into transition matrices. These models rely on the following assumption:

Assumption A.1 The individual rating histories $(Y_{i,t},\ t = 1, \ldots, T, \ldots)$, $i = 1, \ldots, n$, are independent and identically distributed Markov processes, with a common sequence of transition matrices.

Intuitively, this assumption characterizes "homogeneous" populations of funds. Indeed, the joint distribution of individual histories is not modified by permutation of funds; thus, these funds are observationally equivalent. The homogeneity Assumption A.1 can be satisfied by either absolute or relative ratings, but cannot be satisfied for both of them simultaneously, except in very special cases.

The joint dynamics of rating histories of all funds is characterized by the sequence of $K \times K$ time dependent transition matrices Π_t, $t = 1, \ldots, T$. The elements of Π_t are the transition probabilities from rating l to rating k between dates t and $t + 1$:

$$\pi_{kl,t} = P[Y_{i,t+1} = k | Y_{i,t} = l]. \tag{1}$$

The transition probabilities are nonnegative and sum up to one for each row of the transition matrix. Under Assumption A.1, the most recent individual rating contains all relevant information on the rating history of a given firm (when Π_t is given), and the transition can be time varying since the serial homogeneity hypothesis is

generally not made. Moreover, the transition probabilities are the same for two different funds; this follows from the cross-sectional homogeneity assumption on the population of funds.

Under Assumption A.1, the time-varying transition matrices can be easily estimated by following a cross-sectional approach. More precisely, let us consider a given date t, and denote (i) by $n_{k,t}$ the number of funds in rating k at date t and, (ii) by $n_{kl,t}$ the number of funds migrating from rating l to rating k between t and $t+1$. The transition probabilities are estimated by their cross-sectional counterparts:

$$\hat{\pi}_{kl,t} = n_{kl,t}/n_{k,t}. \tag{2}$$

These estimators are the fixed effect maximum likelihood estimators of transition matrices computed at each point in time. The accuracy of this ML estimator depends on both size $n_{k,t}$ of the rating class and transition probabilities ($n_{kl,t}$, l varying) [see e.g. [13]].

The estimated transition probabilities can be used to check if the transition probabilities are time independent, that is, if the Markov chain is time-homogeneous. For instance, [15] accept the time homogeneity restriction for two different rating systems of mutual funds provided by *Standard &Poors* and *Morningstar*, respectively. Other homogeneity tests have been proposed, based on the spectral decompositions of estimated transition matrices [see e.g. [8, 9]]. Applying these tests, [8] reject the null hypothesis of homogeneous Markov chain for credit rating transitions [see also [3]].

Moreover, there exist more fundamental reasons for considering heterogeneous Markov chain. First, as seen later in the application to hedge funds, the null hypothesis of homogeneous Markov chain will be rejected. Second, even if the standard specification with time-independent stochastic transition matrix features migration correlation, the effect of migration correlation stay the same at all horizon. We get a flat term structure of inactivity and migration correlation. Finally, the heterogeneous Markov chain is more appropriate for distinguishing the macro dynamics such as, trends, business cycles or contagion effects, from the idiosyncratic dynamics [see Sect. 2.2.4]. It is especially appealing for the analysis of common factors, i.e. systematic risk.

2.2.2 Stochastic Intensity

This approach was considered by [12] to model migration correlation in the context of corporate credit risk. It corresponds to an extension of the stochastic intensity model introduced for default risk by Vasicek [27], and used in the advanced Basel 2 regulation [2] and for credit derivative pricing [20]. The specification is completed by specifying the distribution of latent transition matrices. Two types of assumptions denoted A.2 and A.2*, respectively, are introduced depending on whether either a static, or a dynamic model for transition probabilities is considered. These assumptions are the following:

Assumption A.2 *(static)*: The transition matrices at horizon 1 are stochastic, independent with identical distribution.

Assumption A.2* *(dynamic)*: The transition matrices define a time-homogeneous Markov process, with transition pdf $f[\Pi_t|\Pi_{t-1}]$.

Since the transition probabilities Π_t are not directly observed (even if they can be estimated by $\hat{\Pi}_t$), the joint distribution of individual rating histories is obtained by integrating out the latent (Π_t). This induces both serial and cross dependence. Model (A.1)–(A.2) [or (A.1)–(A.2*)] is a nonlinear factor model in which the factors are the elements of the transition matrices. The number of linearly independent factors is $K(K-1)$ due to the unit mass restrictions.[2]

In general, the introduction of stochastic transitions in a Markov chain implies a longer memory. This means that the joint process $(Y_{1,t}, \ldots, Y_{n,t})'$ is no longer Markov under (A.1)–(A.2*). However, it has been proved in [[12–14], Appendix 1] that, for static factors, i.e., under (A.1)–(A.2), the joint rating vector $(Y_{1,t}, \ldots, Y_{n,t})'$ is still an homogeneous Markov chain with K^n admissible states. This result is the basis for defining the notion of migration correlation.

2.2.3 Migration Correlation

Under (A.1)–(A.2), the migration correlation can easily be defined as follows. Let us focus on a couple of funds i and j. Their individual transition probabilities are:

$$p_{kk^*} = P[Y_{i,t+1} = k|Y_{i,t} = k^*] = E[\pi_{kk^*,t}], \tag{3}$$

and their joint transition probabilities are:

$$p_{kk^*,ll^*} = P[Y_{i,t+1} = k, Y_{j,t+1} = l|Y_{i,t} = k^*, Y_{j,t} = l^*] = E[\pi_{kk^*,t}\pi_{ll^*,t}], \tag{4}$$

where the expectation is taken with respect to stochastic Π_t. In particular, the joint transitions do not depend on the selected couple of funds. Let us now introduce the state indicators:

$$\mathrm{I}_{Y_{i,t}=k} = 1 \text{ if } Y_{i,t} = k, \ \mathrm{I}_{Y_{i,t}=k} = 0 \text{ otherwise.}$$

The migration correlation is defined as the conditional correlation of such indicator functions:

$$\begin{aligned}\rho_{kk^*,ll^*} &= corr\left[\mathrm{I}_{Y_{i,t+1}=k}\, \mathrm{I}_{Y_{j,t+1}=l}|Y_{i,t} = k^*, Y_{j,t} = l^*\right] \\ &= \frac{p_{kk^*,ll^*} - p_{kk^*}p_{ll^*}}{[p_{kk^*}(1-p_{kk^*})]^{1/2}[p_{ll^*}(1-p_{ll^*})]^{1/2}}.\end{aligned} \tag{5}$$

Matrices $P_2 = (p_{kk^*,ll^*})$ and $\rho_2 = (\rho_{kk^*,ll^*})$ cannot be approximated from the cross-sectional information at time t only, since an accurate estimation of $E[\pi_{kk^*,ll^*}$ $\pi_{kk^*,ll^*}]$ requires observations of transitions at several dates. The quantities p_{kk^*},

[2] $(K-1)^2$ independent factors, if the state "inactive", i.e. state K, is an absorbing state.

p_{kk^*,ll^*}, ρ_{kk^*,ll^*} will be approximated by mixing appropriately cross-sectional and time averaging. Typically, from Eqs. (3), (4), the individual transitions are estimated by:

$$\hat{p}_{kk^*} = \frac{1}{T}\sum_{t=1}^{T} \hat{\pi}_{kk^*,t},$$

and the pairwise migration probability p_{kk^*,ll^*} are estimated by:

$$\hat{p}_{kk^*,ll^*} = \frac{1}{T}\sum_{t=1}^{T} \hat{\pi}_{kk^*,t}\hat{\pi}_{ll^*,t}.$$

The estimated migration correlations are deduced by substituting $\hat{p}_{kk^*}$, $\hat{p}_{kk^*,ll^*}$ to their theoretical counterparts in Eq. (5).

2.2.4 Persistence and Cycle

The dynamic of an homogeneous Markov chain with transition matrix Π is usually analyzed by means of the eigenvalues[3] of Π. Its eigenvalues can be real or complex, but have always a modulus smaller or equal to 1. Moreover, due to the unit mass restriction, 1 is always one of these eigenvalues. In such homogeneous chain, the persistence (resp. the cycles) is analyzed by considering the real eigenvalues close to 1 (resp. the complex eigenvalues). The cycles are created by the complex eignevalues with a frequency corresponding to the period of the cycle.

In the static stochastic migration model (A.1)–(A.2), this analysis can be directly performed on individual transition matrix: $P_1 = (p_{kk^*}) = (E[\pi_{kk^*}])$, and on pairwise matrices: $P_2 = (p_{kk^*,ll^*})$.

The (dynamic) stochastic migration model (A.1)–(A.2*) is more complicated, but allows for a more detailed analysis of persistence and cycle effects. Indeed, we can distinguish idiosyncratic from general persistence effects (resp. cycle effects). Loosely speaking, we can first estimate the transition matrices $\hat{\Pi}_t$. Then, (i) we can perform the spectral decomposition of each matrix $\hat{\Pi}_t$ to analyze the idiosyncratic persistence or cycles, that are conditional on factor values; (ii) This analysis is completed by considering the autocorrelation function of the multivariate series[4] $vec(\hat{\Pi}_t)$ to detect the persistence and cycle effects due to the common factor Π_t.

2.2.5 Ordered Probit Model

In practice, the stochastic transition matrices are generally written as function of a smaller number of factors, i.e. $\Pi_t = \Pi_t(Z_t)$, $t = 1, \ldots, T$; then, the whole

[3] A transition matrix can always be diagonalized.

[4] *vec* means that all the elements of the matrix (K, K) are stacked in a single vector of dimension $(K^2, 1)$.

dependence between individual histories is driven by the common factors Z_t. The static (resp. dynamic) factor models are obtained by assuming i.i.d. (resp. Markov) factors Z_t. In particular, a factor specification for transition matrices proposed in the credit risk literature is the ordered probit model.

The approach assumes a continuous quantitative risk-adjusted performance for each fund, used to define the qualitative class. Let us denote by $s_{it} = \log S_{it}$ the continuous latent risk-adjusted log-performance of fund i at date t, k^* its rating at the beginning of the period, and assume that:

$$s_{it} = \alpha_{k^*} + Z_t \beta_{k^*} + \sigma_{k^*} u_{it},$$

where the error terms u_{it} and the common factors Z_t are independent and u_{it} are standard Gaussian variables. (i) Let us first consider absolute rating defined from risk-adjusted performance and denote $a_1 < a_2 < \cdots < a_K$ the time-independent thresholds defining the rating alternatives:

$$\text{rating} = k, \text{ iff } a_{k-1} < s_{it} < a_k.$$

In a static model, where Z_t are i.i.d. Gaussian variables, we get:

$$p_{kk^*,ll^*} = P\left[a_{k-1} < \alpha_{k^*} + Z_t \beta_{k^*} + \sigma_{k^*} u_{it} < a_k, a_{l-1} < \alpha_{l^*} + Z_t \beta_{l^*} + \sigma_{l^*} u_{it} < a_l\right],$$

which is a bivariate probit formula.
(ii) When relative ratings are considered, for instance by deciles, the thresholds become dependent of both time and parameters α_k, β_k, σ_k, $k = 1, \ldots, K$. More precisely, the cross-sectional residual distribution of the log-performances at time t is a mixture of Gaussian distribution:

$$P_t = \sum_{k=1}^{10} \frac{1}{10} N\left[\alpha_k + \beta_k Z_t, \sigma_k^2\right].$$

The thresholds (a_{kt}) are the deciles of the distribution Pt and depend on parameters α, β, σ^2 and of the value of the factor at date t. In practice, these theoretical deciles are estimated by their empirical cross-sectional counterparts. These estimators are not fully efficient since the estimation approach does not take into account the dependence of the thresholds in parameters α, β, σ^2.

3 Application to Hedge Funds

3.1 Data

We use the November 2008 version of the TASS database for a retrospective period covering November 1977 to October 2008. The TASS database includes various informations related to performance, asset under management and fund

characteristics (fees, location, style) at a monthly frequency and for a large set of hedge funds. This database is frequently used by academic researchers, which will allow for a comparison of our results with other results appeared in the literature.

As usual, the hedge funds databases suffer from various biases [see for instance [11]].

(i) A survivorship bias arises when the returns, performances, rating histories of the currently inactive fund have not been kept in the database. Typically, if the dead funds are the poorly performing ones (an hypothesis which is disputed in the literature as stellar closed-funds may have no interest in reporting to databases either), survivorship bias can lead to an overestimation of returns. For a bias correction of the survivorship, the TASS database is made of two different sub-bases. The Live funds database includes the funds, which are still active in October 2008. The Graveyard database includes information on funds which are no longer active, since they have been either liquidated, restructured, merged with other funds, or simply stopped to report their performance.[5] The comparison of the two databases can be used to correct for survivorship bias of the Live funds database. Following [19], we exclude data prior to January 1994 as the Graveyard database only became active in 1994.

(ii) Backfill bias arises when an additional fund is introduced in the database with its historical track-record, including its history prior to inclusion. Very often, the backfill bias is a form of incubation bias since the funds are proposed to a large public after an incubation period during which the fund has to reach a minimal performance target. Only successful funds reaching the target are made available to the public. The TASS database provides both the inception and inclusion dates. This information can be used for correcting the backfill bias and we can consider that all data prior to inclusion date in the database are due to incubation bias. However, as stated by [11], this might lead to an overestimation of the real incubation bias, since the difference between the inception and inclusion dates can be due to switches of database, from HFR to TASS, or to the merger of TASS and Tremont in September 2001. In our application, we follow Kosowski et al. (2008) and treat the backfilling/incubation bias by excluding the first 12 months of data for each fund.

There are 7,068 funds in the Live database and 5,150 funds in the Graveyard database, respectively. We drop observations when: (i) The currency of denomination is not the USD; (ii) The strategy is not reported, the fund is a fund of hedge funds, or is investing in other funds; (ii) The performance is not net of fees; (iv) The fund is guaranteed; (v) The Asset Under Management (AUM) are not disclosed, or only poorly estimated; (vi) The publication of returns is not monthly; (vii) The fund has less than 24 months of returns history. This latter condition can introduce another type of survivorship bias, but it is required to get enough historical data for computing Sharpe performances. After applying these various filters, we get 1,183 live funds and 1,814 inactive funds.

[5] TASS can take up to nine months before transferring a fund from the Live database to the Graveyard database. This implies that some dead funds can be considered as still living in the end of the sample.

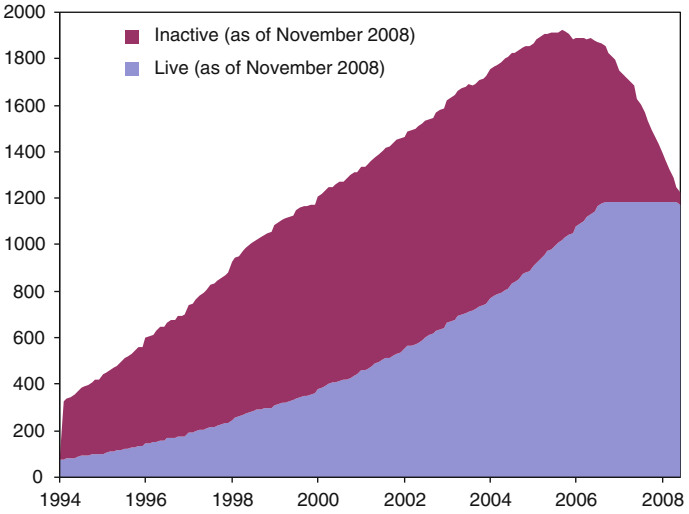


Fig. 1 Number of funds

3.2 Summary Statistics of Returns

The number of funds is displayed in Fig. 1, where we observe that inactive funds are dominating the sample up to 2005, and then naturally decreases.[6] We have a rather large sample, with an average of roughly 1,250 funds per months, almost equally split between active and inactive funds.

Summary statistics of the hedge funds are reported in Table 1. They are given by style and distinguish active and inactive funds. The proportion of inactive funds depends on the style and takes value between 52 % for Emerging Markets to 69 % for Convertible Arbitrage. The average return by style can be either smaller, or larger for inactive funds. For instance, they are smaller for Managed Futures, Global Macro…, larger for Event Driven and Convertible Arbitrage. The most liquid strategies, that are Managed Futures and Global Macro, feature positive skewness, whereas the returns corresponding to other strategies are generally left skewed. This systematic skewness is more pronounced for active funds. As expected, the return distribution has fat tails, with kurtosis significantly larger than 3 corresponding to the normal distribution. Extreme risks are especially seen for the arbitrage strategies, which

[6] The constant number of live funds at the end of the period should be due to lags in funds reporting their performance. Indeed, we observe the performances for October 2008 at the end of November 2008, which implies that only funds with short notice of publication of Net Asset Value (NAV) are incorporated at that moment. Thus, without additional information, the drop in the number of live funds cannot be directly related to an increase in fund mortality, which seems to happen at the end of 2008.

Table 1 Summary statistics

	# funds	# obs	Average # obs by month	Avg (%)	Std (%)	SK	KU	JB (%)
Active funds								
Convertible Arbitrage	44	4201	23.6	0.38	3.03	−2.0	16.9	90.9
Dedicated short bias	9	912	5.1	0.51	5.82	0.1	5.1	77.7
Emerging Markets	103	8385	47.1	0.88	5.16	−0.8	8.6	79.6
Equity market neutral	85	6709	37.7	0.63	2.44	−0.2	7.4	68.2
Event driven	156	14051	78.9	0.69	2.57	−0.7	7.7	80.1
Fixed income arbitrage	60	4556	25.6	0.46	2.53	−1.7	18.5	83.3
Global Macro	52	3721	20.9	1.06	4.51	0.1	5.9	44.2
Long/short eq. hedge	457	39354	221.1	0.71	4.50	−0.3	5.9	68.1
Managed Futures	128	12512	70.3	1.21	5.50	0.4	4.7	49.2
Multi-strategy	89	7512	42.2	0.74	3.27	−0.8	9.2	78.6
Total	**1183**	**101913**	**572.5**	**0.76**	**4.03**	**−0.4**	**7.6**	**70.1**
Inctive funds								
Convertible Arbitrage	102	7251	41.0	0.53	2.05	−0.5	7.3	62.7
Dedicated short bias	21	1531	8.6	−0.01	6.38	0.3	4.4	52.3
Emerging Markets	113	7896	44.6	0.77	6.94	−0.5	7.6	63.7
Equity market neutral	122	6985	39.5	0.47	2.16	0.0	4.7	38.5
Event driven	226	16222	91.6	0.82	2.75	−0.2	6.5	65.0
Fixed income arbitrage	106	6584	37.2	0.46	2.32	−1.5	12.4	83.0
Global Macro	109	6841	38.6	0.47	4.17	0.2	5.8	56.8
Long/short eq. hedge	743	49930	282.1	0.88	5.03	0.1	5.4	51.2
Managed Futures	191	12157	68.7	0.57	6.28v	0.1	5.2	45.5
Multi-strategy	79	4881	27.6	0.74	2.76	−0.3	9.0	63.2
Total	**1812**	**120278**	**679.5**	**0.72**	**4.34**	**−0.1**	**6.3**	**55.6**

Notes The statistics are computed by averaging across all funds in a given category (active/inactive and style). Avg is the mean, Std the standard deviation, SK the skewness, KU the kurtosis. JB denotes the proportion of funds for which the null hypothesis of normality is rejected at the 95 % level for the Jarque-Bera test

are Convertible Arbitrage and Fixed Income Arbitrage. Both skewness and fat tail explain why the normality assumption is rejected by the Jarque-Bera Test given in the last column for a large proportion of individual funds.

3.3 The Ratings

Risk-adjusted Sharpe ratios are calculated with a rolling window of 24 months, and a risk-free rate fixed to 4 % per year.[7] For each month, the funds are assigned into

[7] The size 24 months of the window has been chosen for illustrating the approach. In a more complete analysis, it might be preferable to consider different sizes, then to select the most informative one for predictive purpose and/or the most relevant one for the updating frequency of the investors portfolio.

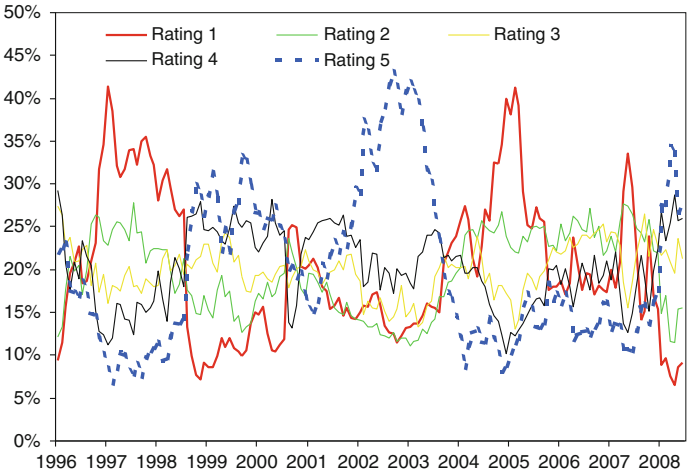


Fig. 2 Proportion of active funds by absolute rating class

six categories according to absolute performance (resp. relative performance). For instance, for absolute rating, the first category includes the 20 % of funds with highest Sharpe ratios, the second category the following 20 %, etc. up to the fifth category where the worst 20 % are represented. The sixth category includes the funds which are inactive in November 2008. This provides for each fund monthly time-series of absolute and relative ratings, evolving between grade 1 and grade 6.

For relative ratings, the proportion of active funds is equal to 20 % for ratings 1, 2, 3, 4, 5 by construction. This proportion depends on time for absolute rating as observed in Fig. 2.

Figure 2 shows clearly different regimes with high performances around 1997, 2001, 2005 and 2007, and much lower performances around 1999, 2003, 2006 and 2008. Generally, when the performance is high, a large proportion of funds feature these performances, e.g. 40 % in 1997, 40 % in 2005 and 35 % in 2007. Symmetrically, in 1999–2000 and 2008, low performances are observed with 35 and 45 %, respectively, for rating 5. This is only around 2006 that the funds show rather heterogeneous performances.

3.4 Time Homogeneous Transition Matrices

The absolute and relative rating histories can be used to construct transition matrices $\hat{\Pi}_t$. Recall that the absolute (resp. relative) ratings are the Sharpe performances transformed by their historical c.d.f. (resp. cross-sectional c.d.f.) [see Sect. 2.2.1]. In this Section, we assume time-homogeneous Markov processes $Pi_t = \Pi$, $\forall t$, and estimate the common Π matrix by averaging matrices $\hat{\Pi}_t$, weighted by the number of funds at each period.

These estimated matrices are computed for investment horizons 1 month, 1 year, 2 years, 3 years and 5 years, respectively. They are given in Table 2 (Panel A–E) for relative rating and in Table 3 (Panel A–E) for absolute rating.

Let us discuss the 1-month transition matrices. The last row 0, 0, 0, 0, 0, 100 % corresponds to the absorbing state NR. The last column provides the proportion of inactive as a function of the previous rating. As expected, this proportion increases when previous rating becomes worse. Absolute and relative persistence are observed, since the significant transition probabilities are large on the main diagonal. Moreover, at 1-month horizon, we can essentially observe upgrades or downgrades of at most one grade.

When the investment horizon increases, transitions to inactivity increase and all types of upgrades and downgrades are observed. In the long run (Tables 2 (Panel E) and 3 (Panel E)), we expect to observe the stationary distribution of the Markov process if the time-homogeneity assumption is satisfied. For instance, in Table 2 (Panel E), we expect similar values for the first five transition probabilities in each row. Clearly, this property is not satisfied in row 1, which shows a rather strong persistence for rating 1.

This analysis is completed by considering the eigenvalues and eigenvectors of the transition matrices. These spectral decompositions are given in Table 4 (Panel A, B).

If the Markov homogeneity assumption is satisfied, the transition matrix at horizon h is equal to the short term transition matrix at power h. In particular, if the eigenvalues of the short term matrix are real positive too and decrease, the eigenvalues of the transition matrices at larger horizon have to be real positive too and have to decrease with the horizon. This effect is observed on the two largest eigenvalues (after the unitary eigenvalue). The short term eigenvalues are close to 1, as usual for a transition at short term horizon, but a part of this persistence can be due to the rolling window of 24 months used in the computation of the risk-adjusted Sharpe ratio.

Under the homogeneity hypothesis, we have seen in Sect. 2.2.2 that the joint rating histories are still Markov, which allow for defining joint probabilities of transition and migration correlations. To avoid transition matrices of a too large dimension, we aggregate the joint transition in the following way: (i) probability of jointly maintaining the initial rating; (ii) probability of jointly being upgraded by one rating category; (ii) probability of being jointly downgraded by one rating category; (iv) probability of becoming jointly not rated; (v) probability of getting a same rating, but not necessary equal to the initial common one. These five events are called Unchanged, Upgrade, Downgrade, Non-rated, Similar Rating, respectively. The joint transition probabilities and the migration correlations are provided in Table 5 (Panel A–H) H for absolute and relative ratings.

We observe very small migration correlations of order less than[8] 0.05. At 2-years horizon, the effect of the rolling window used in computing the adjusted Sharpe ratio disappears and the maximal migration correlations are for extreme ratings, that are

[8] Very small migration correlations have not to be put to zero due to the possible leverage effects in portfolio management strategy. Loosely speaking, even small migration correlations can allow for gain opportunities by using appropriate correlation strategies.

Table 2 Transition matrices at 1-month (*Panel A*), 1-year (*Panel B*), 2-years (*Panel C*), 3-years (*Panel D*), 5-years (*Panel E*) horizon (static factor, relative performance)

	Q1 (%)	Q2 (%)	Q3 (%)	Q4 (%)	Q5 (%)	NR (%)
Panel A						
Q1	89.6	9.5	0.3	0.1	0.0	0.5
Q2	8.6	75.1	15.2	0.6	0.0	0.6
Q3	0.3	13.5	68.8	16.2	0.3	0.9
Q4	0.1	0.4	14.4	71.2	12.6	1.3
Q5	0.0	0.0	0.2	11.0	86.1	2.7
NR	0.0	0.0	0.0	0.0	0.0	100.0
Panel B						
Q1	52.2	21.0	10.4	6.0	3.3	7.2
Q2	18.5	27.4	22.3	15.0	8.6	8.2
Q3	8.0	19.8	24.0	22.2	15.0	11.0
Q4	4.3	12.0	19.9	24.8	23.1	16.0
Q5	1.7	5.0	10.8	20.3	38.9	23.3
NR	0.0	0.0	0.0	0.0	0.0	100.0
Panel C						
Q1	34.1	16.5	12.1	11.3	12.1	13.9
Q2	15.7	17.5	16.9	16.6	17.5	15.9
Q3	10.6	15.7	17.5	17.9	17.4	21.0
Q4	8.1	13.7	17.2	17.9	16.1	26.9
Q5	6.1	11.6	13.7	14.6	16.3	37.7
NR	0.0	0.0	0.0	0.0	0.0	100.0
Panel D						
Q1	29.6	14.5	10.9	10.8	13.5	20.8
Q2	13.4	14.8	15.5	15.4	17.1	23.8
Q3	9.6	14.2	15.5	16.6	14.9	29.3
Q4	8.3	12.9	15.1	14.7	14.0	34.9
Q5	5.7	10.4	12.2	12.1	12.3	47.2
NR	0.0	0.0	0.0	0.0	0.0	100.0
Panel E						
Q1	20.6	14.1	11.0	10.0	9.4	34.8
Q2	10.4	11.9	12.8	13.7	14.4	36.8
Q3	8.5	10.7	11.7	12.5	14.2	42.4
Q4	6.8	10.5	11.2	12.2	13.5	45.9
Q5	6.4	8.2	8.2	8.8	9.9	58.5
NR	0.0	0.0	0.0	0.0	0.0	100.0

Notes Initial ratings are in rows and arrival ratings in columns. Transition matrices are computed for the whole panel of active and inactive funds, as observed over the period 1994–2008

Table 3 Transition matrices at 1-month (*Panel A*), 1-year (*Panel B*), 2-years (*Panel C*), 3-years (*Panel D*), 5-years (*Panel E*) horizon (static factor, absolute performance)

	Q1 (%)	Q2 (%)	Q3 (%)	Q4 (%)	Q5 (%)	NR (%)
Panel A						
Q1	88.7	10.2	0.5	0.1	0.0	0.5
Q2	9.1	73.6	15.8	0.8	0.0	0.7
Q3	0.3	14.4	66.8	17.2	0.4	1.0
Q4	0.1	0.4	15.5	70.0	12.7	1.3
Q5	0.0	0.0	0.2	11.1	86.2	2.6
NR	0.0	0.0	0.0	0.0	0.0	100.0
Panel B						
Q1	48.8	21.0	11.4	6.7	4.4	7.7
Q2	18.1	27.2	21.8	15.1	9.2	8.7
Q3	9.5	18.8	22.8	21.1	15.8	12.1
Q4	4.8	12.9	19.6	24.5	22.8	15.4
Q5	1.6	5.6	11.6	21.6	37.6	22.0
NR	0.0	0.0	0.0	0.0	0.0	100.0
Panel C						
Q1	31.2	17.7	12.8	11.0	12.8	14.6
Q2	14.1	16.6	17.2	17.5	17.7	16.9
Q3	10.4	14.2	16.4	18.3	18.8	21.9
Q4	9.0	13.0	16.5	18.2	17.2	26.1
Q5	8.0	13.1	14.6	14.8	14.1	35.4
NR	0.0	0.0	0.0	0.0	0.0	100.0
Panel D						
Q1	27.5	14.6	11.5	11.0	12.6	22.7
Q2	12.4	14.3	14.8	16.5	17.2	24.8
Q3	9.6	13.6	14.3	16.9	16.8	28.9
Q4	8.9	12.5	15.1	15.5	14.5	33.5
Q5	7.2	12.0	13.0	11.5	11.8	44.5
NR	0.0	0.0	0.0	0.0	0.0	100.0
Panel E						
Q1	21.7	14.0	10.5	9.5	10.5	33.7
Q2	9.8	12.5	13.4	13.6	14.1	36.5
Q3	9.7	11.0	12.1	13.3	13.4	40.4
Q4	8.3	10.9	11.2	11.7	12.5	45.4
Q5	7.8	9.0	9.1	9.0	8.9	56.1
NR	0.0	0.0	0.0	0.0	0.0	100.0

Table 4 Eigenvalues of the transition matrices (static factor, relative performance)

	Eig. 1	Eig. 2	Eig. 3	Eig. 4	Eig. 5	Eig. 6
Panel A						
1-month	1.000	0.988	0.931	0.828	0.674	0.487
1-year	1.000	0.866	0.506	0.234	0.061	0.005
2-years	1.000	0.767	0.231	0.018 +0.004i	0.018+0.004i	−0.001
3-years	1.000	0.685	0.184	−0.007	0.004−0.0006i	0.004+0.0006i
4-years	1.000	0.617	0.155	0.009−0.010i	0.009+0.010i	−0.008
5-years	1.000	0.558	0.107	−0.004	0.001−0.003i	0.001+0.003i
Panel B						
1-month	1.000	0.988	0.928	0.819	0.660	0.458
1-year	1.000	0.865	0.475	0.205	0.054	0.009
2-years	1.000	0.767	0.200	0.011	0.000	−0.013
3-years	1.000	0.688	0.165	0.010	−0.004	−0.026
4-years	1.000	0.629	0.156	0.004	−0.004	−0.024
5-years	1.000	0.575	0.103	−0.002−0.005i	−0.002+0.005i	−0.005

either rating 1 and rating 5. They are also greater for absolute ratings than for relative ones.

3.5 Time Heterogeneous Transition Matrices

In fact, it is easily seen that the transition matrices Π_t depend on time. Table 6 (Panel A–D) provides the 2-years transition matrices in January 1996 and June 2006, for absolute and relative rating. We note that the transition matrices change considerably between January 1996 and June 2006. For instance, the diagonal terms decrease, whereas the transitions to inactivity increase. Moreover, the difference between transition matrices for absolute and relative ratings can become much more important date by date. For instance, the probability to stay in grade 1 for absolute rating is twice the corresponding probability for relative rating.

We also provide in Table 6 (Panel A–D) the eigenvalues of the transition matrices. The spectral decomposition allows for a discussion of the idiosyncratic persistence and cycle phenomena. The 3 first eigenvalues (after the unitary one) are real positive, and possible pseudo periodic effects appear for smaller eigenvalues. The first eigenvalue (after the unitary one) is close to one in 1996 for both absolute and relative ratings, but much smaller in 2006. This shows that idiosyncratic persistence can exist at some date, and disappear at some other ones. Similarly the number of eigenvalues (except the unitary one), which are sufficiently large in modulus (larger than 0.1, say) is equal to 5, 2, 3 and 2, respectively. Thus the estimated rank of the transition matrix is also highly varying. Among these significant eigenvalues, we observe only one

Table 5 Joint transition proba. at a 1-month (*Panel A*) and 2-years (*Panel C*); Migration correlations at a 1-month (*Panel B*) and 2-years (*Panel D*) horizon (static factor, relative performance); Joint transition proba. at a 1-month (*Panel E*) and 2-years (*Panel G*); Migration correlations at a 1-month (*Panel F*) and 2-years (*Panel H*) horizon (static factor, absolute performance)

Joint rating arrivals					
Initial rating	Unchanged (%)	Upgrade (%)	Downgrade (%)	Non-rated (%)	Similar rating (%)
Panel A					
(Q1,Q1)	81.12	0.00	89.91	0.01	82.16
(Q2,Q2)	57.69	6.21	69.27	0.01	61.01
(Q3,Q3)	48.51	9.20	59.70	0.01	53.33
(Q4,Q4)	51.55	10.39	60.48	0.03	55.58
(Q5,Q5)	74.92	9.66	0.00	0.09	76.36
Panel B					
Initial rating		Unchanged		Non-rated	
(Q1,Q1)		−0.04		1.16	
(Q2,Q2)		0.06		1.01	
(Q3,Q3)		0.14		1.10	
(Q4,Q4)		0.05		0.92	
(Q5,Q5)		−0.05		1.10	
Panel C					
Joint rating arrivals					
Initial rating	Unchanged	Upgrade	Downgrade	Non-rated	Similar rating
(Q1,Q1)	11.88	0.00	29.97	1.92	22.17
(Q2,Q2)	3.29	2.85	12.55	2.39	18.22
(Q3,Q3)	3.45	4.92	10.01	4.30	18.46
(Q4,Q4)	3.60	7.50	6.60	6.78	19.51
(Q5,Q5)	3.53	7.32	0.00	13.69	24.05
Panel D					
Initial rating		Unchanged		Non-rated	
(Q1,Q1)		0.71		3.79	
(Q2,Q2)		0.54		3.39	
(Q3,Q3)		0.56		3.37	
(Q4,Q4)		0.59		2.51	
(Q5,Q5)		2.84		0.99	
Panel E					
Joint rating arrivals					
Initial rating	Unchanged	Upgrade	Downgrade	Non-rated	Similar rating
(Q1,Q1)	81.37	0.00	89.71	0.01	82.85
(Q2,Q2)	56.02	6.65	67.43	0.01	60.94
(Q3,Q3)	46.35	9.59	57.48	0.02	53.27
(Q4,Q4)	50.69	10.73	59.29	0.03	56.09
(Q5,Q5)	74.07	9.78	0.00	0.11	75.92

(continued)

Table 5 (continued)

Joint rating arrivals					
Initial rating	Unchanged (%)	Upgrade (%)	Downgrade (%)	Non-rated (%)	Similar rating (%)
Panel F					
Initial rating		Unchanged		Non-rated	
(Q1,Q1)		0.03		1.26	
(Q2,Q2)		0.20		1.15	
(Q3,Q3)		0.31		1.16	
(Q4,Q4)		0.21		1.07	
(Q5,Q5)		−0.01		1.29	
Panel G					
Joint rating arrivals					
Initial rating	Unchanged	Upgrade	Downgrade	Non-rated	Similar rating
(Q1,Q1)	13.82	0.00	30.30	1.96	24.95
(Q2,Q2)	3.08	2.83	11.40	2.40	21.38
(Q3,Q3)	2.99	4.14	9.40	4.61	20.47
(Q4,Q4)	3.75	6.95	7.13	7.35	21.01
(Q5,Q5)	4.55	7.35	0.00	13.90	25.06
Panel H					
Initial rating		Unchanged		Non-rated	
(Q1,Q1)		4.22		3.93	
(Q2,Q2)		1.96		4.32	
(Q3,Q3)		0.88		4.40	
(Q4,Q4)		0.83		3.59	
(Q5,Q5)		4.23		1.54	

possible periodicity for static factor, relative performance. The complex eigenvalues is close to a pure imaginary one ($\pm 0.107\, i$), which correspond to a periodicity of 4 months.

We provide in Figs. 3 and 4 the evolution of several transition probabilities for different horizons, and for absolute /relative rating. These transitions are aggregated on the initial rating class. We typically observe:
(i) An increase of the probability to become inactive in the recent years;
(ii) A large variability of the upgrade and downgrade probabilities;
(ii) More pronounced cycle effects on the absolute ratings. This stylized fact is partly explained by the lack of invariance of the autocovariance with respect to nonlinear transforms [see the discussion in Appendix A.2].

This dynamic descriptive analysis is completed in Fig. 5, where we report 4 times the cross-sectional covariance between performances and relative ratings. This quantity is equal to the Gini index of the cross sectional performance distribution.

Table 6 Transition matrices at 2-years horizon starting 01/96 (*Panel A*) and 06/06 (*Panel B*) (static factor, relative performance); 01/96 (*Panel C*) and 06/06 (*Panel D*) (absolute performance)

	Q1 (%)	Q2 (%)	Q3 (%)	Q4 (%)	Q5 (%)	NR (%)
Panel A						
Q1	33.33	27.27	16.67	10.61	10.61	1.52
Q2	16.67	21.21	22.73	9.09	19.70	10.61
Q3	9.09	13.64	27.27	33.33	10.61	6.06
Q4	4.55	18.18	15.15	18.18	25.76	18.18
Q5	7.69	4.62	13.85	15.38	24.62	33.85
NR	0.00	0.00	0.00	0.00	0.00	100.00
Eigenvalues	1.00	0.851	0.278	0.125	−0.004 −0.1i	−0.004 +0.1i
Panel B						
Q1	19.86	12.54	10.80	10.10	12.54	34.15
Q2	14.29	14.98	16.38	14.63	9.76	29.97
Q3	10.14	11.89	11.19	14.69	11.89	40.21
Q4	5.23	8.71	11.85	9.76	15.33	49.13
Q5	6.29	9.09	11.54	10.49	13.64	48.95
NR	0.00	0.00	0.00	0.00	0.00	100.00
Eigenvalues	1.00	0.529	0.107	0.023	−0.005	−0.024
Panel C						
Q1	64.52	19.35	3.23	6.45	3.23	3.23
Q2	32.50	37.50	10.00	17.50	2.50	0.00
Q3	25.56	17.78	16.67	21.11	8.89	10.00
Q4	7.29	20.83	21.88	28.13	8.33	13.54
Q5	9.72	9.72	15.28	15.28	18.06	31.94
NR	0.00	0.00	0.00	0.00	0.00	100.00
Eigenvalues	1.00	0.936	0.439	0.191	0.082	0.000
Panel D						
Q1	13.44	11.07	13.04	13.44	16.60	32.41
Q2	5.93	11.02	15.25	20.90	14.41	32.49
Q3	4.07	7.85	12.50	13.66	19.48	42.44
Q4	1.01	4.03	14.09	11.41	20.81	48.66
Q5	3.80	4.89	8.70	15.22	17.39	50.00
NR	0.00	0.00	0.00	0.00	0.00	100.00
Eigenvalues	1.00	0.553	0.101	0.050	−0.02−0.02i	−0.02+0.02i

We observe very clearly the cycle effects on the heterogeneity of performances which is more pronounced at the beginning of the period.

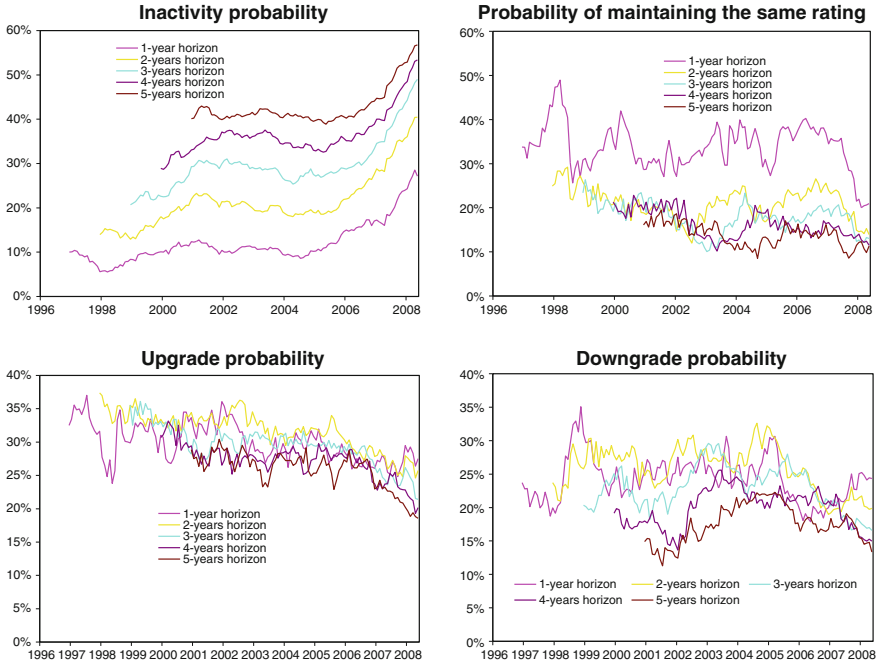


Fig. 3 Inactivity, upgrade and downgrade probabilities (static factor, relative performance)

3.6 Stochastic Transition

We have noted the variability of the different elements of the transition matrices, and also the presence of persistence and of cycle effets. This leads to introduce a stochastic dynamic model for the matrice of transition probabilities. In a first step we can consider Vector Autoregressive (VAR) model on these matrices after vectorization:

$$vec(\Pi_t) = c + \Phi vec(\Pi_{t-1}) + u_t, \tag{6}$$

where $vec(\Pi_t)$ is the vector obtained by stacking the columns of the transition matrix. This vector has dimension 30. The VAR specification is only a first step in the dynamic analysis since it does not account for the constraint $0 \leq \Pi_{j,t} \leq 1, \forall i, j$ to be satisfied by the migration probabilities [see [7, 12] for stochastic migration models taking into account these inequalities]. However the VAR model has the advantage of simplicity and estimation methods easy to implement. In particular the elements of the intercept (30 parameters) and the autoregressive matrix (900 parameters) can be estimated by ordinary least squares. It is not possible to display these elements due to the large dimension, but we provide in Table 7 (Panel A, B) the eigenvalues of matrix $\hat{\Phi}$ for relative and absolute ratings.

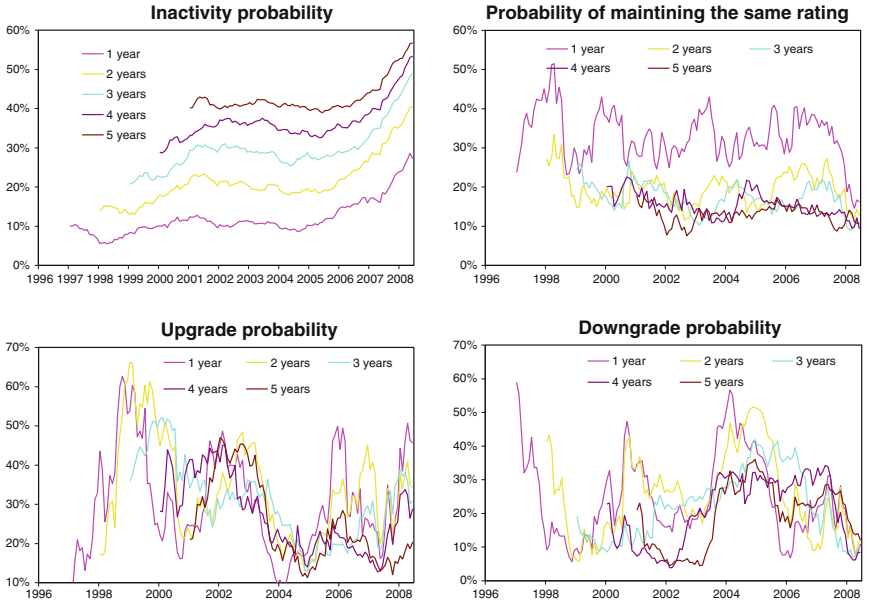


Fig. 4 Inactivity, upgrade and downgrade probabilities (static factor, absolute performance)

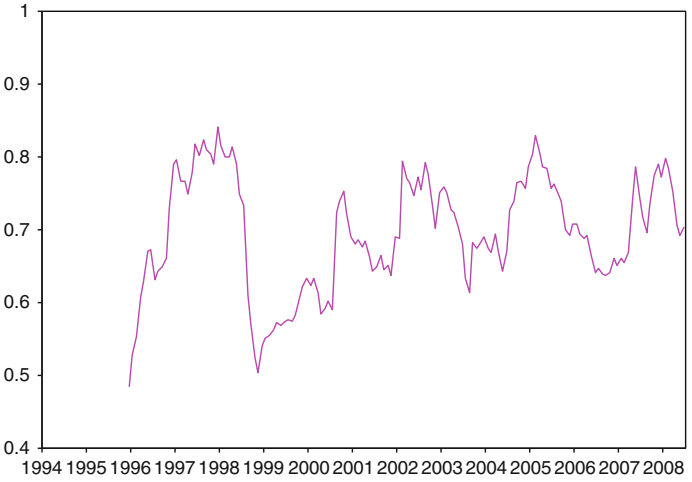


Fig. 5 Gini concentration coefficients (relative ratings)

The eigenvalue with the largest modulus is much larger than 1. This explosive trend captures the increasing evolutions of the transitions to inactivity. The next eigenvalue for relative performance is around 0.8, a test of unit root leads to reject the unit root hypothesis. As expected the computation of relative ratings is

Table 7 Eigenvalues of $\hat{\Phi}$ for the 2-years horizon transition probabilities (static factor, relative performance (*Panel A*) and absolute performance (*Panel B*))

v1	v2	v3	v4	v5
27.323	0.7996	0.6494	0.2083	0.1149
Panel A				
v6	v7	v8	v9	v10
0.0798	0.0534	0.0484	0.0382	0.0291
v11	v12	v13	v14	v15
$0.0260+0.0046$i	$0.0260-0.0046$i	$0.0062+0.0093$i	$0.0062-0.0093$i	0.0182
v16	v17	v18	v19	v20
0.0157	0.0136	0.0085	$-0.0009+0.0051$i	$-0.0009-0.0051$i
v21	v22	v23	v24	v25
$0.0043+0.0027$i	$0.0043-0.0027$i	$-0.0030+0.0003$i	$-0.0030-0.0003$i	0.0016
v26	v27	v28	v29	v30
-0.0005	0	0	$-0.0000+0.0000$i	$-0.0000-0.0000$i
Panel B				
v1	v2	v3	v4	v5
24.7552	2.0518	0.9992	0.651	1 0.2229
v6	v7	v8	v9	v10
0.1149	0.0758	$0.0567+0.0080$i	$0.0567-0.0080$i	$0.0438+0.0078$i
v11	v12	v13	v14	v15
$0.0438-0.0078$i	0.0358	0.0249	$0.0064+0.0125$i	$0.0064-0.0125$i
v16	v17	v18	v19	v20
0.016	$0.0129+0.0078$i	$0.0129-0.0078$i	$0.0072+0.0073$i	$0.0072-0.0073$i
v21	v22	v23	v24	v25
-0.0038	$0.0040+0.0038$i	$0.0040-0.0038$i	0.0036	$-0.0007+0.0011$i
v26	v27	v28	v29	v30
$-0.0007-0.0011$i	0	0	0	0

a way of diminishing the persistence effects, especially when comparing with the results for absolute ratings [see the discussion in Appendix A.2]. Similarly, pseudo-periodic affects are more significant for absolute ratings. The eigenvalues v8 and v9 in Table 7 (Panel B) feature a rather small imaginary part, which indicates rather long periodicities to be compared to the 4-month idiosyncratic periodicities seen before).

3.7 Duration Analysis

The knowledge of transition matrices allows for deriving the distribution of the time of entry in inactivity by simulation. More precisely, let us consider this question for the different models:
(i) For an homogenous Markov chain, the simulations are performed by drawing the sequence of ratings with the estimated matrix $\hat{\Pi}_t$.
(ii) In the static case, where the stochastic transition matrices are independent identically distributed, at each date the next rating will be drawn by using a transition matrix randomly selected among the estimated matrices $\hat{\Pi}_t$, $t = 1, , T$.
(ii) In the dynamic case, the method requires the introduction of a dynamics for the transition matrix.

The duration analysis can be used to compute the Time-at-Risk (TaR), which is the threshold at 5 % for residual lifetime [see [16]].

4 Conclusion

Fund performances are often summarized by means of qualitative ratings computed from adjusted risk performances. These ratings can be absolute when the performance level is taken into account, or relative when the fund is simply ranked against its competitors at a given date. Both absolute and relative ratings are interpretable as qualitative ranks, deduced from the performance by applying the historical and cross-sectional c.d.f., respectively. The aim of our paper was to study and compare the dynamics of absolute and relative ratings by means of stochastic migration models. These models are appropriate for distinguishing the idiosyncratic and systematic persistence (resp. cycle) effects, and to compare the decompositions obtained for absolute and relative ratings.

The stochastic transition models are factor models with transition probabilities as factors. They are easily implemented to derive by simulation the distribution of fund lifetime and analyze how the distribution depends on the current rating and date. These models are preferred to factor models directly written on lifetimes. Indeed, the lifetime is affected by the path of the stochastic transition probabilities during the whole life of the fund.

Acknowledgments We gratefully acknowledge financial support of the chair QUANTVALLEY/Risk Foundation: “Quantitative Management Initiative”, the Global Risk Institute and the chair ACPR: “Regulation and Systemic Risk”.

Appendix A: Properties of the Theoretical Ranks

We discuss in this appendix different properties of the theoretical ranks, such as the link between levels and ranks, or the links between the cross-sectional ranks and the historical ranks. We start by reviewing some basic results without in mind the applications to fund performances and rankings. Then we particularize the results to the ranking of funds.

A.1 Basic Properties

A.1.1 Definition of a Theoretical Rank

Let us consider a one-dimensional continuous random variable X and denote by F its cumulative distribution function: $F(x) = P(X < x)$. If its continuous distribution admits a strictly positive probability density function (p.d.f), then the theoretical rank of X is the random variable:

$$Y = F(X). \tag{1}$$

The variable Y takes values in $[0, 1]$ and follows the uniform distribution on $(0, 1)$.

Example A.1 To understand the definition of the theoretical rank, let us consider a large set of funds, whose performances for period t are $X_{i,t}$, $i = 1, \ldots, n$, say. We can construct the empirical cross-sectional c.d.f. $\hat{F}_{n,t} = \frac{1}{n}\sum_{i=1}^{n} 1_{X_{i,t}<x}$, and the empirical ranks $\hat{Y}_{i,t} = \hat{F}_{n,t}(X_{i,t})$. Under cross-sectional ergodicity, the empirical cross-sectional c.d.f. tends to a limiting one F_t, say, and the empirical rank $\hat{Y}_{i,t}$ to the theoretical rank $Y_{i,t} = F_t(X_{i,t})$.

A.1.2 Linear Links Between Levels and Ranks

Let us now consider a pair of continuous variable (X_1, X_2), with joint and marginal c.d.f. denoted by $F_{1,2}$, F_1 and F_2, respectively: $F_{1,2}(x_1, x_2) = P[X_1 < x_1, X_2 < x_2]$, $F_j = P[X_j < x_j]$, $j = 1, 2$. The associated theoretical ranks are:

$$Y_1 = F_1(X_1), Y_2 = F_2(X_2). \tag{2}$$

Y_1 and Y_2 are marginally uniform on $[0, 1]$, but are dependent if X_1 and X_2 are. This dependence can be characterized by their joint c.d.f.: $C(y_1, y_2) = P[Y_1 < y_1, Y_2 < y_2] = F_{1,2}[F_1^{-1}(y_1), F_2^{-1}(y_2)]$, called the copula of X_1 and X_2. Let us now discuss the links between all these variables, when we focus on linear links measured by the Pearson correlations. The joint variance-covariance matrix of levels and ranks is given by:

$$V\begin{pmatrix} X_1 \\ X_2 \\ Y_1 \\ Y_2 \end{pmatrix} = \begin{pmatrix} VX_1 & Cov(X_1, X_2) & Cov(X_1, Y_1) & Cov(X_1, Y_2) \\ . & VX_2 & Cov(X_2, Y_2) & Cov(X_2, Y_2) \\ . & . & VY_1 & Cov(Y_1, Y_2) \\ . & . & . & VY_2 \end{pmatrix}. \tag{3}$$

The different elements of this variance-covariance matrix are of the form: $Cov[G_1(X_1), G_2(X_2)]$, where G_1, G_2 are two increasing functions, either the identity function, or the c.d.f. We have the following Lemma:

Lemma A.1 *If G_1, G_2 are increasing functions:*

$$Cov[G_1(X_1), G_2(X_2)] = \int\int [F_{1,2}(x_1, x_2) - F_1(x_1)F_2(x_2)]dG_1(x_1)dG_2(x_2). \tag{4}$$

In fact, a functional measure of dependence is the difference between the joint c.d.f. and what would be its expression under the independence assumption, i.e. $F_{1,2}(x_1, x_2) - F_1(x_1)F_2(x_2)$. Lemma A.1 says that any covariance of this type is a weighted average of this functional measure, with weights equal to the derivatives of functions G_j. In particular, the variables X_1 and X_2 are independent iff $F_{1,2}(x_1, x_2) = F_1(x_1)F_2(x_2), \forall x_1, x_2$, or equivalently iff $Cov[G_1(X_1), G_2(X_2)] = 0$, for any increasing functions G_1, G_2.

The formula (A.4) in Lemma A.1. provides comparable expressions of the elements of the joint variance-covariance matrix (A.3). We get:

$$Cov(X_1, X_2) = \int\int [F_{1,2}(x_1, x_2) - F_1(x_1)F_2(x_2)]dx_1dx_2, \tag{5}$$

$$Cov[X_1, F_2(X_2)] = \int\int [F_{1,2}(x_1, x_2) - F_1(x_1)F_2(x_2)]dx_1dF_2(x_2), \tag{6}$$

$$Cov[F_1(X_1), X_2] = \int\int [F_{1,2}(x_1, x_2) - F_1(x_1)F_2(x_2)]dF_1(x_1)dx_2, \tag{7}$$

$$Cov[F_1(X_1), F_2(X_2)] = \int\int [F_{1,2}(x_1, x_2) - F_1(x_1)F_2(x_2)]dF_1(x_1)dF_2(x_2). \tag{8}$$

It is known that the covariance operator is invariant by linear affine transformations of the variables. However we pass from levels to ranks by a nonlinear transform, i.e. the c.d.f. This transformation can imply a loss of information concerning linear links. For instance, by using the Frechet upper bound [10]:

$$F_{1,2}(x_1, x_2) \leq \min[F_1(x_1), F_2(x_2)], \tag{9}$$

we get:

$$Cov[X_1, F_2(X_2)] \leq \int\int [\min[F_1(x_1)F_2(x_2)] - F_1(x_1)F_2(x_2)]\, dx_1 dF_2(x_2)$$
$$\leq Cov[X_1, F_1(X_1)]. \tag{10}$$

Similarly, by considering the Frechet lower bound [[24], Proposition 4], we get an inequality in the other direction. Finally, we know:

$$-Cov[X_1, F_1(X_1)] \leq Cov[X_1, F_2(X_2)] \leq Cov[X_1, F_1(X_1)]. \tag{11}$$

A.1.3 Correlations Between Levels and Ranks

Let us now focus on the correlation matrix:

$$R = Corr\begin{pmatrix} X_1 \\ X_2 \\ F_1(X_1) \\ F_2(X_2) \end{pmatrix} = \begin{pmatrix} 1 & \rho_{12} & \lambda_{11} & \lambda_{12} \\ . & 1 & \lambda_{21} & \lambda_{22} \\ . & . & 1 & r_{12} \\ . & . & . & 1 \end{pmatrix}. \tag{12}$$

$\rho_{1,2}$ is the standard Pearson correlation, $r_{1,2}$ is the correlation of ranks defined by [26], $\lambda_{1,1}$ and $\lambda_{2,2}$ are the L-moments of order 2 introduced by [18], $\lambda_{1,2}$ and $\lambda_{2,1}$ are the L-comoments of order 2 [see e.g. [23, 24]].

By considering the whole matrix R, we perform a joint analysis of these different dependence measures. These correlations are constrained. First, matrix R is a positive semi-definite matrix. Second, additional constraints are due to the deterministic increasing relationship between X_1 and $F_1(X_1)$ [resp. X_2 and $F_2(X_2)$]. For instance, we know that: $\lambda_{11} \geq 0$, $\lambda_{22} \geq 0$, $|\lambda_{12}| \leq \min(\lambda_{11}, \lambda_{22})$, $|\lambda_{21}| \leq \min(\lambda_{11}, \lambda_{22})$, from inequalities (A.12).

These constraints can be explicited in special symmetric cases, where $\lambda_{11} = \lambda_{22} = \lambda$, say, $\lambda_{12} = \lambda_{21} = \mu$, say.

Proposition *The different correlations* $\rho = \rho_{12}$, $r = r_{12}$, $\lambda = \lambda_{11} = \lambda_{22}$, $\mu = \lambda_{12} = \lambda_{21}$ *are constrained as follows:*

(i) *If* $\rho = 1$, *then* $r = 1$, $\mu = \lambda$ *with* $0 \leq \lambda \leq 1$.
(ii) *If* $\rho = -1$, *then* $r = -1$, $\mu = -\lambda$ *with* $0 \leq \lambda \leq 1$.
(iii) *If* $\rho \neq \pm 1$, *we get:*
$0 < |\mu| < \lambda$,
$\lambda^2 + \mu^2 - 2\rho\lambda\mu \leq 1 - \rho^2$,
$\lambda^2 - \mu^2 \leq \sqrt{(1 - r^2)(1 - \rho^2)}$.

The first inequality in (iii) *is equivalent to:*

$$\frac{1}{2(1+\rho)}(\lambda+\mu)^2+\frac{1}{2(1-\rho)}(\lambda-\mu)^2\leq 1.$$

It corresponds to an ellipsoid with the 45° and −45° lines as it main axes. The second inequality corresponds to the interior of an hyperbola with the same axes as the ellipsoid above. The domain of admissible value for λ, μ for given values of r, ρ is the intersection of the hyperbolic and ellipsoidal domain.

A.2 Rank Autocorrelation Functions

A.2.1 The Variables

The results in Appendix A.1 are useful to interpret the joint dynamics of performances, absolute ranks and relative ranks. Recall that these variables are defined as follows:

$S_{i,t}$ denotes the risk-adjusted Sharpe performance for HF i and period t.

For these data, we can define the historical c.d.f. of S computed on averaging on both HF and dates, and deduce the associated absolute ranks as $r^a_{i,t}=\hat{F}(S_{i,t})$.
We can also consider the cross-sectional c.d.f. at date t, obtained by averaging on HF for given t, and deduce the associated relative rank as $r_{i,t}=\hat{F}_t(S_{i,t})$.
Thus the results of Appendix A.1 can be used with the following $[X_1, X_2, F_1(X_1), F_2(X_2)]$ variables:
(i) Joint analysis of levels and absolute ranks at different dates:
$X_1=S_{i,t}$, $X_2=S_{i,t-h}$, $F_1(X_1)=r^a_{i,t}$, $F_2(X_2)=r^a_{i,t-h}$.
(ii) Joint analysis of levels and relatives ranks at different dates:
$X_1=r^a_{i,t}$, $X_2=r^a_{i,t-h}$, $F_1(X_1)=r_{i,t}$, $F_2(X_2)=r_{i,t-h}$.

A.2.2 Autocorrelograms

When the variables X_1, X_2 (resp. $F_1(X_1)$, $F_2(X_2)$) correspond to an observation of the same variable at different dates $t, t-h$, the correlation matrix in (A.12) becomes the autocorrelation at lag h for level and ranks (or for absolute and relative ranks):

$$R(h)=\begin{pmatrix}1 & \rho_h & \lambda(0) & \lambda(h)\\ . & 1 & \lambda(-h) & \lambda(0)\\ . & . & 1 & r(h)\\ . & . & . & 1\end{pmatrix}, \tag{13}$$

where for instance:

$\rho(h)$ is the ACF on absolute ranks,

$r(h)$ is the ACF on relative ranks,

$\lambda(h)$ is the cross ACF between absolute and relative ranks.

References

1. Aragon, G., Nanda, V.: Tournament behavior in hedge funds: high-water marks, fund liquidation and managerial skills. Rev. Financ. Stud. **25**, 937–974 (2012)
2. The new basel capital accords. Second consultative paper, Bank of International Settlements, January (2001)
3. Chakroun, O.: Migration dependence among the U.S. business sectors. DP HEC Montreal (2008)
4. Darolles, S., Gourieroux, C.: Conditionally fitted sharpe performance with an application to hedge fund rating. J. Bank. Financ. **34**, 578–593 (2010)
5. Darolles, S., Gourieroux, C.: The effects of management and provision accounts on hedge fund returns. Part 1: the highwater mark scheme; Part 2: the loss carry forward scheme. In: Huynh, V., et al. (eds.) Advances in Intelligence Systems and Computing, vol. 251, pp. 23–72. Springer, Berlin (2014)
6. Darolles, S., Gourieroux, C., Jasiak, J.: L-performance with an application to hedge funds. J. Empir. Financ. **16**, 671–685 (2009)
7. Feng, C., Gourieroux, C., Jasiak, J.: The ordered qualitative model for credit rating transition. J. Empir. Financ. **15**, 111–130 (2008)
8. Foulcher, S., Gourieroux, C., Tiomo, A.: Term structure of default and ratings. Insur. Risk Manag. **72**, 207–276 (2004)
9. Foulcher, S., Gourieroux, C., Tiomo, A.: Migration correlation: estimation, methods and application to french corporate ratings. Annales d'Economie et Statistique **82**, 71–102 (2007)
10. Frechet, M.: Sur les tableaux de corrélation dont les marges sont données. Annales de l'Université de Lyon, Section A, Série **3**(14), 53–57 (1951)
11. Fung, W., Hsieh, D.: Hedge funds: an industry in its adolescence. Fed. Reserve Bank Atlanta Econ. Rev. **91**, 1–33 (2006)
12. Gagliardini, P., Gourieroux, C.: Stochastic migration models with application to corporate risk. J. Financ. Econom. **3**, 188–226 (2005)
13. Gagliardini, P., Gourieroux, C.: Migration correlation: definition and efficient estimation. J. Bank. Financ. **29**, 865–894 (2005)
14. Gagliardini, G., Gourieroux, C., Rubin, R.: Positional Portfolio Management. CREST DP (2014)
15. Garnier, O., Pujol, T.: Les Etoiles d'Aujourd'hui Préjugent-elles des Etoiles de Demain? Les Cahiers Scientifiques de l'AMF, 3 (2007)
16. Ghysels, E., Gourieroux, C., Jasiak, J.: Stochastic volatility duration models. J. Econom. **119**, 419–433 (2003)
17. Gourieroux, C., Jouneau, F.: Econometrics of efficient fitted portfolios. J. Empir. Financ. **6**, 87–118 (1999)
18. Hosking, J.: L-moments: analysis and estimation of distributions using linear combinations of order statistics. J. R. Stat. Soc. Ser. A **52**, 105–124 (1990)
19. Kosowski, R., Naik, N., Teo, M.: Do Hedge funds deliver alpha? A Bayesian and bootstrap analysis. J. Financ. Econ. **84**, 229–264 (2007)
20. Lando, D.: Credit Risk Modelling: Theory and Applications. Princeton University Press, Princeton (2004)
21. Sharpe, W.: Mutual fund performance. J. Bus. **39**, 119–138 (1966)

22. Sharpe, W.: The Sharpe ratio. J. Portf. Manag. **21**, 49–58 (1994)
23. Stuart, A.: The correlation between variable values and ranks in samples from a continuous distribution. Br. J. Stat. Psychol. **7**, 37–44 (1954)
24. Serfling, R., Xiao, P.: A contribution to multivariate L-moments: L-comoment. Matrices, J. Multivar. Anal. **98**, 1765–1781 (2007)
25. Sortino, F., Price, L.: Performance measurement in a downside risk framework. J. Invest. **3**, 59–65 (1994)
26. Spearman, C.: The proof and measurement of association between two things. Am. Psychol. **15**, 72–101 (1904)
27. Vasicek, O.: Limiting Loan Loss Probability Distribution. DP KMV Corporation (1991)

The Joint Belief Function and Shapley Value for the Joint Cooperative Game

Zheng Wei, Tonghui Wang, Baokun Li and Phuong Anh Nguyen

Abstract In this paper, the characterization of the joint distribution of random set vectors by the belief function is investigated and the joint game in terms of the characteristic function is given. The bivariate Shapley value of a joint cooperative game is obtained through both cores and games. Formulas for the Shapley value derived from two different methods are shown to be identical. For illustration of our main results, several examples are given.

1 Introduction

Random sets can be used to model imprecise observations of random variables where the outcomes are assigned as set valued instead of real valued. The theory of random sets is viewed as a natural generalization of multivariate statistical analysis. Random set data can also be viewed as imprecise or incomplete observations which are frequent in today's technological societies. The distribution of the univariate random set and its properties can be found in Dempster [2], Shafer [13], Nguyen and Wang [8], Nguyen [5], and Li and Wang [4]. Recently, the characterization of joint distributions of random sets on co-product spaces was discussed by Schmelzer [10] and Nguyen [6]. In this paper, this characterization is modified for discrete random set vector.

Z. Wei · T. Wang
Department of Mathematical Sciences, New Mexico State University, Las Cruces, USA
e-mail: weizheng@nmsu.edu

T. Wang (✉)
Innovation Experimental College, Northwest A & F University, Yangling, China
e-mail: twang@nmsu.edu

B. Li
School of Statistics, Southwestern University of Finance and Economy, Chengdu, China
e-mail: bali@swufe.edu.cn

P.A. Nguyen
International University, Vietnam National University, Ho Chi Minh, Vietnam
e-mail: npanh@hcmiu.edu.vn

V.-N. Huynh et al. (eds.), *Econometrics of Risk*, Studies in Computational Intelligence 583, DOI 10.1007/978-3-319-13449-9_8

In the game theory, a cooperative game is a game where groups of players may enforce cooperative behaviour, see Burger [1]. The Shapley value is a solution concept in the cooperative game theory aiming to propose the fairest allocation of collectively gained profits between the several collaborative agents. Let E be a finite set of N players. A game is a mapping $\nu : 2^E \to \mathbb{R}$ such that $\nu(\emptyset) = 0$. The **Shapley value** of player i is given by

$$\phi_i(\nu) = \sum_{S \subseteq E \setminus \{i\}} \frac{|S|!(N - |S| - 1)!}{N!} [\nu(S \cup \{i\}) - \nu(S)],$$

where $|S|$ is the cardinality of the set S. A game ν is said to be strategically equivalent a game ν', if there are positive number α and real numbers c_i, where $i \in E$, such that for any coalition $A \subseteq E$, we have $\nu'(A) = \alpha\nu(A) + \sum_{i \in A} c_i$. It is known that every essential coalitional game is strategically equivalent to one and only one game in the $(0, 1)$-reduced form, i.e., $\nu' : 2^E \to [0, 1]$. Note that if ν' satisfies the property of monotone of infinite order, see [5], it can be treated as special case of belief function. For the computational aspect of the Shapley value in the univariate case, see Li and Wang [4]. It is natural to extend univariate coalitional games into bivariate cases using the bivariate belief functional, which has not been discussed in literature. Also the Shapley value of a bivariate coalitional game can be extended to the cases of the bivariate random sets. In this paper, a formula for calculating the bivariate Shapley value is provided.

In literature, the theory of games have been extended to the game with vector-valued payoffs, see Fernandez et al. [3] and Roemer [9]. The vector-valued game is a game with n players, each of whom has several goals. For example, in the World Figure Skating Championship competition, the total points for the performance of each figure skater or each pair of skaters is the weighted combination of two scores, short program and free skating. In this paper, we will use the proposed joint game to analyze the bivariate vector-valued game and treat each component as the marginal of the joint game. Thus, the Shapley value of the bivariate vector-valued game can be computed as those of marginal games.

The following example can be considered as an application of the joint coalition game. In China's transportation energy market, the gasoline and the natural gas are two main complementary energy sources, both are used in most large cities for cars and buses. Suppose there are n_1 companies producing gasoline and n_2 companies producing nature gas. Each grade of gasoline is produced solely by a company or produced jointly by several companies. Similarly situations are applied to natural gas companies. Let $\nu(A, B)$ be the amount of sales (say in million dollars), where $A \subseteq \{1, \ldots, n_1\}$ and $B \subseteq \{n_1 + 1, \ldots, n_1 + n_2\}$ are coalitions of gasoline companies and natural gas companies, respectively. In each gas station, grades of gasoline and grades of natural gas are sold. The intersection of two sets of companies may not be empty, since some companies produce both gasoline and natural gas. Given the data of total sales of all combinations of grades of gasoline and natural gas, how to estimate the sales attributed to each combination of ith gasoline company and jth natural gas company, besides the sales attributed to each ith gasoline company

and jth natural gas company individually? In this paper, the framework of the joint cooperative game is established.

This paper is organized as follows. The characterization of the joint distribution of random set vector by its joint belief functions is obtained in Sect. 2. As an application of random set vector, the bivariate coalitional game and its properties are investigated in Sect. 3. In Sect. 4, the Shapley value of a joint cooperative game is obtained through cores and games. The formulas for calculating bivariate Shapley value derived by two methods are shown to be identical. To illustrate our main results, several examples are given.

2 The Characterization of the Joint Belief Function of Discrete Random Set Vector

Throughout this paper, let $(\Omega, \mathscr{A}, P)$ be a probability space and let E_1 and E_2 be finite sets, where Ω is sample space, $\mathscr{A}$ is a σ-algebra on subsets of Ω and P is a probability measure.

Recall that a finite random set $\mathscr{S}$ with values in powerset of a finite E is a map $\mathscr{S} : \Omega \to 2^E$ such that $\mathscr{S}^{-1}(\{A\}) = \{\omega \in \Omega : \mathscr{S}(\omega) = A\} \in \mathscr{A}$ for any $A \subseteq E$. Let $f : 2^E \to [0, 1]$ be $f(A) = P(\mathscr{S} = A)$, then f is a probability density function of $\mathscr{S}$ on 2^E. In the following, we will extend this definition to the case of the bivariate random set vector.

Definition 2.1 Let E_1 and E_2 be two finite sets. A bivariate random set vector $(\mathscr{S}_1, \mathscr{S}_2)$ with values in $2^{E_1} \times 2^{E_2}$ is a map $(\mathscr{S}_1, \mathscr{S}_2) : \Omega \to 2^{E_1} \times 2^{E_2}$ such that $\{\omega \in \Omega : \mathscr{S}_1(\omega) = A, \mathscr{S}_2(\omega) = B\} \in \mathscr{A}$, for any $A \subseteq E_1$ and $B \subseteq E_2$. The function $h : 2^{E_1} \times 2^{E_2} \to [0, 1]$ is said to be a **joint probability density function** of $(\mathscr{S}_1, \mathscr{S}_2)$ if $h \geq 0$ and $\sum_{A \subseteq E_1} \sum_{B \subseteq E_2} h(A, B) = 1$, where $h(A, B) = P(\mathscr{S}_1(\omega) = A, \mathscr{S}_2(\omega) = B)$, $A \subseteq E_1$, and $B \subseteq E_2$.

Inspired by the distribution of univariate random sets, we are going to define axiomatically the concept of joint distribution functions of the random set vector $(\mathscr{S}_1, \mathscr{S}_2)$. Let $(\mathscr{S}_1, \mathscr{S}_2)$ be a (nonempty) random set vector on $2^{E_1} \times 2^{E_2}$ and $H : 2^{E_1} \times 2^{E_2} \to [0, 1]$ be

$$H(A, B) = P(\mathscr{S}_1 \subseteq A, \mathscr{S}_2 \subseteq B) = \sum_{C \subseteq A} \sum_{D \subseteq B} h(C, D), \quad A \in 2^{E_1}, \quad B \in 2^{E_2}. \quad (1)$$

It can be shown that H satisfies the following properties:

(i) $H(\emptyset, \emptyset) = H(\emptyset, B) = H(A, \emptyset) = 0$, and $H(E_1, E_2) = 1$;
(ii) H is **monotone of infinite order on each component**, i.e., for any B in 2^{E_2} and any distinct sets $A_1, A_2, \ldots, A_k$ in 2^{E_1}, $k \geq 1$,

$$H\left(\bigcup_{i=1}^{k} A_i, B\right) \geq \sum_{\emptyset \neq I \subseteq \{1,2,\ldots,k\}} (-1)^{|I|+1} H\left(\bigcap_{i \in I} A_i, B\right), \quad (2)$$

and for any $A \in 2^{E_1}$ and any distinct sets $B_1, B_2, \ldots, B_\ell$ in 2^{E_2}, $\ell \geq 1$,

$$H\left(A, \bigcup_{j=1}^{\ell} B_j\right) \geq \sum_{\emptyset \neq J \subseteq \{1,2,\ldots,\ell\}} (-1)^{|J|+1} H\left(A, \bigcap_{j \in J} B_j\right), \tag{3}$$

and

(iii) $H(.,.)$ is **jointly monotone of infinite order**, i.e., for distinct sets $A_1, A_2, \ldots, A_k$ in 2^{E_1} and distinct $B_1, B_2, \ldots, B_\ell$ in 2^{E_2}, where k, ℓ are positive integers,

$$\begin{aligned} H\left(\bigcup_{i=1}^{k} A_i, \bigcup_{j=1}^{\ell} B_j\right) \geq & - \sum_{\emptyset \neq I \subseteq \{1,2,\ldots,k\}} \sum_{\emptyset \neq J \subseteq \{1,2,\ldots,\ell\}} (-1)^{|I|+|J|} H\left(\bigcap_{i \in I} A_i, \bigcap_{j \in J} B_j\right) \\ & + \sum_{\emptyset \neq I \subseteq \{1,2,\ldots,k\}} (-1)^{|I|+1} H\left(\bigcap_{i \in I} A_i, \bigcup_{j=1}^{\ell} B_j\right) \\ & + \sum_{\emptyset \neq J \subseteq \{1,2,\ldots,\ell\}} (-1)^{|J|+1} H\left(\bigcup_{i=1}^{k} A_i, \bigcap_{j \in J} B_j\right). \end{aligned} \tag{4}$$

It turns out that the properties (i), (ii) and (iii) of H above characterize the joint distribution function of a (nonempty) random set vector.

Definition 2.2 A set function $H : 2^{E_1} \times 2^{E_2} \to [0, 1]$ satisfying the properties (i), (ii) and (iii) is said to be the **joint belief function** of random set vector $(\mathscr{S}_1, \mathscr{S}_2)$.

Given any given joint belief function H of $(\mathscr{S}_1, \mathscr{S}_2)$, there exists a probability density function $h : 2^{E_1} \times 2^{E_2} \to [0, 1]$ corresponding to H. In fact, let $H : 2^{E_1} \times 2^{E_2} \to [0, 1]$ be such that

(i) $H(\emptyset, \emptyset) = H(\emptyset, B) = H(A, \emptyset) = 0$, and $H(E_1, E_2) = 1$,

(ii) H is monotone of infinite order on each component, and

(iii) H is joint monotone of infinite order. Then for any $(A, B) \in 2^{E_1} \times 2^{E_2}$, there exists a nonnegative set function $h : 2^{E_1} \times 2^{E_2} \to [0, 1]$, called the **Möbius inverse** of H, such that

$$H(A, B) = \sum_{C \subseteq A} \sum_{D \subseteq B} h(C, D) \tag{5}$$

and

$$\sum_{C \subseteq E_1} \sum_{D \subseteq E_2} h(C, D) = 1. \tag{6}$$

The function $h : 2^{E_1} \times 2^{E_2} \to [0, 1]$ is of the form

$$h(A, B) = \sum_{C \subseteq A} \sum_{D \subseteq B} (-1)^{|A \setminus C| + |B \setminus D|} H(C, D), \tag{7}$$

where $A \setminus C = A \cap C^c$ and C^c is the complement of C.

Given a set function $H : 2^{E_1} \times 2^{E_2} \to [0, 1]$, it is natural to ask whether if it is a well-defined joint belief function. By the conditions in (i), (ii) and (iii) of H, we only need to check that (i)–(iii) hold for all distinct sets $A_1, \ldots, A_k$ and $B_1, \ldots, B_\ell$.

Similar to conditions (i), (ii) and (iii) of H, there is a property called completely monotone in each component, given by Schmelzer [10, 11] and Nguyen [6] as follows.

A set function $H_1 : 2^{E_1} \times 2^{E_2} \to [0, 1]$ is said to be **completely monotone in each component**, if for any $k \geq 2$ and $(A_i, B_i) \in 2^{E_1} \times 2^{E_2}$, $i = 1, 2, \ldots, k$,

$$H_1\left(\bigcup_{i=1}^{k} A_i, \bigcup_{i=1}^{k} B_i\right) \geq \sum_{\emptyset \neq I \subseteq \{1,2,\ldots,k\}} (-1)^{|I|+1} H_1\left(\bigcap_{i \in I} A_i, \bigcap_{i \in I} B_i\right). \tag{8}$$

It can be shown that (8) is equivalent to (ii) and (iii). The difference between two forms is that the (A_i, B_i)'s in (8) are not necessarily distinct sets and may be repeated many times.

3 The Joint Cooperative Game

In the cooperative game theory (see, e.g. Burger [1]), an n-person game G is given by n non-empty sets S_i's, the strategy sets of the players, and u_i's, the pay-off functions of the player $i = 1, 2, \ldots, n$, real-valued functions defined on $S_1 \times S_2 \times \cdots \times S_n$. Such a game is denoted by $G = (S_i, u_i : i \in E = \{1, 2, \ldots, n\})$. The characteristic function of the n-person game G is a set valued function $\nu : 2^E \to \mathbb{R}$ satisfies properties (a) $\nu(\emptyset) = 0$ and (b) Superadditivity: if $A, B \in 2^E$ and $A \cap B = \emptyset$, then $\nu(A) + \nu(B) \leq \nu(A \cup B)$.

Motivated by this, if we consider of the same group of people playing two parts (say, defensive and offensive parts) of a game, we can define the joint characteristic function of the game. Without loss of generality, we assume that $E_1 = \{1, \ldots, n_1\}$ and $E_2 = \{n_1 + 1, \ldots, n_1 + n_2\}$.

Definition 3.1 The bivariate set valued function $\nu : 2^{E_1} \times 2^{E_2} \to \mathbb{R}$ is said to be a **joint characteristic function** for a joint game if it satisfies,

(a) $\nu(\emptyset, \emptyset) = \nu(A, \emptyset) = \nu(\emptyset, B) = 0$, for any $A \subseteq E_1$ and $B \subseteq E_2$;

(b) Superadditivity on each component: for each fixed $A \subseteq E_1$,

$$\nu(A, B_1) + \nu(A, B_2) \leq \nu(A, B_1 \cup B_2) \quad \text{with } B_1, B_2 \subseteq E_2, B_1 \cap B_2 = \emptyset, \tag{9}$$

and for each fixed $B \subseteq E_2$,

$$\nu(A_1, B) + \nu(A_2, B) \leq \nu(A_1 \cup A_2, B) \quad \text{with } A_1, A_2 \subseteq E_2, A_1 \cap A_2 = \emptyset; \quad (10)$$

and,

(c) Joint superadditivity: for any $A_1, A_2 \subseteq E_1$ and $B_1, B_2 \subseteq E_2$, with $A_1 \cap A_2 = \emptyset$, and $B_1 \cap B_2 = \emptyset$,

$$\begin{aligned}\nu(A_1 \cup A_2, B_1 \cup B_2) \geq\ & \nu(A_1 \cup A_2, B_1) + \nu(A_1 \cup A_2, B_2) \\ & + \nu(A_1, B_1 \cup) + \nu(A_2, B_1 \cup B_2) \\ & - \nu(A_1, B_1) - \nu(A_2, B_1) - \nu(A_1, B_2) - \nu(A_2, B_2). \quad (11)\end{aligned}$$

Since there are a great variety of coalitional games, it is desirable to classify them in such a way that those games which belong to the same class will have the same basic properties. This will allow us to consider a single representative game in a class rather the whole class and choose the simplest one, if possible.

Definition 3.2 The joint game $(E_1 \times E_2, \nu)$ is said to be **strategically equivalent** to the game $(E_1 \times E_2, \nu^*)$ if there are positive number α and real numbers c_{ij}, where $i \in E_1$ and $j \in E_2$, such that for any coalition $A \times B \in 2^{E_1 \times E_2}$, we have

$$\nu^*(A, B) = \alpha \nu(A, B) + \sum_{i \in A} \sum_{j \in B} c_{ij}.$$

It is easy to show strategically equivalent relation is an equivalence relation, see Nguyen [7].

A joint game is called **unessential** if $\nu(E_1, E_2) = \sum_{i \in E_1} \sum_{j \in E_2} \nu(\{i\}, \{j\})$, otherwise it is called **essential**, i.e., $\nu(E_1, E_2) > \sum_{i \in E_1} \sum_{j \in E_2} \nu(\{i\}, \{j\})$. Note that for any unessential game $(E_1 \times E_2, \nu)$, it is true that $\nu(A, B) = \sum_{i \in A} \sum_{j \in B} \nu(\{i\}, \{j\})$ for any $A \subseteq E_1, B \subseteq E_2$. Indeed, if not we have $\nu(A, B) > \sum_{i \in A} \sum_{j \in B} \nu(\{i\}, \{j\})$ by super-additivity. Furthermore, we have

$$\begin{aligned}\nu(E_1, E_2) &\geq \nu(E_1, B) + \nu(A, E_2) + \nu(E_1, B^c) + \nu(A^c, E_2) \\ &\quad - \nu(A, B) - \nu(A^c, B) - \nu(A, B^c) - \nu(A^c, B^c) \\ &\geq \nu(A, E_2) + \nu(A^c, E_2) \\ &\geq \nu(A, B) + \nu(A, B^c) + \nu(A^c, B) + \nu(A^c, B^c) \\ &> \sum_{i \in E_1} \sum_{j \in E_2} \nu(\{i\}, \{j\})\end{aligned}$$

which is a contradiction. Also, this implies

$$\nu(E_1, E_2) = \sum_{i \in E_1} \nu(\{i\}, E_2) \quad \text{and} \quad \nu(E_1, E_2) = \sum_{j \in E_2} \nu(E_1, \{j\}).$$

Proposition 3.1 *Every unessential game is strategically equivalent to a trivial game. Every essential game is equivalent to a* $(0, 1)$*-reduced game, i.e., a characteristic function* $v^* : 2^{E_1} \times 2^{E_2} \to [0, 1]$ *such that* $v^*(\{i\}, \{j\}) = 0$ *and* $v^*(E_1, E_2) = 1$.

Proof Let $v : 2^{E_1} \times 2^{E_2} \to \mathbb{R}$ be an unessential game, then for any $A \subseteq E_1$ and $B \subseteq E_2$,

$$v(A, B) = \sum_{i \in A} \sum_{j \in B} v(\{i\}, \{j\}).$$

Let $\alpha = 1$ and $c_{ij} = -v(\{i\}, \{j\})$ for any $i \in E_1$ and $j \in E_2$, v is equivalent to $v^*(A, B) = \alpha v(A, B) + \sum_{i \in A} \sum_{j \in B} v(A, B) = 0$.

Let v be an essential game. Consider the system of equations with unknowns c_{ij} and α given below.

$$v^*(\{i\}, \{j\}) = \alpha v(\{i\}, \{j\}) + c_{ij} = 0$$

and

$$v^*(E_1, E_2) = \alpha v(E_1, E_2) + \sum_{i \in E_1} \sum_{j \in E_2} c_{ij} = 1.$$

The solutions of α and c_{ij}'s are

$$\alpha = \left[v(E_1, E_2) - \sum_{i \in E_1} \sum_{j \in E_2} v(\{i\}, \{j\}) \right]^{-1} \quad \text{and} \quad c_{ij} = -\alpha v(\{i\}, \{j\})$$

so that the joint game $v^*(A, B) = \alpha v(A, B) + \sum_{i \in A} \sum_{j \in B} c_{ij}$ is a game in $(0, 1)$-reduced form. □

Remark 3.1 For any game $v : 2^{E_1} \times 2^{E_2} \to \mathbb{R}$, there exists one unique $(0, 1)$-reduced game $v^* : 2^{E_1} \times 2^{E_2} \to [0, 1]$ such that v^* is strategically equivalent to v. Thus, every v can be uniquely transformed to a characteristic function v^*. Therefore, without loss of generality and for simplicity, we assume that v is the the characteristic function in the $(0, 1)$-reduced form, i.e., $v : 2^{E_1} \times 2^{E_2} \to [0, 1]$. Also note that if we take $A_1, A_2 \subseteq E_1$ and $B_1, B_2 \subseteq E_2$ with $A_1 \cap A_2 = \emptyset$ and $B_1 \cap B_2 = \emptyset$, then (2), (3) and (4) in the definition of the joint belief function will be reduced to (9), (10) and (11), and conditions (i), (ii) and (iii) in the definition of the joint belief function will be reduced to conditions (a), (b) and (c) in the definition of the characteristic function. Therefore, every joint belief function can be treated as a characteristic function. □

Now, assume v is a joint characteristic function, we can define three games. let $S_i = \{A \subseteq E_1 | i \in A\}$ and $T_j = \{B \subseteq E_2 | j \in B\}$, i.e., each player i in E_1, and player j in E_2 has their strategies as subsets to which i and j belongs. Then, we can define payoff functions for two marginal games and a joint game as,

$$u_i^1(s_1, \ldots, s_{n_1}) = \begin{cases} \frac{v(s_i, E_2)}{|s_i| n_2} & \text{if for any } i' \in s_i, s_{i'} = s_i, \\ \frac{v(\{i\}, E_2)}{n_2} & \text{otherwise,} \end{cases}$$

$$u_j^2(t_1,\ldots,t_{n_2}) = \begin{cases} \frac{v(E_1,t_j)}{n_1|t_j|} & \text{if for any } j' \in t_j,\ t_{j'} = t_j, \\ \frac{v(E_1,\{j\})}{n_1} & \text{otherwise,} \end{cases}$$

and the joint payoff functions are defined as,

$$u_{ij}(s_1,\ldots,s_{n_1};t_1,\ldots,t_{n_2})$$

$$= \begin{cases} \frac{v(s_i,t_j)}{|s_i||t_j|} & \text{if for any } i' \in s_i, j' \in t_j,\ s_{i'} = s_i,\ t_{j'} = t_j, \\ \frac{v(s_i,\{j\})}{|s_i|} & \text{if for any } i' \in s_i,\ s_{i'} = s_i \text{ and there exists } j' \in t_j,\ t_{j'} \neq t_j, \\ \frac{v(\{i\},t_j)}{|t_j|} & \text{if for any } j' \in t_j,\ t_{j'} = t_j \text{ and there exists } i' \in s_i,\ s_{i'} \neq s_i, \\ v(\{i\},\{j\}) & \text{otherwise.} \end{cases}$$

It is easy to see that

$$u_{ij}(s_1,\ldots,s_{n_1};E_2,\ldots,E_2) = u_i^1(s_1,\ldots,s_{n_1})$$

and

$$u_{ij}(E_1,\ldots,E_1;t_1,\ldots,t_{n_2}) = u_j^2(t_1,\ldots,t_{n_2}),$$

i.e., the marginals of the joint payoff functions are exactly two univariate marginal payoff functions.

Definition 3.3 An $n_1 \times n_2$ matrix X is called the **joint imputation** in $(E_1 \times E_2, v)$ if it satisfies

(i) **Individual rationality**: $x_{ij} \geq v(\{i\},\{j\})$ for each $i \in E_1$ and $j \in E_2$ and
(ii) **Group rationality**: $\sum_{i\in E_1}\sum_{j\in E_2} x_{ij} = v(E_1,E_2)$.

In terms of characteristic functions in (0, 1)-reduced form, the joint imputation is an $n_1 \times n_2$ matrix X such that $x_{ij} \geq 0$ and $\sum_{i\in E_1}\sum_{j\in E_2} x_{ij} = 1$, i.e., a probability distribution on $E_1 \times E_2$. It is a $n_1n_2 - 1$ dimensional simplex in $\mathbb{R}^{n_1\times n_2}$.

Definition 3.4 Let X and Y be two joint imputations in $(E_1 \times E_2, v)$, and (A, B) be a coalition. We say that X **dominates** Y through (A, B) if

$$y_{ij} < x_{ij} \quad \text{for all } i \in A, j \in B \quad \text{and} \quad \sum_{i\in A}\sum_{j\in B} x_{ij} \leq v(A,B).$$

This partial order relation is denoted by $Y \prec_{(A,B)} X$. Also we say that X **dominates** Y, denoted by $Y \prec X$, if there is a coalition (A, B) such that $Y \prec_{(A,B)} X$. The set of all undominated imputations of $(E_1 \times E_2, v)$ is called the **core** of v, denoted by $\mathscr{C}(v)$.

Now we obtain one solution for a joint game.

Theorem 3.1 *An imputation X is in $\mathscr{C}(v)$ if and only if*

$$v(A,B) \leq \sum_{i\in A}\sum_{j\in B} x_{ij} \quad \textit{for any coalition } (A,B).$$

Proof For "if part", it suffices to consider the game v in its (0, 1)-reduced form. Let $X \in \mathscr{C}(v)$ and suppose that there is a coalition (A, B) such that

$$v(A,B) > \sum_{i\in A}\sum_{j\in B} x_{ij}. \tag{12}$$

Note that if (12) holds, either A or B must have more than one element, otherwise let $A = \{i\}$ and $B = \{j\}$, $v(\{i\}, \{j\}) > x_{ij}$ which contradicts the definition of imputation (See Definition 3.3). Similarly, (12) can not hold for any $(A, B) \neq (E_1, E_2)$ since

$$\begin{aligned}\sum_{i\notin A}\sum_{j\notin B} x_{ij} + \sum_{i\notin A}\sum_{j\in B} x_{ij} + \sum_{i\in A}\sum_{j\notin B} x_{ij} &= v(E_1,E_2) - \sum_{i\in A}\sum_{j\in B} x_{ij} \\ &\geq v(A,B) - \sum_{i\in A}\sum_{j\in B} x_{ij} > 0.\end{aligned}$$

Let $\varepsilon > 0$ such that

$$0 < \varepsilon < \frac{1}{|A||B|}\left[v(A,B) - \sum_{i\in A}\sum_{j\in B} x_{ij}\right].$$

We can construct an imputation Y by setting

$$y_{ij} = x_{ij} + \varepsilon \quad \text{if } i \in A \text{ and } j \in B$$

and

$$y_{ij} = \frac{\sum_{i\notin A}\sum_{j\notin B} x_{ij} + \sum_{i\notin A}\sum_{j\notin B} x_{ij} + \sum_{i\notin A}\sum_{j\notin B} x_{ij} - |A||B|\varepsilon}{|E_1\backslash A||E_2\backslash B| + |E_1\backslash A||B| + |A||E_2\backslash B|}$$

otherwise. Then Y is an imputation, and moreover, $X \prec_{(A,B)} Y$ which contradicts $v \in \mathscr{C}(v)$.

Now for the "only if part", suppose X satisfies the condition for any coalition (A, B), $v(A, B) \leq \sum_{i\in A}\sum_{j\in B} x_{ij}$. If X is dominated by some Y then there exists some coalition (A, B), such that

$$\sum_{i\in A}\sum_{j\in B} x_{ij} < \sum_{i\in A}\sum_{j\in B} y_{ij} \leq v(A,B),$$

violating the above condition. □

4 The Bivariate Shapley Value

In game theory, the Shapley value is a solution concept in cooperative game theory, it assigns a unique distribution (among the players) of a total surplus generated by the coalition of all players [12]. In this section, we derive the bivariate Shapley value through two different methods and we will show the Shapley value formulas derived from these two methods are equivalent. The Shapley value of the univariate game and its computational method based on random set can be found in Li and Wang [4].

4.1 The Bivariate Shapley Value Through the Cores of the Belief Function H

Let $E_1 = \{1, \ldots, n_1\}$ and $E_2 = \{n_1 + 1, \ldots, n_1 + n_2\}$ be two finite sets and $\mathbb{P}$ be the set of all joint probability measures on $E_1 \times E_2$. Let H be a joint belief function function on $2^{E_1} \times 2^{E_2}$. By Remark 3.1, H can be treat as a characteristic function on $2^{E_1} \times 2^{E_2}$. Therefore, the cores of H is well-defined by Definition 3.4. Furthermore, by Theorem 3.1, the cores of H can be rewrite as

$$core(H) \equiv \{P \in \mathbb{P} | H \preceq P\},$$

where $H \preceq P$ means that $H(A, B) \leq P(A, B)$ for all $A \in 2^{E_1}, B \in 2^{E_2}$.

Note that there are $n_1!$ and $n_2!$ different orders of the elements in E_1 and E_2. Specifically, let Σ_1 and Σ_2 denote the set of all permutations of E_1 and E_2, respectively. For each pair $\sigma_i \in \Sigma_i$, $i = 1, 2$, the elements of E_1 and E_2 are indexed as $\{\sigma_1(1), \sigma_1(2), \ldots, \sigma_1(n_1)\}$ and $\{\sigma_2(n_1 + 1), \sigma_2(n_1 + 2), \ldots, \sigma_2(n_1 + n_2)\}$. We associate a density p_{σ_1,σ_2} on $E_1 \times E_2$ as follows.
(i) For all $i \in E_1, j \in E_2$, define,

$$\begin{aligned} p_{\sigma_1,\sigma_2}(1, n_1 + 1) &= H(\{\sigma_1(1)\}, \{\sigma_2(1), \ldots \sigma_2(n_1 + 1)\}), \\ p_{\sigma_1,\sigma_2}(1, j) &= H(\{\sigma_1(1)\}, \{\sigma_2(1), \ldots \sigma_2(j)\}) \\ &\quad - H(\{\sigma_1(1)\}, \{\sigma_2(1), \ldots \sigma_2(j-1)\}), \\ p_{\sigma_1,\sigma_2}(i, n_1 + 1) &= H(\{\sigma_1(1), \ldots, \sigma_1(i)\}, \{\sigma_2(n_1 + 1)\}) \\ &\quad - H(\{\sigma_1(1), \ldots, \sigma_1(i-1)\}, \{\sigma_2(n_1 + 1)\}). \end{aligned} \tag{13}$$

(ii) For all $i \geq 2$ and $j \geq 2$,

$$\begin{aligned} p_{\sigma_1,\sigma_2}(i, j) = &H(\{\sigma_1(1), \ldots, \sigma_1(i)\}, \{\sigma_2(n_1 + 1), \ldots, \sigma_2(j)\}) \\ &- H(\{\sigma_1(1), \ldots, \sigma_1(i-1)\}, \{\sigma_2(n_1 + 1), \ldots, \sigma_2(j)\}) \\ &- H(\{\sigma_1(1), \ldots, \sigma_1(i)\}, \{\sigma_2(n_1 + 1), \ldots, \sigma_2(j-1)\}) \\ &+ H(\{\sigma_1(1), \ldots, \sigma_1(i-1)\}, \{\sigma_2(n_1 + 1), \ldots, \sigma_2(j-1)\}). \end{aligned} \tag{14}$$

Then the associated probability measure P is defined as

$$P_{\sigma_1,\sigma_2}(A,B) = \sum_{i\in A}\sum_{j\in B} p_{\sigma_1,\sigma_2}(i,j), \tag{15}$$

for $A \in 2^{E_1}$ and $B \in 2^{E_2}$. In the following we will show that this associated probability measure P is in the $core(H)$.

Theorem 4.1 *The probability measure P_{σ_1,σ_2} given in (15) is in the core(H), i.e., $H(A,B) \le P(A,B)$ for all $A \in 2^{E_1}, B \in 2^{E_2}$.*

Proof By the definition of $p_{\sigma_1,\sigma_2}(i,j)$, $1 \le i \le n_1$ and $n_1+1 \le j \le n_2$, we know that

$$p_{\sigma_1,\sigma_2}(i,j) \ge 0 \quad \text{and} \quad \sum_{i=1}^{n_1}\sum_{j=n_1+1}^{n_1+n_2} p_{\sigma_1,\sigma_2}(i,j) = 1.$$

To show $H(A,B) \le P(A,B)$ for all $A \in 2^{E_1}, B \in 2^{E_2}$, we need to show that, for any $2 \le i \le n_1$ and $n_1+2 \le j \le n_1+n_2$,

$$\begin{aligned} &H(\{1,\ldots,i\},\{n_1+1,\ldots,j\}) - H(\{1,\ldots,i-1\},\{n_1+1,\ldots,j-1\}) \\ &\quad = P_{\sigma_1,\sigma_2}(\{1,\ldots,i\},\{n_1+1,\ldots,j\}) \\ &\qquad - P_{\sigma_1,\sigma_2}(\{1,\ldots,i-1\},\{n_1+1,\ldots,j-1\}). \end{aligned} \tag{16}$$

Indeed, from the definition of P_{σ_1,σ_2}, we obtain

$$\begin{aligned} &P_{\sigma_1,\sigma_2}(\{1,\ldots,i\},\{n_1+1,\ldots,j\}) - P_{\sigma_1,\sigma_2}(\{1,\ldots,i-1\},\{n_1+1,\ldots,j-1\}) \\ &= \sum_{t=n_1+1}^{n_1+n_2} p_{\sigma_1\sigma_2}(i,t) + \sum_{s=1}^{n_1-1} p_{\sigma_1,\sigma_2}(s,j), \\ &= H(\{1,\ldots,i\},\{n_1+1\}) - H(\{1,\ldots,i-1\},\{n_1+1\}) \\ &\quad + \sum_{t=n_1+2}^{n_1+n_2} (H(\{1,\ldots,i\},\{n_1+1,\ldots,t\}) \\ &\quad - H(\{1,\ldots,i-1\},\{n_1+1,\ldots,t\}) - H(\{1,\ldots,i\},\{n_1+1,\ldots,t-1\}) \\ &\quad + H(\{1,\ldots,i-1\},\{n_1+1,\ldots,t-1\})) \\ &\quad + H(\{1\},\{n_1+1,\ldots,j\}) - H(\{1\},\{n_1+1,\ldots,j-1\}) \\ &\quad + \sum_{s=2}^{n_1-1} (H(\{1,\ldots,s\},\{n_1+1,\ldots,j\}) \\ &\quad - H(\{1,\ldots,s-1\},\{n_1+1,\ldots,j\}) - H(\{1,\ldots,s\},\{n_1+1,\ldots,j-1\}) \\ &\quad + H(\{1,\ldots,s-1\},\{n_1+1,\ldots,j-1\})) \\ &= H(\{1,\ldots,i\},\{n_1+1,\ldots,j\}) - H(\{1,\ldots,i-1\},\{n_1+1,\ldots,j-1\}). \end{aligned}$$

Now, let $i = \min\{s \in E_1 | s \in A^c\}$ and $j = \min\{t \in E_2 | t \in B^c\}$. Define $A' = \{1, \ldots, i\}$ and $B' = \{n_1 + 1, \ldots, j\}$. Note that $A \cup A' = A \cup \{i\}$, $B \cup B' = B \cup \{j\}$ and $A \cap A' = \{1, \ldots, i-1\}$, $B \cap B' = \{n_1 + 1, \ldots, j-1\}$. From (8) and (16), we have

$$\begin{aligned} H(A \cup \{i\}, B \cup \{j\}) &\geq H(A, B) + H(\{1, \ldots, i\}, \{n_1 + 1, \ldots, j\}) \\ &\quad - H(\{1, \ldots, i-1\}, \{n_1 + 1, \ldots, j-1\}) \\ &= H(A, B) + P_{\sigma_1,\sigma_2}(A \cup \{i\}, B \cup \{j\}) \\ &\quad - P_{\sigma_1,\sigma_2}(\{1, \ldots, i-1\}, \{n_1 + 1, \ldots, j-1\}) \\ &\geq H(A, B) + P_{\sigma_1,\sigma_2}(A \cup \{i\}, B \cup \{j\}) - P_{\sigma_1,\sigma_2}(\{A, B\}). \end{aligned}$$

Therefore, we have

$$H(A, B) - P_{\sigma_1,\sigma_2}(\{A, B\}) \leq H(A \cup \{i\}, B \cup \{j\}) - P_{\sigma_1,\sigma_2}(A \cup \{i\}, B \cup \{j\}).$$

Furthermore, viewing $A \cup \{i\}$, and $B \cup \{j\}$ as another A, and B in the above inequality, and using the same argument recursively, we obtain

$$\begin{aligned} H(A, B) - P_{\sigma_1,\sigma_2}(\{A, B\}) &\leq H(A \cup \{i\}, B \cup \{j\}) - P_{\sigma_1,\sigma_2}(A \cup \{i\}, B \cup \{j\}) \\ &\leq \cdots \leq H(E_1, E_2) - P_{\sigma_1,\sigma_2}(\{E_1, E_2\}) = 1 - 1 = 0, \end{aligned}$$

and hence, P_{σ_1,σ_2} is in the $core(H)$. □

Note that for each pair (σ_1, σ_2), $\sigma_i \in \Sigma_i$, $i = 1, 2$, we obtain an element of $core(H)$, denoted as P_{σ_1,σ_2}. There are $n_1! \times n_2!$ of P_{σ_1,σ_2}. These P_{σ_1,σ_2} are extreme points of $core(H)$ and can be used to define the Shapley value for the joint game. The bivariate Shapley value for (i, j) with $i \in E_1$ and $j \in E_2$ is defined as the arithmetic average of all extreme points of $core(H)$, i.e.,

$$\phi_{ij} = \frac{1}{n_1! n_2!} \sum_{\sigma_1 \in \Sigma_1} \sum_{\sigma_2 \in \Sigma_2} p_{\sigma_1,\sigma_2}(i, j), \tag{17}$$

where $p_{\sigma_1,\sigma_2}(i, j)$ are given in (13) and (14).

Example 4.1 Let $E_1 = \{1, 2\}$ and $E_2 = \{3, 4, 5\}$. Consider the one joint belief function H given in Table 1.

By formula given in (17), we can calculate the bivariate Shapley value listed in Table 2.

Table 1 The joint belief function H

H	{3}	{4}	{5}	{3,4}	{3,5}	{4,5}	{3,4,5}
{1}	5/48	13/144	11/144	7/36	13/72	1/6	1/3
{2}	1/12	1/12	1/12	1/6	1/6	1/6	1/3
{1,2}	1/4	1/4	1/4	1/2	1/2	1/2	1

Table 2 The bivariate Shapley value for H

Φ	{3}	{4}	{5}
{1}	0.1736111	0.1689815	0.1597222
{2}	0.1603009	0.1666667	0.1678241

4.2 The Bivariate Shapley Value Through the Joint Game

Besides the Shapley value calculation given above, we propose a procedure corresponds to each joint game $(E_1 \times E_2, \nu)$ with an imputation matrix $\Phi(\nu) = (\phi_{ij}(\nu))$ whose components describe fair pay-offs' to each of the players in a joint game.

For each $\emptyset \neq A \subseteq E_1$ and $\emptyset \neq B \subseteq E_2$, we define a **simple game** $\omega_{A,B}$, a game taking values in {0, 1}, as follows,

$$\omega_{A,B}(C, D) = \begin{cases} 1 & \text{if } C \subseteq A, D \subseteq B \\ 0 & \text{otherwise.} \end{cases}$$

Note that $\omega_{A,B}$'s given above have the following properties,

Lemma 4.2 *If ν is a $(0, 1)$-reduced game, then there are $(2^{n_1} - 1)(2^{n_2} - 1)$ real numbers $c_{A,B}$, for each $\emptyset \neq A \subseteq E_1$ and $\emptyset \neq B \subseteq E_2$, such that*

$$\nu(C, D) = \sum_{A \subseteq E_1} \sum_{B \subseteq E_2} c_{A,B} \omega_{A,B}(C, D).$$

Proof For each $\emptyset \neq A \subseteq E_1$ and $\emptyset \neq B \subseteq E_2$, let

$$c_{A,B} = \sum_{C \subseteq A} \sum_{D \subseteq B} (-1)^{|A \setminus C| + |B \setminus D|} \nu(C, D),$$

be the 2-dimensional Mobius inverse of $\nu(A, B)$. Then,

$$\nu(A, B) = \sum_{C \subseteq A} \sum_{D \subseteq B} c_{C,D}.$$

Therefore, for any C, D,

$$\sum_{A \subseteq E_1} \sum_{B \subseteq E_2} c_{A,B} \omega_{A,B}(C, D) = \sum_{A \subseteq C} \sum_{B \subseteq D} c_{A,B} = \nu(C, D). \qquad \square$$

Suppose that the player i is in a coalition A of E_1. Then by Superadditivity on the first component we have

$$\nu(A, B) \geq \nu(A \setminus \{i\}, B) + \nu(\{i\}, B),$$

for any coalition B of E_2. If $\nu(A, B) = \nu(A \setminus \{i\}, B) + \nu(\{i\}, B)$, for any $i \in A \subseteq E_1$ and any $B \subset E_2$, then player $i \in E_1$ is unessential to any coalition in E_1.

Definition 4.1 The pair of player (i, j) is said to be **dummy** if
(a) for player i in E_1, $v(A, B) = v(A\backslash\{i\}, B) + v(\{i\}, B)$, for any $i \in A \subseteq E_1$ and any $B \subset E_2$ and
(b) for player j in E_2, $v(A, B) = v(A, B\backslash\{j\}) + v(A, \{j\})$, for any $A \subseteq E_1$ and any $j \in B \subset E_2$.
A **carrier** of v is a set of all non-dummy pairs.

Note that if (i, j) is dummy, then $v(\{i\}, B) = \sum_{j\in B} \phi_{ij}$ and $v(A, \{j\}) = \sum_{i\in A} \phi_{ij}$. Also if (T_1, T_2) is a carrier of v, then $v(T_1, T_2) = \sum_{i\in T_1} \sum_{j\in T_2} \phi_{ij}$.

For the computational aspect of the bivariate Shapley value, we need the following lemmas.

Lemma 4.3 *For $\emptyset \neq A \subset E_1$ and $\emptyset \neq B \subset E_2$, we have for any $c > 0$,*

$$\phi_{ij}(c\omega_{A,B}) = \begin{cases} c/(|A||B|) & \text{if } (i, j) \in (A, B), \\ 0 & \text{otherwise.} \end{cases}$$

Proof Clearly $A \times B$ is a carrier of $c\omega_{A,B}$ and $c\omega_{A,B}(\{i\}, \{j\}) = 0$ for $(i, j) \notin A \times B$. We have

$$\sum_{i\in A} \sum_{j\in B} \phi_{ij} = c\omega_{A,B}(A, B) = c,$$

On the other hand, $(i, j) \notin (A, B)$ is a dummy, thus

$$\phi_{ij} = c\omega_{A,B}(\{i\}, \{j\}) = 0.$$

Also, all ϕ_{ij} of $\Phi(c\omega_{A,B})$ for $(i, j) \in (A, B)$ are equal to each other, and the desired result follows. □

Lemma 4.4 *There is a unique function for the bivariate Shapley value. The bivariate Shapley value formula is given as follow*

$$\phi_{ij}(v) = \sum_{C\ni i} \sum_{D\ni j} \left[\frac{(n_1 - |C|)!(|C| - 1)!}{n_1!} \frac{(n_2 - |D|)!(|D| - 1)!}{n_2!} \right] \times \left[v(C, D) - v(C, D\backslash\{j\}) - v(C\backslash\{i\}, D) + v(C\backslash\{i\}, D\backslash\{j\}) \right]. \tag{18}$$

The proof of Lemma 4.4 is given in Appendix. Let $(E_1 \times E_2, v)$ be a joint game, (E_1, v_1) and (E_2, v_2) be its two marginal games. For the marginal game (E_1, v_1), the Shapley value $\Phi(v_1)$ can be obtained by the formula of univariate case (see Nguyen [7]),

$$\phi_i(v_1) = \sum_{C\ni i} \left[\frac{(n_1 - |C|)!(|C| - 1)!}{n_1!} \right] [v_1(C) - v_1(C\backslash\{i\})] \qquad \text{for } i \in E_1.$$

Similarly, the Shapley value $\Phi(v_2)$ of v_2 can be obtained for the marginal game (E_2, v_1). The following result provides the relation between $\phi_{ij}(v)$ and $\phi_i(v_1), \phi_j(v_2)$.

Table 3 The bivariate Shapley value for game ν

$\Phi(\nu)$	{3}	{4}	{5}
{1}	0.1736111	0.1689815	0.1597222
{2}	0.1603009	0.1666667	0.1678241

Proposition 4.1 *For each* $i \in E_1, j \in E_2$,

$$\sum_{j \in E_2} \phi_{ij}(\nu) = \phi_i(\nu_1) \quad \text{and} \quad \sum_{i \in E_1} \phi_{ij}(\nu) = \phi_j(\nu_2).$$

The proof of Proposition 4.1 is similar to proof of Lemma 4.4.

The following example illustrates the above Proposition.

Example 4.2 (Example 4.1 continued) Consider the joint game given in Table 1. By the second formula of the Shapley value (18), we can calculate the bivariate Shapley value in Table 3,

From the last row and the last column of the joint belief function, we obtain marginal games given in Table 4.

Also the Shapley value for each of marginal games ν_1 and ν_2 listed in Table 5.

Example 4.3 (See Fernandez [3] with some values changed) Consider three cell-phone operators (namely O1, O2, and O3) that want to enter a new market. There are two criteria that must be considered in the process. On the one hand, there is the profit that has been estimated from the market analysis. On the other hand, there is the coverage, which is regulated by law. Thus, the percentage of population covered by each operator or by merging is fixed by the government. Cover- age is very important because it is known to improve the return in the medium and long run. Let us assume that profit is measured in millions of dollars and coverage in percent. We represent by vectors with two entries the values obtained by each operator: the first entry is the profit and the second one is the coverage. Let us consider the following data that represent the values obtained in different cooperation situations:

Let set functions ν_1 and ν_2 be the profit and the coverage of each coalition, respectively. Note that ν_1 and ν_2 are converted to standardized games ν_1' and ν_2' which are belief functions, given in Table 6.

Table 4 Marginal games $\nu_1(A)$ and $\nu_2(B)$

A	{1}	{2}	{1,2}	B	{3}	{4}	{5}	{3,4}	{3,5}	{4,5}	{3,4,5}
ν_1	1/3	1/3	1	ν_2	1/4	1/4	1/4	1/2	1/2	1/2	1

Table 5 Marginal shapley value of games $\nu_1(A)$ and $\nu_2(B)$

$\Phi(\nu_1)$	{1}	{2}	$\Phi(\nu_2)$	{3}	{4}	{5}
	1/2	1/2		1/3	1/3	1/3

Table 6 ν_1' and ν_2'

Coalition	{O1}	{O2}	{O3}	{O1,O2}	{O1,O3}	{O2,O3}	{O1,O2,O3}
ν_1'	1/6	1/4	1/3	1/2	1/2	2/3	1
ν_2'	0.2	0.4	0.1	0.7	0.3	0.5	1

Table 7 The joint game ν of ν_1 and ν_2

ν	{O1}	{O2}	{O3}	{O1,O2}	{O1,O3}	{O2,O3}	{O1,O2,O3}
{O1}	0.052	0.076	0.013	0.139	0.065	0.089	1/6
{O2}	0.064	0.107	0.022	0.191	0.086	0.129	1/4
{O3}	0.059	0.13	0.035	0.225	0.094	0.164	1/3
{O1,O2}	0.126	0.213	0.045	0.38	0.171	0.258	1/2
{O1,O3}	0.111	0.206	0.048	0.363	0.159	0.253	1/2
{O2,O3}	0.131	0.266	0.067	0.464	0.199	0.333	2/3
{O1,O2,O3}	0.2	0.4	0.1	0.7	0.3	0.5	1

Table 8 The vector valued game μ

Coalition S	{O1}	{O2}	{O3}	{O1,O2}	{O1,O3}	{O2,O3}	{O1,O2,O3}
$\mu(S)$	$\begin{pmatrix} 2 \\ 20 \end{pmatrix}$	$\begin{pmatrix} 3 \\ 40 \end{pmatrix}$	$\begin{pmatrix} 3 \\ 10 \end{pmatrix}$	$\begin{pmatrix} 6 \\ 70 \end{pmatrix}$	$\begin{pmatrix} 6 \\ 30 \end{pmatrix}$	$\begin{pmatrix} 8 \\ 50 \end{pmatrix}$	$\begin{pmatrix} 12 \\ 100 \end{pmatrix}$

Table 9 The bivariate Shapley value for game ν

$\Phi(\nu)$	{O1}	{O2}	{O3}	$\Phi(\nu_1)$
{O1}	0.08080939	0.1288041	0.03318769	0.2428012
{O2}	0.11365177	0.1866061	0.05838052	0.3586384
{O2}	0.11930673	0.2087663	0.07489487	0.4029679
$\Phi(\nu_2)$	0.3137679	0.5241765	0.1664631	1

If we use the correlation coefficient of the profit and the coverage $\rho = 0.83$, and adopt Farlie-Gumbel-Morgenstern copula $C_\rho(u, v)$ to construct the joint game (or the joint belief function) ν, then we have the following Table 7.

Note that the last row and the last column of ν can be treated as the standardized the vector-valued game μ in Table 8. Furthermore, we can find the Shapley value for this joint game (Table 9).

Acknowledgments The authors would like to thank Professor Hung T. Nguyen for introducing this interesting topic to us and anonymous referees for their helpful comments which led to the big improvement of this paper.

Appendix

Proof of Lemma 4.4 Note that the left hand side of (18) is equal to

$$\begin{aligned}\phi_{ij}(\nu) &= \sum_{A\subseteq E_1}\sum_{B\subseteq E_2} c_{A,B}\phi_{ij}(\omega_{A,B}) = \sum_{A\ni i}\sum_{B\ni j}\frac{c_{A,B}}{|A||B|} \\ &= \sum_{A\ni i}\sum_{B\ni j}\frac{1}{|A||B|}\sum_{C\subseteq A}\sum_{D\subseteq B}(-1)^{|A\setminus C|+|B\setminus D|}\nu(C,D) \qquad (19)\\ &= \alpha_1-\alpha_2-\alpha_3+\alpha_4,\end{aligned}$$

where

$$\begin{aligned}\alpha_1 &= \sum_{C\ni i}\sum_{D\ni j}\left[\sum_{C\subseteq A}\sum_{D\subseteq B}(-1)^{|A\setminus C|+|B\setminus D|}\frac{1}{|A||B|}\right]\nu(C,D)\\ &= \sum_{C\ni i}\sum_{D\ni j}\left[\sum_{s=|C|}^{n_1}\sum_{t=|D|}^{n_2}(-1)^{(s-|C|)+(t-|D|)}\frac{1}{st}\binom{n_1-|C|}{s-|C|}\binom{n_2-|D|}{t-|D|}\right]\nu(C,D),\end{aligned}$$

$$\begin{aligned}\alpha_2 &= \sum_{C\ni i}\sum_{D\not\ni j}\left[\sum_{C\subseteq A}\sum_{D\cup\{j\}\subseteq B}(-1)^{|A\setminus C|+|B\setminus D|}\frac{1}{|A||B|}\right]\nu(C,D)\\ &= \sum_{C\ni i}\sum_{D\not\ni j}\left[\sum_{s=|C|}^{n_1}\sum_{t=|D|+1}^{n_2}(-1)^{(s-|C|)+(t-|D|)}\frac{1}{st}\binom{n_1-|C|}{s-|C|}\binom{n_2-|D|-1}{t-|D|-1}\right]\\ &\quad \nu(C,D),\end{aligned}$$

$$\begin{aligned}\alpha_3 &= \sum_{C\not\ni i}\sum_{D\ni j}\left[\sum_{C\cup\{i\}\subseteq A}\sum_{D\subseteq B}(-1)^{|A\setminus C|+|B\setminus D|}\frac{1}{|A||B|}\right]\nu(C,D)\\ &= \sum_{C\not\ni i}\sum_{D\ni j}\left[\sum_{s=|C|+1}^{n_1}\sum_{t=|D|}^{n_2}(-1)^{(s-|C|)+(t-|D|)}\frac{1}{st}\binom{n_1-|C|-1}{s-|C|-1}\binom{n_2-|D|}{t-|D|}\right]\\ &\quad \nu(C,D),\end{aligned}$$

and

$$\alpha_4 = \sum_{C\not\ni i}\sum_{D\not\ni j}\left[\sum_{C\cup\{i\}\subseteq A}\sum_{D\cup\{j\}\subseteq B}(-1)^{|A\setminus C|+|B\setminus D|}\frac{1}{|A||B|}\right]\nu(C,D)$$

$$= \sum_{C \not\ni i} \sum_{D \not\ni j} \left[\sum_{s=|C|+1}^{n_1} \sum_{t=|D|+1}^{n_2} (-1)^{(s-|C|)+(t-|D|)} \cdot \frac{1}{st} \binom{n_1 - |C| - 1}{s - |C| - 1} \binom{n_2 - |D| - 1}{t - |D| - 1} \right] \nu(C, D).$$

Now by using the following equality,

$$\sum_{s=c}^{n} \frac{1}{s} (-1)^{s-c} \binom{n-c}{s-c} = \frac{(n-c)!(c-1)!}{n!},$$

α_i's can be reduced so that the $\phi_{ij}(\nu)$ is

$$\begin{aligned}
\phi_{ij}(\nu) = & \sum_{C \ni i} \sum_{D \ni j} \left[\frac{(n_1 - |C|)!(|C| - 1)!}{n_1!} \frac{(n_2 - |D|)!(|D| - 1)!}{n_2!} \right] \nu(C, D) \\
& - \sum_{C \ni i} \sum_{D \not\ni j} \left[\frac{(n_1 - |C|)!(|C| - 1)!}{n_1!} \frac{(n_2 - |D| - 1)!|D|!}{n_2!} \right] \nu(C, D) \\
& - \sum_{C \not\ni i} \sum_{D \ni j} \left[\frac{(n_1 - |C| - 1)!|C|!}{n_1!} \frac{(n_2 - |D|)!(|D| - 1)!}{n_2!} \right] \nu(C, D) \\
& + \sum_{C \not\ni i} \sum_{D \not\ni j} \left[\frac{(n_1 - |C| - 1)!|C|!}{n_1!} \frac{(n_2 - |D| - 1)!|D|!}{n_2!} \right] \nu(C, D).
\end{aligned}$$

Therefore, the bivariate Shapley value formula is given, after simplification, by

$$\begin{aligned}
\phi_{ij}(\nu) = & \sum_{C \ni i} \sum_{D \ni j} \left[\frac{(n_1 - |C|)!(|C| - 1)!}{n_1!} \frac{(n_2 - |D|)!(|D| - 1)!}{n_2!} \right] \\
& \times \left[\nu(C, D) - \nu(C, D \backslash \{j\}) - \nu(C \backslash \{i\}, D) + \nu(C \backslash \{i\}, D \backslash \{j\}) \right]. \quad (20)
\end{aligned}$$

Note that Shavley value given in (20) is equivalent to the one given by the cores of the joint belief function H, in (17). □

References

1. Burger, E.: Introduction to the Theory of Games. Prentice-Hall, Englewood Cliffs (1963)
2. Dempster, A.P.: Upper and lower probabilities induced by a multivalued mapping. Ann. Math. Stat. **28**, 325–339 (1967)
3. Fernandez, F.R., Miguel, A.H., Justo, P.: Core solutions in vector-valued games. J. Optim. Theory Appl. **112**, 331–360 (2002)

4. Li, B., Wang, T.: Computational aspects of the coarsening at random model and the Shapley value. Inf. Sci. **16**, 3260–3270 (2007)
5. Nguyen, H.T.: An Introduction to Random Sets. CRC Press, Boca Raton (2006)
6. Nguyen, H.T.: Lecture notes. In: Statistics with Copulas for Applied Research. Department of Economics, Chiang Mai University, Chiang Mai, Thailand (2013).
7. Nguyen, H.T.: Lecture notes. In: Economic Applications of Game Theory. Department of Economics, Chiang Mai University, Chiang Mai, Thailand (2014).
8. Nguyen, H.T., Wang, T.: Belief Functions and Random Sets. The IMA Volumes in Mathematics and Its Applications, pp. 243–255. Springer, New York (1997).
9. Roemer, J.: Games with vector-valued payoffs and their application to competition between organizations. Econ. Bull. **3**, 1–13 (2005)
10. Schmelzer, B.: Characterizing joint distributions of random sets by multivariate capacities. Int. J. Approx. Reason. **53**, 1228–1247 (2012)
11. Schmelzer, B.: Joint distributions of random sets and their relation to copulas. In: Modeling Dependence in Econometrics. Advances in Intelligent Systems and Computing, vol. 251 (2014) doi:10.1007/978-3-319-03395-2-10.
12. Shapley, L.S.: A value for n-person games. In: Kuhn, H.W., Tucker, A.W. (eds.) Contributions to the Theory of Games. Annals of Mathematical Studies, vol. 28, pp. 307–317. Princeton University Press, Princeton (1953)
13. Shafer, G.: A Mathematical Theory of Evidence. Princeton University Press, New Jersey (1976)

Distortion Risk Measures Under Skew Normal Settings

Weizhong Tian, Tonghui Wang, Liangjian Hu and Hien D. Tran

Abstract Coherent distortion risk measure is needed in the actuarial and financial fields in order to provide incentive for active risk management. The purpose of this study is to propose extended versions of Wang transform using skew normal distribution functions. The main results show that the extended version of skew normal distortion risk measure is coherent and its transform satisfies the classic capital asset pricing model. Properties of the stock price model under log-skewnormal and its transform are also studied. A simulation based on the skew normal transforms is given for a insurance payoff function.

1 Introduction

Risk measures are used to decide insurance premiums and required capital for a given risk portfolio by examining its downside risk potential. A widely used risk measure for the risk of loss on a specific portfolio of financial assets is the value at risk (VaR). Mathematically, the VaR is simply a percentile of the distribution of losses.

Unfortunately, because the VaR fails to satisfy the sub-additivity property and ignores the potential loss beyond the confidence level, distortion risk measures given in Wang [22] have overcome these drawbacks.

W. Tian · T. Wang (✉)
Department Mathematical Sciences, New Mexico State University, Las Cruces, USA
e-mail: twang@nmsu.edu

W. Tian
e-mail: xjlaojiu@nmsu.edu

T. Wang
Innovation Experimental College, Northwest A and F University, Xianyang, China

L. Hu
College of Science, Donghua University, Shanghai, China
e-mail: Ljhu@dhu.edu.cn

H.D. Tran
Tan Tao University, Long An, Vietnam
e-mail: hien.tran@ttu.edu.vn

V.-N. Huynh et al. (eds.), *Econometrics of Risk*, Studies in Computational Intelligence 583, DOI 10.1007/978-3-319-13449-9_9

Distortion risk measures were originally applied to a wide variety of insurance problems such as the determination of insurance premiums, capital requirements, and capital allocations. Because insurance and investment risks are closely related, the investment community started to apply distortion risk measures in context of the asset allocation problem. Wang [25] also applied the distortion risk measure to price catastrophe bonds, while Fabozzi and Steel [6] used it to price real estate derivatives.

The properties of a coherent risk measure were given by Azzalini [3]. In order to construct a coherent risk measure, Li et al. [16] proposed two extended versions of Type I Wang transform Wang [22] using two versions of skew normal distributions. In this paper, a new skew normal risk measure and its properties are investigated. It has been shown that our new skew normal risk measure satisfied the capital asset pricing model classic capital asset model (CAPM).

This paper is organized as follows. Distortion risk measures and their properties are discussed in Sect. 2. A new skew normal distortion risk measure based on extended Wang transform is introduced and its properties are studied in Sect. 3. The CAPM is introduced and the new distortion transform method satisfied the CAPM is obtained in Sect. 4. The behavior of stock price model under log-skew normal setting is studied in Sect. 5 and a simulation based on the skew normal transform for a insurance pay-off function is obtained in Sect. 6.

2 Distortion Risk Measures

Let X be a non-negative loss random variable and $F_X(x)$ be its distribution function, where $F_X(x) = P(X \leq x)$ is the probability that $X \leq x$. The survival function $S_X(x) = 1 - F_X(x)$, has a special role in calculating insurance premiums based on the fact that the expect value of X is given by

$$E(X) = \int_0^\infty S_X(y)dy. \tag{1}$$

An insurance layer $X(a, a+m]$, as a payoff function, is defined by

$$X(a, a+m] = \begin{cases} 0 & \text{for } 0 \leq X < a \\ X - a & \text{for } a \leq X < a+m \\ m & \text{for } a+m \leq X, \end{cases} \tag{2}$$

where a is called the attachment point, a point at which excess insurance or reinsurance limit apply, and m is call the limit point, an amount that starts from attachment point. The survival function for the layer $X(a, a+m]$ is related to that of the underlying risk X by

$$S_{X(a,a+m]}(y) = \begin{cases} S_X(a+y) & \text{for } 0 \leq y < m \\ 0 & \text{for } m \leq y. \end{cases} \tag{3}$$

The expected loss for the layer $X(a, a+m]$ can be calculated by

$$E\left(X(a, a+m]\right) = \int_0^{\infty} S_{X(a,a+m]}(y)dy = \int_a^{a+m} S_X(x)dx. \tag{4}$$

For a very small layer $X(a, a+\varepsilon]$, the net premium (expected loss) is $\varepsilon S_X(a)$. This is the reason why S_X is said to be the density of layer net premium. In relation to the expected layer loss cost, Lee [14] provided a detailed account of S_X. Venter [21] showed that, for any given risk, market prices by layers always imply transformed distributions and Wang [22] suggested calculating premium by directly transforming the survival function,

$$H_g(X) = \int_0^{\infty} g\left(S_X(x)\right) dx, \tag{5}$$

where $g(\cdot)$ is the distortion function given below.

Definition 2.1 The increasing and continuous function $g : [0, 1] \to [0, 1]$ with $g(0) = 0$ and $g(1) = 1$ is called a **distortion function**.

Remark 2.1 The distortion function transforms a probability distribution S_X to a new distribution $g(S_X)$. The mean value under the distorted distribution, $H_g(X)$, is given by

$$H_g\left(X(a, a+m]\right) = \int_0^{\infty} g\left(S_{X(a,a+m]}(y)\right) dy = \int_a^{a+m} g(S_X)(x)dx. \tag{6}$$

Lemma 2.1 (Artzner et al. [2]). *The distortion function for pricing insurance layers should meet the following properties:*

(i) $0 \le g(u) \le 1$, $g(0) = 0$, *and* $g(1) = 1$;
(ii) $g(u)$ *is an increasing function* (*where it exists,* $g'(u) \ge 0$);
(iii) $g(u)$ *is concave* (*where it exists,* $g''(u) \le 0$); *and*
(iv) $g'(0) = +\infty$.

Note that these conditions can be explained as follows. (i) For each $x \in [0, 1]$, $g\left(S_X(x)\right)$ defines a valid probability and zero probability events will still have zero probability after applying the distortion operator $g(\cdot)$. (ii) The distorted probability $g(S_X(x))$ defines another distribution and the risk adjusted layer premium decreases as the layer increases for fixed limit. (iii) The risk load is non-negative for every risk or layer and the relative risk loading increases as the attachment point increases for a fixed limit. (iv) Unbounded relative loading at high reinsurance layers seems to be supported by observed market reinsurance premiums (see, e.g., Venter [21] and Artzner [2]).

3 A New Distortion Function Based on the Wang Transform

In many real world applications, normal setting may not be a good fit because data sets collected are not symmetrically distributed. Since data sets are skewed in most applications, the class of skew normal distributions may be the better choice.

Univariate skew normal models have been considered by many authors, see e.g., Azzalini [3], Gupta et al. [7], and Wang et al. [26]. In the last three decades there has been substantial work in the areas of skew normal (SN) and its related distributions. The main feature of this class is that a new skewness parameter α is introduced to control skewness and kurtosis.

Definition 3.1 A random variable X is said to have a **skew normal distribution** with location parameter μ, scale parameter σ^2, and skewness parameter α, denoted by $X \sim SN(\mu, \sigma^2, \alpha)$, if its density function is given by

$$f(x|\mu, \sigma, \alpha) = 2\phi(x|\mu, \sigma^2)\Phi\left[\alpha\left(\frac{x-\mu}{\sigma}\right)\right], \tag{7}$$

where $\alpha \in \Re$, $\phi(\cdot; \mu, \sigma^2)$ and $\Phi(\cdot)$ are the probability density function of $N(\mu, \sigma^2)$ and the distribution function of $N(0, 1)$, respectively.

Note that when $\alpha > 0$, the distribution is skewed to the right and when $\alpha < 0$, the distribution is skewed to the left. Also when $\alpha = 0$, the distribution is reduced to $N(\mu, \sigma^2)$. The basic properties of $X \sim SN(\mu, \sigma^2, \alpha)$ are listed as follows.

Lemma 3.1 (Azzalini [3]). *The mean value, variance, and skewness of random variable* $X \sim SN(\mu, \sigma^2, \alpha)$ *are*

$$E(X) = \mu + \sigma\delta\sqrt{\frac{2}{\pi}}, \quad Var(X) = \sigma^2\left(1 - \frac{2\delta^2}{\pi}\right),$$
$$Sk(X) = \frac{4-\pi}{2}\frac{(\delta\sqrt{2/\pi})^3}{(1-2\delta^2/\pi)^{3/2}},$$

where $\delta = \alpha/\sqrt{1+\alpha^2}$, *and* $Sk(\cdot)$ *represents the skewness for the distribution.*

Note that the skewness parameter α appears in the three equations so that it has effects on all three moments. From Lemma 3.1, we can conclude that the mean value of the family is affected by both skew and scale parameters, the variance is also depend on the skewness parameter, and the skewness of distribution is the only function of the skewness parameter and is free of μ and σ^2.

Let

$$SN\Phi(x|\mu, \sigma^2, \alpha) = \int_{-\infty}^{x} f(t|\mu, \sigma^2, \alpha)dt \tag{8}$$

be the distribution function of $X \sim SN(\mu, \sigma^2, \alpha)$. The special case of it was given by Wang [23] as

$$g_\lambda(x) = \Phi(\Phi^{-1}(x) + \lambda). \tag{9}$$

Note that, the $g_\lambda(x)$ given in (9), called Type I Wang transform, is a pricing formula that recovers the CAPM and Black-Scholes formula under normal asset-return distributions, see Wang [24] and Kijima [13] for details. Corresponding to Type I Wang transform and its extended version given in Li et al. [16], we propose a new extended distortion function of Type I Wang transform defined as follows.

Definition 3.2 For $x > 0$, the skew normal distributed distortion function is defined by

$$g_N(x) = SN\Phi(bSN\Phi^{-1}(x) + \lambda|\alpha), \tag{10}$$

where $SN\Phi(\cdot|\alpha)$ is the distribution function of $SN(0, 1, \alpha)$, $SN\Phi^{-1}(\cdot|\alpha)$ is the inverse function of $SN\Phi(\cdot|\alpha)$, and $\lambda \in \Re$.

For the illustration of $g_N(x)$'s in (10), see Figs. 1 and 2. From Fig. 1, the values of market price risk parameter λ effect the curves of distortion functions. From the Fig. 2, the distortion function curves are effected by the values of the skewness parameter α. Note that, the distortion function of type I Wang transform is a special case of $g_N(x)$ in (10), where $b = 1$ and $\alpha = 0$.

The following result shows that the extended versions of Type I Wang transform given in (10) is coherent.

Theorem 3.1 *The distortion function $g_N(x)$ given in* (10) *satisfies the properties given in Lemma* 2.1.

Proof For (i), we have

$$g_N(0) = \lim_{x\to 0} SN\Phi(bSN\Phi^{-1}(x)|\alpha) = SN\Phi(-\infty|\alpha) = 0,$$

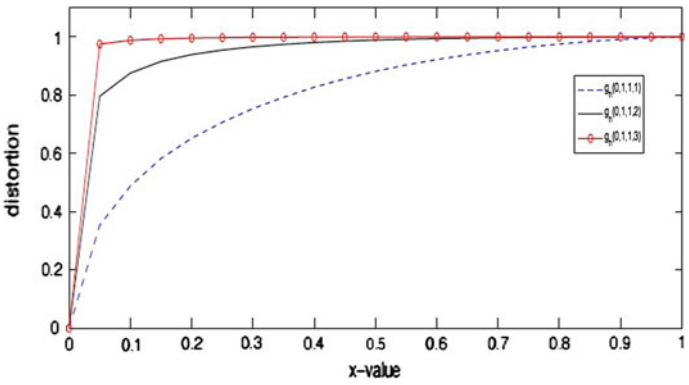


Fig. 1 The curves of $g_N(x)$ for $\alpha = 1$ and $b = 1$ and different values of λ

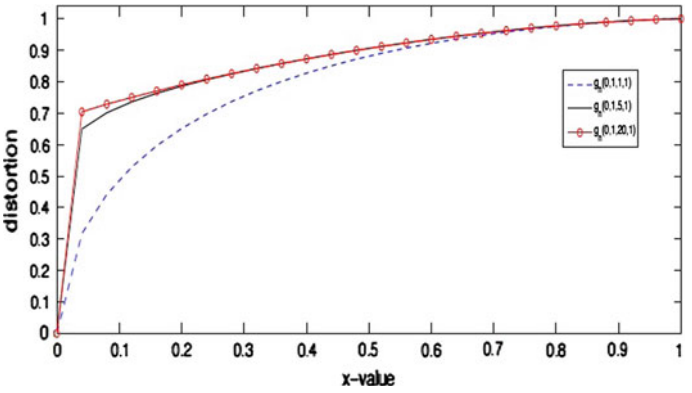


Fig. 2 The curves of $g_N(x)$ for $\lambda = 1$ and $b = 1$ and different values of α

and

$$g_N(1) = \lim_{x \to 1} SN\Phi(bSN\Phi^{-1}(x)|\alpha) = SN\Phi(\infty|\alpha) = 1.$$

Also it is easy to see that $g_N(x)$ is bounded in [0, 1].
For (ii), note that the first derivative of g_N is

$$g'_N(x) = b\frac{\phi(bw+\lambda)\Phi(\alpha(bw+\lambda))}{2\phi(w)\Phi(\alpha w)},$$

where $w = SN\Phi^{-1}(x)$. Clearly, $g'_N(x) > 0$ if $b > 0$.
For (iii), indeed, if we let $f(w) = \phi(w)\Phi(\alpha w)$, then by (ii), we have

$$g'_N(x) = \frac{f(bw+\lambda)}{2f(w)}.$$

Now

$$f'(w) = \frac{\alpha\phi(w)}{2\Phi(\alpha w)} - \frac{w}{2}$$

and

$$f'(bw+\lambda) = \frac{b\alpha\phi\left(\alpha(bw+\lambda)\right)\phi(bw+\lambda)}{2\Phi(\alpha w)\phi(w)} - (bw+\lambda)b\frac{\phi(\alpha(bw+\lambda))\Phi(bw+\lambda)}{2\Phi(\alpha w)\phi(w)}.$$

Without loss of the generality, let $b = 1$ and the second derivative of $g_N(x)$ can be simplified as

$$g''_N(x) = \frac{f'(w+\lambda)\cdot f(w) - f(w+\lambda)\cdot f'(w)}{2f^3(w)}.$$

Note that the numerate of $g_N''(x)$ is

$$\begin{aligned} & f'(w+\lambda)\cdot f(w) - f(w+\lambda)\cdot f'(w) \\ &= \frac{\alpha\phi(w+\lambda)[\phi(\alpha(w+\lambda))\Phi(\alpha w) - \phi(\alpha w)\Phi(\alpha(w+\lambda))]}{2\Phi(\alpha w)} \\ &\quad - \lambda\phi(w+\lambda)\Phi(\alpha(w+\lambda)). \end{aligned} \tag{11}$$

Let $F(w) = \phi(\alpha w)/\Phi(\alpha w)$. It is easy to show $F'(w) < 0$ for $\alpha > 0$ and $w > 0$. Thus (11) is negative with an additional condition $\lambda < 0$.
For (iv),

$$\lim_{x\to 0} g_N'(x) = \lim_{w\to-\infty} \frac{\phi(w+\lambda)\Phi(\alpha(w+\lambda))}{2\phi(w)\Phi(\alpha w)}. \tag{12}$$

Note that

$$\lim_{w\to-\infty} \frac{\phi(w+\lambda)}{2\phi(w)} = \lim_{w\to-\infty} \exp\{-\frac{\lambda^2}{2} - w\lambda\} = +\infty$$

and

$$\lim_{w\to-\infty} \frac{\Phi(\alpha(w+\lambda))}{2\Phi(\alpha w)} \lim_{w\to-\infty} \frac{\phi\alpha(w+\lambda)}{2\phi(\alpha w)} = \lim_{w\to-\infty} \exp\{-\frac{(\lambda\alpha)^2}{2} - w\lambda\alpha^2\} = +\infty.$$

Therefore, (12) tends to ∞ as $x \to 0$ and the desired results follows. □

4 The Capital Asset Pricing Model

The CAPM is a set of predictions concerning equilibrium expected returns on assets. The classic CAPM assumes that all investors have the same one-period horizon, and asset returns have multivariate normal distributions. For a fixed time horizon, let R_i and R_M be the rate-of-return for the asset i and the market portfolio M, with variances σ_i^2 and σ_M^2, respectively. The classic CAPM asserts that

$$E(R_i) = r + \beta_i[E(R_M) - r], \tag{13}$$

where r is the risk-free rate-of-return $\beta_i = \mathrm{Cov}(R_i, R_M)/\sigma_M^2$, and $\mathrm{Cov}(X, Y)$ is the covariance of X and Y. Assume that the asset returns are normally distributed and the time horizon is one period (e.g., one month or one year), one of the key concepts in financial economics is the market price of risk given by

$$\lambda_i = \frac{E(R_i) - r}{\sigma_i}. \tag{14}$$

In asset portfolio management, this is also called the Sharpe Ratio, after William Sharpe [20]. In finance, the Sharpe ratio is a way to examine the performance of

an investment by adjusting its risk. The ratio measures the excess return (or risk premium) per unit of deviation in an investment asset or a trading strategy.

The CAPM provides powerful insight regarding the risk-return relationship, where only the systematic risk deserves an extra risk premium in an efficient market. However, the CAPM and the concept of market price of risk were developed under the assumption of multivariate normal distributions for the asset returns. The CAPM has serious limitations when applied to insurance pricing under loss distributions are not normally distributed.

By (9), we propose a new transform given by

$$F_*(x) \equiv g(F(x)) = SN\Phi[bSN\Phi^{-1}(F(x)) + \lambda], \tag{15}$$

where where $F(x)$ is the distribution function of X. When $b = 1$ and $\alpha = 0$, the above transform is reduced to the universal pricing transform given in Wang [23].

Our main result is given as follows.

Theorem 4.1 *Let F be the distribution function of $X \sim SN(\mu, \sigma^2, \alpha)$. Then $F_*(X) \sim SN(\mu_*, \sigma_*^2, \alpha_*)$, where*

$$\mu_* = \mu - \frac{\lambda\sigma}{b}, \qquad \sigma_* = \frac{\sigma}{b}, \qquad \text{and} \qquad \alpha_* = \alpha.$$

Proof Since $X \sim SN(\mu, \sigma^2, \alpha)$,

$$F(x) = \int_{-\infty}^{x} 2\phi\left(\frac{t-\mu}{\sigma}\right)\Phi\left(\alpha\frac{t-\mu}{\sigma}\right)dt = \int_{-\infty}^{\frac{x-\mu}{\sigma}} 2\phi(u)\Phi(\alpha u)du = SN\Phi\left(\frac{x-\mu}{\sigma}\right).$$

Thus, $(X - \mu/b = SN\Phi^{-1}(F(X))$ and $bSN\Phi^{-1}(F(X)) + \lambda = b(X-\mu)/\sigma + \lambda$. From (15), we have

$$\begin{aligned} F_*(x) &= SN\Phi\left(b\left(\frac{x-\mu}{\sigma}\right) + \lambda\right) \\ &= \int_{-\infty}^{b\left(\frac{x-\mu}{\sigma}\right)+\lambda} 2\phi(t)\Phi(\alpha t)dt \\ &= \int_{-\infty}^{\sigma_*^{-1}(x-\mu_*)} 2\phi(t)\Phi(\alpha t)dt \\ &= \int_{-\infty}^{x} 2\phi\left(\frac{t-\mu_*}{\sigma_*}\right)\Phi\left(\alpha_*\frac{t-\mu_*}{\sigma_*}\right)dt, \end{aligned}$$

which is the distribution function of $X_* \sim SN(\mu_*, \sigma_*^2, \alpha_*)$ and the desired result follows. □

From the skew normal transform $F_*(X)$ we have the following result.

Proposition 4.1 *With λ being the market price of risk for an asset, the skew normal transform $F_*(X)$ given in (15) replicates the classic CAPM.*

Proof By Lemma (3.1), we obtain

$$E(X_*) = (\mu - \frac{\lambda\sigma}{b}) + \frac{\sigma\delta}{b}\sqrt{\frac{2}{\pi}},$$

where $\delta = \alpha/\sqrt{1+\alpha^2}$. Let $r = E(X_*)$, then

$$E(X) - r = \sigma\delta\sqrt{\frac{2}{\pi}}(1 - \frac{1}{b}) + \frac{\lambda\sigma}{b}.$$

Sovling for λ

$$\lambda = \frac{E(X) - r_*}{\sigma_*}, \tag{16}$$

where $\sigma_* = \dfrac{\sigma}{b}$, and $r_* = r + \sigma\delta\sqrt{\dfrac{2}{\pi}}\left(1 - \dfrac{1}{b}\right)$, which is the revised risk-free rate to different the risk-free rate r. □

Note that our formula given in (16) is valid for any $b > 0$. If $b = 1$, λ in (16) is reduced to the one given in (14).

5 The Model for the Behavior of Stock Prices

In practice, we do not observe stock prices following the continuous variables and continuous-time processes. Stock prices are restricted to discrete values and changes can be observed only when the exchange is open. But, any variable whose value changes over time in an uncertain way is said to follow a stochastic process. A *Markov process* is a particular type of stochastic process where only the present value of a variable is relevant for predicting the future. The *Wiener process* is a particular type of Markov stochastic process with the mean change zero and the variance rate of 1. A *generalized Wiener process* for a variable x can be defined in terms of dz as follows:

$$dx = adt + bdz, \tag{17}$$

where a and b are constants. In general, for a and b are functions of the value of the underlying variable x and time t, the generalized Wiener process will be known as *Itô process*, such as

$$dx = a(x,t)dt + b(x,t)dz. \tag{18}$$

Lemma 5.1 (Hull [8]). *Suppose that the value of a variable x follows the Itô process, then a function G of x and t follows an Itô process*

$$dG = \left(\frac{\partial G}{\partial x}a + \frac{\partial G}{\partial t} + \frac{1}{2}\frac{\partial^2 G}{\partial x^2}b^2\right)dt + \frac{\partial G}{\partial x}bdz, \tag{19}$$

where a and b are functions of x and time t.

If we assume stock prices are modeled by log-skewnormal distributions, then stock returns should be modeled by skew normal distributions. Therefore, the equivalent results can be obtained by applying Wang transform either to the stock price distribution, or, to the stock return distribution. Now we introduce the log-skewnormal distribution.

Definition 5.1 The positive random variable X in the $\Re^+$ has a univariate log-skewnormal distribution if the transformed variable $Y = log(X) \sim SN(\mu, \sigma, \alpha)$, denoted by $X \sim LSN(\mu, \sigma, \alpha)$. The probability density function of X is given by

$$f(x; \mu, \sigma, \alpha) = \frac{2}{x}\phi(\log x; \mu, \sigma)\Phi\left(\frac{\alpha(\log x - \mu)}{\sigma}\right). \tag{20}$$

Note that it is easy to show that if $X \sim LSN(0, 1, \alpha)$, then the moment generating function of X is

$$M_X(t) \equiv E(e^{tX}) = 2e^{t^2/2}\Phi(\delta t), \qquad \delta = \frac{\alpha}{\sqrt{1+\alpha^2}}.$$

Thus we have the following result.

Theorem 5.1 *Assume the asset price $X_i(t)$ for an individual stock satisfies the stochastic differential equation*

$$\frac{dX_i(t)}{X_i(t)} = \mu_i dt + \sigma_i dW_i, \tag{21}$$

where $dW_i \sim SN(0, dt, \alpha)$. Then for any future time T

$$\frac{X_i(T)}{X_i(0)} \sim LSN\left(\mu_i T - \sigma_i^2 T/2,\ \sigma_i^2 T,\ \alpha\right), \tag{22}$$

where $X_i(0)$ is the current asset price.

Proof Clearly, from (21) we have,

$$dX_i(t) = \mu_i X_i(t)dt + \sigma_i X_i(t)dW_i.$$

Let $G = \log X_i(t)$, then from Lemma 5.1, we obtain

$$dG = \left(\frac{\partial G}{\partial X_i(t)}\mu X_i(t) + \frac{\partial G}{\partial t} + \frac{1}{2}\frac{\partial^2 G}{\partial X_i(t)^2}\sigma_i^2 X_i(t)^2\right)dt + \frac{\partial G}{\partial X_i(t)}\sigma_i X_i(t)dW_i. \quad (23)$$

Note that

$$\frac{\partial G}{\partial X_i(t)} = \frac{1}{X_i(t)}, \quad \frac{\partial G}{\partial t} = 0, \quad \text{and} \quad \frac{\partial^2 G}{\partial X_i(t)^2} = -\frac{1}{X_i(t)^2}.$$

Therefore (23) is reduced to

$$dG = (\mu_i - \frac{\sigma_i^2}{2})dt + \sigma_i dW_i. \quad (24)$$

Since $dW_i \sim SN(0, dt, \alpha)$, then the change in $\log X_i(t)$ between 0 and future time T is skew normally distributed, i.e.,

$$\log X_i(T) - \log X_i(0) \sim SN\left(\mu_i T - \frac{1}{2}\sigma_i^2 T,\ \sigma_i^2 T,\ \alpha\right),$$

which is equivalent to

$$\frac{X_i(T)}{X_i(0)} \sim LSN\left(\mu_i T - \frac{1}{2}\sigma_i^2 T,\ \sigma_i^2 T,\ \alpha\right). \quad \square$$

The following result gives the relationship between the transform F_* and the F of X and its proof is similar to that of Theorem 4.1.

Theorem 5.2 *If $F(x)$ be the distribution function of $X \sim LSN(\mu, \sigma^2, \alpha)$, then $F_*(x)$ is the distribution function of $X_* \sim LSN(\mu_*, (\sigma_*)^2, \alpha_*)$, where parameters*

$$\mu_* = \mu - \frac{\lambda\sigma}{b}, \qquad \sigma_* = \frac{\sigma}{b}, \qquad \textit{and} \quad \alpha_* = \alpha.$$

Remark 5.1 From Theorem 5.2, if we let $b = 1$, then

$$\frac{X_i^*(T)}{X_i(0)} \sim LSN\left(\mu_i T - \lambda\sigma_i\sqrt{T} - \frac{1}{2}\sigma_i^2 T,\ \sigma_i^2 T,\ \alpha\right).$$

Now, define an implicated parameter value

$$\lambda_i(T) = \frac{(\mu_i - r)}{\sigma_i}\sqrt{T} = \lambda_i\sqrt{T},$$

where r is the same risk-free rate as the one defined previously. And the λ_i above coincides with the market price if the asset i is defined in Hull [8]. Note that λ_i is also consistent with continuous-time CAPM (Merton [18]).

6 Simulation Results

Consider a ground-up liability risk X with a Pareto severity distribution

$$S_X(x) = \left(\frac{2{,}000}{2{,}000 + x}\right)^{1.2}, \quad \text{for } x > 0.$$

To compare the risk loading by the layer, assume that the ground-up frequency is exactly one claim. We apply the pricing formula (6) to the severity distribution. For the numerical illustration, we choose a loading parameters $\alpha = 1$, $\lambda = 0.1$, and $b = 1$ for the skew normal distortion Transform in (15), and $\mu = 0, \sigma = 1$ in both transform methods. If the loss is capped by a basic limit of \$50,000, the expected loss is \$4,788 and the risk-adjusted premium is \$5,600, implies a 16.96 % loading increases. As shown in Table 1, the relative loading increases at higher layers.

Two comparisons can be made with Wang transform and proportional hazards (PH) transform given in Wang [23]. A loading parameter $\lambda = 0.1126$ is selected for Wang transform and index $r = 0.913$ is selected for PH transform to yield the same relative loading (16.96 %) for the basic limit layer (\$0, 50,000).

Table 1 shows that the PH transform method produces a risk loading that increases much faster than when using Wang transform distortion function and the new skew normal distortion function. And also the new distortion function produced a slow increase risk loading as the layers increase. As the layers increase to infinity, both the new skew normal distortion function and Wang distortion function should have a same pattern.

Table 1 Risk load by layer under new skew-normal distortion, Wang distortion and PH-transform

Layer's in (1,000)	Expected loss	PH premium	Relative loading (%)	Wang premium	Relative loading (%)	New premium	Relative loading (%)
(0, 50]	4,788	5,600	17.0	5,600	17.0	5,600	17.0
(50, 100]	657	956	45.5	873	32.9	816	28.8
(100, 200]	582	908	56.0	797	36.9	764	31.3
(200, 300]	307	508	65.5	430	40.0	412	34.0
(300, 400]	203	349	71.9	290	42.9	279	37.4
(400, 500]	150	265	76.7	216	44.0	209	39.3
(500, 1,000]	428	785	84.4	624	45.8	614	41.1
(1,000, 2,000]	373	739	98.1	558	49.6	550	47.2
(2,000, 5,000]	420	906	115.7	646	53.8	641	51.7

7 Conclusion

Like the distortion function of Wang [22], the new skew normal distortion function still satisfied the classic CAPM. Furthermore, after introducing the behavior of stock price model under log-skewnormal distribution, our skew normal distortion functions are still consistent with the continuous-time CAPM. It promotes a unified approach to pricing financial and insurance risks. With great promise in theoretical development and practical application, more research is needed to further explore the properties of this pricing formula.

Acknowledgments The authors would like to thank Ying Wang for proofreading of this paper and anonymous referees for their valuable comments which let the improvement of this paper.

References

1. Artzner, P., Delbaen, F., Eber, J.M., Heath, D.: Thinking coherently: generalised scenarios rather than var should be used when calculating regulatory capital. Risk-Lond.-Risk Mag. Ltd. **10**, 68–71 (1997)
2. Artzner, P., Delbaen, F., Eber, J.M., Heath, D.: Coherent measures of risk. Math. Financ. **9**(3), 203–228 (1999)
3. Azzalini, A.: A class of distributions which includes the normal ones. Scand. J. Stat. **12**, 171–178 (1985)
4. Basak, S., Shapiro, A.: Value-at-risk-based risk management: optimal policies and asset prices. Rev. Financ. stud. **14**(2), 371–405 (2001)
5. Butsic, R.P.: Capital Allocation for property-liability insurers: a catastrophe reinsurance application. In: Casualty Actuarial Society Forum, pp. 1–70 (1999)
6. Fernández, C., Steel, M.F.: On Bayesian modeling of fat tails and skewness. J. Am. Stat. Assoc. **93**(441), 359–371 (1998)
7. Gupta, A.K., Chang, F.C., Huang, W.J.: Some skew-symmetric models. Random Oper. Stoc. Equ. **10**, 133–140 (2002)
8. Hull, J.C.: Options, Futures and Other Derivatives. Pearson Education, New York (1999)
9. Hürlimann, W.: On stop-loss order and the distortion pricing principle. Astin Bull. **28**, 119–134 (1998)
10. Hürlimann, W.: Distortion risk measures and economic capital. N. Am. Actuar. J. **8**(1), 86–95 (2004)
11. Hürlimann, W.: Conditional value-at-risk bounds for compound Poisson risks and a normal approximation. J. Appl. Math. **2003**(3), 141–153 (2003)
12. Jones, M.C., Faddy, M.J.: A skew extension of the t-distribution, with applications. J. R. Stat. Soc.: Ser. B (Stat. Methodol.) **65**(1), 159–174 (2003)
13. Kijima, M.: A multivariate extension of equilibrium pricing transforms: the multivariate Esscher and Wang transforms for pricing financial and insurance risks. Astin Bull. **36**(1), 269 (2006)
14. Lee, Y.S.: The mathematics of excess of loss coverages and retrospective rating-a graphical approach. PCAS LXXV **49** (1988)
15. Lin, G.D., Stoyanov, J.: The logarithmic skew-normal distributions are moment-indeterminate. J. Appl. Probab. **46**(3), 909–916 (2009)
16. Li, B., Wang, T., Tian, W.: Risk measures and asset pricing models with new versions of Wang transform. In: Uncertainty Analysis in Econometrics with Applications, pp. 155–167. Springer, Berlin (2013)

17. Ma, Y., Genton, M.G.: Flexible class of skew-symmetric distributions. Scand. J. Stat. **31**(3), 459–468 (2004)
18. Merton, R.C.: An intertemporal capital asset pricing model. Econom.: J. Econom. Soc. **41**, 867–887 (1973)
19. Nekoukhou, V., Alamatsaz, M.H., Aghajani, A.H.: A flexible skew-generalized normal distribution. Commun. Stat.-Theory Methods **42**(13), 2324–2334 (2013)
20. Sharpe, W.F.: The sharpe ratio. J. Portf. Manag. **21**(1), 4958 (1994)
21. Venter, G.G.: Premium calculation implications of reinsurance without arbitrage. Astin Bull. **21**(2), 223–230 (1991)
22. Wang, S.S.: Premium calculation by transforming the layer premium density. Astin Bull. **26**, 71–92 (1996)
23. Wang, S.S.: A class of distortion operators for pricing financial and insurance risks. J. Risk Insur. **67**, 15–36 (2000)
24. Wang, S.S.: A universal framework for pricing financial and insurance risks. Astin Bull. **32**(2), 213–234 (2002)
25. Wang, S.S.: Cat bond pricing using probability transforms. Geneva Papers: Etudes et Dossiers, special issue on Insurance and the State of the Art in Cat Bond Pricing 278, 19–29 (2004)
26. Wang, T., Li, B., Gupta, A.K.: Distribution of quadratic forms under skew normal settings. J. Multivar. Anal. **100**(3), 533–545 (2009)
27. Wirch, J.L., Hardy, M.R.: A synthesis of risk measures for capital adequacy. Insur.: Math. Econ. **25**(3), 337–347 (1999)
28. Wirch, J.L., Hardy, M.R.: Distortion risk measures. Coherence and stochastic dominance. In: International Congress on Insurance: Mathematics and Economics, pp. 15–17 (2001)

Towards Generalizing Bayesian Statistics: A Random Fuzzy Set Approach

Hien D. Tran and Phuong Anh Nguyen

Abstract This paper proposes a realistic way of assessing prior probabilistic information on population parameters in an effort of making Bayesian statistics more robust. The approach is based upon viewing the unknown parameter as a random fuzzy set. To achieve this point of view, we elaborate on the concept of coarsening schemes for gathering experts' opinion, how to combine experts' opinion, and how to define rigorously the concept of random fuzzy sets.

1 Introduction

Bayesian statistics become again a popular statistical approach in econometrics, not only because of reasons such as: statisticians (or engineers) could seek experts' knowledge to assess prior information in addition to observations; when maximum likelihood method in traditional statistics is difficult to solve, say, in high dimensions, or model structures are complicated, but mainly because of computational problems which could be resolved by using the method of Markov Chain Monte Carlo (MCMC) (which "revolutionized" Bayesian statistics).

However, the Bayesian approach to statistical inference is one of the most controversal approaches, due essentially to its subjective nature and automatic inference engine. There is no clear cut as to which approaches (frequentist or Bayesian) an applied statistician should choose. In practice, it looks like the choice is usually based upon each problem at hand.

Applied econometricians will choose a tool which is the most appropriate to analyze the problem under investigation. If the tool happens to be Bayes, are there

H.D. Tran (✉)
Tan Tao University, Long An, Vietnam
e-mail: hien.tran@ttu.edu.vn

P.A. Nguyen
Ho Chi Minh City International University, Ho Chi Minh City, Vietnam
e-mail: npanh@hcmiu.edu.vn

P.A. Nguyen
Vietnam National University, Ho Chi Minh City, Vietnam

V.-N. Huynh et al. (eds.), *Econometrics of Risk*, Studies in Computational Intelligence 583, DOI 10.1007/978-3-319-13449-9_10

anything which could "bother" you? Perhaps, you may say: I am not quite comfortable with specifying a precise prior probability measure on the model parameter space. Perhaps you feel more comfortable to assess that your "true" prior probability measure π_o lies in a class of possible probability measures $\mathcal{P}$, rather than specifying just one probability measure. This will clearly be a generalization of Bayesian methodology concerning assessing prior information. It will be a generalization to make traditional Bayes more realistic, and not just a generalization per se! This natural way to generalize Bayesian prior information assignment is consistent with the spirit of robust statistics in general. See also, robust Bayesian statistics [1].

From the knowledge of $\mathcal{P}$, we can obtain bounds on the true π_o, namely

$$F(A) = \inf\{P(A) : P \in \mathcal{P}\} \leq \pi_o(A) \leq T(A) = \sup\{P(A) : P \in \mathcal{P}\}$$

Thus, research efforts have been focusing on weakening a specific assignment of prior by allowing it to belong to a class of plausible probability measures instead. Specifically, if the random variable of interest has a density $f(x, \theta)$ with $\theta \in \Theta$, then the prior probability law of θ (viewing as a random variable) is only specified to lie in a class $\mathcal{P}$ of probability measures on Θ. But then, knowing only $\mathcal{P}$, we are forced to work with the set function $F(.) = \inf\{P(.) : P \in \mathcal{P}\}$ as a lower bound (or $T(A) = \sup\{P(A) : P \in \mathcal{P}\} = 1 - F(A^c)$, as an upper bound). This approach to generalizing Bayesian statistics is referred to as imprecise probabilities in the literature. In general, these set functions are not additive (i.e., not probability measures) and as such, it is still not clear how the familiar machinery of Bayesian statistics (Bayes formula and posterior expectation) is going to be extended.

In the above framework, the true, but unknown parameter θ_o is viewed as a *random variable*, taking values in Θ. We will "argue" that a more realistic point of view (the point of view is everything!) is regarding θ_o as a *random set, or more generally, a random fuzzy set*, i.e., a random element taking sets or fuzzy sets as values. The main rationale of this view is that we will be working (towards generalizing Bayesian statistics) completely within probability theory, but at a higher "level", namely, set-valued rather than point-valued random elements. The paper is devoted to providing the foundations for this point of view.

2 Coarsening Schemes for Experts' Knowledge

It is a common feature of human intelligence to coarsen a precise domain to arrive at decisions, in everyday life. If we cannot guess the precise age of someone, or cannot measure with accuracy, with our eyes, the distance to an obstacle to stop our car, we coarsen the domain of ages (or the domain of measurements) such as "young, middle age, old" (or "close, very close"). In statistics, coarse data refer to data with low quality (see, e.g., [2]). A *coarsening scheme* is a procedure for transforming a domain into a collection of subsets of it, including the special case of a collection of subsets forming a partition (fuzzy or not) of it.

Suppose a doctor tries to identify the cause of a symptom observed on a patient. After making a series of medical tests, while he knows the set $\Theta = \{\theta_1, \theta_2, \ldots, \theta_n\}$ of all possible causes pertinent to the observed symptom, he is still uncertain about the correct cause. Of course, he has a "better" idea about the possible cause! If you try to "extract" his expert's knowledge, you could ask "what is the probability that the real cause is θ_i", for $i = 1, 2, \ldots, n$. The doctor might try to answer you but with much less ease than if you ask, instead, "what is the probability that the real cause is among A?" where A is a subset of Θ. By doing so, you have presented the doctor with a coarsening scheme to facilitate his answers. In either case, you get an expert's opinion!

The true cause, while unknown, is not a random variable, but since it is uncertain (unknown), we could *view* it as a random variable, exactly like the Bayesians, and we model its uncertainty by a probability distribution on Θ, which is the expert's knowledge. There are two levels of "randomness": on Θ and on 2^{Θ} which 2^{Θ} corresponds to a coarsening scheme. In a rigorous setting, we have a *random variable* taking values in Θ, and a *random set* taking values in 2^{Θ}, both are bona fide *random elements*, defined on appropriate probability spaces.

Thus, if we *generalize* the standard view of Bayesians, namely, viewing an unknown model parameter as a random variable, to a random set, we are actually generalizing Bayesian methodology concerning prior information as random variables are special cases of random sets by identifying random variables with singleton-valued random sets.

In generalizing Bayesian methodology through coarsening schemes (i.e., using random sets to model prior information on the parameter), we remain entirely within the theory of probability, as now, the prior information is a probability measure, not on some σ-field of subsets of Θ, but at a higher level, namely, on some σ-field of subsets of 2^{Θ}. The Bayesians might not object to this modeling!

Now, if it is appropriate to coarsen the domain Θ with fuzzy subsets of it, we can obtain, either probabilities of fuzzy events, or more generally, consider *random fuzzy sets* (see e.g. [3, 4]) as prior information modeling, combining randomness and fuzziness. We elaborate all technical details of this approach in the next section.

3 Random Sets

Let X be an observable random element, defined on $(\Omega, \mathcal{A}, P)$, taking values in $\mathcal{X}$ which is equiped with a σ-field $\sigma(\mathcal{X})$, i.e., X is $\mathcal{A} - \sigma(\mathcal{X})$—measurable. The probability law P_X is assumed to be a member of the statistical model $\mathcal{P}_{\Theta} = \{P_{\theta} : \theta \in \Theta\}$, where each probability measure P_{θ} is defined on $\sigma(\mathcal{X})$.

Statistical inference regarding X can be carried out once the law P_X is "discovered". Usually, we use observations on X to estimate the true (but unknown) θ_o corresponding to P_X. The Bayesian methodology seeks additional information on θ_o prior to making observations on X. Since we do not know θ_o, there is uncertainty about its value in the parameter space Θ. According to the Bayesian approach, the

uncertainty about the true parameter is modeled as a (prior) probability measure π on the parameter space Θ. Formally, θ is viewed as a random variable, defined, say, on $(\Omega, \mathcal{A}, P)$, with values in Θ.

We generalize this standard Bayesian approach as follows.

Let $\mathcal{C} \subseteq 2^{\Theta}$ be a collection of subsets of Θ, representing a *coarsening scheme* of Θ. We equip $\mathcal{C}$ with some σ-field $\sigma(\mathcal{C})$ of its subsets. A random element $S : \Omega \to \mathcal{C}$ is called a *random set* as it takes subsets of Θ as values. It is $\mathcal{A} - \sigma(\mathcal{C})$—measurable, with probability law $P_S = PS^{-1}$, as usual. This probability measure P_S will play the role of probabilistic (prior) information on the parameter θ. In other words, we view the parameter θ as a random set S taking values in $\mathcal{C}$.

We specify the above framework in three popular settings: Θ is a finite set with $\mathcal{C} = 2^{\Theta}$, $\Theta \subseteq \mathbb{R}^d$ (or more generally, a subset of a Hausdorff, locally compact and separable topological space) with $\mathcal{C}$ being the class of closed sets, and with $\mathcal{C}$ being the class of upper-semi continuous functions (fuzzy closed sets). A unified framework for all these settings could be built using *Lawson topology on continuous lattices*.

3.1 Finite Random Sets

The natural coarsening scheme of a finite set Θ is its power set 2^{Θ} with $2^{2^{\Theta}}$ as its σ-field.

A random set S on Θ is a map $S : \Omega \to 2^{\Theta}$ such that , for any $A \subseteq \Theta$, $S^{-1}(\{A\}) = \{\omega \in \Theta : S(\omega) = A\} \in \mathcal{A}\}$, i.e., an $\mathcal{A} - 2^{2^{\Theta}}$—measurable map. Its probability law on $2^{2^{\Theta}}$ is

$$P_S(\mathbb{A}) = PS^{-1}(\mathbb{A}) = \sum_{A \in \mathbb{A}} P(S = A)$$

If we let $f : 2^{\Theta} \to [0, 1]$, with $f(A) = P(S = A)$, then $f(.)$ is a bona fide probability density function on 2^{Θ}, i.e., $f(.) \geq 0$ and $\sum_{A \subseteq \Theta} f(A) = 1$. If $f(\varnothing) = 0$, then we say that S is a nonempty random set. Clearly, X can be characterized either by P_S, its density $f(.)$, or *its distribution function* $F(.) : 2^{\Theta} \to [0, 1]$:

$$F(A) = P(S \subseteq A) = \sum_{B \subseteq A} f(B)$$

Note that, in the definition of $F(.)$ for random sets, the partial order relation (set inclusion $\subseteq$) replaces the partial order relation $\leq$ on $\mathbb{R}^d$.
By Mobius inversion (see, e.g., [5]), where $|A|$ denotes the cardinality of the set A:

$$f(A) = \sum_{B \subseteq A} (-1)^{|A \setminus B|} F(B)$$

Remark 1 The distribution function F of a (nonempty) random set S can be defined axiomatically (just like distribution functions of random vectors) as follows. $F : 2^{\Theta} \to [0, 1]$ is the distribution function of some nonempty random set on a finite Θ if and only if if satisfies the axioms:

1. $F(\varnothing) = 0$, $F(\Theta) = 1$
2. For any $k \geq 2$, and $A_1, A_2, \ldots, A_k$ subsets of Θ,

$$F(\cup_{j=1}^{k}) \geq \sum_{\varnothing \neq I \subseteq \{1,2,\ldots,k\}} (-1)^{|I|+1} F(\cap_{i \in I} A_i)$$

Here are some examples of distributions of finite random sets.

Example 1 Let $\Theta = \{\theta_1, \theta_2, \theta_3\}$. The true P_o on Θ is known only up to the extent that $P_o(\theta_1) = \frac{1}{3}$, and then, of course, $P_o(\{\theta_2, \theta_3\}) = \frac{2}{3}$. Let $\mathcal{P}$ denote the set of all probability measures P having this condition. If we define $f(\theta_1) = \frac{1}{3}$, $f(\{\theta_2, \theta_3\}) = \frac{2}{3}$, and $f(A) = 0$ for all other subsets on Θ, then

$$F(A) = \sum_{B \subseteq A} f(A) = \inf\{P(A) : P \in \mathcal{P}\}$$

noting also that $\mathcal{P} = \{P : F \leq P\}$.

Example 2 Let $\Theta = \{1, 5, 10, 20\}$. Let

$$\mathcal{P} = \{P : P(1) \geq 0.4, P(5) \geq 0.2, P(10) \geq 0.2, P(20) \geq 0.1\}$$

Let $f(1) = 0.4$, $f(5) = 0.2$, $f(10) = 0.2$, $f(20) = 0.1$, and $f(\{1, 5, 10, 20\}) = 0.1$, then

$$\mathcal{P} = \{\mathcal{P} : F \leq P\} \text{ where } F(A) = \textstyle\sum_{B \subseteq A} f(B),$$

so that $F = \inf \mathcal{P}$.

Example 3 Let $\Theta_1, \Theta_2, \ldots, \Theta_k$ be a partition of the finite Θ. Let

$$\mathcal{P} = \{P : P(\Theta_i) = \alpha_i, i = 1, 2, \ldots, k\},$$

then

$$F = \inf \mathcal{P} = \inf\{P : F \leq P\}$$

which is a distribution of some random set.

In general, $F = \inf \mathcal{P}$ might not be a distribution function of a random set, and hence does not have a random set interpretation. But what interesting is this. If F is the distribution function of a random set, then it is a *lower probability*, i.e., there

exists a class $\mathcal{P}$ of probability measures on Θ such that $F(A) = \inf\{Q(A) : Q \in \mathcal{P}\}$, implying that the random set approach to modeling prior information is consistent with the main stream of generalizing Bayesian inference.

For ease of reference, here is the proof of the above fact. We denote by $\mathcal{P}_F$ the set of probability measures Q on Θ such that $Q(.) \geq F(.)$. Let f be the density on 2^Θ derived from F (via Mobius inversion). From f we can construct probabilities densities on Θ in a natural way: assigning nonnegative values to elements of $A \in 2^\Theta$ so that their sum is equal to $f(A)$. An *allocation* of f is a function $\alpha : \Theta \times 2^\Theta \to [0, 1]$ such that $\sum_{\theta \in \Theta} \alpha(\theta, A) = f(A)$ for all $A \in 2^\Theta$.

Example 4 Let $p : \Theta$ (finite)$\to [0, 1]$ be a probability density. Let Q be its associated probability measure: $Q(A) = \sum_{\theta \in A} p(\theta)$. Then

$$\alpha(\theta, A) = \frac{f(A)}{Q(A)} p(\theta)$$

is an allocation. In particular, if $p(.)$ is uniform, i.e., $p(\theta) = \frac{1}{|\Theta|}$, then $\alpha(\theta, A) = \frac{f(A)}{|A|}$, for any $\theta \in \Theta$.

Now observe that, for each allocation α, the map

$$g_\alpha : \theta \in \Theta \to \sum_{A \in 2^\Theta} \alpha(\theta, A)$$

is a probability density on Θ. Let $\mathcal{D}$ denote the set of densities on Θ arising from allocations of f.

Let σ be the map which takes a density on Θ to the corresponding probability measure on Θ. Then $\sigma : \mathcal{D} \to \mathcal{P}_F$. Indeed, let α be an allocation of f, and let Q_α be the probability measure on Θ that α induces:

$$Q_\alpha(A) = \sum_{\theta \in A} g_\alpha(\theta) = \sum_{\theta \in A} \sum_{B \in 2^\Theta} \alpha(\theta, B) = \sum_{\theta : \theta \in A} \sum_{B : \theta \in B} \alpha(\theta, B)$$

Now,

$$F(A) = \sum_{B : B \subseteq A} f(B) = \sum_{B : B \subseteq A} \sum_{\theta : \theta \in B} \alpha(\theta, B)$$

so that clearly $F(A) \leq Q_\alpha(A)$, implying that $Q_\alpha \in \mathcal{P}_F$.

In fact, $\sigma : \mathcal{D} \to \mathcal{P}_F$ is onto, i.e., $\mathcal{P}_F$ consists of probability measures on Θ coming from allocations. The proof of this fact relies on Shapley' theorem in game theory (See [6], pp. 101–102). Let $A \in 2^\Theta$. Let α be an allocation of f such that for $\theta \in A$, for all B not contained in A, allocate 0 to $\alpha(\theta, B)$. Then $F(A) = Q_\alpha(A)$. It follows that

$$F(A) = \inf\{Q(A) : Q \in \mathcal{P}_F\}$$

Thus, if we insist on having prior information about the model parameter θ as its prior probability measure π on Θ, then we can only have a lower bound $F(.)$ (and, of course, an upper bound $T(A) = 1 - F(A^c)$) which is a nonadditive set function. Should we "view" the distribution F or its associated *probability measure* P_S, where

$$P_S(\mathbb{A}) = \sum_{A \in \mathbb{A}} \sum_{B \subseteq A} (-1)^{|A \setminus B|} F(B)$$

as a generalization of π?

Note that when working with F, we could use Choquet integral in the Bayesian machinery for reaching posterior expectation. We will discuss this technical issue elsewhere.

3.2 Random Closed Sets

We look now at the case where our coarsening scheme consists of a collection of subsets of a general topological space Θ such as $\mathbb{R}^d$, or more generally, a Hausdorff, second countable, locally compact space. In view of [7], we take our coarsening scheme on $\Theta = \mathbb{R}^d$ as the collection $\mathcal{F}$ of all closed subsets of $\mathbb{R}^d$.

In order to define rigorously the concept of *random closed sets* as bona fide random elements taking values in $\mathcal{F}$, besides having a probability space $(\Omega, \mathcal{A}, P)$ in the background, we need to equip $\mathcal{F}$ with some σ-field of its subsets. A standard way to accomplish this is to topologize $\mathcal{F}$ and take its Borel σ-field. This was precisely what Matheron did.

But since we will next consider a more general situation, namely "*random fuzzy closed sets",* we will employ a general construction method which produces both types of random sets, namely that of *Lawson's topology* in the theory of *continuous lattices* (see [8]). Note that the Matheron's construction of the so-called "hit-or-miss" topology for $\mathcal{F}$ cannot be extended to fuzzy closed sets in a straightforward manner.

Recall that a coarsening scheme on a set Θ aims at facilating answers for experts in the knowledge extraction process. As such, each subset $A \subseteq 2^\Theta$ contains *localization information* about the true (but unknown) parameter θ_o. In this sense, smaller sets contain more information than larger ones. This is exactly what *information theory* is all about (for example, if the information provided by the realization of an event A is $I(A) = -\log P(A)$, then $A \subseteq B \Longrightarrow I(A) \geq I(B)$). Note that "less informative" also means "less specific" as far as localization of an object is concerned.

Thus, in the context of coarsening, the partial order relation on 2^Θ "less informative than" is nothing else than the reverse of set inclusion $\subseteq$, i.e., $\supseteq$. The poset $(2^\Theta, \supseteq)$ is a complete lattice. For $\Theta = \mathbb{R}^d$, we will look at the class of its closed subsets $\mathcal{F}$, rather than the whole $2^{\mathbb{R}^d}$. The poset $(\mathcal{F}, \supseteq)$ is also a complete lattice. Indeed,

$$\wedge(F_i \in \mathcal{F} : i \in I\} = \quad \text{the closure of} \quad \cup_{i \in I} F_i$$

$$\vee\{F_i \in \mathcal{F} : i \in I\} = \cap_{i \in I} F_i$$

Moreover, it is a continuous lattice, i.e., for every $F \in \mathcal{F}$, we have $F = \vee\{G \in \mathcal{F} : G \Subset F\}$ where the finer relation $\Subset$ ("much less informative than" or "*way below*") is defined as: $G \Subset F$ (G is way below F, meaning G is much less informative than F) if for every collection D of elements of $\mathcal{F}$, for which $F \supseteq \vee D$, there is $d \in D$ such that $G \supseteq d$.

The *Lawson topology* τ for the continuous lattice $(\mathcal{F}, \supseteq, \Subset)$ has as a subbase the sets $\{G \in \mathcal{F} : G \Subset F\}$ and $\{G \in \mathcal{F} : G \not\supseteq F\}$ for all $F \in \mathcal{F}$, i.e., open sets are taken to be arbitrary unions of finite intersections of these sets. The Borel σ-field $\sigma(\tau)$ is precisely the Matheron's hit-or-miss σ-field $\sigma(\mathcal{F})$. More specifically, the Lawson topology τ on $\mathcal{F}$ is the Matheron's hit-or-miss topology, since τ is generated by the subbase consisting of subsets of the from $\{F \in \mathcal{F} : F \cap K = \varnothing\}$ and $\{F \in \mathcal{F} : F \cap G \neq \varnothing\}$, for K, G compact and open, respectively. For details, see, e.g., [4].

Thus, by a *random closed set* on $\mathbb{R}^d$, we mean a map $X : \Omega \to \mathcal{F}$ such that $X^{-1}(\sigma(\mathcal{F})) \subseteq \mathcal{A}$. Its probability law is a probability measure P_X on $\sigma(\mathcal{F})$. Unlike general infinitely dimensional spaces, there is a counter part of the Lebesgue-Stieltjes theorem for random closed sets, namely the Choquet theorem that characterizes probability measures on $\sigma(\mathcal{F})$ by distribution functions of random closed sets. See, e.g., [6]. Specifically, the dual of a distrubution function, i.e., the capacity functional $T : \mathcal{K}$ (class of compact sets of $\mathbb{R}^d$) $\to [0, 1]$, of a random closed set, defined by $T(K) = P(X \cap K \neq \varnothing) = P_X(\mathcal{F}_K)$, where

$$\mathcal{F}_K = \{F \in \mathcal{F} : F \cap K \neq \varnothing\}$$

characterizes P_X. It is this (nonadditive) set function on $\mathcal{K}$ which plays the role of distribution functions of random finite sets.

3.3 Random Fuzzy Closed Sets

Often coarsening schemes could be formed by using natural language, such as "small", "medium", "large".... The quantification of these semantics can be provided by membership functions of fuzzy sets (see, e.g., [9]). In the Bayesian spirit, we view the unknown model parameter as a random element taking fuzzy subsets of, say, $\mathbb{R}^d$. We restrict ourself to the case of fuzzy closed sets. The indicator function 1_F of an (ordinary) closed subset F of $\mathbb{R}^d$ is upper-semi continuous, i.e., $\{x \in \mathbb{R}^d : 1_F(x) \geq \alpha\}$ is a closed set, for every $\alpha \in \mathbb{R}$. Thus, a fuzzy subset is said to be closed (a *fuzzy closed set*) if its membership function $f : \mathbb{R}^d \to [0, 1]$ is upper-semi continuous: $\{x \in \mathbb{R}^d : f(x) \geq \alpha\}$ is a closed set, for every $\alpha \in \mathbb{R}$. To define a *random fuzzy closed set*, we will topologize the class of all fuzzy closed sets, denoted as $\mathcal{J}$, using Lawson topology, as in the case of ordinary closed sets.

As in the case of closed sets, in order to obtain a continuous lattice for $\mathcal{J}$, we consider the partial order relation "f is less informative than g" if $f(x) \geq g(x)$ for all $x \in \mathbb{R}^d$. Then $(\mathcal{J}, \geq)$ is a complete and continuous lattice. Its Lawson topology

provides a Borel σ-field $\sigma(\mathcal{J})$ for $\mathcal{J}$ from which we can fomulate the rigorous notion of random fuzzy sets. A choquet theorem for $\mathcal{J}$ can be obtained by embedding $\mathcal{J}$ into the closed sets of $\mathbb{R}^d \times [0, 1]$ via hypographs and using Choquet theorem for random closed sets on it. Details were published elsewhere.

Remark 2 Perhaps a question that could come to everybody's mind is this. *What are the benefits of having a random set representation in the process of generalizing Bayesian statistics?* We simply say this here. The random set representations provide an appropriate (and rigorous/logical) framework for

(a) combining experts' opinion in assessing prior information,
(b) specifying (generalized) Bayesian priors in multivariate statistical models.

The key ingredient for the above problems is *copulas for random sets.* Let's elaborate a bit here on specifying (generalized) Bayesian priors in multivariate statistical models.

In one dimension case, it seems not too difficult to "assign" values to the mass or density f, just like in univariate statistical modeling. If $\Theta \subseteq \mathbb{R}^d$ with d "high", it is not clear how practitioners would assign a mass function on 2^{Θ}. Note that, what we are facing is nothing else than multivariate statistical modeling, where, instead of dealing with random vectors, we are dealing with random *set* vectors, but we still entirely with probability theory, since random sets are bona fide random elements.

For concreteness, consider two random set representations X and Y corresponding to two distribution functions F and G, respectively, on, say, two finite domains U and V. The pair (X, Y) is a *vector of random sets* which takes values in $2^U \times 2^V$ instead of $2^{U\times V}$. If we identify $(A, B) \in 2^U \times 2^V$ with $A \times B \in 2^{U\times V}$, then a vector of random sets is a special case of bivariate (multivariate) random sets. Note that there is no difference between multivariate random *variables* (point-valued maps) and vectors of random variables. The situation is different with set-valued random elements (random sets). A vector of random sets, or random set vector, is a special multivariate random set, namely it only takes "rectangles" as values rather than arbitrary subsets of $U \times V$. As such, the distribution of a vector (X, Y) on $2^U \times 2^V$ is referred to as a *joint distribution of a random set vector*, to distinguish with function a multivariate distribution function in general.

Specifically, the joint distribution function associated with (X, Y) is just the two-dimensional version of the univariate one, namely, it is a set function $H : 2^U \times 2^V \to [0, 1]$, given by

$$H(A, B) = P(X \subseteq A, Y \subseteq B)$$

First of all, since $2^U \times 2^V$ is finite, the "joint" probability measure $P_{(X,Y)}$ on the power set of $2^U \times 2^V$ is completely determined by its joint density $h : 2^U \times 2^V \to [0, 1]$, where

$$h(A, B) = P(X = A, Y = B)$$

Thus, $H(.,.)$ characterizes the probability law of the vector (X, Y). Now,

$$H(A, B) = P(X \subseteq A, Y \subseteq B)$$

$$= \sum_{A' \subseteq A, B' \subseteq B} P(X = A, Y = B) = \sum_{A' \subseteq A, B' \subseteq B} h(A', B')$$

The density h is recovered from its distribution H simply by Mobius inversion on the product poset $(2^U \times 2^V, \leq)$ where the partial order $\leq$ on $2^U \times 2^V$ is defined pointwise: $(A', B') \leq (A, B)$ means $A' \subseteq A$ and $B' \subseteq B$. As such, it is well-known that the Mobius function on $(2^U \times 2^V, \leq)$ is simply the product, namely

$$\mu : (2^U \times 2^V) \times (2^U \times 2^V) \to \mathbb{Z}$$

$$\mu[(A', B'), (A, B)] = (-1)^{|A \setminus A'| + |B \setminus B'|}$$

and the Mobius inverse of H is

$$h(A, B) = (F * \mu)(A, B) = \sum_{(A', B') \leq (A, B)} \mu[(A', B'), (A, B)] F(A', B')$$

$$= \sum_{(A', B') \leq (A, B)} (-1)^{|A \setminus A'| + |B \setminus B'|} F(A', B')$$

What are the characteristic properties of a joint distribution $H(A, B) = P(X \subseteq A, Y \subseteq B)$ of a random set vector, assuming that both X and Y are *nonempty* random sets?

The recent work of [10] provides an axiomatic concept of joint distribution of (nonempty) random sets. Specifically, *a joint distribution function* of a random set vector (X, Y) is a set function $H : 2^U \times 2^V \to [0, 1]$, satisfying the conditions (i), (ii) and (iii) below:

(i) $H(\varnothing, \varnothing) = 0$
(ii) $H(U, V) = 1$
(iii) $H(.,.)$ is *jointly* monotone of infinite order.

Remark 3 Note that $(2^U, \subseteq)$ and $(2^V, \subseteq)$ are (locally) finite partially ordered sets (posets). The product $2^U \times 2^V$ is then equiped with the natural order $(A', B') \leq (A, B)$ if and only if $A' \subseteq A$ and $B' \subseteq B$, so that $(2^U \times 2^V, \leq)$ is a finite poset. It is well-known that the Mobius function of $(2^U \times 2^V, \leq)$ can be obtained from the Mobius functions on the posets $(2^U, \subseteq)$ and $(2^V, \subseteq)$, namely

$$\mu[(A', B'), (A, B)] = (-1)^{|A \setminus A'| + |B \setminus B'|}$$

As H is a real-valued function, defined on $(2^U \times 2^V, \leq)$, its Mobius inverse is $h = H * \mu$, resulting in $h : 2^U \times 2^V \to [0, 1]$

$$h(A, B) = \sum_{(A', B') \leq (A, B)} (-1)^{|A \setminus A'| + |B \setminus B'|} H(A', B')$$

with

$$H(A, B) = \sum_{(A', B') \leq (A, B)} h(A', B')$$

Thus, to see whether or not a function H, satisfying (i), (ii) and (iii) above admits a random set representation, it suffices to find out whether or not h is a bona fide joint probability density on $2^U \times 2^V$ with $h(\varnothing, \varnothing) = 0$.

If $H : 2^U \times 2^V \to [0, 1]$ satisfies

(i) $H(\varnothing, \varnothing) = 0$
(ii) $H(U, V) = 1$
(iii) $H(., .)$ is jointly monotone of infinite order, then there exist $(\Omega, \mathcal{A}, P)$ and $(X, Y) : \Omega \to 2^U \times 2^V$, a nonempty random set vector, such that $H(A, B) = P(X \subseteq A, Y \subseteq B)$.

4 Concluding Remarks

A realistic way to assess a prior probability distribution is to seek experts' opinion. This process may require at least two things: facilitating the acquisition of experts' knowledge (this can be achieved by designing appropriate coarsening schemes on the parameter space), and combining experts' opinion (this can be achieved by using copulas to obtain joint distributions of random fuzzy sets, in a rigorous fashion). The combined distribution of a (finite) random fuzzy set can be served as a realistic prior assessment for the unknown parameter. Since the analysis is entirely within probability theory, it is expected that the Bayesian machinery for deriving posterior distribution could be carried out. This will be our future work.

References

1. Huber, P.J.: The use of Choquet capacities in statistics. Bull. Inst. Int. Stat. XLV, Book **4**, 181–188 (1973)
2. Heitjan, D.F., Rubin, D.B.: Ignorability and coarse data. Ann. Stat. **19**, 2244–2253 (1991)
3. Nguyen, H.T., Tran, H.: On a continuous lattice approach to modeling of coarse data in system analysis. J. Uncertain Syst. **1**(1), 62–73 (2007)

4. Nguyen, H.T., Kreinovich, V., Xiang, G.: Random fuzzy sets. In: Wang, H.-F. (ed.) Intelligent Data Analysis: Developing New Methodologies Through Pattern Discovery and Recovery, pp. 18–44. ICI Global, Hershey, Pennsylvania (2008)
5. Rota, G.C.: On the foundations of combinatorial theory I: theory of Mobius functions. A. Wahrsch. **2**, 340–368 (1964)
6. Nguyen, H.T.: An Introduction to Random Sets. Chapman and Hall/CRC Press (2006)
7. Matheron, G.: Random Sets and Integral Geometry. Wiley, New York (1975)
8. Gierz, G., et al.: A Compendium of Continuous Lattices. Springer, Berlin (1980)
9. Nguyen, H.T., Walker, E.A.: A First Course in Fuzzy Logic. Chapman and Hall/CRC Press (2005)
10. Schmelzer, B.: Characterizing joint distributions of random sets by multivariate capacities. Int. J. Approx. Reason. **53**, 1228–1247 (2012)

Local Kendall's Tau

P. Buthkhunthong, A. Junchuay, I. Ongeera, T. Santiwipanont and S. Sumetkijakan

Abstract We introduce two local versions of Kendall's tau conditioning on one or two random variable(s) varying less than a fixed distance. Some basic properties are proved. These local Kendall's taus are computed for some shuffles of Min and the Farlie-Gumbel-Morgenstern copulas and shown to distinguish between complete dependence and independence copulas. A pointwise version of Kendall's tau is also proposed and shown to distinguish between comonotonicity and countermonotonicity for complete dependence copulas.

1 Introduction and Preliminaries

Let (X, Y) and (X', Y') be independent and identically distributed random vectors. Then the population version of the Kendall's tau of X, Y is defined as

$$\tau(X, Y) \equiv P\left((X' - X)(Y' - Y) > 0\right) - P\left((X' - X)(Y' - Y) < 0\right).$$

If X and Y are continuous random variables with joint distribution function $F_{X,Y}$ then there exists a unique copula $C = C_{X,Y}$ such that $F_{X,Y}(u, v) = C(F_X(u), F_Y(v))$ for all $u, v \in [0, 1]$ where F_X and F_Y are the distribution functions of X and Y, respectively. A copula can be defined as the restriction onto I^2 of a joint distribution

P. Buthkhunthong · A. Junchuay · I. Ongeera ·
T. Santiwipanont · S. Sumetkijakan (✉)
Department of Mathematics and Computer Science, Faculty of Science, Chulalongkorn University, Bangkok, Thailand
e-mail: songkiat.s@chula.ac.th

P. Buthkhunthong
e-mail: pitchaya.bu@gmail.com

A. Junchuay
e-mail: anusorn-j@outlook.com

I. Ongeera
e-mail: ilada.ogr@gmail.com

T. Santiwipanont
e-mail: tippawan.s@chula.ac.th

V.-N. Huynh et al. (eds.), *Econometrics of Risk*, Studies in Computational Intelligence 583, DOI 10.1007/978-3-319-13449-9_11

of two uniform (0, 1) random variables. See [6] for a systematic treatment of the theory of copulas. In this setting, the Kendall's tau can be written in terms of the copula C as

$$\tau(C) = \tau(X, Y) = 4 \iint_{I^2} C(u, v)\, dC(u, v) - 1.$$

Kendall's tau is a measure of concordance in the sense of Scarsini [8], i.e. $\tau(X, Y)$ is defined for all continuous random variables X and Y; τ attains value in $[-1, 1]$; τ is equal to 0 if X and Y are independent; τ is symmetric ($\tau(Y, X) = \tau(X, Y)$); τ is coherence ($C_{XY} \leq C_{X'Y'}$ implies $\tau(X, Y) \leq \tau(X', Y')$); $\tau(-X, Y) = -\tau(X, Y)$; and τ is continuous with respect to convergence in distribution.

However, there are copulas of dependent random variables whose Kendall's tau is zero. An extreme example is the copula $S_{1/2}$, defined in Example 2.1, of continuous random variables that are completely dependent by an injective function which is strictly increasing on two disjoint intervals separated at the median. Its Kendall's tau is zero because of cancellation between local concordance and global discordance. Local dependence has been studied in [1, 3, 4]. To bring more local dependence into focus, we propose two local versions of Kendall's tau called uni- and bi-conditional local Kendall's taus in Sects. 2 and 3. Their formulas for shuffles of Min and FGM copulas are given and their basic properties are proved. In particular, $S_{1/2}$ has non-zero local Kendall's taus.

Both uni- and bi-conditional local Kendall's taus are conditioning on one or two random variables varying less than a fixed small distance. They are measures of local concordance/discordance between two random variables without restriction on the range of the conditioning random variable(s). In a sense, they detect local dependence globally. In order to detect true local dependence, we introduce a pointwise Kendall's tau in Sect. 4. Its empirical version was first introduced in [2] and shown to detect monotonicity of two random variables. We show that the pointwise Kendall's tau can distinguish between comonotonicity and countermonotonicity for complete dependence copulas.

2 Uni-conditional Local Kendall's Tau

Let X and Y be continuous random variables with the copula C. We will consider the difference between the probabilities of concordance and discordance conditioning on the event that X is varying less than a fixed distance regardless of the value of X. Since the same amount of variation of X could reflect different interpretations depending on the X-value, we assume that X and Y are uniformly distributed on $[0, 1]$. Let $0 < \varepsilon \leq 1$. The *uni-conditional local Kendall's tau* of X and Y given that X varies less than ε is defined as

$$\tau_\varepsilon(C) = \tau_\varepsilon(X, Y) = P\left((X - X')(Y - Y') > 0 || X - X'| < \varepsilon\right) - P\left((X - X')(Y - Y') < 0 || X - X'| < \varepsilon\right)$$

where (X', Y') is an independent copy of (X, Y). The following quantity shall be needed in computing local Kendall's tau.

$$T_C(a, b) = \iint_{I^2} C(u - a, v - b)\, dC(u, v) \quad \text{for } -1 \le a, b \le 1. \tag{1}$$

Note that $T_C(a, b) = P(X' - X > a, Y' - Y > b)$. By uniform continuity of C, the function T_C is continuous on $[-1, 1]^2$ and in particular $T_C(\cdot, 0) : a \mapsto T_C(a, 0)$ is continuous on $[-1, 1]$.

Proposition 2.1 1. *The uni-conditional local Kendall's tau of a copula C can be computed by the formula*

$$\tau_\varepsilon(C) = \frac{4T_C(0, 0) - 2\left[T_C(-\varepsilon, 0) + T_C(\varepsilon, 0)\right]}{\varepsilon(2 - \varepsilon)}. \tag{2}$$

2. *The mapping $\varepsilon \mapsto \tau_\varepsilon(C)$ is continuous on* $(0, 1]$.

Proof 1. It is straightforward to verify that $P(|X - X'| < \varepsilon) = 2\varepsilon - \varepsilon^2$,

$$\begin{aligned} P\left(-\varepsilon < X - X' < 0, Y - Y' < 0\right) &= P\left(0 < X' - X < \varepsilon, 0 < Y' - Y\right) \\ &= T_C(0, 0) - T_C(\varepsilon, 0), \quad \text{and} \\ P\left(-\varepsilon < X - X' < 0, 0 < Y - Y'\right) &= T_C(-\varepsilon, 0) - T_C(0, 0). \end{aligned}$$

Therefore,

$$\begin{aligned} P\left((X - X')(Y - Y') > 0, |X - X'| < \varepsilon\right) &= 2\left(T_C(0, 0) - T_C(\varepsilon, 0)\right), \\ P\left((X - X')(Y - Y') < 0, |X - X'| < \varepsilon\right) &= 2\left(T_C(-\varepsilon, 0) - T_C(0, 0)\right), \end{aligned}$$

and the formula follows.

2. This is clear from Eq. (2) and the continuity of $T_C(\cdot, 0)$.

Recall the definition of three important copulas Π, M and W: for $u, v \in [0, 1]$, $\Pi(u, v) = uv$, $M(u, v) = \min(u, v)$ and $W(u, v) = \max(u + v - 1, 0)$. Then it can be shown via the Eq. (2) that $\tau_\varepsilon(\Pi) = 0$, $\tau_\varepsilon(M) = 1$, $\tau_\varepsilon(W) = -1$ for all $\varepsilon \in (0, 1]$.

Example 2.1 Let us consider simple shuffles of Min introduced in [5]. Let S_α be the shuffle of Min whose support is illustrated in Fig. 1 and defined by

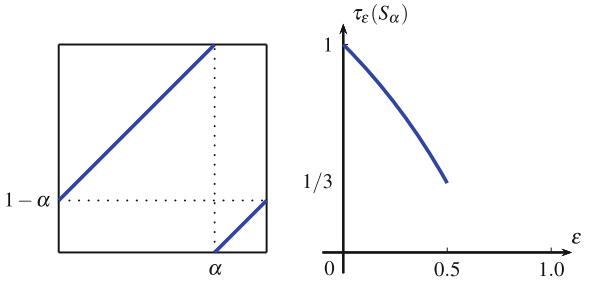


Fig. 1 The support of S_α and the uni-conditional local Kendall's tau of S_α

$$S_\alpha(x,y) = \begin{cases} 0 & \text{if } 0 \le x \le \alpha, 0 \le y \le 1-\alpha, \\ \min(x, y-(1-\alpha)) & \text{if } 0 \le x \le \alpha, 1-\alpha < y \le 1, \\ \min(x-\alpha, y) & \text{if } \alpha < x \le 1, 0 \le y \le 1-\alpha, \\ x+y-1 & \text{if } \alpha < x \le 1, 1-\alpha < y \le 1. \end{cases} \tag{3}$$

Since S_α is supported on the lines $\ell_1 \colon y = x + (1-\alpha), 0 \le x \le \alpha$ and $\ell_2 \colon y = x - \alpha, \alpha \le x \le 1$, we have

$$\iint_{I^2} f(x,y)\, dS_\alpha(x,y) = \int_0^\alpha f(x, x+1-\alpha)\, dx + \int_\alpha^1 f(x, x-\alpha)\, dx. \tag{4}$$

Because $S_\alpha(x-a, y-b)$ has positive value on rectangles $(a,b)+R_i$ where $R_1 \equiv [\alpha,1]\times[0,1-\alpha]$, $R_2 \equiv [0,\alpha]\times[1-\alpha,1]$, $R_3 \equiv [\alpha,1]\times[1-\alpha,1]$, $R_4 \equiv [1,\infty]\times[0,1]$, $R_5 \equiv [0,1]\times[1,\infty]$, and $R_6 \equiv [1,\infty]\times[1,\infty]$, it can be derived that

$$T_{S_\alpha}(a,b) = \int_{L_1} S_\alpha(x-a, (x+1-\alpha)-b)\, dx + \int_{L_2} S_\alpha(x-a, (x-\alpha)-b)\, dx \tag{5}$$

where each L_1 and L_2 is a union of six non-overlapping possibly empty intervals.[1] For $\frac{1}{2} \le \alpha < 1$ and $0 < \varepsilon \le \min(\alpha, 1-\alpha)$, $T_{S_\alpha}(0,0) = \frac{1}{2} - \alpha(1-\alpha)$, $T_{S_\alpha}(\varepsilon,0) = \frac{1}{2} - \alpha - \varepsilon + \alpha^2 + \varepsilon^2$, $T_{S_\alpha}(-\varepsilon,0) = \frac{1}{2} - \alpha + \alpha^2 + \frac{\varepsilon^2}{2}$, and hence by (2),

$$\tau_\varepsilon(S_\alpha) = \frac{2-3\varepsilon}{2-\varepsilon}$$

[1] $L_1 = [\max(0,\alpha+a,\alpha-1+b), \min(\alpha,1+a,b)] \cup [\max(0,a,b), \min(\alpha,\alpha+a,\alpha+b)] \cup [\max(0,\alpha+a,b), \min(\alpha,1+a,b+\alpha)] \cup [\max(0,1+a,\alpha-1+b), \min(\alpha,b+\alpha)] \cup [\max(0,\alpha+b,a), \min(\alpha,1+a)] \cup [\max(0,\alpha+b,1+a),\alpha]$ and $L_2 = [\max(1-\alpha,\alpha+a,\alpha+b), \min(1,1+a,1+b)] \cup [\max(1-\alpha,a,1+b), \min(1,\alpha+a,\alpha+1+b)] \cup [\max(1-\alpha,\alpha+a,1+b), \min(1,1+a,\alpha+1+b)] \cup [\max(1-\alpha,1+a,\alpha+b), \min(1,\alpha+1+b)] \cup [\max(1-\alpha,a,\alpha+1+b), \min(1,1+a)] \cup [\max(1-\alpha,1+a,\alpha+1+b), 1]$.

as shown in Fig. 1. Surprisingly, $\tau_\varepsilon(S_\alpha)$ is independent of α when ε is sufficiently small. But it is not unexpected that we obtain $\lim_{\varepsilon\to 0^+} \tau_\varepsilon(S_\alpha) = 1$.

For any given copula C, we then investigate the limit of $\tau_\varepsilon(C)$ as ε goes down to 0, denoted by $\tau_{\text{loc}}(C)$:

$$\tau_{\text{loc}}(C) = \lim_{\varepsilon\to 0^+} \tau_\varepsilon(C)$$

wherever the limit exists. The left-hand and right-hand derivatives are denoted respectively by

$$\partial_1^- C(u,v) = \lim_{\varepsilon\to 0^-} \frac{C(u+\varepsilon,v) - C(u,v)}{\varepsilon} \text{ and}$$
$$\partial_1^+ C(u,v) = \lim_{\varepsilon\to 0^+} \frac{C(u+\varepsilon,v) - C(u,v)}{\varepsilon};$$

and $\partial_1 C(u,v) = \partial_1^+ C(u,v) = \partial_1^- C(u,v)$ wherever the one-sided derivatives exist and are equal. Let μ_C denote the doubly stochastic measure on $[0,1]^2$ induced by C.

Theorem 2.1 *Let $0 < \varepsilon < 1$ and C be a copula. Then*

$$\tau_{\text{loc}}(C) = \iint_{I^2} \left(\partial_1^- C(u,v) - \partial_1^+ C(u,v)\right)\, dC(u,v)$$

provided that the set of points (u,v) where the left and right partial derivatives $\partial_1^- C(u,v)$ and $\partial_1^+ C(u,v)$ exist has C-volume one.

Proof By Proposition 2.1, $\tau_\varepsilon(C)$ can be reformulated as

$$\tau_\varepsilon(C) = \frac{2}{2-\varepsilon} \iint_{I^2} \left[\frac{C(u,v) - C(u-\varepsilon,v)}{\varepsilon} - \frac{C(u+\varepsilon,v) - C(u,v)}{\varepsilon}\right] dC(u,v).$$

Note that both quotients are bounded by 1 and the integral is with respect to a finite measure μ_C. Applying the dominated convergence theorem on the set of points (u,v) where the first quotient converges to $\partial_1^- C(u,v)$ and the second quotient converges to $\partial_1^+ C(u,v)$, we have the desired identity.

Corollary 2.1 *Let C be a copula. If $\partial_1 C$ exists for μ_C-almost everywhere on I^2 then $\tau_{\text{loc}}(C) = 0$.*

Example 2.2 Let C be a copula. If $\partial_1 C$ exists everywhere then $\tau_{\text{loc}}(C) = 0$. In particular, if C_θ is a Farlie-Gumbel-Morgenstern (FGM) copula, then $\tau_{\text{loc}}(C_\theta) = 0$ for all θ.

Example 2.3 We show that $\tau_{\text{loc}}(S) = 1$ for all straight shuffles of Min S. Let (u,v) be in the support of S. From the assumption that S is a straight shuffle of Min, there

exists $\delta > 0$ such that $\partial_1 S(t, v) = 1$ for $u - \delta < t < u$ and $\partial_1 S(t, v) = 0$ for $u < t < u + \delta$. Since S is continuous, we have $\partial_1^- S(u, v) = 1$ and $\partial_1^+ S(u, v) = 0$ at all (u, v) in the support of S. Hence

$$\tau_{\text{loc}}(S) = \iint_{I^2} \left(\partial_1^- S(u, v) - \partial_1^+ S(u, v)\right) dS(u, v) = \mu_S(\text{supp } S) = 1.$$

Similar arguments show that if S is a flipped shuffle of Min then $\tau_{\text{loc}}(S) = -1$ and for any shuffle of Min S,

$$\tau_{\text{loc}}(S) = \lambda(I_S) - \lambda(D_S)$$

where I_S (D_S) is the set of points u for which the support of S is a line of slope 1 (-1) in a neighbourhood of $(u, v) \in S$.

3 Bi-conditional Local Kendall's Tau

For uniform (0, 1) random variables X and Y with copula C and any $0 < \varepsilon \leq 1$, the *bi-conditional local Kendall's tau* of X and Y, or equivalently of C, given that both X and Y vary less than ε is defined as

$$\begin{aligned}\tau_{[\varepsilon]}(C) = {} & P\left((X_1 - X_2)(Y_1 - Y_2) > 0 \,||X_1 - X_2| < \varepsilon, |Y_1 - Y_2| < \varepsilon\right) \\ & - P\left((X_1 - X_2)(Y_1 - Y_2) < 0 \,||X_1 - X_2| < \varepsilon, |Y_1 - Y_2| < \varepsilon\right).\end{aligned}$$

Theorem 3.1 *The bi-conditional local Kendall's tau can be computed by*

$$\tau_{[\varepsilon]}(C) = \frac{2\left[2T_C(0, 0) - 3T_C(\varepsilon, 0) + T_C(\varepsilon, \varepsilon) - T_C(-\varepsilon, 0) + T_C(\varepsilon, -\varepsilon)\right]}{T_C(\varepsilon, \varepsilon) - 2T_C(\varepsilon, -\varepsilon) + T_C(-\varepsilon, -\varepsilon)}.$$

Proof This results from a long calculation similar to the proof of Proposition 2.1 but much more tedious.

Theorem 3.2 $\tau_{[\varepsilon]}(C)$ *is a continuous function of* $\varepsilon \in (0, 1]$.

Proof This is because T_C is continuous on $[-1, 1]^2$.

Let us define $\tau_{[\text{loc}]}(C) = \lim_{\varepsilon \to 0^+} \tau_{[\varepsilon]}(C)$.

Example 3.1 $\tau_{[\varepsilon]}(\Pi) = 0$, $\tau_{[\varepsilon]}(M) = 1$, $\tau_{[\varepsilon]}(W) = -1$ for all $\varepsilon \in [0, 1]$.

Example 3.2 For $\alpha \geq \frac{1}{2}$ and $0 < \varepsilon < \frac{1}{2}\min(\alpha, 1 - \alpha)$, lengthy computations give $T_{S_\alpha}(\varepsilon, \varepsilon) = \frac{1}{2} - \alpha - \varepsilon + \alpha^2 + \varepsilon^2$, $T_{S_\alpha}(\varepsilon, -\varepsilon) = \frac{1}{2} - \alpha - \varepsilon + \alpha^2 + 3\varepsilon^2/2$ and $T_{S_\alpha}(-\varepsilon, -\varepsilon) = \frac{1}{2} - \alpha + \varepsilon + \alpha^2 + \varepsilon^2/2$ and hence

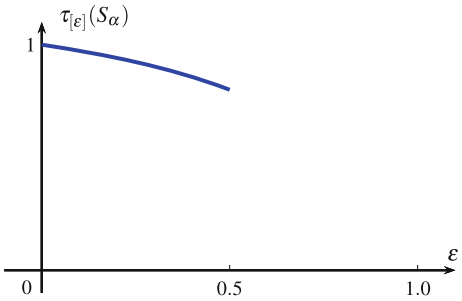


Fig. 2 The bi-conditional local Kendall's tau of S_α

$$\tau_{[\varepsilon]}(S_\alpha) = \frac{4-4\varepsilon}{4-3\varepsilon},$$

which is again independent of α. Its graph is shown in Fig. 2. So $\tau_{[\mathrm{loc}]}(S_\alpha) = \lim_{\varepsilon \to 0^+} \tau_{[\varepsilon]}(S_\alpha) = 1$.

Example 3.3 Let C_θ be the FGM copulas. Then

$$\tau_{[\varepsilon]}(C_\theta) = \frac{2(3-2\varepsilon)^2\varepsilon^2\theta}{9(2-\varepsilon)^2 + (1-\varepsilon)^4(2+\varepsilon)\theta^2} \quad \text{and} \quad \tau_{[\mathrm{loc}]}(C_\theta) = 0.$$

4 Pointwise Kendall's Tau

Definition 4.1 Let X and Y be continuous random variables on a common sample space with marginal distributions F and G, respectively. Let (X_1, Y_1) and (X_2, Y_2) be independent random vectors with identical joint distribution as (X, Y). Let $t \in (0, 1)$ and $r \in (0, \min(t, 1-t))$. Then *a population version of the local Kendall's tau around a point t for X and Y* is defined as

$$\begin{aligned}\tau_{X,Y,r}(t) = {} & P\left[(X_1 - X_2)(Y_1 - Y_2) > 0 \mid -\underline{r} < X_i - F^{-1}(t) < \overline{r},\ \forall i = 1, 2\right] \\ & - P\left[(X_1 - X_2)(Y_1 - Y_2) < 0 \mid -\underline{r} < X_i - F^{-1}(t) < \overline{r},\ \forall i = 1, 2\right]\end{aligned}$$

where $-\underline{r} = \triangle F^{-1}_{-r}(t) = F^{-1}(t-r) - F^{-1}(t)$ and $\overline{r} = \triangle F^{-1}_{r}(t) = F^{-1}(t+r) - F^{-1}(t)$. Note that

$$\begin{aligned}\tau_{X,Y,r} &= \tau_{F(X),G(Y),r} \\ &= P\left[\mathrm{Conc} \mid |FX_i - t| < r,\ \forall i = 1, 2\right] \\ &\quad - P\left[\mathrm{Disc} \mid |FX_i - t| < r,\ \forall i = 1, 2\right]\end{aligned}$$

where

$$\text{Conc} = \{(FX_1 - FX_2)(GY_1 - GY_2) > 0\} \text{ and}$$
$$\text{Disc} = \{(FX_1 - FX_2)(GY_1 - GY_2) < 0\}.$$

The following theorem shows that local Kendall's tau around a point depends only on the copula of continuous random variables X and Y. It can be proved straightforwardly.

Theorem 4.1 *Let X and Y be continuous random variables with copula C. Let t be in $(0, 1)$ and r be in $(0, \min(t, 1 - t))$. Then a population version of the local Kendall's tau around a point t for X and Y is given by*

$$\tau_{X,Y,r}(t) = \frac{1}{r^2} \iint\limits_{(t-r,t+r)\times[0,1]} \left(C(x, y) - \frac{C(t - r, y) + C(t + r, y)}{2} \right) dC(x, y).$$

Since $\tau_{X,Y,r}$ depends only on the copula C, it is also called *the local Kendall's tau around a point of* C and denoted by $\tau_{C,r}$. The *pointwise Kendall's tau of* C *at* t is given by

$$\tau_C(t) = \lim_{r \to 0^+} \tau_{C,r}(t). \tag{6}$$

Similar to its empirical counterpart introduced in [2], $\tau_C(t) = \tau_{X,Y}(t)$ can detect monotonicity at t at least in the case when Y is completely dependent on X.

Theorem 4.2 *If C is the complete dependence copula $C_{U,f(U)}$ for some measure preserving function f and uniform $(0, 1)$ random variable U, then for every continuity point t of f, $\tau_C(t) = \text{sgn}(f'(t))$.*

Proof Since f is measure preserving and continuous on $(t - \delta, t + \delta)$ for some $\delta > 0$, it must be affine on $(t - \delta, t + \delta)$ with slope $m \neq 0$. Assume without loss of generality that $m > 0$. Put $s = f(t)$ so that $f(t + \triangle t) = s + m\triangle t$ if $|\triangle t| < \delta$. By a theorem in [7], $\partial_1 C(x, y) = 0$ for $y < f(x)$ and $\partial_1 C(x, y) = 1$ for $y > f(x)$. So, for $r < \delta$, $x \mapsto C(x, y)$ is constant on $[t - r, t + r]$ whenever $y \geq s + mr$ or $y \leq s - mr$.

Let $r < \delta$. Then

$$\begin{aligned}
\tau_{C,r}(t) &= \frac{1}{r^2} \int_{s-mr}^{s+mr} \int_{t-r}^{t+r} \left(C(x, y) - \frac{C(t - r, y) + C(t + r, y)}{2} \right) C(dx, dy) \\
&= \frac{1}{r^2} \int_{t-r}^{t+r} \Bigg(C(x, s + m(x - t)) \\
&\qquad - \frac{C(t - r, s + m(x - t)) + C(t + r, s + m(x - t))}{2} \Bigg) dx
\end{aligned}$$

where, in the last equality, we use the fact that $\iint_{I\times J} g(x,y)\,dC(x,y) = \int_I g(x, f(x))dx$ if C is supported on the line $y = f(x)$. Since $C(\cdot, s + m(x-t))$ is constant on $[x, t+r]$ and affine of slope 1 on $[t-r, x]$, we have

$$\begin{aligned}\tau_{C,r}(t) &= \frac{1}{2r^2}\int_{t-r}^{t+r} C(x, s+m(x-t)) - C(t-r, s+m(x-t))\,dx \\ &= \frac{1}{2r^2}\int_{t-r}^{t+r} (x-t+r)\,dx = 1.\end{aligned}$$

Hence $\tau_C(t) = 1$. Derivation for the case $m < 0$ gives $\tau_C(t) = -1$.

Acknowledgments The authors thank the anonymous referee for comments and suggestions. The last author would also like to thank the Commission on Higher Education and the Thailand Research Fund for the support through grant no. RSA5680037.

References

1. Bairamov, I., Kotz, S., Kozubowski, T.J.: A new measure of linear local dependence. Statistics **37**(3), 243–258 (2003)
2. Ghosal, S., Sen, A., Van Der Vaart, A.W.: Testing monotonicity of regression. Ann. Stat. **28**(4), 1054–1082 (2000)
3. Hufthammer, K.O.: Some measures of local and global dependence. Master's thesis, University of Bergen (2005)
4. Jones, M.C.: The local dependence function. Biometrika **83**(4), 899–904 (1996)
5. Mikusinski, P., Sherwood, H., Taylor, M.D.: Shuffles of min. Stochastica **13**, 61–74 (1992)
6. Nelsen, R.B.: An Introduction to Copulas, 2nd edn. Springer Series in Statistics. Springer, New York (2006)
7. Ruankong, P., Santiwipanont, T., Sumetkijakan, S.: Shuffles of copulas and a new measure of dependence. J. Math. Anal. Appl. **398**(1), 398–402 (2013)
8. Scarsini, M.: On measures of concordance. Stochastica **8**(3), 201–218 (1984)

Estimation and Prediction Using Belief Functions: Application to Stochastic Frontier Analysis

Orakanya Kanjanatarakul, Nachatchapong Kaewsompong, Songsak Sriboonchitta and Thierry Denœux

Abstract We outline an approach to statistical inference based on belief functions. For estimation, a consonant belief functions is constructed from the likelihood function. For prediction, the method is based on an equation linking the unobserved random quantity to be predicted, to the parameter and some underlying auxiliary variable with known distribution. The approach allows us to compute a predictive belief function that reflects both estimation and random uncertainties. The method is invariant to one-to-one transformations of the parameter and compatible with Bayesian inference, in the sense that it yields the same results when provided with the same information. It does not, however, require the user to provide prior probability distributions. The method is applied to stochastic frontier analysis with cross-sectional data. We demonstrate how predictive belief functions on inefficiencies can be constructed for this problem and used to assess the plausibility of various assertions.

1 Introduction

Many problems in econometrics can be formalized using a parametric model

$$(Y, Z)|x \sim f_{\theta,x}(y, z), \tag{1}$$

where Y and Z are, respectively, observed and unobserved random vectors, x is an observed vector of covariates and $f_{\theta,x}$ is the conditional probability mass or density function of (Y, Z) given $X = x$, assumed to be known up to a parameter vector $\theta \in \Theta$.

O. Kanjanatarakul
Department of Economics, Faculty of Management Sciences,
Chiang Mai Rajabhat University, Chiang Mai, Thailand

N. Kaewsompong · S. Sriboonchitta
Faculty of Economics, Chiang Mai University, Chiang Mai, Thailand

T. Denœux (✉)
Heudiasyc (UMR 7253), Université de Technologie de Compiègne and CNRS,
Compiègne, France
e-mail: thierry.denoeux@utc.fr

V.-N. Huynh et al. (eds.), *Econometrics of Risk*, Studies in Computational Intelligence 583, DOI 10.1007/978-3-319-13449-9_12

For instance, in the standard linear regression model, $Y = (Y_1, \ldots, Y_n)$ is a vector of n independent observations of the response variable, with $Y_i \sim \mathcal{N}(x_i'\beta, \sigma^2)$, $Z = Y_{n+1}$ is an independent random value of the response variable distributed as $\mathcal{N}(x_{n+1}'\beta, \sigma^2)$, $x = (x_1, \ldots, x_{n+1})$ and $\theta = (\beta, \sigma^2)$. Having observed a realization y of Y (and the covariates x), we often wish to determine the unknown quantities in the model, i.e., the parameter θ (assumed to be fixed) and the (yet) unobserved realization z of Z. The former problem is referred to as *estimation* and the latter as *prediction* (or *forecasting*).

These two problems have been addressed in different ways within several theoretical frameworks. The three main theories are frequentist, Bayesian and likelihood-based inference. In the following, we briefly review these three approaches to introduce the motivation for the new method advocated in this paper.

Frequentist methods provide *pre-experimental* measures of the accuracy of statistical evidence. A procedure (for computing, e.g., a confidence or prediction interval) is decided before observing the data and its long-run behavior is determined by averaging over the whole sample space, assuming it is repeatedly applied to an infinite number of samples drawn from the same population. It has long been recognized that such an approach, although widely used, does not provide a reliable measure of the strength of evidence provided by specific data. The following simple example, taken from [6], illustrates this fact. Suppose X_1 and X_2 are iid with probability mass function

$$\mathbb{P}_\theta(X_i = \theta - 1) = \mathbb{P}_\theta(X_i = \theta + 1) = \frac{1}{2}, \quad i = 1, 2, \tag{2}$$

where $\theta \in \mathbb{R}$ is an unknown parameter. Consider the following confidence set for θ,

$$C(X_1, X_2) = \begin{cases} \frac{1}{2}(X_1 + X_2) & \text{if } X_1 \neq X_2 \\ X_1 - 1 & \text{otherwise.} \end{cases} \tag{3}$$

It is a minimum length confidence interval at level 75 %. Now, let (x_1, x_2) be a given realization of the random sample (X_1, X_2). If $x_1 \neq x_2$, we know for sure that $\theta = (x_1 + x_2)/2$ and it would be absurd to take 75 % as a measure of the strength of the statistical evidence. If $x_1 = x_2$, we know for sure that θ is either $x_1 - 1$ or $x_1 + 1$, but we have no reason to favor any of these two hypotheses in particular. Again, it would make no sense to claim that the evidence support the hypothesis $\theta = x_1 - 1$ with 75 % confidence. Although frequentist procedures do provide usable results in many cases, the above example shows that they are based on a questionable logic if they are used to assess the reliable of given statistical evidence, as they usually are. Moreover, on a more practical side, confidence and prediction intervals are often based on asymptotic assumptions and their true coverage probability, assuming it is of interest, may be quite different from the nominal one for small sample sizes.

The other main approach to statistical inference is the Bayesian approach, which, in contrast to the previous approach, implements some form of *post-experimental*

reasoning. Here, all quantities, including parameters, are treated as random variables, and the inference aims at determining the probability distribution of unknown quantities, given observed ones. With the notations introduced above, the estimation and prediction problems are to determine the posterior distributions of, respectively, θ and Z, given x and y. Of course, this is only possible if one provides a prior probability distribution $\pi(\theta)$ on θ, which is the main issue with this approach. There has been a long-standing debate among statisticians about the possibility to determine such a prior when the experimenter does not know anything about the parameter before observing the data. For lack of space, we cannot reproduce all the arguments of this debate here. Our personal view is that no probability distribution is truly non-informative, which weakens the conclusions of Bayesian inference in situations where no well-justified prior can be provided.

The last classical approach to inference is grounded in the likelihood principle (LP), which states that all the information provided by the observations about the parameter is contained in the likelihood function. A complete exposition of the likelihood-based approach to statistical inference can be found in the monographs [6, 8] (see also the seminal paper of Barnard et al. [3]). Birnbaum [7] showed that the LP can be derived from the two generally accepted principles of sufficiency and conditionality. Frequentist inference does not comply with the LP, as confidence intervals and significance tests depend not only on the likelihood function, but also on the sample space. Bayesian statisticians accept the LP, but claim that the likelihood function does not make sense in itself and needs to be multiplied by a prior probability distribution to form the posterior distribution of the parameter given the data. The reader is referred to Refs. [6, 8] for thorough discussions on this topic. Most of the literature on likelihood-based inference deals with estimation. Several authors have attempted to address the prediction problem using the notion of "predictive likelihood" [4, 8, 18]. For instance, the predictive profile likelihood is defined by $L_x(z) = \sup_\theta f_{\theta,x}(y, z)$. However, this notion is quite different conceptually from the standard notion of likelihood and, to some extent, arbitrary. While it does have interesting theoretical properties [18], its use poses some practical difficulties [6, p.39].

The method described in this paper builds upon the likelihood-based approach by seeing the likelihood function as describing the plausibility of each possible value of the parameter, in the sense of the Dempster-Shafer theory of belief functions [9, 10, 20]. This approach of statistical inference was first proposed by Shafer [20] and was later investigated by several authors (see, e.g., [1, 23]). It was recently justified by Denœux in [11] and extended to prediction in [16, 17]. In this paper, we provide a general introduction to estimation and prediction using belief functions and demonstrate the application of this inference framework to the stochastic frontier model. In this model, the determination of the production frontier and disturbance parameters is an estimation problem, whereas the determination of the inefficiency terms is a prediction problem. We will show, in particular, how this method makes it possible to quantify both estimation uncertainty and random uncertainty, and to evaluate the plausibility of various hypothesis about both the production frontier and the efficiencies.

The rest of this paper is organized as follows. The general framework for inference and prediction will first be recalled in Sect. 2. This framework will be particularized to the stochastic frontier model in Sects. 3 and 4 will conclude the paper.

2 Inference and Prediction Using Belief Functions

Basic knowledge of the theory of belief functions will be assumed throughout this paper. A complete exposition in the finite case can be found in Shafer's book [20]. The reader is referred to [5] for a quick introduction on those aspects of this theory needed for statistical inference. In this section, the definition of a belief function from the likelihood function and the general prediction method introduced in [16] will be recalled in Sects. 2.1 and 2.2, respectively.

2.1 Inference

Let $f_{\theta,x}(y)$ be the marginal probability mass or density function of the observed data Y given x. In the following, the covariates (if any) will be assumed to be fixed, so that the notation $f_{\theta,x}(y)$ can be simplified to $f_\theta(y)$. Statistical inference has been addressed in the belief function framework by many authors, starting from Dempster's seminal work [9]. In [20], Shafer proposed, on intuitive grounds, a more direct approach in which a belief function Bel_y^Θ on Θ is built from the likelihood function. This approach was further elaborated by Wasserman [23] and discussed by Aickin [1], among others. It was recently justified by Denœux in [11], from three basic principles: the likelihood principle, compatibility with Bayesian inference and the least commitment principle [21]. The least committed belief function verifying the first two principles, according to the commonality ordering [12] is the consonant belief function Bel_y^Θ defined by the contour function

$$pl_y(\theta) = \frac{L_y(\theta)}{\sup_{\theta' \in \Theta} L_y(\theta')}, \tag{4}$$

where $L_y(\theta) = f_\theta(y)$ is the likelihood function. The quantity $pl_y(\theta)$ is interpreted as the plausibility that the true value of the parameter is θ. The corresponding plausibility and belief functions can be computed from pl_y as:

$$Pl_y^\Theta(A) = \sup_{\theta \in A} pl_y(\theta), \tag{5a}$$

$$Bel_y^\Theta(A) = 1 - \sup_{\theta \notin A} pl_y(\theta), \tag{5b}$$

for all $A \subseteq \Theta$. The focal sets of Bel_y^Θ are the levels sets of $pl_y(\theta)$ defined as follows:

$$\Gamma_y(\omega) = \{\theta \in \Theta | pl_y(\theta) \geq \omega\}, \tag{6}$$

for $\omega \in [0, 1]$. These sets may be called plausibility regions and can be interpreted as sets of parameter values whose plausibility is greater than some threshold ω. When ω is a random variable with a continuous distribution $\mathscr{U}([0, 1])$, $\Gamma_y(\omega)$ becomes a random set equivalent to the belief function Bel_y^{Θ}, in the sense that

$$Bel_y^{\Theta}(A) = \mathbb{P}_\omega(\Gamma_y(\omega) \subseteq A) \tag{7a}$$

$$Pl_y^{\Theta}(A) = \mathbb{P}_\omega(\Gamma_y(\omega) \cap A \neq \emptyset), \tag{7b}$$

for all $A \subseteq \Theta$ such that the above expressions are well-defined.

Example 1 Let us consider the case where $Y = (Y_1, \ldots, Y_n)$ is an i.i.d. sample from a normal distribution $\mathscr{N}(\theta, 1)$. The contour function on θ given a realization y of Y is

$$pl_y(\theta) = \frac{(2\pi)^{-n/2} \exp\left(-\frac{1}{2}\sum_{i=1}^n (y_i - \theta)^2\right)}{(2\pi)^{-n/2} \exp\left(-\frac{1}{2}\sum_{i=1}^n (y_i - \overline{y})^2\right)} \tag{8a}$$

$$= \exp\left(-\frac{n}{2}(\theta - \overline{y})^2\right), \tag{8b}$$

where $\overline{y}$ is the sample mean. The plausibility and belief that θ does not exceed some value t are given by the upper and lower cumulative distribution functions (cdfs) defined, respectively, as

$$Pl_y(\theta \leq t) = \sup_{\theta \leq t} pl_x(\theta) \tag{9a}$$

$$= \begin{cases} \exp\left(-\frac{n}{2}(t - \overline{y})^2\right) & \text{if } t \leq \overline{y} \\ 1 & \text{otherwise} \end{cases} \tag{9b}$$

and

$$Bel_y(\theta \leq t) = 1 - \sup_{\theta > t} pl_x(\theta) \tag{10a}$$

$$= \begin{cases} 0 & \text{if } t \leq \overline{y} \\ 1 - \exp\left(-\frac{n}{2}(t - \overline{y})^2\right) & \text{otherwise.} \end{cases} \tag{10b}$$

The focals sets (6) are closed intervals

$$\Gamma_y(\omega) = \left[\overline{y} - \sqrt{\frac{-2\ln\omega}{n}}, \overline{y} + \sqrt{\frac{-2\ln\omega}{n}}\right]. \tag{11}$$

When ω has a uniform distribution on [0, 1], $\Gamma_y(\omega)$ is a closed random interval. The cdfs of its lower and upper bounds are equal, respectively, to the lower and upper cdfs (10a, 10b) and (9a, 9b). □

2.2 Prediction

The prediction problem can be defined as follows: having observed the realization y of Y with distribution $f_\theta(y)$, we wish to make statements about some yet unobserved data $Z \in \mathbb{Z}$ whose conditional distribution $f_{y,\theta}(z)$ given $Y = y$ also depends on θ. The uncertainty on Z has two sources: (1) the randomness of the generation mechanism of Z given θ and y and (2) the estimation uncertainty on θ. In the approach outlined here, the latter uncertainty is represented by the belief function Bel_y^Θ on θ obtained by the approach described in the previous section. The random generation mechanism for Z can be represented by a sampling model such as the one used by Dempster [9] for inference. In this model, the new data Z is expressed as a function of the parameter θ and an unobserved auxiliary random variable ξ with known probability distribution independent of θ:

$$Z = \varphi(\theta, \xi), \tag{12}$$

where φ is defined in such a way that the distribution of Z for fixed θ is $f_{y,\theta}(z)$.

When Z is a real random variable, a canonical model of the form (12) can be obtained as $Z = F_{y,\theta}^{-1}(\xi)$, where $F_{y,\theta}$ is the conditional cumulative distribution function (cdf) of Z given $Y = y$, $F_{y,\theta}^{-1}$ is its generalized inverse and ξ has a continuous uniform distribution in [0, 1]. This canonical model can be extended to the case where Z is a random vector. For instance, assume that Z is a two-dimensional random vector (Z_1, Z_2). We can write

$$Z_1 = F_{y,\theta}^{-1}(\xi_1) \tag{13a}$$

$$Z_2 = F_{y,\theta,Z_1}^{-1}(\xi_2), \tag{13b}$$

where $F_{y,\theta}$ is the conditional cdf of Z_1 given $Y = y$, F_{y,θ,Z_1} is the conditional cdf of Z_2 given $Y = y$ and Z_1 and $\xi = (\xi_1, \xi_2)$ has a uniform distribution in $[0, 1]^2$.

Equation (12) gives us the distribution of Z when θ is known. If we only know that $\theta \in \Gamma_y(\omega)$ and the value of ξ, we can assert that Z is in the set $\varphi(\Gamma_y(\omega), \xi)$. As ω and ξ are not observed but have a joint uniform distribution on $[0, 1]^2$, the set $\varphi(\Gamma_y(\omega), \xi)$ is a random set. It induces belief and plausibility functions defined as

$$Bel_y(Z \in A) = \mathbb{P}_{\omega,\xi}\left(\varphi(\Gamma_y(\omega), \xi) \subseteq A\right), \tag{14a}$$

$$Pl_y(Z \in A) = \mathbb{P}_{\omega,\xi}\left(\varphi(\Gamma_y(\omega), \xi) \cap A \neq \emptyset\right), \tag{14b}$$

for any $A \subseteq \mathbb{Z}$.

Example 2 Continuing Example 1, let us assume that $Z \sim \mathcal{N}(\theta, 1)$ is a yet unobserved normal random variable independent of Y. It can be written as

$$Z = \theta + \Phi^{-1}(\xi), \tag{15}$$

where Φ is the cdf of the standard normal distribution. The random set $\varphi(\Gamma_y(\omega), \xi)$ is then the random closed interval

$$\varphi(\Gamma_y(\omega), \xi) = \left[\overline{y} - \sqrt{\frac{-2 \ln \omega}{n}} + \Phi^{-1}(\xi), \overline{y} + \sqrt{\frac{-2 \ln \omega}{n}} + \Phi^{-1}(\xi) \right]. \tag{16}$$

Expressions (14a, 14b) for the belief and plausibility of any assertion about Z can be approximated by Monte Carlo simulation [16]. □

As remarked by Bjornstad [8], a prediction method should have at least two fundamental properties: it should be invariant to any one-to-one reparametrization of the model and it should be asymptotically consistent, in a precise sense to be defined. An additional property that seems desirable is compatibility with Bayesian inference, in the sense that it should yield the same result as the Bayesian approach when a prior distribution on the parameter is provided. Our method possesses these three properties. Parameter invariance follows from the fact that it is based on the likelihood function; compatibility with Bayes is discussed at length in [16] and consistency will be studied in greater detail in a forthcoming paper.

3 Application to Stochastic Frontier Analysis

In this section, we apply the above estimation and prediction framework to the stochastic frontier model (SFM). To keep the emphasis on fundamental principles of inference, only the simplest case of cross-sectional data will be considered. The model as well as the inference method will be in introduced in Sect. 3.1 and an illustration with simulated data will be presented in Sect. 3.2.

3.1 Model and Inference

The SFM [2] defines a production relationship between a p-dimensional input vector $\mathbf{x}_i$ and output Y_i of each production unit i of the form

$$\ln Y_i = \boldsymbol{\beta}' \ln \mathbf{x}_i + V_i - U_i, \tag{17}$$

where $\boldsymbol{\beta}$ is a vector of coefficients, V_i is an error term generally assumed to have a normal distribution $\mathcal{N}(0, \sigma_v^2)$ and U_i is a positive inefficiency term. Usual models for

U_i are the half-normal distribution $|\mathcal{N}(0, \sigma_u^2)|$ (i.e., the distribution of the absolute value of a normal variable) and the exponential distribution. The SFM is thus a linear regression model with asymmetric disturbances $\varepsilon_i = V_i - U_i$. The inefficiency terms U_i are not observed but are of particular interest in this setting.

Assuming U_i to have a half-normal distribution, let $\lambda = \sigma_u/\sigma_v$ and $\sigma^2 = \sigma_u^2 + \sigma_v^2$ be new parameters to be used in place of σ_u^2 and σ_v^2. Although the variance of U_i is not σ_u^2 but $(1 - 2/\pi)\sigma_u^2$, λ has an intuitive interpretation as the relative variability of the two sources of error that distinguish firms from one another [2]. Using the notations defined in Sect. 1, we have $Y = (Y_1, \ldots, Y_n)$, $Z = (U_1, \ldots, U_n)$ and $\theta = (\boldsymbol{\beta}, \sigma, \lambda)$. The determination of the inefficiency terms is thus a prediction problem.

3.1.1 Parameter Estimation

Assuming the two error components U_i and V_i to be independent, the log-likelihood function is [14, p.540]

$$\ln L_y(\theta) = -n \ln \sigma + \frac{n}{2} \log \frac{2}{\pi} - \frac{1}{2} \sum_{i=1}^{n} \left(\frac{\varepsilon_i}{\sigma} \right)^2 + \sum_{i=1}^{n} \ln \Phi \left(-\frac{\varepsilon_i \lambda}{\sigma} \right). \tag{18}$$

The maximum likelihood estimate (MLE) $\widehat{\theta}$ can be found using an iterative nonlinear optimization procedure. Parameter β may be initialized by the least squares estimate, which is unbiased and consistent (except for the constant term) [14]. However, it may be wise to restart the procedure from several randomly chosen initial states, as the log-likelihood function may have several maxima for this problem. Once $\widehat{\theta}$ has been found, the contour function (4) can be computed. The marginal contour function for any subset of parameters is the relative profile likelihood function. For instance, the marginal contour function of λ is

$$pl_y(\lambda) = \sup_{\boldsymbol{\beta}, \sigma} pl_y(\theta). \tag{19}$$

3.1.2 Prediction

The main purpose of stochastic frontier analysis is the determination of the inefficiency terms u_i, which are not observed. The usual approach is to approximate u_i by $\mathbb{E}(U_i|\varepsilon_i)$, which is itself estimated by plugging in the MLEs and by replacing ε_i by the residuals $\widehat{\varepsilon}_i$. The main result is due to Jondrow et al. [15], who showed that the conditional distribution of U given ε_i, in the half-normal case, is that of a normal $\mathcal{N}(\mu_*, \sigma_*^2)$ variable truncated at zero, with

$$\mu_* = -\frac{\sigma_u^2 \varepsilon_i}{\sigma^2} = -\frac{\varepsilon_i \lambda^2}{1+\lambda^2} \tag{20a}$$

$$\sigma_* = \frac{\sigma_u \sigma_v}{\sigma} = \frac{\lambda \sigma}{1+\lambda^2}. \tag{20b}$$

The conditional expectation of U_i given ε_i is

$$\mathbb{E}(U_i|\varepsilon_i) = \frac{\lambda\sigma}{1+\lambda^2}\left[\frac{\phi(\lambda\varepsilon_i/\sigma)}{1-\Phi(\lambda\varepsilon_i/\sigma)} - \frac{\lambda\varepsilon_i}{\sigma}\right], \tag{21}$$

where ϕ and Φ are, respectively, the pdf and cdf of the standard normal distribution. As noted by Jondrow et al. [15], when replacing the unknown parameter values by their MLEs, we do not take into account uncertainty due to sampling variability. While this uncertainty becomes negligible when the sample size tends to infinity, it certainly is not when the sample is of small or moderate size.

To implement the approach outlined in Sect. 2.2 for this problem, we may write the cdf of U_i as

$$F(u) = \frac{\Phi[(u-\mu_*)/\sigma_*] - \Phi(-\mu_*/\sigma_*)}{1-\Phi(-\mu_*/\sigma_*)} \mathbb{1}_{[0,+\infty)}(u). \tag{22}$$

Let $\xi_i = F(U_i)$, which has a uniform distribution $\mathscr{U}([0,1])$. Solving the equation $\xi_i = F(U_i)$ for U_i, we get

$$U_i = \mu_* + \sigma_* \Phi^{-1}\left[\xi_i\left(1-\Phi\left(-\frac{\mu_*}{\sigma_*}\right)\right) + \Phi\left(-\frac{\mu_*}{\sigma_*}\right)\right]. \tag{23}$$

Replacing μ_* and σ_* by their expressions as functions of the parameters, we have

$$U_i = \varphi(\theta, \xi_i) = \frac{\lambda}{1+\lambda^2}\left\{-\varepsilon_i\lambda + \sigma\Phi^{-1}\left[\xi_i + \Phi\left(\frac{\varepsilon_i\lambda}{\sigma}\right)(1-\xi_i)\right]\right\} \tag{24}$$

with $\varepsilon_i = \ln y_i - \boldsymbol{\beta}' \ln \mathbf{x}_i$, which gives us an equation of the same form as (12), relating the unobserved random variable U_i to the parameters and the auxiliary variable ξ_i.

To approximate the belief function on $Z = (U_1, \ldots, U_n)$, we may use the Monte Carlo method described in [16]. More specifically, we randomly generate N $n+1$-tuples $(\omega^{(j)}, \xi_1^{(j)}, \ldots, \xi_1^{(n)})$ for $j = 1, \ldots, N$ uniformly in $[0,1]^{n+1}$. For $i = 1$ to n and $j = 1$ to N, we compute the minimum and the maximum of $\varphi(\theta, \xi_i^{(j)})$ w.r.t. θ under the constraint

$$pl_y(\theta) \geq \omega^{(j)}. \tag{25}$$

Let $[\underline{u}_i^{(j)}, \overline{u}_i^{(j)}]$ be the resulting interval. The belief and plausibility of any statement $Z \in A$ for $A \subset \mathbb{R}^n$ as defined by (14a, 14b) can be approximated by

$$Bel_y(Z \in A) \approx \frac{1}{N} \# \left\{ j \in \{1, \ldots, N\} | [\underline{u}_1^{(j)}, \overline{u}_1^{(j)}] \times \cdots \times [\underline{u}_n^{(j)}, \overline{u}_n^{(j)}] \subseteq A \right\}, \tag{26a}$$

$$Pl_y(Z \in A) \approx \frac{1}{N} \# \left\{ j \in \{1, \ldots, N\} | [\underline{u}_1^{(j)}, \overline{u}_1^{(j)}] \times \cdots \times [\underline{u}_n^{(j)}, \overline{u}_n^{(j)}] \cap A \neq \emptyset \right\}, \tag{26b}$$

where # denotes cardinality. We can also approximate the belief function on any linear combination $\sum_{i=1}^n \alpha_i u_i$ by applying the same transformation to the intervals $[\underline{u}_i^{(j)}, \overline{u}_i^{(j)}]$, using interval arithmetics. For example, the belief and plausibility of statements of the form $u_i - u_k \leq c$ can be approximated as follows:

$$Bel_y(u_i - u_k \leq c) \approx \frac{1}{N} \# \left\{ j \in \{1, \ldots, N\} | \overline{u}_i^{(j)} - \underline{u}_k^{(j)} \leq c \right\}, \tag{27a}$$

$$Pl_y(u_i - u_k \leq c) \approx \frac{1}{N} \# \left\{ j \in \{1, \ldots, N\} | \underline{u}_i^{(j)} - \overline{u}_k^{(j)} \leq c \right\}. \tag{27b}$$

3.2 Simulation Experiments

To illustrate the behavior of our method, we simulated data from model (17) with $p = 1$, $\boldsymbol{\beta} = (1, 0.5)'$, $\sigma_v = 0.175$ and $\sigma_u = 0.3$. We thus have, for this model, $\lambda = 1.7143$ and $\sigma = 0.3473$. Figure 1 displays the marginal contour functions of β_0, β_1, σ and λ for a simulated sample of size $n = 100$. These plots show graphically the plausibility of any assertion of the form $\theta_j = \theta_{j0}$. For instance, we can see from Fig. 1d that the plausibility of the assertion $\lambda = 0$ is around 0.6: consequently, the hypothesis that inefficiencies are all equal to zero is quite plausible, given the data.

Figures 2 and 3 show the true and estimated inefficiencies for 20 individuals in the above simulated sample of size $n = 100$ and in a simulated sample of size $n = 1{,}000$, respectively. For the belief function estimation, we give the quantile intervals for $\alpha = 5\,\%$ and $\alpha = 25\,\%$. The lower bound of the quantile interval [16] is the α quantile of the lower bounds $\underline{u}_i^{(j)}$ of the prediction intervals, while the upper bound is the $1 - \alpha$ quantile of the upper bounds $\overline{u}_i^{(j)}$. The larger intervals in the case $n = 100$ reflect the higher estimation uncertainty.

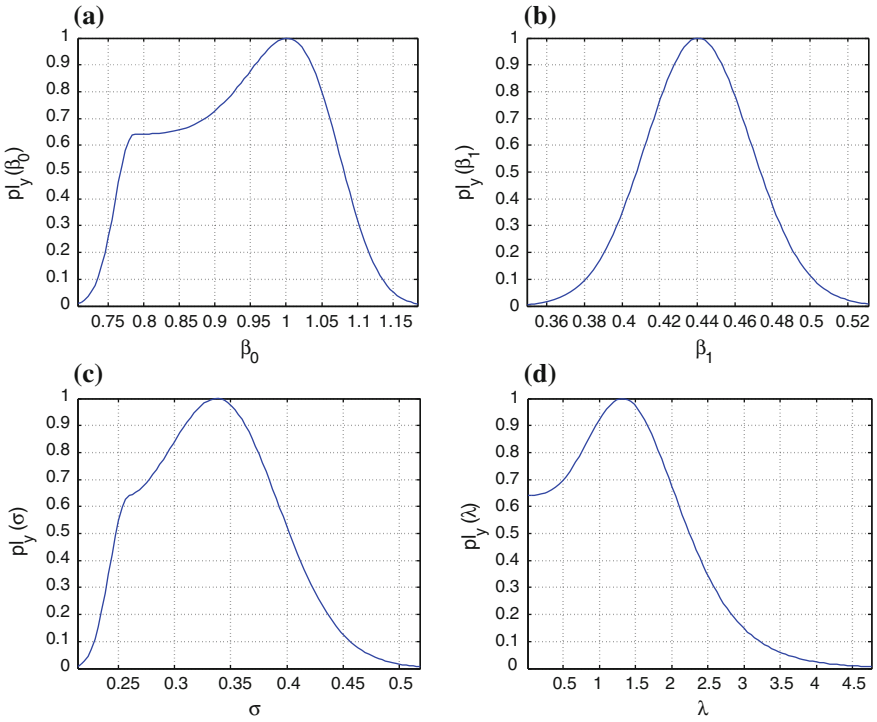


Fig. 1 Marginal contour functions for a simulated sample of size $n = 100$

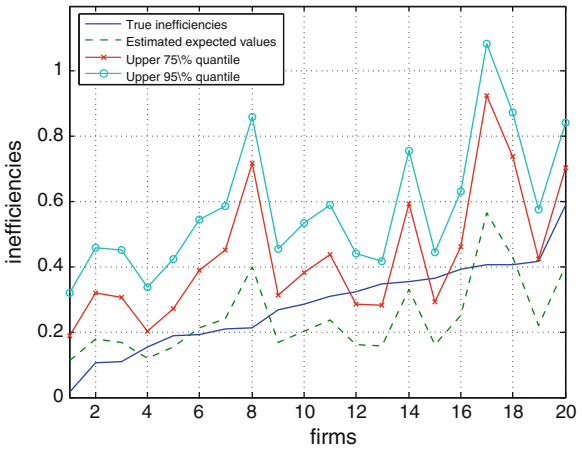


Fig. 2 True and predicted inefficiencies for 20 individuals in a simulated sample of size $n = 100$

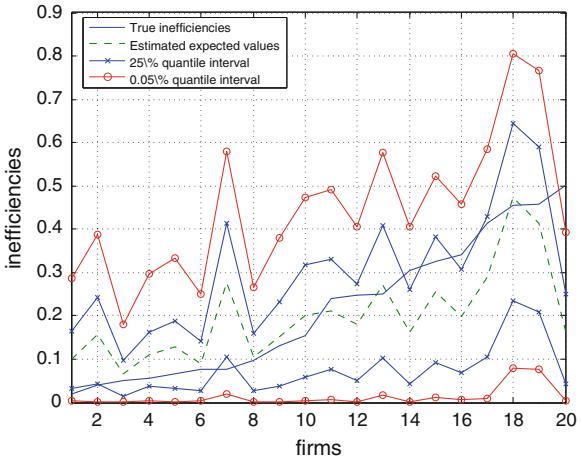


Fig. 3 True and predicted inefficiencies for 20 individuals in a simulated sample of size $n = 1{,}000$

4 Conclusions

We have shown how the estimation and prediction problems may be solved in the belief function framework, and illustrated these solutions in the case of the stochastic frontier model with cross-sectional observation. In the case of this model, the estimation problem concerns the determination of the model parameters describing the production frontier and the distributions of the noise and inefficiency terms, while the prediction problem consists in the determination of the unobserved inefficiency terms, which are of primary interest in this analysis. In our approach, uncertainties about the parameters and the inefficiencies are both modeled by belief functions induced by random sets. In particular, the random set formulation allows us to approximate the belief or plausibility of any assertion about the inefficiencies, using Monte Carlo simulation.

We can remark that parameters and realizations of random variables (here, inefficiencies) are treated differently in our approach, whereas there are not in Bayesian inference. In particular, the likelihood-based belief functions in the parameter space are consonant, whereas predictive belief functions are not. This difference is not due to conceptual differences between parameters and observations, which are just considered here as unknown quantities. It is due to different natures of the evidence from which the belief functions are constructed. In the former case, the evidence consists of observations that provide information on parameters governing a random process. In the latter case, evidence about the data generating process provides information about unobserved observations generated from that process.

The evidential approach to estimation and prediction outlined in this paper is invariant to one-to-one transformations of the parameters and compatible with Bayesian inference, in the sense that it yields the same result when provided with the

same initial information. It is, however, more general, as it does not require the user to supply prior probability distributions. It is also easily implemented and does not require asymptotic assumptions, which makes it readily applicable to a wide range of econometric models.

The preliminary results reported in this paper need to be completed in several ways. First, a detailed comparison, based on underlying principles, with alternative approaches such as, e.g., the empirical Bayes method [19] or imprecise probabilities [22] remains to be performed. Secondly, it would be interesting to study experimentally how users interpret the results of the belief function analysis to make decisions in real-world situations. Finally, theoretical properties of our method, such as asymptotic consistency, are currently being studied.

References

1. Aickin, M.: Connecting Dempster-Shafer belief functions with likelihood-based inference. Synthese **123**, 347–364 (2000)
2. Aigner, D.J., Lovell, C.A.K., Schmidt, P.: Formulation and estimation of stochastic frontier production function models. J. Econom. **6**, 21–37 (1977)
3. Barnard, G.A., Jenkins, G.M., Winsten, C.B.: Likelihood inference and time series. J. R. Stat. Soc. **125**(3), 321–372 (1962)
4. Bayarri, M.J., DeGroot, M.H.: Difficulties and ambiguities in the definition of a likelihood function. J. Ital. Stat. Soc. **1**(1), 1–15 (1992)
5. Ben Abdallah, N., Mouhous Voyneau, N., Denœux, T.: Combining statistical and expert evidence using belief functions: application to centennial sea level estimation taking into account climate change. Int. J. Approx. Reason. **55**, 341–354 (2014)
6. Berger, J.O., Wolpert, R.L.: The likelihood principle: A Review, Generalizations, and Statistical Implications, 2nd edn. Lecture Notes-Monograph Series, vol 6. Institute of Mathematical Statistics, Hayward (1988)
7. Birnbaum, A.: On the foundations of statistical inference. J. Am. Stat. Assoc. **57**(298), 269–306 (1962)
8. Bjornstad, J.F.: Predictive likelihood: a review. Stat. Sci. **5**(2), 242–254, 05 (1990)
9. Dempster, A.P.: Upper and lower probabilities induced by a multivalued mapping. Ann. Math. Stat. **38**, 325–339 (1967)
10. Dempster, A.P.: A generalization of Bayesian inference (with discussion). J. R. Stat. Soc. B **30**, 205–247 (1968)
11. Denœux, T.: Likelihood-based belief function: justification and some extensions to low-quality data. Int. J. Approx. Reason. **55**(7), 1535–1547 (2014)
12. Dubois, D., Prade, H.: A set-theoretic view of belief functions: logical operations and approximations by fuzzy sets. Int. J. Gen. Syst. **12**(3), 193–226 (1986)
13. Edwards, A.W.F.: Likelihood (expanded edn.). The John Hopkins University Press, Baltimore (1992)
14. Greene, W.H.: Econometric Analysis, 6th edn. Prentice Hall, Upper Saddle River (2008)
15. Jondrow, J., Lovell, C.A.K., Materov, I.S., Schmidt, P.: On the estimation of technical efficiency in the stochastic production function model. J. Econom. **19**, 233–238 (1982)
16. Kanjanatarakul, O., Sriboonchitta, S., Denœux, T.: Forecasting using belief functions: an application to marketing econometrics. Int. J. Approx. Reason. **55**(5), 1113–1128 (2014)
17. Kanjanatarakul, O., Lertpongpiroon P., Singkharat S. and Sriboonchitta, S.: Econometric forecasting using linear regression and belief functions, In: Cuzzolin, F. (ed.), Belief Functions: Theory and Applications. Oxford, Springer, LNAI 8764, pp. 304–312 (2014)

18. Mathiasen, P.E.: Prediction functions. Scand. J. Stat. **6**(1), 1–21 (1979)
19. Robbins, H.: An empirical Bayes approach to statistics. In: Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability. Volume 1: Contributions to the Theory of Statistics, pp. 157–163 (1956)
20. Shafer, G.: A Mathematical Theory of Evidence. Princeton University Press, Princeton (1976)
21. Smets, Ph: Belief functions: the disjunctive rule of combination and the generalized Bayesian theorem. Int. J. Approx. Reason. **9**, 1–35 (1993)
22. Walley, P.: Statistical Reasoning with Imprecise Probabilities. Chapman and Hall, London (1991)
23. Wasserman, L.A.: Belief functions and statistical evidence. Can. J. Stat. **18**(3), 183–196 (1990)

The Classifier Chain Generalized Maximum Entropy Model for Multi-label Choice Problems

Supanika Leurcharusmee, Jirakom Sirisrisakulchai, Songsak Sriboonchitta and Thierry Denœux

Abstract Multi-label classification can be applied to study empirically discrete choice problems, in which each individual chooses more than one alternative. We applied the Classifier Chain (CC) method to transform the Generalized Maximum Entropy (GME) choice model from a single-label model to a multi-label model. The contribution of our CC-GME model lies in the advantages of both the GME and CC models. Specifically, the GME model can not only predict each individual's choice, but also robustly estimate model parameters that describe factors determining his or her choices. The CC model is a problem transformation method that allows the decision on each alternative to be correlated. We used Monte-Carlo simulations and occupational hazard data to compare the CC-GME model with other selected methodologies for multi-label problems using the Hamming Loss, Accuracy, Precision and Recall measures. The results confirm the robustness of GME estimates with respect to relevant parameters regardless of the true error distributions. Moreover, the CC method outperforms other methods, indicating that the incorporation of the information on dependence patterns among alternatives can improve prediction performance.

1 Introduction

The *discrete choice problem* describes how an individual chooses an alternative from $M \geq 2$ available ones. Empirically, the problem is similar to the *single-label classification problem*, in which objects are classified into M classes. However, in many situations, we observe an individual choosing more than one alternative simultaneously. This problem is then empirically equivalent to the *multi-label classification* problem, in which one object can be associated with a subset of classes. In this study, we extend existing single-label choice models to multi-label choice models.

S. Leurcharusmee (✉) · J. Sirisrisakulchai · S. Sriboonchitta
Faculty of Economics, Chiang Mai University, Chiang Mai, Thailand
e-mail: supanika.econ.cmu@gmail.com

T. Denœux
Université de Technologie de Compiègne CNRS, UNR 7253 Heudiasyc, cedex, France

V.-N. Huynh et al. (eds.), *Econometrics of Risk*, Studies in Computational Intelligence 583, DOI 10.1007/978-3-319-13449-9_13

Since the development of the random utility model, which explains individuals' decision making process, the parameters in the empirical models can be linked to those in the utility functions. The knowledge of the model parameters can contribute to behavior explanation and policy implications. Consequently, the objectives of our multi-label choice model are not only to predict a set of alternatives that each individual chooses, but also to estimate the model parameters that describe factors determining each individual's decisions.

Common methods to study discrete choice problems are the Logit and Probit models [16]. These models are limited in the sense that, being likelihood-based, they require distributional assumptions for the errors. [4, 15] introduced the Maximum Entropy (ME) model for discrete choice problems. [5] added the error components to the model and extended it to the Generalized Maximum Entropy (GME) model for multinomial choice problem to improve efficiency. The traditional discrete choice models are for single-label classification. There are a few Logit and Probit models that were developed to explain the multi-label choice problem in which each individual purchases a bundle of products. As discussed in [2], commonly used models are the Label Powerset model with multinomial Logit and Probit estimation and the multivariate Probit or Logit models [1–3]. Although both models allow each individual to choose more than one alternatives, none of them can cope with large choice sets.

Existing methodologies to analyze the multi-label classification problem in computer science follow two main approaches, referred to as *problem transformation* and *algorithm adaptation* [17]. The strategy of the former approach is to transform the multi-label problem into single-label one in order to apply traditional classification methods. Problem transformation methods include Binary Relevance, Label Powerset, Random k-labelsets, Classifier Chains, Pruned Sets, Ensemble of Classifier Chains and Ensemble of Pruned Sets [7, 13]. The algorithm adaptation approach, in contrast, tackles the multi-label problem directly. Algorithm adaptation methods include Multi-label k-Nearest Neighbors, Back-Propagation Multi-label Learning and Decision Trees [7, 13].

Since the objective of this study is to extend the single-label choice model to multi-label choice, we focus on the problem transformation approach. As discussed in [7], the problem transformation approach is generally simpler, but it has a disadvantage of not incorporating the dependence among alternatives. However, this is not true for the Classifier Chain (CC) method, which can capture the dependence pattern among alternatives. Since the choices that each individual makes are usually correlated, this study focuses on the CC method. As for the base single-label choice model, the Logit, Probit and GME models were all developed with a main objective to estimate model parameters that describe factors determining each individual's decisions. However, the GME estimates are robust to distributional assumptions. In addition, the GME model can generally estimate under-determined problems most efficiently. In other words, the GME method yields the estimated parameters with the smallest possible variances [6]. Therefore, to robustly estimate all relevant parameters and capture the dependence pattern among alternatives, we propose the CC-GME model.

For the experimental part of this study, we used Monte-Carlo simulations to compare the CC method with the Binary Relevance (BR) and Label Powerset (LP)

methods and we compare the GME method with the Logit and Probit methods. Specifically, we empirically assessed the performances CC-GME model against CC-Logit, CC-Probit, BR-GME, BR-Logit, BR-Probit, LP-GME, LP-Logit and LP-Probit models. To test the robustness of the estimations, we applied all the methods to three simulated datasets with normal, logistic and uniform errors. Moreover, we also applied all the methods to a real dataset to explain factors determining the set of occupational hazards that each individual faces. Performance measures used in this study include Hamming Loss, Accuracy, Precision and Recall [7, 13, 18]. The results show that the forecasting performances are more sensitive to the choice of the problem transformation method than to the choice of single-label estimation methods. That is, the CC model outperformed the BR and LP models with respect to all evaluation measures expect the Precision. For the parameter estimates, the GME based methods yielded smaller Mean Squared Error (MSE) than the Logit and Probit base methods.

This paper is organized as follows. The original GME model for single-label choice model is recalled in Sect. 2 and the multi-label CC-GME model is introduced in Sect. 3. In Sect. 4, the CC-GME model is evaluated using Monte-Carlo simulations. Section 5 provides an empirical example using occupational hazard data. Finally, Sect. 6 presents our conclusions and remarks.

2 The Single-Label GME Model

The concept of entropy was introduced by [14] to measure the uncertainty of a set of events. The *Shannon entropy function* is $H(p) = \sum_j p_j \log(p_j)$, where p_j is the probability of observing outcome j [8, 9]. Proposed the Maximum Entropy (ME) principle, stating that the probability distribution that best represents the data or available information is the one with the largest entropy. From the ME principle, [4, 15] developed the ME model for discrete choice problems [5]. Added error components to the model and extended it to the GME model for discrete choice problems.

Consider a problem in which each of N individuals chooses his or her most preferred choice from M alternatives. From the data, we observe dummy variables y_{ij} which equal 1 if individual i chooses alternative j and 0 otherwise. Moreover, we observe K characteristics of each individual x_{ik}, where $k = 1, \ldots, K$. The objective of the GME multinomial choice model is to predict $p_{ij} = Pr\{y_{ij} = 1|x_{ik}\}$ for all i and j, which is the probability of individual i choosing each alternative j given the set of his or her characteristics x_{ik}. That is, we want to recover p_{ij} from the observed data y_{ij} and x_{ik}.

In the GME model, the observed data y_{ij} is assumed to be decomposed into the signal component p_{ij} and error component e_{ij},

$$y_{ij} = p_{ij} + e_{ij}. \tag{1}$$

The error component is supposed to be the expected value of a discrete random variable with support $\{v_h\}$ and probabilities $\{w_{ijh}\}$: $e_{ij} = \sum_h v_h w_{ijh}$. Following [11], the error support is constructed using the three sigma rule, which states that the error support should be symmetric around zero and the bounds should be $-3\sigma_y$ and $3\sigma_y$ where σ_y is the empirical standard deviation of the dependence variable. The number of values for the error is usually fixed at 3 or 5. That is, the error support is usually set to $\{-3\sigma_y, 0, 3\sigma_y\}$ or $\{-3\sigma_y, -1.5\sigma_y, 0, 1.5\sigma_y, 3\sigma_y\}$ [6, 11]. Premultiplying (1) with x_{ik} and summing across i, we have MK stochastic moment constraints,

$$\sum_i x_{ik} y_{ij} = \sum_i x_{ik} p_{ij} + \sum_{ih} x_{ik} v_h w_{ijh}, \quad \forall j = 1, \ldots, M, \forall k = 1, \ldots, K. \tag{2}$$

From the principle of ME, p_{ij} that best represents the data must maximize the entropy function

$$\max_{p,w} H(p_{ij}, w_{ijh}) = -\sum_{ij} p_{ij} \log(p_{ij}) - \sum_{ijh} w_{ijh} \log(w_{ijh}) \tag{3}$$

subject to constraints (2) and the following normalization constraints

$$\sum_j p_{ij} = 1, \quad \forall i = 1, \ldots, N \tag{4}$$

$$\sum_h w_{ijh} = 1, \quad \forall i = 1, \ldots, N, \forall j = 1, \ldots, M. \tag{5}$$

This maximization problem can be solved using the Lagrangian method. It should be noted that we can estimate p_{ij} and w_{ijh} without making any functional form or distributional assumptions. However, to analyze marginal effects of each characteristic x_{ik} on p_{ij}, let us assume that

$$y_{ij} = p_{ij} + e_{ij} = G(x_i \beta_j) + e_{ij} \tag{6}$$

for some function G and coefficients β_j. Unlike the Logit or Probit-based models, the GME model only makes the linear assumption on $x_i \beta_j$, but it does not need to make assumption on function G. However, [5] show that the estimated Lagrange multiplier for each stochastic moment constraint λ_j is equal to $-\beta_j$ and the marginal effect can be calculated using the information from the λ_j.

3 The Multi-label CC-GME Model

Let Ω be a choice set that contains M alternatives. Let us observe a set of dummy variables y_{ij} where $y_{ij} = 1$ when individual i chooses alternative j. For the multi-label model, each individual may choose more than one alternative. In other words, it is

possible that $y_{ij} = 1$ for more than one j. Therefore, there are at most 2^M possible outcomes.

3.1 The CC Model

The multi-label CC model was introduced by [12]. The objective of the multi-label choice model is to estimate $Pr\{\underline{y}_i = A|x_{ik}\}$ where $\underline{y}_i$ is the set of all alternatives that individual i chooses and $A \subseteq 2^{\Omega}$. To allow the probability of choosing each alternative to be correlated, the CC method uses Bayes' rule to expand $Pr\{\underline{y}_i|x_{ik}\}$ as follows,

$$\begin{aligned} Pr\{\underline{y}_i|x_{ik}\} &= Pr\{y_{i1} = 1|x_{ik}\}Pr\{y_{i2} = 1|y_{i1}, x_{ik}\}\ldots \\ &\quad Pr\{y_{iM} = 1|y_{i1}, y_{i2}, \ldots, y_{1(M-1)}, x_{ik}\}, \end{aligned} \tag{7a}$$

which can be denoted as

$$Pr\{\underline{y}_i|x_{ik}\} = \prod_{j=1}^{M} Pr\{y_{ij} = 1|\tilde{x}_{ij}\} = \prod_{j=1}^{M} G(\tilde{x}_{ij}\beta_j), \tag{7b}$$

where $\tilde{x}_{ij} = (y_{i1}, \ldots, y_{ij}, x_{i1}, \ldots, x_{iK})$ for all $j = 2, \ldots, M$ and $\tilde{x}_{i1} = (x_{i1}, \ldots, x_{iK})$. In (7a, 7b), notice that the multi-label problem is decomposed into a series of conditionally independent binary choice problems $Pr\{y_{ij} = 1|\tilde{x}_{ij}\}$ for all $j = 1, \ldots, M$. The CC method reduces the dimension of the problem significantly, as 2^{Ω} grows exponentially with the size of the choice set Ω.

Notice that different sequences of the choices y_{ij} yield different estimates and predictions. The criterion to select the sequence of the choice depends on the method used to estimate $Pr\{y_{ij} = 1|\tilde{x}_{ij}\}$. When GME is used, the criterion is to choose the sequence that maximizes the total entropy. When the Logit or Probit models are used, the criterion is to maximize the likelihood.

3.2 The CC-GME Model

To estimate the probability $Pr\{\underline{y}_i|x_{ik}\}$ of individual i choosing a set A, we need to estimate all the components of the Bayes' decomposition in Eq. (7a, 7b). In this section, we address the problem of estimating the parameters for each of the binomial choice problems $Pr\{y_{ij} = 1|\tilde{x}_{ij}\}$ for all $j = 1, \ldots, M$ using the multinomial choice GME model with two alternatives in the choice set. In this case, the y_{ij} only can take values 0 or 1. Therefore, the two alternatives are whether individual i chooses alternative j or not.

Let $y_{ij} = \tilde{p}_{ij} + e_{ij} = G(\tilde{x}_{ij}\beta_j) + e_{ij}$, where $e_{ij} = \sum_h v_h w_{ijh}$. Let k_j be the index for elements in $\tilde{x}_{ij}$. To simultaneously estimate $\tilde{p}_{ij}$ and w_{ijh} for all j, the GME model

can be written as

$$\max_{\tilde{p},w} H(\tilde{p}_{ij}, w_{ijh}) = -\sum_{ij} \tilde{p}_{ij} \log(\tilde{p}_{ij}) - \sum_{ijh} w_{ijh} \log(w_{ijh}) \tag{8}$$

subject to

$$\sum_{i} \tilde{x}_{ijk_j} y_{ij} = \sum_{i} \tilde{x}_{ijk_j} \tilde{p}_{ij} + \sum_{ih} \tilde{x}_{ijk_j} v_h w_{ijh}, \quad \forall j = 1, \ldots, M, \quad \forall k_j = 1, \ldots, (K + j - 1) \tag{9}$$

$$\sum_{h} w_{ijh} = 1, \quad \forall i = 1, \ldots, N, \forall j = 1, \ldots, M, \tag{10}$$

where (8) is the entropy function, (9) are the stochastic-moment constraints and (10) are normalization constraints. From the maximization problem, the Lagrangian can be expressed as

$$\begin{aligned} L(\tilde{p}_{ij}, w_{ijh}) = & -\sum_{ij} \tilde{p}_{ij} \log(\tilde{p}_{ij}) - \sum_{ijh} w_{ijh} \log(w_{ijh}) \\ & + \sum_{jk} \lambda_{jk} \left[\sum_{i} \tilde{x}_{ijk_j} y_{ij} - \sum_{i} \tilde{x}_{ijk_j} \tilde{p}_{ij} - \sum_{ih} \tilde{x}_{ijk_j} v_h w_{ijh} \right] \\ & + \sum_{ij} \delta_{ij} \left[1 - w_{ijh} \right]. \end{aligned} \tag{11}$$

The solutions to the above Lagrangian problem are

$$\hat{p}_{ij} = \exp\left(-1 - \sum_{k} \lambda_{jk} \tilde{x}_{ijk_j} \right) \tag{12a}$$

and

$$\hat{w}_{ijh} = \frac{\exp(-\sum_k \hat{\lambda}_{jk} \tilde{x}_{ijk_j} v_h)}{\sum_h \exp(-\sum_k \hat{\lambda}_{jk} \tilde{x}_{ijk_j} v_h)}. \tag{12b}$$

3.2.1 The Concentrated CC-GME Model

Following [5], the GME model can be reduced to the *concentrated GME model*, which is the model with the minimum number of parameters that represents the original GME model. From the Lagrangian (11) and the GME solutions (12a, 12b), we

can derive the objective function for the concentrated GME model as

$$M(\lambda_{jk_j}) = \sum_{ijk_j} \lambda_{jk_j} \tilde{x}_{ijk_j} y_{ij} + \sum_{ij} \left[\exp(-1 - \sum_{k} \lambda_{jk_j} \tilde{x}_{ijk_j}) \right] + \sum_{ij} \left[\log \sum_{h} \exp(- \sum_{k_j} \lambda_{jk_j} \tilde{x}_{ijk_j} v_h) \right]. \tag{13}$$

The concentrated GME model minimizes expression (13) with respect to λ_{jk_j}. The gradient can be written as

$$\frac{\partial M}{\partial \lambda_{jk_j}} = \sum_{i} \tilde{x}_{ijk_j} y_{ij} - \sum_{i} \tilde{x}_{ijk_j} \tilde{p}_{ij} - \sum_{i} \tilde{x}_{ijk_j} v_h w_{ijh}. \tag{14}$$

Notice that the objective function of the concentrated model is no longer a function of $\tilde{p}_{ij}$ and w_{ijh}, but only a function of λ_{jk_j}. As discussed in [5], the interpretation of λ_{jk_j} from the concentrated model can be compared to that of the β_{jk_j} parameters. Specifically, it can be shown mathematically that $\beta_{jk_j} = -\lambda_{jk_j}$.

3.3 Result Analysis

The multi-label CC-GME model can capture the marginal effects of an individual characteristics on his or her decisions and the dependence pattern of the decisions on all available alternatives.

3.3.1 Marginal Effects

The marginal effects measure the effect of a change in an individual characteristic on an individual's choice decisions. For this multi-label model, the marginal effects are situated at two levels. The first level is to analyze the effect of a change in x_k on the probability that the individual will choose an alternative $j \in \Omega$. This marginal effects in this level is

$$\frac{\partial Pr\{y_j|\tilde{x}_j\}}{\partial x_k} = \beta_{jk} G'(\tilde{x}_j \beta_j). \tag{15}$$

The second level is to analyze the effect of a change in x_k on the probability that the individual will choose a set of alternatives $A \in 2^{\Omega}$. From Eq. (7a, 7b), the marginal effect of x_k on $Pr\{\underline{y}|x\}$ is

$$\frac{\partial Pr\{\underline{y}|x\}}{\partial x_k} = \sum_{j} \left[\beta_{jk} G'(\tilde{x}_j \beta_j) \prod_{q \neq j} G(\tilde{x}_q \beta_q) \right]. \tag{16}$$

3.3.2 Dependence of the Alternatives

In the multi-label model, an individual can choose multiple alternatives. The decisions of choosing each of those alternatives or not can be dependent. The dependence between an alternative j and another alternative q, where the index $q < j$, can be captured from the marginal effects of the change in y_q on $Pr\{y_j|\tilde{x}_j\}$, which is

$$\frac{\partial Pr\{y_j|\tilde{x}_j\}}{\partial y_q} = \beta_{j(K+q)} G'(\tilde{x}_j \beta_j). \tag{17}$$

3.3.3 Model Evaluations

The evaluation of multi-label choice problems requires different measures from those of single-label problems. In contrast to the single-label prediction, which can either be correct or incorrect, the multi-label prediction can be partially correct [13]. Summarized several measures to evaluate multi-label classification models. Commonly used measures include the Hamming Loss, Accuracy, Precision and Recall. The Hamming Loss measures the symmetric difference between the predicted and the true choices with respect to the size of the choice set. The other three methods measures the number of correct predicted choices. The difference is in the normalizing factors. The Accuracy measures the number of correct predicted choices with respect to the sum of all correct, incorrect and missing choices. The Precision and Recall measure the number of correct predicted choices with respect to the number of all predicted choices and the number of all true choices, respectively. The formulas for these four measures are

$$Hamming\ Loss = \sum_{i=1}^{N} \frac{|\hat{\mathbf{y}}_i \Delta \mathbf{y}_i|}{NM} \tag{18}$$

$$Accuracy = \sum_{i=1}^{N} \frac{|\hat{\mathbf{y}}_i \cap \mathbf{y}_i|}{|\hat{\mathbf{y}}_i \cup \mathbf{y}_i|} \tag{19}$$

$$Precision = \sum_{i=1}^{N} \frac{|\hat{\mathbf{y}}_i \cap \mathbf{y}_i|}{|\hat{\mathbf{y}}_i|} \tag{20}$$

$$Recall = \sum_{i=1}^{N} \frac{|\hat{\mathbf{y}}_i \cap \mathbf{y}_i|}{|\mathbf{y}_i|}. \tag{21}$$

where $|\cdot|$ is the number of elements in the set, Δ is the symmetric difference between the two sets, $\cap$ is the intersection of the two sets and $\cup$ is the union of the two sets.

4 Monte-Carlo Experiment

In this section, we used Monte-Carlo simulations to empirically evaluate our multi-label CC-GME model using three simulated datasets with normal, logistic and uniform errors. We compared the performance of the CC-GME model with some selected multi-label estimations including CC-Logit, CC-Probit, BR-GME, BR-Logit, BR-Probit, LP-GME, LP-Logit and LP-Probit models.

The BR model simplifies the multi-label model to an independent series of binary single-label choice models. For example, the BR-GME model applies the GME single-label model to estimate the probability that individual i chooses alternative j, $Pr\{y_{ij} = 1|x_{ik}\}$. The probability that individual i chooses the set of alternatives A is then $Pr\{\underline{y}_i = A|x_{ik}\} = \prod_{j=1}^{K} Pr\{y_{ij} = 1|x_{ik}\}$. The LP model transforms the multi-label problem into a single-label problem of 2^{Ω} alternatives. For example, the LP-GME model applies the GME single-label model to estimate $Pr\{\underline{y}_i = A|x_{ik}\}$ where $A \in 2^{\Omega}$.

4.1 Simulation

For simplicity, we assumed $N = 1{,}000$ individuals, $M = 3$ alternatives and $K = 2$ individual characteristics. The simulation procedures are composed of two main steps. The first step is to generate all characteristics x_i, the true parameters β_{ik}^0 and the error e_{ij}. Given the information from x_i and β_{1k}^0, we calculated the latent variable $y'_{i1} = \sum_k \tilde{x}_{i1k}\beta_{1k} + \varepsilon_{i1}$. We then generated the choice variable y_{i1} by letting $y_{i1} = 1$ when $y'_{i1} \geq 0$ and $y_{ij} = 0$ otherwise. Once we have y_{i1}, we can simulate $y_{i2}, \ldots, y_{iM}$. This first step provided us with the simulated data (y_{ij}, x_i) and true parameters β_{ik}^0.

The second step is to use the data from the first step and apply the CC-Logit, CC-Probit, BR-GME, BR-Logit, BR-Probit, LP-GME, LP-Logit and LP-Probit models. After computing the parameter estimates $\hat{\beta}_{ik}$, the predicted probability of individual i choosing choice j, $\hat{p}_{ij}$, and the corresponding predicted choices, $\hat{\underline{y}}_i$, can be obtained. Using Monte-Carlo simulation, the standard deviation of each estimated parameter and statistics can be estimated.

4.2 Results

The Monte-Carlo simulations allowed us to compare the performances of the CC-Logit, CC-Probit, BR-GME, BR-Logit, BR-Probit, LP-GME, LP-Logit and LP-Probit models.

Table 1 shows the true parameters and the estimated parameters from all the CC and the BR models. It should be noted that the LP models can also provide estimates

Table 1 True and estimated parameters for the CC and BR models

Alternative	Regressor	TRUE	Classifier chains			Binary relevance		
			GME	Logit	Probit	GME	Logit	Probit
Normal error								
y_1	x_1	0.318	0.513 (0.071)	0.516 (0.072)	0.319 (0.044)	0.513 (0.071)	0.516 (0.072)	0.319 (0.044)
	x_2	0	0.003 (0.068)	0.003 (0.068)	0.002 (0.042)	0.003 (0.068)	0.003 (0.068)	0.002 (0.042)
y_2	x_1	−0.223	−0.382 (0.073)	−0.382 (0.076)	−0.228 (0.045)	−0.331 (0.069)	−0.333 (0.071)	−0.199 (0.042)
	x_2	−0.659	−1.100 (0.076)	−1.108 (0.077)	−0.665 (0.044)	−1.073 (0.071)	−1.098 (0.074)	−0.660 (0.043)
	y_1	0.243	0.404 (0.101)	0.382 (0.139)	0.228 (0.082)	–	–	–
y_3	x_1	0.706	1.190 (0.103)	1.205 (0.107)	0.700 (0.060)	0.841 (0.071)	1.092 (0.095)	0.640 (0.053)
	x_2	−0.360	−0.629 (0.092)	−0.633 (0.100)	−0.368 (0.058)	−0.645 (0.068)	−0.849 (0.092)	−0.498 (0.052)
	y_1	0.551	0.965 (0.150)	0.992 (0.177)	0.574 (0.102)	–	–	–
	y_2	0.844	1.407 (0.152)	1.442 (0.183)	0.837 (0.106)	–	–	–
MSE			0.001	0.143	0.005	0.001	0.112	0.006
Logistic error								
y_1	x_1	0.318	0.324 (0.070)	0.325 (0.070)	0.203 (0.043)	0.324 (0.070)	0.325 (0.070)	0.203 (0.043)
	x_2	0	0.006 (0.060)	0.006 (0.061)	0.004 (0.038)	0.006 (0.060)	0.006 (0.061)	0.004 (0.038)
y_2	x_1	−0.223	−0.227 (0.068)	−0.228 (0.069)	−0.139 (0.042)	−0.209 (0.068)	−0.208 (0.069)	−0.127 (0.042)
	x_2	−0.659	−0.665 (0.072)	−0.669 (0.073)	−0.409 (0.043)	−0.656 (0.071)	−0.666 (0.073)	−0.408 (0.043)
	y_1	0.243	0.244 (0.099)	0.241 (0.131)	0.148 (0.080)	–	–	–
y_3	x_1	0.706	0.703 (0.079)	0.707 (0.080)	0.422 (0.046)	0.593 (0.068)	0.676 (0.076)	0.407 (0.044)
	x_2	−0.360	−0.348 (0.075)	−0.348 (0.079)	−0.208 (0.047)	−0.390 (0.067)	−0.452 (0.077)	−0.272 (0.046)
	y_1	0.551	0.563 (0.129)	0.581 (0.146)	0.348 (0.087)	–	–	–
	y_2	0.844	0.832 (0.133)	0.852 (0.184)	0.511 (0.109)	–	–	–

(continued)

Table 1 (continued)

Alternative	Regressor	TRUE	Classifier chains			Binary relevance		
			GME	Logit	Probit	GME	Logit	Probit
MSE			0.000	0.012	0.043	0.000	0.007	0.032
Uniform error								
y_1	x_1	0.318	1.665	1.680	1.008	1.665	1.680	1.008
			(0.095)	(0.097)	(0.054)	(0.095)	(0.097)	(0.054)
	x_2	0	−0.022	−0.023	−0.013	−0.022	−0.023	−0.013
			(0.082)	(0.082)	(0.048)	(0.082)	(0.082)	(0.048)
y_2	x_1	−0.223	−1.284	−1.351	−0.780	−0.758	−0.877	−0.505
			(0.148)	(0.173)	(0.096)	(0.103)	(0.124)	(0.069)
	x_2	−0.659	−3.812	−4.011	−2.320	−3.353	−3.769	−2.172
			(0.199)	(0.237)	(0.125)	(0.170)	(0.216)	(0.115)
	y_1	0.243	1.431	1.492	0.859	–	–	–
			(0.184)	(0.293)	(0.160)			
y_3	x_1	0.706	3.581	4.488	2.552	1.359	2.801	1.580
			(0.183)	(0.347)	(0.188)	(0.068)	(0.184)	(0.097)
	x_2	−0.360	−1.818	−2.240	−1.280	−1.163	−2.456	−1.385
			(0.182)	(0.288)	(0.155)	(0.072)	(0.168)	(0.089)
	y_1	0.551	2.811	3.506	1.997	–	–	–
			(0.293)	(0.468)	(0.253)			
	y_2	0.844	4.307	5.434	3.085	–	–	–
			(0.294)	(0.533)	(0.288)			
MSE			0.047	7.161	1.728	0.017	3.480	0.783

[1] Standard deviations in parentheses

of the parameters. However, the parameters in the LP model are not comparable to the true parameters generated in this Monte-Carlo experiment. It should be noted that the data simulation process was based on the CC model. When the errors are normally distributed, the true model is the CC-Probit model. Therefore, the Probit-based models performed better than the Logit-based models. When the errors are logistically distributed, the true model is the CC-Logit model and the Probit models performed better than the Logit models. However, regardless of the error distributions, the GME models always have the lowest MSE.

Figure 1 shows the prediction regions for an individual's decision on each alternative y_{ij} given his or her characteristics x_1 and x_2 using the CC-GME model. Each of the three lines represents the combinations of x_1 and x_2 such that $Pr\{y_{ij} = 1|\tilde{x}_{ij}\} = 0.5$. Regions (1) to (8) represent the choices $y_i = (0, 0, 0)$, $(0, 0, 1)$, $(0, 1, 0)$, $(0, 1, 1)$, $(1, 0, 0)$, $(1, 0, 1)$, $(1, 1, 0)$ and $(1, 1, 1)$, respectively. Therefore, the result shows that individuals with high value x_1 and low value of x_2 are more likely to choose all three alternatives. Individuals with lower x_1 and high x_2 are likely to choose none of the alternatives.

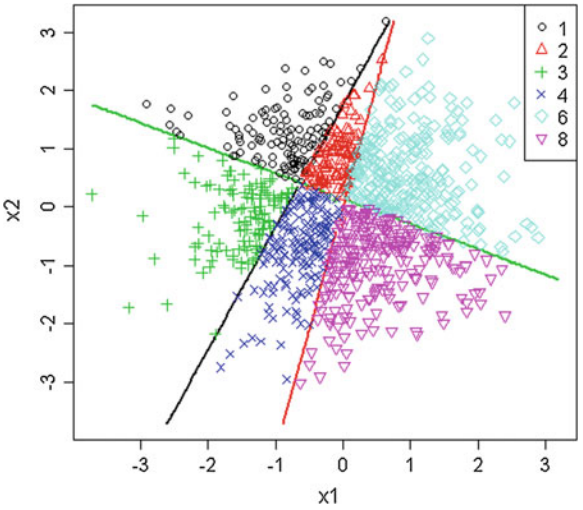


Fig. 1 Prediction regions for all possible sets of alternatives from the CC-GME estimation for the simulation with logistic errors

Table 2 reports the Hamming Loss, Accuracy, Precision and Recall statistics for all the CC, BR and LP models. The results show that the forecasting performance depends on the choice of the problem transformation methods, but not on the choice of single-label estimation methods. That is, the CC model outperformed the BR and LP models with respect to all evaluation measures, except Precision. The CC-GME, CC-Logit and CC-Probit models yielded similar results.

5 Occupational Hazards Empirical Example

Consider a problem in which an individual chooses a job with multiple job attributes. This problem can be viewed as an individual choosing a set of job attributes. In this empirical example, the job attributes are a set of occupational hazards. Therefore, each individual will choose the hazards from which he or she gains the least disutility. In this section, we apply the CC-GME model to predict a set of occupational hazards that an individual faces and the factors determining his or her choices of hazards. For the performance evaluation, we applied the five-fold cross validation method to compare the out-sample prediction performance between the CC-GME model and other models [10].

5.1 Data Description

The dataset is from *The Informal Worker Analysis and Survey Modeling for Efficient Informal Worker Management Project*, which aims at studying the structure and

Table 2 Model comparison for the simulated data

Evaluation	Classifier chains			Binary relevance			Label powerset		
Methods	GME	Logit	Probit	GME	Logit	Probit	GME	Logit	Probit
Normal error									
Hamming loss	**0.304**	**0.303***	**0.303***	0.341	**0.314**	**0.315**	0.326	0.327	0.331
	(0.008)	(0.009)	(0.008)	(0.008)	(0.008)	(0.008)	(0.009)	(0.009)	(0.010)
Accuracy	**0.589**	**0.590***	**0.590***	0.516	**0.581**	**0.581**	0.543	0.541	0.528
	(0.011)	(0.013)	(0.013)	(0.009)	(0.013)	(0.013)	(0.013)	(0.013)	(0.016)
Precision	0.726	0.725	**0.725**	0.737	0.713	0.713	**0.736**	**0.737**	**0.745***
	(0.010)	(0.009)	(0.009)	(0.011)	(0.008)	(0.008)	(0.013)	(0.013)	(0.015)
Recall	**0.757**	**0.760***	**0.760***	0.633	**0.758**	**0.758**	0.674	0.671	0.645
	(0.009)	(0.017)	(0.017)	(0.008)	(0.017)	(0.017)	(0.017)	(0.017)	(0.025)
Logistic error									
Hamming loss	**0.364**	**0.363***	**0.364**	0.391	**0.372**	**0.372**	0.383	0.383	0.388
	(0.010)	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)	(0.011)	(0.011)	(0.011)
Accuracy	**0.521***	**0.521***	**0.521***	0.455	**0.516**	**0.516**	0.475	0.473	0.460
	(0.013)	(0.017)	(0.017)	(0.010)	(0.018)	(0.018)	(0.018)	(0.018)	(0.020)
Precision	**0.666**	**0.659**	**0.659**	**0.666**	**0.648**	**0.648**	**0.665**	**0.666**	**0.670***
	(0.014)	(0.010)	(0.010)	(0.014)	(0.010)	(0.010)	(0.016)	(0.015)	(0.017)
Recall	**0.713**	**0.714**	**0.714**	0.589	**0.717**	**0.718***	0.624	0.620	0.596
	(0.014)	(0.029)	(0.029)	(0.009)	(0.030)	(0.030)	(0.031)	(0.030)	(0.035)
Uniform error									
Hamming loss	**0.156**	**0.155***	**0.155***	0.216	0.174	0.174	0.184	0.184	0.185
	(0.005)	(0.006)	(0.006)	(0.006)	(0.007)	(0.007)	(0.006)	(0.006)	(0.006)
Accuracy	**0.768**	**0.769***	**0.769***	0.668	0.745	0.745	0.724	0.723	0.719
	(0.008)	(0.008)	(0.008)	(0.007)	(0.009)	(0.009)	(0.008)	(0.008)	(0.009)
Precision	**0.866**	**0.867**	0.867	**0.879***	0.849	0.849	**0.867**	**0.870**	**0.874**
	(0.007)	(0.005)	(0.005)	(0.008)	(0.006)	(0.006)	(0.007)	(0.007)	(0.008)
Recall	**0.871***	**0.871***	**0.871***	0.736	**0.860**	**0.859**	**0.814**	**0.811**	**0.802**
	(0.059)	(0.006)	(0.006)	(0.007)	(0.007)	(0.006)	(0.007)	(0.007)	(0.009)

Standard deviations in parentheses. Statistics with * represent estimation methods that are the 'best' with respect to each evaluation metric. Statistics in bold represent estimation methods with the prediction power not statistically different from the 'best' estimation method

nature of the informal sector in Chiang Mai, Thailand in 2012. In the survey, each respondent was asked whether he or she faced each of the three types of occupational hazards, namely, (1) physical and mechanical hazards, (2) ergonomic and psychosocial hazards and (3) biological and chemical hazards. The survey also provides data for each individual's demographic, employment and financial status. Explanatory variables used in this study include (1) age, (2) number of children, (3) total income and dummy variables for (4) female, (5) high school, (6) college and (7) agricultural household.

Table 3 Model comparison for the occupational hazards data

Evaluation	Classifier chains			Binary relevance			Label powerset		
Methods	GME	Logit	Probit	GME	Logit	Probit	GME	Logit	Probit
Hamming loss	**0.263***	**0.297**	0.372	**0.284**	0.382	0.383	**0.315**	**0.325**	–
	(0.049)	(0.016)	(0.022)	(0.039)	(0.032)	(0.032)	(0.083)	(0.020)	
Accuracy	**0.702***	0.541	**0.658**	**0.675**	**0.652**	**0.652**	**0.676**	0.529	–
	(0.065)	(0.012)	(0.041)	(0.055)	(0.053)	(0.053)	(0.080)	(0.015)	
Precision	**0.753**	**0.759***	**0.758**	**0.755**	**0.748**	**0.748**	**0.701**	**0.723**	–
	(0.080)	(0.058)	(0.047)	(0.084)	(0.044)	(0.044)	(0.102)	(0.063)	
Recall	**0.914**	**0.848**	**0.844**	0.870	**0.849**	**0.850**	**0.956***	**0.883**	–
	(0.023)	(0.096)	(0.112)	(0.045)	(0.131)	(0.131)	(0.040)	(0.131)	

Standard deviations in parentheses. The LP-Probit model fails to converge. Statistics with * represent estimation methods that are the 'best' with respect to each evaluation metric. Statistics in bold represent estimation methods with the prediction power not statistically different from the 'best' estimation method

5.2 Results

For the choice of problem transforming methods, the results are similar to the simulation exercises in the sense that the CC model outperformed the BR and LP models in most measures (see Table 3). Specifically, the CC model is superior than the BR and LP models with respect to the Hamming Loss, Accuracy and Precision criteria. For the choice of single-label estimation methods, the GME model outperformed the Logit and Probit models with respect to the Hamming Loss, Accuracy and Recall measures.

6 Concluding Remarks

The empirical results obtained in this study show that the forecasting performance depends on the choice of the problem transformation methods, but not on the choice of single-label estimation methods. Specifically, the CC model outperformed the BR and LP models with respect to all evaluation measures expect the Precision. For the parameter estimates, the GME-based methods yielded smaller MSE than those of the Logit and Probit-based methods.

Although the Bayes' rule implies that $Pr\{\underline{y}_i|x_{ik}\} = \prod_{j=1}^{M} Pr\{y_{ij} = 1|\tilde{x}_{ij}\}$, it does not imply directly that $Pr\{\underline{y}_i|x_{ik}\} = \prod_{j=1}^{M} G(\tilde{x}_{ij}\beta_j)$. The CC-GME model still relies on the linearity assumption of $\tilde{x}_{ij}\beta_j$ when we set

$$Pr\{y_{ij} = 1|\tilde{x}_{ij}\} = G(\tilde{x}_{ij}\beta_j). \tag{22}$$

Therefore, other methods to incorporate the dependency among alternatives into the multi-label classification problem with weaker assumptions could potentially improve the performance.

Acknowledgments We are highly appreciated and would like to acknowledge the financial support from the Thailand Research Fund through the Royal Golden Jubilee Ph.D. Program (Grant No. PHD/0211/2556).

References

1. Baltas, G.: A model for multiple brand choice. Eur. J. Oper. Res. **154**(1), 144–149 (2004)
2. Bhat, C.R., Srinivasan, S., Sen, S.: A joint model for the perfect and imperfect substitute goods case: application to activity time-use decisions. Transp. Res. Part B: Methodol. **40**(10), 827–850 (2006)
3. Bhat, C.R., Srinivasan, S.: A multidimensional mixed ordered-response model for analyzing weekend activity participation. Transp. Res. Part B: Methodol. **39**(3), 255–278 (2005)
4. Denzau, A.T., Gibbons, P.C., Greenberg, E.: Bayesian estimation of proportions with a cross-entropy prior. Commun. Stat.-Theory Methods **18**(5), 1843–1861 (1989)
5. Golan, A., Judge, G., Perloff, J.M.: A maximum entropy approach to recovering information from multinomial response data. J. Am. Stat. Assoc. **91**(434), 841–853 (1996)
6. Golan, A.: Information and Entropy Econometrics: A Review and Synthesis. Now Publishers Inc. (2008)
7. Heath, D., Zitzelberger, A., Giraud-Carrier, C.G.: A multiple domain comparison of multi-label classification methods. In: Working Notes of the 2nd International Workshop on Learning from Multi-label Data at ICML/COLT, 21–28 (2010)
8. Jaynes, E. T.: Information theory and statistical mechanics. Phys. rev. **106**(4), 620 (1957a)
9. Jaynes, E. T.: Information theory and statistical mechanics. II. Phys. rev. **108**(2), 171 (1957b)
10. Mosteller, F., Tukey, J.W.: Data Analysis, Including Statistics. The Collected Works of John W. Tukey: Graphics pp. 1965–1985, vol. 5 (123) (1988)
11. Pukelsheim, F.: The three sigma rule. Am. Stat. **48**(2), 88–91 (1994)
12. Read, J., Pfahringer, B., Holmes, G., Frank, E.: Classifier chains for multi-label classification. Mach. Learn. **85**(3), 333–359 (2011)
13. Santos, A., Canuto, A., Neto, A.F.: A comparative analysis of classification methods to multi-label tasks in different application domains. Int. J. Comput. Inform. Syst. Indust. Manag. Appl **3**, 218–227 (2011)
14. Shannon, C.E.: A mathematical theory of communication. ACM SIGMOBILE Mob. Comput. Commun. Rev. **5**(1), 3–55 (2001)
15. Soofi, E.S.: A generalizable formulation of conditional logit with diagnostics. J. Am. Stat. Assoc. **87**, 812–816 (1992)
16. Train, K.: Data analysis. Including Statistics Discrete Choice Methods with Simulation. Cambridge University Press, Cambridge (2009)
17. Tsoumakas, G., Katakis, I.: Multi-label classification: an overview. Int. J. Data Warehous. Min. **3**(3), 1–13 (2007)
18. Zhang, M.L., Zhou, Z.H.: ML-KNN: a lazy learning approach to multi-label learning. Pattern Recognit. **40**(7), 2038–2048 (2007)

Part II
Applications

Asymmetric Volatility of Local Gold Prices in Malaysia

Mohd Fahmi Ghazali and Hooi Hooi Lean

Abstract This study investigates the volatility of local gold prices in Malaysia using daily data over the period of July 2001–May 2014. Specifically, this paper analyzes the asymmetric reaction of gold in different weights to negative and positive news on average at all times as well as during extreme decreases in stock market. The former provides potential evidence for hedge, while the latter tests for the existence of a safe haven characteristic. We find that the local gold returns demonstrate an inverted asymmetric reaction to positive and negative innovations respectively. Positive shock increases the gold returns volatility more than the negative shock in full sample as well as the stock market downside, thus supporting the hedge and safe haven properties of gold investment in Malaysia.

1 Introduction

Despite the importance of gold as a hedge or safe haven asset [10, 21], studies that investigating the volatility of gold prices are infrequent. With the exception of [8], we are unaware of any study that analyzes the volatility of gold prices and the asymmetric effect of market shocks on the gold prices volatility. Thus far, many researches only focus on volatility in stock markets and its asymmetric behaviour to negative and positive shocks. For instance, [11, 13, 14] report a larger increase of volatility in response to negative shocks than the positive shocks. This asymmetric is explained with the leverage of firms and volatility feedback effect.

On the other hand, [37] focus on estimating several asymmetric power GARCH (APGARCH) models and the role of US dollar in influencing gold prices. [7] model the volatility of a gold futures price, while [9, 10, 21] focus on hedge and safe haven

M.F. Ghazali
Labuan Faculty of International Finance, Universiti Malaysia Sabah,
87000 Federal Territory Labuan, Malaysia
e-mail: fahmi@ums.edu.my

M.F. Ghazali · H.H. Lean (✉)
Economics Program, School of Social Sciences, Universiti Sains Malaysia,
11800 USM Penang, Malaysia
e-mail: hooilean@usm.my

V.-N. Huynh et al. (eds.), *Econometrics of Risk*, Studies in Computational Intelligence 583, DOI 10.1007/978-3-319-13449-9_14

characteristics of gold. These studies analyze some structures of the volatility of gold prices but do not investigate the volatility asymmetry.

If the volatility of gold price increases following the bad news in economic and financial markets, investors can transmit the increase in volatility and uncertainty to the gold market. This scenario is interpreted by investor as a hedge and safe haven purchase. In other words, if the empirical results show a volatility clustering of gold returns and an inverted asymmetric reaction (positive shock increases the volatility more than the negative shock), this effect is related to the hedge and safe haven characteristics of gold. Accordingly, if the gold return decreases in times of stock market and economic condition rising, investors can transmit the decrease in volatility and uncertainty to the gold market. In other words, if the price of gold falls, uncertainty and thus volatility is lower. This can lead to an asymmetric reaction of the volatility of gold price which is different from the effect that observed in stock market: the volatility of gold price increases by more with positive shocks than with negative shocks [8].

Another explanation for the volatility asymmetry of gold price is the role of inventory and storage. If inventory level of an asset is low, the risk of inventory exhaustion increases and will lead to higher price and volatility. In contrast, if the inventory level of an asset is high, the risk of inventory exhaustion decreases and will lead to lower price and volatility. If inventory level is important for gold, it could establish an asymmetric reaction of the volatility of gold price to past price shocks [8].

This study investigates the asymmetric reaction of gold returns on its volatility for two major reasons. First, if the volatility of gold return exhibits different reactions to shocks than stock, the (desired) low or negative correlation of gold with other assets might be compromised in certain conditions, therefore complements the correlation analysis of gold with other assets. Second, the economic explanations for asymmetric volatility for stocks, namely, time-varying risk premium or volatility feedback effect and leverage effect are not applicable for gold. This is due to volatility feedback as discussed by, for example, [13, 19] see that an anticipated increase in volatility would raise the required rate of return, in turn necessitating an immediate stock-price decline to allow for higher future returns. In other words, the volatility feedback effect justifies how an increase in volatility may result in negative returns. The fundamental difference between the leverage and volatility feedback explanations lies in the causality, where leverage effect explains why a negative return leads to higher subsequent volatility. Although the term is arguably a misnomer, [11] defines leverage effect as a negative correlation between current return and future volatility, which means bad news will cause violent fluctuations compare to the good ones. This model is not consistent with an inverted asymmetric reaction of gold due to the amplification of positive returns.

This study also contributes to the literature by investigating the volatility of gold at different weights, i.e. 1 ounce, 1/2 ounces and 1/4 ounces. Introduction of gold with different weights by the Malaysian government is to encourage retail investors to invest in gold through regular small purchases as well as get protection from gold prices volatility. Relatively, investors pay a higher premium to gold with a smaller

size because the cost of making gold is fixed. The different in gold price can be explained by economies of scale, that is, the cost advantage that arises from the bigger weight of gold. Per-unit fixed cost is lower because it is shared over a larger weight of gold. Variable costs per unit are reducing due to operational efficiencies and synergies. We expect different weights of gold will yield different returns and thus display how it can influence the size and magnitude of asymmetric effect. A smaller weight of gold is more volatile because the gold is affordable and able to be purchased relatively cheap by retail investors.

2 Literature Review

There is a large literature studying the volatility in equity markets and its asymmetric behaviour to negative and positive shocks. As one of the primary restrictions of generalized autoregressive conditional heteroscedasticity (GARCH) that they enforce a symmetric response of volatility to positive and negative shocks; many previous studies using asymmetric volatility in their estimations, particularly in equity markets. Many studies examine the asymmetric volatility in equity markets with different methods [11, 13, 14]. These studies report a larger increase of volatility in response to negative shocks than positive shocks. This asymmetry in the reaction to shocks is explained with the leverage of firms and volatility feedback effect.

Despite the importance of gold as a hedge and safe haven asset, studies that investigating the volatility of gold are rare. [25] use GARCH(1,1) model and find that gold is negatively correlated with the S&P 500 when realized stock market volatility is more than two standard deviations from the mean. The diversification is most important to investors when equity markets are experiencing high volatility and poor performance. Using the same model, [10] report the conditional volatility of gold for a 30-year period and analyze the safe haven hypothesis for different volatility regimes. The results show that gold is a hedge in European markets, Switzerland and the US and a safe haven in periods of increase volatility (90 and 95 % thresholds) in most markets. But gold does not work as a safe haven in spells of extreme volatility (99 % threshold) or uncertainty except for the US and China. [34] analyze the effect of macroeconomic news on commodities and gold using GARCH(1,1) model. Some commodity prices are influenced by the surprise element in macroeconomic news, with evidence of a pro-cyclical bias, particularly when control for the effect of the US dollar. Commodities tend to be less sensitive than financial assets, for instance, crude oil shows no significant responsiveness to almost all announcements. Nevertheless, as commodity markets become financialized, their sensitivity appears to rise somewhat to both macroeconomic news and surprise interest rate changes. The gold price is sensitive to a number of macroeconomic announcements in the US and Euro area, including retail sales, non-farm payrolls, and inflation. High sensitivity of gold to real interest rates and its unique role as a safe haven and store of value typically leads to a counter-cyclical reaction to surprise news, in contrast to their commodities. It also shows a particularly high sensitivity to negative surprises that might lead investors to become more risk-averse.

In the same issue, [6] investigate the impact of macroeconomic and financial market variables on precious metals markets in a framework that utilizes monthly data. While many studies rely on autoregressive conditional heteroscedasticity (ARCH)/GARCH estimation to establish possible volatility relationships in precious metals markets, this study provides further information on the long-term trends prevalent in these markets as well as identifying structural linkages which hitherto remains uncovered. Based on the vector autoregressive (VAR) model, they find that gold volatility does not respond to changes in equity volatility, but is instead sensitive only to monetary variables during the period of 1986–1995. From 1996 to 2006, nevertheless, gold volatility does have a positive relation to equity volatility. The volatility of silver is not sensitive to either financial or monetary shocks in either period. [35] also use a VAR model to examine the relation between volatility index (VIX), oil, and metals. Using VIX as a proxy for global risk perceptions, they find that gold, exchange rates, and VIX lead oil prices. VIX and oil are negatively related. VIX has an economically significant long-run effect on oil, gold, and silver, and is itself affected by oil and silver in the long-run. In another study, [7] analyze the volatility structure of gold, trading as a futures contract on the Chicago Board of Trade using intraday data from January 1999 to December 2005. They use GARCH modeling and the Garman Klass estimator and find significant variations across the trading days consistent with microstructure theories, although volatility is only slightly positively correlated with volume when measured by tick-count.

Some studies analyze some features of the volatility of gold but do not focus on volatility asymmetry and its importance for safe haven property. [22] study the relation of return and volatility in the commodity and the stock market for indices without an explicit analysis of gold. They estimate an exponential GARCH (EGARCH) specification using the Goldman Sachs and JP Morgan Commodity Indices to find that commodity returns in general display asymmetric volatility, where the relationship between return and volatility in the commodity markets is the inverse of that observe in the stock markets. The inverse relationship may exist under specific circumstances. For instance, geopolitical concerns would pull the stock market down, but prove positive for energy and metals because of the potential for supply disruption. Further, changes in commodity prices do impact the share prices of companies, whether they are commodity producers or consumers. For instance, a rise in steel prices would be positive for producers but not for consumers. [9] find that while conditional asymmetric volatility is significant, gold has a negative and significant relation to equities in bear markets, but not in bull markets. [37] specify an asymmetric component in their APGARCH model but find that the asymmetry is statistically insignificant. They find that gold volatility is largely determined by prior period gold volatility itself. This is true both overall and during the crisis periods of 1987 and 2001.

Parallel to research on the relationship between the asymmetric volatility and gold returns, [8] finds that the volatility of gold returns exhibits an asymmetric reaction to positive and negative gold returns. The asymmetric nature of this reaction allows for its characterization as abnormal or inverted when compared to its parallel in equity markets. In addition, since this asymmetric reaction is ten times larger for gold than for any other commodity, [8] argues that investors interpret positive gold returns as

an indication of future adverse conditions and uncertainty in other asset markets. [8] further argues that this interpretation effectively introduces uncertainty into the gold market and brings about increased volatility of gold returns.

Others examine the interaction between oil and the precious metals, for instance [24] where they find persistent volatility effects of oil shocks (as well as interest rate shocks) on both gold and silver. These effects are long-lasting, as indicated by significant persistence in EGARCH (2,2) and GARCH (2,2) models. Similar with [37], they find no evidence of volatility asymmetry in their study. [31] in their research apply threshold GARCH (TGARCH) models in order to describe the spillover effects of oil and gold price returns on industrial sub-indices returns and GARCH models in order to examine volatility spillover effects on each other. By conducting the research, they conclude that the volatility of gold price returns has no effect on oil price returns, while oil price return volatility has spillover effects on the volatility of gold price returns, which means that investors can monitor gold prices by reviewing oil prices returns. Another finding that they conclude is that previous volatility of oil price returns spillover the volatility of Electronics and Rubber sub-indices returns. Meanwhile, the volatility of Chemistry, Cement, Automobile, Food and Textiles sub-indices are affected by previous volatility of gold price returns. Another research that is focusing on the spillover effects of gold return volatility on stock return volatility is made by [36]. They study the spillover between gold, the US bonds and stocks from 1970 to 2009 by applying VAR model together with variance decomposition. The findings are that previous gold return has low correlation and low spillover effect on stocks and bonds returns. Therefore, they suggest that gold can hardly be a useful predictor for stocks and bonds; however, low correlation feature makes gold a useful portfolio diversifier.

The findings of prior studies prove that researches on asymmetric volatility of gold are scarce, suggesting further studies are needed to shed light on the issue. Our sample at 2001–2014 is also the most up-to-date and takes into account the recent global financial crisis. In addition, this study contributes to the empirical literature by investigating the issue using different weights of gold via two types of asymmetric GARCH model, that is, TGARCH and EGARCH to see whether the difference can change the asymmetric pattern of gold volatility.

3 Volatility Model

The common econometric approach to volatility modelling is the GARCH framework that was pioneered by [12, 17]. The models are handy if we model the time-varying volatility of a financial asset. Therefore it becomes the bedrock of the dynamic volatility models [2]. The advantage of these models is that they are practically easy to estimate in addition to allow us to perform diagnostic tests [16]. Nevertheless, the normal GARCH model cannot account for the entire leptokurtosis in data [28] and a better fit is obtained using non-normal distributions such as Students t, Gen-

eralized Error Distribution (GED), normal-Poisson mixture and normal-lognormal distributions or the EGARCH model [4, 5, 18, 20, 27, 30].

On the other hand, several studies [14, 23, 33] point out asymmetric responses in the conditional variance, suggesting the leverage effect and differential financial risk depending on the direction of price change movements. In response to the weakness of symmetric assumption, [23, 33] model a conditional variance formulation that capture asymmetric response in the conditional variance.

3.1 TGARCH Model

The model is introduced by [23]. The mean equation of both systematic and threshold analyses in this model is specified in Eqs. (1) and (2) respectively. The specification of conditional volatility of gold return is estimated by the variance equation in Eq. (3):

$$r_{g,t} = \alpha_0 + \alpha_1 \sum_{i=1}^{k} r_{g,t-i} + \beta_1 \sum_{j=0}^{l} r_{s,t-j} + \varepsilon_t \tag{1}$$

$$r_{g,t} = \alpha_0 + \alpha_1 \sum_{m=1}^{n} r_{g,t-m} + \beta_1 \sum_{p=0}^{q} r_{s,t-p(q)} + \mu_t \tag{2}$$

$$h_t = \delta + \alpha \varepsilon_{t-1}^2 + \vartheta d_{t-1} \varepsilon_{t-1}^2 + \rho h_{t-1} \tag{3}$$

$$\varepsilon_t \sim GED(0, h + t) \tag{4}$$

where $r_{g,t}$ denotes gold return at time t. In the systematic analysis (Eq. 1), the gold returns are regressed on a constant, α_0 and its own lagged returns as well as the contemporaneous and lagged stock market shocks which are captured by $r_{s,t-j}$. In the threshold analysis which is specified in Eq. (2), we include lagged returns of gold with the extreme stock return movements. In order to obtain the extreme values of stock returns, we estimate the threshold value of stock return via quantile regression. Then, we analyze the volatility of gold in times of stock market stress by including the regressors that contain stock returns in the q % lower quantile (10, 5, 2.5 and 1 %) in the mean equation.[1] If the stock market returns exceed a certain (lower tail) threshold given by the q %, the value of dummy variable $D(\ldots)$ is one and zero otherwise.

In Eq. (3), h_t is known as the conditional variance of error at time t. Parameters to be estimated are $\delta, \alpha, \vartheta$ and ρ. The constant volatility is estimated by δ, the effect of lagged return shocks of gold on its volatility (ARCH) is estimated by α and an asymmetry or leverage term if the return shock is negative is captured by ϑ. When

[1] The severity of the shock is taken into account by looking at a range of lower quantiles of stock returns. The choice of the quantiles is arbitrary to some degree. Nevertheless, these quantiles have also been analyzed in other studies, such as [3, 10, 21].

shock is positive, the effect on volatility is α; when the shock is negative, the effect on volatility is $\alpha + \vartheta$. If there is a symmetric effect of lagged shocks on the volatility of gold, ϑ is zero, and the equation is a standard GARCH form. In contrast, if lagged negative shocks augment the volatility by more than lagged positive shocks ($\vartheta > 0$), there is an asymmetric effect which is typically associated with a leverage effect or a volatility feedback effect. If lagged negative shocks decrease the volatility of gold ($\vartheta < 0$) the asymmetric effect typically found for equity is inverted, that is, positive shocks of gold increase its volatility by more than negative shocks. The influence of the previous periods conditional volatility level (h_{t-1}) on the current period is given by ρ (GARCH). Since the results can be influenced by the distributional assumptions regarding the error distribution, we estimate the asymmetric GARCH model of a GARCH (1,1)-type by maximum-likelihood with a Gaussian error distribution and a student-t error distribution [8]. The innovations in Eq. (4) are assumed to follow the GED (Generalized Error Distribution). We employ the GED because of its ability to accommodate leptokurtosis (fat tails) that usually observed in the distribution of financial time series [33].

3.2 EGARCH Model

The EGARCH model which is introduced by [33] has an advantage that it requires no non-negativity restrictions of the parameters in the variance equation as in the case of the traditional GARCH (1,1) specification. Similar to TGARCH model, it allows positive and negative shocks to have asymmetric influences on volatility [32]. The mean equation of systematic analysis and threshold analysis as well as the innovations of gold return in EGARCH model are specified the same as in Eqs. (1), (2), and (4) respectively. However, the corresponding time-varying conditional variance of the regression residuals is written as follows:

$$h_t = \alpha_1 + \alpha_2 \left| \frac{\varepsilon_{t-1}}{\sqrt{h_t}} \right| + \alpha_3 \frac{\varepsilon_{t-1}}{\sqrt{h_t}} + \alpha_4 h_{t-1} \tag{5}$$

where h_t is the conditional variance of error at time t which permits the coefficients to be negative, and thus allows the positive and negative innovations to have different impacts on the conditional variance. The parameters to estimate are α_1, α_2, α_3, and α_4. The constant volatility is estimated by α_1. The parameter of α_4 measures the persistence in conditional volatility irrespective of anything happening in the market. It captures the influence of past volatility on the current gold return volatility. When α_4 is relatively large, then volatility takes a long time to die out following a crisis in the market [1]. $\frac{\varepsilon_{t-1}}{\sqrt{h_t}}$ is the standardized value of the lagged residual and it helps in interpreting the magnitude and the persistence of the shocks. The α_3 parameter measures the asymmetry or the leverage effect, the parameter of importance so that the EGARCH model allows for testing of asymmetries. If α_3 equals to zero and

significant, then the model is symmetric. If α_3 not equal to zero and significant, the impact is asymmetric. When α_3 less than zero, the positive shocks (good news) generate less volatility than negative shocks (bad news), that is, leverage effect is present. When α_3 greater than zero, it implies that positive innovations are more destabilizing than negative innovations. In other word, positive shock to gold time series is likely to cause volatility to rise by more than a negative shock of the same magnitude. Thus, term α_3 is the asymmetrical effect term. On the other hand, the α_2 represents a magnitude effect or the symmetric effect of the model, the GARCH effect. In the presence of a positive shock (the term $\frac{\varepsilon_{t-1}}{\sqrt{h_t}}$ is positive), the shock impact on the conditional variance is $(\alpha_2 + \alpha_3)$ and when this term is negative (the leverage effect is present) the impact is $(\alpha_2 - \alpha_3)$.

4 Empirical Analysis

This section describes the data, report the descriptive statistics, econometric analysis and discuss the main findings.

4.1 Data

Daily data are gathered from various sources ranging from July 18, 2001 to May 30, 2014. This study uses the prices of 1 ounce, 1/2 ounces and 1/4 ounces of Malaysian official gold bullion, *Kijang Emas* which is denominated in local currency, *Ringgit* Malaysia to represent the local gold prices. The prices of *Kijang Emas* are determined by the prevailing international gold market price.[2] Unlike the United States and the United Kingdom, Malaysia does not have important role in the gold trading market. Nevertheless, Malaysia is chosen due to the deep interest in gold shown by Malaysian policy makers and investors in the face of 1997/1998 Asian financial crisis as it is seen as a stable and profitable tool for successful investments. The recent and on-going financial crisis and the attendant strength of gold price, also cause profound interest in this precious metal in Malaysia. Thus, answering the question of the asymmetric volatility of gold prices would provide important information to Malaysian investors.

The Kuala Lumpur Composite Index (KLCI) is used to represent stock market. *Kijang Emas* prices are collected from Central Bank of Malaysia and KLCI is from Datastream International. Daily gold returns and stock return are computed using continuous compounded return. A problem arises with missing observations due to different holidays in stock market and gold market. We follow [26, 29] by adopting the method of Occam's razor (using the previous day's price). Hence, it is desirable to fill in estimate-based information from an adjacent day.

[2] Source: Central Bank of Malaysia.

4.2 Descriptive Statistics

Figure 1 shows the time series plot of domestic gold prices from 2001 to 2014. The figure shows that of the local gold prices have shown a dramatic growth for about 300 %. The performance of gold is more impressive given the losses suffered in other asset classes during the 2007/2008 financial crisis. In times of uncertainty, due to investors unwillingness to trade, asset values become ambiguous. However, the trades on gold may increase because the relative simplicity of gold market [15]. Fears of global recession sent a stock market plummeting in October 2008. As shown in the figure, since October 2008, gold prices have surged, indicating a positive response to the intensification of financial crisis. Gold prices show a downward trend since 2012 due to investors shift from gold market to stock market as economy around the globe improving.

Table 1 shows that the daily gold returns significantly outperform stock return over the sample period. The volatility of stock return is significantly lower than gold returns. KLCI return also exhibits more extreme positive and negative values than the gold returns. The coefficient of variation which is measured by dividing standard deviation by the mean return is lower for gold returns than the stock return. This indicates that in measuring the degree of risk in relation to return, gold gives a better risk-return tradeoff. Kurtosis exhibits a leptokurtic distribution and as clearly shown by the Jarque-Bera statistics, both gold and stock returns are not normally distributed at the 1 % significance level.

Comparing between the three different weights of gold, the results illustrate that 1/2 ounces of Kijang Emas generally exhibits slightly higher average return and risk and 1 ounce of *Kijang Emas* displays a most extreme value. On the other hand, the coefficient of variation for 1/4 ounces *Kijang Emas* is lower than the other two counterparts.

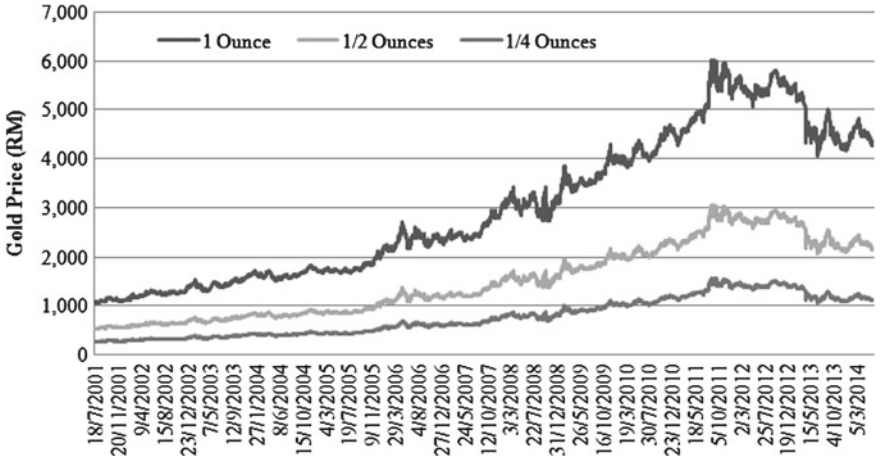


Fig. 1 The evolution of local gold prices from july 2001 to may 2014

Table 1 Descriptive statistics of daily returns

	1 ounce	1/2 ounces	1/4 ounces	KLCI
Mean	0.04195	0.04196	0.04192	0.03229
Maximum	12.46452	12.45939	12.36455	19.86049
Minimum	−11.19943	−11.18483	−11.15078	−19.24639
Standard deviation	1.17802	1.17884	1.17686	1.04034
Coefficient of variation	28.08021	28.09366	28.07182	32.21478
Skewness	−0.21607	−0.21737	−0.21563	−0.17763
Kurtosis	12.96410	12.90894	12.80586	127.2971
Jarque-Bera	13652.26***	13502.12***	13222.79***	2120499***

Notes (***) denotes significant at 1 %

We also analyze the descriptive statistics between these asset classes by quantiles of stock returns, to see if the results differ from the average during the periods of extreme stock market declines. We focus on four thresholds from the left of the return distribution i.e. 1, 2.5, 5 and 10 %. The results in Table 2 shows that the daily gold returns always outperforms stock returns during extreme negative stock market returns. With the exception of 10 % quantile, gold returns also less volatile than the stock returns in all extreme stock market conditions. In this sense, gold are more efficient asset relative to stock during this period.

Table 2 Statistics during extreme stock market returns

Panel A		Obs	Mean	Max	Min	Std.dev	CV
KLCI< 10 %	1 ounce	330	0.09409	12.46452	−7.81820	1.73658	18.45722
Quantile	1/2 ounces	330	0.09580	12.45939	−7.82660	1.73892	18.15104
	1/4 ounces	330	0.09957	12.36455	−7.74268	1.73394	17.41289
	KLCI	330	−1.53281	−0.78124	−19.2464	1.627877	−1.06202
Panel B							
KLCI< 5 %	1 ounce	165	0.02021	7.54567	−7.81820	1.92342	95.16238
Quantile	1/2 ounces	165	0.02319	7.53827	−7.82660	1.92314	82.92971
	1/4 ounces	165	0.01554	7.53603	−7.74268	1.92107	123.589
	KLCI	165	−2.12107	−1.15759	−19.2464	2.14672	−1.0121
Panel C							
KLCI< 2.5 %	1 ounce	83	−0.13213	6.72496	−7.81820	1.85340	−14.0271
Quantile	1/2 ounces	83	−0.12605	6.69395	−7.82660	1.85182	−14.6912
	1/4 ounces	83	−0.13908	6.68942	−7.74268	1.85198	−13.316
	KLCI	83	−2.88806	−1.58140	−19.2464	2.82975	−0.97981
Panel D							
KLCI< 1 %	1 ounce	33	0.02037	6.72496	−4.39708	1.95570	96.01817
Quantile	1/2 ounces	33	0.01979	6.69395	−4.38426	1.95153	98.61706
	1/4 ounces	33	0.02605	6.68942	−4.43920	1.95560	75.08251
	KLCI	33	−4.49702	−2.22536	−19.2464	4.00549	−0.8907

Notes Obs and CV are observation and coefficient of variation, respectively

Comparing between different weights of gold, the data show that 1/4 ounces *Kijang Emas* displays higher returns in the 1 and 10 % quantiles, while 1/2 ounces provide the higher return in 2.5 and 5 % quantiles. 1/4 ounces *Kijang Emas* also give the better risk-return tradeoff.[3] Meanwhile, 1 ounce *Kijang Emas* provides the highest maximum value for all quantiles. Nevertheless, in term of risk, with the exception of 10 % quantile, 1 ounce *Kijang Emas* is more volatile if compared with 1/2 and 1/4 ounces.

4.3 Econometrics Analysis

This section describes the estimation results of the asymmetric GARCH models and discusses the main findings. The results of the systematic model are presented followed by a threshold analysis of that model.

4.3.1 Systematic and Threshold Analyses Based on TGARCH Model

The main findings are presented in Tables 3 and 4. The systematic model analyzes how conditional error variance of gold reacts to shocks in average while the threshold model analyzes the reaction of conditional error variance to shocks under extreme stock market conditions. The analysis only uses observations on the periods when the stock returns below the 10, 5, 2.5 and 1 % quantiles.

The table contains three panels (1 ounce, 1/2 ounces and 1/4 ounces) of the coefficient estimates and z-statistics of the asymmetric volatility model parameters specified in Eq. (2). The estimation results can be summarized and interpreted as follows. The systematic and threshold TGARCH models exhibit highly significant of past shocks or innovations (α) and past volatilities (ρ) on volatility behaviour of gold. But the value of past volatilities (around 0.931) strongly dominates the value of past shocks (around 0.07) in all equations. This implies that past volatilities (but not the shocks) should be used to predict volatility in the future. This also suggests that the volatility is extremely persistent over time or takes a long time to die out, that is, the volatility is likely to remain high over several periods. Consequently, the volatility (information) are slowly assimilated to the gold market, thus slowly converges to long-run equilibrium. This is due to the fact that gold is a precious metal, and is influenced by factors that affect the demand for jewelry and recycling but less by shocks which have short impact duration, which normally reported in industrial metal [24].[4]

[3] Since all variables in the denominator of the calculation are negative in 2.5 % quantile, the ratio will not make sense. Therefore, we ignore the coefficient of variation for 5 % threshold.

[4] [24] opine that industrial metal is more cyclical and has relatively lower volatility persistence than gold. This is because of its lower transitory persistence partly due to the stronger impact of the short lived shocks on it.

Table 3 Estimation results of variance equation: systematic analysis in the TGARCH model

	1 ounce		1/2 ounces		1/4 ounces	
	Coefficient	z-stat	Coefficient	z-stat	Coefficient	z-stat
δ	0.01956	4.10576***	0.01943	4.12738***	0.01958	3.90454***
α	0.07209	5.81661***	0.07219	5.85104***	0.07174	5.69845***
ϑ	−0.03552	−2.72960 ∗ ∗∗	−0.03642	−2.83059 ∗ ∗∗	−0.03369	−2.51983 ∗ ∗
ρ	0.93196	105.6566***	0.93242	106.6097***	0.93122	102.5347***

Note (*), (**) and (***) denote significant at 1, 5 and 10 %, respectively

Table 4 Estimation results of variance equation: threshold analysis based on TGARCH model

		1 ounce		1/2 ounces		1/4 ounces	
		Coefficient	z-stat	Coefficient	z-stat	Coefficient	z-stat
Panel	δ	0.01965	4.11717***	0.01972	4.15080***	0.01981	3.92561***
A 10 %	α	0.07256	5.85423***	0.07270	5.87640***	0.07206	5.72002***
	ϑ	−0.03586	−2.75353 ∗ ∗∗	−0.03625	−2.80316 ∗ ∗∗	−0.03344	−2.49331 ∗ ∗
	ρ	0.93157	105.6021***	0.93156	105.9561***	0.93057	102.0996***
Panel	δ	0.01963	4.11936***	0.01987	4.16372***	0.01980	3.92797***
B 5 %	α	0.07236	5.85036***	0.07269	5.87576***	0.07214	5.72516***
	ϑ	−0.03559	−2.73992 ∗ ∗∗	−0.03585	−2.77063 ∗ ∗∗	−0.03355	−2.50401 ∗ ∗
	ρ	0.93164	105.7006***	0.93123	105.5946***	0.93057	102.1781***
Panel	δ	0.01992	4.12131***	0.02155	4.22829***	0.02123	3.99943***
C 2.5 %	α	0.07264	5.81447***	0.07474	5.81570***	0.07483	5.70874***
	ϑ	−0.03552	−2.68073 ∗ ∗∗	−0.03427	−2.53434 ∗ ∗	−0.03404	−2.43928 ∗ ∗
	ρ	0.93104	104.3244***	0.92726	99.77087***	0.92721	97.61037***
Panel	δ	0.01964	4.11796***	0.01960	4.14713***	0.01965	3.91368***
D 1 %	α	0.07223	5.83209***	0.07219	5.85194***	0.07178	570253***
	ϑ	−0.03562	−2.73853 ∗ ∗∗	−0.03610	−2.80306 ∗ ∗∗	−0.03365	−2.51518 ∗ ∗
	ρ	0.93179	105.5930***	0.93208	106.3333***	0.93110	102.4654***

Note (*), (**) and (***) denote significant at 1, 5 and 10 %, respectively

In the systematic variance model, the asymmetric coefficients are negative, highly significant and large in magnitude for 1/2 ounces (–0.03642), followed by 1 ounce (–0.03552) and 1/4 ounces (–0.03369). The asymmetric terms in the threshold models also display same results, where 1/2 ounces demonstrate the highest magnitude except in quantile 2.5 %. The negative coefficient implies that negative shocks exhibit a smaller impact on the volatility of gold than positive shocks. In other words, positive shocks of gold increase the volatility of gold more than negative shocks. For example, the increase in volatility with positive shocks (0.07) is two times larger than the increase in volatility with negative shocks (0.03). Specifically, comparing between different weights of gold, the increase in volatility with positive shock has the largest magnitude for 1/2 ounces (0.07219) in the systematic analysis. The results are consistent in threshold analysis at 5 and 10 % quantiles. While during the negative shocks

in gold returns, the increase in volatility is smaller for 12 ounces and 1 ounce in systematic and conditional analysis (for example, the effect on volatility in systematic analysis is 0.07219 (+) − 0.03642 = 0.03577 for 1/2 ounces, 0.03657 for 1 ounce and 0.03805 for 1/4 ounces).

The results reveal that the volatility of gold displays an inverted asymmetry of positive and negative shocks relative to the asymmetric volatility which normally reported in stock markets. Since financial leverage and volatility feedback cannot describe this effect, we believe this effect is related to the hedge and safe haven characteristics of gold. If the price of gold increases in times of financial or macroeconomic uncertainty, then investors buy gold and transmit the volatility and uncertainty to the gold market. The price and volatility of gold increase simultaneously. If the price of gold decreases in tranquil times, investors sell gold, thereby signaling that financial and macroeconomic uncertainty is low. This leads to a smaller increase of volatility compared with the positive gold price change [8, 10].

4.3.2 Systematic and Threshold Analysis Based on EGARCH Model

The results of EGARCH model are in line with the TGARCH model. The conditional current volatility is affected by past news (shocks), past volatility and asymmetric parameter. The parameters on the lagged conditional variances (α_4) are relatively large (around 0.985) for all models, thus the volatilities are extremely persistent and take a long time to die out.

The presence of both the standardized value of the lagged residual $\frac{\varepsilon_{t-1}}{\sqrt{h_t}}$ and and its absolute value implies that innovations (news) have asymmetric effects. Nevertheless, since the coefficients of symmetric effects and asymmetric effects are positive, the good news ($\varepsilon_{t-1} > 0$) are more destabilizing and have a greater impact on the conditional variance of error, and hence the volatility of gold, than the bad news ($\varepsilon_{t-1} < 0$). While an unanticipated fall in stock market return signals "bad" news and an unanticipated increase of stock return implies "good" news. However, it is the other way in the gold market. Positive shocks in the gold market imply "bad" financial and macroeconomic news. This demonstrates gold as a hedge and a safe haven asset during market turmoil, thus making gold a better investment for risk-averse investors (Tables 5 and 6).

Looking at the specific weights of gold, the results are consistent with the TGARCH model, where in the presence of a positive shocks, the shock impact on the conditional variance is relatively large for 1/2 ounces. For example, the effect on volatility in systematic analysis is 0.12589 + 0.02855 = 0.15444 for 1/2 ounces, 0.15337 for 1 ounce and 0.15341 for 1/4 ounces (Table 5).

Table 5 Estimation results of variance equation: systematic EGARCH model

	1 ounce		1/2 ounces		1/4 ounces	
	Coefficient	z-stat	Coefficient	z-stat	Coefficient	z-stat
α_1	−0.08875	−8.59436 ∗ ∗∗	−0.08921	−8.52451 ∗ ∗∗	−0.08962	−8.47210 ∗ ∗∗
α_2	0.12523	8.76053***	0.12589	8.71628***	0.12620	8.63656***
α_3	0.02814	2.82446***	0.02855	2.88260***	0.02721	2.65044***
α_4	0.98541	270.1471***	0.98534	271.3851***	0.98570	261.0569***

Notes (*), (**) and (***) denote significant at 1, 5 and 10 %, respectively

Table 6 Estimation results of variance equation: threshold analysis based on EGARCH model

		1 ounce		1/2 ounces		1/4 ounces	
		Coefficient	z-stat	Coefficient	z-stat	Coefficient	z-stat
Panel A	α_1	–0.08913	–8.64037***	–0.08806	–8.54827***	–0.08904	–8.52184***
10 %	α_2	0.12572	8.80592***	0.12411	8.74695***	0.12517	8.68702***
	α_3	0.02895	2.90993***	0.02944	3.01405***	0.02818	2.77009***
	α_4	0.98534	269.6158***	0.98553	276.1395***	0.98581	264.0566***
Panel B	α_1	–0.08866	–8.59083***	–0.08858	–8.52649***	–0.08932	–8.48081***
5 %	α_2	0.12511	8.75951***	0.12489	8.72369***	0.12571	8.64800***
	α_3	0.02856	2.87184***	0.02885	2.93541***	0.02758	2.69584***
	α_4	0.98544	270.6441***	0.98547	274.4849***	0.98575	262.5662***
Panel C	α_1	–0.08674	–8.56030***	–0.08942	–8.52505***	–0.08949	–8.46603***
2.5 %	α_2	0.12225	8.72956***	0.12617	8.71438***	0.12596	8.62899***
	α_3	0.02936	2.99567***	0.02846	2.86814***	0.02757	2.68149***
	α_4	0.98586	277.4642***	0.98528	270.7433***	0.98574	261.4751***
Panel D	α_1	–0.08688	–8.57104***	–0.08728	–8.47284***	–0.08795	–8.45290***
1 %	α_2	0.12242	8.74469***	0.12307	8.67556***	0.12365	8.62355***
	α_3	0.02925	2.98944***	0.02901	2.97494***	0.02805	2.77313***
	α_4	0.98584	277.3846***	0.98566	277.6297***	0.98602	267.1033***

Notes (*), (**) and (***) denote significant at 1, 5 and 10 %,respectively

5 Conclusion

This study empirically examines the volatility dynamics of gold returns in Malaysia via TGARCH and EGARCH models. At a first glance, we perceive that current conditional volatility of gold prices is significantly impacted by past shocks (news) and past volatilities. Furthermore, we find that the volatility of gold returns display an asymmetric reaction to positive and negative shocks, which can be characterized as inverted compared with the findings for the volatility in stock markets. The findings of this study are also in line with [8] that identifies evidence of asymmetric volatility in the US, British, European, Switzerland and Australia; thus supporting the argument that domestic and international gold prices follow the same volatility characteristics. This is due to the price of domestic gold is determined by the prevailing international

gold market price. Therefore, the usual explanation for asymmetric volatility, that is, financial leverage and volatility feedback, cannot be applied to gold. The finding is primarily related to the hedge and safe haven characteristics of gold. One of the reasons why gold can be a good hedge and safe haven tool is because it is considered as a homogeneous class of asset, whose returns are driven by an unobservable factor. Gold, unlike property, is easily traded in a continuously open market. Nevertheless, since rising interest rates have a dampening effect on the gold market, economic policy makers around the world can pursue tightening monetary policy to dampen volatilities. On comparison of the strength of asymmetric volatility effect between different weights of gold, we find that 1/2 ounces of gold exhibits a slightly larger magnitude if compared with 1 ounce. In other words, this result provides evidence that 1/2 ounces of gold gives more return on average and during the financial turmoil, thus encourages investors accumulate gold over time or retain a capacity to sell their gold investment in small amounts in the future.

Acknowledgments The authors would like to acknowledge the Fundamental Research Grant Scheme 203/PSOSIAL/6711417 by Ministry of Education Malaysia and Universiti Sains Malaysia.

References

1. Alexander, C.: Practical Financial Econometrics. Wiley Ltd, Chichester (2009)
2. Alexander, C., Lazar, E.: Normal mixture GARCH (1,1): application to exchange rate modelling. J. Appl. Econ. **21**(3), 307–336 (2006)
3. Bae, K.-H., Karolyi, G. A., Stulz, R. M.: A new approach to measuring financial contagion. Rev. Financ. Stud. **16**(3), 717–763 (2003)
4. Baillie, R.T., Bollerslev, T.: The message in daily exchange rates: a conditional-variance tale. J. Bus. Econ. Stat. **7**(3), 297–305 (1989)
5. Baillie, R.T., Bollerslev, T.: Intra-day and inter-market volatility in foreign exchange rates. Rev. Econ. Stud. **58**(3), 565–585 (1991)
6. Batten, J.A., Ciner, C., Lucey, B.M.: The macroeconomic determinants of volatility in precious metals markets. Resour. Policy **35**(2), 65–71 (2010)
7. Batten, J.A., Lucey, B.M.: Volatility in the gold futures market. Appl. Econ. Lett. **17**(2), 187–190 (2010)
8. Baur, D.G.: Asymmetric volatility in the gold market. J. Altern. Invest. **14**(4), 26–38 (2012)
9. Baur, D.G., Lucey, B.M.: Is gold a hedge or a safe haven? An analysis of stocks, bonds and gold. Financ. Rev. **45**(2), 217–229 (2010)
10. Baur, D.G., McDermott, T.K.: Is gold a safe haven? International evidence. J. Banking Finan. **34**(8), 1886–1898 (2010)
11. Black, F.: Studies of stock price volatility changes. Paper presented at the proceedings of the 1976 meetings of the american statistical association, Bus. Econ. Stat. (1976)
12. Bollerslev, T.: Generalized autoregressive conditional heteroskedasticity. J. Econ. **31**(3), 307–327 (1986)
13. Campbell, J.Y., Hentschel, L.: No news is good news: an asymmetric model of changing volatility in stock returns. J. Financ. Econ. **31**(3), 281–318 (1992)
14. Christie, A.A.: The stochastic behavior of common stock variances: value, leverage and interest rate effects. J. Financ. Econ. **10**(4), 407–432 (1982)
15. Dee, J., Li, L., Zheng, Z.: Is gold a hedge or a safe haven? Evidence from inflation and stock market. Int. J. Dev. Sustain. 2(1): (In Press) (2013)

16. Drakos, A.A., Kouretas, G.P., Zarangas, L.P.: Forecasting financial volatility of the Athens stock exchange daily returns: An application of the asymmetric normal mixture GARCH model. Int. J. Finan. Econ. **15**(4), 331–350 (2010)
17. Engle, R.F.: Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. Econometrica **50**(4), 987–1007 (1982)
18. Engle, R.F., Ito, T., Lin, W.-L.: Meteor showers or heat waves? Heteroskedastic intra-daily volatility in the foreign exchange market. Econometrica **58**(3), 525–542 (1990)
19. French, K.R., Schwert, G.W., Stambaugh, R.F.: Expected stock returns and volatility. J. Finan. Econ. **19**(1), 3–29 (1987)
20. Gau, Y-F., Engle, R.F.: Conditional volatility of exchange rates under a target zone. Department of Economics Discussion Paper Series 06. University of California. San Diego (1997)
21. Ghazali, M.F., Lean, H.H., Bahari, Z.: Is gold a hedge or a safe haven? An empirical evidence of gold and stocks in Malaysia. Int. J. Bus. Soc. **14**(3), 428–443 (2013)
22. Giamouridis, D.G., Tamvakis, M.N.: The relation between return and volatility in the commodity markets. J. Altern. Invest. **4**(1), 54–62 (2001)
23. Glosten, L.R., Jagannathan, R., Runkle, D.E.: On the relation between the expected value and the volatility of the nominal excess return on stocks. J. Financ. **48**(5), 1779–1801 (1993)
24. Hammoudeh, S., Yuan, Y.: Metal volatility in presence of oil and interest rate shocks. Energ. Econ. **30**(2), 606–620 (2008)
25. Hillier, D., Draper, P., Faff, R.: Do precious metals shine? Invest. Perspect. Finan. Anal. J. **62**(2), 98–106 (2006)
26. Hirayama, K., Tsutsui, Y.: Threshold effect in international linkage of stock prices. Jpn. World Econ. **10**(4), 441–453 (1998)
27. Hsieh, D.A.: Modeling heteroscedasticity in daily foreign-exchange rates. J. Bus. Econ. Stat. **7**(3), 307–317 (1989)
28. Hsieh, D.A.: Testing for nonlinear dependence in daily foreign exchange rates. J. Bus. **62**(3), 339–368 (1989)
29. Jeon, B.N., Furstenberg, G.M.V.: Growing international co-movement in stock price indexes. Q. Rev. Econ. Financ. **30**(3), 15–30 (1990)
30. Johnston, K., Scott, E.: GARCH models and the stochastic process underlying exchange rate price changes. J. Financ. Strateg. Decis. **13**(2), 13–24 (2000)
31. Liao, S-J., Chen, J.T.: The relationship among oil prices, gold prices and the individual industrial sub-indices in Taiwan. Paper presented at the Int. Conf. Bus. Inf. Seoul, South Korea (2008)
32. Lobo, B.J.: Asymmetric effects of interest rate changes on stock prices. Financ. Rev. **35**(3), 125–144 (2000)
33. Nelson, D.B.: Conditional heteroskedasticity in asset returns: a new approach. Econometrica **59**(2), 347–370 (1991)
34. Roache, S.K., Rossi, M.: The effects of economic news on commodity prices: is gold just another commodity? IMF Working Paper WP 09/140 (2009)
35. Sari, R., Soytas, U., Hacihasanoglu, E.: Do global risk perceptions influence world oil prices? Energy Econ. **33**(3), 515–524 (2011)
36. Sumner, S., Johnson, R., Soenen, L.: Spillover effects among gold, stocks, and bonds. J. CENTRUM Cathedra **3**(2), 106–120 (2010)
37. Tully, E., Lucey, B.M.: A power GARCH examination of the gold market. Res. Int. Bus. Financ. **21**(2), 316–325 (2007)

Quantile Regression Under Asymmetric Laplace Distribution in Capital Asset Pricing Model

Kittawit Autchariyapanitkul, Somsak Chanaim and Songsak Sriboonchitta

Abstract We used a quantile regression under asymmetric Laplace distribution for predicting stock returns. Specifically, we apply this method to the classical capital asset pricing model (CAPM) to estimate the beta coefficient which measure risk in the portfolios management analysis at given levels of quantile. Quantile regression estimation is equivalent to the parametric case where the error term is asymmetrically Laplace distributed. Finally, we use the method to measures the volatility of a portfolio relative to the market.

1 Introduction

Capital asset pricing model is the tool for evaluating portfolios investment. The basic concept behind CAPM is that investors need to be compensated time value of money and risk. The time value of money is represented by the risk-free rate (R_F) and compensates the investors for placing money in any investment over a period of time rather than put money into the risk free rate asset (e.g. government bonds, T-bill). The rest of the model represents risk and find the value of compensation the investor needs for taking on additional risk when they choose to invest in risky asset (e.g. stocks, corporate bonds). This is calculated by taking a risk measure that compares the returns of the asset to the market over a period of time and to the market premium ($R_M - R_F$).

K. Autchariyapanitkul (✉) · S. Sriboonchitta
Faculty of Economics, Chiang Mai University, Chiang Mai, Thailand
e-mail: kittawit_autchariya@cmu.ac.th

S. Chanaim
Department of Mathematics Faculty of Science, Chiang Mai University, Chiang Mai, Thailand

V.-N. Huynh et al. (eds.), *Econometrics of Risk*, Studies in Computational Intelligence 583, DOI 10.1007/978-3-319-13449-9_15

Normally in CAPM, asset returns are imposed normally distributed random variables. But this is not seem to be correct. For example, high peak, large swings, swings as big as 2 to 4 standard deviation from the standard mean, occur in the market more regularly than we would expect in a normal distribution. CAPM assumes that the variance of returns adequately measures risk. This may be true if returns were distributed normally. However other risk measurements are probably better for showing investors' preferences. In this paper we propose to use quantile regression with asymmetric Laplace distribution (ALD).

Quantile regression is a very popular tool since the distinguished work of Koenker and Gilbert [3]. Unlike mean regression model, quantile regression can capture the entire conditional distribution of the outcome variable, and is more robust to outliers and not satisfied the error distribution. For quantile regression to modelling and testing the CAPM, the reader is referred to e.g., Barnes and Hughes [1]. In their studied, that the market price of beta risk is significant in both tails of the conditional distribution of returns. In Chen et al. [2], the authors used a couple of methods to obtain the time varying market betas in CAPM to analyze stock in the Dow Jones Industrial for several quantiles. The results indicated that smooth transition quantile CAPM-GARCH model is strongly preferred over the method of sharp threshold transition and a symmetric CAPM-GARCH model.

We present a likelihood-based approach to estimation of the regression quantiles based on the asymmetric Laplace distribution. In Sánchez et al. [10], the paper investigated the distribution under the asymmetric Laplace law by using currency exchange rates give that the ALD is successful in capturing the peakedness, leptokurticity, and skewness, inherent in such data. In Kotz and Drop [5], the authors studied refinements of the project evaluation and review technique, and developed a reparameterization of the ALD and found that it is a useful tool for extending and improving various three-point approximations of continuous distributions by specifying the values of two quantiles and the mode. Similarly, the works from Linden [6] showed that the Laplace distribution has a geometric stability for the weekly and the monthly distribution of the stock return and also ignores the high peak, the fat tail and the skewness of the return but ALD is insufficient enough to measures the negative skew, but it explains well with the positive skewness of the stock returns for S&P500 and FTSE100 although it captures the high peak and the fat tail of the stock return.

It should be noted that a complete study of QR models with various error distributions is of great interests for applications in financial analysis.

The remainder of the paper is organized as follows. Section 2 provides a basics knowledge of quantile regression and asymmetric Laplace distribution. Section 3 provides the theoretical concepts of the prediction. Section 4 discusses the empirical discovering and the solutions of the forecasting problem. The last section gives the conclusion and extension of the paper.

2 Quantile Regression Model

First, recall the classical situation of linear mean regression model: suppose we can observe another variable X and wish to use X to predict Y, a predictor of Y is some function of X, say, $\varphi(X)$. If we use mean square error (MSE), then we seek $\varphi^*(X)$ minimizing $E[Y - \varphi(X)]^2$ over all $\varphi(X)$. With respect to this square loss function, it is well-known that the conditional mean $E(Y|X)$ is optimal.

Now, since random evolution do not obey specific "laws of motion" as dynamical systems in physics, we need to rely on "plausible statistical models" to proceed. A plausible model for $E(Y|X)$ is the linear model θX where θ is some unknown constant. The associated (additive noise) statistical model is

$$Y = \theta' X + \varepsilon, \tag{1}$$

where ε is a random variable representing random error due to all sources of randomness other than X. For this model (for regressing Y from X) to be "consistent" with $E(Y|X) = \theta X$, we must have $E(\varepsilon|X) = 0$, since from $Y = \theta X + \varepsilon$, we have

$$E(Y|X) = E(\theta X|X) + E(\varepsilon|X) = \theta X + E(\varepsilon|X). \tag{2}$$

Mean linear models are motivated by best predictor $E(Y|X)$ in the sense of MSE. When we consider Least Absolute Deviation (LAD) instead, we talk about *median regression*, and in particular linear median regression. Then we have to consider about conditional median of Y given X, denoted as $q_{1/2}(Y|X)$. Essentially, we replace the familiar concept of conditional expectation (mean) by the concept of conditional median and more generally conditional quantiles. (Like conditional mean, conditional quantiles are random variables whose existence also follows from Radon-Nikodym theorem in probability theory.)

Recall that the $\alpha - quantile$ of a distribution F of a random variable X is $F^{-1}(\alpha) = \inf\{y \in \mathbb{R} : F(y) \geq \alpha\}$ which we write as $q_\alpha(Y)$.

Unlike the expectation operator, $q_\alpha(.)$ is not additive but for $a > 0$ and $b \in \mathbb{R}$, we do have,

$$q_\alpha(aX + b) = aq_\alpha(X) + b. \tag{3}$$

Let (X, Y) be a random vector. The conditional distribution of Y given $X = x$ is $F_{Y|X=x}(y) = P(X \leqslant y|X = x)$. The $\alpha - quantile$ of Y given $X = x$ is called the conditional $\alpha - quantile$ of Y given $X = x$, and is simply the $\alpha - quantile$ of the conditional distribution $F_{Y|X=x}$ so that $P(Y \leqslant q_\alpha(Y|x)|X = x) = \alpha$.

The conditional distribution of Y given X is

$$F_{Y|X}(x) = P(Y \leqslant y|X) = E[1_{Y \leqslant y}|Y]. \tag{4}$$

The $\alpha - quantile$ of X given Y is defined to be

$$q_\alpha(Y|X) = \inf\{y \in \mathbb{R} : F_{Y|X}(y) \geqslant \alpha\}. \tag{5}$$

Thus, since $E[1_{Y \leqslant y}|X]$ is a random variable, so is $q_\alpha(Y|X)$ which is a function of X, referred to as a *quantile regression function*. Given in this equation,

$$q_\alpha(Y|X) = F^{-1}_{Y|X}(\alpha) = \varphi_\alpha(X) = \theta_\alpha(X). \tag{6}$$

Just like the median (for special case of $\alpha = 1/2$) and similar to conditional mean, the quantile $q_\alpha(Y|X)$ is the best predictor for Y based on X (i.e., as a function of X) in the sense of the predictor error $E\rho_\alpha(Y - (\varepsilon|X))$, where $\rho_\alpha(u) = u(\alpha - 1_{(u<0)})$. i.e., $q_\alpha(Y|X)$ minimizes $E\rho_\alpha(Y - (\varepsilon|X))$ over all possible $\varphi(X)$. Indeed, using the same proof for unconditional quantiles, $q_\alpha(Y|X = x)$ minimizes $E[\rho_\alpha(Y - a)|X = x]$ so that the function $x \to q_\alpha(Y|X = x)$ minimizes $E\rho_\alpha(Y - (\varepsilon|X))$.

In general, the "quantile regression function" $q_\alpha(Y|X)$(which is a function of X) is nonlinear. A model for it is the linear model $q_\alpha(Y|X) = X'\beta_\alpha$, where β_α is a vector $(k \times 1)$ of unknown parameters of interest. This is what we call quantile regression, which is in fact a semi-parameter linear model relating a response variable Y to the explanatory variable X via "a quantile parameter". We need an associated statistical model representing (consistently) this linear quantile model. If we denote by ε_α that disturbance in the relationship between Y and $X'\beta_\alpha$, then we could write the quantile regression model as

$$Y = X'\beta_\alpha + \varepsilon_\alpha. \tag{7}$$

From this model we have, symbolically,

$$Y|X = (X'\beta_\alpha + \varepsilon_\alpha). \tag{8}$$

We have

$$q_\alpha(Y|X) = q_\alpha[(X'\beta_\alpha + \varepsilon_\alpha)|X] = q_\alpha(X'\beta_\alpha + \varepsilon_\alpha|X) = X'\beta_\alpha + q_\alpha(\varepsilon_\alpha|X), \tag{9}$$

and since given X, $X'\beta_\alpha$ is a constant. Thus, $q_\alpha(\varepsilon_\alpha|X) = 0$ the counterpart of the standard condition $E(\varepsilon|X) = 0$ in mean linear regression model.

Note that $q_\alpha(\varepsilon_\alpha|X) = 0$ means that 0 is the α-conditional quantile of the "noise" ε_α, i.e.,

$$P(\varepsilon_\alpha \leqslant 0|X) = \alpha. \tag{10}$$

If ε_α is independent of X, then the α-quantile of the noise ε_α is zero, that is, $\int_{-\infty}^{0} dF_{\varepsilon_\alpha}(u) = \alpha$. For $q_\alpha(Y|X) = X'\beta_\alpha$, we see that β_α minimize $E[\rho_\alpha(Y - X'\beta)]$

over β. Thus, given i.i.d $(X_i, Y_i), i = 1, 2, \ldots, n$, a plausible estimator of β_α proceeds by minimizing

$$\widehat{\beta_\alpha} = \arg\min \frac{1}{n} \left\{ \sum_{i=1}^{n} \rho_\alpha(Y_i - X_i'\beta) \right\}, \tag{11}$$

$\rho_\alpha(\cdot)$ is so called check (or loss) function defined by $\rho_\alpha(u) = u(\alpha - 1_{(u<0)})$, with $1_{(u<0)}$ denoting the usual indicator function and this estimator is called the LAD estimator.

Thus, suppose that the error term ε_α is distributed as an asymmetric Laplace distribution (ALD) then the LAD estimator of β_α is a MLE. For the MLE, the minimization of $E\rho_\alpha(Y_i - X_i'\beta)$ is the same as maximization of ALD. Given $Y \sim ALD(X_i'\beta_\alpha, \sigma, \alpha), i = 1, 2, ..., n$ are independent. Then, the likelihood function for β, σ (see, Koenker [4] and Sánchez [10]) is

$$L(\beta_\alpha, \sigma|(X_i, Y_i) = \frac{\alpha^n(1-\alpha)^n}{\sigma^n} \exp\left\{ -\sum_{i=1}^{n} \rho_\alpha\left(\frac{Y_i - X_i'\beta_\alpha}{\sigma}\right)\right\}. \tag{12}$$

3 Validating Linear Quantile Models

To see whether the linear model is a good approximation of $\varphi(.)$, we can use variance as a measure of variation and hence as we already know the ratio $\frac{a^2 Var(X)}{Var(Y)} \in [0, 1]$ can be used as indication of goodness-of-fit: the higher this ratio, the more adequate the linear model, in the sense that the linear model captures reasonable well the relationship between X and Y. A consistent of ratio $\frac{a^2 Var(X)}{Var(Y)}$ is simply the ratio of $\frac{\Sigma_i(\hat{y}_i - \bar{y})^2}{\Sigma_i(y_i - \bar{y})^2}$ which is what we call R^2 (coefficient of determination). When X and Y have finite variance, R^2 is a consistent estimator of $\frac{a^2 Var(X)}{Var(Y)}$, i.e., sufficiently close to $\frac{a^2 Var(X)}{Var(Y)}$, as $n \to \infty$, with probability one.

Consider now a linear quantile regression model. To validate such a model, it is natural to use the idea of coefficient of determination in linear mean regression. In general a link between X and Y is given by $\varphi(X) = E(Y|X)$. The nonparametric coefficient of determination is

$$R^2 = \frac{Var\varphi(X)}{Var(Y)} = 1 - \frac{E|Y - \varphi(X)|^2}{E|Y - EY|^2}. \tag{13}$$

Which is used for assessing the prediction power of a mean regression model. In linear model, $\varphi(X) = aX$.

Note that R^2 cannot answer questions such as "Does X exert any utilize any significant effect on the tail of the distribution of Y?" Thus, in risk management where large values of Y are of interest, we should use quantile regression instead.

In the LAD sense, the best predictor of Y given X, at the $\alpha - quantile$ level, is the conditional quantile $q_\alpha(Y|X)$ which is a function of X, as a conditional expectation. A validation measure for general quantile regression model is

$$Q(\alpha) = 1 - \frac{E[\rho_\alpha(Y - q_\alpha(Y|X))]}{E[\rho_\alpha(Y - q_\alpha(Y))]}, \tag{14}$$

by replacing $E(Y|X)$ and $E(Y)$ by $q_\alpha(Y|X)$ and $q_\alpha(Y)$, respectively. The empirical $Q_n(\alpha)$ is obtained by estimated quantiles. We have

$$Q_n(\alpha) = 1 - \frac{\sum_{i=1}^{n}[\rho_\alpha(Y_i - q_\alpha(Y_i|X_i))]}{\sum_{i=1}^{n}[\rho_\alpha(Y_i - \hat{\theta}_\alpha X_i)]}. \tag{15}$$

For more results on this measure for validating quantile regression models, including asymptotics, we refer the reader to the paper by Noh et al. [9].

4 An Application to the Stock Market

4.1 Capital Asset Pricing Model:CAPM

The Capital Asset Pricing Model (CAPM) was developed by Sharpe [11] and John Lintner [7]. The CAPM measures the sensitivity of the expected excess return on security to expected market risk premium. The equation of CAPM is a linear function of the security market line:

$$E(R_A) - R_F = \beta_0 + \beta_1 E(R_M - R_F). \tag{16}$$

where $E(R_A)$ is the expected return of the asset, R_M is the expected market portfolio return, β_0 is the intercept and R_F is the risk free rate. $E(R_M - R_F)$ is the expected risk premium, and β_1 is the equity beta, denoting market risk. To measure the systematic risk of each stock via the beta take form:

$$\beta_1 = \frac{cov(R_A, R_M)}{\sigma_M^2}. \tag{17}$$

where σ_M^2 is the variance of the expected market return. Given that, the CAPM predicts portfolio's expected return should be relative to its risk and the market return. In this paper, we calculate the beta coefficients under the ALD assumption by using quantile regressions via belief function procedures.

4.2 Beta estimation

From above conventional method in equation (16), we estimate the β coefficient through the quantile with ALD instead. Suppose we have observed the historical return of stock $R_A = (r_{a1}, \ldots, r_{an})$ and return from market $R_M = (r_{m1}, \ldots, r_{mn})$ at specific location over n years. These observations will be assumed an i.i.d. random innovation from $ALD(\alpha, \mu, \sigma)$. In this case, we consider $\mu = 0$. From quantile equation.

$$R_A = R'_M \beta_\alpha + \varepsilon_\alpha \tag{18}$$

So that,

$$\varepsilon_\alpha = R_A - R'_M \beta_\alpha \tag{19}$$

Then, probability density function is

$$f(\varepsilon_\alpha, \sigma) = \frac{\alpha(1-\alpha)}{\sigma} \exp\{-\rho_\alpha(\frac{\varepsilon_\alpha}{\sigma})\} \tag{20a}$$

$$= \frac{\alpha(1-\alpha)}{\sigma} \exp\{-(\frac{\varepsilon_\alpha(\alpha - 1_{\varepsilon_\alpha<0})}{\sigma})\}. \tag{20b}$$

And CDF is

$$F(\varepsilon_\alpha) = \int_{-\infty}^{\varepsilon_\alpha} \frac{\alpha(1-\alpha)}{\sigma} \exp\{-(\frac{\varepsilon_\alpha(\alpha - 1_{\varepsilon_\alpha<0})}{\sigma})\} d\varepsilon_\alpha \tag{21a}$$

$$= \int_{-\infty}^{\varepsilon_\alpha} -\frac{\alpha(1-\alpha)}{\alpha - 1_{\varepsilon_\alpha<0}} \exp\{-(\frac{\varepsilon_\alpha(\alpha - 1_{\varepsilon_\alpha<0})}{\sigma})\} d(\cdot) \tag{21b}$$

$$= -\frac{\alpha(1-\alpha)}{\alpha - 1_{\varepsilon_\alpha<0}} \exp\{-(\frac{\varepsilon_\alpha(\alpha - 1_{\varepsilon_\alpha<0})}{\sigma})\}|_{-\infty}^{\varepsilon_\alpha}. \tag{21c}$$

where $d(\cdot) = -\dfrac{\varepsilon_\alpha(\alpha - 1_{\varepsilon_\alpha<0})}{\sigma}$

We get

$$F(\varepsilon_\alpha) = \begin{cases} \alpha \cdot \exp\left[\dfrac{(1-\alpha)\varepsilon_\alpha}{\sigma}\right] & : \varepsilon_\alpha < 0 \\ 1 + (\alpha - 1) \cdot \exp\left[\dfrac{-\alpha\varepsilon_\alpha}{\sigma}\right] & : \varepsilon_\alpha \geq 0 \end{cases} \tag{22}$$

If we need random number for this CDF then, let $F(\varepsilon_\alpha) = u \sim \text{uniform}(0, 1)$

$$u = \begin{cases} \alpha \cdot \exp\left[\dfrac{(1-\alpha)\varepsilon_\alpha}{\sigma}\right] & : \varepsilon_\alpha < 0 \\ 1 - (1 - \alpha) \cdot \exp\left[\dfrac{-\alpha\varepsilon_\alpha}{\sigma}\right] & : \varepsilon_\alpha \geq 0 \end{cases} \tag{23}$$

Table 1 Data descriptive and statistics

	MUR	XOM	S&P500
Mean	–0.0009	–0.0011	0.0024
Median	–0.0019	–0.0011	0.0032
Maximum	0.0756	0.1183	0.0713
Minimum	–0.0729	–0.1534	–0.0746
Std. Dev	0.0248	0.0429	0.0218
Skewness	–0.1848	–0.3827	–0.3864
Kurtosis	3.8899	4.4453	4.4495
DW-test	1.8838	1.8755	
Obs	209		

Source All values are the log return

$$F^{-1}(u) = \varepsilon_\alpha = \begin{cases} \dfrac{\sigma(\ln u - \ln \alpha)}{(1-\alpha)} & : 0 < u \leq \alpha \\ -\dfrac{\sigma}{\alpha}(\ln \dfrac{(1-u)}{(1-\alpha)}) & : \alpha < u < 1. \end{cases} \tag{24}$$

Using (12), (16) the corresponding to the observed data through the CAPM model using quantile regression with ALD is a realization that generate likelihood function is

$$L_\alpha(\beta_0^\alpha, \beta_1^\alpha, \sigma|(r_{mi}, r_{ai})) = \frac{\alpha^n(1-\alpha)^n}{\sigma^n} \exp\left\{-\sum_{i=1}^{n} \rho_\alpha\left(\frac{r_{ai} - r_{mi}\beta_1^\alpha - \beta_0^\alpha}{\sigma}\right)\right\} \tag{25}$$

4.3 Empirical Results

The data contain of 209 weekly returns during 2010–2013 are obtained from Yahoo to compute the log returns on the following securities. Integrated oil and gas company—A company that participates in every aspect of the oil or gas business, which includes the discovering, obtaining, producing, refining, and distributing oil and gas. The integrated oil and gas company in this paper contains of two companies: Exxon Mobil Corp. (XOM) and Murphy Oil (MUR). Due to high turn over volume and market capitalization. Table 1 displays a summary of the variables.

The appropriate risk-free rate for the CAPM in this paper, we use Treasury bills -the bill with the shortest maturity not less than one month as a proxy belong to Chen et al. [2] and Mukherji [8] indicated that Treasury bills are better proxies for the risk-free rate than longer-term Treasury securities regardless of the investment horizon, only related to the U.S. market.

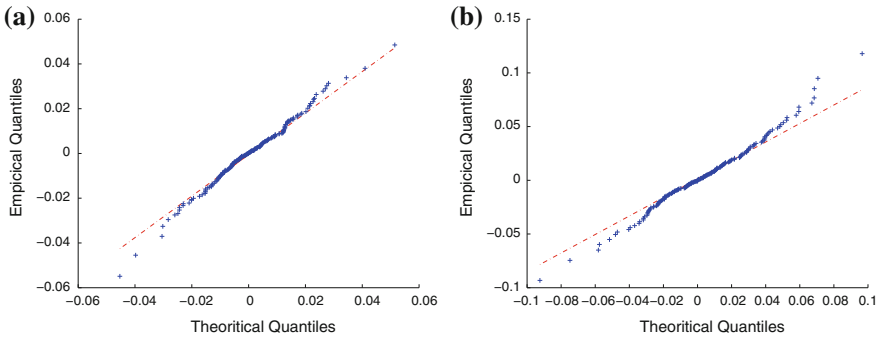


Fig. 1 The Q-Q plots of error from asymmetric Laplace distribution. **a** *ALD* Q-Q plot of XOM $\alpha = 0{:}50$. **b** *ALD* Q-Q plot of MUR $\alpha = 0{:}50$

The daily returns of the treasury bills are adjusted to the weekly returns and be used in this manner by using compound interest that take form:

$$I_{wj} = \left\{ \prod_{i=1}^{N} (1 + I_{di}) \right\} - 1 \tag{26}$$

where $I_{wj}, j = 1, 2, \ldots, N$ is the weekly interest rate and $I_{di}, i = 1, 2, \ldots, N$ is the daily interest rate.

The Q-Q plots shown in Fig. 1 are based on the distribution of AL, given in (25). In support of the Kolmogorov-Smirnov test (KS-test) gives the value of MUR, $D_n = 0.0425$ and the value of XOM, $D_n = 0.0500$ compare with the critical value $\frac{K(\alpha'=0.05)}{\sqrt{n}} = 0.0941$ confirm that none of the marginals rejects the null hypotheses of the KS at the 5 % level. The lines in these figures represent the 0.5th quantile. These figures and KS values are clearly show that the asymmetric Laplace distribution provides a good-fit to the sample data set.

The values in the Table 2 exhibit the results of the parameters estimation for the CAPM via the quantile regression under asymmetric Laplace distributions at given level of α. The results for ALD assumption performs well for the given quantile of these two stocks.

Now, Fig. 2 shows parameter estimation for the entire quantile. From these picture, we summarize the results as follows:

1. The intercepts $\beta_0^{0.5}$ increase with α, and are negative under low quantile levels and positive under high quantile levels. They are close to 0 under $\alpha = 0.5$.

2. XOM has less risky than the market except for the quantile lower than 0.07. And more risk less at higher quantile levels, The risks decrease as the stock return increase, but the relationship is non-monotonic.

3. MUR is more risky than the market under all quantile levels, and more risk under bull market than under bear market. We cannot find any monotonic relationship between risks and excess returns for MUR.

Table 2 Parameter estimated results

Parameter	Stock Name	$\alpha = 0.10$	$\alpha = 0.40$	$\alpha = 0.50$	$\alpha = 0.60$
β_0	XOM	–0.0155	–0.0032	–0.0006	0.0023
		(0.0002)	(0.0001)	(0.0001)	(0.0000)
	MUR	–0.0290	–0.0073	–0.0010	0.0049
		(0.0000)	(0.0003)	(0.0000)	(0.0001)
β_1	XOM	0.9479	0.9344	0.9319	0.9234
		(0.0572)	(0.0029)	(0.0012)	(0.0021)
	MUR	1.5355	1.4734	1.4588	1.3877
		(0.0006)	(0.0017)	(0.0003)	(0.7638)
σ	XOM	0.0024	0.0048	0.0050	0.0049
		(0.0002)	(0.0009)	(0.0003)	(0.0004)
	MUR	0.0046	0.0098	0.0102	0.0101
		(0.0003)	(0.0039)	(0.0005)	(0.0007)
$Q_n(\alpha)$	XOM	0.4836	0.4699	0.4716	0.4603
	MUR	0.4348	0.3725	0.3648	0.3500
LL	XOM	547.7220	606.6534	607.8101	603.1395
	MUR	414.6052	458.8301	459.2612	454.0436

Consistent standard errors() is in parenthesis

Most quantile levels have asymmetric behavior in market betas. Hence, we refer that the data have asymmetric effect. Notice that risk parameter (β_1) for some quantile at the lower and higher regime conflict with others points because that may be come from the ALD not capture well at the tails of distribution.

4.4 Measures the volatility of stock

After we get all the parameters estimated. We plug them into (16). In Fig. 3: The lower line is $\alpha = 0.05$, the middle line is $\alpha = 0.50$ and the upper line is $\alpha = 0.95$. The slope of the line, which is a measure of systematic risk (β_1), determines the tradoff between risk and return. The high beta may be appropriate for high risk aggressive investors. The other way around, low beta may be suitable for low risk defensive investors. It quite be crucial that if an investment were to lie above or below that straight line, then an opportunity for riskless arbitrage would exist.

5 Conclusions and Extension

In this paper, we demonstrate our method of quantile CAPM with ALD for the two stocks in $S\&P500$. This method can be used to study the linear relationship between the expected returns on a stock and its asymmetric market risk over various quantile levels. However, only a systematic risk is calculated through the model, we neglect

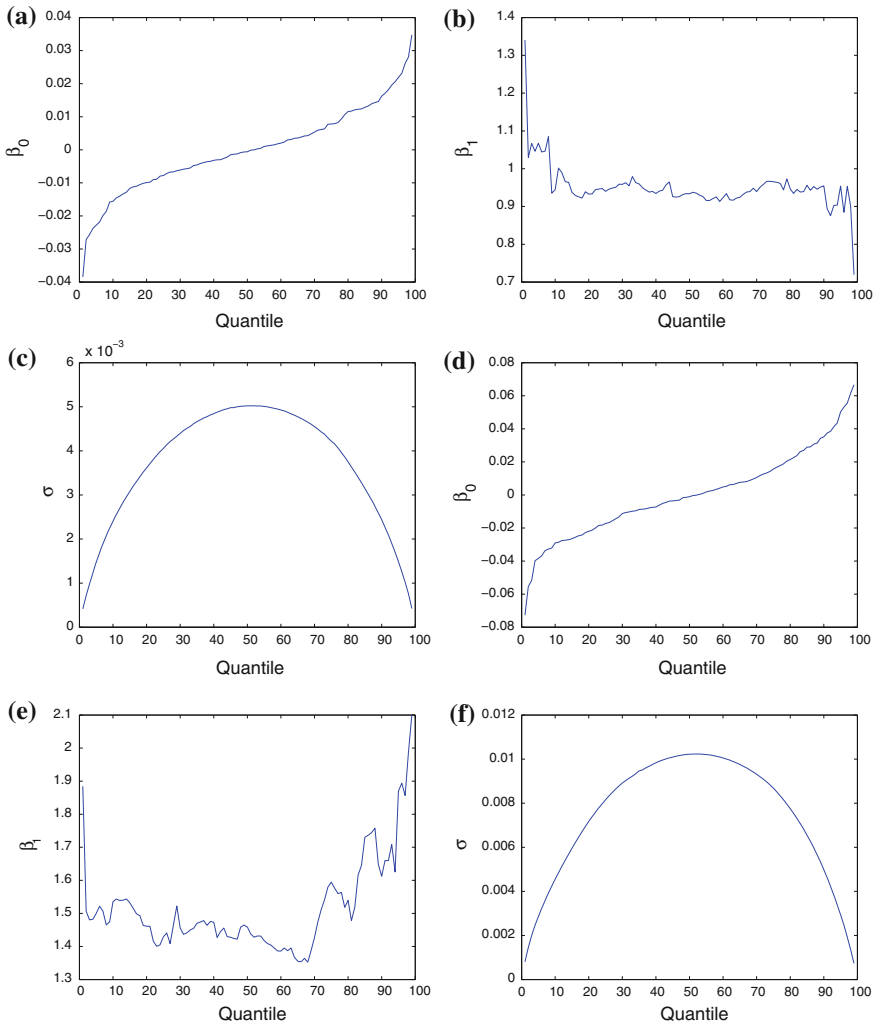


Fig. 2 Marginal parameter plots of two stocks at difference quantile(α). **a** β_0 XOM, **b** β_1 XOM, **c** σ XOM, **d** β_0 MUR, **e** β_1 MUR, **f** σ MUR

the unsystematic risk under CAPM assumption. CAPM concludes that the expected return of a security or a portfolio equals the rate on a risk-free security plus a risk premium. If this expected return does not meet or beat the required return, then the investment should not be undertaken.

The empirical diagnostic exhibit that this method captures the stylized factors in financial data to describe the stock returns under most quantile levels, especially under the middle quantile levels. Clearly, there is no monotonic relationship between

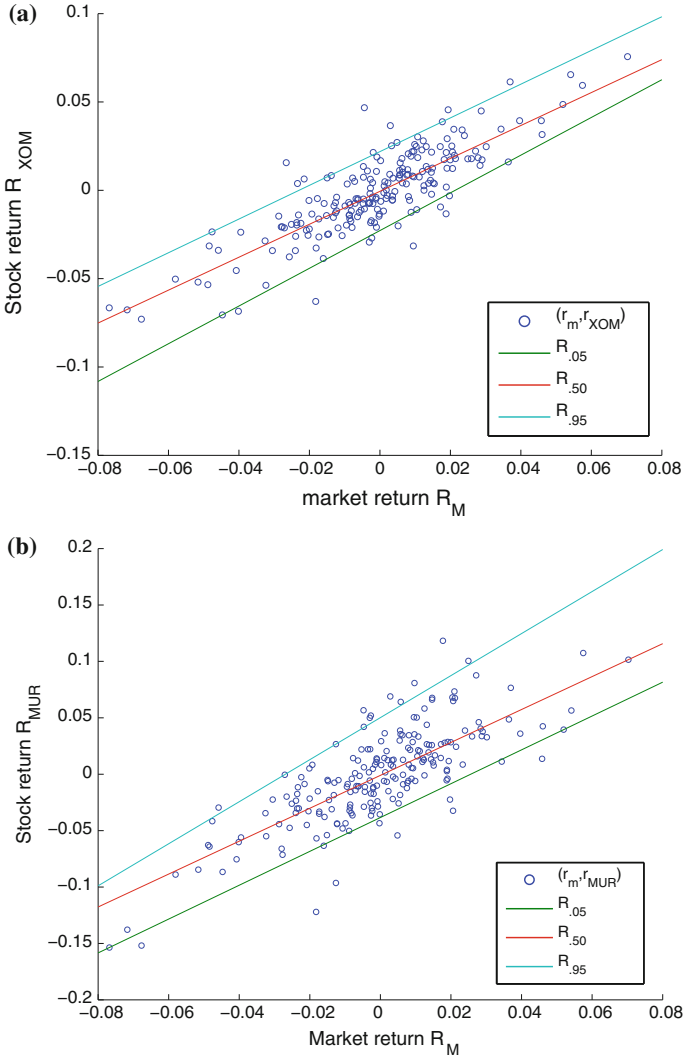


Fig. 3 Securities characteristic line at difference quantile(α). **a** Security line for XOM. **b** Security line for MUR

risk and stock returns for the three stocks. Stochastic intercepts are large for extreme quantile levels and small for the middle quantile.

For future works, we are interesting to extend this model to the time series model such as ARMA, GARCH model. A single factor, β is used on CAPM to compare a portfolio with the market as a whole. Moreover, we can add factors to the model

to give a better fit. The famous approach is the three factor model developed by Fama and French in 1993. A factor model can be extended the CAPM by adding size and value factors in addition to the market risk. So, multifactor models will be considered.

Acknowledgments The authors thank Prof. Dr. Hung T. Nguyen for his helpful comments and suggestions. We would like to thank referee's comments and suggestions on the manuscript.

References

1. Barnes, L.M., Hughes, W.A.: A Quantile Regression Analysis of the Cross Section of Stock Market Returns. Federal Reserve Bank of Boston, Working Paper (2002)
2. Chen, W.S.C., Lin, S., Yu, L.H.P.: Smooth transition quantile capital asset pricing models with Heteroscedasticity. Comput. Econ. **40**, 19–48 (2012)
3. Koenker, R., Gilbert, B.J.: Regression quantiles. Econom.: J. Econ. Soc. 33–50 (1978)
4. Koenker, R.: Quantile Regression, vol. 38, Cambridge University Press (2005)
5. Kotz, S., Drop, van R.J.: Link between two-sided power and asymmetric Laplace distributions: with applications to mean and variance approximations. Stat. Probab. Lett. **71**, 383–394 (2005)
6. Linden, M.: A model for stock return distribution. Inter. J. Financ. Econ. **6**, 159–169 (2001)
7. Lintner, J.: The valuation of risk assets and the selection of risky investments in stock portfolios and capital budgets. Rev. Econ. Stat. **47**(1), 12–37 (1965)
8. Mukherji, S.: The capital asset pricing model's risk-free rate. Inter. J. Bus. Financ. Res. **5**, 793–808 (2011)
9. Noh, H., Ghouch, A., Keilegom, I.: Quality of Fit Measures in the Framework of Quantile. Université catholique de Louvain, Discussion Paper (2011)
10. Sánchez, B.L., Lachos, H.V., Labra, V.F.: Likelihood based inference for quantile regression using the asymmetric Laplace distribution. J. Stat. Comput. Simul. **81**, 1565–1578 (2013)
11. Sharpe, William F.: Capital asset prices: a theory of market equilibrium under conditions of risk. J. Financ. **19**(3), 425–442 (1964)

Evaluation of Portfolio Returns in Fama-French Model Using Quantile Regression Under Asymmetric Laplace Distribution

Kittawit Autchariyapanitkul, Somsak Chanaim and Songsak Sriboonchitta

Abstract We applied the method of quantile regression under asymmetric Laplace distribution to predicting stock returns. Specifically, we used this method in the Fama and French three-factor model for the five industry portfolios to estimate the beta coefficient, which measure risk in the portfolios management analysis at given levels of quantile. In many applications, we are concerned with the changing effects of the covariates on the outcome across the quantiles of the distribution. Inference in quantile regression can be proceeded by assigning an asymmetric Laplace distribution for the error term. Finally, we use the method to measures the volatility of a portfolio relative to the market, size and value premium. It should be noted that a complete study of quantile regression models with various error distributions is of great interests for applications.

1 Introduction

The portfolio theory was first purposed by Markowitz in 1952, a simple idea that described the return of the portfolio by mean and variance. These concepts were essential to development of the famous capital asset pricing model (CAPM). CAPM was introduced by Sharpe [19] and Lintner [14]. The classical CAPM is predicting the return of the asset by using only market return to evaluate the return in portfolio management. But it is only 70 % given by the CAPM (within sample) explains of the diversified portfolios returns compared with the Fama-French three-factor model can explains over 90 % (see, Fama and French [6]). The three-factor model, two more factors namely, size and value variables are added into the original CAPM. Both CAPM and the Fama-French model are use ordinary least square (OLS) to obtain the beta parameters and usually assumed error term to be jointly normally distributed.

K. Autchariyapanitkul (✉) · S. Sriboonchitta
Faculty of Economics, Chiang Mai University, Chiang Mai 52000, Thailand
e-mail: kittawit_autchariya@cmu.ac.th

S. Chanaim
Department of Mathematics, Faculty of Science, Chiang Mai University, Chiang Mai, Thailand

V.-N. Huynh et al. (eds.), *Econometrics of Risk*, Studies in Computational Intelligence 583, DOI 10.1007/978-3-319-13449-9_16

However, this is not always true in the financial market. The Fama-French assumes that the variance of returns adequately measures risk. This may be true if returns are distributed normally. In this paper we introduce the quantile regression with an asymmetric Laplace distribution (ALD) to estimate the parameters of the model and predict portfolios returns.

Quantile regression can explain the entire conditional distribution of the outcome variable, and is more robust to outliers and wrong assumption of the error distribution. For the application of quantile regression to Fama-French model, we have not seen much about applying the quantile in the three-factor model.

Quantile regression estimation is equivalent to the parametric case where the error term is asymmetrically Laplace distributed. The beneficial of parametric estimation is that we can have all the properties of Maximum Likelihood Estimates (MLE). Estimators derived by the method of maximum likelihood have some desirable properties such as, sufficiency, consistency, efficiency, asymptotic normally (Fisher Information), invariance, etc.

Many studies on the Fama-French model is wildly used to study the diversification of the risk parameter and the performance of protfolios, which we can found in the studied from Abhakorn et al. [1], they used standard C-CAPM by including two additional factor associated with Fama-French [7]. Same as in the studied of Bartholdy and Peare [4], compared the performance of CAPM and Fama-French models for individual stocks. In the support of Gaunt [11] tests validity between Three factor model and CAPM, all their results shown that Fama-French model provides a better explanation of stock returns than the CAPM model. Lin et al. [15] studied the relation between the Fama-French factors and the latent risk factors in Chinese market. More related work using the Fama-French model, we refer the reader to the works of Mwalla and Karasneh [2], Eraslan [5], Faff et al. [10]. Grauer and Janmatt [11]

This study extends the standard Fama-French three factor model by present a likelihood-based approach to the estimation of regression quantiles based on the asymmetric Laplace distribution. In [18], the authors construct the distribution of currency exchange rates using the asymmetric Laplace distribution, which successfully captures the peakedness, fat tail and skewness inherent in such data. Similarly, it is shown in [16] that the Laplace distribution has a geometric stability for the weekly and monthly distributions of stock returns and also captures the high peak, fat tail and the skewness of the stock returns.

The useful of regression with Fama-French model have been mentioned in Kent [13] i.e., First, The Fama-French model can explains much more of the variation observed in realized returns. Second, it is show that a positive alpha observed in a CAPM regression is merely a result of exposure to either SMB or HML factors, rather than actual manager performance.

Hence, the main objective of this study is to illustrate the method of quantile regression under asymmetric Laplace distribution. we estimate five industrial portfolios returns based on quantile regression under asymmetric Laplace distribution to evaluate the returns of the portfolios. Thus, the contribution of this paper is using the

method of quantile regression under ALD to obtained the value of betas parameter under various market situation.

The remainder of the paper proceeds as follows. Section 2 gives the overview of quantile regression with asymmetric Laplace distribution and Fama-Frnch three factor model. Meanwhile, Sect. 3 described the empirical method using qunatile regression under asymmetric Laplace distribution. Section 4 exhibits the empirical solutions. The last gives the conclusion of the paper.

2 Quantile Regression and Fama-French Model

2.1 Quantile Regression with an Asymmetric Laplace Distribution

Quantile regression (QR) supplies information about the relationship between response and the covariates at the tails of the response distribution. In a linear QR model $Y = X'\beta_\alpha + \varepsilon_\alpha$, the parameter β_α is estimated by minimizing the empirical objective function $\sum_{i=1}^{n}[\rho_\alpha(Y - X'\beta)]$ over β. Thus, given i.i.d (X_i, Y_i), a plausible estimator of β_α is

$$\widehat{\beta_\alpha} = \arg\min \frac{1}{n}\left\{\sum_{i=1}^{n}\rho_\alpha(Y_i - X_i'\beta)\right\}. \tag{1}$$

Function $\rho_\alpha(\cdot)$ is the so called check (or loss) function defined by $\rho_\alpha(u) = u(\alpha - 1_{(u<0)})$, with $1_{(u<0)}$ denoting the usual indicator function. This estimator is called the Least Absolute Deviation (LAD) estimator. Just as minimizing a loss is associated with normal errors, minimizing check function corresponds to assuming a distribution called asymmetric Laplace distribution (ALD) for the error ε_α. Note that, just like in mean regression model, while the OLS method provides estimators for the model parameters, to make tests and set up confidence intervals, we need to make an assumption about the distribution of the error term.

Thus, suppose that Y_i is distributed as ALD $(\beta_\alpha X_i, \sigma, \alpha)$, $i = 1, 2, \ldots, n$. Then the likelihood is

$$L(\beta_\alpha, \sigma|Y_1, \cdots, Y_n) = \frac{\alpha^n(1-\alpha)^n}{\sigma^n}\exp\left\{-\sum_{i=1}^{n}\rho_\alpha\left(\frac{Y_i - X_i'\beta_\alpha}{\sigma}\right)\right\}. \tag{2}$$

Maximizing L with respect to β_α is equivalent to minimizing $\sum_{i=1}^{n}[\rho_\alpha(Y - X'\beta)]$. Note that, the ALD of the error ε_α is

$$f_{\varepsilon_\alpha}(u) = \frac{\alpha(1-\alpha)}{\sigma}\exp\left\{-\rho_\alpha\left(\frac{u}{\sigma_\alpha}\right)\right\}. \tag{3}$$

The validation measure for general quantile regression model is

$$Q_n(\alpha) = 1 - \frac{\sum_{i=1}^{n} \rho_\alpha(Y_i - q_\alpha(Y_i|X_i))}{\sum_{i=1}^{n} \rho_\alpha(Y_i - \hat{\theta}_\alpha X_i)} \tag{4}$$

The empirical general quantile regression is obtained by estimated quantiles. For more details on this measure for validating quantile regression models, see, [17].

2.2 Fama-French Three-Factor Model

The three-factor model was purposed by Fama and French [6] and has been applied in various issue (see, [7–9]). This model provides an extended version of the CAPM for evaluation of the portfolio. The original CAPM model is described by the linear regression as follows

$$r_A = r_f + \beta_A(r_M - r_f) + \varepsilon, \tag{5}$$

In the three-factor model, two additional factors are added to explain excess return; "size" and "value" to be the most significant factors. Thus, for each portfolio can be estimate the return by the following regression:

$$r_A = r_f + \beta_A(r_M - r_f) + s_A\ SMB + h_A\ HML + \varepsilon, \tag{6}$$

where r_A is the total return of portfolio, r_f is the risk free rate, r_M is the market return and ε is the error term. SMB which is so called "Small Minus Big" accounting for the *size premium*, is designed to measure the difference in return between investing in small and big capitalization stocks, s_A represents the level of exposure to size risk. The words "Small and Big" are refer to the size of the market equity (ME) which is the multiplication of share price and number of shares outstanding. "High Minus Low" (HML) represents the *value premium*, is invented to measure the excess returns for investing in high book-to-market values (BE/ME) and low BE/ME companies and h_A shows the level of exposure to value risk.

Note that, SMB is the average return on the three small portfolios minus the average return on the three big porfolios. HML is the average return on the two value portfolios minus the average return on the two growth portfolios. SMB and HML are calculated from the combinations of Small Value (SV), Small Neutral (SN), Small Growth (SG), Big Value (BV), Big Neutral (BN) and Big Growth (BG). Thus, we have

$$SMB = \frac{1}{3}(SV + SN + SG) - \frac{1}{3}(BV + BN + BG) \tag{7a}$$

$$HML = \frac{1}{2}(SV + BV) - \frac{1}{2}(SG + BG) \tag{7b}$$

3 Simulated Data for ALD

Consider the linear model $Y = X\beta_\alpha + \varepsilon_\alpha$, for $0 < \alpha < 1$, with $F_{\varepsilon_\alpha|X}(0) = \alpha$. Such that this condition entails that the conditional $\alpha - quantile$ of Y given X is $X\beta_\alpha$. Recall the $\alpha -$ loss function

$$\rho_\alpha(u) = u[\alpha - 1_{(u<0)}] = \begin{cases} u(\alpha - 1) & : u < 0 \\ u\alpha & : u \geqslant 0 \end{cases} \tag{8}$$

Thus, $\rho_\alpha(\frac{u}{\sigma}) = \frac{\rho_\alpha(u)}{\sigma}$ when $\sigma > 0$. Suppose the conditional distribution of Y given X is a $ALD(\mu, \sigma, \alpha)$, where the location parameter $-\infty < \mu < \infty$, the scale parameter $\sigma > 0$, and the skew parameter $0 < \alpha < 1$. Given the density of ε_α in (3). For simulations of ε from this distribution where we know α and σ, we seek its distribution function F_{ε_α} to carry out the usual procedure by setting $F_{\varepsilon_\alpha} = U$, uniformly distributed on [0,1], so that $\varepsilon_\alpha = F^{-1}_{\varepsilon_\alpha}(U)$ we have

$$F_{\varepsilon_\alpha}(x) = \begin{cases} \alpha \exp\{\frac{(1-\alpha)x}{\sigma}\} & : x < 0 \\ 1 - (1-\alpha)\exp\{-\frac{\alpha x}{\sigma}\} & : x \geqslant 0 \end{cases} \tag{9}$$

From which, we get

$$u = \alpha \exp\left\{\frac{(1-\alpha)x}{\sigma}\right\} \Rightarrow x = \frac{\sigma(\log u - \log \alpha)}{1-\alpha} : u < \alpha \tag{10}$$

which is less than 0 when $\log u - \log \alpha < 0$ and

$$u = 1 - (1-\alpha)\exp\{-\frac{\alpha x}{\sigma}\} \tag{11a}$$

$$x = \frac{\sigma[\log(1-\alpha) - \log(1-u)]}{\alpha} \geqslant 0 : u \geqslant \alpha \tag{11b}$$

4 Application to Portfolio Evaluation

4.1 Model and Parameters Estimation

The Fama and French three-factor asset pricing model provides an option to CAPM as an improvement to poor performance of the CAPM. With this method, the estimation of expected excess return on portfolios will be calculated by adding two more factors

namely; SMB and HML into the classic CAPM model. Suppose we have observed the past data of stock return $r_{ai}, r_{mi}, SMB_i, HML_i, i = 1, 2, \cdots, n$ over past n years. These observations are assumed an i.i.d. random noise from $ALD(\alpha, \mu, \sigma)$. In this case we consider $\mu = 0$.

The equation of the three-factor model under asymmetric Laplace distribution at given level of α quantile, using (2) and (6) the corresponding to the historical data via the three-factor model is a realization that generate likelihood function is

$$L_\alpha(\beta_0^\alpha, \beta_1^\alpha, \beta_2^\alpha, \beta_3^\alpha, \sigma | (r_{mi}, r_{ai}, SMB_i, HML_i)) = \frac{\alpha^n(1-\alpha)^n}{\sigma^n} \exp\left\{-\sum_{i=1}^{n} \rho_\alpha\left(\frac{r_{ai} - HML_i\beta_3^\alpha - SMB_i\beta_4^\alpha - r_{mi}\beta_1^\alpha - \beta_0^\alpha}{\sigma}\right)\right\}. \tag{12}$$

4.2 Experimental Results

The data contain of 1050 monthly returns during 1926–2013 are original obtained from Center for Research in Security Prices (CRSP) to compute the log returns on the following asset. The data consist of the returns from the five industry portfolios, Consumer (Cnsmr), Manufacturing (Manuf), Hi-Technologies (HiTec), Health care (Hlth) and Other (Other), such as Mines, Transportation etc. Market returns (r_{Mt}) includes all New York Stock Exchange (NYSE), American Stock Exchange (AMEX) and NASDAQ Stock Market (NASDAQ) firms.

Data for SMB and HML were obtained from French's homepage. Table 1 gives the summary of the variables.

The Q-Q plots shown in Fig. 1 are based on the distribution of AL, given in (2). The lines in these figures represent the 0.5th quantile. These figures are clearly show that the asymmetric Laplace distribution provides a good-fit to the sample data set. Moreover, the Kolmogorov-Smirnov test (KS-test) in Table 2 ensure that all the

Table 1 Summary statistics

	Cnsmr	Manuf	HiTec	Hlth	Other	r_{Mt}	SMB	HML
Mean	0.726	0.696	0.655	0.803	0.627	0.649	0.234	0.394
Median	0.965	0.935	0.930	0.775	0.975	1.030	0.065	0.230
Maximum	43.750	41.310	33.850	37.030	58.790	38.040	37.450	34.080
Minimum	−28.590	−29.880	−26.780	−34.140	−29.960	−29.100	−16.390	−12.680
Std. dev	5.363	5.552	5.652	5.649	6.514	5.414	3.232	3.512
Skewness	0.120	0.334	−0.183	0.114	0.882	0.157	2.060	1.920
Kurtosis	10.570	11.082	6.566	9.589	15.985	10.392	23.558	18.722
Obs	1050							

All values are the log return

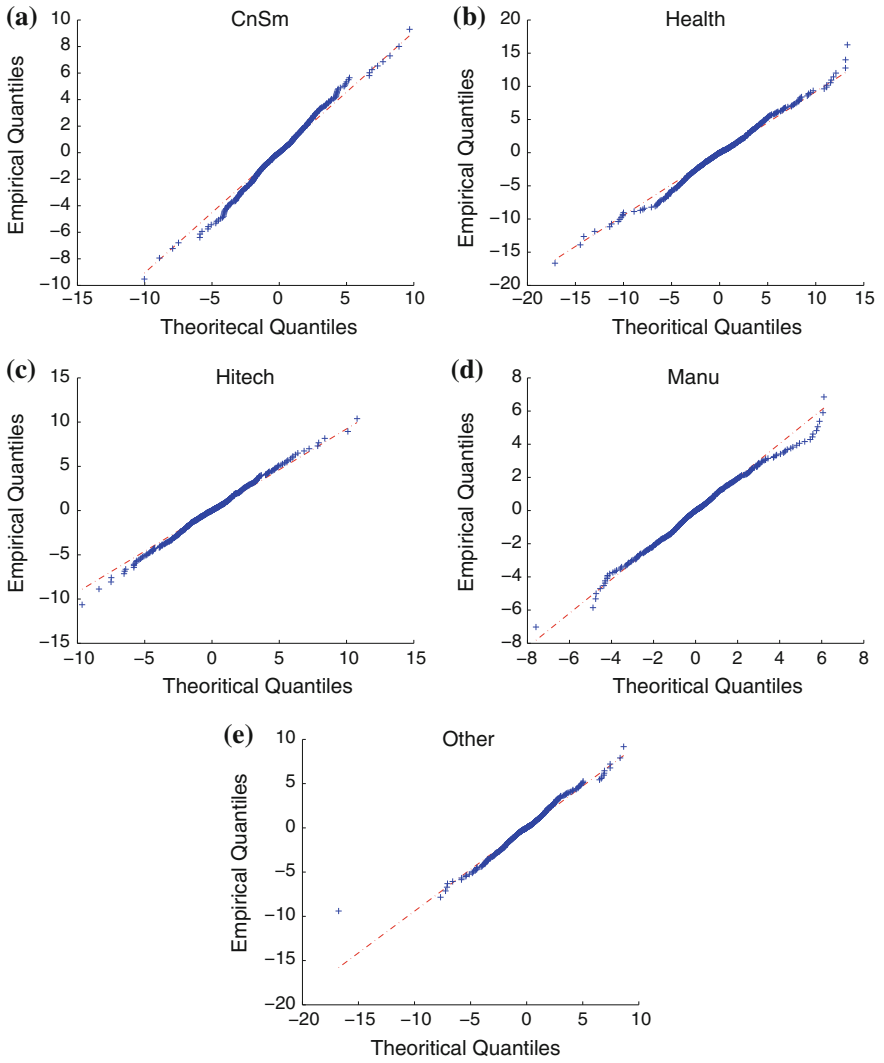


Fig. 1 The Q-Q plots of error under asymmetric Laplace distribution. **a** *ALD* Q-Q plot of Cnsm $\alpha = 0.50$. **b** *ALD* Q-Q plot of Hlth $\alpha = 0.50$. **c** *ALD* Q-Q plot of HiTec $\alpha = 0.50$. **d** *ALD* Q-Q plot of Manuf $\alpha = 0.50$. **e** *ALD* Q-Q plot of Other $\alpha = 0.50$

marginals are follow the ALD compare with the critical value $\frac{K(\alpha'=0.05)}{\sqrt{n}} = 0.042$, all of the marginals accepts the null hypotheses of the KS-test at 5 % level.

The values in the Table 2 exhibit the results of the parameters estimation for the Fama and French three-factor model via the quantile regression under asymmetric Laplace distributions at given level of α. The results for ALD assumption performs well for the given quantile of these five industrial portfolios.

Table 2 Parameter estimated results

Industries	Parameters	$\alpha = 0.20$	$\alpha = 0.40$	$\alpha = 0.50$	$\alpha = 0.60$
Cnsmr	β_0	−1.1814	−0.2704	0.1281	0.4767
		(0.0002)	(0.0144)	(0.0125)	(0.0008)
	β_1	0.9164	0.9386	0.9396	0.9342
		(0.0003)	(0.0366)	(0.0053)	(0.0007)
	β_2	0.0822	0.0653	0.0457	0.0344
		(0.0003)	(0.0463)	(0.0041)	(0.0003)
	β_3	−0.0577	−0.0593	−0.0666	−0.0535
		(0.0001)	(0.0000)	(0.0347)	(0.0044)
	σ	0.4941	0.6563	0.6766	0.6601
		(0.0151)	(0.0190)	(0.0203)	(0.0192)
	$Q_n(\alpha)$	0.6561	0.6474	0.6419	0.6359
	LL	2233.9	2106.2	2095.3	2112.2
KS-test(D_n)	0.0318				
Hlth	β_0	−1.8832	−0.3510	0.2227	0.9299
		(0.0011)	(0.0362)	(0.0023)	(0.3184)
	β_1	0.8881	0.8784	0.8642	0.8608
		(0.0008)	(0.0757)	(0.0013)	(0.4387)
	β_2	−0.1341	−0.0911	−0.1090	−0.1035
		(0.0008)	(0.0869)	(0.0022)	(1.278)
	β_3	−0.2062	−0.2528	−0.2521	−0.2857
		(0.0011)	(0.0391)	(0.0009)	(0.4881)
	σ	0.8609	0.6563	0.6766	0.6601
		(0.0224)	(0.0255)	(0.0337)	(0.1057)
	$Q_n(\alpha)$	0.4191	0.4161	0.4091	0.4017
	LL	2816.9.1	2687.6	2682.2	2706.9
KS-test(D_n)	0.0374				
HiTech	β_0	−1.3758	−0.2920	0.1262	0.5110
		(0.0002)	(0.3468)	(0.0483)	(0.0034)
	β_1	0.9535	0.9655	0.9701	0.9852
		(0.0001)	(0.0014)	(0.6291)	(0.0060)
	β_2	0.0246	0.0702	0.0567	0.0422
		(0.0001)	(0.0601)	(0.4999)	(0.0245)
	β_3	−0.3052	−0.2797	−0.2735	−0.2829
		(0.0001)	(0.0015)	(0.9118)	(0.0152)
	σ	0.5540	0.7351	0.7562	0.7379
		(0.0180)	(0.0587)	(0.0219)	(0.0266)
	$Q_n(\alpha)$	0.6456	0.6327	0.6269	0.6212
	LL	2354.0	2225.2	2212.0	2229.2
KS-test(D_n)	0.0291				

(continued)

Table 2 (continued)

Industries	Parameters	$\alpha = 0.20$	$\alpha = 0.40$	$\alpha = 0.50$	$\alpha = 0.60$
Manuf	β_0	−0.8635	−0.2670	−0.0480	0.2131
		(0.0156)	(0.0237)	(0.0001)	(0.0003)
	β_1	0.9935	1.0060	0.9955	0.9982
		(0.0097)	(0.0009)	(0.0024)	(0.0000)
	β_2	−0.1037	−0.1053	−0.1090	−0.1021
		(0.0155)	(0.0068)	(0.0024)	(0.0001)
	β_3	0.1482	0.1367	0.1508	0.1529
		(0.0103)	(0.0015)	(0.9118)	(0.0152)
	σ	0.3696	0.4814	0.4987	0.4923
		(0.0089)	(0.0587)	(0.0219)	(0.0266)
	$Q_n(\alpha)$	0.7497	0.7449	0.7393	0.7324
	LL	1928.8	1780.9	17775.0	1804.4
KS-test(D_n)	0.0328				
Other	β_0	−1.4527	−0.5941	−0.1880	0.1755
		(0.0006)	(0.0756)	(0.0002)	(0.0362)
	β_1	1.0704	1.0584	1.0606	1.0539
		(0.0001)	(0.5700)	(0.0003)	(0.0217)
	β_2	0.0684	0.0936	0.0936	0.1172
		(0.0003)	(0.8370)	(0.0017)	(0.0184)
	β_3	0.3311	0.3379	0.3512	0.3394
		(0.0002)	(0.5059)	(0.0005)	(0.0114)
	σ	0.4932	0.6509	0.6680	0.6480
		(0.0150)	(0.0092)	(0.0185)	(0.0179)
	$Q_n(\alpha)$	0.7129	0.7024	0.6983	0.6948
	LL	2521.5	2225.2	2212.0	2229.2
KS-test(D_n)	0.0378				

Consistent standard errors() is in parenthesis

Now, Fig. 2 shows parameter estimation for the entire quantile. From these picture, we summarize the results, e.g. the return from Cnsmr portfolio as follows:

For the quantile lower than 0.90, Cnsmr has less risky than the market and more risk at higher quantile levels, The risks decrease as the stock return increase, but we cannot find any monotonic relationship between risks and excess reurns.

For the lower quantile less than 0.9, positive exposure to size risk increases the average excess return while negative exposure to size risk reduces the average excess retrun regarding medium and small size portfolios. We conclude that the size factor SMB is not effect on large scale of portfolio returns.

For every quantile, HML capturing the value risk effect of Cnsmr portfolio on average excess returns. Since, it has a negative value, it is expected that high book to market value (BE/ME) decrease average excess return more than the low (BE/ME) one.

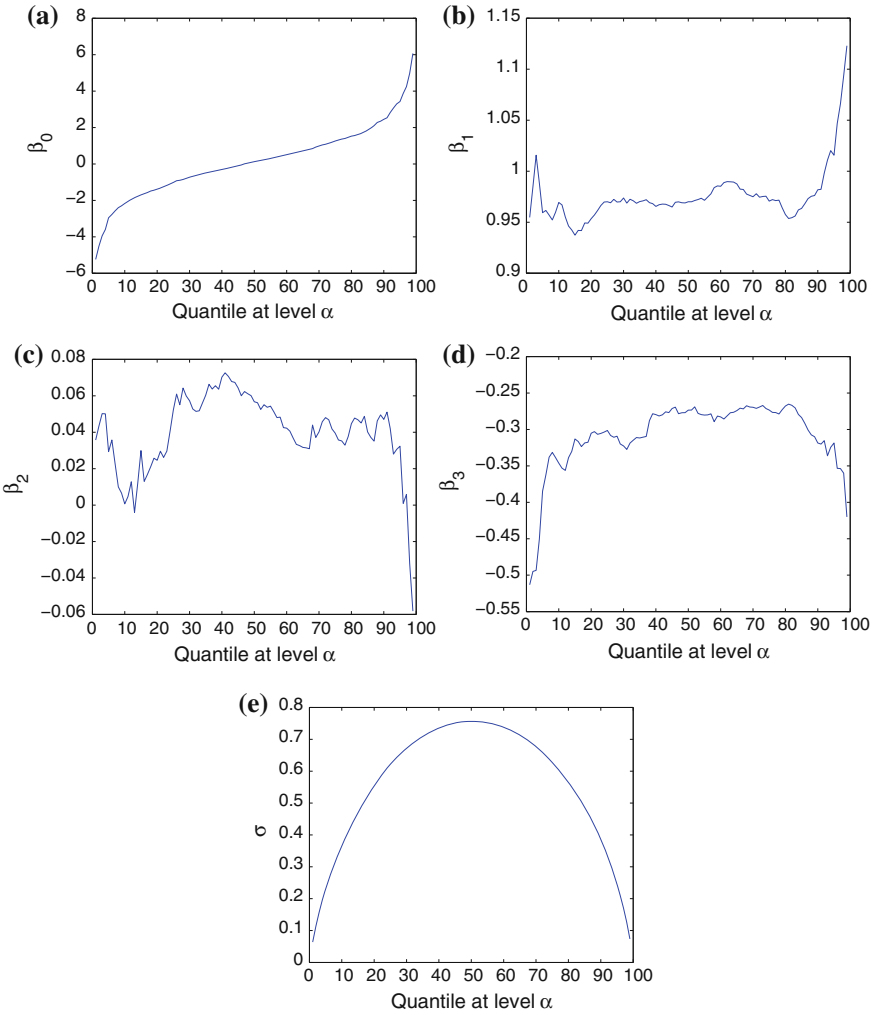


Fig. 2 Marginal parameter plots of Cnsmr portfolio at difference quantile(α). **a** β_0 Cnsmr. **b** β_1 Cnsmr. **c** β_2 Cnsmr. **d** β_3 Cnsmr. **e** σ Cnsmr

4.3 In Sample prediction

To predict the in-sample expected return of the asset $\hat{r}_{a,n}$ for a given market portfolio return $r_{m,n}$, we compute the estimated values of $r_{a,n}$ given $r_{m,n}$ at fixed α by

$$\hat{r}_{Cnsmr} = r_f + \beta_0 + \beta_1(r_M - r_f) + \beta_2\, SMB + \beta_3\, HML + \varepsilon_i, \qquad (13)$$

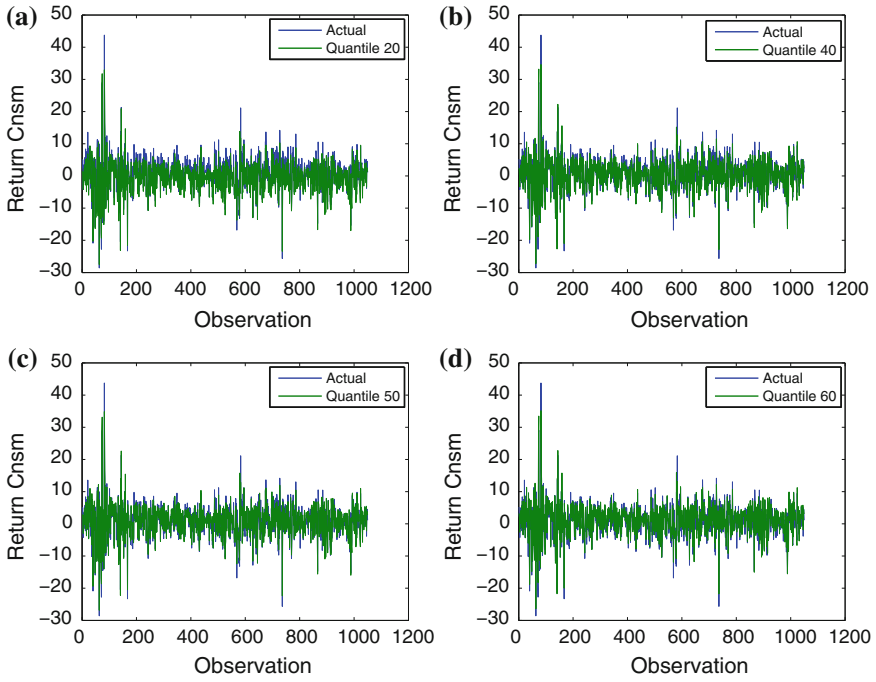


Fig. 3 In-sample prediction at difference quantile (α). **a** Prediction at $\alpha = 0.20$. **b** Prediction at $\alpha = 0.40$. **c** Prediction at $\alpha = 0.50$. **d** Prediction at $\alpha = 0.60$

where ε_i is asymmetric Laplace distribution. Figure 3 displays the in-sample prediction at different quantile. It is clearly that the predicted values are very close to actual values for the given level of quantile under ALD.

5 Conclusions

In this paper, we used method of quantile with ALD assumption applied to the Fama and French three factor model for the five industry portfolios stocks markets, which includes all New York Stock Exchange (NYSE), American Stock Exchange (AMEX) and NASDAQ Stock Market (NASDAQ) firms. The Fama and French model is an extension of original CAPM model by adding two more important variables namely, "size premium" and "value premium" into the model. This method can be used to evaluate the linear relationship between the expected returns on a portfolio and its asymmetric market risk with size and value variables over various quantile levels.

The empirical results show that the method of quantile regression under ALD can captures the stylized factors in financial data to describe the stock returns under most quantile levels, especially under the middle quantile levels. Clearly, there is no

monotonic relationship between risk and stock returns for the these portfolios. This suggests that during that time frame, the ability to increase the returns of portfolios beyond the risk exposure would be achieve by using quantile regression for the risk measurement.

Acknowledgments The authors thank Prof. Dr. Hung T. Nguyen for his helpful comments and suggestions. We would like to thank referee(s) comments and suggestions on the manuscript.

References

1. Abhakorn, P., Peter, N.: Smith and Michael R. Wickens. What do the FamaFrench factors add to C-CAPM? J. Empir. Financ. **22**, 113–127 (2013)
2. Al-Mwalla, M., Karasneh, M.: Fama and French three factor model: evidence from emerging market. Eur. J. Econ. **41**, 132–140 (2011)
3. Barnes, L.M., Hughes, W.A.: A Quantile Regression Analysis of the Cross Section of Stock Market Returns, working paper, Federal Reserve Bank of Boston (2002)
4. Bartholdy, J., Paula, P.: Estimation of expected return: CAPM versus Fama and French. Int. Rev. Financ. Anal. **14**(4), 407–427 (2005)
5. Eraslan, V.: Fama and French three-factor model: evidence from Istanbul stock exchange. Bus. Econ. Res. J. **4**(2), 11–22 (2013)
6. Fama, E.F., French, k: The cross-section of expected stock returns. J. Financ. **47**(2), 427–465 (1992)
7. Fama, E.F., French, K.: Common risk factors in the returns on stocks and bonds. J. Financ. Econ. **33**, 3–56 (1993)
8. Fama, E.F., French, K.: Size and book-to-market factors in earnings and returns. J. Financ. **50**, 131155 (1995)
9. Fama, E.F., French, K.: Size, value, and momentum in international stock returns. J. Financ. Econ. **105**(3), 457–472 (2012)
10. Faff, R., Gharghori, P., Nguyen, A.: Non-nested tests of a GDP-augmented FamaFrench model versus a conditional FamaFrench model in the Australian stock market. Int. Rev. Econ. Financ. **29**, 638–672 (2014)
11. Gaunt, C.: Size and book to market effects and the Fama French three factor asset pricing model: evidence from australian stock market. Account. Financ. **44**(1), 22–44 (2004)
12. Grauer, R., Johannus, J.A.: Cross-sectional tests of the CAPM and FamaFrench three-factor model. J. Bank. Financ. **34**(2), 457–470 (2010)
13. Kent, L.W., Zhang, Y.: Understanding risk and return, the CAPM, and the Fama-French three-factor model. Tuck Case 03-111 (2003)
14. Lintner, J.: The valuation of risk assets and the selection of risky investments in stock portfolios and capital budgets. Rev. Econ. Stat. **47**(1), 12–37 (1965)
15. Lin, J., Wang, M., CaiAre, L.: The FamaFrench factors good proxies for latent risk factors? Evidence from the data of SHSE in China. Econ. Lett. **116**(2), 265–268 (2012)
16. Linden, M.: A model for stock return distribution. Int. J. Financ. Econ. **6**, 159–169 (2001)
17. Noh, H., Ghouch, A., Keilegom, I.: Quality of Fit Measures in the Framework of Quantile.Université catholique de Louvain, Discussion Paper (2011)
18. Sánchez, B.L., Lachos, H.V., Labra, V.F.: Likelihood based inference for quantile regression using the asymmetric Laplace distribution. J. Stat. Comput. Simul. **81**, 1565–1578 (2013)
19. Sharp, William F.: Capital asset prices: a theory of market equilibrium under conditions of risk. J. Financ. **19**(3), 425–442 (1964)

Analysis of Branching Ratio of Telecommunication Stocks in Thailand Using Hawkes Process

Niwattisaiwong Seksiri and Napat Harnpornchai

Abstract The aim of the research is to study the branching ratios of the telecommunication stocks in Thailand, ADVANC and DTAC, both listed on the Stock Exchange of Thailand (SET). The branching ratio is the parameter defined in the Hawkes process and directly measures the influential degree of endogeneity. The results indicate to what extent the stock price changes are affected by internal factors. The study found that the branching ratio of ADVANC is at 29 %, which means ADVANCs price change is only 29 %, caused by internal factors, while the remaining 71 % derives from external factors. Meanwhile, DTACs branching ratio is at 55 %, meaning DTACs price change is 55 % due to internal factors and 45 % due to external. Knowing to what extent the stock price is affected by external factors can strengthen investor strategy. Stocks with a low branching ratio are more speculative than those having a high branching ratio.

1 Introduction

Thailand stock market plays an important role for Thai economy, especially in terms of its influence on industrial growth and country development, for many reasons. First of all, it provides companies a way to raise money through issuing corporate bonds or shares. According to Stanlake [17], a stock exchange is of high importance because without one, it will be difficult for companies to find share buyers and sellers, thus making this type of securities illiquid. The companies are only able to raise funds from selling their IPO shares in a primary market. Therefore, share price changes rely only on trading in a secondary marketa stock exchange. Investors profit from price increases and dividend pay-outs while companies gain money from selling IPO shares only. Having a stock exchange means that companies have a long-term source of funding for their business expansion. The companies also pay attention to the changes

N. Seksiri · N. Harnpornchai (✉)
Faculty of Economics, Chiang Mai University, Chiang Mai, Thailand
e-mail: napateconcmu@gmail.com

N. Seksiri
e-mail: s.niwattisaiwong@alumni.lse.ac.uk

V.-N. Huynh et al. (eds.), *Econometrics of Risk*, Studies in Computational Intelligence 583, DOI 10.1007/978-3-319-13449-9_17

in their share price because, according to Shauna [16], most of their executives are the major shareholders. When the price is up, these executives gain higher profits. Moreover, the companies often offer a certain amount of shares to their executives as a year-end bonus. When the share price continues to increase, executives have an incentive to work harder. Second, equity investment is a savings alternative which requires no use of intermediaries like financial institutions and commercial banks. Investors are also able to diversify risks more efficiently when investing through stock markets. Third, a stock market is a key driver of a countrys economic growth. In Thailand, the number of stock traders is relatively low considering the number of citizens, due to the lack of understanding and wrong attitudes toward stock markets. However, a stock exchange is still substantially influential to the economy. Foreign investors are more interested in stock investment in emerging markets, including Thailand, as it provides higher returns than when investing in developed countries. McGregor [10] stated that the stock market is an indicator of listed companies financial health. The healthier these companies are in terms of finance, the more foreign investors think the country has a solid financial status. Note that a strong financial health reflects well on the companys performance and solvency, which makes financial institutes confident in giving out loans. Consequently, investors are more likely to buy the companys shares, which leads to an increase in its share price.

However, Engel and Rangel [4] found that stock markets in developing countries are different from those in developed countries. Stock markets in developing countries have been growing rapidly over the past two and a half decades. Because these countries still have lots of volatility, their stock markets remain sensitive to external factors such as domestic political situations, economic circumstances (both domestic and overseas), and other macro factors.

According to Oyama [13], in the short run, share prices are greatly influenced by market expectations, which usually contradict actual economic fundamentals. The Thailand Security Institute (TSI) [18] noted that share prices tend to change drastically in the short run due to a number of factors, such as economic conditions, corporate fundamentals, and psychological factors. Investors tend to overreact and sell out their shares upon hearing good/bad news, causing the share price to increase/decrease rapidly. Other investors then panic and begin increasing/dumping their shares accordingly, which leads to a continued slide in the share price.

From the viewpoint of risk management, it is thus crucial to comprehend the price movement whether it is dominated by the endogenous or exogenous factors. Accordingly, the so-called branching ratio which is a parameter in the Hawkes process can be used for such a purpose. The branching ratio give the information about the influences of endogenous and exogenous factors on the price movement [5].

Since telecommunication stocks are one of major stocks contributing the SET market, the study of branching ratios for important telecommunication stock prices is highly informative for both the companies and investors for appropriate financial decisions.

This paper investigates the effects of endogenous and exogenous influences on the stock prices of leading telecommunication companies via the branching ratio of the Hawkes process. The content of the paper is as follows. After this introduction section, the second section is a review of the key literature related to the Hawkes process and its application in finance. Discussed in the third section includes the univariate Hawkes process, the parameter estimation using Maximum Likelihood Estimation (MLE), and the compensator of the Hawkes process. The test procedure for the goodness-of-fit of the resulting model is also described in this section. In the fourth section, the investigation of the ADVANC and DTAC stock prices are shown. Finally, the fifth section provides a conclusion of this study and also discusses topics that can be addressed in future research.

2 Literature Review

The Hawkes process was initially adapted to model earthquake aftershocks. As we know, an earthquake is followed constantly by aftershocks, so researchers use the Hawkes process to predict how long the aftershocks will occur. The Hawkes process is commonly used in the area of seismology for this reason—it considers the influence of past events on current conditional intensity. Although it was invented in the 1970s, the Hawkes process was first adapted in finance just around 2007.

Among the first research works to use the Hawkes process in the financial market was Bowsher [2], Modelling Security Market Events in Continuous Time: Intensity Based, Multivariate Point Process Models. The studys objective is to develop a generalized Hawkes model, described in terms of its vector conditional intensity, and apply the bivariate version of it to explain the durations of trades and mid-quote changes. Bowsher [2] used the dataset of trades and changes to the mid-quote that occurred for General Motors Corporation (GM) for a period of 40 trading days in 2000. He found that there is a two-way interaction between trades and changes in the mid-price where the occurrence of a trade increases the intensity of mid-price changes, and mid-price changes increase trade intensity.

Another widely-recognized study using the Hawkes Process in finance is "Clustering of Order Arrivals, Price Impact and Trade Path Optimisation", conducted by Hewlett [7]. The aim was to use a bivariate Hawkes process in modelling an order flow in the FX market. Hewlett proposed that one could predict future trading intensity that depends on the pattern of past trading modelled via the Hawkes process. Moreover, the model he developed demonstrates how the liquidity taker should behave, given the reaction of the market maker, by assuming that the process of order arrival follows a bivariate Hawkes process. The dataset used in this study consists of records of market orders maintained on EBS over two months. According to the study, the more risk-averse the traders, the more quickly they dispose of their inventory. The traders were assumed to start with an inventory of 10 units at time zero. Moreover, for risk-averse, impatient traders, the price changes are expected to overshoot its equilibrium value of –10, and the average price received is considerably lower than

this equilibrium value. For 'average' traders, the average price paid is around the same level as the equilibrium value. For patient traders, the average price paid is lower than the equilibrium value.

Toke and Pomponio's [19] "Modelling Trades-Through in a Limit Order Book Using Hawkes Processes" sought to model trades-through using a simple bivariate Hawkes process. They used the Thomson-Reuters tick-by-tick data on the Euronext-Paris stock, called the BNP Paribas (BNPP.PA), trading for 108 days from June to October 2010. The result shows that the dataset fits the Hawkes process well.

Filimonov and Sornette [5] conducted a study entitled "Quantifying Reflexivity in Financial Markets: Towards a Prediction of Flash Crashes". They proposed a market endogeneity measurement which indicates whether price changes are a result of exogenous events such as the companys general news or economic situations, or rather certain events that occurred endogenously. These endogenous events might happen due to market movements causing positive feedback mechanisms that induce correlation among price changes. Filimonov and Sornette [5] used the branching ratio of a Hawkes process model as a proxy for market endogeneity and provided an estimation using Maximum Likelihood. The dataset is extracted from the E-mini S&P 500 Futures for the period 1998–2010. According to the study, the branching ratio was about 0.3 before 2000, showing that the market endogeneity was relatively low. After 2004, the branching ratio increased to almost 0.9. In addition, the branching ratio during the period was quite stable as long as it was satisfied by some exogenous news. Filimonov and Sornette [5] cited the downgraded debt ratings of Greece and Portugal on April 27, 2010 as an example of such news. Nevertheless, the branching ratio increased dramatically during the crash of May 6, 2010—widely known as the Flash Crash—when stock markets fell without any relevant exogenous news. Filimonov and Sornette [5] also noticed that the increase in the branching ratio coincides with the rise in activity by high-frequency traders. The Flash Crash itself, despite there being no evidence that it was triggered by high-frequency trading, was to some extent associated with the presence of high-speed, automated trading systems that might have exacerbated the extreme market movements observed on that day.

Lorenzen's [9] "Analysis of Order Clustering Using High Frequency Data: A Point Process Approach" used a Hawkes model and high-frequency stock market data to estimate durations, trades, and quotes. The study follows the estimation performed by Filimonov and Sornette [5], but instead of using the duration of mid-price changes of the E-mini S&P 500 contract, it used the data on equity markets obtained from 2 different databases: (1) the Trades and Quotes (TAQ), comprising stocks traded in the US exchanges (stock selected was Yahoo Inc. [YHOO]); and (2) the Thomson Reuters database, comprising stocks traded in European exchanges (stock selected was Vodafone [VOD]). Another difference between the work of Lorenzen [9] and that of Filimonov and Sornette [5] is that the former's data do not fully rely on the randomization of timestamps, which sets up a strong assumption. Lorenzen [9] attempted to assess the impact of the randomization of timestamps on the estimates and the fit of the Hawkes model. The result showed that the best robustness check one can perform to assess the impact of the randomization of timestamps in the estimates and in the fit of the Hawkes process is use data that distinguishes among events

at the same second. This check is conducted by constructing a randomized dataset from a 'real' dataset that has a precision higher than one second. This randomized dataset is obtained by simply rounding to the nearest second the timestamps with high precision and adding a randomized precision component to the rounded data.

3 Methodology

3.1 Hawkes Process

As the Hawkes process combines endogenous responses and exogenous influences, the process is employed to measure the ratio of the price movements due to endogenous and exogenous effects. In particular, the Hawkes process defines the branching ratio, which is the parameter directly measures the influential degree of endogeneity. In other words, the branching ratios obtained through parameter estimation in the Hawkes process displays factors affecting the price changes, indicating by how many percentage points the price movements are affected by endogenous factors (the stock's fundamentals).

The Hawkes process is a generalization of the nonhomogeneous Poisson process, whose intensity λ_t depends not only on time t but also on the past effects. It is thus a continuous-time process. The Hawkes process was first introduced by Hawkes in 1971 [6]. Accordingly, the intensity λ_t where $t = 1, 2, 3, \ldots$ given by:

$$\lambda_t = \mu_0(t) + \Sigma_{t_i<(t)} g(t - t_i) \tag{1}$$

where λ_t is a conditional intensity and t_i is some random variable that satisfies $t_1 < t_2 < \cdots < t_N$.

The first term $\mu_0(t)$ is the original intensity of the model that determines the arrival rate of the first-order event per unit of time. $\mu_0(t)$ represents a background intensity that accounts for exogenous events (i.e., independent on history).There is no specific function form for $\mu_0(t)$, but it is generally assumed time-invariant. Consider the case of the first-order event-the stock price falls for the first time in 5 trading days. The fall in the stock price is the first-order event while the period of 5 trading days is a unit of time. The second term is a summation of the response functions $g(t-t_i)$ which explain the events following the first-order event. According to the stock price example, aftershocks i.e. the fall of stock price in the following days due to the investor's panic in falling in the stock price in the previous day, are the response functions which can occur countless times. The summation of these aftershocks constitutes the clustering property of the model. From Bowsher [2] and Hewlett [7], the response functions are in the form of exponential function, $g(t - t_i) = \alpha e^{(-\beta(t-t_i))}$.

As a result, the intensity of the Hawkes process is given by

$$\lambda_t = \mu_0(t) + \Sigma_{(t_i<t)} \alpha e^{(-\beta(t-t_i))} \tag{2}$$

A branching ratio ζ is derived as

$$\zeta = \int_0^\infty \alpha e^{-\beta t} dt = \frac{\alpha}{\beta} \tag{3}$$

If $\zeta < 1$, the process is in the sub-critical regime and if $\zeta = 1$, the process is in the critical regime. The branching ratio can be used to measure the proportion of all events that depend on the first-order event (see, Filimonov and Sornette [5]). Basically speaking, the branching ratio can measure the proportion of all aftershocks. If $\zeta < 1$, φ will be higher than 1, meaning that aftershocks will occur for a certain period of time. If $\zeta = 1$, φ would equal $+\infty$, which means aftershocks will keep occurring infinitely and endlessly.

3.2 Parameter Estimation of Hawkes Process

According to Ozaki [13], the Hawkes Process can be estimated by using the Maximum Likelihood Estimation. Meanwhile, Ogata [11] found that the asymptotic properties of the Maximum Likelihood estimator in the Hawkes Process is in the form of a log-likelihood function with a response function. This can be written as

$$\log L(t_1, t_2, \ldots, t_N) = -\int_{-\infty}^{t_N} \lambda(t|\theta) dt + \int_0^{t_N} \log \lambda(t|\theta) dN(t) \tag{4}$$

where $\lambda(t|\theta)$ is the conditional intensity of the process.

In this paper, the event is the up-crossing of the stock price across the mid-price of the stock of interest. An index i which runs from 1 to N will label each event. Note that the times when an event takes place must satisfy $t_1 < t_2 < \cdots < t_N$. Ozaki [13] demonstrated that the Hawkes Process's log-likelihood function described by Eq. (2) can be shown as

$$\log L(t_1, \ldots, t_N|\theta) = -\mu t_N + \sum_{i=1}^{N} \frac{\alpha}{\beta}(e^{-\beta(t_N - t_i)} - 1) + (i = 1)^N log(+(i)) \tag{5}$$

where $\Omega(i)$ is derived from

$$\Omega(i) = \begin{cases} \sum_{t_j < t_i} e^{-\beta(t_i - t_j)}, & \text{for} \quad i \geq 2 \\ 0, & \text{otherwise} \quad t_{i+1}. \end{cases} \tag{6}$$

3.3 Compensator of Hawkes Process

The compensator of Hawkes Process is defined as the integral of the intensity over the overall history of the time process.

$$\Lambda(t) = \int_0^t \lambda(r)dr \tag{7}$$

If we look at only a certain period of time, such as from the period t_i to t_{i+1}, the solution can be written

$$\Lambda(t_i, t_{t+1}) = \int_{t_i}^{t_{t+1}} \lambda(r)dr \tag{8}$$

The time changes in the (11) process is random. Therefore, the result from (11) can be called the residual process.

The compensator of the Hawkes process with the intensity (2) from the period t_i to t_{i+1} is given by

$$\Lambda(t_i, t_{t+1}) = \int_{t_i}^{t_{i+1}} \mu(r)dr + \int_{t_i}^{t_{t+1}} \sum_{t_k<r} \alpha e^{-\beta(r-t_k)}dr \tag{9}$$

When $\mu(r) = \mu$, then

$$\Lambda(t_i, t_{t+1}) = \mu(t_{i+1} - t_i) - \sum_{k=1}^{i} \frac{\alpha}{\beta}[e^{-\beta(t_{i+1}-t_k)} - e^{-\beta(t_i-t_k)}] \tag{10}$$

3.4 Goodness of fit

Because the compensator follows a unit-rate exponential distribution, the QQ-plot between the simulated univariate Hawkes process and the estimated compensator is used as a measure of goodness of fit here. An alternative approach introduced by Daley and Vero-Jones [3] is the Kolmogorov-Smirnov.

4 Empirical Results

The telecommunication stock prices that are considered in this study are ADVANC and DTAC. Both stocks are listed on the SET and are major communications stocks traded by many investors. The average prices of ADVANC and DTAC in the period between January 4, 2012 and March 18, 2014 are used as mid-prices in this study. A total of 543 data points for the period January 4, 2012 to March 18, 2014 are analysed. Figure 1 shows the historical daily average prices of ADVANC, and Fig. 2 portrays the historical daily average prices of DTAC.

It should be noted that the time scale is represented by day collapse of which the starting day is 4 Jan 2012 and set to be day 1 although the Day-axis begins from 0. The mid-price is defined as the average value of the data and is obtained as 216.31 Bath for ADVANC stock and 93.29 Baht for DTAC. The interested events in this paper are those in which the daily-average prices pass the mid-price in an upward direction.

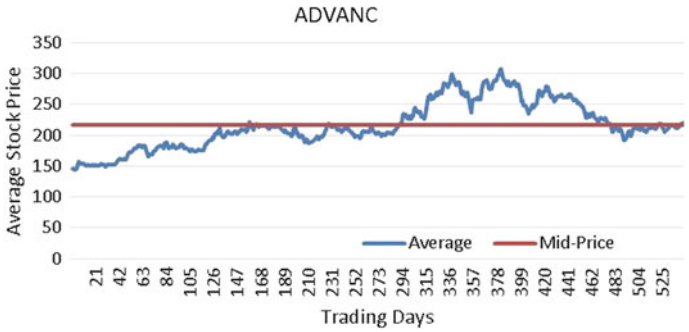


Fig. 1 Historical daily average stock prices of ADVANC for the period January 4, 2014–March 18, 2014. *Source* the Stock Exchange of Thailand [15]

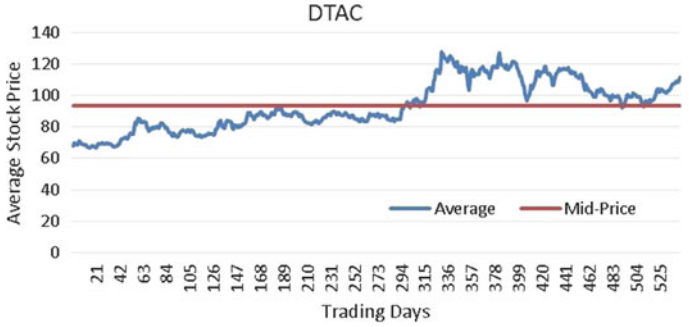


Fig. 2 Historical daily average stock prices of DTAC for the period January 4, 2014–March 18, 2014. *Source* the Stock Exchange of Thailand [15]

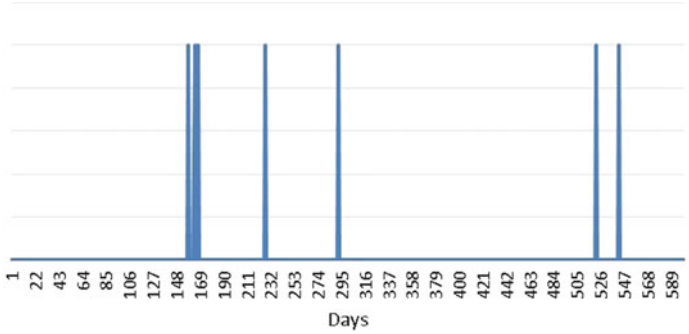


Fig. 3 The days on which the interested events occur for ADVANC

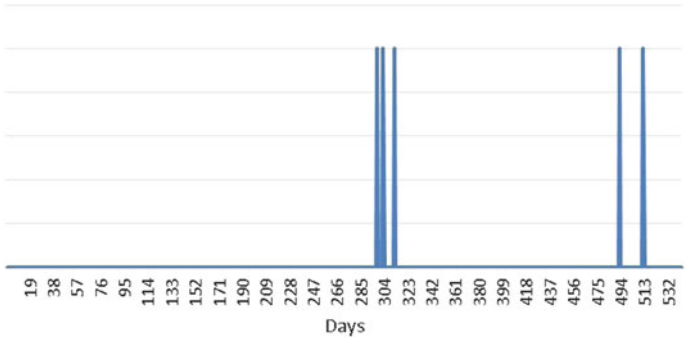


Fig. 4 The days on which the interested events occur for DTAC

Figures 3 and 4 display the days on which the interested events occur. ADVANC shows the events for seven days, namely Day 158, 164, 167, 227, 292, 522, and 542. DTACs events occur for five days, namely Day 298, 303, 312, 493, and 512.

The parameter estimation of the corresponding Hawkes process is carried out \based on the maximization of the following likelihood function [11, 14]:

$$\log L(t_1, \ldots, t_N|\theta) = -\mu t_N + \sum_{i=1}^{N} \frac{\alpha}{\beta}(e^{-\beta(t_N - t_i)} - 1) + \sum_{i=1}^{N} \log(\mu + \alpha\Omega(i)) \quad (11)$$

where $A(i) = \sum_{t_j < t_i} e^{-\mu(t_i - t_j)}$ for $i \geq 2$ and t_i denotes the time of occurrence of the i denotes the time of occurrence of the ith event $A(1) = 0$.

The parameters to be estimated are μ, α, and β, respectively. The determination of those parameters is carried out by the genetic algorithm in Mathlab program. The results are shown in Table 1.

Table 1 demonstrates that both ADVANC and DTAC have μ at 0.01 whereas μ means the base rate the process returns to. Meanwhile, α shows the rise of intensity after an event occurrence. α of ADVANC is lower than that of DTAC. β shows the exponential intensity decay which is higher in ADVANC than in DTAC. Moreover,

Table 1 Estimated parameters

	μ	α	β
ADVANC	0.01	0.04	0.14
DTAC	0.01	0.06	0.11

Table 1 shows $\alpha < \beta$ which confirms that the intensity decreases quicker than when it increases with the occurrence of the new events, otherwise the process could explode.

Afterwards, the estimators obtained are used to create a model:

$$\text{ADVANCE } = \lambda_t = 0.01(t) + \sum_{t_i < t} 0.04e^{-0.14(t-t_i)} \tag{12}$$

$$\text{DTAC } = \lambda_t = 0.01(t) + \sum_{t_i < t} 0.06e^{-0.11(t-t_i)} \tag{13}$$

According to the model, the time-dependent intensity $\lambda(t)$ of ADVANC and DTAC can be calculated as shown in Figs. 5 and 6, respectively.

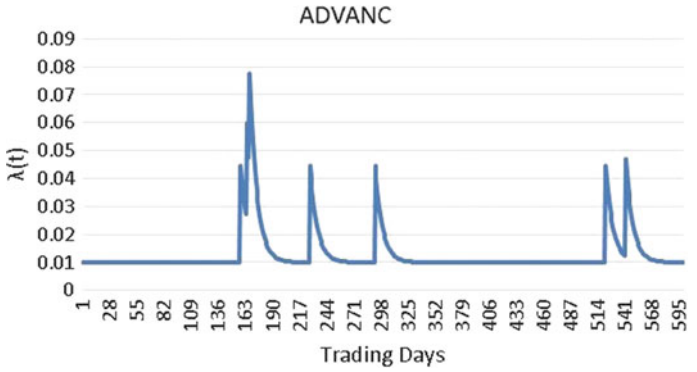


Fig. 5 The time-dependent intensity according to the ADVANC data

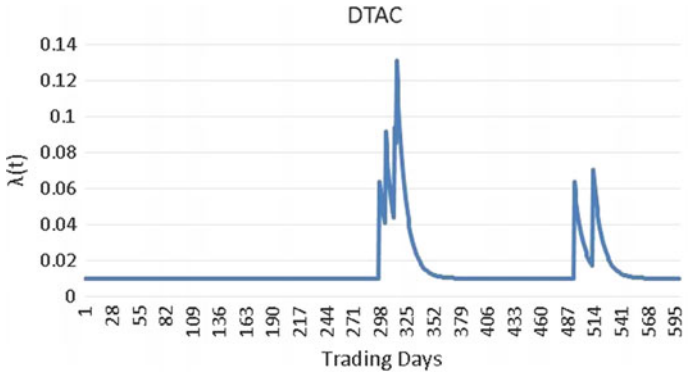


Fig. 6 The time-dependent intensity according to the DTAC data

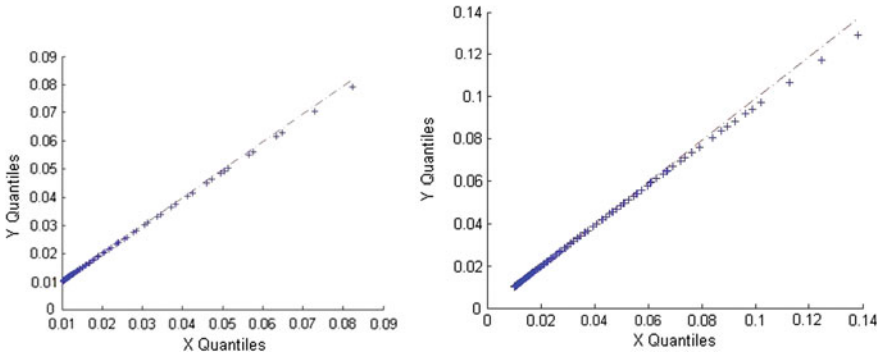


Fig. 7 The QQ plots

At this point, the estimators and the model are obtained, but there is no telling whether the created model is good enough, to which extent it can describe the events, and how accurate it can predict the future. Therefore, the goodness of fit of the model has to be checked.

The goodness of fit can be evaluated through a number of methods. One is by comparing Akaike information criterion (AIC) with a homogenous Poisson model, and another is by evaluating the residuals. According to Lorenzen [9], the residual process should be homogenous and have inter-event times, if the model is a good fit. As a result, a QQ-plot against an exponential distribution confirms this. Below is the plot that shows an excellent fit as it is closely to a 45-degree line (Fig. 7).

Branching Ratio

The estimators acquired in Table 1 can be used to calculate the branching ratios through Eq. 5 written as

$$\varsigma = \int_0^\infty \alpha e^{-\beta t} dt = \frac{\alpha}{\beta} \tag{14}$$

Low branching ratios indicate that the price change is more influenced by external factors than internal, an information which leads to different trading strategies performed by different types of investors. Stocks with a low branching ratio are more appropriate for speculation than those having a high branching ratio, as they are more likely to be volatile due to external factors than their own fundamentals.

According to Eq. (5), the branching ratio of ADVANC equals 0.29, while DTAC gives the branching ratio of 0.55. This result indicates that ADVANCs price change is only 29 % caused by internal factors, while the rest 71 % of the change is caused by external factors. Meanwhile, the change of DTAC price is 55 % due to internal factors and 45 % external. Accordingly, ADVANC is more speculative than DTAC, because external factors influence as much as 71 % of the price change, while DTAC, which has a 55 % branching ratio, is only 45 % affected by external factors.

When comparing the branching ratios calculated by the aforementioned method against the economic and political events in Thailand, the results obtained are in the same direction. ADVANC is greatly influenced by external factors, especially political factors. This is because the major shareholder of ADVANC is Shin Corporation Plc, the company founded by former Prime Minister Thaksin Shinawatra. When Thai and foreign traders believe Thaksin or a member of the Shinawatra family is likely to become the next prime minister, ADVANC receives a positive impact. When a political event occurs, ADVANCs price change tends to be more obvious than DTAC. If the event shows a positive impact on the Shinwatra family, ADVANC price usually goes up. On the contrary, when the situation negatively affects the family, ADVANC usually falls, while DTAC and TRUE are on the rise. Consider the news from Manager Newspaper [1], which reported that the entrance of Yingluck Shinawatra, former Prime Minister Thaksin Shinawatra's youngest sister, to Government House had caused the Shinawatra-related stocks to surge. One of the stocks was ADVANC which ended at 85 baht a share on December 30, 2010, before rising to 140 baht a year later. ADVANC's price soared to 209 baht as of December 30, 2012, but fell to 175.94 baht on December 27, 2013 following the anti-Yingluck movement. The share price continued to drop when it was becoming more and more obvious that Yingluck would be ousted. In addition, according to a KKTrade analysis [8], the National Broadcasting and Telecommunications Commission (NBTC) states that the political conflicts have led to a campaign against using AIS mobile numbers. The NBTCs inspection during February 20–21 shows that the mobile number transfers from AIS to other mobile network operators increase from 700 numbers a day to 1,400 numbers a day. Of the total, 70 % switch to DTAC and 30 % to TRUE. The event affects the stock sentiment. In the short run, concerns over the effects from the number transfers draw a psychological impact on the stock price, but a slight effect on ADVANCs fundamentals.

5 Conclusion and Further Study

This study focuses on the branching ratios acquired through the Hawkes process and uses the data from two telecommunication stocks listed on the Stock Exchange of Thailand, namely ADVANC and DTAC. The data set has 543 data points. The parameters are estimated in the Hawkes process using the MLE technique. The estimators obtained are inserted into an equation. Later, the QQ-plot is created to check whether the model fits the empirical data. The estimators are used to find the branching ratios. The results indicate how much the stock price changes are affected by internal factors. The study finds that the branching ratio of ADVANC is at 29 %, which means ADVANCs price change is only 29 % caused by internal factors, while the rest 71 % derives from external factors. Meanwhile, DTACs branching ratio is at 55 %, meaning DTACs price change is 55 % due to internal factors and 45 % external. Knowing to which extent the stock price is affected by external factors can strengthen investors strategy. Stocks with a low branching ratio are more speculative than those having a high branching ratio.

A further study will address the external factors that affect telecommunication stocks. Currently, we believe political events are a much influential factor affecting ADVANCs price. However, we will explore in details our expectation based on statistical data. We are also interested in finding the strategies investors should use with stocks having different branching ratios.

References

1. ASTV Manager Online. Retrieved from 1 January 2013, Manager Online: http://www.manager.co.th/Home/ViewNews.aspx?NewsID=9550000157897 (2013)
2. Bowsher, C.: Modelling securities market events in continuous time: intensity based, multivariate point process models. J. Econom. **141**(2) (2007)
3. Daley, D.J., Vere-Jones, D.: An Introduction to the Theory of Point Processes, vol. I. Springer, Heidelberg (2003)
4. Engle, R.F., Rangel, J.G.: The spline GARCH model for unconditional volatility and its global macroeconomic causes, vol. 13, pp. 1–28. CNB Working Papers Series (2005)
5. Filimonov, V., Sornette, D.: Quantifying reflexivity in financial markets: towards a prediction of flash crashes. arXiv:1201.3572v1.pdf (2012)
6. Hawkes, G.A.: Point spectra of some mutually exciting point processes. J. R. Stat. Soc. **33**(2) (1971)
7. Hewlett, P.: Clustering of order arrivals, price impact and trade path optimization. Workshop on Financial Modelling with Jump Processes, Ecole Polytechnique (2006)
8. KKtrade.: Advanced Info Service: ADVANCSocial Sanction: Retrieved from 24 February 2014, http://portal.settrade.com/brokerpage/IPO/Research/upload/2000000233022/Analyst20NoteADVANC240214.pdf (2006)
9. Lorenzen, F.: Analysis of Order Clustering using High Frequency Data: A Point Process Approach. Thesis. Tilburg School of Economics and Management Finance Department (2012)
10. McGregor, R.: The Mechanics of the Johannesburg Stock Exchange. Juta and Co Ltd, Cape Town (1989). ISBN 0-7021-2248-3
11. Ogata, Y.: The asymptotic behaviour of maximum likelihood estimators for stationary point processes. Ann. Inst. Stat. Math. **30**(1), 243–261 (1978)
12. Ogata, Y.: Statistical models for earthquake occurrences and residual analysis for point processes. J. Am. Stat. Assoc. **83**(401), 9–27 (1988)
13. Oyama, T.: Determinants of Stock Prices: The Case of Zimbabwe, Policy Development and Review Department, IMF Working Paper (1997)
14. Ozaki, T.: Maximum likelihood estimation of Hawkes self-exciting point processes, Part B. Ann. Inst. Stat. Math. **31**, 145–155 (1979)
15. SetSmart. (n.d.).: SET Market Analysis and Reporting Tools. Retrieved from SetSmart. http://www.setsmart.com/ism/login.jsp (2014)
16. Shauna, C.: Why do companies care about their stock prices? Available online at http://www.investopedia.com/articles/basics/03/020703.asp (2003)
17. Stanlake, G.F.: Introductory Economics. Longman House, Essex p. 222. (1993). ISBN 0582354846
18. Thailand Securities Institute: Impact Factors on Futures Price. Retrieved from Thailand Securities Institute (TSI). Available online at http://www.tsi-thailand.org/index.phpoption=comcontenttask=viewid=197
19. Toke, I.M., Pomponion, F.: Modelling trades-through in a limit order book using Hawkes processes. Econ.: Open-Access, Open-Assess. E-J. **6**, 2012–2022 (2012)

Forecasting Risk and Returns: CAPM Model with Belief Functions

Sutthiporn Piamsuwannakit and Songsak Sriboonchitta

Abstract This paper presents a CAPM model with a belief function approach for forecasting the Integrated Oil and Gas Company (CHK) stock and the S&P500 index. The approach composed of two steps. First, we estimate the systematic risk or the beta coefficient in the CAPM model using the maximum likelihood method. Second, to improve the forecasting performance, we incorporate the likelihood-based belief function method. Likelihood-based belief functions are calculated from the historical data. The data set contains of 209 weekly returns during the period of 2010–2013. The finding shows evidence on systematic risk which is associated by the belief function derived from the distribution likelihood function given the market return. Finally, we use the method to predict the return of a particular stock.

1 Introduction

Most investors focus on the stock market return forecasting. The aim is to gain high profit by using the best trading strategies. The more successful in stock return prediction, the more profitable it becomes in stock market investment. The uncertainty and volatility of stock prices have an effect on the investor's decision. The knowledge on the dependence pattern between stock and market returns can help portfolio investors to diversify their assets better as well as reducing their risk at the suitable moments. The Capital Asset Pricing Model (CAPM) is a foundation and widely used model for evaluating the risk of a portfolio of assets with respect to the market risk which was introduced by Sharpe [21]. The CAPM is a linear model that estimates asset prices using the information on the risk free rate and the market returns. The CAPM takes into account the non-diversifiable risk, which is captured by the parameter β. The CAPM non-diversifiable risk depends on the correlation between particular stock

S. Piamsuwannakit (✉) · S. Sriboonchitta
Faculty of Economics, Chiang Mai University, Chiang Mai, Thailand
e-mail: sutthiporn_p@cmu.ac.th

S. Sriboonchitta
e-mail: songsakecon@gmail.com

V.-N. Huynh et al. (eds.), *Econometrics of Risk*, Studies in Computational Intelligence 583, DOI 10.1007/978-3-319-13449-9_18

and overall stock market. Essentially, the standard CAPM model depends on the assumptions of normality of returns and quadratic utility functions of investors.

However, the numerous empirical evidences that have been carried out to analyze the applicability of CAPM in different stock markets have failed to maintain this relationship due to the inadequacy of the market beta alone in explaining the variations in stock returns and the assumptions of CAPM model. For example, Isa et al. [11] applied CAPM in the Malaysian stock market by using the linear regression method, which was carried out on four models. The result indicated that both of the standard CAPM models with constant beta and time varying beta are statistically insignificant. On the other hand, the CAPM models conditional on segregating positive and negative market risk premiums are statistically significant. Nikolaos [18] evaluated of CAPM's validity in the British Stock Exchange. The result showed that under the two steps procedure, the CAPM does not have a statistical significance in portfolio selection. Choudhary and Choudhary [6] applied the CAPM model for the Indian stock. There is a lack of substantiating the theory's basic result illustrating that there is higher risk (beta) is associated with higher levels of return. Masood et al. [15] examined the validity of the CAPM in the capital markets of the Pakistan. The least squares method (OLS) is used to find the beta of the stocks in the first step and then searches for the regression equations in second step. The result showed that there is no support with the CAPM. The intercept term is equal to zero. Also, there is a positive relation between the risk and return. In addition, the market risk premium is a significant explanatory variable for the determining to see if the stock's risk premium are rejected. Zhang and Meng [22] analyzed the CAPM model in the Chinese stock market. The main problem of their studies was found that the effective test method did not exist.

From the above literature reviewed, CAPM is a useful tool to estimate the stock market return in different stock index. It can be concluded that there is no one model that can claim to have the absolute ability to predict the expected stock return by using the standard CAPM model. Then, there is a need of accurate forecast model that consistently predict uncertainty and volatility of the stock market prices. The stock market investor would be able to make decisions on the investment that is more informed and accurate. Therefore, various techniques are used for handling the uncertainty data. One such method applied is the Dempster-Shafer belief function theory, which is a useful tool for forecasting. Many studies have applied the belief function model to predict the uncertainty data. For instance, Nampak et al. [17] used the belief function model in order to forecast groundwater of specific area in Malaysia. Abdallah et al. [3] cooperated the statistical judgements with expert evidence by using belief function for prediction the future centennial sea level which climate change is considered. Kanjanatarakul et al. [13] used the Bass model for innovation diffusion together with past sales data and the formalism of belief functions to quantify the uncertainty on future sales. In their studies, a piece of evidence as a belief function was considered which can be viewed as the distribution of a random set. Furthermore, two main reasons for using the belief function formalism in this paper are the following:

(1) The belief function approach does not require the statistician to arbitrarily provide a prior probability distribution when prior knowledge is not available.
(2) We wish to measure the weight of statistical evidence that pertains to some specific questions, whereas confidence and prediction intervals are related to sequences of trials.

For more discussion on the comparison bet the belief function approach and classical methods of inference, the reader can find more information with the regards to the work done by Kanjanatarakul et al. [13].

In this contributions, we propose and alternative method for drawing inference via a likelihood based on a belief function approach for estimation of linear regression of CAPM. The objectives of this study are to (1) analyze the dependence pattern between the CHK stock and market returns and to (2) forecast the CHK stock returns using belief functions.

The remainder of the paper is organized as follows. Section 2 provides the Maximum Likelihood Estimation of capital asset pricing model and Sect. 3 introduces the prediction machinery using belief functions. Section 4 discusses the empirical solutions to the forecasting problem. The last section summarizes the paper.

2 Maximum Likelihood Estimation of Capital Asset Pricing Model

The CAPM represents a positive and linear relationship between asset return and systematic risk relative the overall market. The linear regression model is defined as

$$E(R_i) - R_f = \alpha + \beta E(R_m - R_f) \tag{1}$$

where $E(R_i)$ is the expected return of the asset, R_m is the expected market portfolio return, R_f is the risk free rate, α is the intercept and β is the equity beta, representing market risk. The observed the historical returns of stock $R_i = (r_{i1}, \ldots, r_{in})$ and returns from market $R_m = (r_{m1}, \ldots, r_{mn})$. The estimator of β is a measure of risk for financial analysis and also for risk and portfolio managers. The parameter β estimation procedure is defined by Arellano-Valle et al. [1] Let us consider in Eq. (1) has extended into Eq. (2) as follow:

$$r_i - r_f = \alpha + \beta(r_{mi} - r_f) + \varepsilon_i \tag{2}$$

or

$$y_i = \alpha + \beta x_i + \varepsilon_i \tag{3}$$

where r_i denotes the return of stock i, r_m is the market return and r_f corresponds to the risk free return, so that

$$y_i = r_i - r_f \tag{4}$$

and

$$x_i = r_m - r_f \tag{5}$$

represent the return of an asset in excess of risk free rate and the excess return of the market portfolio of assets.

The estimation method with the considering in the financial model is based on the least squares theory under the assumption of the random errors $\varepsilon_1, \ldots \varepsilon_n$ are independent and identically distributed according to the normal distribution.

$$N(\varepsilon_i, 0, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} exp\left\{\frac{-1}{2\sigma^2}(y - x\beta)^2\right\} \tag{6}$$

The likelihood function is given by

$$L = \Pi_{n=1}^{n} N(y_i; x_i, \beta, \sigma^2) = (2\pi\sigma^2)^{\frac{-n}{2}} exp\left\{\frac{-1}{2\sigma^2}(y - x\beta)'(y - x\beta)\right\} \tag{7}$$

3 Statistical Inference and Prediction Using Belief Functions

3.1 Belief Functions

The theory of belief function is a formalism for reasoning with the uncertain, inaccurate and incomplete information. It was developed by Dempster [9] and later formalized by Shafer [20]. The model comprises several functions including Bel(degree of belief), Dis(degree of disbelief), Unc(degree of uncertainty) and Pls(degree of plausibilty), in range of [0, 1]. Belief function can be defined on finite set and infinite set. Let us begin with finite case.

3.1.1 Belief Functions on Finite Set

In the formalism of belief functions, we assign probabilities to sets (Pearl) [12]. The belief model as given below, see Frikha [10], Liu et al. [14], and Nampak et al. [17].

Let Θ be a finite set, Θ is called frame of discernment of the problem of consideration. The power set of Θ, denoted by 2^{Θ}.

A basic probability assignment (BPA) is a function $m(.)$ from 2^{Θ} to [0, 1] that assigns a number [0, 1] to each subset A of Θ. The quantity $m(A)$, called the mass of A, which represents the degree of belief attributed exactly to A, and to no one of its subsets. This function satisfies the following condition:

$$0 \leq m(A) \leq 1,\ m(\phi) = 0,\ \sum_{A \subseteq \Theta} m(A) = 1 \tag{8}$$

When $m(A) > 0$, A is called focal element of m. To each BPA, we can associate a belief function and a plausibility function are a mapping $Bel(A) : 2^{\Theta} \rightarrow [0, 1]$ and $Pl(A) : 2^{\Theta} \rightarrow [0, 1]$ respectively, defined as:

$$Bel(A) = \sum_{B \subseteq A} m(B) \tag{9}$$

$$pl(A) = \sum_{A \cap B \neq \phi} m(B) \tag{10}$$

$Bel(A)$ measures the total belief completely attributed to $A \subseteq \Theta$. It is interpreted as the lower bound of probability of A. $Pl(A)$ is interpreted as the upper bound of probability of A.

The two functions satisfied the following properties:

$$Bel(A) \leq Pl(A) \tag{11}$$

$$Pl(A) = 1 - Bel(\bar{A}) \tag{12}$$

where $\overline{A}$ is the complement of A and $Bel(\bar{A})$ is called a degree of disbelief in A.

$$Pl(A) - Bel(A) = Unc \tag{13}$$

Equation (13) represents the difference between belief and plausibility.

If $Unc = 0$, then $Bel(A) = Pl(A)$.

Figure 1 shows a schematic description of the relationship between belief, disbelief and uncertain functions.

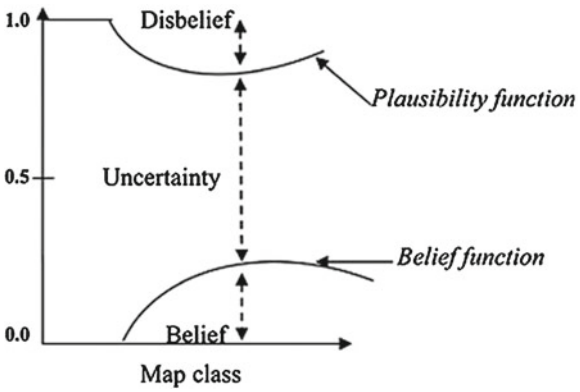


Fig. 1 Schematic description of the relationship between belief, disbelief and uncertainty [4]

3.1.2 Belief Functions on Infinite Set

In an infinite case, there may not be a mass function associated with completely monotone function as in the finite case, Denoeux [7]. The definitions are provided which defined by Denoeux [7] as following;

Let (Ω, B) be a measurable space (i.e., B is a sigma-field, that is a non-empty subset of 2^{Θ} closed under complementation and countable union). A belief function on B is a function $Bel : B \to [0, 1]$ verifying the following three conditions:

1. $Bel(\phi) = 0$
2. $Bel(\Omega) = 1$
3. For any $k \leq 2$ and any collection $B_1, \ldots, B_k$ of elements of B,

$$Bel(U_{i=1}^{k} B_i) \geq \sum_{\phi \neq I\,(1,\ldots,k)} (-1)^{|I|+1} Bel(\cap_{i \in I} B_i) \tag{14}$$

Furthermore, a belief function Bel on (Ω, B) is continuous if for any decreasing sequence $B_1 \supset B_2 \supset B_3 \supset \cdots$ of elements of B,

$$\lim_{i \to +\infty} Bel(B_i) = Bel(\cap_{i \in I} B_i) \tag{15}$$

3.2 Likelihood-based Belief Functions

The likelihood-based belief functions have been derived by Shafer [20]. They have been applied by Abdallah et al. [3], among others, and justified by Denoeux [8].

Let $x \in X$ be the observable data with a probability density function (pdf) $p_\theta X$, where $\theta \in \Theta$ is an unknown parameter. In this paper, we use the method proposed by Shafer [20]. The belief function be derived from the Likelihood Principle and Least Commitment Principle(LCP). The information about Θ can be represented by the likelihood function which is defined by $L_x(\theta) = p_\theta X$ for all $\theta \in \Theta$. The likelihood ratio is meant to be a "relative plausibility", which can be written as:

$$\frac{pl_x(\theta_1)}{pl_x(\theta_2)} = \frac{L_x(\theta_1)}{L_x(\theta_2)} \tag{16}$$

for all $(\theta_1, \theta_2) \in \Theta^2$ or, equivalently, $pl_x(\theta) = cL_x(\theta)$
for all $\theta \in \Theta$ and some positive constant c. From LCP, it can be implied that the highest possible value of c is $\frac{1}{sup_{\theta \in \Theta}} L(\theta|x)$. Thus, the contour function is defined as follow:

$$pl(\theta; x) = \frac{L(\theta; x)}{sup_{\theta \in \Theta} L(\theta; x)} \tag{17}$$

The information about θ are expressed by the belief function Bel_A^{Θ} with contour function pl_x, i.e., with corresponding plausibility function $pl_x^{\Theta}(A) = sup_{\theta \in A} pl_x(A)$, for all $A \subseteq \Theta$ The focal sets of Bel_A^{Θ} are the levels sets of pl_x defined as follows:

$$\Gamma_x(\omega) = \{\theta \in \Theta | pl_x(\theta) \geq \omega\} \tag{18}$$

for $\theta \in [0, 1]$. Equation (18) is called plausibility regions. With the inducing of the Lebesgue measure λ on [0,1] and multi-valued mapping Γ_x from $[0, 1] \rightarrow \Theta^2$ the belief function is equivalent to the random set, see Kanjanatarakul et al. [13]. We remark that the MLE of θ is the value of θ with highest plausibility.

3.3 Incorporating the Belief Functions

The objective of this section is to forecast the risk premium of the return of stock i, $y_i = r_i - r_f$. The methodology to incorporate the belief function framework into the prediction procedure follows Kanjanatarakul et al. [13]. From the CAPM equation from the previous section, the return equation can be written as:

$$y_i = \alpha + \beta x + \sigma F^{-1}(u) \tag{19}$$

where $F \sim Normal(0, 1)$ and $U \sim Uniform(0, 1)$

As discussed in Kanjanatarakul et al. [13], the forecasting problem is the inverse problem of the regular inference problem. Given the knowledge on the set of parameters $\theta = (\alpha, \beta, \sigma)$ and the distribution $F(.)$, the future value of y_i can be forecasted.

Belief function framework allows us to forecast an interval $[y_i^L, y_i^U]$ for the future value of y_i. The estimation of $[y_i^L, y_i^U]$ can be done using Monte Carlo method. Given a set of two independently $Uniform(0, 1)$ random variables (u_s, ω_s), in each simulation s, the lower bound $y_{i,s}^L$ and the upper bound $y_{i,s}^U$ solve the following optimization problems respectively,

$$y_{i,s}^L = min_{\theta} \alpha + \beta x + \sigma F^{-1}(u_s) \tag{20}$$

subject to

$$pl(\theta) \geq \omega_s \tag{21}$$

and

$$y_{i,s}^U = max_{\theta} \alpha + \beta x + \sigma F^{-1}(u_s) \tag{22}$$

subject to

$$pl(\theta) \geq \omega_s \tag{23}$$

In the constraints, the plausibility function $pl(\theta)$ can be derived from the likelihood function. Therefore, using the likelihood function in Eq. (7), the plausibility function is as follows:

$$pl(\theta) = \frac{L(\theta)}{L(\theta^*)} \tag{24}$$

where θ^* is such that $L(\theta^*) \geq L(\theta)$, $\forall\theta$. The belief and the plausibility functions corresponding to a given set A can be calculated by:

$$Bel(A) = \frac{1}{N}\sharp\{s \in \{1, \cdots, N\} | [y_{i,s}^L, y_{i,s}^U] \subset A\} \tag{25a}$$

$$Pl(A) = \frac{1}{N}\sharp\{s \in \{1, \cdots, N\} | [y_{i,s}^L, y_{i,s}^U] \cap A \neq \emptyset\} \tag{25b}$$

The lower and the upper of the expectation for y_i is, thus,

$$\hat{y_i^L} = E(y_{i,s}^L) = \frac{1}{N}\Sigma y_{i,s}^L \tag{26a}$$

$$\hat{y_i^U} = E(y_{i,s}^U) = \frac{1}{N}\Sigma y_{i,s}^U \tag{26b}$$

4 An Application to Stock Market

4.1 Data

The data contain of 209 weekly returns during the period of 2010–2013: they were obtained from Yahoo Finance to compute the log returns on integrated oil and gas company (CHK) stock. The log returns prices by using the formula:

$$r_t = ln(\frac{P_t}{P_{t-1}}) \tag{27}$$

where P_t and P_{t-1} are the weekly closing prices at time t and $t-1$ respectively. Mukherji [16] indicated that the treasury bills are better proxies for the risk-free rate than longer-term treasury securities regardless of the investment horizon, which is only related to the U.S. market. In this paper, the treasury bills stand for the risk free rate. The daily returns of the treasury bills are adjusted to the weekly returns and can be used in this manner by using the compound interest that take form:

$$I_{wj} = \{\prod_{i=1}^{N}(1 + I_{di})\} - 1 \tag{28}$$

Table 1 Parameter estimation results

Stock name	Parameters	
CHK	β_0	-0.001**(0.0031)
	β_1	1.436**(0.1417)
	σ^2	.0020**$(1.91739e^{-4})$

The ** shows significant at 5 % level. Standard errors in parentheses

where I_{wj}, $j = 1, N$ is the weekly interest rate and I_{di}, $i = 1, N$ is the daily interest rate. The Maximum Likelihood estimates of the parameters are shown in Table 1

Figure 2 displays two-dimensional marginal contour functions, with one of the three parameters fixed to its MLE.
Figure 3 shows the marginal contour functions for parameters β_0, β_1, σ^2. These three plausibilities will be used to perform plausibility intervals for each of the three parameters.

To predict the expected return of the asset $y_{i,n+1}$ for a new market portfolio return $X_{i,n+1}$ we compute the minimum and maximum of $y_{i,n+1}$ given $X_{i,n+1}$ by

$$y_{i,n+1} = \beta_0 + \beta_1 X_{i,n+1} + \sigma F^{-1}(u_s) \tag{29}$$

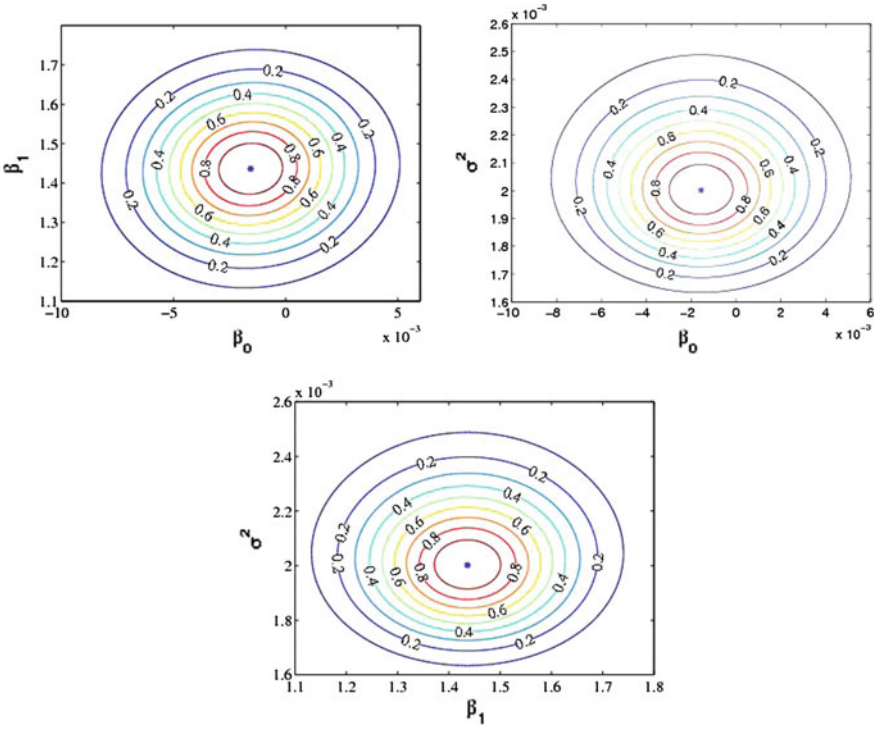


Fig. 2 Displays two-dimensional marginal contour functions

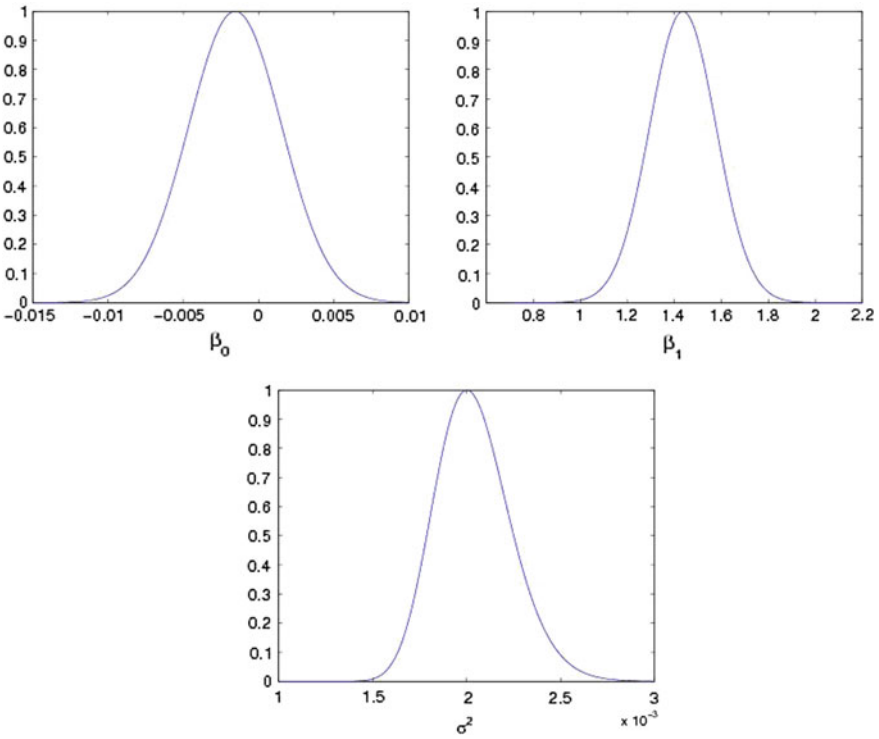


Fig. 3 Marginal plausibility of β_0, β_1 and σ^2

under the constraint $pl(\theta) \geq \omega_s$, where $F^{-1}(u_s)$is the inverse cumulative distribution function (cdf) of the normal distribution and u, ω are independent random variables with the same uniform distribution U ([0, 1]). Given (29), we randomize independently N pairs of the random number, (u_s, ω_s); $s = 1, 2, N$ resulting in N intervals $[y^L_{i,s}(u_s, \omega_s), y^U_{i,s}(u_s, \omega_s]$ For any $A \subset R$, the stock returns $Bel_{y_i}(A)$ and $Pl_{y_i}(A)$ can be estimated by Eq. (3). The estimated lower and upper expectations of $r_{a,n+1}$ are then:

$$\overline{y}^L_{i,s} = \sum_{s=1}^{N} \frac{y^L_s(u_s, \omega_s)}{N} \tag{30}$$

$$\overline{y}^U_{i,s} = \sum_{s=1}^{N} \frac{y^U_s(u_s, \omega_s)}{N} \tag{31}$$

Figure (4) displays the lower and upper cdfs $Bel_{y_i}([-\infty, y_i])$ and $Pl_{y_i}([-\infty, y_i])$. This function give us the summary of the predictive belief function Bel_{y_i}.
Figure (5) shows the upper and lower bound of stock return via CAPM using belief function.

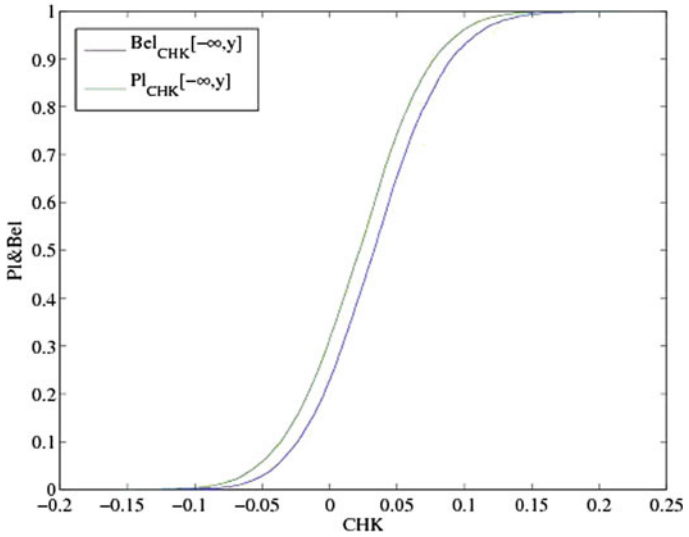


Fig. 4 *Lower* and *Upper* cumulative distribution function

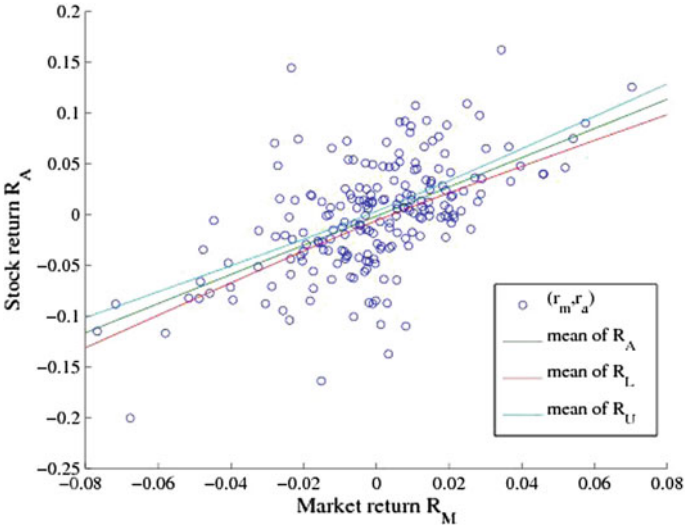


Fig. 5 *Lower* and *Upper* interval of stock return via CAPM using belief function

The another representation of uncertainty prediction can be defined as the lower-upper expectations of stock returns, the uncertainty and randomness estimation are considered. From the empirical result, the gap between the lower and upper cdfs is quite narrow, which shows that estimation uncertainty is small as compared to random uncertainty. Therefore, the investor can use these results to increase the gain of portfolio investment Autchariyapanikul et al. [2].

5 Conclusions

In this paper, we presented the method of standard CAPM with normal distribution for CHK stock in S&P500 in the belief function framework. The Dempster-Shafer belief function theory was used in order to identify the uncertainty. The statistical prediction based on historical data and a financial model. This method consists of two steps. First, a belief function is defined from the normalized likelihood function given the past data which is refered to the uncertainty on the parameter vector θ. Second, the return of stock y_i is illustrated as $\varphi(\theta, u)$, where u is a stochastic variable with known distribution. Then, belief on θ and u are transferred through φ, resulting in a belief function on y_i. This approach has been adapted to the prediction of the stock returns. A possible extension of this work is to consider uncertainty on the independent variable r_m, which can also be expressed as a belief function and combined with other uncertainties to compute a belief function on y_i.

Acknowledgments We would like to thank Prof. Dr. Thierry Denoeux for his comments and suggestions, and Somsak Chanaim, a research assistant at Faculty of Economics, Chaing mai University for his helpful program running output.

References

1. Arellano-Valle, R.B., Bolfarine, H., Iglesias, P.L., Viviani, P.: Portfolio selection: an application to the Chilean stock market. Chil. J. Stat. **1**(2), 3–15 (2010)
2. Autchariyapanitkul, K., Chanaim, S., Sriboonchitta S., Denoeux, T.: Predicting stock returns in the capital asset pricing model using quantile regression and belief functions. In: Proceedings of the 3rd International Conference on Belief Functions (BELIEF 2014). Oxford, Springer (2014)
3. Ben Abdallah, N., Mouhous-Voyneau, N., Denoeux, T.: Combining statistical and expert evidence using belief functions: Application to centennial sea level estimation taking into account climate change. Int. J. Approx. Reason. **55**(1), 341–354 (2014)
4. Carranza, E.J.M., Woldai, T., Chikambwe, E.M.: Application of data-driven evidential belief functions to prospectivity mapping for aquamarine-bearing pegmatites, Lundazi district Zambia. Nat. Resour. Res. **14**(1), 47–63 (2005)
5. Chang, E.C., Cheng, J.W., Khorna, A.: (2005) An examination of Herd behavior in equity markets: an international perspective? J. Bank. Financ. **24**, 1651–1679 (2000)
6. Choudary, K., Choudhary, S.: Testing capital asset pricing model: empirical evidences from Indian equity market. Eurasian J. Bus. Econ. **3**(6), 127–138 (2010)
7. Denoeux, T.: Belief functions on infinite spaces[class handout]. Faculty of Economics, Chiang Mai University, Chiang Mai (2013)
8. Denoeux, T.: Likelihood-based belief function: justication and some extensions to low-quality data. Int. J. Approx. Reason. **55**(537), 15–1547
9. Dempster, A.P.: Upper and lower probabilities induced by a multivalued mapping. Ann. Math. Stat. **38**, 325–339 (1967)
10. Frikha, A.: On the use of a multi-criteria approach for reliability estimation in belief function theory. Inf. Fusion **18**, 20–32 (2014)
11. Isa, M., Hassan, A., Puah, C.H., Yong, Y.K.: Risk and return nexus in Malaysian stock market: Empirical evidence from CAPM. Online at http://mpra.ub.uni-muenchen.de/12355/ (2008)

12. Pearl, J.: Reasoning with belief functions: an analysis of compatibility. Int. J. Approx. Reason. **4**(5), 363–389 (1990)
13. Kanjanatarakul, O., Sriboonchitta, S., Denoeux, T.: Forecasting using belief functions: an application to marketing econometrics. Int. J. Approx. Reason. **55**(5) 1113–1128 (2014)
14. Liu, Z.G., Pan, Q., Dezert, J., Mercier, G.: Credal classification rule for uncertain data based on belief functions. Pattern Recognit. **47**(7), 2532–2541 (2014)
15. Masood, S., Saghir, G., Muhammad, W.: The Capital Asset Pricing Model: Empirical Evidence from Pakistan. Online at http://mpra.ub.uni-muenchen.de/41961/ (2012)
16. Mukherji, S.: The capital asset pricing model's risk-free rate. Int. J. Bus. Financ. Res. **5**, 793–808 (2011)
17. Nampak, H., Pradhan, B., Manap, M.A.: Application of GIS based data driven evidential belief function model to predict groundwater potential zonation. J. Hydrol. **513**, 283–300 (2014)
18. Nikolaos, L.: An empirical evaluation of CAPM?s validity in the British stock exchange. Intern. J. Appl. Math. Inf. **3**(1), 1–8 (2009)
19. Shalit, H., Yitzhaki, S.: Estimating beta. Rev. Quant. Financ. Account. **18**(2), 95–118 (2002)
20. Shafer, G.: Math. Theory Evid. Princeton University Press, Princeton, NJ (1976)
21. Sharpe, W.F.: Capital asset prices? A theory of market equilibrium under conditions of risk. J. Financ. **19**(3): 425–442 (1964)
22. Zhang, P., Meng, X.: The market application analysis of CAPM model in China's securities. In: 2nd International Conference System Engineering and Modeling (ICSEM-13). Atlantis Press (2013)

Correlation Evaluation with Fuzzy Data and its Application in the Management Science

Berlin Wu, Wei-Shun Sha and Juei-Chao Chen

Abstract How to evaluate an appropriate correlation with fuzzy data is an important topic in the educational and psychological measurement. Especially when the data illustrate uncertain, inconsistent and incomplete type, fuzzy statistical method has some promising features that help resolving the unclear thinking in human logic and recognition. Traditionally, we use Pearson's Correlation Coefficient to measure the correlation between data with real value. However, when the data are composed of fuzzy numbers, it is not feasible to use such a traditional approach to determine the fuzzy correlation coefficient. This study proposes the calculation of fuzzy correlation with three types of fuzzy data: interval, triangular and trapezoidal. Empirical studies are used to illustrate the application for evaluating fuzzy correlations. More related practical phenomena can be explained by this appropriate definition of fuzzy correlation.

1 Introduction

Traditional statistics reflects the results from a two-valued logic world, which often reduces the accuracy of inferential procedures. To investigate the population, people's opinions or the complexity of a subjective event more accurately, fuzzy logic should be utilized to account for the full range of possible values. Especially, when

B. Wu (✉)
Department of Mathematical Sciences, National Cheng Chi University, Taipei, Taiwan
e-mail: berlin@nccu.edu.tw

W.S. Sha
Graduate Institute of Business Administration, Fu Jen Catholic University, New Taipei, Taiwan
e-mail: P581122@ms57.hinet.net

J.C. Chen
Department of Statistics and Information Science, Fu Jen Catholic University, New Taipei, Taiwan
e-mail: 006884@mail.fju.edu.tw

V.-N. Huynh et al. (eds.), *Econometrics of Risk*, Studies in Computational Intelligence 583, DOI 10.1007/978-3-319-13449-9_19

dealing with psychometric measures, fuzzy statistics provides a powerful research tool. Since Zadeh [1] developed fuzzy set theory, its applications have been extended to traditional statistical inferences and methods in social sciences, including medical diagnosis or stock investment systems. For example, a successive series of studies demonstrated approximate reasoning methods for econometrics [2–4] and a fuzzy time series model to overcome the bias of stock markets was developed [5].

Within the framework of classical statistical theory, observations should follow a specific probability distribution. However, in practice, the observations are sometimes described by linguistic terms such as "Very satisfactory", "Satisfactory", "Normal", "Unsatisfactory", "Very unsatisfactory", or are only approximately known, rather than equating with randomness. How to measure the correlation between two variables involving fuzziness is a challenge to the classical statistical theory. The number of studies which focus on fuzzy correlation analysis and its application in the social science fields has been steadily increasing [6–9]. For example [9, 10] define a correlation formula to measure the interrelation of intuitionist fuzzy sets. However, the range of their defined correlation is from 0 to 1, which contradicts with the conventional awareness of correlation which should range from -1 to 1. An article [11] also has the same problems of lying the correlations between 0 and 1 for the interval valued fuzzy numbers. In order to overcome this issue, [12] takes random sample from the fuzzy sets and treat the membership grades as the crisp observations. Their derived coefficient is between -1 and 1; however, the sense the fuzziness is gone [8] calculated the fuzzy correlation coefficient based on Zadeh's extension principles. They used a mathematical programming approach to derive fuzzy measures based on the classical definition of the correlation coefficient. Their derivation is quite promising, but in order to employ their approach, the mathematical programming is required.

In addition, most previous studies deal with the interval fuzzy data, their definitions cannot deal with triangle or trapezoid data. In addition, formulas in these studies are quite complicated or required some mathematical programming which really limited the access of some researchers with no strong mathematical background. In this study, we give a simple solution of a fuzzy correlation coefficient without programming or the aid of computer resources. In addition, the provided solutions are based on the classical definition of Pearson correlation which should quite easy and straightforward. The definitions provided in this study can also be used for interval-valued, triangular and trapezoid fuzzy data.

Traditionally, if one wishes to understand the relationship between the variables x and y, the most direct and simple way is to draw a scatter plot, which can approximately illustrate the relationship between these variables: positive correlation, negative correlation, or zero correlation. The issue at hand is how to measure the relationship in a rational way. Statistically, the simplest way to measure the linear relationship between two variables is using Pearson's correlation coefficient, which expresses both the magnitude and the direction of the relationship between the two variables with a range of values from 1 to -1. However, Pearson correlations can only be applied to variables that are real numbers and is not suitable for a fuzzy dataset.

When considering the correlation for fuzzy data, two aspects should be considered: centroid and data shape. If the two centroids of the two fuzzy dataset are close, the correlation should be high. In addition, if the data shape of the two fuzzy sets is similar, the correlation should also be high. An approach to dealing with these two aspects simultaneously will be presented later in this study. Before illustrating the approach of calculating fuzzy correlations, a review of fuzzy theory and fuzzy datasets are presented in the next section.

2 Fuzzy Theory and Fuzzy Data

Traditional statistics deals with single answers or certain ranges of the answer through sampling surveys, but it has difficulty in reflecting people's incomplete and uncertain thoughts. In other words, these processes often ignore the intriguing and complicated yet sometimes conflicting human logic and feeling. For example, we would like to investigate the a person's favorite topics. In this case, consider a fuzzy set of favorite topics for a person as shown in Table 1. Note that in the extreme cases when a degree is given as 1 or 0, that is "like" or "dislike", a standard "yes" and "no" are in a complementary relationship, as in binary logic. Let A_1 represent for "favorite topics", A_2 "dislike the topics".

Based on the analysis of binary logic, we can find that he likes culture, religions and finance but dislikes politics and recreation. On the other hand, the fuzzy statistical result can be represented as:

$$\begin{aligned}\mu_{A_1} &= 0I_{\text{politics}}(x) + 0.8I_{\text{culture}}(x) \\ &\quad + 0.6I_{\text{religions}}(x) + 0.9I_{\text{finance}}(x) + 0.3I_{\text{recreation}}(x); \\ \mu_{A_2} &= 1I_{\text{politics}}(x) + 0.2I_{\text{culture}}(x) \\ &\quad + 0.4I_{\text{religions}}(x) + 0.1I_{\text{finance}}(x) + 0.7I_{\text{recreation}}(x).\end{aligned}$$

This means that the person likes the topic of politics 0 %, culture 80 %, and religion 60 %. etc. He dislikes the topic of finance 10 %, dislikes culture 20 %, dislikes religion 40 %, and dislikes recreation with 70 %. The percentages for each category represent the degree of their perceptions based on their own concept.

Table 1 Comparing fuzzy numbers with crisp numbers

	Fuzzy	Logic	Binary	Logic
Favorite topics	A_1 = like	A_2 = dislike	A_1 = like	A_2 = dislike
Politics	0	1		V
Culture	0.8	0.2	V	
Religions	0.6	0.4	V	
Finance	0.9	0.1	V	
Recreation	0.3	0.7		V

Therefore, based on the binary (like or dislike) logic, we can see only the superficial feeling about people's favorite topics. With the information of fuzzy response we will see a more detailed data representation. In illustrating human feelings with degrees, we encounter the problems that measurement cast the uncertainty and fuzzy property. Hence, a precise explanation about fuzzy numbers is illustrative and convincing.

2.1 Continuous Fuzzy Data

Continuous fuzzy data has been widely used in many applications. It can be classified into several types, such as interval-valued numbers, triangular numbers, trapezoid numbers, and exponential numbers. Typically, the nomenclature is based on the shape of the membership function. Even though there are various types of fuzzy numbers, here we limit the discussion to three usual types: interval-valued numbers, triangular numbers and trapezoid numbers. The definitions of the three types of fuzzy data are given as follows.

Definition 1 A fuzzy number $X = [a, b, c, d]$ defined on the universe set U of real numbers R with its vertices $a \leq b \leq c \leq d$, is called a trapezoidal fuzzy number if its membership function is given by

$$u_A(x) = \begin{cases} \dfrac{x-a}{b-a}, & a \leq x \leq b \\ 1, & b \leq x \leq c \\ \dfrac{d-x}{d-c}, & c \leq x \leq d \\ 0, & \text{otherwise} \end{cases}$$

When $b = c$, X is called a triangular fuzzy number; when $a = b$ and $c = d$, X is called an interval-valued fuzzy number.

2.2 Collecting Continuous Fuzzy Data

Respondents choose one single answer or certain range of the answer in traditional sampling surveys. But traditional methods are not able to truly reflect the complex thoughts of each respondent. If people can express the degree of their feelings by using membership functions, the answer presented will be closer to real human thoughts. But unfortunately scholars disagree in opinion about the construction of continuous fuzzy data. Many studies use continuous fuzzy without describing the construction method. The core of all the questions is fuzzy data determined by its membership function, but the construction of membership function is quite subjective. To reflect this, we ask the respondents to determine the membership function on Geometer's Sketchpad (GSP).

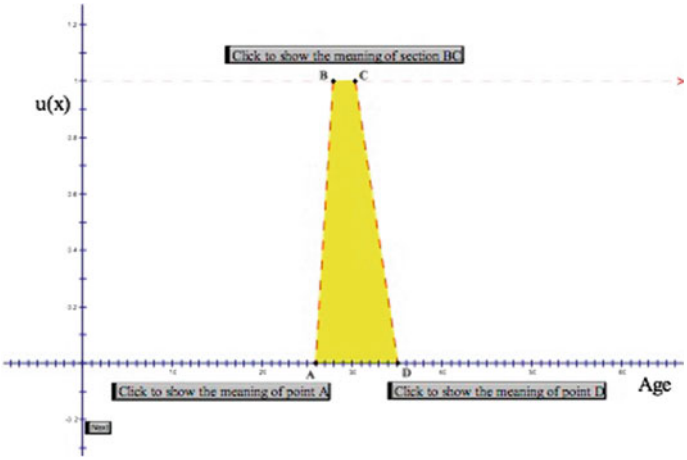


Fig. 1 A fuzzy answer for the expected marriage age

Figure 1 is the image of a fuzzy questionnaire item querying the prime time for marriage. Before answering the fuzzy questionnaire, respondents could click the three buttons to realize the meaning of each section and points. For example, people may decide: $\overline{AB}$ which represents the desire for marriage grows continuously for 2 years from 26 to 28, $\overline{BC}$ represents the desire for optimal marriage is 28–30, $\overline{CD}$ represents the desire for marriage falling continuously from 30 until it reaches 35.

Respondents can decide their own membership function of the prime time for marriage by moving the four points A, B, C, and D. By moving the four points, the age corresponding to the points will be changed automatically. There are probably three types of fuzzy data: The first is trapezoid; the second is triangular; the third is interval-valued type. Figure 2 illustrates these three kinds of fuzzy data. Triangular

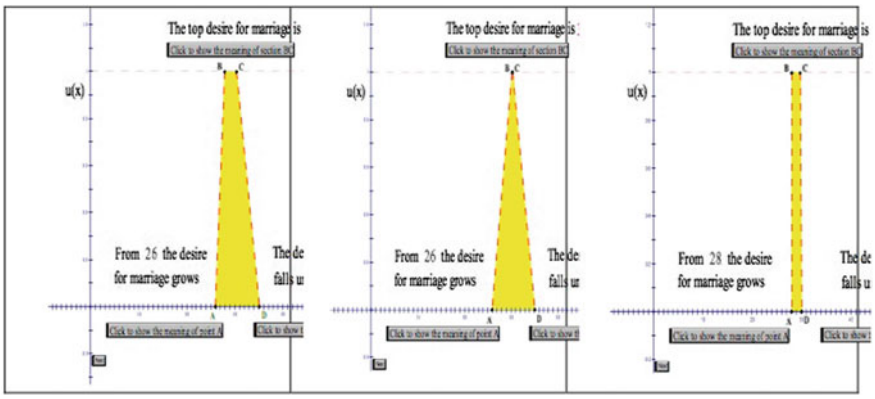


Fig. 2 Fuzzy observation for idea marriage year

data is a special case of trapezoid when point B equals to point C. It represents the prime time for marriage is only 30. The interval valued data shows the prime time for marriage is 28–30.

3 Fuzzy Correlation

The correlation coefficient is a commonly used statistics that presents a measure of how two random variables are linearly related in a sample. The population correlation coefficient, which is generally denoted by the symbol ρ is defined for two variables x and y by the formula:

$$\rho = \frac{\sigma_{X,Y}}{\sigma_X \sigma_Y} = \frac{Cov(X, Y)}{\sigma_X \sigma_Y}.$$

In this case, the more positive ρ is, the more positive the association is. This also indicates that when ρ close to 1, an individual with a high value for one variable will likely have a high value for the other, and an individual with a lower value for one variable will likely to have a low value for the other. On the other hand, the more negative ρ is, the more negative the association is, this also indicate that an individual with a high value for one variable will likely have a low value for the other when ρ is close to -1 and conversely. When ρ is close to 0, this means there is little linear association between two variables. In order to obtain the correlation coefficient, we need to obtain σ_X^2, σ_Y^2, and the covariance of x and y. In practice, these parameters for the population are unknown or difficult to obtain. Thus, we usually use r_{xy}, which can be obtained from a sample, to estimate the unknown population parameter. The sample correlation coefficient is expressed as:

$$r_{xy} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2}}, \tag{1}$$

where (x_i, y_i) is the ith pair observation value, $i = 1, 2, \ldots, n$, $\bar{x}$ and $\bar{y}$ are sample means for x and y respectively.

Pearson correlation is a straightforward approach to evaluate the relationship between two variables. However, if the variables considered are not real numbers, but fuzzy data, the formula above is problematic. For example, Mr. Smith is a new graduate from college; his expected annual income is 50,000 dollars. However, he can accept a lower salary if there is a promising offer. In his case, the annual income is not a definite number but more like a range. Mr. Smith's acceptable salary range is from 45,000 to 50,000. We can express his annual salary as an interval [45,000, 50,000]. In addition, when Mr. Smith has a job interview, the manager may ask how many hours he can work per day. In this case, Mr. Smith may not be able to provide

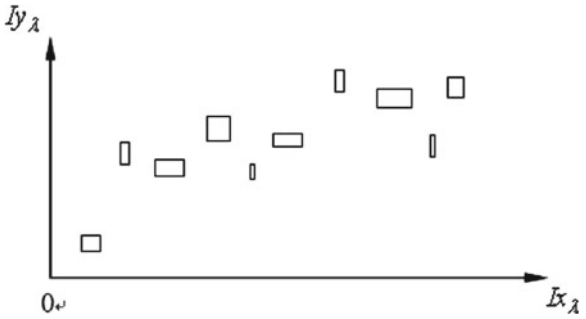


Fig. 3 Fuzzy correlation with interval data

a definite number since his everyday schedule is different. However, Mr. Smith may tell the manger that his expected working hours per day is an interval [8, 10].

We know Mr. Smith's expected salary ranges from [45,000, 50,000] and his expected working hours are [8, 10]. If we collect this kind of data from many new graduates, how can we use this data and calculate the correlation between expected salary and working hours? Suppose I_x is the expected salary for each new graduate, I_y is the number of working hours they desired, then the scatter plot for these two sets of fuzzy interval numbers would approximate that shown in Fig. 3.

For the interval valued fuzzy number, we need to take out samples from population X and Y. Each fuzzy interval data for sample X has centroids x_i, and for sample Y has centroids y_i. For the interval data, we also have to consider whether the length of interval fuzzy data are similar or not. In Mr. Smith's example, if the correlation between the expected salary and working hours are high, then we can expect two things: (1) the higher salary the new employee expects, the more working hours he can endure; (2) the wider the range of the expected salary, the wider the range of the working hours should be. However, how should one combine the information from both centroid and length? In addition, the effect of length should not be greater than the impact of centroids. In order to get the rational fuzzy correlations, we used natural logarithms to make some adjustments.

Let $(X_i = [a_i, b_i, c_i, d_i], Y_i = [e_i, f_i, g_i, h_i]; i = 1, 2, \ldots, n)$ be a sequence of paired trapezoid fuzzy sample on population Ω with its pair of centroids (cx_i, cy_i) and pair of areas $(\|x_i\| = area(x_i), \|y_i\| = area(y_i))$. The adjusted correlation for the pair of area will be

Definition 2 Let $(X_i = [a_i, b_i, c_i, d_i], Y_i = [e_i, f_i, g_i, h_i]; i = 1, 2, \ldots, n)$ be a sequence of paired trapezoid fuzzy sample on population Ω with its pair of centroids (cx_i, cy_i) and pair of areas $\|x_i\| = area(x_i), \|y_i\| = area(y_i)$. Let

$$cr_{xy} = \frac{\sum_{i=1}^{n}(x_i - \overline{cx})(cy_i - \overline{cy})}{\sqrt{\sum_{i=1}^{n}(cx_i - \overline{cx})^2}\sqrt{\sum_{i=1}^{n}(cy_i - \overline{cy})^2}};$$

$$ar_{xy} = \frac{\sum_{i=1}^{n}(\|x_i\| - \overline{\|x\|})(\|y_i\| - \overline{\|y\|})}{\sqrt{\sum_{i=1}^{n}(\|x_i\| - \overline{\|x\|})^2}\sqrt{\sum_{i=1}^{n}(\|y_i\| - \overline{\|y\|})^2}}. \quad (2)$$

Then fuzzy correlation is defined as as

$$FC = \beta_1 cr_{xy} + \beta_2 ar_{xy}, \quad (\beta_1 + \beta_2 = 1).$$

We choose a pair of (β_1, β_2) depending on the weight of practical use. For instance, if we think the location correlation is much more important than that of area scale, $\beta_1 = 0.7$ and $\beta_2 = 0.3$ will be a good suggestion.

Example 1 Suppose we have the following data as shown in Table 2.

In this case, the correlation between the two centroids is

$$cr_{xy} = \frac{\sum_{i=1}^{n}(x_i - 26.62)(cy_i - 1.7)}{\sqrt{\sum_{i=1}^{n}(cx_i - 26.62)^2}\sqrt{\sum_{i=1}^{n}(cy_i - 1.7)^2}} = 0.17;$$

Table 2 Numerical example for interval-valued, triangular, and trapezoidal fuzzy data

X			
Student	Data	Centroid	Area (length)
A	[23, 25]	24	2
B	[21, 23, 26]	23.3	2.5
C	[26, 27, 29, 35]	28.3	5.5
D	[28, 30]	29	2
E	[25, 26, 28, 35]	28.5	6
(fuzzy) mean	[24.6, 25.12, 29, 30.2]	26.62	3.6
Y			
Student	Data	Centroid	Area (length)
A	[1, 2]	1.5	1
B	[0, 2, 3]	1.7	1.5
C	[0, 1]	0.5	1
D	[1, 2, 4]	2.3	1.5
E	[1, 2, 3, 4]	2.5	2
(fuzzy) mean	[0.6, 1.4, 2, 2.8]	1.7	1.4

$$ar_{xy} = \frac{\sum_{i=1}^{n}(\|x_i\| - 3.6)(\|y_i\| - 1.4)}{\sqrt{\sum_{i=1}^{n}(\|x_i\| - 3.6)^2}\sqrt{\sum_{i=1}^{n}(\|y_i\| - 1.4)^2}}.$$

Considering the contribution of (area) length correlation to the fuzzy correlation, the idea of correlation interval is proposed. Suppose we fix the (area) length correlation by the following adjusted values

$$\lambda ar_{xy} = 1 - \frac{\ln(1 + |ar_{xy}|)}{|ar_{xy}|}.$$

Since $-1 \leq ar_{xy} \leq 1$, the range of λar_{xy} will be $0 < \lambda ar_{xy} < 0.3069$.

We will have the following definition for fuzzy correlation interval.

Definition 3 Let $(X_i = [a_i, b_i, c_i, d_i], Y_i = [e_i, f_i, g_i, h_i]; i = 1, 2, \ldots, n)$ be a sequence of paired trapezoid fuzzy sample on population Ω with its pair of centroids (cx_i, cy_i) and pair of areas $\|x_i\| = area(x_i)$, $\|y_i\| = area(y_i)$. Let

$$cr_{xy} = \frac{\sum_{i=1}^{n}(x_i - \overline{cx})(cy_i - \overline{cy})}{\sqrt{\sum_{i=1}^{n}(cx_i - \overline{cx})^2}\sqrt{\sum_{i=1}^{n}(cy_i - \overline{cy})^2}};$$

$$ar_{xy} = \frac{\sum_{i=1}^{n}(\|x_i\| - \overline{\|x\|})(\|y_i\| - \overline{\|y\|})}{\sqrt{\sum_{i=1}^{n}(\|x_i\| - \overline{\|x\|})^2}\sqrt{\sum_{i=1}^{n}(\|y_i\| - \overline{\|y\|})^2}},$$

and

$$\lambda ar_{xy} = 1 - \frac{\ln(1 + |ar_{xy}|)}{|ar_{xy}|}.$$

Then fuzzy correlation is defined as follows:

(i) When $cr_{xy} \geq 0$ and $\lambda ar_{xy} \geq 0$, fuzzy correlation $= (cr_{xy}, \min(1, cr_{xy} + \lambda ar_{xy}))$.
(ii) When $cr_{xy} \geq 0$ and $\lambda ar_{xy} < 0$, fuzzy correlation $= (cr_{xy} - \lambda ar_{xy}, cr_{xy})$.
(iii) When $cr_{xy} < 0$ and $\lambda ar_{xy} \geq 0$, fuzzy correlation $= (cr_{xy}, cr_{xy} + \lambda ar_{xy})$.
(iv) When $cr_{xy} < 0$ and $\lambda ar_{xy} < 0$, fuzzy correlation $= (\max(-1, cr_{xy} - \lambda ar_{xy}), cr_{xy})$.

Example 2 Suppose we have the following data as shown in Table 2.

In this case, the correlation between the two centroids is $cr_{xy} = 0.17$. Similarly, the correlation between two lengths is $ar_{xy} = 0.32$, so

$$\lambda ar_{xy} = 1 - \frac{\ln(1 + 0.32)}{0.32} = 0.13.$$

Since the centroids correlation $ar_{xy} \geq 0$, and the area (length) correlation $\lambda ar_{xy} \geq 0$, thus, fuzzy correlation $= (ar_{xy}, \min(1, ar_{xy} + \lambda ar_{xy})) = (0.17, \min(1, 0.30)) = (0.17, 0.30)$. This implied that the relationship between the X and Y are quite small.

4 Empirical Studies

In this section, 11 samples (5 girls and 6 boys) are collected from a middle high school at Taipei city in Taiwan. We want to investigate which factors will impact their academic achievement. The results present the correlation for fuzzy data and in comparison with the traditional person correlation to demonstrate the difference. Suppose we are interesting in measuring the strength of the linear relationship between the students: sleeping hours per day (X), play hours on Internet per day (Y), studying hours in exercising mathematics per day (Z), and grades (range) of mathematical tests in last two months (T), as shown in Table 2.

The data set consists of interval-valued, triangular and trapezoidal fuzzy numbers. For example, for variable X, the data [8, 8.5, 9.5] represents a triangular fuzzy number, which represents that normal sleeping hours per day is 8.5 h, but the range of his/her sleeping hours is 8 to 9.5 h. Similarly, the data [9, 10.5, 11, 12] represented a trapezoidal fuzzy data, in this case, the normal sleeping hours is 10.5–11, and the range of sleeping time falls from 9 to 12 h (Table 3).

Table 3 Survey of fuzzy data

Sample	X	Y	Z	T
1	[8, 8.5, 9.5]	[1, 1.5]	[2, 2.5]	[90, 95]
2	[7, 7.5]	[1, 2, 3.5]	[2, 3.5, 4]	[92, 96]
3	[9, 10.5, 11, 12]	[1, 3]	[1, 2]	[85, 87]
4	[8, 8.5]	[1.5, 2.5]	[0, 0.5, 1]	[70, 72]
5	[6, 7.5]	[1, 1.5]	[2, 3]	[90, 97]
6	[10, 11, 13]	[1, 2, 4]	[0.5, 1]	[56, 63]
7	[7, 8]	[3, 3.5, 5]	[0, 1]	[35, 67]
8	[8, 10, 11]	[1, 2]	[1.5, 2]	[80, 85]
9	[6.5, 8]	[0, 1.5, 2, 2.5]	[2, 2.5, 3]	[92, 100]
10	[7.5, 8.5]	[2, 2.5, 4]	[0.5, 1]	[35, 55]
11	[8, 8.5]	[1, 2]	[1, 1.5]	[60, 67]
Fuzzy mean	8.50	2.02	1.58	75.9

Table 4 Correlation with fuzzy data

Fuzzy corr	X	Y	Z	T
X	1	$[-0.01, 0.09]$	$[-0.38, -0.27]$	$[-0.18, -0.17]$
Y		1	$[-0.59, -0.49]$	$[-0.73, -0.66]$
Z			1	$[0.87, 0.96]$
T				1

Table 5 Pearson correlations based on centroids

Pearson corr. (center)	X	Y	Z	T
X	1	-0.01	-0.38	-0.18
Y		1	-0.59	-0.73
Z			1	0.87
T				1

Based on Tables 4 and 5, we have the following findings. First, besides the correlation of studying hours in exercising mathematics per day (Z) and grades (range) of mathematical tests last two month (T) is positive, all of the other measures were negatively correlated.

Second, the correlation between X and Z is close to 0. This means there is almost no relationship between sleeping hours per day and studying hours in exercising mathematics per day. Third, the correlations between Y and Z and between Y and T are moderately negative. This means if the students spend more time on internet, then they will have less time study mathematics. In addition, the more time they spend on internet, the lower math grade will be. Fourth, the correlations between X and Z and between X and T are slightly negative. This means the relationship between student's sleeping hours and time study on mathematics are weakly related. The relationship between the sleeping hours and student math grades are also weakly related.

Table 4 is the fuzzy correlation, and the correlations in Table 4 are fuzzy numbers. This overcomes the deficiency of those studies which the correlation coefficients calculated are crisp values, rather than the intuitively believed fuzzy numbers. Table 5 is the Pearson correlation, which calculated based on the centroids of two dataset. It is found that the results of Tables 4 and 5 are quite close, the difference is the correlations in Table 4 are fuzzy numbers, and in Table 5 are crisp values. This is because the calculation of fuzzy correlation considered not only the correlations of centroids, but also the correlation between the area (length) of two dataset, and the fuzzy correlation expands based on the direction of the two dataset's area correlation. For example, the Pearson correlation between two centroids of Y and Z is -0.59. However, after considering the area (length) of two fuzzy dataset, the fuzzy correlation becomes $[-0.59, -0.49]$. This is due to the area(length) correlation of two dataset are positive, and this positive effect push the actual fuzzy correlation to the

positive side. On the other hand, the Pearson correlation between X and T is -0.18, and the fuzzy correlation is $[-0.18, -0.17]$. The range of this fuzzy correlation is quite narrow compare to the correlation between Y and Z. This is because the area correlation between two fuzzy dataset is quite small, thus the fuzzy correlation are mainly impacted by the correlations between two centroids.

5 Conclusions

This paper uses a simple way to derive fuzzy measures based on the classical definition of Pearson correlation coefficient which are easy and straightforward. Moreover, the range of the calculated fuzzy coefficient is a fuzzy number with domain $[-1, 1]$, which consist with the conventional range of Pearson correlation. In the formula we provided, when all observations are real numbers, the developed model becomes the classical Pearson correlation formula.

There are some suggestions for future studies. First, the main purpose of this study is to provide the formula of calculating fuzzy correlations. Only few samples are collected to illustrate how to employ the formula. Future interested researchers can use formula and collect a large-scale fuzzy questionnaires to make this formulas implement in practice. Second, when calculating the fuzzy correlation, we adopt λar_{xy} to adjust the correlations, but researchers can set up their own λar_{xy} values if there are defensible reasons. However, it is suggested that the impact of length correlation should not exceed the impact of centroid correlation. Third, this study only considered the fuzzy correlation for continuous data. It would be interested to investigate the fuzzy correlation for discrete fuzzy data.

In practice, many applications are fuzzy in nature. We can absolutely ignore the fuzziness and make the existing methodology for crisp values. However, this will make the researcher over confident with their results. With the methodology developed in this paper, a more realistic correlation is obtained, which provides the decision maker with more knowledge and confident to make better decisions.

References

1. Zadeh, L.A.: Fuzzy sets. Inf. Control **8**, 338–353 (1999)
2. Dubois, D., Prade, H.: Fuzzy sets in approximate reasoning, part 1: inference with possibility distributions. Fuzzy Sets Syst. **40**, 143–202 (1991)
3. Lowen, R.: A fuzzy language interpolation theorem. Fuzzy Sets Syst. **34**, 33–38 (1990)
4. Ruspini, E.: Approximate reasoning: past, present, future. Inf. Sci. **57**, 297–317 (1991)
5. Wu, B., Hsu, Y.: The use of Kernel set and sample memberships in the identification of nonlinear time series. Soft Comput. **8**(3), 207–216 (2002)
6. Bustince, H., Burillo, P.: Correlation of interval-valued intuitionistic fuzzy sets. Fuzzy Sets Syst. **74**, 237–244 (1995)
7. Hong, D.: Fuzzy measures for a correlation coefficient of fuzzy numbers under T_w—(the weakest t-norm)-based fuzzy arithmetic operations. Fuzzy Sets Syst. **176**, 150–160 (2006)

8. Liu, S., Kao, C.: Fuzzy measures for correlation coefficient of fuzzy numbers. Fuzzy Sets Syst. **128**, 267–275 (2002)
9. Yu, C.: Correlation of fuzzy numbers. Fuzzy Sets Syst. **55**, 303–307 (1993)
10. Hong, D., Hwang, S.: Correlation of intuitionistic fuzzy sets in probability space. Fuzzy Sets Syst. **75**, 77–81 (1995)
11. Wang, G., Li, X.: Correlation and information energy of interval-valued fuzzy numbers. Fuzzy Sets Syst. **103**, 169–175 (1999)
12. Chiang, D.A., Lin, N.P.: Correlation of fuzzy sets. Fuzzy Sets Syst. **102**, 221–226 (1999)

Empirical Evidence Linking Futures Price Movements of Biofuel Crops and Conventional Energy Fuel

Jianxu Liu, Songsak Sriboonchitta, Roland-Holst David, Zilberman David and Aree Wiboonpongse

Abstract This study proposes a dynamic vine copula based ARMAX-GARCH model to explore the dependence structures between energy futures and agricultural futures, and between corn future and soybean future conditional on energy futures etc. The more important thing is that we employ the empirical results of dynamic vine copulas to forecast the expected shortfall (ES) and the optimal portfolio weights (OPW) based on minimum ES and Monte Carlo simulation method results showed that the appropriate margins were skewed student t distribution for soybean future return, and student t distribution for crude oil, palm oil and corn future returns, and the time-varying copulas T copula, R-BB8(180°), R-BB8(180°), Gaussian copula, R-Joe(180°) and T copula can preferably capture the dependences compared with static copulas in C-vine copula structure. Moreover, we found that the values of ES will converge to −0.0121, −0.0145 and −0.0164 at period t+1 under 5, 2 and 1 % level, respectively. Meanwhile, As long as we invest in strict accordance with the optimal portfolio weights, the ES will reduce 56, 54 and 53 % at 5, 2 and 1 % level, respectively.

J. Liu · S. Sriboonchitta (✉)
Faculty of Economics, Chiang Mai University, Chiang Mai 50200, Thailand
e-mail: songsakecon@gmail.com

J. Liu
e-mail: liujianxu1984@163.com

R. David · Z. David
Department of Agricultural and Resource Economics,
University of California at Berkeley,
207 Giannini Hall, Berkeley, CA 94720-3310, USA

A. Wiboonpongse
Faculty of Agriculture, Department of Agricultural Economics
and Agricultural Extension, Chiang Mai University,
Chiang Mai 50200, Thailand

A. Wiboonpongse
Faculty of Economics, Prince of Songkla University, Songkla, Thailand

V.-N. Huynh et al. (eds.), *Econometrics of Risk*, Studies in Computational Intelligence 583, DOI 10.1007/978-3-319-13449-9_20

1 Introduction

With the rapid development of world economic integration, there are the increasingly close ties between the various commodities. Crude oil and other energy commodities also manifest increasing the influence to non-energy commodities, and the spillover effect is more pronounced compared to the last century, especially after the financial crisis in 2008. Crude oil futures prices were more than $147 per barrel on July 11, 2008, and after 3 months (2008-09-10) the same forward contracts had fallen back below the $100 mark and stayed there until the following January. At the same time, the prices of biofuel feedstocks such palm oil, corn, and soybeans all displayed strong volatility. Kilian [1] showed that oil price shocks impacted the U.S. and global economy significantly. Du et al. [2] have shown empirically that volatility spillovers among crude oil, corn, and wheat markets occurred after the Fall of 2006, while Ji [3] demonstrated the existence of significant volatility spillovers between crude oil and non-energy commodity markets. Despite this evidence of price co-movement between agricultural commodities and conventional energy fuel, conditional and time-varying co-movement with nonlinear correlations have still not been formally estimated. In other words, we have observed apparent financial linkage between the agricultural and energy commodities, but we do not know how much correlation is there, nor how it changes over time? Thus for a given trend in energy future prices are known, we would want to know the direction, magnitude, and timing of impacts on biofuel feedstock futures.

A good literature already exists on volatility and dependence between energy and agriculture futures. Pindyck and Rotemberg [4] found that the prices of raw commodities have a persistent tendency to move together. Palaskas and Varangis [5] confirmed co-movement between all commodity pairs using cointergration tests. Babula and Somwaru [6] studied the dynamic impact of a shock in crude oil prices on agricultural chemical and fertilizer prices using Vector Auto-Regression (VAR) methods, and Noel [7] revealed linkage between crude oil prices and agricultural employment using Granger causality tests. Campiche et al. [8] investigated links between crude oil and agriculture commodities using cointegration and VEC methods, with results suggesting crude oil is more strongly correlated with soy than corn. Nazlioglu and Soytas [9] examined the short-run and long-run interdependence between world oil prices, exchange rates, and Turkish agricultural prices, including cotton, soybean, wheat, and maize. Guo et al. [10] confirmed interaction between crude oil and agricultural commodities prices and applied Granger causality tests. Malliaris and Urrutia [11] examined correlation between futures prices across agriculture commodities. Chang et al. [12] estimated a long memory volatility model for 16 different agricultural commodity future prices. Choudhry [13] applied six GARCH models to analyze the volatilities of agricultural future prices.

Although VAR, VEC, cointegration tests, and Grange causality models are efficient and feasible for analyzing volatility and price interaction, copula models have emerged rapidly for application to the same class of problems. In particular, copula based GARCH models have been used to investigate exchange rate dependence [14], conditional dependence of internet stock prices [15], the co-movement of oil prices

and exchange rates [16] and option prices [17]. Furthermore, vine copulas (more flexible and transparent than their multivariate counterparts) have been used to elucidate the conditional interdependence of asset returns. Various studies have demonstrated their properties, classifications, structures and merits [18–23].

This paper complements the above literature by applying vine copula based GARCH model to investigate the direct linkages among conventional energy fuels (proxied by crude oil) and biofuel feedstock crops, i.e. palm oil, corn and soybean futures. The purposes of this paper are to examine the volatility of crude oil and palm oil future prices, and agricultural commodities future prices that include corn and soybean, and study their co-movements, especially to shed new light on what is the co-movement of conditional on crude oil and palm oil between corn and soybean. In addition, our analysis takes fully account of the time-varying characteristics nonlinear correlations (Kendall's tau parameter). This approach enables us to estimate expected shortfalls and identify risk characteristics. Finally, and just as importantly, we can estimate optimal invested portfolio weights based on the best copula model for next period. The main contributions of the paper are three: (1) We apply the vine copula based GARCH model to analyze the conditional dependences, and consider the time-varying Kendall's tau of conditional dependences. (2) We show how this framework can be used to estimate expected shortfalls and optimal portfolio weights using the results of copulas and Monte Carlo simulation method. (3) The optimal portfolio weights of selected assets are constructed under the minimum expected shortfall framework, allowing for global optimization via a Differential Evolution algorithm.

The remainder of this paper is organized as follows. Section 2 introduces the basic estimation framework, including static and time-varying copula models. Section 3 presents vine copulas. Section 4 presents the data and discusses our empirical results. Section 5 presents our ES and OPW estimation results. Section 6 offers conclusions.

2 Copula Based ARMAX-GARCH Models

The whole idea of a copula is that it provides information beyond the marginals. In the present application we use a ARMAX-GARCH model as the marginal distribution with respect to the copula for multiple asset returns. This approach allows, for example, more general symmetry characteristics for joint distributions, as well as more detailed and decomposable treatment of covariability and dependence structures. In the sub-section, the ARMAX-GARCH model, copulas and time-varying copulas are considered sequentially.

2.1 ARMAX-GARCH Model

The essential characteristic of copula model is that any multivariate distribution function can be decomposed into marginal distributions. Therefore, the ARMAX-GARCH

model is used to describe the marginal distribution of each series. Following Lee and Lin [24], the ARMAX-GARCH model can be expressed as:

$$r_t = c + \sum_{i=1}^{p} \phi_i r_{t-i} + \sum_{i=1}^{q} \psi_i \varepsilon_{t-i} + \varphi_i X_{it} + \varepsilon_t \tag{1}$$

$$\varepsilon_t = h_t \cdot \eta_t \tag{2}$$

$$h_t^2 = \omega + \sum_{i=1}^{k} \alpha_i \varepsilon_{t-i}^2 + \sum_{i=1}^{l} \beta_i h_{t-i}^2 \tag{3}$$

where $\sum_{i=1}^{p} \phi_i < 1$, $\omega > 0$, $\alpha_i \geq 0$, $\beta_i \geq 0$, and $\sum_{i=1}^{k} \alpha_i + \sum_{i=1}^{l} \beta_i < 1$. η_t is the standardized residual, which can be assumed for any distribution. Normally, we assumed that it is Gaussian, student t or skewed-t distribution. In particular, skewed-t distribution can capture characteristics of heavy tail and asymmetry, which was proposed by Hansen [25] whose model has two parameters λ and υ are the asymmetry and kurtosis parameters, respectively, and symmetric heavy tails can be captured by student-t distribution.

2.2 Copulas

After an early example by Sklar [26], copula methods have recently enjoyed rapidly growing interest in econometrics, economics, finance. Let $x = (x_1, x_2, \ldots, x_n)$ be a random vector with joint distribution function H and marginal distribution $F_1, F_2, \ldots, F_n$, then there exists a function C that is called copula:

$$F(x_1, x_2, \ldots, x_n) = C(F_1(x_1), F_2(x_2), \ldots, F_n(x_n)) \tag{4}$$

In the light of formula (4), the copula function can be expressed as:

$$C(u_1, u_2, \ldots, u_n) = F(F_1^{-1}(u_1), F_2^{-1}(u_2), \ldots, F_n^{-1}(u_n)) \tag{5}$$

If F_i is an absolutely continuous with strictly increasing, we have the density function as

$$\begin{aligned} f(x_1, \ldots\ldots, x_n) &= \frac{\partial F(x_1, \ldots\ldots, x_n)}{\partial x_1, \ldots, \partial x_n} \\ &= \frac{\partial C(u_1, \ldots, u_n)}{\partial u_1, \ldots, \partial u_n} \times \prod \frac{\partial F(x_i)}{\partial x_i} \\ &= c(u_1, \ldots, u_n) \times \prod f_i(x_i) \end{aligned}$$

The joint distribution F contains all statistical information about $x = (x_1, x_2, \ldots, x_n)$. In particular, marginal distributions of the components are derived as

$$F_i(x_i) = F(\infty, \infty, \ldots, x_i, \infty, \ldots, \infty) \tag{6}$$

Patton [14] extended (unconditional) copulas to conditional copulas, and applied them to time-varying conditional dependence for the analysis of exchange rates.

For concreteness, consider an example where $r_{o,t}$, $r_{p,t}$, $r_{c,t}$, $r_{s,t}$ represent asset returns to biofuel feedstock futures, e.g. crude oil, palm oil, corn and soybean forward contracts, with marginal conditional CDF $u_{o,t} = F\left(r_{o,t}|\psi_{t-1}\right) = F\left(\eta_{o,t}\right)$, $v_{p,t} = F\left(r_{p,t}|\psi_{t-1}\right) = F\left(\eta_{p,t}\right)$, $w_{c,t} = F\left(r_{c,t}|\psi_{t-1}\right) = F\left(\eta_{c,t}\right)$ and $z_{s,t} = F\left(r_{s,t}|\psi_{t-1}\right) = F\left(\eta_{s,t}\right)$, where ψ_{t-1} represent historical information. Thus, for a sample interaction between crude oil and palm oil, the corresponding bivariate conditional copula can be written as

$$F\left(r_{o,t}, r_{p,t}|\psi_{t-1}\right) = C\left(u_{o,t}, v_{p,t}|\psi_{t-1}\right) \tag{7}$$

It should be noted that the probability distributions will not be uniform over [0, 1] if we derive the marginal distribution using a misspecified model. In this paper, we experimented with Gaussian copula, T copula, Clayton copula, Frank copula, Gumbel copula, Joe copula, BB1 copula, BB6 copula, BB7 copula, BB8 copula and rotate copulas to elucidate the energy price the dependence structures. The Gaussian copula can reflect positive and negative correlation, and the Pearson correlation ρ can be transformed to Kendall's tau, which equals 2/arcsin(ρ). Clayton copula has the capacity to capture lower tail dependence. Frank copula can describe both positive and negative dependence. Gumbel copula is an asymmetric copula of the Archimedean family, allowing for upper tail dependence. Joe copula also helps explain upper tail dependence. BBX copulas are two parameter copulas, and BB8 can capture the upper tail dependence. But both BB1 and BB7 copulas can capture upper tail and lower tail dependence.

2.3 Time-Varying Copulas

Patton [14] observed that it is very difficult to identify the determinants of copula model parameters, and assumed they were generated in Gaussian and SJC copula by an ARMA (1,10) process. Manner and Reznikova [27], Wu [16], Liu and Sriboonchitta [28] and Ng [29] extended this approach without a definitive specification. Our study relies on this precedence, but also extends it with new time-varying, two parameter copulas described below:

(1) Time-varying Gaussian copula

$$\rho_t = \tilde{\Lambda}(\omega_N + \beta_{N1} \cdot \rho_{t-1} + \cdots + \beta_{Np} \cdot \rho_{t-p} + \alpha_N \cdot \frac{1}{q} \sum_{j=1}^{q} \Phi^{-1}(u_{1t-j}) \cdot \Phi^{-1}(u_{2t-j})) \tag{8}$$

where Φ represents standard normal distribution, $\tilde{\Lambda}$ is a logistic transformation which is defined as follows: $\Lambda\tilde{(x)} = (1-e^{-x})(1+e^{-x})^{-1}$. The purpose of using this logistic transformation is to keep the correlation coefficient ρ belonging to $(-1, 1)$.

(2) Time-varying T copula

$$\rho_t = \tilde{\Lambda}(\omega_T + \beta_{T1} \cdot \rho_{t-1} + \cdots + \beta_{Tp} \cdot \rho_{t-p} + \alpha_T \cdot \frac{1}{q} \sum_{j=1}^{q} T_v^{-1}(u_{1t-j}; v) \cdot T_v^{-1}(u_{2t-j}; v)) \tag{9}$$

where T_v represents student-t distribution with degree of freedom v, T copula has two parameters that are Pearson correlation ρ and degree of freedom v_c. Obviously, assume that fixed the degree of freedom, just let the correlation be change with time.

(3) Time-varying (rotate) Gumbel copula

$$\tau_t = \Lambda(\omega_G + \beta_{G1} \cdot \tau_{t-1} + \cdots + \beta_{Gp} \cdot \tau_{t-p} + \alpha_G \cdot \frac{1}{q} \sum_{j=1}^{q} \left|u_{1t-j} - u_{2t-j}\right|) \tag{10}$$

where $\Lambda(x) = (1 + e^{-x})^{-1}$. This guarantees that the Kendall's tau will be between -1 and 1, and the time varying Joe copula employ the same form as it.

(4) Time-varying (rotate) Clayton copula

$$\tau_t = \Lambda(\omega_C + \beta_{C1} \cdot \tau_{t-1} + \cdots + \beta_{Cp} \cdot \tau_{t-p} + \alpha_{C1} \cdot |u_{1t-1} - u_{2t-1}| + \cdots + \alpha_{Cq} \cdot \left|u_{1t-q} - u_{2t-q}\right|) \tag{11}$$

(5) Time-varying BBX copula

BB1, BB7 and BB8 are two parameter copula specifications, we assume that the parameters in each copula vary over, following an ARMA (p, q) type process. Specially, our case is an ARMA (1, 20) type process, which performs better than the original ARMA (1, 10) proposed by Patton [14].

$$\theta_t = \mathrm{H_{BBX}} \left(\omega_{BBX} + \beta_{BBX} \cdot \theta_{t-1} + \cdots + \beta_{BBXp} \cdot \theta_{t-p} + \alpha_{BBX} \cdot \frac{1}{q} \sum_{j=1}^{q} \left|u_{1t-j} - u_{2t-j}\right| \right) \tag{12}$$

$$\delta_t = \tilde{\mathrm{H}}_{\mathrm{BBX}} \left(\omega_{BBX} + \beta_{BBX} \cdot \delta_{t-1} + \cdots + \beta_{BBXp} \cdot \delta_{t-p} + \alpha_{BBX} \cdot \frac{1}{q} \sum_{j=1}^{q} \left| u_{1t-j} - u_{2t-j} \right| \right) \tag{13}$$

where the $H_{BBX}(x)$ and $\tilde{H}_{BBX}(x)$ are the logistic transformations, $H_{BB1}(x) = 1/e^{-x}$, $\tilde{H}_{BB1}(x) = 1/e^{-x} + 1$, $H_{BB1}(x) = \tilde{H}_{BB7}(x)$, $\tilde{H}_{BB1}(x) = H_{BB7}(x) = H_{BB8}(x)$, and $\tilde{H}_{BB8}(x) = \Lambda(x)$.

3 Vine Copulas

A bivariate copula specification is called a pair-copula construction or a vine copula. Standard multivariate copulas do not allow for different dependency structures between pairs of variables. While vine approach is more flexible, as we can select bivariate copulas from a wide range of (parametric) families. Vine copulas involve marginal conditional distributions that can be expressed by the form $F(r_t|\upsilon)$. Joe [30] showed that

$$F(r|\upsilon) = \frac{\partial C_{r,\upsilon_j|\upsilon_{-j}}(F(r|\upsilon_{-j}), F(\upsilon_j \,|\upsilon_{-j}))}{\partial F(\upsilon_j|\upsilon_{-j})} \tag{14}$$

where υ denotes all conditional variables. If the υ is univariate, the marginal conditional distribution is a special case that can be written as

$$F(r_1|r_2) = \frac{\partial C_{r_1,r_2}(F(r_1), F(r_2))}{\partial F(r_2)} \tag{15}$$

Bedford and Cooke [20, 31] gave the definition of the regular vine copula. The class of regular vines comprises a large number of pair-copula decompositions. In this section, we focus on the two special cases of regular vine copulas, the so-called Canonical vine and Drawable vine, which we present in a four-dimensional framework.

(1) C-vine copula. The four-dimensional C-vine structure can be written as:

$$\begin{aligned} f(x_1, x_2, x_3, x_4) = {} & f(x_1) \cdot f(x_2) \cdot f(x_3) \cdot f(x_4) \cdot c_{12}(F(x_1), F(x_2)) \\ & \cdot c_{13}(F(x_1), F(x_3)) \cdot c_{14}(F(x_1), F(x_4)) \\ & \cdot c_{23|1}(F(x_{2|1}), F(x_{3|1})) \cdot c_{24|1}(F(x_{2|1}), \\ & F(x_{4|1})) \cdot c_{34|12}(F(x_{3|12}), F(x_{4|12})) \end{aligned} \tag{16}$$

where lowercase c denotes the density function of copulas. $F(x_{2|1})$, $F(x_{3|1})$ and $F(x_{4|1})$ can be derived from expression 4 above. The function $c_{23|1}$ represents the joint density function (copula) of x_2 and x_3, conditional on x_1, and so on. According to the formula (14), the marginal distributions of $x_{3|12}$ and $x_{4|12}$ are

$$F(x_{3|12}) = \frac{\partial C_{23|1}(F(x_{3|1}), F(x_{2|1})}{\partial F(x_{2|1})} \text{ and } F(x_{4|12}) = \frac{\partial C_{24|1}(F(x_{4|1}), F(x_{2|1})}{\partial F(x_{2|1})} \tag{17}$$

(2) D-vine copula. The four-dimensional D-vine structure can be shown as:

$$\begin{aligned} f(x_1, x_2, x_3, x_4) = & f(x_1) \cdot f(x_2) \cdot f(x_3) \cdot f(x_4) \cdot c_{12}(F(x_1), F(x_2)) \\ & \cdot c_{23}(F(x_2), F(x_3)) \cdot c_{34}(F(x_3), F(x_4)) \\ & \cdot c_{13|2}(F(x_{1|2}), F(x_{3|2})) \cdot c_{24|3}(F(x_{2|3}), F(x_{4|3}) \\ & \cdot c_{14|23}(F(x_{1|23}), F(x_{4|23})) \end{aligned} \tag{18}$$

Comparing the structure of D-vine with C-vine, we see that the C-vine has a central variable x_1, but the structure of D-vine is ordered by the sequence of the variables. The small letter x_1, x_2, x_3 and x_4 is here used to denote returns of crude oil, palm oil, corn and soybean, respectively.

4 The Data and Empirical Results

4.1 The Data

This study uses crude oil and US dollar index futures from ICE market, corn and soybean futures from CBOT market and palm oil from MDEX market to analyze the issues that we have proposed in introduction part. Our sample covers the period January 2, 2008 to May 3, 2012, and, to eliminate spurious correlation arising from holidays, we drop those observations for any holiday associated with least one index. The asset returns are calculated by using the difference between logarithmic closing prices for each contract type.

Table 1 provides descriptive statistics for this data. It shows that crude oil, palm oil, corn and the dollar index all exhibit slight negative skewness and excess kurtosis, while soybean contracts show strong negative skewness and excess kurtosis. In addition, these results show that corn futures has the maximum return, soybean futures the minimum return, during the period considered. Standard deviations show clear evidence that crude oil more price volatile than the other commodities. The Jarque-Bera test results show that the distribution of asset returns strongly rejects an assumption of normality, while the unconditional correlation coefficients high degree of linear dependence between future corn and soybean returns and negative dependency between dollar return and the other asset returns.

Table 1 Data descriptive and statistics

	Crude oil	Palm oil	Corn	Soybean	Dollar
Mean	0.0001	0	0.0001	0.0001	0
Median	0.0003	0	0	0.0007	−0.0001
Maximum	0.0551	0.0424	0.0554	0.0356	0.0119
Minimum	−0.0475	−0.0479	−0.0452	−0.0891	−0.0132
Std. dev	0.0113	0.0090	0.0104	0.0098	0.00281
Skewness	−0.2717	−0.3165	0.0514	−1.334	0.0054
Kurtosis	5.9813	6.6814	4.9561	11.5958	4.4304
Jarque-Bera	393.3751	597.6917	164.3429	3469.7810	87.6443
Correlations					
Crude oil	1	0.2683	0.3774	0.3284	−0.3910
Palm oil	0.2683	1	0.2114	0.2057	−0.1731
Corn	0.3774	0.2114	1	0.5097	−0.3122
Soybean	0.3284	0.2057	0.5097	1	−0.2387
Dollar	−0.3910	−0.1731	−0.3122	−0.2387	1

4.2 The Results of ARMAX-GARCH Model

Table 2 shows the estimated results for the ARMAX-GARCH model. To insure the marginal distributions are correctly specified, we assume that crude oil, palm oil and corn are all t-distribution because of non-normality, the slight skewness and kurtosis, but on the contrary, soybean is assumed to follow a skewed student-t distribution. The sum of ARCH and GARCH terms can be interpreted to measure the persistence of volatility. Each value of this sum is greater than 0.97, which means unforeseen shock will enhance volatility for a long time or, alternatively conditional variance convergence to long-term variance can be thought to take longer. Moreover, the ARCH LM tests are calculated for testing whether squared residuals remain serially correlated up to lags 10 and 20. In copula functions, we assume that the marginal distribution must be iid uniformly on (0, 1). The tests are performed by Box-Liung and Kolmogorov-Smirnov (KS) test, and the test results are reported in Table 3. The Box-Liung test evaluates whether u, v, w and z are serially correlated, i.e. we examine the serial correlation for first four moments of each asset return. The p-value from Box-Liung and KS test reported in Table 3 generally do not reject null hypothesizes, meaning that all series satisfy iid-uniform (0, 1) condition.

4.3 Results for the Static and Time-Varying C-Vine Copula

Tables 4 and 5 report parameter estimates for the appropriate C-vine copula from a set of static and time-varying possible copula families according to the AIC, respectively.

Table 2 Fitted asset returns from a ARMAX-GARCH model

Parameters	Crude oil	Palm oil	Corn	Soybean
c	0.0005***	0.0002	0.0003	0.0001
	(0.0002)	(0.0002)	(0.0003)	(0.0002)
φ	−1.4224***	−0.3791***	−0.9541***	0.6742***
	(0.1001)	(0.0748)	(0.1011)	(0.0856)
ω	7.75E-07*	3.42E-07	5.22E-06	(6.19E-07)
	(4.09E-07)	(2.39E-07)	(3.75E-06)	(6.19E-07)
α	0.0553***	0.0645***	0.0833***	0.0391***
	(0.0108)	(0.0160)	(0.0368)	(0.0129)
β	0.9363***	0.9307***	0.8681***	0.9500***
	(0.0114)	(0.0164)	(0.0665)	(0.0165)
ν	7.0032***	8.7328***	5.894***	5.1281***
	(1.4964)	(2.0847)	(0.9893)	(0.6685)
λ				−0.1492***
				(0.0405)
logL	3419	3599	3335	3453
Long-term var	9.23E-05	7.13E-05	0.0001	9.36E-05
AIC	−6827	−7185	−6659	−6892
BIC	−6797	−7156	−6629	−6857
LM(10)	0.3164	0.7307	0.9263	0.9876
LM(20)	0.7420	0.9038	0.9792	0.9924

Note Signif. codes are as follows: 0 ***0.001 **0.01 * 0.05. The numbers in the parentheses are the standard deviations

Note that this copula selection proceeds tree by tree, since the conditional pairs in trees 2 and 3 depend upon the specification of the previous trees through formula 3 in Sect. 4. We implemented the most common single parameter copula families, such as the Gaussian, Clayton, Gumbel, Frank and Joe, and five copula families with two parameters, namely, student-t, BB1, BB6, BB7 and BB8. Moreover, the rotated copulas such as Clayton, Gumbel, Joe, BB1, BB6, BB7 and BB8 are put to use as well. Furthermore, the variable order was chosen to correspond to crude oil, palm oil, corn and soybean future returns. According to the AIC, the optimal choices are T copula, R-BB8 (180°) copula and R-BB8 (180°) copula for tree 1, Gaussian copula, R-Joe (180°) copula for tree 2 and T copula for tree 3. We can see that the dependency parameter of the bivariate T copula between crude oil and palm oil has a low value of 0.22, Kendall's tau equals to 0.14, and estimated tail dependence is very low (0.0053) as well, while the dependency parameter of T copula between corn and soybean conditional on crude oil and palm oil has the largest correlation 0.47, tail dependence equals 0.18, and Kendall'tau is 0.31. Comparing values of AIC and BIC in Table 4 with Table 5, all the time-varying copula dependence structures exhibit better explanatory ability than static copulas. The value of the β parameter in $C_{23|1}$ is close to unity, implying the time-varying dependence between palm oil and corn

Table 3 Serial correlation tests

Box-Liung test					
Crude oil	X-squared	p-value	Palm oil	X-squared	p-value
First moment	4.8287	0.9023	First moment	16.794	0.07905
Second moment	2.0944	0.9956	Second moment	10.3769	0.4081
Third moment	4.4029	0.9273	Third moment	17.3687	0.06659
Fourth moment	3.7289	0.9588	Fourth moment	10.3148	0.4133
Corn	X-squared	p-value	Soybean	X-squared	p-value
First moment	14.5332	0.15	First moment	6.2957	0.7898
Second moment	5.9983	0.8154	Second moment	9.2183	0.5115
Third moment	8.7697	0.5541	Third moment	4.4453	0.925
Fourth moment	3.966	0.9489	Fourth moment	8.5842	0.572
KS test					
	Statistics	p-value		Statistics	p-value
Crude oil	0.001	1	Corn	0.001	1
Palm oil	0.001	1	Soybean	0.001	1

Table 4 The results of C-vine copula and Kendall'tau

	Parameters	Tail dependence	Kendall'tau	AIC	BIC
T copula (C_{12})	0.2200***	0.0053	0.1412	−48.8391	−38.9703
	(0.0310)				
	15.1672	0.0053			
	(8.3718)				
R-BB8 (180) (C_{13})	1.8180***	0	0.2055	−112.0967	−102.2279
	(0.2426)				
	0.8746***	0			
	(0.0700)				
R-BB8 (180) (C_{14})	1.4542***	0	0.1637	−82.3986	−72.5298
	(0.1203)				
	0.9559***	0			
	(0.0374)				
Gaussian copula ($C_{23\|1}$)	0.0975***	0	0.0622	−7.6219	−2.6875
	(0.0310)				
R-Joe (180) ($C_{24\|1}$)	1.0827***	0	0.04541	−12.3508	−7.4164
	(0.0287)				
T copula ($C_{34\|12}$)	0.4730***	0.1832	0.3136	−284.0692	−274.2004
	(0.0262)				
	5.2572***	0.1832			
	(1.0628)				

Note Signif. codes are as follows: 0 ***0.001 **0.01 *0.05. The numbers in the parentheses are the standard deviations

Table 5 The results of time-varying C-vine copula

Copulas	W	β	α	AIC	BIC
T copula (C_{12})	0.7085***	−0.1381***	−1.45189***	−49.5000	−42.5669
	(0.0102)	(0.0100)	(0.0456)		
R-BB8 (180) (C_{13})	−2.1098***	9.9297***	2.5683***	−122.1628	−117.2284
	(0.0215)	(0.0919)	(0.0116)		
	1.5112***	−3.149***	0.1486***		
	(0.0329)	(0.0677)	(0.0081)		
R-BB8 (180) (C_{14})	4.0864***	−1.284***	0.2346***		
	(0.0513)	(0.2725)	(0.0451)		
	−0.284***	0.2841***	0.6007***		
	(0.0037)	(0.0130)	(0.0042)		
Gaussian copula ($C_{23\|1}$)	0.2974***	0.8730***	−2.0388***	−14.2621	−8.3277
	(0.0031)	(0.0110)	(0.0002)		
R-Joe (180) ($C_{24\|1}$)	1.4806***	−0.2543***	−1.4272***	−209.5979	−203.6635
	(0.0161)	(0.0077)	(0.0297)		
T copula ($C_{34\|12}$)	1.1411***	0.3789***	−1.6895***	−286.9409	−275.0721
	(0.0051)	(0.0044)	(0.0131)		

Note Signif. codes are as follows: 0 ***0.001 **0.01 *0.05. The numbers in the parentheses are the standard deviations

returns persists through time under the condition of crude oil, while the autoregressive parameter β in $C_{24|1}$ and $C_{34|12}$ shows slight autocorrelation. Figures 1, 2, and 3 show the time-varying dependence structures in $C_{23|1}$, $C_{24|1}$ and $C_{34|12}$, respectively. The lines of dashes are the average values of time-varying dependence. Although $C_{23|1}$ and $C_{34|12}$ are Gaussian and student-t copulas that can capture linear dependence, we transform them into nonlinear correlations Kendall's tau as described in Figs. 2 and 3. Meanwhile, the parameter in $C_{24|1}$ is transformed into Kendall's tau. The fluctuation range for $C_{23|1}$, $C_{24|1}$ and $C_{34|12}$ is from −0.18 to 0.31, from 0.17 to 0.31 and from 0.25 to 0.6, respectively.

Note: the best bivariate copula between corn and soybean is student-t copula as well. The parameter estimates of T copula equal 0.5194 for correlation, 5.4592 for the degree of freedom, 0.1995 for the tail dependence and 0.3477 for the Kendall's tau. Therefore, this is significant evidence that crude oil and palm oil have affected the dependence structure between corn and soybean. Compare C_{34} with $C_{34|12}$, the dependence fall by 8.93 % and the Kendall's tau fall by 9.81 %.

5 Forecasting of the ES and Optimal Portfolio

In last section, we examined the volatility and dependence structures between futures returns on energy and biofuel feedstocks, and between corn and soybean futures returns conditional on crude oil and palm oil futures returns, using a copula based

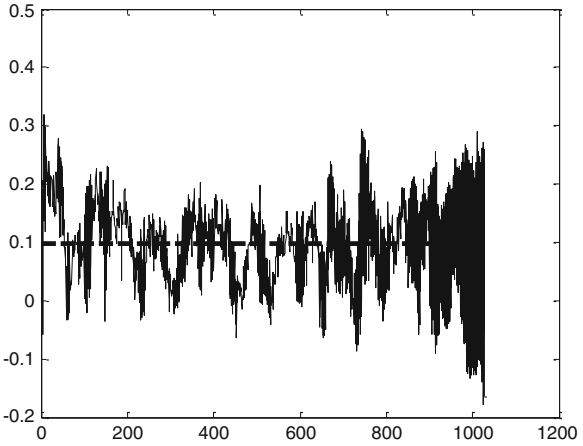


Fig. 1 Time-varying gaussian copula of 23 given 1

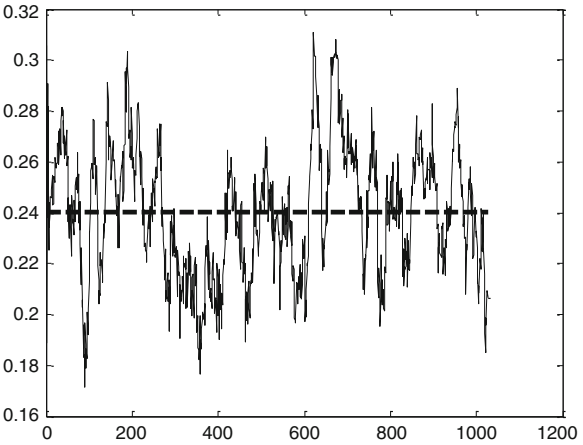


Fig. 2 Time-varying Joe copula of 24 given 1

ARMAX-GARCH model. However, we must take note of estimation results that may not have an economically useful application. Hence, in this section, we use Monte Carlo simulation for the copula based ARMAX-GARCH model to calculate the expected shortfall of an equally weighted portfolio. After that, optimal portfolio weights of selected assets are constructed under a minimum expected shortfall (ES) framework, using global optimization with the differential evolution algorithm.

First, we summarize the five steps for calculating ES and optimal portfolio weights:

(1) Use of the estimation results of bivariate copulas (C_{12}, C_{13} and C_{14}) to generate random number 1027, the length of our observations.

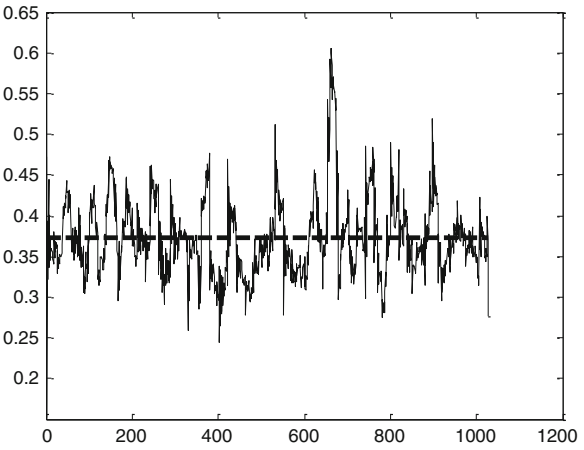


Fig. 3 Time-varying T copula of 34 given 12

(2) Use inverse functions of the Student's t (for crude oil, palm oil and corn) and skewed Student's-t distributions (for soybean) to get the standardize residuals η_{oil}, η_{palm}, η_{corn} and η_{soy} of each variable.
(3) Forecast the values of $r_{oil,t+1}$, $r_{palm,t+1}$, $r_{corn,t+1}$ and $r_{soy,t+1}$ for subsequent periods using the ARMAX-GARCH model, giving equal weight (0.25) to each, i.e., $r = 0.25 * r_{oil,t+1} + 0.25 * r_{palm,t+1} + 0.25 * r_{corn,t+1} + 0.25 * r_{soy,t+1}$.
(4) Calculate the quantiles of r at 5, 2 and 1 % level, respectively. The "expected shortfall at q % level" is the expected return on the portfolio in the worst q % of the cases.

Repeat steps (1)–(4) 1,000, 2,000 and 5,000 times, and calculate the average ES. Generally, each mean of ES should be close to its true value.

After this procedure, we calculated optimal portfolio weights using the results of copula based ARMAX-GARCH model with the following strategy:

(1) Repeat the first two steps above, calculating ES for optimal portfolio weights.
(2) Now consider an investor who wants to minimize ES at 50, 25, 5, 2 and 1 % subject to achieving a particular expected return. Let w_i be weight vector of portfolio weights on risky assets, namely, crude oil, palm oil, corn and soybean future returns. The investor solves the following optimization problem:
Min $ES = E[r|r \leq r_\alpha]$
Subject to
$r = w_1 \times r_{\text{oil},t+1} + w_2 \times r_{\text{palm},t+1} + w_3 \times r_{\text{corn},t+1} + w_4 \times r_{\text{soy},t+1}$
$w_1 + w_2 + w_3 + w_4 = 1$
$0 \leq w_i \leq 1, i = 1, 2, 3, 4$

(3) Global optimization can then solve this problem with maximum iterations to be 25 and only 10 repetitions. Even at this small simulation scale, the estimated weights still converge.

Table 6 Expected shortfall of equally weighted portfolio

ES	5 %	2 %	1 %
1,000 times	−0.01212	−0.01454	−0.01640
2,000 times	−0.01208	−0.01452	−0.01641
5,000 times	−0.01208	−0.01454	−0.01639

Table 7 Optimal portfolio weights based on minimum ES with MC simulation given copulas

	w_1	w_2	w_3	w_4	O.P.R
50 %	0.1677	0.5224	0.2053	0.1045	−0.0021
25 %	0.1775	0.5410	0.2048	0.0767	−0.0031
5 %	0.1935	0.4646	0.1639	0.1780	−0.0053
2 %	0.2172	0.3984	0.1773	0.2071	−0.0067
1 %	0.2328	0.4491	0.2041	0.1139	−0.0077
C.E	0.1842	0.4982	0.2346	0.0830	−0.0052

Table 6 presents the ES at levels of 5, 2 and 1 %. We can see that the estimated ES will converge to −0.0121, −0.0145 and −0.0164 at period t+1 under 5, 2 and 1 % level, respectively. Table 7 reports the optimal portfolio weighting estimates results at period t+1. We can find that the optimal weight of crude oil futures rises with gradually increasing risk, meaning crude oil future has high risk and high return, significant evidence that crude oil futures have higher volatility than others. The weight of palm oil accounts for about 50 % at each risk level, suggesting this asset has lower investment risk. The above-mentioned conclusions are further substantiated by the standard deviation results in Table 1. As long as we invest in strict accordance with the optimal portfolio weights, the ES will mitigate risk by 56, 54 and 53 % at 5, 2 and 1 %, respectively. This shows clear evidence of the strategy's hedging potential.

6 Conclusions

This paper used a vine copula based ARMAX-GARCH model to model dependencies between energy futures and those of biofuel feedstock commodities, and between corn future and soybean futures, conditional on energy futures, etc. When returns are non-normal, it is often difficult to specify the multivariate distribution relating two or more return series, in spite of the bivariate Student's-t distribution imposes a symmetric dependence structure, but the assumed degrees of freedom of that is assumed are the same.

We provide the ARMAX-GARCH model to fit the marginal distributions. These perform quite well, capturing fundamental properties of skewness, kurtosis, heavy tail and volatility. In this context, we apply a static and time-varying vine copula to join these complicated univariate margins. Then, to enhance the practical application of this methodology, we calculated the expected shortfall and optimal portfolio allocation for this model. Empirical evidence reveals that all the dependencies between energy and agricultural future returns are time varying and, in particular, the time-varying dependence structure between palm oil and corn returns persists under the

condition of crude oil, while $C_{24|1}$ and $C_{34|12}$ show lesser persistence. Finally, the ES results for ES and optimal portfolio weights are very suggest a wide range of practical application to hedging and other investment strategies and dynamic asset allocation problems.

References

1. Kilian, L.: The economic effects of energy price shock. J. Econ. Lit. **46**, 871–909 (2008)
2. Du, X., Yu, C.L., Hayes, D.J.: Speculation and volatility spillover in the crude oil and agricultural commodity markets: a Bayesian analysis. Energy Econ. **33**, 497–503 (2011)
3. Ji, Q., Fan, Y.: How does oil price volatility affect non-energy commodity markets? Appl. Energy **89**, 273–280 (2012)
4. Pindyck, R.S., Rotemberg, J.J.: The excess co-movement of commodity prices. Econ. J. **100**(403), 1173–1189 (1990)
5. Palaskas, T.B., Varangis, P.N.: Is There Excess Co-movement of Primary Commodity Prices? A Cointegration Test. International Economics Department, The World Bank (1991)
6. Babula, R.A., Somwaru, A.: Dynamic impacts of a shock in crude oil price on agricultural chemical and fertilizer prices. Agribusiness **8**(3), 243–252 (1992)
7. Noel, D.U.: Crude oil price volatility an unemployment in the United States. Energy **21**(1), 29–38 (1996)
8. Campiche, J., Bryant, H., Richardson, J., Outlaw, J.: Examining the evolving correspondence between petroleum prices and agricultural commodity prices. American Agricultural Economics Association (New Name: Agricultural and Applied Economics Association). Available at http://ideas.repec.org/p/ags/aaea07/9881.html (2007)
9. Nazlioglu, S., Soytas, U.: World oil prices and agricultural commodity prices: evidence from an emerging market. Energy Econ.—ENERG ECON **33**(3), 488–496 (2011)
10. Guo, H., Li, F., Yang, W., Yang, Y.: Analysis of price trends of crude oil, agricultural commodities and policy choices of biofuels in developing countries. Afr. J. Bus. Manag. **6**(2), 538–547 (2012)
11. Malliaris, A.G., Urrutia, J.L.: Linkages between agricultural commodity futures contracts. J. Futures Mark. **16**(5), 595–609 (1996)
12. Chang, C.L., McAleer, M., Tansuchat, R.: Modelling Long Memory Volatility in Agricultural Commodity Futures Returns. Department of Economics and Finance, University of Canterbury, Christchurch (2012)
13. Choudhry, T.: Short-run deviations and time-varying hedge ratios: evidence from agricultural futures markets. Int. Rev. Financ. Anal. **18**(1–2), 58–65 (2009)
14. Patton, A.J.: Modelling asymmetric exchange rate dependence. Int. Econ. Rev. **47**(2), 527–556 (2006)
15. Jondeau, E., Rockinger, M.: The copula-GARCH model of conditional dependencies: an international stock market application. J. Int. Money Financ. **25**, 827–853 (2006)
16. Wu, C.C., Chung, H., Chang, Y.H.: The economic value of co-movement between oil price and exchange rate using copula-based GARCH models. Energy Econ. **34**(1), 270–282 (2012)
17. Zhang, J., Guégan, D.: Pricing bivariate option under GARCH processes with time-varying copula. Math. Econ. **42**, 1095–1103 (2008)
18. Nikoloulopoulos, A.K.: Vine copulas with asymmetric tail dependence and applications to financial return data. Comput. Stat. Data Anal. **56**, 3659–3673 (2012)
19. Kurowicka, D., Cooke, R.M.: Uncertainty Analysis with High Dimensional Dependence Modelling. Wiley, New York (2006)
20. Bedford, T., Cooke, R.M.: Vines—a new graphical model for dependent random variables. Ann. Stat. **30**, 1031–1068 (2002)

21. Joe, H., Li, H., Nikoloulopoulos, A.K.: Tail dependence functions and vine copulas. J. Multivar. Anal. **101**, 252–270 (2010)
22. Aas, K., Czado, C., Frigessi, A., Bakken, H.: Pair-copula constructions of multiple dependence. Insur. Math. Econ. **44**, 182–198 (2009)
23. Bedford, T., Cooke, R.M.: Monte Carlo simulation of vine dependent random variables for applications in uncertainty analysis. In: 2001 Proceedings of ESREL, Turin, Italy (2001).
24. Lee, W.C., Lin, H.N.: Portfolio value at risk with copula-ARMAX-GJR-GARCH model-evidence from the gold and silver futures. Afr. J. Bus. Manag. **5**(5), 1650–1662 (2011)
25. Hansen, B.: Autoregressive conditional density estimation. Int. Econ. Rev. **35**, 705–730 (1994)
26. Sklar, A.: Fonctions de Réparitition à n dimensions et Leurs Marges. Publ l'Inst Stat l'Univ Paris **8**, 229–231 (1959)
27. Manner, H., Reznikova, O.: A survey on time-varying copulas specification simulations and application. Econom. Rev. **31**(6), 654–687 (2012)
28. Liu, J., Sriboonchitta, S.: Analysis of Volatility and Dependence between the Tourist Arrivals from China to Thailand and Singapore: A Copula-based GARCH Approach. Uncertainty Analysis in Econometrics with Applications Advances in Intelligent Systems and Computing, pp. 283–294. Springer, Heidelberg (2012)
29. Ng, W.L.: Modeling duration clusters with dynamic copulas. Financ. Res. Lett. **5**, 96–103 (2008)
30. Joe, H.: Families of m-variate distributions with given margins and m(m−1)/2 bivariate dependence parameters. In: Rüschendorf L., Schweizer B., Taylor M.D. (ed.) Distributions with Fixed Marginals and Related Topics (1996)
31. Bedford, T., Cooke, R.M.: Probability density decomposition for conditionally dependent random variables modeled by vines. Ann. Math. Artif. Intell. **32**, 245–268 (2001b)

Optimal Portfolio Selection Using Maximum Entropy Estimation Accounting for the Firm Specific Characteristics

Xue Gong and Songsak Sriboonchitta

Abstract The estimated return and variance for the Markowitz mean-variance optimization have been demonstrated to be inaccurate; thereafter it could make the traditional mean-variance optimization inefficient. This paper applied the Maximum Entropy (ME) principle in portfolio selection while accounting for firm specific characteristics; they are the firm size, return on equity and also lagged 12 months return. Since these characteristics are found not only related to the stock's expected return, variance and correlation with other stocks, they can be good variables to estimate the weights. Furthermore, this method used Generalized Cross Entropy to shrink portfolio weights to the equal weights; therefore solving the problem of concentrated weights in Markowitz mean-variance framework. Also in our empirical study, six stocks are used to investigate the effect of maximum entropy based methods. The results show that the in-sample forecasts that are in comparison with other traditional methods are good, however, in the out-of-sample forecasts the results are mixed.

1 Introduction

The Markowitz's portfolio selection, which is based on mean and variance, is one of the most important models of normative investment behavior in modern finance [1, 2]. However, optimal portfolio selection requires knowledge of expected return and variance. As Merton [3] argued, rather than the variance and covariance becoming

X. Gong (✉) · S. Sriboonchitta
Faculty of Economics, Chiang Mai University, Chiang Mai 50200, Thailand
e-mail: gongxue.cmu@gmail.com

S. Sriboonchitta
e-mail: songsakecon@gmail.com

V.-N. Huynh et al. (eds.), *Econometrics of Risk*, Studies in Computational Intelligence 583, DOI 10.1007/978-3-319-13449-9_21

stable over time, the expected returns are not easy to detect. When we only approximately know that these values are using the historical information or parametric method, the portfolio optimized for not-exactly-correct value may lead to a poor performance [4].

Therefore, it leads to several critical problems of optimal mean-variance portfolio; first, the optimal weights are concentrated on some stocks due to the dominated mean and variance [2]. However, the principle of sharing risk is diversification. Second, it is well known that the out-of-sample forecasting results are not so good [5]. The drawbacks of traditional mean-variance method have raised up a lot of attentions for further studies from the scholars and practitioners [4, 6–8].

Frost and Savarino [6] and Jorion [7] used empirical Bayesian approaches to derive those shrinkage estimators of return and variance, to make the portfolio more diversified. Later, the resampling scheme was first proposed by [8]. This method used a bootstrap to find more "truth" on the data; it makes more to be diversified and intuitively less risky than one on a corresponding Markowitz efficient frontier. Bai et al. [4] developed new bootstrap-corrected estimations for the optimal return and its asset allocation. The study shows that these bootstrap-corrected estimates are proportionally consistent with their theoretic counterparts.

On the other hand, the stock characteristics, such as the firms' market capitalization, return on equity (ROE), and lagged return etc., are found not only related to the stock's expected return, variance and correlation with other stocks. Based on the results, it can explain optimal portfolio weights [9, 10]. The traditional way to incorporate the firm's characteristics into the function of mean, variance and covariance is very complicated and creates unstable results, since it needs to estimate a large number of parameters [8, 9]. Therefore, in practice, the traditional method based on firm characteristics is rarely applied, although it may largely improve the accuracy of the estimates.

To best utilize the information, and build a correct portfolio weights, in this paper we modified a maximum entropy method to the portfolio selection problem accounting for the firm characteristics. The previous work of Bera and Park [11] has applied the cross entropy method to the portfolio selection, and uses the results of out-of-sample forecasting to show that the new method outperform the other popular nonparametric methods. Our study parameterize the weight invested in each stock as a function of the firm's characteristics, due to the fact that these characteristics fully capture the information of the joint distribution of returns that are all relevant for forming optimal portfolios. We maximized the Generalized Cross Entropy (GCE) problem subject to certain constraints. This estimation improved the previous works since it adds more information when considering the characteristics of each firm, and shrinks to the equal weights and makes a diversified portfolio.

Moreover, we implemented this information-based method to several six firm stocks as a portfolio and we evaluated the performance of our entropy based method with firm characteristics and related method. The results show that the two cross entropy method outperform other traditional methods in terms of in-sample case,

and have mixed results when considering the out-of-sample forecasts. The objective of this study is to apply the maximum entropy principle to optimal portfolio selection among different stocks with firm characteristics variables.

The rest of the paper is organized as follows. In Sect. 2, we provide a critical review of the existing methodologies. In Sect. 3, we discuss portfolio selection procedures using the ME principle based on the CE measure methodologies. In Sect. 4, we provide an empirical application using six stocks portfolio. The conclusion of this paper is in Sect. 5.

2 Literature Review

2.1 The Problem of Portfolio Selection Weights

In Markowitz's [1] mean-variance (MV) optimization, the portfolio weights are gained from two steps: the first step is to obtain the first moment and second moment of each stocks, and then assume the estimates are true parameters, the mean variance portfolio selection problem is solved separately: we either maximize the expected return under given bounds on risk or, equivalently, we minimize risk under the constraint that the expected returns should be at least a given value.

This approach works well if we know the expected returns of each financial instrument and we know the correlation between these returns. In practice, the problem of these steps is that the samples estimates are treated as the accurate values, which is absolutely not true [3, 7, 12]. The expected returns are not easy to detect. And, as a result, the portfolio optimized for not-exactly-correct value may lead to a poor performance.

Therefore, to improve traditional method and make the portfolio more robust, we introduce the maximum entropy-based method.

2.2 The Firm Characteristics Influence on Optimal Weights

There is several literatures investigating the relationship between the optimal portfolio weights and a set of firm characteristics variables [9, 13]. The channel of the firm characteristics impact on the portfolio weights is the following: the characteristics are important factors of the expected returns, variances, covariances, and even higher order moments of returns, and then these factors all affect the distribution of the optimized portfolio's returns and finally the investor's utility. As Brandt et al. [9] stated, the deviations of the optimal weights from the benchmark weights decrease with the firms' size, increases with its book-to-market ratio, and increases with firms' lagged

12-month returns. Nigmatullin [14] used a nonparametric method to incorporate and model the weights in each asset class. But he adopted different variables which are specific to the asset classes as a separate function.

3 Methodology

3.1 Introduction to Maximum Entropy Method

As Jaynes [15] suggest that 'a certain probability distribution maximizes entropy subject to certain constraints representing a piece of incomplete information, is the fundamental property which justifies use of that distribution for inference; it agrees with everything that is known, but carefully avoids assuming anything that is not known' (p. 1). When we want to approximate this unknown probability distribution, what should be the best approximation? Jaynes [16] gave a general answer to this question:'the best approach is to ensure that the approximation satisfies any constraints on the unknown distribution that we are aware of, and that subject to those constraints, the distribution should have maximum entropy'. This is known as the maximum-entropy principle. The Shannon entropy of π is defined as:

$$SE(\pi) = -\sum_{i=1}^{N} \pi_i \ln(\pi_i) \tag{1}$$

To better understand the approach we will conduct, we introduce the classical dice problem which had been first stated by Jaynes [17]. In a classical dice problem, we know the empirical mean value (first moment) of a six-sided die, say μ_0. Suppose we would like to predict the probabilities $\pi = (\pi_1, \pi_2, \ldots, \pi_6)$ for each possible outcomes of a six-sided die for the next throw. We also know that the probability is appropriate, that is, the sum of the probabilities is one $\sum \pi$. Undoubtedly, there are infinite number of sets of values of π that satisfy the conditions. The difficulty of this problem is clear, there are six values to predict but only two observed value: the mean and the sum of the outcomes [18]. Using the maximum entropy principle, we can solve the optimization problem easily. The problem can be formed as the following:

$$\max SE(\pi) = -\sum_{i=1}^{6} \pi_i \ln(\pi_i) \tag{2}$$

$$\text{st.} \sum \pi_k x_k = y \quad \text{and} \quad \sum_k \pi_k = 1 \tag{3}$$

where $x_k = 1, \ldots, 6$ for $k = 1, \ldots, 6$ respectively. The solution for the problem is constructed by the Lagrangean function:

$$L = -\sum_{k=1}^{6} \pi_k \ln \pi_k + \lambda(y - \sum_k \pi_k x_k) + \mu(1 - \sum_k \pi_k) \tag{4}$$

$$\hat{\pi} = \frac{\exp(-\hat{\lambda} x_k)}{\sum_{k=1}^{6} \exp(-\hat{\lambda} x_k)} \tag{5}$$

3.2 The Maximum Entropy Method to Portfolio Selection Accounting for Firm Characteristics

Optimal portfolio weights for different stocks can be regarded as probabilities in the above dice problem. What we should do next is to find what is known and what is unknown and give constraints to the problem. However, if we use the cross entropy as the dice problem, we cannot allow short-selling and also cannot incorporate the firm characteristic into the weight function. Based on these two points, the Generalized Cross Entropy method is adopted in our study. Let's illustrate a little bit about cross entropy. We define the portfolio allocation with the unknown probability distribution $\pi = (\pi_1, \pi_2, \ldots, \pi_N)'$ among N risky assets, with proper constraint that $\sum \pi_i = 1$. Also, we can add some side constraints such as the variance ($\sigma^2 = \pi' \sum \pi$) is less than certain values or the utility function of the investor ($\pi' R - \frac{\lambda}{2}\sigma$) is greater than some values. Here we adopt the latter one.

Consider a diversified portfolio, we have the prior $q_i = 1/N$, which is an equal chance on each stock. Suppose a portfolio weight changes from π_i to q_i. In our analysis we will emphasize the minimization of cross entropy $CE(\pi, q)$ for a given q. Thus, we start from an initial portfolio allocation, by minimization of CE we can obtain a more diversified portfolio. The objective function and constraint function are the following:

$$\min_{\pi} CE(\pi|q) = -\sum_{i=1}^{N} \pi_i \ln(\pi_i/q_i) \tag{6}$$

$$U(\mu, \sigma^2|\pi, \lambda) = \pi' R - \frac{\lambda}{2}\sigma^2 \geq U_0, \pi \geq 0 \; and \; \pi' 1_N = 1 \tag{7}$$

The disadvantage of the cross entropy method is obvious: Any negative weights cannot be estimated in ln(.), therefore this method constrain short-selling. Therefore, we turn to the Generalized Cross entropy method with short-selling and firm characteristic variables. The details of this approach can be found in Golan et al. [18]. The weights of the optimal portfolio can be expressed by firm characteristics:

$$w = X\beta \tag{8}$$

where w is a $M \times 1$ vector of portfolio weights, X is an $M \times K$ matrix of firm characteristic variables, and β is a $K \times 1$ vector of unknown parameters.

We can estimate the unknown probabilities by the maximum entropy method in a discrete probability distribution $p_i = (p_{i1}, p_{i2}, \ldots, p_{iM})'$ over a set of support points z_i for each unknown parameter. Here we assume that both the unknown parameters and unknown errors should be bounded a priori. The maximum and minimum support point of z_i is the possible largest and smallest the value of β. The parameter supports can be based on prior information or economic theory. In our study, we use the support points between $[-3, 3]$. Let z_k be the $M \times 1$ support vector for the kth parameter and let p_k be the associated $M \times 1$ vector of probabilities of these support points. The unknown parameter vector β:

$$\beta = Zp = \begin{pmatrix} z_1' & 0 & \ldots & 0 \\ 0 & z_2' & \ldots & 0 \\ \vdots & \vdots & \vdots & \ddots \\ 0 & 0 & \ldots & z_k' \end{pmatrix} \begin{pmatrix} p_1 \\ p_2 \\ \vdots \\ p_k \end{pmatrix}$$

In our case, k = 3 for firm's size, ROE and also lag return. Therefore we can consider the following GCE minimization problem which allowing the short-selling and incorporate firms' characteristics:

$$\min \sum_{i=1}^{N} \sum_{m=1}^{M} p_{im} \ln(p_{im}/w_{im}) \tag{9}$$

$$(X\beta)'\hat{m} - \frac{\lambda}{2}(X\beta)'\hat{\Sigma}(X\beta) \geq \hat{G}^{-1}(r) \tag{10}$$

$$p_i' 1_M = 1 \tag{11}$$

$$(X\beta)' 1_N = 1 \tag{12}$$

Note that the prior of w_{im} can come from either minimization of variance problem or equal weights. It is not difficult to solve the optimization problem, but how to find out an explicit value of U_0. In Bera and Park [11]'s work, they use bootstrap to get the utility. However, we use the utility value:

$$U_0 = r \times (\hat{\pi}'R - \frac{\lambda}{2}\hat{\pi}' \sum \hat{\pi})) \tag{13}$$

where λ is equal to 1, and $\hat{\pi}'$ from minimization of variance problem. In our empirical study, we give r different values, r = 1.0 and 1.1. Therefore, e_r represents an investor's strength of belief, when r is larger, the investor has less uncertainty interval. The similar explanation can be found in detail in Bera and Park [11]. The differences between our work and theirs are as following aspects: first, they use bootstrap to

obtain the threshold of utility function values, however it is easy to use the utility from the minimized variance method. In our study, we try this out and find that this utility can make the calculation easier and also useful. And second, we add the firm' characteristics into the weight directly, the results show that this change improved the out-of-sample forecasting.

3.3 Discussion of Advantage and Disadvantage

We use the whole sample span, which will be introduced in the next section to test our method. In Fig. 1, the point A and B are minimize the variance, as well as efficient portfolio by setting a target return and also minimizing the variance. The dash line is the efficient frontier. In addition, C is cross entropy method with short-sell constraints, and D is the generalized cross entropy with firm's characteristic variables. The comparison of different methods is explicitly shown in the Table 1. Means and standard deviations of A, B, C and portfolios are (0.03, 3.253), (0.506, 4.887), (−0.0826, 3.596), and (0.021, 3.256), respectively. Table 2 gives the supports and resulting probabilities (p). It can be seen that the firm size has negative effects on the weights, the return on equity has positive effects and the lagged return has mixed effects in different companies. Since we assume that the effects of the firm characteristic variables are constant over the time. When we forecast one-month ahead weights, the weights sometimes do not sum to one. In such a case, we normalize it to one. In addition, we try different supports range, the results are almost consistent.

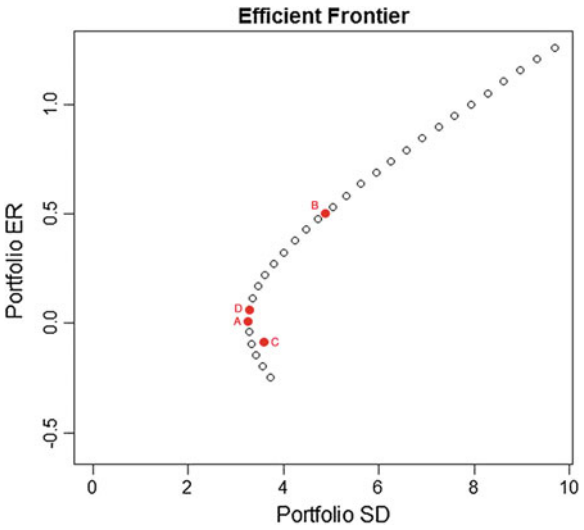


Fig. 1 Efficient frontier

Table 1 The weights of different portfolio

	Naive	MV1	MV2	CE	GCE
AMD	0.167	−0.085	−0.360	0.060	−0.100
APPL	0.167	0.074	0.243	0.122	0.085
BOA	0.167	0.048	−0.327	0.125	0.038
CISCO	0.167	0.840	1.39	0.460	0.850
JCP	0.167	0.100	−0.134	0.161	0.100
SIRI	0.167	0.022	0.190	0.069	0.027
Mean	−0.125	0.003	0.506	−0.082	0.021
SD	4.244	3.253	4.887	3.596	3.256
S.R	−0.029	0.001	0.104	−0.007	0.007
CEQ	−9.132	−5.288	−11.435	−6.450	−5.260

Note S.R is the sharpe ratio and CEQ is the certainty equivalent return

Table 2 The support points and probabilities of three firm characteristics

	Supports	−3	−1.5	0	1.5	3	Mean
	AMD	0.283	0.239	0.199	0.159	0.12	−0.609
	APPL	0.469	0.033	0.000	0.097	0.402	−0.105
Size	BOA	0.363	0.093	0.137	0.181	0.226	−0.279
	CISCO	0.357	0.025	0.000	0.518	0.1	−0.032
	JCP	0.431	0.158	0.146	0.136	0.128	−0.942
	SIRI	0.335	0.067	0.133	0.198	0.267	−0.007
	Supports	−3	−1.5	0	1.5	3	Mean
	AMD	0.222	0.163	0.107	0.057	0.452	0.531
	APPL	0.151	0.205	0.21	0.215	0.219	0.219
ROE	BOA	0.249	0.224	0.2	0.176	0.151	−0.366
	CISCO	0.159	0.205	0.208	0.212	0.216	0.1815
	JCP	0.198	0.196	0.199	0.202	0.205	0.03
	SIRI	0.191	0.194	0.199	0.206	0.211	0.078
	Supports	−10	−5	0	5	10	Mean
	AMD	0.231	0.214	0.199	0.185	0.171	−0.745
	APPL	0.13	0.206	0.214	0.221	0.229	1.065
Lag Ret	BOA	0.231	0.214	0.199	0.185	0.171	−0.745
	CISCO	0.221	0.197	0.196	0.194	0.192	−0.305
	JCP	0.068	0.063	0.058	0.054	0.756	6.835
	SIRI	0.205	0.201	0.199	0.198	0.197	−0.095

4 Empirical Application

4.1 Data Description

To illustrate the effectiveness of our maximum entropy based method, we present an empirical application of six stocks from the period from January 2004 to August 2014, totaling 127 observations. The six companies are: Advanced Micro Devices, INC. (AMD), Apple Computer, INC (AAPL), Bank of America (BOA), Cisco Systems, INC (CISCO), J.C. Penney Company, INC (JC), and Sirius XM Holdings INC (Siri), which are in the area of microprocessor, electronic computers, finance, routing system, department stores, broadcasting and cable TV; they cover different business. The summary of the data can be found in Table 3 and Fig. 2. Table 4 presents the correlation and covariance matrix of the stocks. It can be seen that the correlation among stocks are generally significant and positive.

For each firm, we collected the data from the annual report (2004–2014) to construct the following variables: log of total asset (the size of the firms), and the firm's return on equity (ROE), defining as net income divided by shareholder's equity (total asset minus liability), which represent a firm's profitability revealing the profit a company generates with the money that the shareholders have invested in. For the last eight months in 2014, we used the quarterly report instead of the annual one if the annual report had not been published on time. And for the third variable lag in one-year return (lagret), we used the return in the same month of last year. For every characteristics variable, we standardized them into zero mean and unit standard deviation. Similar characteristics are commonly used in the literatures [9, 10]. Figure 3, with three sub figures provides further details about the firm-level data.

4.2 The Results of Out-of-Sample Forecasts

We compared the performance of the above portfolio allocation models, they are MV_1 efficient method (with minimizing variance functions); MV_2 efficient method

Table 3 The description of six stocks

	AMD	APPL	BOA	CISCO	JCP	SIRI
Mean	−0.43	0.51	−0.57	−0.01	−0.33	0.09
Min	−21.59	−83.32	−33.04	−10.37	−18.91	−36.80
Max	18.30	13.11	23.72	8.99	18.44	33.99
S.D	7.44	9.02	6.72	3.43	5.87	8.09
Skewness	−0.10	−6.51	−1.38	−0.20	−0.18	−0.30
Kurtosis	0.14	56.53	6.88	0.42	0.96	5.49
JB statistics	0.39	18402.32	303.92	2.04	6.27	169.00
Probabiltiy	0.82	0.00	0.00	0.36	0.04	0.00

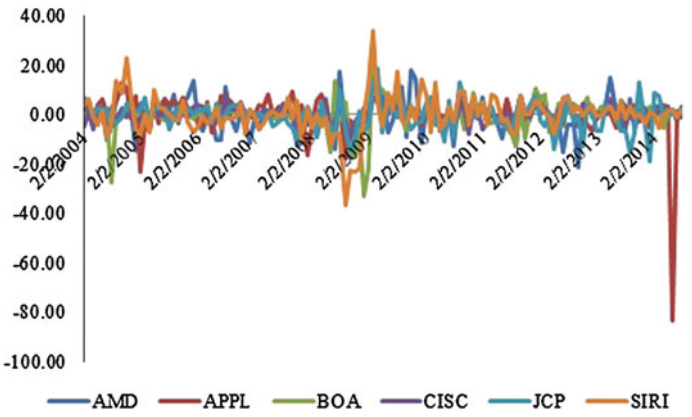


Fig. 2 The return of six stocks from 2004 to 2014

Table 4 The correlation and covariance of six stocks

	AMD	APPL	BOA	CISCO	JCP	SIRI
AMD	X	9.690	14.917	13.980	15.926	22.568
APPL	0.144	X	4.037	5.042	6.206	14.566
BOA	0.298	0.066	X	9.283	11.969	17.662
CISCO	0.547	0.162	0.402	X	8.504	9.193
JCP	0.364	0.117	0.303	0.422	X	13.917
SIRI	0.374	0.199	0.324	0.331	0.293	X

Note The upper triangle is the covariance matrix, the lower triangle is the correlation matrix

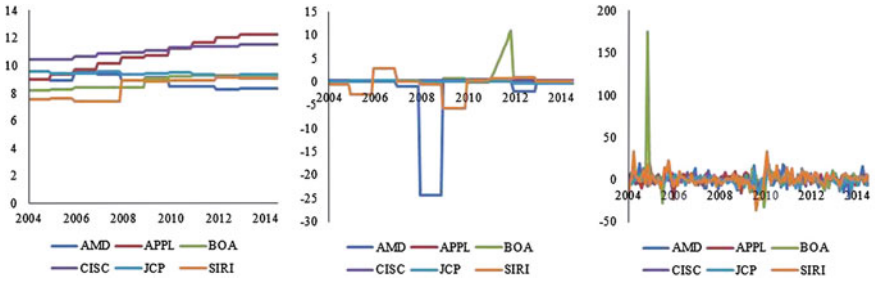


Fig. 3 The three firm characteristics: size, ROE and lag return (*Note* From *left* to *right*, they are size, ROE and lag return, respectively)

(with minimizing variance and set a certain target); naive method (equally weighted portfolio) and CE_1 (with short-selling constraints), GCE_2 (with firm characteristics). In the MV_1 method, the optimizing problem is built as:

$$\min(\pi' \sum \pi) \quad s.t. \quad \pi' R = \bar{\mu} \tag{14}$$

And in the MV_2 method, the objective function(criterion) changes into:

$$\min(\pi' \sum \pi) \tag{15}$$

With the help of program GAMS to handle the optimization problem [19], in order to analyze the portfolio performance with different methods we used the "rolling window" scheme. We considered two window lengths, W = 25, 49 months, they are two and four year time span. To evaluate the performance of each model, we used two evaluation measures; they are the Sharpe ratio (SR) and the certainty equivalent return (CEQ). The Sharpe ratio is widely used in finance, generally speaking, it is the ratio between the portfolio mean and standard deviation. Therefore, the meaning is very clear, when there is a greater number, the portfolio becomes better since it has higher return but lower risk [20, 21]. The equation is the following:

$$Sharpe\ Ratio = \frac{1}{T-W} \sum_{t=W}^{T} \frac{\hat{\pi}_t' \hat{m}_t}{(\hat{\pi}_t' \hat{\sum}_t \hat{\pi}_t)^{1/2}} \tag{16}$$

$$CEQ = \hat{\pi}_t' R - \frac{\lambda}{2} \hat{\pi}_t' \hat{\sum}_t \hat{\pi}_t \tag{17}$$

Here the only unknown parameter is λ. Actually we should evaluate portfolio performances by several different risk aversion parameter to represent different groups, λ can be any number in (0, 1), when λ approach to zero, the investor is the risk-lover, when λ approach to one, the investor is risk adverse. We present the results of $\lambda = 1$ due to the results are not much different. In Table 5, we show the rolling window length W = 25 and 49. And the portfolio weights are plot in Fig. 3. There are some interesting findings: first, CE performs the best in terms of both SR and CEQ among

Table 5 The in-sample and out-of-sample forecasts

In-Sample(25)							
Measures	Naive	MV_1	MV_2	CE_1	GCE_1	CE_2	GCE_2
SR	0.229	0.206	0.645	0.400	0.476	0.371	0.466
CEQ	−5.016	−2.114	−2.535	−1.876	−1.874	−2.063	−2.062
Out-of-Sample(25)							
SR	−0.078	−0.106	−0.022	−0.082	−0.117	−0.072	−0.127
CEQ	−10.079	−12.201	−10.607	−9.681	−9.631	−8.561	−10.630
In-Sample(49)							
SR	−0.036	0.088	0.294	0.129	0.137	0.103	0.104
CEQ	−7.253	−4.283	−6.186	−2.366	−2.372	−2.603	−2.597
Out-of-Sample(49)							
SR	−0.057	−0.141	−0.069	−0.127	−0.050	−0.132	−0.048
CEQ	−12.403	−22.081	−29.609	−23.062	−13.609	−23.581	−14.976

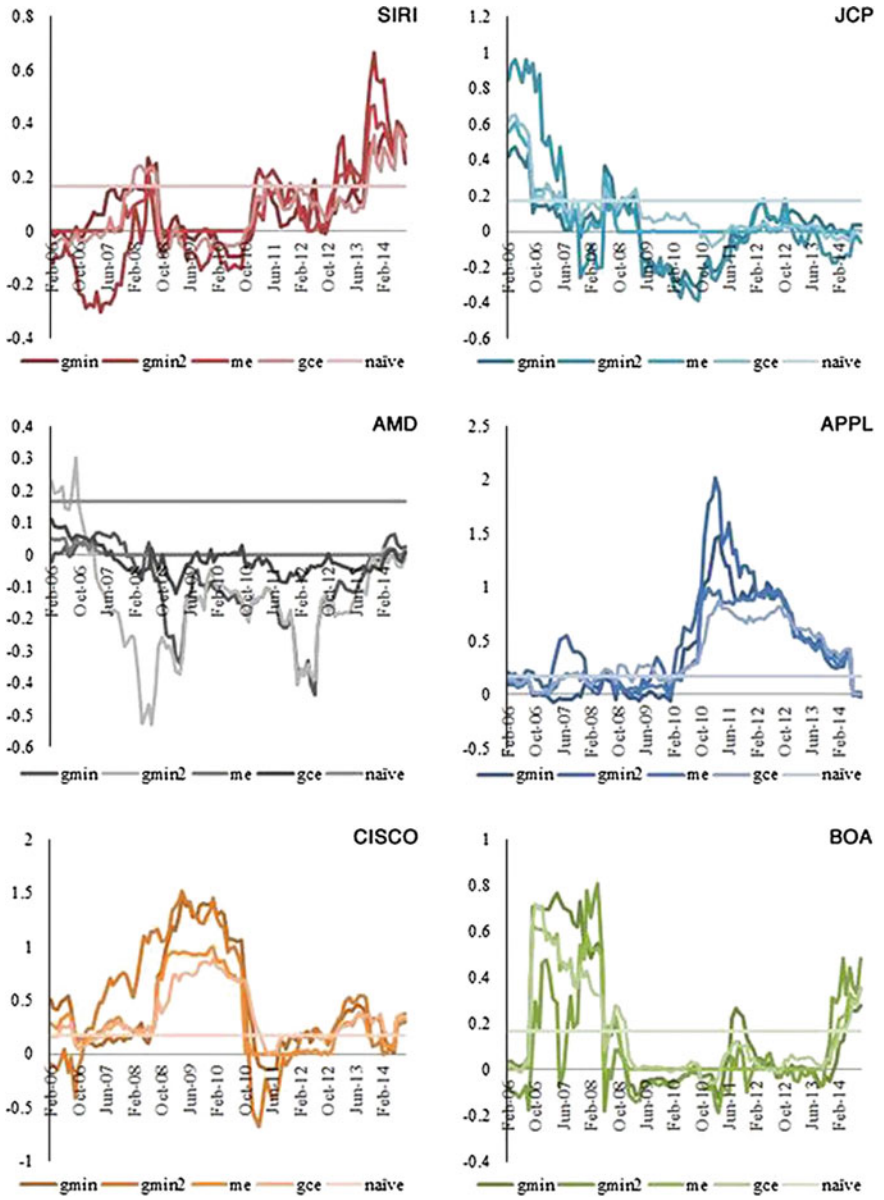


Fig. 4 The weights of optimal portfolios using different models

all considered models for in-sample case. Second, the out-of-sample of traditional MV method are not good, compared with the CE_1 and GCE_2. These results agree with those of Jorion [7], he compares the MV method with other nonparametric methods. Third, the results of out-of-sample forecasting are generally worse than the in-sample one and the naive method is better than others (Fig. 4).

The out-of-sample CEQ of CE_1 for r = 1.1, W = 24 is −8.561 which is the highest value among all considered models. On the other hand, the poor out-of-sample performance of GCE_2 shows that choosing MV portfolio with minimizing variance as q is not enough to improve the performance. As we increase the window length W from 25 to 49, we find that CEQ values of MV_1 are lower than that of CE_1 (Table 5). This better performance of MV is due to increased accuracy of the sample covariance estimates with relatively larger number of observations. For larger value of W, the performance of CE_2 is also much improved due to firm characteristics variable and shrinkage towards the equal weights. When W = 49, GCE_2 (r = 1) has the better out-of-sample SR and CEQ than CE_1.

5 Conclusion

This paper applies maximum entropy estimation in portfolio selection of six stocks, while accounting for the firm characteristics which complements the work of Bera and Park [11] proposing the cross entropy measure to portfolio diversification. This method combines the idea of shrinkage and firm characteristics variable, and overcomes the disadvantage of traditional Markowitz MV method. When the sample mean and variance are estimated incorrectly, it usually makes the portfolio concentrates on some stock and creates poor out-of-sample performances.

When we introduced the idea of firm characteristics to estimate the portfolio optimization using generalized cross entropy, the firm specific information about the stocks are brought into estimation. Furthermore, the cross entropy can be thought of as a direct shrinkage to the benchmark weights or MV portfolio weights. We also simplified the previous works and adopted an easy utility value instead of the bootstrap one. Although our method is simple, our empirical results show that it is equally efficient.

Moreover, in the empirical study, we analyzed the portfolio including six stocks. The results show that the Cross Entropy based model outperforms the traditional MV method; this demonstrates the model produces a better out-of-sample forecast. Although it is not better than the naive method, it can serve as a alternative tool in portfolio selection for investors, stock managers, and shareholders in the future.

References

1. Markowitz, H.M.: Portfolio Selection: Efficient Diversification of Investments, vol. 16. Yale University Press (1970)
2. Rubinstein, M.: Markowitzs "portfolio selection": a fifty year retrospective. J. Financ. **57**(3), 1041–1045 (2002)
3. Merton, R.C.: An analytic derivation of the efficient portfolio frontier. J. Financ. Quant. Anal. **7**(04), 1851–1872 (1972)

4. Bai, Z., Liu, H., Wong, W.K.: Enhancement of the applicability of Markowitz's portfolio optimization by utilizing random matrix theory. Math. Financ. **19**(4), 639–667 (2009)
5. Jagannathan, R., Ma, T.: Risk reduction in large portfolios: why imposing the wrong constraints helps. J. Financ. **58**(4), 1651–1684 (2003)
6. Frost, P.A., Savarino, J.E.: An empirical Bayes approach to efficient portfolio selection. J. Financ. Quant. Anal. **21**(03), 293–305 (1986)
7. Jorion, P.: Bayes-Stein estimation for portfolio analysis. J. Quant. Anal. **21**(03), 279–292 (1986)
8. Michaud, R.O.: The Markowitz optimization enigma: is 'optimized' optimal? Financ. Anal. J. **45**, 31–42 (1989)
9. Brandt, M.W., Santa-Clara, P., Valkanov, R.: Parametric portfolio policies: exploiting characteristics in the cross-section of equity returns. Rev. Financ. Stud. hhp003 (2009)
10. Fama, E.F., French, K.R.: Multifactor explanations of asset pricing anomalies. J. Financ. **51**(1), 55–84 (1996)
11. Bera, A.K., Park, S.Y.: Optimal portfolio diversification using the maximum entropy principle. Econom. Rev. **27**(4–6), 484–512 (2008)
12. Barry, C.B.: Portfolio analysis under uncertain means, variances, and covariances. J. Financ. **29**(2), 515–522 (1974)
13. AitSahalia, Y., Brandt, M.W.: Variable selection for portfolio choice. J. Financ. **56**(4), 1297–1351 (2001)
14. Nigmatullin, E.A.: Bayesian model averaging for moment conditions models. Working Paper, University of Wisconsin-Madison 2003
15. Jaynes, E.T.: Notes on present status and future prospects. In: Maximum Entropy and Bayesian Methods, pp. 1–13. Springer, The Netherlands (1991)
16. Jaynes, E.T.: Information theory and statistical mechanics. Phys. Rev. **106**(4), 620 (1957)
17. Jaynes, E.T.: New Engineering Applications of Information Theory. In: Proceedings of the first symposium on Engineering. Wiley (1963)
18. Golan, A., Judge, G.G., Miller, D.: Maximum entropy econometrics (No. 1488) (1996)
19. Rosenthal, R.E.: GAMS—a user's guide (2004)
20. Sharpe, W.F.: The sharpe ratio. Streetwise Best J. Portf. Manag. 169–185 (1998)
21. Cenesizoglu, T., Timmermann, A.: Do return prediction models add economic value. J. Bank. Financ. **36**(11), 2974–2987 (2012)

Risk, Return and International Portfolio Analysis: Entropy and Linear Belief Functions

Apiwat Ayusuk and Songsak Sriboonchitta

Abstract In this study, we analyze the international portfolio with respect to risk and return aspects. We applied entropy methods to find the optimal portfolio weights. In this method, we used entropy as the objective function and we also compared our results with the conventional method. Moreover, we use the linear belief function to build a portfolio, which can represent market information and financial knowledge and then we use matrix sweepings to integrate the knowledge for evaluating portfolio performance. Overall, our empirical analysis indicates that all entropy methods performed better than Markowitz method, and the finding also suggests that the investor should take the benefit from ASEAN market.

1 Introduction

Risk and return are important factors when investing in the capital market. According to the risk and return trade-off, the capital invested in the market cannot make higher returns without the possibility of investment loss. In classical work, Markowitz [14] is a well-known for the foundation of modern portfolio theory; which is mean and variance based method to find the optimal portfolio weights. Several researches were extensively studied in both theoretical and empirical works. (See, Tobin [19], Markowitz [15], Hakansson [6], Zenios and Kang [20], Konno and Kobayashi [10], etc.)

The modern world of economic globalization has a quick changing impact on the capital markets that contributed to an increase in international capital flow across countries. Many researches on topics related to international diversification, including Cavaglia et al. [1], Li [18], Fletcher and Marshall [4], Chiou [2] and Herrero and Vzquez [7] recommend that international diversification improves

A. Ayusuk (✉) · S. Sriboonchitta
Faculty of Economics, Chiang Mai University, Chiang Mai 50200, Thailand
e-mail: pai.mr.flute@gmail.com

A. Ayusuk
Department of Business Economics, Faculty of Liberal Arts and Management Sciences, Prince of Songkla University, Suratthani 84000, Thailand

V.-N. Huynh et al. (eds.), *Econometrics of Risk*, Studies in Computational Intelligence 583, DOI 10.1007/978-3-319-13449-9_22

portfolio performance. During the last decade, The Chinese economy has been rapidly developed and played an important role in Asia and the world. Jayasuriya [8] found evidence that the stock market behavior of China had an impact on the stock market behavior of the East Asia and Pacific. Zhou et al. [21] found that the impact of the Chinese stock market on Asian markets had become increasingly powerful after 2005. Glick and Hutchison [5] also found that the strength of the correlation of stock markets between China and other Asia countries has increased markedly during 2008–2010 and has remained high in 2010–2012. In 2015, the ASEAN Economic Community (AEC) will induce regional economic integration, which provides a competitive advantage and economic benefits. Hence, to take advantage of investment diversification an international level, this study focuses on the stock markets in ASEAN, China and The U.S., which is the world major stock market.

In our review of the literature, we found two specific research questions. First, what is the most efficient tool for portfolio allocation under international risk and return strategy? Second, which portfolio should we invest in? Therefore, the primary objective of this research is to suggest new portfolio selection methods under risk and return using an information theory to select the optimal portfolio and the linear belief function to combine evidence. The secondary objective is to evaluate international portfolio performance.

The remainder of this paper is organized as follows. We give more details about the portfolio optimization methods and portfolio analysis in the system of the linear belief function in Sect. 2. We examine the data selection, descriptive statistics and the results of portfolio analysis in Sect. 3. Finally provides a brief conclusion.

2 Methodology

2.1 Portfolio Selection Methods

In this section we present four different methods to determine the optimal portfolio based on risk-return framework. The basic notations are defined by: $r_{i,t}$ is the return of market i at time t, μ_i is the expected return of market i, $\sigma_{i,j}$ is the covariance between the market of i and j, p_i and p_j are the weights assigned to markets i and j, μ_p is the expected return of portfolio, σ_p^2 is the portfolio risk. While m denote number of markets in portfolio, then expected return and variance of return of portfolio can be described by $\mu_p = \sum_i^m p_i \mu_i$ and $\sigma_p^2 = \sum_{i=1}^m \sum_{j=1}^m p_i p_j \sigma_{i,j}$ respectively.

2.1.1 Mean-Variance Markowitz Method

The conventional work of MV method is well known for the portfolio optimization approach. The goal of an investor is to find the optimal weight determinations in a portfolio by minimizing risk subjecting to the expected return of the portfolio being

greater than or equal to risk free rate. The problem can be stated as:

$$\begin{aligned} &Minimize\ \sigma_p^2 \\ &st.\ \sum_i^m p_i\mu_i \geq \mu_0, \quad \sum_i^m p_i = 1 \end{aligned} \tag{1}$$

2.1.2 Mean Entropy Method

Entropy is a one of the methods to measure uncertainty in random variables. This study uses the Shannon entropy, $S(p) = -\sum_i^m p_i ln(p_i)$ under the principle of maximum entropy introduced by Jaynes [9]. The optimization problem is to choose the probability (or weight) in a portfolio by maximizing entropy function subject to the expected return (mean) of the portfolio being greater than or equal to risk free rate.

$$\begin{aligned} &Maximize -\sum_i^m p_i ln(p_i) \\ &st.\ \sum_i^m p_i\mu_i \geq \mu_0, \quad \sum_i^m p_i = 1 \end{aligned} \tag{2}$$

2.1.3 Mean-Variance Entropy Method

As the constraint of the principle of maximum entropy can be flexible, then we can provide more information. This optimization problem becomes maximizing the Shannon entropy subject to the expected return condition and the risk limitation strategy.

$$\begin{aligned} &Maximize -\sum_i^m p_i ln(p_i) \\ &st.\ \sum_i^m p_i\mu_i \geq \mu_0, \quad \sum_{i=1}^m \sum_{j=1}^m p_i p_j \sigma_{i,j} \leq \sigma_p^2, \quad \sum_i^m p_i = 1 \end{aligned} \tag{3}$$

2.1.4 Sharpe Ratio Entropy Method

The Sharpe ratio is introduced by Sharpe [16] to measure the portfolio performance that is described in unit of return per unit of risk. Therefore, we propose this methodology based on the Sharpe ratio into the principle of maximum entropy. The optimization problem is maximizing the Shannon entropy with an additional criterion of excess return per unit of risk.

$$Maximize - \sum_{i}^{m} p_i ln(p_i)$$

$$st.\ \frac{\mu_p}{\sigma_p} \geq \frac{\mu_0}{\sigma_0}, \quad \sum_{i}^{m} p_i = 1 \tag{4}$$

After we complete the solutions from each method, we use the Sharpe ratio to compare the performance of the portfolio selection methods.

2.2 Linear Belief Function

According to Dempster [3] and Liu [11], Linear belief functions is a special type of belief functions in expert system such as linear equations, linear regressions and Kalman filters, and also including Gaussian distributions that explain probabilistic knowledge on a set of variables in the continuous case. Liu et al. [13] used matrix sweepings to combine information from the linear belief function. In this study, we extended by using the linear time series belief function. Consequently, this study considers the linear time series to model portfolio investment by using the reduced form vector autoregressive (VAR) model with market returns as follows:

$$r_{i,t} = \Phi_0 + \Phi_1 r_{i,t-1} + e_t \tag{5}$$

where $r_{i,t} = [r_{1,t}, r_{2,t}, r_{3,t}]'$ is a 3×1 market return vector at time t that consists of ASEAN (AS), Chinese (CN) and U.S. markets, respectively. Φ_0 is a 3×1 vector of intercepts, Φ_1 is a time-invariant 3×3, e_t is a 3×1 vector of error terms and assume Gaussian distribution $e_t \sim N(0, \Sigma)$ with satisfying $E(e_t) = 0$, $E\left(e_t, e'_{t-1}\right) = 0$ is no serial correlation in error term and $E\left(e_t, e'_t\right) = \Sigma_{ij}$ is the variance-covariance matrix of error term that allowing non-zero correlation between error terms. The parameters of the VAR model can be estimated consistently by the OLS method when sample size is large. We construct a graphical structure for international portfolio analysis as follows:

In Fig. 1 we used the linear relationship from a VAR model combining with optimal portfolio weight to construct a graphical structure of international portfolio. There are ten variable nodes: $AS_t, CN_t, US_t, AS_{t-1}, CN_{t-1}, US_{t-1}, E_{AS_t}, E_{CN_t}, E_{US_t}, P$ and ten belief function nodes. Four linear belief functions represent the relationship between the variables, e.g. $Bel(AS_t, AS_{t-1}, CN_{t-1}, US_{t-1}, E_{AS_t})$ is a linear belief function of the ASEAN return that depended on the first lag of past returns of itself, China, and the U.S. and a residual variable. $Bel(P, AS_t, CN_t, US_t)$ is a linear belief function of portfolio return that is integrated with three market returns and optimal portfolio weights from best methods. And six linear belief functions represent an individual variable, e.g. $Bel(E_{AS_t})$ is the true value of a residual variable from

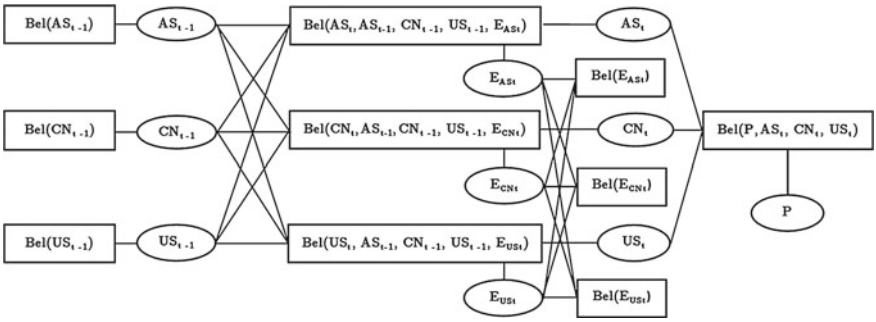


Fig. 1 A graphical structure of international portfolio

first function. Thus, we can analyze the linear belief function into the moment matrix approach.

The concept of Dempster's rule used to combine multiple focal elements that are independent evidence from several sources. Liu [11] proved the combination rule in Gaussian linear belief function of variable space was equivalent to that of Dempster [3]'s for continuous case. Liu [12] also proved that combination and marginalization of Gaussian linear belief function satisfies the axioms of Shenoy and Shafer [17] and showed Dempster's rule for the combination could be interpreted by matrix sweepings.

According to the matrix sweeping technique for Gaussian linear belief function is a matrix operation or a matrix transformation that including forward sweep and reverse sweep to consider. We can sweep a matrix from variance and covariance matrix move to conditional representation. Let r_i be a random variable representing the market returns that are assuming Gaussian distribution with expected mean: $E(r_i) = \mu_i$, variance: $Var(r_i) = \Sigma_{ii}$ and covariance $Cov(r_i, r_j) = \Sigma_{ij}, i, j = 1, 2, \ldots, n$, then we can write the moment matrix as $m = \begin{pmatrix} \mu_j \\ \Sigma_{ij} \end{pmatrix}$, This matrix represents the distribution of the random variables. We can define the operation on moment matrices by definitions below.

Definition 1 (*Marginalization*) Liu [12], the marginalization of a linear belief function is simply a projection in variable space. Let r_1 and r_2 are two random variables in the moment matrix: $m(r_1, r_2)$, its marginal to r_1 as

$$m^{\downarrow r_1}(r_1, r_2) = \begin{pmatrix} \mu_j \\ \Sigma_{ij} \end{pmatrix} \tag{6}$$

where $m^{\downarrow r_1}$ is the marginalization of the moment matrix that represent to the conditional moment matrix of linear regression coefficient.

Definition 2 (*Forward sweep*) Liu [12], Forward sweeping is the transformation of the moment matrix to be the conditional moment matrix. Let n market returns in

portfolio and then we can operate a forward sweep of $m(r_1, , r_n)$ from r_s as follows:

$$m(r_1, \ldots, r_{s-1}, \overrightarrow{r_s}, r_{s+1}, \ldots, r_n) = \begin{pmatrix} \mu_{j,s} \\ \Sigma_{ij,s} \end{pmatrix} \tag{7}$$

where

$$\mu_{j,s} = \begin{cases} \mu_j - \mu_s \Sigma_{ss}^{-1} \Sigma_{sj}, & for\ j \neq s \\ \mu_s \Sigma_{ss}^{-1}, & for\ j = s \end{cases}$$

$$\Sigma_{ij,s} = \begin{cases} -\Sigma_{ss}^{-1}, & for\ i = s = j \\ \Sigma_{is} \Sigma_{ss}^{-1}, & for\ j = s \neq i \\ \Sigma_{ss}^{-1} \Sigma_{sj}, & for\ i = s \neq j \\ \Sigma_{ij} - \Sigma_{is} \Sigma_{ss}^{-1} \Sigma_{sj}, & for\ \ otherwise \end{cases}$$

Definition 3 (*Reverse sweep*) Liu [12], Let n market returns in portfolio and then we can operate a reverse sweep of $m(\overrightarrow{r_1}, , \overrightarrow{r_n})$ from r_s as follows:

$$m(\overrightarrow{r}_1, \ldots, \overrightarrow{r}_{s-1}, r_s, \overrightarrow{r}_{s+1}, \ldots, \overrightarrow{r}_n) = \begin{pmatrix} \mu_{j,s} \\ \Sigma_{ij,s} \end{pmatrix} \tag{8}$$

where

$$\mu_{j,s} = \begin{cases} \mu_j - \mu_s \Sigma_{ss}^{-1} \Sigma_{sj}, & for\ j \neq s \\ -\mu_s \Sigma_{ss}^{-1}, & for\ j = s \end{cases}$$

$$\Sigma_{ij,s} = \begin{cases} -\Sigma_{ss}^{-1}, & for\ i = s = j \\ -\Sigma_{is} \Sigma_{ss}^{-1}, & for\ j = s \neq i \\ -\Sigma_{ss}^{-1} \Sigma_{sj}, & for\ i = s \neq j \\ \Sigma_{ij} - \Sigma_{is} \Sigma_{ss}^{-1} \Sigma_{sj}, & for\ \ otherwise \end{cases}$$

Definition 4 (*The combined linear belief function*) Liu [12] The combination of two linear belief functions is the sum of fully swept matrices: $\overrightarrow{m} = \overrightarrow{m_1} \oplus \overrightarrow{m_2}$ and then we can write this combination as follows:

$$\overrightarrow{m} = \overrightarrow{m_1} \oplus \overrightarrow{m_2} = \begin{pmatrix} \overrightarrow{\mu_1} + \overrightarrow{\mu_2} \\ \overrightarrow{\Sigma_1} + \overrightarrow{\Sigma_2} \end{pmatrix} \tag{9}$$

3 An Application to International Portfolio Evaluation

The research is performed as follows: Firstly, we calculate the optimal weights of international portfolio. There are different methods to optimize the portfolio selection problem; Mean-Variance Markowitz (MV) method, Mean Entropy (ME) method, Mean-Variance Entropy (MVE) method. Secondly, we use the Sharpe ratio to measure the portfolio performance and select the best performance method. Thirdly, we

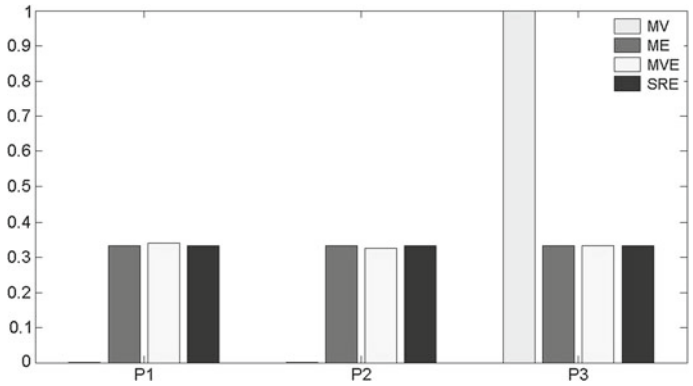


Fig. 2 The Optimal portfolio weight for four methods

construct the network of portfolio structure by using linear time series belief function. Finally, applying matrix sweepings to integrate the knowledge and information from second and third to evaluate the portfolio performance.

We collected daily data from January of 2009 to December of 2013 from Data Stream. As mentioned before, we considered a portfolio selection problem from three attractive markets, which are ASEAN (FTSE/ASEAN index), China (The Shanghai Composite index) and the U.S. (The S&P 500 index) markets.

Figure 2 presents the optimal weights of portfolios that are computed from four different methods. Table 1 presents the performances of the portfolio selection methods. From the results of Sharpe ratios, ME, MVE and SRE perform better than MV. MVE is better than other considered methods. Its optimal weights are 34.0 %, 32.6 % and 33.4 % in ASEAN, China and The U.S. markets respectively.

Table 2 shows the results of the parameter estimates in a VAR model. It represents the relationship between international markets, which should have an influence on each other. The results show that the first lag of the U.S. has high influence in ASEAN and China.

According to the financial information and market knowledge from the above results, we use MVE to optimize the portfolio weights because this method performs better than others. The optimal weight can represent by using the partially swept

Table 1 Comparison results for the portfolio selection methods

Methods	Portfolio returns	Portfolio variance	Sharpe ratios
MV	0.000549	0.000145	0.0373
ME	0.000421	0.000066	0.0396
MVE	0.000424	0.000065	0.0402
SRE	0.000421	0.000066	0.0396

MV Mean-Variance Markowitz method, *ME* Mean Entropy method,
MVE Mean-Variance Entropy method, *SRE* Sharpe Ratio Entropy method

Table 2 Estimates of a VAR model

Methods	ASEAN	China	U.S
ASEAN(−1)	−0.010810	−0.067352	0.004051
	[0.02801]	[0.04710]	[0.04308]
China(−1)	−0.040520	0.002412	0.015236
	[0.01770]	[0.02977]	[0.02722]
U.S.(−1)	0.330312	0.221640	−0.092744
	[0.01870]	[0.03145]	[0.02876]
Constant	0.000431	0.000031	0.000596
	[0.00022]	[0.00037]	[0.00033]
R-squared	0.202821	0.036987	0.008427
Schwarz SC	−6.846264	−5.806809	−5.985372

In parentheses are standard errors of the coefficient estimates

matrix as

$$m(P, \overrightarrow{AS}_t, \overrightarrow{CN}_t, \overrightarrow{US}_t) = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0.3403 & 0.3255 & 0.3255 \\ 0.3403 & 0 & 0 & 0 \\ 0.3255 & 0 & 0 & 0 \\ 0.3255 & 0 & 0 & 0 \end{pmatrix} \tag{10}$$

Figure 1 $Bel(AS_t, AS_{t-1}, CN_{t-1}, US_{t-1}, E_{AS_t})$, we can define by the partially swept matrix from the first equation in a VAR model, $m(AS_t, \overrightarrow{AS}_{t-1}, \overrightarrow{CN}_{t-1}, \overrightarrow{US}_{t-1}, \overrightarrow{E}_{AS_t})$ with $Var(e_{AS_t}) = 0.00006$ is a variance of residual and $Cov(e_{AS_t}, e_{CNt_t}) = 0.00004$, $Cov(e_{AS_t}, e_{CN_t}) = 0.00004$ are covariance of residual as follows:

$$m(AS_t, \overrightarrow{AS}_{t-1}, \overrightarrow{CN}_{t-1}, \overrightarrow{US}_{t-1}, \overrightarrow{E}_{AS_t}) = \begin{pmatrix} 0.000431 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & -0.01081 & -0.04052 & 0.33031 & 1 & 1 & 1 \\ -0.01081 & 0 & 0 & 0 & 0 & 0 & 0 \\ -0.04052 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0.33031 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0.00006 & 0.00004 & 0.00004 \\ 1 & 0 & 0 & 0 & 0.00004 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0.00004 & 0 & 0 \end{pmatrix} \tag{11}$$

Therefore, to analyze the portfolio performance when the markets are related by using the linear belief function. We use six step method of Liu, Shenoy and Shenoy [13] to integrate knowledge using the combination of matrix sweeping.

Table 3 The results for moment matrix in portfolio

	Portfolio	ASEAN	China	U.S
Return	0.000584	0.0005942	0.000110	0.000551
Var-Cov	0.000051			
	0.000058	0.000077		
	0.000007	0.000011	0.000180	
	0.000031	−0.000005	−0.000003	0.000146
Sharpe ratio	0.067433			

Table 3 presents portfolio performance using the linear belief function. The result shows risks and returns in a portfolio: the ASEAN return 0.0594 % is highest, the standard deviation of ASEAN 0.8775 % is smallest, and the portfolio return is 0.0584 % with the standard deviation 0.7141 %.

4 Conclusions

This study provided empirical example for ASEAN, China and the U.S. markets between January 2009 and December 2013. We use three entropy methods base on Sharnon measure, which are ME, MVE and SRE, to select the optimal weights and compare its performances with conventional method, which is MV. Moreover, we use the linear belief function to extend portfolio evaluation because the belief function method can allow us to add information on the conditions of the relationship between international markets. There are two main findings from this study. First, all entropy methods perform better than MV because the entropy method well handles uncertainty information from simulations. Moreover, we found that MVE has higher performed than ME and SRE since MVE has added more information in constrains. Second, after integrating the information between the optimal portfolio strategy and international market relationship using linear belief function, we found that the portfolio risk is decreased and the portfolio return is increased. This implies that the relationship of international markets affects portfolio performance and this finding suggests that an investor should increase the investment proportion in the ASEAN market.

Acknowledgments The authors are very grateful to Prof. Thierry Denoeux for his comments and Prof. Amos Golan for the concept of Entropy Econometrics. This study was supported from Prince of Songkla University-PhD Scholarship.

References

1. Cavaglia, S., Hodrick, R., Vadim, M., Zhang, X.: Pricing the Global Industry Portfolios. Working Paper, National Bureau of Economic Research (2002)
2. Chiou, W.J.P.: Benefits of international diversification with investment constraints: an over-time perspective. J. Multinatl. Financ. Manag. **19**(2), 93–110 (2009)

3. Dempster, A.P.: Normal Belief Functions and the Kalman Filter Research Report Department of Statistics. Harvard University, Cambridge (1990)
4. Fletcher, J., Marshall, A.: An empirical examination of the benefits of international diversification. J. Int. Financ. Mark. **15**(5), 455–468 (2005)
5. Glick, R., Hutchison, M.: Chinasfinancial linkages with Asia and the globalfinancial crisis. J. Int. Money Financ. **39**, 186–206 (2013)
6. Hakansson, N.: Capital growth and the mean-variance approach to portfolio selection. J. Financ. Quant. Anal. **6**(1), 517–557 (1971)
7. Herrero, A.G., Vázquez, F.: International diversification gains and home bias in banking. J. Bank. Financ. **37**, 2560–2571 (2013)
8. Jayasuriya, S.A.: Stock market correlations between China and its emerging market neighbors. Emerg. Mark. Rev **12**(4), 418–431 (2011)
9. Jaynes, E.T.: Information theory and statistical mechanics. In: Statistical Physics, New York pp. 181–218 (1963)
10. Konno, H., Kobayashi, K.: An integrated stock-bond portfolio optimization model. J. Econ. Dyn. Control **21**, 1427–1444 (1997)
11. Liu, L.: A theory of Gaussian belief functions. Int. J. Approx. Reason. **14**(2–3), 95–126 (1996)
12. Liu, L.: Local computation of Gaussian belief functions. Int. J. Approx. Reason. **22**(3), 217–248 (1999)
13. Liu, L., Shenoy, C., Shenoy, P.P.: Knowledge representation and integration for portfolio evaluation using linear belief functions. IEEE Trans. Syst. Man, Cybern. Ser. A **36**(4), 774–785 (2006)
14. Markowitz, H.: Portfolio selection. J. Financ. **7**(1), 77–91 (1952)
15. Markowitz, H.: Portfolio Selection: Efficient Diversification of Investments. Wiley, New York (1959)
16. Sharpe, W.F.: Mutual fund performance. J. Bus. **39**(S1), 119–138 (1966)
17. Shenoy, P.P., Shafer, G.: Axioms for probability and belief-function propagation. In: Shachter, R.D., Levitt, T.S., Kanal, L.N., Lemmer, J.F. (eds.) Uncertainty in Artificial Intelligence, vol. 4, pp. 169–198. North-Holland, Amsterdam (1990)
18. Li, L.: An economic measure of diversification benefits. Working Paper, Yale International Center for Finance (2003)
19. Tobin, J.: Liquidity preference as behavior towards risk. Rev. Econ. Stud. **25**(2), 65–86 (1958)
20. Zenios, S.A., Kang, P.: Mean-absolute deviation portfolio optimization for mortgage backed securities. Ann. Oper. Res. **45**, 433–450 (1993)
21. Zhou, X., Zhang, W., Zhang, J.: Volatility spillovers between the Chinese and world equity markets. Pacific-Basin Financ. J. **20**(2), 247–270 (2012)

Forecasting Inbound Tourism Demand to China Using Time Series Models and Belief Functions

Jiechen Tang, Songsak Sriboonchitta and Xinyu Yuan

Abstract Modeling uncertainty is a key issue in forecasting. In the tourism area, forecasts are used by governments, airline companies and operators to design tourism policies and they should include a quantification of uncertainties. This paper proposed a new approach to forecast the tourism demand, which is time series models combined with belief functions. We used this method to predict the demand for China international tourism, with an explicit representation of forecast uncertainty. The monthly data of international tourist arrival cover the period from January 1991 to June 2013. The result show that time series models combined with belief functions is a computationally simple and effective method.

1 Introduction

In the last three decades, a large number of studies focused on tourism demand forecasting. The forecasting methods that have remained the most popular over the years in tourism are Times series models, which explain a variable with regard to its own past and a random disturbance term (e.g., [1–9]). Although these empirical can forecast the tourism demand, they do not consider the some measure of uncertainty. In order to fill this drawback, the purpose of this paper is to demonstrate the use of time series models combined with belief functions to forecast Chinese international tourism demand.

China has become one of most important tourism destinations among the world tourism. According to UNWTO Tourism Highlights [10], China ranks third in the world by arrivals with 57.7 million. Furthermore, China shares 5.57 % tourist arrivals in the world's market of 2012. Since Deng Xiaoping's economic reforms in 1978, inbound tourism demand, or tourism arrivals, to China has experienced

J. Tang (✉) · S. Sriboonchitta
Faculty of Economics, Chiang Mai University, Chiang Mai 50200, Thailand
e-mail: tangjiechen1002@163.com

X. Yuan
Faculty of Economics and Management, Yunnan Normal University, Yunnan, China
e-mail: ynandre@hotmail.com

V.-N. Huynh et al. (eds.), *Econometrics of Risk*, Studies in Computational Intelligence 583, DOI 10.1007/978-3-319-13449-9_23

dramatic changes due to economic reform and political environment.[1] Specially, China international arrivals reached 57.7 million in 2012, up from 55.89 million in 1978. This information highlights that China inbound tourism market play important role in international tourism industry. Hence, we select China to study.

This paper is organized as follows. Section 2 briefly introduces our methodological approach, including time series models and belief functions. Section 3 presents a study of several time series models applied to tourism data. The prediction method using time series models and belief functions is then investigated in Sect. 4. Finally, conclusions are presented in the last section.

2 Methodology

2.1 Time Series Models

In the past four decades, time series models have been widely used for tourism demand forecasting [5]. In this study, three different time series models were applied to forecast the monthly Chinese international tourist arrivals: Autoregressive integrated moving average (ARIMA), seasonal autoregressive integrated moving average (SARIMA) and generalized autoregressive conditional heteroskedasticity (GARCH) models.

The general expression of an ARIMA(p, d, q) model is the following:

$$\phi(L)\Delta^d y_t = c + \theta(L)\varepsilon_t \tag{1}$$

where L is the backward-shift operator, $\phi(L) = (1 - \phi_1 L^1 - \phi_2 L^2 - \cdots - \phi_p L^p)$ is a regular autoregressive polynomial, $\theta(L) = (1 + \theta_1 L^1 + \theta_2 L^2 + \cdots + \theta_q L^q)$ is a regular moving average polynomial, Δ^d is the regular difference operator, ε_t is an independent and identically distributed innovation term. The seasonal ARIMA (p, d, q) model can be written as SARIMA$(p, d, q) \times (P, D, Q)$, defined by the following expression:

$$\Phi_s\left(L^s\right)\phi\,(\mathrm{L})\,\Delta_s^D\Delta^d y_t = c + \Theta_s\left(L^s\right)\theta\,(L)\,\varepsilon_t \tag{2}$$

where $\Phi_s\,(L^s) = \left(1 - \Phi_{1s}L^{1s} - \Phi_{2s}L^{2s} - \cdots - \Phi_{Ps}L^{Ps}\right)$ is a seasonal autoregressive polynomial, $\Theta_s\,(L^s) = \left(1 + \Theta_{1s}L^{1s} + \Theta_{2s}L^{2s} + \cdots + \Theta_{Qs}L^{Qs}\right)$ a seasonal moving average polynomial, Δ_S^D is the seasonal difference operator, s is the periodicity of the considered time series.

[1] Although the founding of New China is in 1949, the ban on inbound travel for any purpose was enforced between 1949 and 1976. Since Deng Xiaoping's economic reforms in 1978, inbound tourism in China rapidly developed due to change in this policy (Lim and Pan [11]).

The general expression of the ARIMA(p, d, q) -GARCH(P, Q) model [12] can be written as:

$$\begin{aligned} \phi(L)\Delta^d y_t &= \theta(L)\varepsilon_t \\ \sigma_t^2 = \alpha_0 + \sum_{i=1}^{P} \alpha_i \varepsilon_{t-i}^2 + \sum_{j=1}^{Q} \beta_j \sigma_{t-j}^2 \\ \varepsilon_t &= z_t \sigma_t \end{aligned} \tag{3}$$

where r is the lag length of the moving average ARCH term, m is the lag length of the autoregressive GARCH term, σ_t is the conditional variance of volatility of ε_t, α_i is the ARCH parameter associated with ε_{t-i}^2, β_j is the GARCH parameter associated with σ_{t-j}^2, α_i and β_j are required to be positive, and z_t to be the standardized residual. The sum $\sum \alpha_i + \sum \beta_j$ should be less than unity to satisfy stationary conditions.

In order to apply these three models with forecasting purposes, we designed an algorithm that identifies that best suited model, including the necessary number of differences D and d and the number of lags that should be included in the model. Following [5], the best-fit models were identified based on the lowest Akaike Information Criteria (AIC).

2.2 Likelihood-Based Belief Function

In this section, we followed Denoeux [13], Abdallah et al. [14] and Kanjanatarakul et al. [15] to recall the definition of a belief functions from the likelihood function and its justification. Let $x \in X$ denote the observable data, $\varphi \in \Psi$ the parameter of interest and $f_\varphi(x)$ the probability mass or density function of X.

Denoeux [13] proved that the least committed belief function verifying $pl_x(\phi) \propto L_x(\phi)$ is the consonant belief function Bel_x^Ψ, whose contour function is the relative likelihood function[2]:

$$pl_x(\varphi) = \frac{L_x(\varphi)}{\sup_{\varphi' \in \Psi} L_x(\varphi')} \tag{4}$$

[2] According to the likelihood principle, $L_x(\phi)$ is the likelihood function defined by $L_x(\phi) = \xi f_\phi(x)$, for all $\phi \in \Psi$, where ξ is any positive multiplicative constant. On the basis of compatibility with Bayesian inference, the contour function $pl_x(\phi)$ associated to Bel_x^Ψ should be proportional to the likelihood function: $pl_x(\phi) \propto L_x(\varphi)$. More details can be found in Denoeux (2014).

This belief function is called the likelihood based belief function on Ψ induced by x. From $pl_x(\phi)$, we compute the corresponding plausibility function as:

$$Pl_x^{\Psi}(A) = \sup_{\varphi \in A} pl_x(\varphi) \tag{5}$$

where all $A \in \Psi$. The focal sets of Bel_x^{Ψ} are the levels sets of $pl_x(\phi)$, defined as:

$$\Gamma_x(\omega) = \{\varphi \in \Psi \,|\, pl_x(\varphi) \geq \omega\} \tag{6}$$

where $\omega \in [0, 1]$. The sets $\Gamma_x(\omega)$ are called plausibility regions and can be interpreted as sets of parameter values whose plausibility is greater than some threshold ω Denoeux [13]. According to Nguyen [16], the belief function Bel_x^{Ψ} is equivalent to to the random set induced by the Lebesgue measure λ on the [0, 1] and the multi-valued mapping Γ_x form [0, 1] to 2^{Ψ}. In particular, the following equalities hold:

$$\begin{aligned} Bel_x^{\Psi}(A) &= \lambda(\{\omega \in [0,1] \,|\, \Gamma_x(\omega) \subseteq A\}) \\ Pl_x^{\Psi}(A) &= \lambda\left(\left\{\omega \in [0,1] \,|\, \Gamma_x(\omega) \cap^A \neq \emptyset\right\}\right) \end{aligned} \tag{7}$$

for all $A \subseteq \Psi$ such that the above expressions are well-defined.

3 Estimation and Comparison of Time-Series Models

3.1 Data Description

The data set examined in this study is monthly total tourist arrivals in China.[3] The data used in this study are monthly inbound tourist arrivals in China from January 1991 to June 2013, which are obtained from EcoWin. Figure 1 shows that monthly international tourist arrivals exhibit a deterministic pattern of long-term upward trend. Note that the number of tourist arrivals has been increasing and the series appear to be non-stationary in that the mean is increasing over time. In order to satisfy the assumption of constant error variance, we first transformed the data using the log transformation.

To analyze China's inbound tourism demand, the descriptive statistics of the international tourist arrivals and logarithm international tourist arrivals are reported in Table 1. From Table 1, we find that the means of the international tourist arrivals and logarithm international tourist arrivals are 7485.7050 and 8.8194, respective. Second, JB test (Jarque and Bera [17]) show that the international tourist arrivals and logarithm international tourist arrivals are rejected to be normally distributed at

[3] Total tourist arrivals are all those traveling China on non-Chinese passports, include holders of Hong Kong, Macau and Republic of China (Taiwan) passports and travel documents.

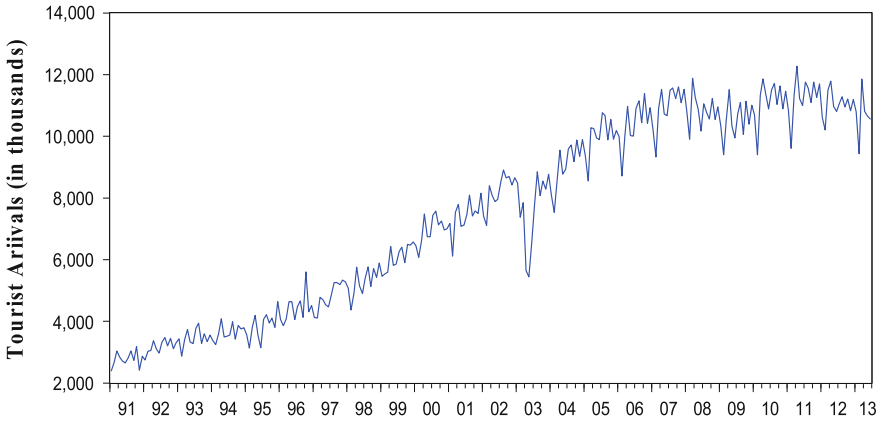


Fig. 1 Monthly international tourist arrivals to China

Table 1 Summary statistics of international tourist arrivals (in thousands) and result of unit root test

Data description						
	Mean	Maximum	Minimum	Skewness	Kurtosis	JB
y_t	7485.7050	12276.7000	2392.8000	−0.1174	1.5057	25.7393***
$\ln(y_t)$	8.8194	9.4155	7.7802	−0.4914	1.8267	26.3524***
Unit root test						
	ADF			KPSS		
	$\ln(y_t)$	$\Delta \ln(y_t)$	$\ln(y_t)$	$\Delta \ln(y_t)$	$\ln(y_t)$	$\Delta \ln(y_t)$
	2.0588	−3.5802***	2.0588	−3.5802***	2.0588	−3.5802***

Note JB is JarqueBera test. $\ln(y_t)$, $\Delta \ln(y_t)$, $\Delta\Delta_{12} \ln(y_t)$ denote logarithm monthly tourist arrivals, first difference of logarithm monthly tourist arrivals and first-twelfth difference of logarithm monthly tourist arrivals, respectively. This table shows the values of t-statistics and LM-statistics for ADF and KPSS unit root tests, respectively. ***, ** and * denote rejection of the null hypothesis at the 1%, 5%, and 10% significance levels, respectively

the 1 % significance level, inferring that the normal distribution is not appropriate to model international tourist arrivals and logarithm international tourist arrivals.

Moreover, QQ plot is plotted in Fig. 2, which also suggest tourist arrivals, is not normally distributed. Hence, we introduced the student-t distribution in the time series models and belief functions.

3.2 Empirical Results

The most popular time series models to forecast tourism demand are ARIMA and SARIMA. Although the ARIMA-GARCH model is uncommon in tourism

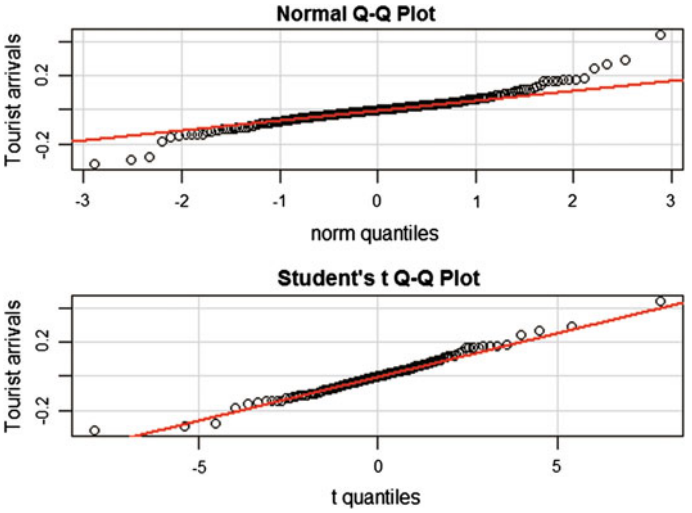


Fig. 2 QQ Plot of tourist arrivals

forecasting studies, it has been very popular in forecasting research in Finance. Hence, the ARIMA, SARIMA, and ARIMA-GARCH models have been considered. It is rare for the order (P, Q) of a GARCH model to be high, indeed the literature suggests that the parsimonious GARCH(1, 1) is often adequate for capturing volatility in time series data (Coshall [18]). For that reason, we used the ARIMA-GARCH (1, 1). This study follow the Box-Jenkins methodology to identify the most suitable ARIMA and SARIMA, ARIMA-GARCH(1, 1) models based on the estimation sample for tourist arrivals series.[4] The autocorrelation function (ACF) and partial autocorrelation function (PACF) were used to select the order of ARIMA, SARIMA and ARIMA-GARCH (1, 1) models. In additional, the best-fit models were identified based on the lowest AIC. In the ARIMA model and ARIMA-GARCH model, monthly tourist arrivals are processed by taking the first-order regular difference (denoted by ∇^1) in order to remove the growth trend and make sure the data is stationary. In the SARIMA model, monthly tourist arrivals are processed by taking the first-order regular difference (denoted by ∇^1) and the first seasonal differencing (denoted by ∇^1_{12}) in an effort to remove the growth trend and the seasonality characteristics and make sure the data is stationary. The best-fit ARIMA, ASRIMA and ARIMA-GARCH models for monthly tourist arrivals in China were estimated using on maximum likelihood estimation procedure. For monthly tourist arrivals series, the best fit of ARIMA, SARIMA and ARIMA-GARCH models generated from the datasets are presented below:

[4] Training data (insample data) is form January 1991 to June 2012. The remaining period from July 2013 to June 2013 are testing data (out-of-sample data).

ARIMA model:

$$\left(1-\phi_1 L^1-\phi_3 L^3-\phi_{12} L^{12}-\phi_{13} L^{13}-\phi_{15} L^{15}-\phi_{24} L^{24}\right) \Delta^1 \ln y_t = \left(1+\theta_2 L^2+\theta_6 L^6+\theta_{12} L^{12}\right) \varepsilon_t$$

SARIMA model:

$$\left(1-\phi_1 L^1\right)\left(1-\Theta_{12} L^{12}\right) \Delta^1 \Delta_{12}^1 \ln y_t = \left(1+\theta_1 L^1\right) \varepsilon_t$$

ARIMA-GARCH model:

$$\left(1-\phi_{12} L^{12}\right) \Delta^1 \ln y_t = \left(1+\theta_1 L^1+\theta_{12} L^{12}+\theta_{13} L^{13}\right) \varepsilon_t$$

$$\sigma_t^2 = c+\alpha_1 \varepsilon_{t-1}^2+\beta_1 \sigma_{t-1}^2$$

where ε_t follows a Student-t distribution.

The results of the best-fit ARIMA, ASRIMA and ARIMA-GARCH models are given in Table 2, where the estimated parameters providing all of the AR and MA terms are statistically significant at the 10 % level. In order to confirm the adequacy of the selected models, we used the ACF and PACF diagnostic correlograms as well as the Ljunge Box (LB) test to verify that the residuals can be considered as independent. As shown in Table 1, the p-value of the LB test is greater than 0.05. Therefore, we cannot reject the null hypothesis that the residual are independent. Moreover, we used QQ plots to check that the residuals have approximately a student-t distribution.[5]

It is important to check the forecasting accuracy of three selected models using out-of-sample data between July 2012 and June 2013. The overall forecasting performances of SARIMA, ARIMA and ARIMA-GARCH are shown in Table 3 of Panel A. Along with these three modes, the forecasts with minimal absolute forecast error are bold-faced. Except July, August and October of 2012, and March of 2013, SARIMA is the best model for all other monthly forecasts. Moreover, accurate forecast is very important for business planning. A variety of measures have been used to assess forecasting accuracy on studies of international tourism demand The mean absolute percentage error (MAPE) and the root mean squared percentage error (RMSPE) are amongst the most commonly used measures of error magnitude. Hence, we used these measures to assess forecasting accuracy (out-of-sample) in this study.[6] MAPE and RMSE measures were used to compare the accuracy of the forecasts obtained from SARIMA, ARIMA and ARIMA-GARCH models were given in Table 3 of Panel B. Numerical results obtained from the best model are bold-faced. From the Panel B of Table 3, we found that the MAPE of ARIMA, SARIMA and ARIMA-GARCH

[5] The QQ plots of error term are available on request.

[6] The MAPE and RMSE are defined as:

$$\text{MAPE} = \frac{1}{K} \sum_{t=N+1}^{N+K} \left| \frac{\hat{y}_t - y_t}{y_t} \right| \times 100, \text{RSME} = \sqrt{\frac{1}{K} \sum_{t=N+1}^{N+K} \left(\hat{y}_t - y_t\right)^2}.$$

Table 2 Results of best-fit ARIMA, ASRIMA and ARIMA-GARCH models for monthly tourist arrivals

Dependence variable: ln(tourist arrival)					
ARIMA$(p, 1, q)$		SARIMA$(p, 1, q) \times (P, 1, Q)_{12}$		ARIMA$(p, 1, q)$-GARCH(P, Q)	
Explanatory variables	Coefficients	Explanatory variables	Coefficients	Explanatory variables	Coefficients
ϕ_1	−0.5143***	ϕ_1	0.3933***	ϕ_{12}	0.9781***
	(0.0453)		−0.082		−0.0148
ϕ_2	−0.0782**	Θ_{12}	−0.4212***	θ_1	−0.6150***
	(0.0375)		−0.0537		−0.0491
ϕ_{12}	0.6826***	θ_1	−0.8637***	θ_{12}	−0.7315***
	(0.0490)		−0.0435		−0.0425
ϕ_{13}	0.4315***	DoF	2.7755***	θ_{13}	0.4389***
	(0.0455)		−0.5112		−0.0541
ϕ_{15}	0.0834**	Variance	0.0704***	c	0.0001*
	(0.0351)		−0.0046		0
ϕ_{24}	0.2035***			α_1	0.1275**
	(0.0424)				−0.0586
θ_2	−0.3992***			β_1	0.8375***
	(0.0476)				−0.0577
θ_6	−0.1833***			DoF	4.0651***
	(0.0504)				−0.0586
θ_{12}	−0.2380***				
	(0.0581)				
DoF	2.5880***				
	(0.5461)				
Variance	0.0049**				
	(0.0019)				
AIC	−789.5928	AIC	−787.6734	AIC	−782.3853
BIC	−746.4117	BIC	−766.0829	BIC	−753.5979
LB(10)	0.4274	LB(10)	0.2744	LB(10)	0.6079

Note $*$, $**$, and $***$ indicate that the test is significant at the 0.10, 0.05, and 0.01 significance level, respectively. Stand errors of the coefficients are in parentheses

models were 4.32, 3.75 and 4.77, respectively. Moreover, the RMSE measures were 545.36 (ARIMA), 485.67 (SARIMA) and 583.79 (ARIMA-GARCH). These results indicated that the SARIMA model outperforms the ARIMA and ARIMA-GARCH models in terms of forecasting accuracy.

Table 3 Forecasting performance and comparison of prediction error of three models

Periods	Actual	ARIMA model		SARIMA model		ARIMA-GARCH model	
Panel A: forecasting performance							
		Forecast	Forecast error (%)	Forecast	Forecast error (%)	Forecast	Forecast error (%)
Jul 2012	11056.3	11685.45	−5.69	11619.04	−5.09	11600.39	−4.92
Aug 2012	11282.9	11487.33	−1.81	11669.10	−3.42	11698.24	−3.68
Sep 2012	10953.7	11064.53	−1.01	11039.11	−0.78	11126.60	−1.58
Oct 2012	11206.5	11724.08	−4.62	11744.03	−4.80	11799.28	−5.29
Nov 2012	10831.1	11189.15	−3.31	11082.25	−2.32	11201.40	−3.42
Dec 2012	11186.9	11667.94	−4.30	11569.18	−3.42	11673.70	−4.35
Jan 2013	10798.7	10880.20	−0.75	10812.95	−0.13	11019.19	−2.04
Feb 2013	9430.8	10200.58	−8.16	9900.10	−4.98	10157.77	−7.71
Mar 2013	11853.9	11695.97	1.33	11446.05	3.44	11631.75	1.87
Apr 2013	10813.4	11926.28	−10.29	11923.90	−10.27	12066.57	−11.59
May 2013	10652.5	11247.49	−5.59	11074.50	−3.96	11290.24	−5.99
Jun 2013	10562.3	11086.68	−4.96	10809.71	−2.34	11074.43	−4.85
Panel B: forecasting comparison							
		ARIMA model		SARIMA model		ARIMA-GARCH model	
MAPE		4.32		3.75		4.77	
RSME		545.36		485.67		583.79	

Note The forecasted absolute tourist arrivals have been transformed from the forecasted values of ln (monthly tourist arrivals). The actual and forecasted values are stated in thousands

4 Forecast Using the Belief Function Approach

In Sect. 3, it was shown that SARIMA is the best fitting model. Hence, this model was chosen to compute a predictive belief function for the tourism arrival data, using the methodology introduced by Kanjanatarakul et al. [15]. The best fitting model is:

$$\left(1-\phi_1 L^1\right)\left(1-\Theta_{12} L^{12}\right)\nabla^1\nabla_{12}^1 \ln y_t = \left(1+\theta_1 L^1\right)\varepsilon_t$$

Introducing the notation $x_t = \nabla^1\nabla_{12}^1 \ln y_t$, we transform the general expression of this model to

$$x_t = \phi_1 x_{t-1} + \Theta_{12}\left(x_{t-12} - \phi_1 x_{t-13}\right) + \theta_1 \varepsilon_{t-1} + \varepsilon_t$$

where ε_t has a student-t distribution $\varepsilon_t \sim iid \;\; t_\lambda(0, \sigma^2)$, with λ degrees of freedom and variance σ. To keep the computation simple, we assume λ to be known.

The likelihood function is thus

$$L_{x_t}(\beta) = \prod_{t=13}^{T} f(x_t|\,\varepsilon_t;\beta) = \prod_{t=13}^{T} f(x_t|\,\varepsilon_t;\phi_1,\theta_1,\Theta_{12},\sigma^2)$$

The contour function of the parameter vector $\beta = \left(\phi_1, \theta_1, \Theta_{12}, \sigma^2\right)$ is

$$pl_{y_t}(\beta) = \frac{L_{x_t}(\beta)}{L_{x_t}\left(\hat{\beta}\right)} = \frac{L_{x_t}\left(\phi_1,\theta_1,\Theta_{12},\sigma^2\right)}{L_{x_t}\left(\hat{\phi}_1,\hat{\theta}_1,\hat{\Theta}_{12},\hat{\sigma}^2\right)}$$

Then, the future data in $t+i$ can be written as:

$$\begin{aligned} x_{t+\mathrm{i}} &= \psi(\phi_1,\theta_1,\Theta_{12},\sigma) \\ &= \phi_1 x_{t+i-1} + \Theta_{12}\left(x_{t+i-12} - \phi_1 x_{t+i-13}\right) + \theta_1\varepsilon_{t+i-1} + \sigma F_{\lambda}^{-1}(v) \end{aligned}$$

and

$$\varepsilon_{t+i-1} = x_{t+i-1} - \phi_1 x_{t+i-2} - \Theta_{12}(x_{t+i-13} - \phi_1 x_{t+i-14}) - \theta_1\varepsilon_{t+i-2}$$

Hence

$$\begin{aligned} x_{t+\mathrm{i}} &= \psi(\phi_1,\theta_1,\Theta_{12},\sigma) \\ &= \phi_1 x_{t+i-1} + \Theta_{12}\left(x_{t+i-12} - \phi_1 x_{t+i-13}\right) \\ &\quad + \theta_1\left(x_{t+i-1} - \phi_1 x_{t+i-2} - \Theta_{12}(x_{t+i-13} - \phi_1 x_{t+i-14}\right) - \theta_1\varepsilon_{t+i-2}) + \sigma F_{\lambda}^{-1}(v) \end{aligned}$$

where $F_{\lambda}^{-1}(v)$ is inverse cdf of $t_{\lambda}(0,1)$ and $v \sim iid\ \ U(0,1)$. For any (w,v) in $|0,1| \times \mathbb{R}$, the set $\Psi(\Gamma_{x+i}(\omega), v)$ is the interval $\left[x_{t+i}^{L}(\omega,v), x_{t+i}^{U}(\omega,v)\right]$ defined by the following lower and upper bounds:

$$\begin{aligned} x_{t+\mathrm{i}}^{L} &= \min_{\{\beta|pl_{x_{t+i-1}}(\beta)\geq w\}} \psi(\phi_1,\theta_1,\Theta_{12},\sigma,v) \\ x_{t+\mathrm{i}}^{U} &= \max_{\{\beta|pl_{x_{t+i-1}}(\beta)\geq w\}} \psi(\phi_1,\theta_1,\Theta_{12},\sigma,v) \end{aligned}$$

with $w \sim iid\ U[0,1]$. We used a constrained nonlinear optimization algorithm to compute the intervals $[x_{t+i}^{L}, x_{t+i}^{U}]$. Using Monte Carlo simulation, simulating independently N pairs $(\omega_n, v_n), n = 1,\ldots,N$, we obtained N intervals $[x_{t+i}^{L}, x_{t+i}^{U}]$. Then we transfer the x to y. we we obtained N intervals $[y_{t+i}^{L}, y_{t+i}^{U}]$. For any $B \subset \mathbb{R}$, the quantities $Bel(B)$ and $Pl(B)$ can then be estimated as

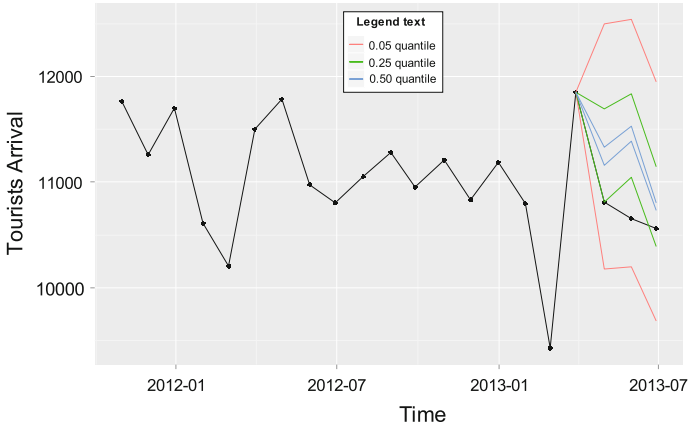


Fig. 3 The predictions for the number of tourist arrivals and α-quantile intervals with $\alpha \in (0.05, 0.25, 0.5)$

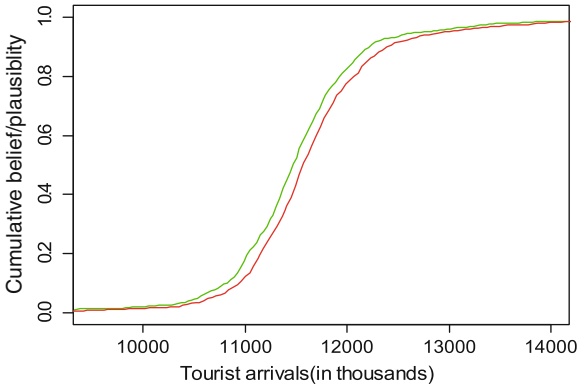


Fig. 4 Lower and upper cdfs for the number of Chinese total tourist arrivals in June 2013

$$\widehat{Bel}(B) = \frac{1}{N}\#\left\{n \in \{1, \cdots, N\} \mid \left[y_{t+i}^{L}, y_{t+i}^{U}\right] \subseteq B\right\}$$
$$\widehat{Pl}(B) = \frac{1}{N}\#\left\{n \in \{1, \cdots, N\} \mid \left[y_{t+i}^{L}, y_{t+i}^{U}\right] \cap B \neq \emptyset\right\}$$

When $\mathbb{R}$ is the real line, the lower and upper predictive cdfs of Y are defined as follow:

$$F^{L}(y) = \widehat{Bel}((-\infty, y])$$
$$F^{U}(y) = \widehat{Pl}((-\infty, y])$$

For any $y \subset \mathbb{R}$.

The approach described above was applied to China Inbound Tourism Demand data. Figure 3 shows the predictions for the number of tourist arrivals in the periods April, May and June 2013. We also plot α-quantile intervals with $\alpha \in (0.05, 0.25, 0.5)$ in Fig. 3. From the Fig. 3, we find that the observed data are contained in the 0.05-quantile intervals. The lower and upper cdfs F^L and F^U are plotted in Fig. 4. Figure 4 shows show additional information about the predictive belief function concerning the forecasted number of tourist arrival for June 2013. Form the figure, we know that when the tourist arrivals for January are less than 11,600,000, possibility of belief is 60.2 % and possibility of plausibility is 53.1 %. Moreover, under the conditional belief and plausibility of 90 %, the number tourist arrivals in China for June 2013 less than [12, 242, 939, 12, 391, 306]. The empirical result show that time series models combined with belief functions is a computationally simple and effective method.

5 Conclusions

The purpose of this paper was to demonstrate the application of time series models combined with belief functions to forecast the demand for China international tourism. The analysis was conducted using monthly inflow of international tourist arrivals covering the period from January 1991 to June 2013. The method of this study was divided into two steps. In the first step, three time series model were considered, namely, ARIMA, SARIMA and ARIMA-GARCH models. These three models were used to forecast tourism demand based on observed data. The MAPE and RMSE measures were used to compare the forecast accuracies and the best fitting model was selected. In the second step, the best fitting model was combined with belief function to forecast Chinese international tourism demand, taking into account both estimation uncertainty and random variability. The empirical result show that time series models combined with belief functions is a computationally simple and effective method.

Modeling uncertainty is a key issue in forecasting. In the tourism area, forecasts are used by governments, airline companies and operators to design tourism policies and they should include a quantification of uncertainties. In this paper, we have shown that the predictive belief function approach introduced by Kanjanatarakul et al. [15] is a computationally simple and effective method that can be used with time series models.

References

1. Chang, C.L., Sriboonchitta, S., Wiboonpongse, A.: Modelling and forecasting tourism from East Asia to Thailand under temporal and spatial aggregation. Math. Comput. Simul. **79**, 1730–1744 (2009)

2. Cho, V.: A comparison of three different approaches to tourist arrival forecasting. Tour. Manag. **24**, 323–330 (2003)
3. Chu, F.L.: A fractionally integrated autoregressive moving average approach to forecasting tourism demand. Tour. Manag. **29**, 79–88 (2008)
4. Chu, F.L.: Forecasting tourism demand with ARMA-based methods. Tour. Manag. **30**, 740–751 (2009)
5. Claveria, C., Torra, S.: Forecasting tourism demand to Catalonia: neural networks vs. time series models. Econ. Model. **36**, 220–228 (2014)
6. Goh, C., Law, R.: Modelling and forecasting tourism demand for arrivals with stochastic nonstationarity seasonality and intervention. Tour. Manag. **23**, 499–510 (2002)
7. Lim, C., McAleer, M.: Time series forecasts of international travel demand for Australia. Tour. Manag. **23**, 389–396 (2002)
8. Song, H., Witt, S.F.: Forecasting international tourist flows to Macau. Tour. Manag. **27**, 214–224 (2006)
9. Turner, L.W., Witt, S.F.: Forecasting tourism using univariate and multivariate structural time series models. Tour. Econ. **7**, 135–147 (2001)
10. UNWTO: UNWTO tourism highlights 2013th edn. UNWTO, Madrid (2013)
11. Lim, C., Pan, G.W.: Inbound tourism developments and patterns in China. Math. Comput. Simul. **68**, 499–507 (2005)
12. Bollerslev, T.: Generalized autoregressive conditional heteroskedasticity. J. Econom. **31**(3), 307–327 (1986)
13. Denoeux, T.: Likelihood-based belief function: justification and some extensions to low-quality data. Int. J. Approx. Reason. (2014)
14. Abdallah, N.B., Voyneau, N.M., Denoeux, T.: Combining statistical and expert evidence using belief functions: application to centennial sea level estimation taking into account climate change. Int. J. Approx. Reason. **55**, 341–354 (2014)
15. Kanjanatarakul, O., Sriboonchitta, S., Denoeux, T.: Forecasting using belief functions: an application to marketing econometrics. Int. J. Approx. Reason. **55**(5), 1113–1128 (2014)
16. Nguyen, H.T.: An Introduction to Random Sets. Chapman and Hall/CRC Press, Boca Raton (2006)
17. Jarque, C., Bera, A.: A test for normality of observations and regression residuals. Int Statis Rev. **55**, 163–172 (1987)
18. Coshall, J.T.: Combining volatility and smoothing forecasts of UK demand for international tourism. Tour. Manag. **30**, 495–511 (2009)

Forecasting Tourist Arrivals to Thailand Using Belief Functions

Nyo Min, Jirakom Sirisrisakulchai and Songsak Sriboonchitta

Abstract This paper applies the belief function approach to statistical forecasting of tourist arrivals to Thailand. Seasonal autoregressive integrated moving average (SARIMA) model was applied to forecast the tourists arrivals to Thailand using the time series data during the period of 1997–2013. To quantify the uncertainty of statistical forecasting, we used the method proposed by Kanjanatarakul et al. [5]. We utilized the statistical model, SARIMA to obtain parameter space which was constructed from the normalized likelihood given the observed data. Then, we rewrote the forecasting equation as a function of parameters and an auxiliary random variable with known distribution not depending on the parameters in prediction stage. Combining beliefs about parameters and auxiliary random variable gave us a predictive belief function for tourist arrivals. The finding supports the statement that the method can be used with any parametric model such as linear regression and time series models including SARIMA.

1 Introduction

Travel and Tourism play a very important role for growth of nation not only for Thailand but for the countries around the world. Tourism affects the countries with both direct and indirect contributions to their GDPs. According to the World Travel and Tourism Council (WTTC), Thailand has growths in tourism in terms of GDP, employment in tourism, visitor exports, and investment in tourism. The tourism contributed 9.0 %($THB1,074.0bn$) of total GDP in 2013 directly and 20.2 %($THB2,401.1bn$) of GDP in total including indirect impact in 2013. The WTTC forecasted that the contribution of tourism to rise by 6.7 % pa, from 2014–2024, to $THB2,046.7bn$ (10.4 % of total GDP) in 2024 [12]. This calculation reveals that the forecasting in tourists' arrival is essential analytical tool for the policy makers and other stakeholders for the policy, planning, and

N. Min (✉) · J. Sirisrisakulchai · S. Sriboonchitta
Faculty of Economics, Chiang Mai University, 50200 Chiang Mai, Thailand
e-mail: nyo.min@gmail.com

V.-N. Huynh et al. (eds.), *Econometrics of Risk*, Studies in Computational Intelligence 583, DOI 10.1007/978-3-319-13449-9_24

development. Many researchers have tried to get more and more reliable forecast interval of tourists' arrival, but all could have concluded for the better model by comparing with others. As result, the forecasting in tourism is an open issue for researchers to work on for the best model.

In the paper, we intend to handle highly uncertainty situations to become the reliable forecast by applying the Dempster-Shafer Theory [12]. The motivation for using Dempster-Shafer Theory is about its characteristics relating to the uncertainty. The theory is a highly developed among non-traditional theories linking to uncertainties. Although it is a new concept, it links with traditional uncertainty theories and set theory. In practice, this theory has been approved in the engineering areas. Dempster's rule of combination can handle different types of evidence, and can deal with many conflicts that can be incorporated when combining the multi sources of information. One of the most important features of Dempster-Shafer (DS) theory is that the model is designed to cope with varying levels of precision due to the information and no further assumptions are needed to represent the information. It also allows for the direct representation of uncertainty of system responses where an imprecise input can be characterized by a set or an interval and the resulting output is a set or an interval.

In this paper we analyzed nature of data set, tourist arrivals to Thailand between 1997 January and 2013 September to know whether the data are stationary or not. Due to small break points the data set is non-stationary. Then, we took log return to make the data set stationary. As GARCH family had been famous in forecasting tourists' arrival, we checked if the GARCH family is applicable to our work. The results showed that the GARCH family is not relevant with our work and finally we chose the best fit model among potential SARIMA models. To select the best fit SARIMA model we tested the data not only with Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) values, but also with root mean square error (RMSE) and mean absolute percentage error (MAPE). After getting the statistical model, we applied belief functions such as likelihood-based belief function, plausibility function, and Dempster's rule of combination on this statistical function to obtain reliable forecasting interval. We favored the belief function approach to quantify the uncertainty pertaining to statistical forecasts for the following reasons. First, the method of forecasting and inference using the belief function approach generalizes Bayesian inference. However, in the belief function approach, we did not have to provide the information on prior distribution. Second, this approach could be used in the situations where the data are scarce and imprecise. Finally, we used this approach further by considering the combination of the statistical evidence and the expert opinion, which is very important in the policy analysis. The advantages of the belief function approach were discussed at length in Kanjanatarakul et al. [5] and Denoeux [3]. The reader is referred to these references for further details and discussion about this method of inference.

The paper is organized as follows. In the next section, we provide a brief review on some definitions of belief function, the literature review on related model and concepts of the paper, and methodology. Section 3 gives details of data, application that we used in this paper, and empirical results we obtained, and Sect. 4: provides

discussion for how to combine or verify the results generated from historical data with experts opinion. Our conclusions are drawn in Sect. 5.

2 Definitions, Literature Reviews and Methodology

2.1 Basics of Belief Functions

We briefly explain the theory of belief function in this section. We first start with the belief function in a finite domain. Suppose that θ is a variable taking values in a finite domain Θ, called the frame of discernment. We can assign a mass function $m : 2^{\Theta} \longrightarrow [0, 1]$, such that $m(\emptyset) = 0$ and $\sum_{A \in \Theta} m(A) = 1$, to represent the uncertain evidence about θ. Any subset A of Θ, such that $m(A) > 0$, is called a focal set of m. The interpretation of $m(A)$ is a degree of belief attached to the proposition $\theta \in A$ ([11]). Shafer used the following framework to explain the meaning of the degrees of belief. Let $\Omega = \{\omega_1, \ldots, \omega_n\}$ be a set of codes, one of which is selected at random. We know exactly the list of possible codes and the probability p_i attached to each code. The way using code to decode the encoded message gives us a message of the form $\theta \in A_i$ for some $A_i \subseteq \Theta$. Thus, the probability that the original message $\theta \in A$ can be computed by $m(A) = \sum_{(1 \leq i \leq n : A_i = A)} P_i$. This setting consists of a probability measure P on a set Ω and a multi-valued mapping $\Gamma : \Omega \longrightarrow 2^{\Theta} \backslash \{\emptyset\}$ such that $A_i = \Gamma(\omega_i)$ for all i. From a mathematical point of view, the triple (Ω, P, Γ) defines a finite random set [7]. The interpretation of a random set usually means a random experiment in which the outcome is a set. However, the interpretation of mass function is different. Here $m(A)$ can be viewed as the chance of the evidence meaning that $\theta \in A$ [9].

We formally define belief and plausibility function corresponding to a mass function m for Ω as a function: $2^{\Theta} \longrightarrow [0, 1]$ such that

$$Bel(A) = P(\omega \in \Omega \mid \Gamma(\omega) \in A) = \sum_{B \in A} m(B), \tag{1}$$

$$Pl(A) = P(\omega \in \Omega \mid \Gamma(\omega) \cap A = \emptyset) = \sum_{B \cap A \neq \emptyset} m(B), \tag{2}$$

for all $A \in \Theta$.

We note here that $Pl(A) \geq Bel(A)$, for all $A \subseteq \Theta$, and $Pl(A) = 1 - Bel(A^c)$, where A^c is the complement of A in Θ. The function $pl : \Theta \longrightarrow [0, 1]$ such that $pl(\Theta) = Pl(\{\theta\})$ for all $\theta \in \Theta$ is called the contour function associated to m. If all focal set elements of *Bel* are singleton subsets, then $Pl(A) = Bel(A)$, for all $A \subseteq \Theta$, a probability function. In the infinite domain of Θ, there may not be a mass function associated with a completely monotone function, thus we have to define a belief function axiomatically from its properties. Let (Θ, B) be a measurable

space, where is a sigma-field, and which is a non-empty subset of 2^{Θ} closed under complementation and countable union. A belief function on (Θ, B) is a function Bel: $2^{\Theta} \longrightarrow [0, 1]$ satisfying the following conditions,

1. $Bel(\emptyset) = 0, Bel(\Theta) = 1$.
2. For any $k \geq 2$ and any collection $A_1, \ldots, A_k$ of subset of Θ,

$$Bel(\bigcup_{i=1}^{k} A_i) \geq \sum_{\emptyset \neq I \subseteq 1,\ldots,k} (-1)^{|I|+1} Bel(\bigcap_{i \in I} A_i).$$

where $| I |$ is cardinality of the set I.

2.2 Likelihood-Based Belief Function

Let x be the observed data of the random variable X. The random variable X has a probability density function (pdf) $p(x; \theta)$, where $\theta \in \Theta$ is an unknown parameter. [11] proposed a method to represent the statistical evidence of x on θ by using belief function. This belief function can be derived from the Likelihood and Least Commitment Principles (see, Denoeux [3] and Kanjanatarakul et al. [5]). According to the Likelihood Principle, the information about θ is supposed to be represented by the likelihood function, $L(\theta; x) = p(x, \theta)$. In statistics, the likelihood ratio is meant to be a relative plausibility, which can be compared with the likelihood ratio in the belief function framework as follows (see, Denoeux [3] and Kanjanatarakul et al. [5]):

$$\frac{pl(\theta_1; x)}{pl(\theta_2; x)} = \frac{L(\theta_1, x)}{L(\theta_2; x)}, \tag{3}$$

for all $(\theta_1, \theta_2) \in \Theta^2$ or, equivalently,

$$pl(\theta; x) = cL(\theta; x), \tag{4}$$

for all $\theta \in \Theta$ and some positive constants, c. According to the Least Commitment Principle, we can get the highest possible value of constant c, i.e., defining $pl(\theta; x)$ as the relative likelihood:

$$pl(\theta; x) = \frac{L(\theta; x)}{sup_{\theta'; x} L(\theta'; x)}, \tag{5}$$

and we can represent information about θ by the belief function $Bel(\cdot; x)$ via the contour function $pl(\cdot; x)$, i.e.,

$$Pl(A; x) = \sup_{\theta \in A} pl(\theta; x) = \frac{\sup_{\theta \in A} L(\theta; x)}{\sup_{\theta' \in \Theta} L(\theta'; x)}, \tag{6}$$

for all $A \subseteq \Theta$. Note that the likelihood function has to be bounded in order to carry out the analysis, which is usually the case for most parametric model. The focal sets of $Bel(\cdot; x)$ are the level sets of $pl(\theta; x)$ defined as follows:

$$\Gamma_x(\omega) = \{\theta \in \Theta \mid pl(\theta; x) \geq \omega\}, \tag{7}$$

for $\omega \in [0, 1]$. These sets are called plausibility regions. The corresponding belief function is equivalent to the random set induced by the Lebesgue measure λ on $[0, 1]$ and the multi-valued mapping $\Gamma_x : [0, 1] \longrightarrow 2^{\Theta}$. The belief and plausibility function can be expressed in these following equations:

$$Bel_x(A) = \lambda(\omega \in [0, 1] \mid \Gamma_x(\omega) \subseteq A), \tag{8}$$

$$Pl_x(A) = \lambda(\omega \in [0, 1] \mid \Gamma_x(\omega) \bigcap A \neq \emptyset), \tag{9}$$

for all $A \subseteq \Theta$. We remark that the maximum likelihood estimation can be represented as the value of highest plausibility.

2.3 Forecasting Using Belief Functions

In this section, we describe how to use the belief function defined in the previous section for prediction. Let X have a probability density function $f(x; \theta)$, for $\theta \in \Theta$. Suppose that we have observed $X = x$. With some evidences given about θ through $Bel(\Theta; x)$, we can predict the future value of a random variable Y whose probability density function $g(y; \theta)$ also depends on the same θ. Kanjanatarakul et al. [5] proposed a model for prediction using belief function theory. In this model, Y can be written in a function of the parameter θ, observed data $X = x$ and an unobserved auxiliary variable $u \in U$ with known probability measure μ, but not depending on θ:

$$Y = \varphi(\theta; u), \tag{10}$$

From Eq. (10), we can derive the multi-valued mapping $\Gamma_x : [0, 1] \longrightarrow 2^{\Theta}$ with $\varphi(\theta; u)$ to get a new multi-valued mapping $\Gamma_x' : [0, 1] \times U \longrightarrow 2^{Y}$ defined as

$$(\omega, u) \longrightarrow \varphi(\Gamma_x(\omega), u), \tag{11}$$

Then we can construct the predictive belief Bel_x and plausibility Pl_x functions on the sample space $\mathscr{Y}$ of Y as follows,

$$Bel_x(A) = (\lambda \otimes \mu)((\omega, u) \in [0, 1] \mid \varphi(\Gamma_x(\omega, u) \subseteq A), \tag{12}$$

$$Pl_x(A) = (\lambda \otimes \mu)((\omega, u) \in [0, 1] \mid \varphi(\Gamma_x(\omega, u) \bigcap A \neq \emptyset), \tag{13}$$

for all measurable subset A of $\mathscr{Y}$, where $\lambda \otimes \mu$ is the product measure on $[0, 1] \times U$.

2.4 Review of the Seasonal ARIMA Model

A time series $Z_t \mid t = 1, 2, \ldots, k$ is generated by a SARIMA $(p, d, q)(P, D, Q)_s$ process with mean of the Box—Jenkins's model if

$$\Phi(B)\Phi(B^s)(1 - B)^d(1 - B^s)^D(Z_t - \mu) = \theta(B)\theta(B^s)a_t, \tag{14}$$

where p, d, q, P, D and Q are integers, s is periodicity, $\Phi(B)$, $\Phi(B^s)$, $\theta(B)$ and $\theta(B^s)$ are polynomials in B of degrees p, q, P and Q; B is the backward shift operator, d is the number of regular differences; D is the number of seasonal differences, and Z_t denotes observed value of time series data, $t = 1, 2, \ldots, k$.

Generally, observations for SARIMA should be at least 50 and preferably 100 observations or more. Due to uncertainty and rapid changes, forecasting future situations should be for a short time-span by using little data, but it is difficult to verify that the data have a normal distribution [13].

3 Application to Tourist Arrivals to Thailand

3.1 Data

This paper used the monthly tourist arrivals to Thailand for model estimation and evaluation. The data range covered from January 1997 to September 2013. All data that we used for the paper were taken from EcoWin data base and the paper intentionally ignored the seasonality adjusted data to prove the effectiveness of the model and belief functions. To overcome seasonality, the paper used SARIMA model with seasonal difference on log return and chose the best fit model by comparing values of Akaike's Information Criterion (AIC), Schwarz's Bayesian information criterion (BIC), mean absolute percentage error (MAPE) and root mean squared error (RMSE). According to the data, the tourist arrivals varied all the time and fluctuated up and down. The data revealed that tourist arrivals to Thailand were increasing as a trend but there was no significant structural change (Fig. 1).

201 observations of tourist arrival to Thailand between January 1997 and September 2013 showed us two potential breaks in May 2003 and May 2010.

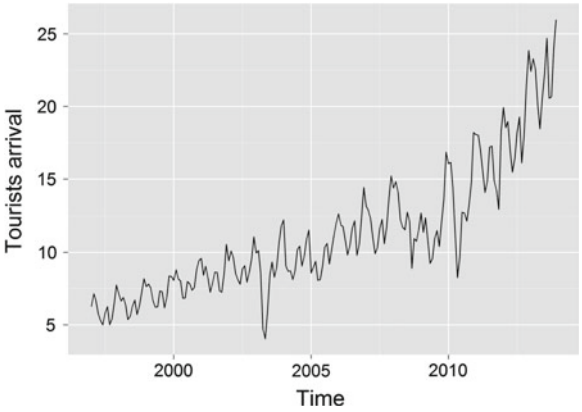


Fig. 1 Tourist arrivals to Thailand, 1997 to 2013

In 2003 there was Severe Acute Respiratory Syndrome (SARS) in Asia region and it affected on tourism of Thailand. In May 2010, Bangkok dangerous street fights also affected tourism in Thailand.

3.2 SARIMA Models

Due to the breaks and slightly structural change in the data observed, the sample size for the paper was non-stationary. Therefore, we took log return to make the data stationary. The Generalized Autoregressive Conditional Heteroskedasticity (GARCH) family had been prominent in forecasting field and many researchers including [1, 2, 4, 6] applied GARCH models to forecast the tourists arrival, so we tested to know whether or not GARCH models are applicable to our data observed. All tests led us to reject predicting the data with GARCH family members that we checked. Therefore, we emphasized our work to some different types of SARIMA models, ARIMA models and seasonal patterns displayed by seasonal ARIMA model. When we sought for the best model, we tailored data, from raw data, to log return, and to seasonal log return, etc (Fig. 2).

Table 1 presented four models for the tourists' arrival to Thailand, tested by EViews 7 Econometric software and these models were chosen in accordance with ACF and PACF results. Among these four models, Model 1 has lowest values of AIC and SBC. Therefore we assumed that Model 1 is the best fitting model to describe the tourist arrivals to Thailand during the period we observed. To make sure that Model 1 is the best fitted model for our work, we checked residual diagnostic tests, and results showed favor for the model we adopted. For the test result for ACF and PACF, we got all Q statistics were favorable and significant. When we tested for LM test the

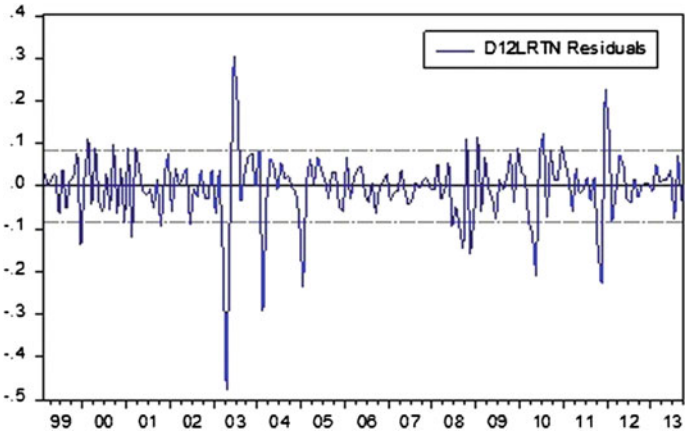


Fig. 2 Errors distribution of selected model, $(12, 0, 12)(0, 1, 0)_{12}$

Table 1 Estimates of selected ARIMA models for tourist arrivals to Thailand with normal distribution

Variable	Coefficient	t-Statistics	AIC/SBC	LM(SC)
Model 1				
C	0.0009	0.5990		
AR(4)	−0.1975	−2.4084		
AR(12)	−0.1788	−2.6611	$AIC = -2.1037$	F-statistics = 1.0893
MA(12)	−0.9399	−41.5408	$SC = -2.0317$	$P = 0.3536$
Model 2				
C	0.0067	3.1900		
AR(6)	−0.2559	−3.3704		
MA(2)	−0.4638	−7.0800	$AIC = -1.5616$	F-statistics = 3.3334
SMA(3)	−0.3888	−5.3617	$SC = -1.4942$	$P = 0.0000$
Model 3				
C	0.0065	3.3405		
AR(6)	−0.2748	−3.8548		
AR(9)	−0.1690	−2.3934		
SMA(3)	−0.3303	−4.6118	$AIC = -1.5896$	F-statistics = 3.0782
MA(2)	−0.4746	−7.0558	$SC = -1.5045$	P=0.0000
Model 4				
C	0.0065	2.2366		
AR(6)	−0.2814	−3.9902		
AR(9)	−0.2626	−3.6626	$AIC = -1.4983$	F-statistics = 3.8674
MA(2)	−0.4533	−6.6844	$SC = -1.4302$	$P = 0.0000$

Note: AIC and SBC are the Akaike Information Criterion and Schwarz Bayesian Criterion, respectively. LM(SC) refers to the Lagrange multiplier test for serial correlation

results showed that F statistics was 2.524235 with probability 0.0045, but normality test showed that the error doesn't distribute normally. Therefore we continued our test with student's t distribution. The results from Kolmogorov-Smirnov test showed that the error distributed as t distribution. Thus, we used the model 1 with the t distribution for random error as the forecasting model in the next section (Table 2).

As traditional methods, we checked all error terms of different models with RMSE and MAPE. RMSE stands for the root-mean-square deviation (RMSD) or root-mean-square error (RMSE). It measures the differences between values predicted by a model or an estimator and the values actually observed. Generally, the RMSD represents the sample standard deviation of the differences between predicted values and observed values. MAPE stands for the mean absolute percentage error (MAPE), also known as mean absolute percentage deviation (MAPD). It measures accuracy of a method for constructing fitted time series values in statistics, specifically in trend estimation. According to the RMSE and MAPE values, the model (1) that we had chosen, had lowest error values, RMSE 0.0826 and MAPE −0.0028 and it means the model we had chosen was the best fit model among the models we had compared (Table 3).

The tourist arrivals to Thailand between 1997 and 2013 had two significant breaks and one potential structural change. Therefore we tested to know if the whole sample is good enough for the prediction or two sub-samples due to deterministic breaks provide better prediction. We used Chow Breakpoint Test on break point of May 2003 and found that there was no break at specific break point with F-statistic 1.4438 and Probability 0.2217. For a second time, we checked on break point, May 2010 and also found that there was no break at specific break point with F-statistic 1.818735 and Probability 0.1275.

Saleh et al. [8] stated in their conference paper that exogenous shocks such as the September 11 2001, SARS outbreak in 2003, war in Iraq in 2003, global recession (in early 2000s) and Asian financial crisis (during 1997–1998) had only temporary effect on the number of arrivals to Thailand from the ten countries [8]. They also commented that the tourism industry recovered strongly in a short period of time. In theory, the effects of breaks and structural changes are likely to come up again

Table 2 Estimates of selected SARIMA model for tourist arrivals to Thailand with Student's t distribution

Variable	Coefficient	SE
Model 1		
C	0.0004	0.0009
D12LRTN(-4)	0.0126	0.0597
D12LRTN(-12)	0.1105	0.0684
MA(12)	−0.8913	0.0391

Table 3 RMSE and MAPE values of the models compared

Method	*Model*1	*Model*2	*Model*3	*Model*4
RMSE	0.0826	0.1086	0.1065	0.1120
MAPE	−0.0028	$-3.40057E-05$	0.0009	0.0008

in the future forecast horizon, but fortunately, our work is only forecasting a period head. Therefore we preceded our empirical work with the selected model to the period we focused.

3.3 Approach with Belief Functions

Once we had chosen the best fitted statistical model, we continued to estimation and prediction by using belief functions. In first stage, we applied observed data with normalized likelihood in order for building belief function on parameter space. Then, we predicted forecast for nearest future on observed data. In this prediction stage, we rewrote our forecast model as the function of parameters and a random auxiliary variable in accordance with Dempster's statistical inference.

In order to implement our empirical work, we stated the SARIMA model as follows,

$$(1 - \beta_1 B^4 - \beta_2 B^{12}) \bigtriangledown_{12} Y_t = (1 + \gamma L_{12} \varepsilon_t). \tag{15}$$

This model is equivalent to,

$$\bigtriangledown_{12} Y_t = \alpha + \beta_1 \bigtriangledown_{12} Y_{t-4} + \beta_2 \bigtriangledown_{12} Y_{t-12} + \gamma \varepsilon_{t-12} + \varepsilon_t, \tag{16}$$

where Y_t = log return of tourist arrivals to Thailand α = Constant term, β_s = coefficients for AR terms, where $s = 1, 2\gamma$ = coefficient for MA terms, ε = residual. From the adopted model, we can get

$$\varepsilon_t = \bigtriangledown_{12} Y_t - \alpha - \beta_1 \bigtriangledown_{12} Y_{t-4} - \beta_2 \bigtriangledown_{12} Y_{t-12} - \gamma \varepsilon_{t-12}. \tag{17}$$

Then, we wrote likelihood function as follows,

$$L(\alpha_s; v; \theta; Y_t) = \prod_{t=1}^{T} \frac{\Gamma[\frac{(v+1)}{2}]}{\sqrt{(v-2)\pi}\,\Gamma[\frac{v}{2}](\frac{(1+\varepsilon_t^2)}{v})^{\frac{(v+1)}{2}}}, \tag{18}$$

where v = degree of freedom for $v > 2$, and Γ = Gamma function.

By using the value of likelihood function, we developed 'belief function' on our parameter space. This belief function is called the likelihood-based belief function on parameter space Θ induced by observed data. The belief function on $\theta = (\alpha, \beta_1, \beta_2, \gamma)$ is defined by the contour function,

$$pl(\theta; Y_t) = \frac{L(\theta; Y_t)}{L(\hat{\theta}; Y_t)}. \tag{19}$$

3.4 Forecasting Using Belief Functions

Forecasting is the use of historic data to predict future trends and no one can guarantee the forecasting is accurate because it connects to uncertainty. There are two types of forecasting methods, qualitative and quantitative. In this paper we applied the quantitative method.

Based on parameters, α, β_s and γ, we predicted forecast value for the tourist arrivals for a period ahead. We had the statistical model for the paper,

$$\nabla_{12} Y_t = \alpha + \beta_1 \nabla_{12} Y_{t-4} + \beta_2 \nabla_{12} Y_{t-12} + \gamma \varepsilon_{t-12} + \varepsilon_t. \tag{20}$$

Then, we extended the model in order to forecast as follows,

$$\nabla_{12} Y_{t+1} = \alpha + \beta_1 \nabla_{12} Y_{t-3} + \beta_2 \nabla_{12} Y_{t-11} + \gamma \varepsilon_{t-11} + \varepsilon_{t+1}. \tag{21}$$

It could be written down in short as

$$\nabla_{12} Y_{t+1} = \varphi(\alpha, \beta_s, \gamma, \sigma; Z), \tag{22}$$

where $\varepsilon_{t+1} = \sigma Z$, $Z = F_t^{-1}(u)$, and F_t^{-1} is the inverse cumulative distribution function of the standard t distribution with degree of freedom v and u is independent random variable with the uniform distribution $U(0, 1)$.

To predict the log return of the tourists arrival $\nabla_{12} Y_{t+1}$, we computed the minimum and maximum of the function $\varphi(\alpha, \beta_s, \gamma, \sigma; Z)$

$$\binom{Max}{Min} \varphi(\alpha, \beta_s, \gamma, \sigma; Z), \tag{23}$$

under the constraint

$$pl(\alpha, \beta_s, \gamma, \sigma; Y_t) \geq \omega.$$

where ω is an independent random variable with the uniform distribution $U(0, 1)$.

From the above optimization problem, we randomized independently N pairs of the random number $(\omega_i, u_i), i = 1, 2, 3, \ldots, N$ resulting in N intervals $[\nabla_{12} Y_{t+1}^L, \nabla_{12} Y_{t+1}^U]$. For any $A \subset R$, the belief function $(\hat{Bel}_x^Y)$ and plausibility function $(\hat{Pl}_x^Y)$ defined by Eqs. (12) and (13) can be approximated by:

$$\hat{Bel}_x^{\mathscr{Y}} = \frac{1}{N} \# \left\{ i \in \{1, \ldots, N\} \,\middle|\, [\nabla_{12} Y_{t+1}^L, \nabla_{12} Y_{t+1}^U] \subseteq A \right\}, \tag{24}$$

$$\hat{Pl}_x^{\mathscr{Y}} = \frac{1}{N} \# \left\{ i \in \{1, \ldots, N\} \,\middle|\, [\nabla_{12} Y_{t+1}^L, \nabla_{12} Y_{t+1}^U] \cap A \neq \phi \right\}. \tag{25}$$

Figure 3 displays the lower and upper cumulative distribution functions $\hat{Bel}_x^{\mathscr{Y}}$ and $\hat{Pl}_x^{\mathscr{Y}}$ of one period ahead forecasting tourist arrivals. These functions gave us the summary of the predictive belief function. The most plausible forecasting values

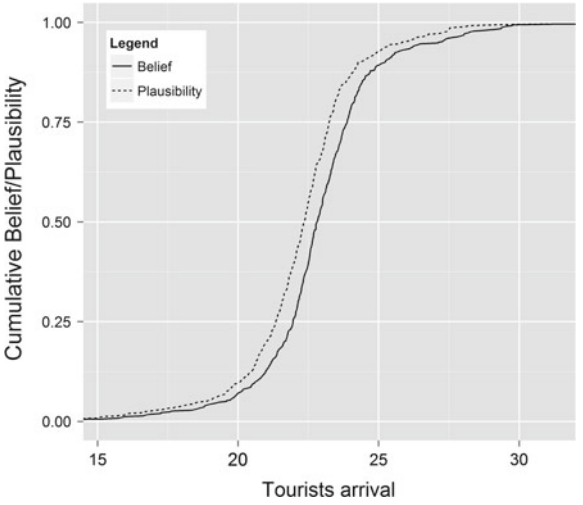


Fig. 3 Lower and upper cumulative distribution functions for the number of tourist arrivals in October 2013, forecasted in September 2013

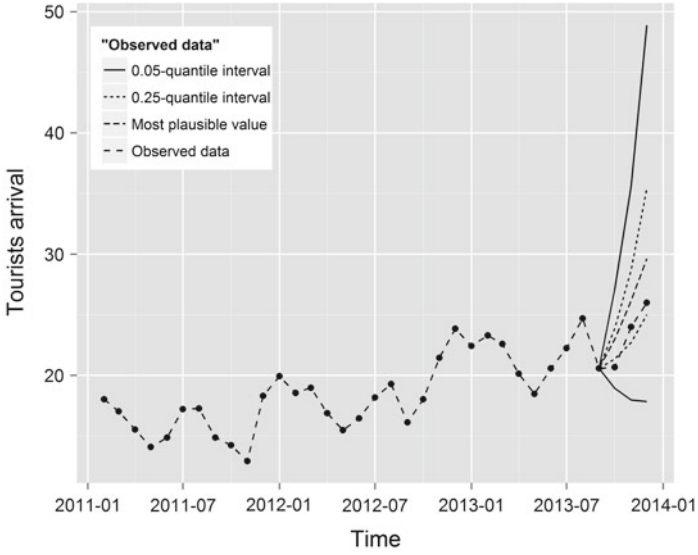


Fig. 4 Forecasting made in September, 2013 for the tourist arrivals in the period October to December, 2013

$\tilde{y}$ are displayed in Fig. 4, together with α-quantile intervals with $\alpha \in \{0.05, 0.25\}$. Assuming that Y was continuous, we could compute its lower and upper predictive quantiles at level α, for any $\alpha \in (0, 1)$, by $q_\alpha^L = (F_x^U)^{-1}(\alpha)$ and $q_\alpha^U = (F_x^L)^{-1}(\alpha)$. We approximated the most plausible value $\tilde{y}$ by the maximum value of $\hat{Pl}_x^{\mathscr{Y}}(\{y\})$. The $\tilde{y}$

can be taken as a point forecasting of Y. We saw that the numbers of tourist arrivals to Thailand in October to December, 2013 were slightly overestimated by $\tilde{y}$. However, the observed data were contained in the 0.05-quantile intervals. The most plausible forecasting values in October to December, 2013 were 2,251,701, 2,549,605, and 2,969,196 persons in comparison with the real observations 2,065,518, 2,399,240, 2,598,015 persons, respectively.

4 Discussion

The data we used for the empirical work, projected some breaks and a slight structural change and led us to check stability test to validate the model of the paper. Since we took log on the number of tourists arrival to Thailand, the data had been stationary. To make ensure we validated the statistical model with Chow's test, and the dynamic stability test. The results revealed that the data set became stable and overcame breaks and a slight structural change. Since we used seasonality in the model, we did not need to use dummy variables.

Although model selection for the statistical model was a little tedious and troublesome task as the data distribution were not informative, the analytical part to find out if the belief function was applicable for forecasting in tourism sector, and was proved relevant with better forecasted interval than traditional methods. We compared the forecasted interval resulted by belief functions with accrual tourist arrivals in October, 2013. Even though the result was not a point, but an interval, it was more efficient than the traditional forecasting methods and the forecasting interval could capture actual number of tourist arrivals. In addition, we showed that point forecasts for October, November, and December 2013 were overestimated, but these forecasts were contained in the 0.05-quantile intervals.

In case to improve the results with Experts' opinion, the paper suggested requesting some experts' opinion for their point estimate and interval estimate. Then, the paper advised combining these data with Dempster's rules. Since there was no expert opinion, we did not apply this idea to the paper. In case, one combines statistical result and expert opinion, finally, the relatively reliable forecast interval would be resulted.

4.1 Combining Historical Data with Expert Opinions

The expert's opinion is relatively informal technique, but it has many advantages as it can cover the time constant and time varying information. In addition, it can take into account some information that the econometric models could not catch up. To get support under complex situation, expert's opinion can be applied.

Dempster's rule can apply to combine not only for intervals but also for triangle data with different degree of confidence. The combination for intervals is simple and follows traditional rule i.e. Dempster's rule of combination. The combination for triangle data is the extension of the Dempster's rule. In this case there might be

debate for the way of assigning degree of confidence. To construct the triangle data, researcher needs to ask the experts opinions for interval and point forecast. Then the researcher can assign confidence between 0 and 1. The point forecast must be nearly 0 confidence as it has high degree of potential to be wrong and interval forecast can be assigned between 1 and 99 % or 100 % for sure as this interval is large. Reasonably, the experts will give larger interval than the forecast interval in their mind to avoid wrong prediction.

For tourist arrivals to Thailand, we recommend to get experts' opinions from some tourism experts from different nature of tour operator or destination management organization (DMO). Then, one can combine these experts' opinions with the statistical results of the paper by using Dempster's rule of combination, since the idea of Dempster's rule is that two independent bodies of evidences make pooling them similarly combining two stochastically independent randomly coded messages [10].

5 Concluding Remarks

The paper focused on the tourist arrivals to Thailand and used the statistical model, SARIMA $(12, 0, 12)(0, 1, 0)_{12}$. This paper overcame the potential bias relating to seasonality with a SARIMA model, and belief functions including Dempster's rule of combination. In applying belief functions, there are two stages, estimation stage and prediction stage. The paper followed formalism of Dempster-Shafer Theory as this method is general and can be applied almost all statistical models. This theory is an extension of Bayesian probability theory and the paper suggests updating the belief with more and more information if accessible as per Bayesian theory.

The results pointed out that the tourist arrivals to Thailand can be forecasted by applying belief functions. In addition, the paper suggested improving the confidence level of the result by using experts opinions on the tourist arrival since Dempster's rule of combination is applicable to combine these intervals. By following the forecasting method mentioned in the paper, we can extend our forecasting not only to the nearest future period but to some near future by depending on the length of lags for the model. For that reason, the method is very useful and efficient for the time series data.

The question may occur for the assigning weight to the degree of confidence to the different expert opinions. Can we assign the weight uniformly to all experts' forecast intervals, or can we assign different weight or degree of confidence to experts' opinion for different time? What are the driving facts to assign confidence level? These questions are also a contribution of the paper to further debate for belief function users.

References

1. Abdallaha, N.B., Voyneau, N.M., Denux, T.: Combining statistical and expert evidence using belief functions: application to centennial sea level estimation taking into account climate chnage. Int. J. Approx. Reason. **55**, 341–354 (2014)
2. Chang, C.L., McAleer, M., Slottje. D.: Modelling international tourist arrivals and volatility: an application to taiwan. In: E-Prints Complutense. The Complutense University of Madrid. Available. http://eprints.ucm.es/8592/1/0906.pdf (2009). 27 June 2014
3. Denoeux, T.: Likelihood-based belief function: Justification and some extensions to low-quality data. Int. J. Approx. Reason. **55**(7), 1535–1547 (2014)
4. Hoti, S., Leon, C., McAleer, M.: Modelling the uncertainty in international tourist arrivals to the Canary Islands. In: School of Economy, The University of Adelaide. Available via DIALOG. http://economics.adelaide.edu.au/events/archive/2004/Modelling-the-Uncertainty-in-International-Tourist-Arrivals-to-the-Canary-Islands.pdf. Cited 27 June 2014 (2004)
5. Kanjanatarakul, O., Sriboonchitta, S., Deneux, T.: Forecasting using belief functions: an application to marketing econometrics. Int. J. Approx. Reason. **55**, 1113–1128 (2014)
6. Neupane, H.S., Shrestha, C.L., Upadhyaya, T.P.: Modelling monthly international tourist arrivals and its risk in Nepal. NRB Econ. Rev. **24**, 28–47 (2011)
7. Nguyen, H.T.: An Introduction to Random Sets. Chapman and Hall/CRC Press, Boca Raton (2006)
8. Saleh, S.A., Verma, R., Ihalanayake, R.: Research online. In: 2010 Oxford Business and Economics Conference Program. Available via DIALOG. http://ro.uow.edu.au/commpapers/1426 (2010). Assessed 5 May 2014
9. Shafer, G.: Constructive probability. Synthesis **48**(1), 1–60 (1981)
10. Shafer, G.: Belief function and parametric models. J. R. Stat. Soc. **44**(3), 322–352 (1982)
11. Shafer, G.: A Mathematical Theory of Evidence. Princeton University Press. 1976. http://fitelson.org/topics/shafer.pdf. Cited 26 June 2014 (1992)
12. Shafer, G.: Dempster-Shafer theory. In: Shafer. Available via DIALOG http://fitelson.org/topics/shafer.pdf (1992). Assessed 26 Jun 2014
13. Tsenga, F., Tzeng, G.: A fuzzy seasonal ARIMA model for forecasting. Fuzzy Sets Syst. **126**(3), 367–376 (2002)
14. Turner., R.: Travel tourism economic impact 2014 world. In: Economic Impact Analysis. World Travel & Tourism. Available via DIALOG http://wttc.org/focus/research-for-action/economic-impact-analysis/country-reports/ (2014). Accessed 15 June 2014

Copula Based Polychotomous Choice Selectivity Model: Application to Occupational Choice and Wage Determination of Older Workers

Anyarat Wichian, Jirakom Sirisrisakulchai and Songsak Sriboonchitta

Abstract This paper aims to estimate the occupational choice equation simultaneously with wage equation for older workers in Thailand by applying the copula approach to a polychotomous choice selectivity model. Several of the copula functions, such as the Frank and Student's t copulas are compared to the standard model which is restricted to a joint normality assumption. This paper demonstrates that a polychotomous choice selectivity model based on the copula approach performs better than the standard one. And, it is evident that among the copula based model, the Frank copula-based model provides the best fit. Also, these results show the presence of highly significant dependency of unobservable factors between the occupational choice regression and the wage regression for unskilled and skilled workers, which implies that the selectivity bias exists. The empirical results show that the older workers who live in Bangkok earn higher wage than those who live outside Bangkok. The gender variable has no impact on wages for the high skilled older workers. For the unskilled, the male worker earns more. A surprised result is that the experience has the significantly negative impact on the wages regardless of the level of the skill of the older workers. This needs serious further research in depth to explain this phenomena.

1 Introduction

Labor supply is a key component in economic growth, especially, the labor force participation of workers aged 15–59. However, at the present, the demographic structure is undergoing a change because of an aging society. In Thailand, between 2000 and 2010, the proportion of people aged 0–14 has been on a decline, while the proportion

A. Wichian (✉) · J. Sirisrisakulchai · S. Sriboonchitta
Faculty of Economics, Chiang Mai University, Chiang Mai, Thailand
e-mail: aunwichian@hotmail.com

J. Sirisrisakulchai
e-mail: sirisrisakulchai@hotmail.com

S. Sriboonchitta
e-mail: songsakecon@gmail.com

V.-N. Huynh et al. (eds.), *Econometrics of Risk*, Studies in Computational Intelligence 583, DOI 10.1007/978-3-319-13449-9_25

of people aged 60 and above has been showing an increase. The proportion of the latter has increased from 9.5 % in 2000 to 12.9 % in 2010 [1]. These data point out that in the future, older workers will inevitably become significant contributors to Thailand's economic activity. Our idea of interest is that the earnings are different in each choice of occupation. The older workers, who are highly skilled workers, may have relatively better earnings than those who are low skilled workers. This leads to selectivity bias, which occurs when unobserved disturbances of occupation choice equation are correlated with those of wage equation. Thus, we could not estimate the wage equation separately. There are two main approaches to correct the selectivity bias; the applications are the two-step procedure and full information maximum likelihood (FIML). The latter method is more efficient and performs better when compared to the former one (see Nawata [2], Nawata and Li [3]). Also, the two-step procedure may have biased estimators due to the collinearity problem between the regression in the selection equation and the outcome equation (Bohara and krieg [4]; Puhani [5]);. However, the FIML method has some crucial drawbacks; namely, it has strong assumptions of bivariate normality for the joint distribution, which leads to incorrect conclusion about the existence of sample selection bias. Thus, econometricians have tried to find the best procedure which can relax the above assumptions and attain the robust estimators. The early important article is due to Lee [6] who allowed the marginal distribution of the disturbances to be non-normal and then transformed it into normal distribution. Although Lee tried to avoid the strong assumption, this method still maintains the bivariate normal distribution, which implies linear dependence between the disturbances.

In recent times, the copula approach has been widely used in the sample selection framework due to many benefits (see Trivedi and Zimmer [7]). The main usefulness is that the joint distributions can be derived when the marginal distributions are allowed to be non-normal margins. Moreover, the concepts and measures of dependence of this approach go beyond the linear correlation which can be developed. Several researchers have tried to apply this approach to the sample selection model due to its main usefulness. Firstly, Smith [8] suggested the general form for the self-selection model using the properties of copula to measure the dependence of the disturbances in the FIML. Actually, Lee's [6] method cited above is among the early works in the copula framework, but the term "copula" is not explicitly used (see Trivedi and Zimmer [7]). The term of the copula has obviously been employed in Smith's [8] paper. Subsequently, there are several papers which have followed Smith's [8] procedure (for example, Smith [9], Genius and Strazzera [10], Eberth and Smith [11], Hasebe and Vijverberg [12], Chinnakum et al. [13], etc.). Nevertheless, these papers applied the copula approach whether it's on the binary choice selectivity or endogenous switching model. Actually, copula approach could be applied to polychotomous choice selectivity model, which Lee [6] generalized the binary choice selectivity model to illustrate the polychotomous choice problems with mixed continuous and discrete dependent variables in the same paper that had been cited above. There are few studies that follow Lee's [6] approach with regards to the polychotomous choice selectivity model. For example, Spissu et al. [14] developed

a copula-based joint multinomial discrete-continuous model of vehicle type choice and miles of travel in field of transportation.

In this current paper we are mainly concerned with the selectivity bias. Therefore, we try to estimate the wage equation of older workers in each of occupation choice by applying the copula approach to the polychotomous choice selectivity model and optimize the data with the FIML. To our knowledge, this current paper is regarded as the first application of copula-based polychotomous choice selectivity model in the context of labor economics. These results will be useful for policy makers to promote the campaigns that can increase the earnings in each of the occupational choices. Furthermore, following this method will attain the efficient estimators, which could reduce the risk that may occur from some strong assumptions.

This paper is organized as follows: Sect. 2 reviews the related literature. Section 3 describes the copula theory, the definitions, and the main properties, bounds of copulas, related measures of dependence, and some examples of copula. Section 4 describes the polychotomous choice selectivity model, as well as explaining on how to apply the copula functions. Section 5 describes the data. Section 6 is devoted to the application of the copula based polychotonomous choice selectivity model to estimate the wage determination according to the occupational choice for older workers. Finally, Sect. 7 provides the conclusion of this paper.

2 Literature Review

Lee [6] extended the binary choice selectivity model to model the polychotomous choice problems with mixed continuous and discrete dependent variables. Importantly, Lee proposed a two-step method, which allowed the marginal distribution of the disturbances to be non-normal, and then transformed it into normal distribution. Another three procedures were proposed by Dubin and McFadden [15], Dahl [16] and Bourguignon et al. [17], which are different in the form of selectivity correction term. Most of the studies on labor economics have followed Lee's [6] two-step procedure; examples are from Dolton and Kidd [18], Tansel [19], Hoyos [20], Demoussis et al. [21]. Although, Lee [6] proposed the FIML method used in the papers cited above, most empirical studies pointed out that the FIML estimator is more efficient and outperforms other methods when it is compared to the two-step estimator (see Lee and Trost [22], Nawata [2], Oya [23]). However, few studies have paid attention on the full information maximum likelihood method (FIML). One such case is the study from Bohara and Krieg [4] that estimated multinomial logit migration equation simultaneously with a system of earnings equations for individuals by migration status. Actually, FIML method has some crucial drawbacks; namely, it has strong assumptions of bivariate normality for the joint distribution, which leads to incorrect conclusion about the existence of sample selection bias. Thus, econometricians have tried to find the best procedure that can relax this strong assumption. Fortunately, in recent times, the copula approach has been widely used in the sample selection framework. Due to various advantages (see Trivedi and Zimmer [7]). Actually, Lee's [6]

this method is among the early work in the copula framework, but the term "copula" is not explicitly used (see Trivedi and Zimmer [7]). The properties of the copulas have obviously been employed in Smith's [8, 9] papers which suggest the general form for the self-selection model using the properties of copula to measure the dependence of the disturbances in the FIML. There are several papers which have followed Smith's [8] procedure; for example, Genius and Strazzera [10], Bhat and Eluru [24], Eberth and Smith [11], Hasebe and Vijverberg [12], an Chinnakum et al. [13], etc.

The previous studies found that copula-based model performs better than the traditional model which is restricted to normality distribution (for example, Smith [9], Genius and Strazzera [10], Chinnakum et al. [13], etc.). However, these papers focused on a sample selection or type 2 Tobit model and endogenous switching model. Since Lee's [6] work, few studies have been based on polychotomous choice selectivity model; for example the study of Spissu et al. [14]. This study developed a copula-based joint multinomial discrete-continuous model for transportation.

Finally, the copula approach has various advantages that had been mentioned above. Also, there are few studies done on the polychotomous choice selectivity model. Therefore, this current paper aims to apply the copula approach to a polychotomous choice selectivity model and estimate the wage determination according to the occupational choice for older workers in Thailand. To our knowledge, this is the first application of a copula-based polychotomous choice selectivity model in labor economics.

3 Copula Theory

3.1 Definition and Properties

The term "copula" and theorem were introduced by Sklar's work in 1959 and 1973, respectively, according to Trivedi and Zimmer [7]. Recently, copula approach has been widely used in various topics in the econometrics fields since it has several advantages. Copula is defined as functions that link or connect multivariate distributions to their one-dimensional margins (see Trivedi and Zimmer [7]). We begin with a bivariate copula function, a simple case, which is defined as the following (see Nelsen [25, p. 10]):

Definition 1 A copula is a function $C : [0, 1]^2 \rightarrow [0, 1]$ with the following properties

1. For every u,v in $[0, 1]$, $C(u, 0) = 0 = C(0, v)$ and $C(u, 1) = u$ and $C(1, v) = v$
2. For every u_1, u_2, v_1, v_2 in [0,1]such that $u_1 \leq u_2$ and $v_1 \leq v_2$,
$C(u_2, v_2) - C(u_2, v_1) - C(u_1, v_2) + C(u_1, v_1) \geq 0$.

Essentially, the theoretical foundation is provided by Sklar's theorem, as given below (see Nelsen [25, p. 18]):

Theorem 1 *Sklar's theorem. Let X and Y be random variables and F be a joint distribution function with margins F_1 and F_2, which are the cumulative distribution functions of the random variables X and Y, respectively. Then, there exists a copula C such that for all real number x, y.*

$$F(x, y) = C(F_1(x), F_2(y)) \tag{1}$$

If F_1 and F_2 are continuous, then C is unique; otherwise, C is uniquely determined on $Ran(F_1) \times Ran(F_2)$, where $Ran(F_i)$ is the range of a funtion F_i. Conversely, if C is a copula and F_1 and F_2 are distribution functions, then the function F defined by Eq. (1) is a joint distribution function with margins F_1 and F_2.

By Sklar's theorem and the method of inversion, the corresponding copula can be generated by using the unique inverse transformations $x = F_1^{-1}(u)$ and $y = F_2^{-1}(v)$. Therefore,

$$C(u, v) = F(F_1^{-1}(u), F_2^{-1}(v)), \tag{2}$$

where u and v are standard uniform variates.

3.2 Empirical Application of Copula

The Sklar's theorem implies that the copula can be used to specify multivariate distribution in terms of its marginal distributions (Trivedi and Zimmer [7]). When the univariate marginal distribution functions are given, copulas allow researchers to bind together joint distribution function. Thus a two-variate function with margins F_1 and F_2, the copula associated with F is a distribution function $C : [0, 1]^2 \rightarrow [0, 1]$ that satisfies

$$F(x, y) = C(F_1(x), F_2(y); \theta), \tag{3}$$

where θ is a parameter of the copula called the dependence parameter, which measures the dependence between the marginals (see Trivedi and Zimmer [7]). Furthermore, the dependence parameter can be used to denote the families of the copulas as notation $C_\theta(u, v)$. There are several examples of families of copulas, such as the Gaussian (Normal) copula, the FGM (Farlie-Gumbel-Morgenstern) copula, the Plackett copula, etc.

3.3 Bounds of Copula

Since copulas relate to the dependence parameter, so defining the bounds of copula has been concentrated on. Usually, the copula lies between two bounds, which is the Fréchet-Hoeffding lower bound, it corresponds to negative dependence and the

Fréchet-Hoeffding upper bound which corresponds to positive dependence. Application of the Fréchet-Hoeffding bounds to a copula in the bivariate case, for any copula C and for all u, v in [0,1], is given by

$$W(u, v) = max(u + v - 1, 0) \leq C(u, v) \leq min(u, v) = M(u, v), \tag{4}$$

where W is the Fréchet-Hoeffding lower bound, and M is the Fréchet-Hoeffding upper bound. In addition, in special cases of copulas, the product copula can be defined if the margins are independent (see Schmidt [26], Trivedi and Zimmer [7]). Some families of copulas are called comprehensive if they include both Fréchet-Hoeffding bounds and product copula, such as the Gaussian and the Frank copulas. While the FGM, Clayton, Gumbel, and Joe copulas are not comprehensive, which make it necessary to calculate the measures of dependence, as described below.

3.4 Measures of Dependence

The measure of dependence can be used to assess the coverage of the copula, which is not comprehensive. The most familiar and often used method is the linear correlation, such as the Pearson's product moment correlation coefficient. But this measure has some drawbacks: first, in general zero correlation, it does not imply independence. Second, it is not defined for the heavy-tailed distribution whose second moments does not exist. Third, it is not invariant under strictly increasing non-linear transformations (see Trivedi and Zimmer [7]). The alternative methods are the concordance measures, such as Kendall's τ and Spearman's ρ_S which the statistician usually uses for application. The former is defined as the following:

$$\tau = P((X - X')(Y - Y') > 0) - P((X - X')(Y - Y') < 0), \tag{5}$$

and the latter is defined as follows:

$$\rho_S = 3(P((X - X')(Y - Y'') > 0) - P((X - X')(Y - Y'') < 0)), \tag{6}$$

where (X, Y),(X', Y'),and (X'', Y'') are independent random vectors, and each vector has a joint distribution function $F(.,.)$ whose margins are F_1 and F_2. Since (X,Y) are continuous random variables whose copula is $C_\theta(u, v)$, the Kendall's τ can be expressed in terms of copulas (see Nelson [27, p. 129]):

$$\tau = 4 \int\int_{[0,1]^2} C_\theta(u, v) dC_\theta(u, v) - 1 = 4E(C_\theta(U, V)) - 1, \tag{7}$$

where the second expression is the expected value of the function $C_\theta(U, V)$ of uniform (0,1) random variables U and V with a joint distribution function C. Also,

Spearman's ρ_S can be simplified thus, in terms of copulas:

$$\rho_S = 12 \int \int_{[0,1]^2} uvdC_\theta(u, v) - 3 = 12E(UV) - 3, \tag{8}$$

where $U = F(X)$ and $V = F(Y)$ are uniform (0,1) random variables with a joint distribution function $C_\theta(u, v)$. Both of the concordance measures are bounded between -1 and 1, and zero under the product copula.

3.5 Some Bivariate Copulas

There are several families of copulas, which are different in functional forms, characteristics and distribution shapes such as symmetric or asymmetric, left or right skewness, thin or fat tails, etc. Table 1 gives the functional forms and characteristics of some copulas.

It can be crudely concluded that the Gaussian, or Normal copula, was proposed by Lee [6] for selectivity models. This copula is comprehensive since it includes the product copula and both of the Fréchet-Hoeffding bounds, and captures both positive and negative dependences. Also, it is radially symmetric in its dependence structure and strong central dependency. In addition, the range of dependence parameter is allowed to $-1 \leq \theta \leq 1$ and for Kendall's τ to $-1 \leq \tau \leq 1$. This copula is given by

$$C(u, v; \theta) = \Phi_2(\Phi^{-1}(u), \Phi^{-1}(v); \theta) \tag{9}$$

or

Table 1 Functional forms and characteristics of bivariate copulas

Copula	Function C(u,v)	Generation function	Range of θ	Range of Kendall's τ
Gaussian	$\Phi_2(\Phi^{-1}(u), \Phi^{-1}(v); \theta)$	–	$-1 \leq \theta \leq 1$	$-1 \leq \tau \leq 1$
Student's t	$t_{\upsilon,\rho}(t^{-1}(u), t^{-1}(v); \upsilon, \rho)$	–	$-1 \leq \theta \leq 1$	$-1 \leq \tau \leq 1$
FGM	$uv(1 + \theta(1-u)(1-v))$	–	$-1 \leq \theta \leq 1$	$-2/9 \leq \tau \leq 2/9$
AMH	$uv/(1 - \theta(1-u)(1-v))$	$log\frac{1-\theta(1-t)}{t}$	$-1 \leq \theta \leq 1$	$-0.18 \leq \tau < 1/3$
Clayton	$(u^{-\theta} + v^{-\theta} - 1)^{-1/\theta}$	$(1/\theta)(t^{-\theta} - 1)$	$0 < \theta < \alpha$	$0 < \tau < 1$
Frank	$-\frac{1}{\theta} ln\{1 + \frac{(e^{-\theta u}-1)(e^{-\theta v}-1)}{e^{-\theta}-1}\}$	$-ln[(e^{\theta t} - 1)(e^{\theta} - 1)]$	$-\alpha < \theta < \alpha$	$-1 \leq \tau \leq 1$
Gumbel	$exp(-[(-lnu)^{\theta} + (-lnv)^{\theta}]^{1/\theta})$	$(-lnt)^{\theta}$	$1 \leq \theta < \alpha$	$0 \leq \tau < 1$
Joe	$1 - [(1-u)^{\theta} + (1-v)^{\theta} - (1-u)^{\theta}(1-v)^{\theta}]^{1/\theta}$	$-ln[1 - (1-t)^{\theta}]$	$1 \leq \theta < \alpha$	$0 \leq \tau < 1$

Source The copula function are given as presented in Trivedi and Zimmer [7] and Smith [8]

$$C(u,v;\theta)=\int_{-\infty}^{\Phi^{-1}(u)}\int_{-\infty}^{\Phi^{-1}(v)}\frac{1}{2\pi(1-\theta^2)^{1/2}}\times\{\frac{-(s^2-2\theta st+t^2)}{2(1-\theta^2)}\}dsdt \tag{10}$$

where Φ^{-1} is the inverse of cdf of standard normal distribution, and $\Phi_2(u,v)$ is the standard bivariate normal distribution with dependence parameter θ.

The Student's t copula is similar to the Gaussian copula but have two dependence parameters, degrees of freedom (υ) and correlation (ρ) (Trivedi and Zimmer [7]). Moreover, this copula allows for joint fat tails (Aas [28]). This copula is given by (see Embrechts et al. [29])

$$C(u,v;\upsilon,\rho)=t_{\upsilon,\rho}(t^{-1}(u),t^{-1}(v);\upsilon,\rho) \tag{11}$$

or (see Trivedi and Zimmer [7])

$$C(u,v;\upsilon,\rho)=\int_{-\infty}^{t_\upsilon^{-1}(u)}\int_{-\infty}^{t_\rho^{-1}(v)}\frac{1}{2\pi(1-\rho^2)^{1/2}}\times\{1+\frac{(s^2-2\rho st+t^2)}{\upsilon(1-\rho^2)}\}^{-(\upsilon+2)/2}dsdt \tag{12}$$

where $t_\upsilon^{-1}(u)$ is the inverse of cdf of standard t-distribution with υ degrees of freedom, which controls the tails heaviness. And $t_{\upsilon,\rho}$ is the bivariate t distribution with degrees of freedom, υ and dependence parameter, ρ.

Another family of copula is FGM (Farlie-Gumbel-Morgenstern) copula. Although this copula is radially symmetric it is similar to the Gaussian copula. But the dependence structure is weaker than that of the Gaussian copula. In addition, this copula is only useful in cases of moderate dependency (see Trivedi and Zimmer [7]). Moreover, it is not comprehensive because it includes only the product copula, and not the Fréchet-Hoeffding lower and upper bounds. The range for Kendall's τ is restricted to $-2/9 \le \tau \le 2/9$. The importance class of copula is the Archimedean copulas for example; the Clayton, Frank, Gumbel and Joe copulas which are popular in empirical works for several reasons. These copulas can display a wide range of dependence properties for different choices of generator function (see Trivedi and Zimmer [7]). Furthermore, Smith [8] pointed out that it makes estimation of the maximum likelihood and calculation of the score function relatively easy. In order to better understand the Archimedean copulas, we need to mention some properties of these copulas. The bivariate Archimedean copulas can be generated in the following form:

$$C_\theta(u,v)=\varphi^{-1}[\varphi(u)+\varphi(v)], \tag{13}$$

where $\varphi : [0,1] \to [0,\alpha]$ is a generator function which satisfies the following properties: $\varphi(1)=0, \varphi'(t)<0$, and $\varphi''(t)>0$ for $0<t<1$. In addition, if $\varphi(0)=\alpha$, then the inverse function φ^{-1}exists. The above form can be written as follows:

$$\varphi(C_\theta(u,v))=\varphi(u)+\varphi(v), \tag{14}$$

Taking the differential with respect to v in the above equation, we obtain the result which will be used in the sample selection model, which can be given as

$$\frac{\partial(C(u,v))}{\partial(v)} = \frac{\varphi'(v)}{\varphi'(C(u,v))}, \tag{15}$$

Also, for the Archimedean copula, Kendall's τ can be described in simple form, as follows:

$$\tau = 1 + 4\int_0^1 \frac{\varphi(t)}{\varphi'(t)}dt, \tag{16}$$

where $\varphi'(t) = \partial\varphi(t)/\partial(t)$.

It can be concluded that there is only Frank copula in the Archimedean class which is comprehensive, symmetric and central dependence, and similar to Gaussian copula. However, the dependence structure is stronger than that of the Gaussian copula. Also, the range for Kendall's τ is allowed to $-1 \leq \tau \leq 1$. In contrast, the Clayton, Gumbel and Joe copulas are not comprehensive and asymmetric. The Clayton copula has strong left tail dependence, and is opposite to Joe and Gumbel copula. Nevertheless, the Gumbel is weaker dependence than the Joe copula. All of them have their ranges for Kendall's τ restricted to $0 \leq \tau < 1$.

In recent times, the bivariate copulas such as Gaussian copula, Student's t copula, etc. have been widely used in several economic research fields such as financial economics (for examples: Patton [30], Boonyanuphong and Sriboonchitta [31] etc.), tourism economics (for examples Puarattanaarunkorn and Sriboonchitta [32] etc.) and agricultural economics (for examples Sriboonchitta et al. [33] Xue and Sriboonchitta [34] etc.). Moreover, the Archimedean copulas such as the Clayton, Frank, Gumbel and Joe copulas, have been extensively used in empirical work (for example, Genius and Strazzera [10], Sener and Bhat [35], Hasebe and Vijverberg [12], Chinnakum et al. [13]). These copulas are different in the generation function, which lead to differences in the functional forms (which are demonstrated in Table 1) and essential dependence structures.

4 Copula Based Polychotomous Choice Selectivity Model

In this current paper, we estimate a multinomial logit equation simultaneous with a wage equation for individual by occupational choice: (1) high skilled worker e.g., managers, senior officials, and professionals (2) skilled worker e.g., skilled agriculture, machine operators (3) unskilled workers and use the FIML. Lee [6] called polychotomous choice selectivity model or polychotomous choice model with mixed continuous and discrete data (Maddala [36, p. 275]). This model relates to correct the selectivity bias, which occurs when unobserved disturbances of occupational choice

equation are correlated with those of wage equation. This can be demonstrated by the system of equations below:

Consider the multinomial logit model, with M categories and wage equation in each category. (see Lee [6], Maddala [36, p. 275]).

$$\ln w_{si} = x_{si}\beta_s + u_{si}(s = 1, 2, 3) \tag{17}$$

$$I_{si}^* = z_{si}\gamma_s + \eta_{si} \tag{18}$$

where $\ln w_{si}$ is the natural log of daily wage for each of ith individual, x_s and z_s are explanatory variables; $E(u_s|x_s, z_s) = 0$, $\ln w_{si}$ is observed only if the sth occupational choice is chosen; I is a polychotomous variables with value 1 to 3 and $I = s$ if the sth occupational choice is chosen. The ith individual will choose the sth occupational choice if

$$I_{si}^* > Max\ I_{ji}^*(j = 1, 2, 3, j \neq s) \tag{19}$$

Let $\varepsilon_{si} = Max\ I_{ji}^* - \eta_{si}$.

Thus we can write below:

$I = s$ iff $\varepsilon_s < z_s\gamma_s$, where the index i has been quit, for easiness.

As shown in Domencich and McFadden [37], $\eta_j(j = 1, 2, ..., M)$ are assumed to be independently and identically distributed, with the type I extreme value distribution. Then (see Lee [6], Maddala [36, p. 275])

$$Prob(\varepsilon_s < z_s\gamma_s) = Prob(I = s) = \frac{exp(z_s\gamma)}{\sum_j exp(z_s\gamma)} \tag{20}$$

Thus,

$$F_s(\varepsilon) = Prob(\varepsilon_s < \varepsilon) = \frac{exp(\varepsilon)}{exp(\varepsilon) + \sum_{j=1,2,..M, j\neq s} exp(z_j\gamma)} \tag{21}$$

Lee [6] suggested a general transformation to normality. Consider transformation as below:

$\varepsilon_s^* = J_s(\varepsilon_s) = \Phi^{-1}[F_s(\varepsilon_s)]$

$u_s^* = G_s(u_s) = \Phi^{-1}[G_s(u_s)]$

where $J_s(\varepsilon_s)$ and $G_s(u_s)$ are the distribution functions of ε_s and u_s. $\Phi(.)$ is the distribution function of the standard normal. $\Phi^{-1}(.)$ is the inverse of the standard normal distribution function. Thus ε_s^* and u_s^* have $N(0, 1)$ distributions. Then the multinomial logit and regression equation are estimated jointly by using the full information maximum likelihood method (FIML) (Lee [6]).

The general joint likelihood function can be written as below: (Spissu et al. [14])

$$L = \prod_{i=1}^{T} \left[\prod_{s=1}^{M} \{ P(\ln w_{si} | z_{si}\gamma_s > \varepsilon_s) \times P(z_{si}\gamma_s > \varepsilon_s) \}^{R_{si}} \right] \tag{22}$$

where R_{si} is a dichotomous variable that take the value 0 and 1, with $R_{si} = 1$ if the sth occupational choice is chosen by the ith individual and $R_{si} = 0$ otherwise.

Let $F_{\varepsilon s}(.)$ and $F_{us}(.)$ be the cumulative distribution functions of the disturbance terms ε_s and u_s, respectively. Spissu et al. [14] used normal distribution functions for the marginal $F_{\varepsilon s}(.)$ and $F_{us}(.)$ And the Eq. (22) can be expressed in terms of the copula approach as below (see Spissu et al. [14])

$$L = \prod_{i=1}^{T} \left[\prod_{s=1}^{M} \left\{ \frac{1}{\sigma_{us}} \times \frac{\partial C_{\theta s}(u, v)}{\partial v} f_{us} \left(\frac{\ln w_{si} - x_{si}\beta_s}{\sigma_{us}} \right) \right\}^{R_{si}} \right] \tag{23}$$

where $C_{\theta s}(.,.)$ is the copula equivalent to $F_{\varepsilon s,us}(u, v)$ with $u = F_{\varepsilon s}(z_s\gamma_s)$ and $v = F_{us}\left(\frac{\ln w_{si} - x_{si}\beta_s}{\sigma_{us}}\right)$. f_{us} is the probability density function of u_{si}, and σ_{us} is the scale parameter of u_{si}. The expression for the component of $\partial C_{\theta s}(u, v)/\partial v$ can be simplified by using Eq. (15). And this is given by the selected families of copulas, as presented in Table 2.

The one appealing advantage of a copula approach is that, it allows for flexibility in the families of copulas for the dependence between the disturbance terms in the multinomial logit equation (ε_s) and those in the wage equations (u_s). However, some of copulas such as Clayton, Gumbel and Joe copulas are not comprehensive and cannot account for negative dependence. Although the FGM and AMH copula can capture both positive and negative dependences, but the range for Kendall's τ are restricted to $-2/9 \leq \tau \leq 2/9$ and $-0.18 \leq \tau < 1/3$, respectively. Thus in the

Table 2 Expressions for $\frac{\partial}{\partial v} C_\theta(u, v)$

Copula	Expressions for $\frac{\partial}{\partial v} C_\theta(u, v)$
Gaussian	$\Phi\{(\Phi^{-1}(u) - \theta\Phi^{-1}(v))/\sqrt{1-\theta^2}\}$
Student's t	$t_{\upsilon+1}\left(\frac{t_\upsilon^{-1}(u) - \rho t_\upsilon^{-1}(v)}{\sqrt{\frac{(\upsilon + (t_\upsilon^{-1}(v))^2)(1-\rho^2)}{\upsilon+1}}}\right)$
FGM	$u\,[1 + \theta(1-u)(1-2v)]$
Clayton	$u^{-(\theta+1)}\left(u^{-\theta} + v^{-\theta} - 1\right)^{-\frac{1+\theta}{\theta}}$
Frank	$\left[1 - e^{\theta C_\theta(u,v)}\right](1 - e^{\theta v})^{-1}$
Gumbel	$v^{-1}(-\ln v)^{\theta-1} C_\theta(u, v)\left[(-\ln u)^\theta + (-\ln v)^\theta\right]^{(\frac{1}{\theta}-1)}$
Joe	$\overline{v}^{\theta-1}(1 - \overline{u}^\theta)\left[\overline{u}^\theta + \overline{v}^\theta - \overline{u}^\theta\overline{v}^\theta\right]^{\frac{1}{\theta}-1}$

Notes: (1) The expressions are presented in Bhat and Eluru [24] and Aas et al. [38]. (2) $\overline{u} = 1 - u, \overline{v} = 1 - v$

current paper, we consider 3 different families of copulas such as Gaussian, Student's t and Frank copula. Moreover, the family of Student $-t_v$ distribution is the appealing and appropriate one for wage density, according to Heckman et al. [39]. Thus we specify Student$-t_v$ distribution for margin F_{us}.

Last, but not the least, the AIC (Akaike information criterion) and the BIC (Bayesian information criterion) can be used to select between the competing copula models. The AIC and the BIC values are equal to $-2\ln(L) + 2K$ and $-2\ln(L) + \ln(Q)K$, respectively, where $\ln(L)$ is the log-likelihood value at convergence, K is the number of parameters, and Q is the number of observations. The better copula-based model is identified by the lowest values of AIC or BIC. Fortunately, if the competing copula models have same exogenous variables and univariate margins fixed across the model, choosing based on these selection criteria is equivalent to selection based on the maximized log-likelihood (see Smith [9], Bhat and Eluru [24], Hasebe [40]). Bhat and Eluru [24] also concluded that in the case of non-nest models, the BIC is the most widely used approach to select from among the competing models.

5 Data

The data set used for this analysis is a sample from the "The Labor Force Survey Whole Kingdom, Quarter 3: July–September 2012" conducted by the National Statistical Office. The sample used consisted of 1,513 observations regarding older workers, 477 and 130 of whom decided to hire as production workers and managers respectively.

This current paper uses occupational choices as the dependent variables for the multinomial logit equation, which we assume that individuals face three occupational choices, which are unskilled workers ($j = 0$), skilled worker ($j = 1$) e.g., skilled agriculture, machine operators and high skilled worker ($j = 2$) e.g., managers, senior officials, and professionals. And we use logarithm of current wage per day of the individual (lwpd) as the dependent variables for the wage equations. As far as the multinomial logit equation was concerned to investigate the choice of occupations, the regressing of the dependent variable was done on gender and education (years). The regressors of the wage equation were as follows: experience (years), gender and region.

6 Results

This current paper aims to apply the copula approach to a polychotomous choice selectivity model and estimate the wage determination according to occupational choice for older workers in Thailand. Importantly, we estimated the dependence parameters between the disturbance terms in the occupational choices equation

(ε_s) and those in the three wage equations (u_s). In other words we estimated the dependence parameters between the residuals of the occupational choices equation and those of wage regressions for unskilled workers, skilled workers and high skilled workers or ρ_0, ρ_1 and ρ_2, respectively. In this current paper we use the same copula dependence structure for those three dependence structures such as Gaussian-Gaussian-Gaussian (Gaussian copula based-model or standard model), Frank-Frank-Frank (Frank copula-based model) and Student's t- Student's t- Student's t (Student's t copula-based model). Also, we estimated the independence model to confirm the existence of the selectivity bias. Furthermore, occupational choices and wage equations are estimated jointly using FIML. The results for the standard polychotomous choice selectivity model and the copula-based model are presented in Table 3.

The main result shows that all of the copula-based models perform better than the standard one, which is restricted to the normal distribution assumption, based on the evaluated AIC and BIC criteria-especially the Frank copula-based model. The log-likelihood value at convergence and the BIC value of the Frank copula-based model are −2073.131 and 4321.987, respectively (as shown in Table 3). Moreover, the log-likelihood value at for the independent model is −2099.725. The estimated dependence parameter between the residual of the occupational choices equation and those of wage regressions for unskilled workers, and skilled workers or ρ_0, and ρ_1, respectively are significantly different from zero in all of the copula-based models. Although ρ_2 or the dependence parameter between the residuals of the occupational choices equation and those of wage regression for high skilled workers is insignificant in standard and Student's t copula-based model, it was found to be significant at 10 % level for the Frank copula-based model. This implies that there exists significant dependence between these disturbance terms, which explains the existence of the selectivity bias. Importantly, this result shows the significant dependence parameter of ρ_2 only in the Frank copula-based model. In addition, this implies that all of the regimes are suitable for central dependence, it is not suitable in the cases of clustering of values in the tail dependence, regardless of whether it is left or right tail dependence.

The estimated parameters from both the standard model and the candidate copula-based model are illustrated in Table 3. The main results are the following: First, the estimated parameters are similar to all of the models. These findings are similar to those obtained in several of the previous studies (for example, Smith [9], Genius and Strazzera [10]).

Consider the Frank copula-based model results in Table 3. The estimated parameters in the occupational choice equations indicate that gender has a significantly negative impact on occupational choice, while it is the opposite in the case of the variable of education level, whatever the choice of occupations. The female workers have a lower probability of being skilled or highly skilled workers than male workers. However, older workers with higher education have a significantly higher probability of being skilled or highly skilled workers.

Table 3 also shows the results for wage regressions, these results surprisingly indicate that years of experience have a significantly negative impact on wages,

Table 3 Estimates of Independent, Standard and Copula-based Models

Variable	Independent	Standard	Frank copula	Student's t copula
Occupations choice equation (for skilled workers)				
Constant	−0.919***(0.099)	−1.141***(0.010)	−1.084***(0.100)	−1.097***(0.010)
Gender	−1.671***(0.148)	−1.826***(0.151)	−1.705***(0.146)	−1.906***(0.155)
Education	0.179***(0.017)	0.232***(0.017)	0.220***(0.018)	0.227***(0.017)
Occupations choice equation (for high skilled workers)				
Constant	−3.188***(0.189)	−3.134***(0.187)	−3.165***(0.186)	−3.127***(0.183)
Gender	−2.669***(0.324)	−2.563***(0.317)	−2.716***(0.323)	−2.869***(0.331)
Education	0.321***(0.022)	0.317***(0.023)	0.326***(0.023)	0.330***(0.023)
Wage equation (for unskilled workers)				
Constant	6.626***(0.142)	6.383***(0.148)	6.194***(0.144)	6.823***(0.159)
Experience	−0.021***(0.003)	−0.019***(0.003)	−0.017***(0.003)	−0.027***(0.003)
Gender	−0.171***(0.025)	−0.099***(0.029)	−0.036***(0.028)	−0.084***(0.031)
Region	0.579***(0.062)	0.555***(0.060)	0.540***(0.057)	0.549***(0.060)
σ_0	0.419***(0.013)	0.431***(0.015)	0.454***(0.017)	0.444***(0.016)
ρ_0	–	−0.373***(0.067)	−4.488***(0.624)	−0.402***(0.084)
τ_0	–	−0.243	−0.423	−0.264
υ_0	–	–	–	4.948(4.188)
Wage equation (for skilled workers)				
Constant	9.799***(0.206)	9.266***(0.225)	9.605***(0.247)	9.543***(0.236)
Experience	−0.075***(0.004)	−0.056***(0.005)	−0.063***(0.006)	−0.062***(0.005)
Gender	0.058(0.069)	0.330***(0.082)	0.207***(0.083)	0.337***(0.080)
Region	0.326***(0.075)	0.409***(0.085)	0.314***(0.082)	0.369***(0.082)
σ_1	0.554***(0.023)	0.733***(0.061)	0.678***(0.061)	0.687***(0.053)
ρ_1	–	0.791***(0.068)	5.023***(1.515)	0.786***(0.060)
τ_1	–	0.581	0.458	0.576
υ_1	–	–	–	2.389(2.883)
Wage equation (for high skilled workers)				
Constant	8.5289***(0.446)	8.488***(0.450)	8.560***(0.448)	9.282***(0.525)
Experience	−0.049***(0.009)	−0.051***(0.010)	−0.045***(0.010)	−0.058***(0.020)
Gender	0.031(0.189)	−0.034(0.214)	0.086(0.195)	0.051(0.312)
Region	1.012***(0.215)	1.052***(0.222)	0.992***(0.214)	0.924***(0.249)
σ_2	0.653***(0.051)	0.646***(0.051)	0.680***(0.064)	0.629***(0.102)
ρ_2	–	−0.164(0.232)	1.618*(1.290)	0.285(0.662)
τ_2	–	−0.105	0.175	0.184
υ_2	–	–	–	6.581***(3.447)
$LogL$	−2099.725	−2080.803	−2073.131	−2080.051
AIC	4241.449	4209.606	4194.262	4214.102
BIC	4353.208	4337.331	4321.987	4357.792

Notes: The standard errors are given in the brackets. The significance levels are the following:
* 10 %,
** 5 %,
*** 1 %

regardless of the choice of occupation. These results need in depth further research to explain this important phenomena. The variable that is also significantly positive impact on wages is the region. The older workers who live in Bangkok province are likely to have higher wages than outside Bangkok. Gender has no impact on the wages of the high skilled older workers, while the female has lower wage in the unskilled older workers. This is not a surprised result.

7 Conclusion

This current study aims to estimate the wage determination according to occupational choice for older worker in Thailand by applying the copula approach to a polychotomous choice selectivity model and using "The Labor Force Survey of Whole Kingdom, Quarter 3: July-September 2012" data set. The main results are as follows: First, based on the log-likelihood value and the criterion of BIC, the copula approach to a polychotomous choice selectivity model (which allows for flexibility in the dependence of the disturbance terms in the occupational choices equation (ε_s) and those in the three wage equations (u_s) performed better than the standard one (which is restricted by the normal distribution assumption). Also it is evident that among the copula based models, the Frank copula-based model provides the best fit.

Second, these results show the presence of significant dependency of unobservable factors between the occupational choice regression and the wage regressions for unskilled and skilled workers, which implies that the selectivity bias exists. However, we found that the estimated dependence parameter for the wage regression of high skilled workers is lowly significant.

Third, female workers earn less than the males in the unskilled older workers, while there is no difference in the high skilled. Fourth, surprisingly, the experience has the significantly negative impact on the wage earned, which needs to do more in depth research to explain this phenomena. Fifth, older workers who live in Bangkok earn higher wage than those live outside Bangkok.

Acknowledgments We would like to thank the referees for their comments and suggestions on the manuscript and National Statistical Office for database. The first author was supported by CHE PhD scholarship.

References

1. National Statistical Office. The Labor Force Survey Whole Kingdom Quarter 1/2013. http://web.nso.go.th/en/pub/data_pub/130712_percent20LaborForce12.pdf. (2013)
2. Nawata, K.: Estimation of the female labor supply models by Heckman's two-step estimator and the maximum likelihood estimator. Math. Comput. Simul. **64**, 385–392 (2004)

3. Nawata, K., Li, M.: Estimation of the labor participation and wage equation model of Japanese married women by the simultaneous maximum likelihood method. J. Jpn. Int. Econ. **18**, 301–315 (2004)
4. Bohara, A.K., Krieg, R.G.: A simultaneous multinomial logit model of indirect internal migration and earnings. JRAP **28**(1), 60–71 (1998)
5. Puhani, P.: The Heckman correction for sample selection and its critique. J. Econ. Surv. **14**(1), 53–68 (2000)
6. Lee, L.F.: Generalized econometric models with selectivity. Econometrica **51**, 507–512 (1983)
7. Trivedi, P.K., Zimmer, D.M.: Copula modeling: an introduction for practitioners. Found. Trends Econom. 1(1), 1–111, Now Publishers (2005).
8. Smith, M.D.: Modelling sample selection using Archimedean copulas. Econom. J. **6**, 99–123 (2003)
9. Smith, M.D.: Using Copulas to model switching regimes with an application to child labour. Econ. Rec. **81**(255), S47–S57 (2005)
10. Genius, M., Strazzera, E.: Applying the Copula approach to sample selection modeling. Appl. Econ. **40**(11), 1443–1455 (2008)
11. Eberth, B., Smith, M.D.: Modelling the participation decision and duration of sporting activity in Scotland. Econ. Model. **27**, 822–834 (2010)
12. Hasebe, T., Vijverberg, W.: A Flexible Sample Selection Model: A GTL-Copula Approach. IZA Discussion Paper No. 7003, (2012).
13. Chinnakum, W., Sriboonchitta, S., Pastpipatkul, P.: Factors affecting economic output in developed countries: a Copula approach to sample selection with panel data. Int. J. Approx. Reason. **54**(6), 809–824 (2013)
14. Spissu, E., Pinjari, A.R., Pendyala, R.M., Bhat, C.R.: A Copula-based joint multinomial discrete-continuous model of vehicle type choice and miles of travel. Transp. **36**, 403–422 (2009)
15. Dubin, J.A., McFadden, D.L.: An econometric analysis of residential electric appliance holdings and consumption. Econom. **52**, 345–362 (1984)
16. Dahl, G.B.: Mobility and the returns to education: testing a Roy model with multiple markets. Econom. **70**, 2367–2420 (2002)
17. Bourguignon, F., Fournier, M., Gurgand, M.: Selection bias corrections based on the multinomial logit model: monte-carlo comparisons. J. Econ. Surv. **21**(1), 174–205 (2007)
18. Dolton, P.J., Kidd, M.P.: Occupational Access and Wage Discrimination. Oxf. Bull. Econ. Stat. **56**(4), 457–474 (1994)
19. Tansel, A.: Formal Versus Informal Sector Choice of Wage Earners and Their Wages in Turkey. Economic Research Forum for the Arab Countries, Iran and Turkey (1999)
20. De Hoyos, R.E.: Structural modelling of female labour participation and occupation decisions. Faculty of Economics (DAE), University of Cambridge, Cambridge Working Papers in Economics (2006).
21. Demoussis, M., Giannakopoulos, N., Zografakis, S.: Native-immigrant wage differentials and occupational segregation in the Greek labour market. Appl. Econ. **42**, 1015–1027 (2010)
22. Lee, L.F., Trost, R.P.: Estimation of some limited dependent variable models with application to housing demand. J. Econom. **8**, 357–382 (1978)
23. Oya, K.: Properties of estimators of count data model with endogenous switching. Math. Comput. Simul. **68**, 539–547 (2005)
24. Bhat, C.R., Eluru, N.: A Copula-based approach to accommodate residential self-selection effects in travel behavior modeling. Transp. Res. Part B **43**(7), 749–765 (2009)
25. Nelsen, R.B.: An Introduction to Copulas, 2nd edn. Springer, New York (2006)
26. Schmidt, T.: 'Coping with Copula'. Forthcoming in Risk Books: Copulas-From Theory to Applications in Finance. Department of Mathematics, University of Leipzig, (2006).
27. Nelsen, R.B.: An Introduction to Copulas. Lecture Notes in Statistics. Springer-Verlag, New York (1999)
28. Aas, K.: Modelling the dependence structure of financial assets: a survey of four copulas. Norwegian Computing Center, Oslo (2004)

29. Embrechts, P., Lindskog, F., McNeil, A.: Handbook of heavy tailed distributions in finance. In: Rachev, S.T. (ed.) Modelling Dependence with Copulas and Applications to Risk Management. Handbooks in Finance: Book 1, pp. 329–385. Elsevier Science B.V. North-Holland (2003).
30. Patton, A.J.: Modelling asymmetric exchange rate dependence. Int. Econ. Rev. **47**(2), 527–556 (2006)
31. Boonyanuphong, P.: Sriboonchitta, S. In: Huynh, V.N., Kreinovich, V., Sriboonchitta, S. (eds.) Modeling Dependence in Econometrics in Advances. Intelligent Systems and Computing. Springer, Heidelberg (2014)
32. Puarattanaarunkorn, O., Sriboonchitta, S.: Copula based GARCH dependence model of chinese and korean tourist arrivals to thailand: implications for risk management. In: Huynh, V.N., Kreinovich, V., Sriboonchitta, S. (eds.) Modeling Dependence in Econometrics in Advances. Intelligent Systems and Computing, pp. 343–366. Springer, Heidelberg (2014).
33. Sriboonchitta, S., Nguyen, H.T., Wiboonpongse, A., Liu, J.: Modeling volatility and dependency of agricultural price and production indices of Thailand: static versus time-varying copulas. Int. J. Approx. Reason. **54**(6), 793–808 (2013)
34. Xue, G., Sriboonchitta, S.: Co-movement of prices of energy and agricultural commodities in biofuel Era: a period-GARCH Copula approach. In: Huynh, V.N., Kreinovich, V., Sriboonchitta, S. (eds.) Modeling Dependence in Econometrics in Advances. Intelligent Systems and Computing, pp. 505–520. Springer, Heidelberg (2014).
35. Sener, I.N., Bhat, C.R.: A Copula-based sample selection model of telecommuting choice and frequency. Environ. Plan. A **43**(1), 126–145 (2009)
36. Maddala, G.S.: Limited-Dependent and Qualitative Variables in Econometrics. Cambridge University Press, Cambridge (1983) (Econometric Society Monographs No.3).
37. Domencich, T., McFadden, D.: Urban Travel Demand: A Behavioral Analysis. North Holland Publishing Company, Amsterdam (1975)
38. Aas, K., Czado, C., Frigessi, A., Bakken, H.: Pair copula construction of multiple dependence. Insur. Math. Econ. **44**, 182–198 (2009)
39. Heckman, J., Tobias, J.L., Vytlacil, E.: Simple estimators for treatment parameters in a latent variable framework with an application to estimating the returns to schooling, NBER Working Paper W7950, (2001).
40. Hasebe, T.: Copula-based maximum-likelihood estimation of sample-selection models. Stata J. **13**(3), 547–573 (2013)

Estimating Oil Price Value at Risk Using Belief Functions

Panisara Phochanachan, Jirakom Sirisrisakulchai and Songsak Sriboonchitta

Abstract We consider extreme value theory to study extreme price movements in crude oil market. Autoregressive-Moving-Average models are developed to describe daily log return of crude oil price. Peak-over-threshold models are then used to model the log return forecasting errors (residuals). The maximum residuals are expressed in terms of value-at-risk or return level corresponding to accepted levels of risk so that appropriate risk measures can be taken. A likelihood-based belief function is constructed to quantify estimation uncertainty. As a result, we can assess the plausibility of various assertions about the value-at-risk of the idiosyncratic shocks in the world crude oil market.

1 Introduction

Oil prices have increased dramatically and fluctuated wildly in the past decade. Many world events have led to oil disruption. The oil crisis started in October 1973 following political and military turmoil, especially in the middle-East with the Arab oil embargo and the conflict between Egypt and Syria against Israel. The 1979–1980 (or second) oil crisis originated from the Iranian revolution. Particularly, the 1980 outbreak of the Iran–Iraq War and the 1990–1991 Persian Gulf war resulted in oil crises. The third oil crisis in 2003–2008 and the inflation resulted in a slight increase in the price of a barrel of crude oil on NYMEX.

In the past, prices in the oil market have been volatile and difficult to predict. Therefore, there is a need of protection against market risk. In this paper, we study oil price fluctuations and implement an effective tool for oil price risk management. Particularly, Extreme value theory (EVT) has been successfully applied to the

P. Phochanachan(✉) · J. Sirisrisakulchai · S. Sriboonchitta
Faculty of Economics, Chiang Mai University, Chiang Mai, Thailand
e-mail: Panisara_pho@hotmail.com

J. Sirisrisakulchai
e-mail: sirisrisakulchai@hotmail.com

S. Sriboonchitta
e-mail: songsakecon@gmail.com

V.-N. Huynh et al. (eds.), *Econometrics of Risk*, Studies in Computational Intelligence 583, DOI 10.1007/978-3-319-13449-9_26

measurement of risk in finance. Even though many studies have been conducted to forecast the return of crude oil prices, there has been little work the application of on EVT-based calculation of value-at-risk (VaR) on crude oil market. We can mention the study of price risk in NYMEX energy complex and the application of the conditional and unconditional factors to estimate VaR by [14], the study of the daily spot of Brent and WTI oil prices by applying both unconditional and conditional EVT models to forecast Value at Risk by [15], and the study of the application of EVT by Peak Over Threshold method to daily returns of Canadian spot oil price and measuring VaR and ES by [19].

In this study, we use the Dempster-Shafer theory introduced by [5, 20] to construct a belief function on the VaR. The Generalized Pareto distribution (GPD) is used for its ability to capture the tail distribution in the financial market. We propose an alternative method to assess the plausibility of extreme value based on the belief functions. The plausibility obtained can be used in further study to combine other evidence or expert opinions.

The main classical approaches to statistical inference are (1) the frequentist approach, relying on confidence intervals and significance testing, (2) Bayesian inference and (3) the likelihood-based approach [3, 8]. A detailed discussion of the limitations of these methods, which motivate the introduction of new inference procedures based on belief functions, is beyond the scope of this paper. In short, the confidence levels and p-values computed using frequentist methods are pre-experimental measures that relate to sequences of trails rather than to specific questions [6]. Bayesian inference does provide post-experimental analysis but relies on prior probability distributions, which are often not available in practice. The inference procedures used in this paper are more in the spirit of likelihood-based inference in that a belief function in the parameter space is derived directly from the likelihood function. The method is also compatible with Bayesian inference, as it provides the same results when a probabilistic prior is provided [7]. However, the belief function approach do not require any information on prior. The advantages of the belief function approach as compared to the frequentist and Bayesian approaches are discussed at length in [7] and [13], among others. The reader is referred to these references for a detailed exposition of the motivations underlying this method of inference.

This paper is organized as follows. Section 2 discusses the methodology and Sect. 3 reports empirical results. Conclusions are presented in Sect. 4.

2 Methodology

The objective of this study is to estimate the Value at Risk (VaR) of the returns from investing in oil price using an Autoregressive Moving Average (ARMA) model, extreme value theory and belief functions. The ARMA model was applied to the mean equation of the returns. The extreme value method then allowed us to model the unobserved shocks, which were basically the residuals from the ARMA equation.

Finally, a likelihood-based belief function was constructed to quantify the uncertainty on estimated parameter. As a result, we obtained a plausibility function on the VaR of the idiosyncratic shocks in the world crude oil market.

2.1 ARMA Model

The ARMA model is based on the assumption that the current value of a time-series is a linear combination of its previous values plus a combination of current and previous values of the residuals [4]. Thus, the general linear ARMA(p, q) model for the conditional mean is expressed as

$$r_t = c + \sum_{i=1}^{p} \beta_i r_{t-i} + \sum_{j=1}^{q} \alpha_j \varepsilon_{t-j} + \varepsilon_t, \tag{1}$$

where r_t is the dependent variable at time t, c is a constant, p is the order of the autoregressive (AR) part, q is the order of moving average (MA) part, the β_i are AR coefficients, the α_j are MA coefficients and ε_t is the error. To estimate coefficients c, β_t and α_t the error terms ε_t are assumed to be independent and to have a normal distribution with zero mean and constant variance. The orders of the AR and MA processes can be determined from the Autocorrelation Function (ACF) and the Partial Autocorrelation Function (PACF). A diagnostic check can be performed by analyzing the residuals using the Ljung-Box Q-statistics, which should not show any significant autocorrelations in the residuals. Moreover, the best model can be selected using the Akaike Information Criterion (AIC) or the Bayesian Information Criterion (BIC) [11].

2.2 Extreme Value Theory

Extreme Value Theory (EVT) is used to model the extreme shocks of the returns. Our variable of interest here is the residual from the ARMA model ε_i. We assume that the shocks are independent and identically distributed (i.i.d.) to satisfy the EVT assumptions.

The EVT method was designed to analyze maxima or minima of some events. The method is commonly used to study the tail behavior of a variable distribution in financial markets. This is because the method relies on the asymptotic distribution of the tail and is robust to distributional assumption of the variable. EVT has two main approaches [9, 18]. The first approach is the Block Maxima method (BMM), which divides the data into blocks and examines the behaviors of the maximum or minimum observations from each of the block. The BMM uses the Generalized Extreme Value (GEV) distribution as it is the limiting distribution of normalized maxima or minima

for a series of i.i.d random variables. The second approach is the Peak Over Threshold (POT) method, which examines the behavior of the exceedances, i.e., all observations that exceed a given threshold. The parametric POT method is based on the extreme value theorem by [17], which states that the exceedances are asymptotically distributed the Generalized Pareto distribution (GPD), following [16]. Therefore, the GPD is used to describe the tail behavior [10, 12]. As the BMM requires large data sets, we used the POT approach in this study.

Let the exceedances over the threshold be $y_t = \varepsilon_t - u$, where u is the threshold. The conditional excess distribution function $F_u(y)$ above of threshold u from the random variable ε_t is defined as

$$F_u(y) = P\left(\varepsilon_t - u \leq y \mid \varepsilon_t > u\right). \tag{2}$$

The selection of the suitable threshold in this study follows [1, 17]. To choose the threshold, we face a trade-off between variance and bias. When using a lower threshold, the number of observations above the threshold increases; as a result, the parameter estimates have smaller variance but are biased. On the other hand, choosing a high threshold decreases the number of observations, thus reducing the bias but making the estimator more volatile. The standard approach for threshold selection is the empirical Mean Excess (ME) plot as discussed in [9]. In [1] and [17], the authors developed the extreme value theorem for the POT method stating that, for any distribution $F_u(.)$, the conditional distribution of the exceedances asymptotically converges to the generalized Pareto distribution (GPD) [9]. That is, the exceedance variable Y has the following distribution:

$$G_{\xi,\sigma,\mu}(y) = \begin{cases} 1 - \left(1 + \xi \frac{y-u}{\sigma}\right)^{-1/\xi}, & \xi \neq 0, \\ 1 - \exp\left(\frac{-(y-u)}{\sigma}\right), & \xi = 0, \end{cases} \tag{3}$$

where $G_{\xi,\sigma,\mu}(.)$ is the GPD function, $\sigma \geq 0$, and $y \geq 0$ when $\xi \geq 0$ and $0 \leq y \leq -\sigma/\xi$ when $\xi < 0$.

The tail index ξ and scale parameter σ are estimated by maximizing the likelihood function for the sample exceeding the threshold u. The individual probability density function is derived from G as follows:

$$g(y_i) = \frac{1}{\sigma}\left(1 + \xi\left(\frac{y_i - u}{\sigma}\right)\right)^{\frac{-1-\xi}{\xi}}. \tag{4}$$

For $\xi = 0$, the individual density function is obtained as

$$g(y_i) = \frac{1}{\sigma}\exp\left(\frac{-1}{\sigma}\left(y_i - u\right)\right). \tag{5}$$

The likelihood function $L(\xi, \sigma \mid y_i)$ for the GPD is the joint density of the n observations. It is defined as

$$L(\xi, \sigma \mid y_i) = \begin{cases} \sigma^{-n} \prod_{i=1}^{n} \left(1 + \xi(\frac{y_i - u}{\sigma})\right)^{\frac{-1-\xi}{\xi}}, & \xi \neq 0, \\ \sigma^{-n} \prod_{i=1}^{n} \exp\left(\frac{-1}{\sigma}(y_i - u)\right), & \xi = 0. \end{cases} \tag{6}$$

2.2.1 Measure of Extreme Risk

According to [22], the Value at risk (VaR) is a measure of extreme risk in term of the possible gains or losses given the distribution F. The VaR is the α-th quantile of the distribution F,

$$VaR_\alpha = F^{-1}(\alpha), \tag{7}$$

where $F^{(-1)}$is the quantile function, which is the inverse of the distribution function F.

Using Eq. (2), $F_u(y) = \frac{F(y+u)-F(u)}{1-F(u)}$ and setting $\varepsilon_t = u + y$ for $\varepsilon_t \geq u$ we can express the model in term of the tail of the underlying distribution $F(\varepsilon_t)$,

$$F(\varepsilon_t) = (1 - F(u))G_{\xi,\sigma}(\varepsilon_t - u) + F(u) \tag{8}$$

Replacing F_u by the generalized Pareto distribution (GPD) and F_u by the estimate $(n-N_u)/n$, where n is the number of observations and N_u the number of observations above the threshold u, we have

$$\hat{F}(\varepsilon_t) = 1 - \frac{N_u}{n}\left(1 + \frac{\hat{\xi}}{\hat{\sigma}}(\varepsilon_t - u)\right)^{\frac{-1}{\hat{\xi}}}, \tag{9}$$

from which we get the *VaR* q_α,

$$q_\alpha = u + \frac{\hat{\sigma}}{\hat{\xi}}\left[\left(\frac{n}{N_u}\right)(1-\alpha)^{-\xi} - 1\right]. \tag{10}$$

2.3 Representation of Statistical Evidence Using Belief Function

2.3.1 Basics of Belief Function

In this section, we briefly present the theory of belief functions, which can be used as a formal framework for reasoning and making decisions under uncertainty. It originates from the work of [5, 6] who introduced a new approach to statistical inference based on lower and the upper probabilities induced by a multi-valued mapping. Later, Shafer [20] showed that belief functions can be used to represent uncertain information or data in a very general setting. This framework is usually referred to as Dempster-Shafer theory.

We can define a belief function on an arbitrary measurable space $(\Theta, \mathscr{B})$ as a function Bel satisfying the following conditions [20],

1. $Bel(\emptyset) = 0$, $Bel(\Theta) = 1$;
2. For any $k \geq 2$ and any collection $A_1, \ldots A_k$ of elements of $\mathscr{B}$,

$$Bel\left(\bigcup_{i=1}^{k} A_i\right) \geq \sum_{\emptyset \neq I \subseteq \{1,\ldots,K\}} (-1)^{|I|+1} Bel\left(\bigcap_{i \in I} A_i\right), \tag{11}$$

where $| I |$ is the cardinality of the set I.

When Θ is finite, we can represent the uncertain evidence about $\theta \in \Theta$ by a mass function m on Θ, defined as a function $m : 2^{\Theta} \longrightarrow [0, 1]$ such that $m(\emptyset) = 0$ and $\sum_{A \subseteq \Theta} m(A) = 1$. Any subset A of Θ such that $m(A) > 0$ is called a focal set of m. Shafer [20] interpreted each number $m(A)$ as a degree of belief attached specifically to the proposition $\theta \in A$ and to no more specific proposition. The function

$$Bel(A) = \sum_{B \subseteq A} m(B), \tag{12a}$$

for all $A \subseteq \Theta$ is then a belief function, and

$$Pl(A) = \sum_{B \cap A \neq \emptyset} m(B), \tag{12b}$$

is a plausibility function, related to Bel by the following relation: $Pl(A) = 1 - Bel(A^c)$ for all A, where A^c is the complement of A in Θ. The quantities $Bel(A)$ and $Pl(A)$can be interpreted, respectively, as the degree to which the evidence support A and the degree to which the evidence is not contradictory with A. The function $pl : \Theta \longrightarrow [0, 1]$ such that $pl(\{\Theta\}) = Pl(\{\theta\})$ for all $\theta \in \Theta$ is called the contour function associated to m. If the focal sets are nested, Bel is said to be consonant.

In the case where Θ is infinite, a belief function can generally not be defined from (12a) because a mass function may not exist. However, we can conveniently define a belief function from a multi-valued mapping Γ from a probability space $(S, \mathscr{A}, \mu)$ to 2^{Θ} such that, for all $B \in \mathscr{B}$, $\{s \in S | \Gamma(s) \cap B \neq \emptyset\}$ belongs to $\mathscr{A}$. We then have, for all $B \in \mathscr{B}$,

$$Pl(B) = \mu(\{s \in S | \Gamma(s) \cap B \neq \emptyset\})$$

and $Bel(B) = 1 - Pl(B^c)$. In the analysis using belief function described in the next section, the frame of discernment Θ is assumed to be a closed interval of the real line and the multi-valued mapping Γ will define a random closed interval, $\Gamma(s) = [U(s), V(s)]$. We then have

$$Bel(B) = \mu(\{s \in S \mid [U(s), V(s)] \subseteq B\}) \tag{13a}$$

$$Pl(B) = \mu(\{s \in S \mid [U(s), V(s)] \cap B \neq \emptyset\}), \tag{13b}$$

for all elements B of the Borel sigma-algebra $\mathscr{B}(\mathbb{R})$ [6].

2.3.2 Likelihood-Based Belief Function

In this section, we briefly recall how to represent the statistical evidence using belief function. Suppose that we observe a sample of random vector X with probability density function $f(x;\theta)$, where $\theta \in \Theta$ is an unknown parameter. According to the likelihood principle, all the information about θ is supposed to be contained in the likelihood function, which is $L(\theta; x) = f(\theta; x)$ for all $\theta \in \Theta$. Shafer [20] proposed to represent the information about Θ from the observed sample by a consonant likelihood-based belief function, whose contour function equals the normalized likelihood function:

$$pl(\theta; x) = \frac{L(\theta; x)}{\sup_{\theta' \in \Theta} L(\theta'; x)}, \tag{14}$$

assuming the denominator in (15) to be finite. Thus, the corresponding plausibility function can be defined as:

$$pl(A; x) = \sup_{\theta \in A} pl(\theta; x) = \frac{\sup_{\theta \in A} L(\theta; x)}{\sup_{\theta' \in \Theta} L(\theta'; x)}, \tag{15}$$

for all measurable subsets A of Θ. This method was shown to follow from the Likelihood and Least Commitment principles by Denoeux [7].

2.4 Application to Oil Price Value-at-Risk Estimation

In this section, we describe how to apply the belief function theory to represent the statistical evidence from the estimation of $VaR_\alpha(\varepsilon)$. In other words, the contour function of a consonant belief function with the normalized likelihood can be viewed as the estimation uncertainty of $VaR_\alpha(\varepsilon)$. The contour function is estimated in the belief function framework using a procedure similar to the one proposed in [2].

From the Sect. 2.2, we have $y \backsim G(\theta)$ where $G(.)$ is the GPD with the parameters $\theta = (u, \xi, \sigma) \in \Theta$. Assuming that $u = 0$ and $\xi \neq 0$, we have

$$G_{\xi,\sigma}(y) = 1 - \left(1 + \xi \frac{y}{\sigma}\right)^{\frac{-1}{\xi}}. \tag{16}$$

The VaR is the quantile of variable y, i.e., $VaR_\alpha = q_\alpha$. Therefore,

$$G(q_\alpha) = 1 - \left(1 + \xi \frac{q_\alpha}{\alpha}\right)^{\frac{-1}{\xi}} = \alpha. \tag{17}$$

Solving the above equation, we get

$$\sigma = \frac{\xi q_\alpha}{(1-\alpha)^{-\xi-1}}. \tag{18}$$

As the parameter of interest is q, Eq. (19) can be used to reparameterize the likelihood function. Thus the log-likelihood function becomes

$$L(\xi, q|y_t) = \left(\frac{\xi q}{(1-\alpha)^{-\xi-1}}\right)^{-n} \prod_{i=1}^{n} \left(1 + \left(\frac{y_t(1-\alpha)^{-\xi-1}}{q}\right)\right)^{\frac{-1-\xi}{\xi}} \tag{19}$$

The plausibility function $pl(\xi, q|y_t)$ can be derived from the likelihood function as

$$pl(\xi, q \mid y_i) = \frac{L(\xi, q \mid y_t)}{\sup_{\xi,q} L(\xi, q \mid y_t)}. \tag{20}$$

Since we are interested to learn about the plausibility of q, parameter ξ can be marginalized out by

$$pl(q \mid y_t) = \sup_{\xi} pl(\xi, q \mid y_t). \tag{21}$$

3 Data and Results

3.1 Data

In this section, we study the spot daily oil price of West Texas Intermediate (WTI). The data set of WTI covers the period from January 21, 1986 to November 21, 2013 (7025 observations). As shown in Fig. 1, the time series is not stationary after early 2008. To overcome the non-stationary issue, we computed the log-returns (lower graph of Fig. 1).

Figure 2 shows some descriptive statistics for the return of WTI time series of data. The data has a left-skewed shape and excess kurtosis. In addition, the Jarque-Bera test indicates that the returns are not normally distributed at the 99 % confidence interval.

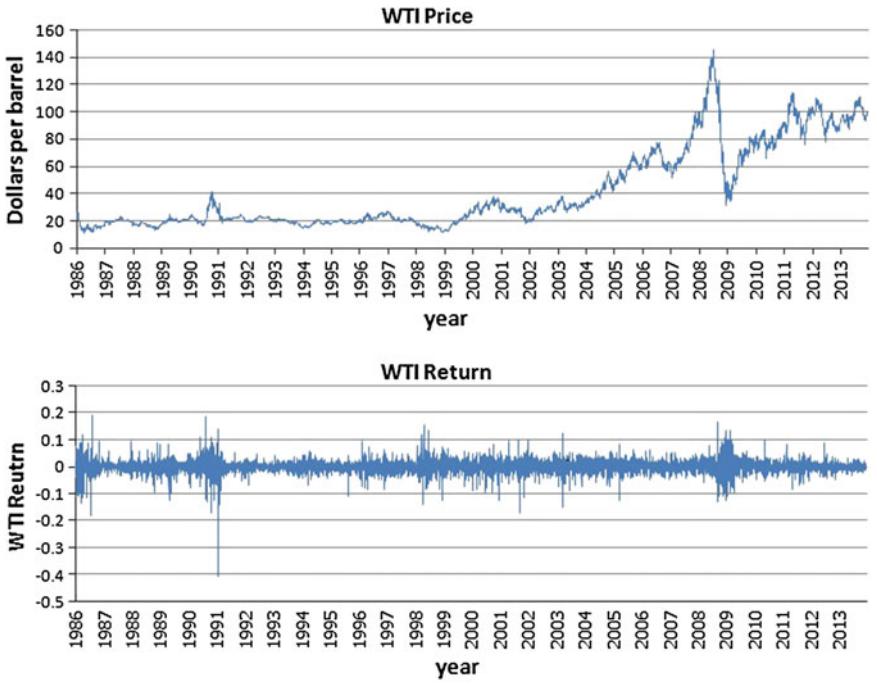


Fig. 1 Daily spot prices and return on WTI crude oil

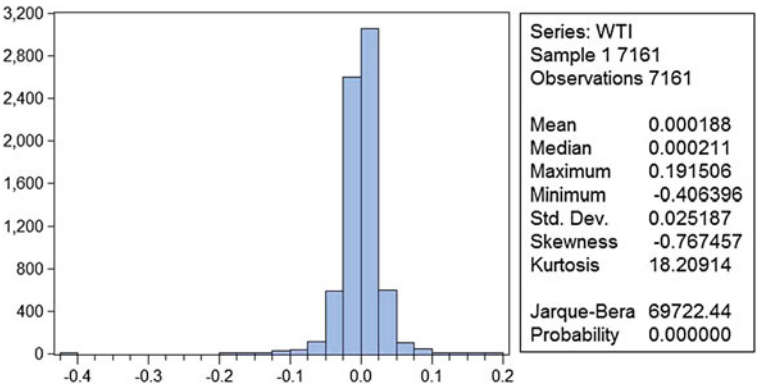


Fig. 2 Summary of Descriptive Statistic; daily return crude oil price of West Texas Intermediate (WTI)

3.2 Results

In this study, we examined both positive and negative extreme returns from investing in the WTI market. That is, we first applied the ARMA model to fit the deterministic

Table 1 Maximum likelihood parameter estimates for both return shocks

Parameter estimation	Positive return shock $u = 0.05$	Negative return shock $u = 0.10$
$\hat{\sigma}$	0.0197	0.0255
	(0.0025)	(0.0065)
$\hat{\xi}$	0.1290	0.2487
	(0.1002)	(0.1823)

Standard errors in parentheses

trend of the returns and computed the residuals from the model. Then, we applied the extreme value and belief function methods to model and estimate the positive and negative extreme shocks separately.

3.2.1 Deterministic Trend

The ARMA model requires the dependent variable or the return r_t to be stationary. In this part, we first tested stationarity using the augmented Dicky Fuller test. The test confirms that the return is stationary at the 1 % significant level.

For the WTI return data, the AIC and BIC statistics suggest that we use the ARMA model with the order 4 and 8. That is, the deterministic trend of the returns can be captured using the following ARMA(4,8) model,

$$\begin{aligned} r_{WTI} = {} & 0.0002 + 1.0957r_{t-1} - 0.5456r_{t-2} + 1.1445r_{t-3} - 0.8937r_{t-4} \\ & - 1.1106\varepsilon_{t-1} + 0.5154\varepsilon_{t-2} - 1.1326\varepsilon_{t-3} + 0.9312\varepsilon_{t-4} - 0.0261\varepsilon_{t-5} \\ & + 0.0201\varepsilon_{t-6} - 0.06212\varepsilon_{t-7} + 0.0480\varepsilon_{t-8}. \end{aligned}$$

3.2.2 Estimation of Value at Risk

The residuals from the ARMA model represent the return shocks. The extreme value model requires the shock variable to be independent and identically distributed (iid). Therefore, we tested the serial correlation using the Breusch-Godfrey Serial Correlation LM Test. The result suggests that there is no autocorrelation in the residuals. The estimates of the scale parameter σ and the shape parameter ξ of the GPD are shown in Table 1. The diagnostic plots for positive and negative shocks are shown in Figs. 3 and 4. For positive return shocks, none of the plots gives any cause for concern about the quality of fitted model. However, the goodness-of-fit in the probability plot for the fitted model of the negative return shock seems unconvincing. There is outlier that can also be seen in the return level plot.

As compared to the traditional point estimation of VaR_α, the belief function method provides us with the plausibility function of the VaR shown in Fig. 5. We can see that the plausibility is not symmetric. Therefore, information on the mean of

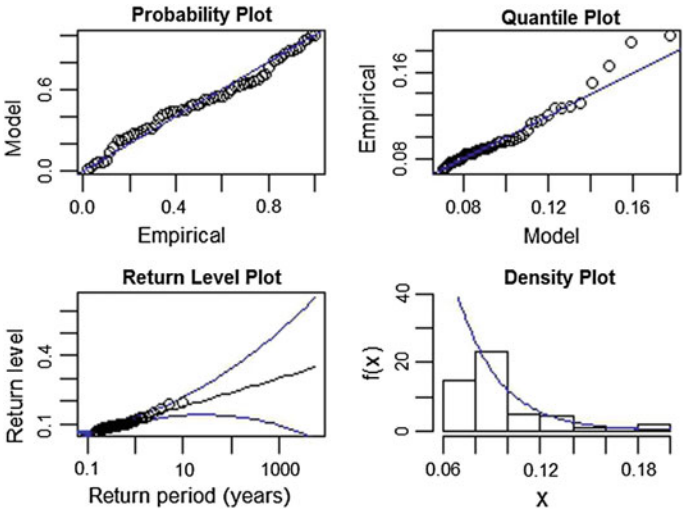


Fig. 3 Diagnostic plots for POT model fitted to positive return shock

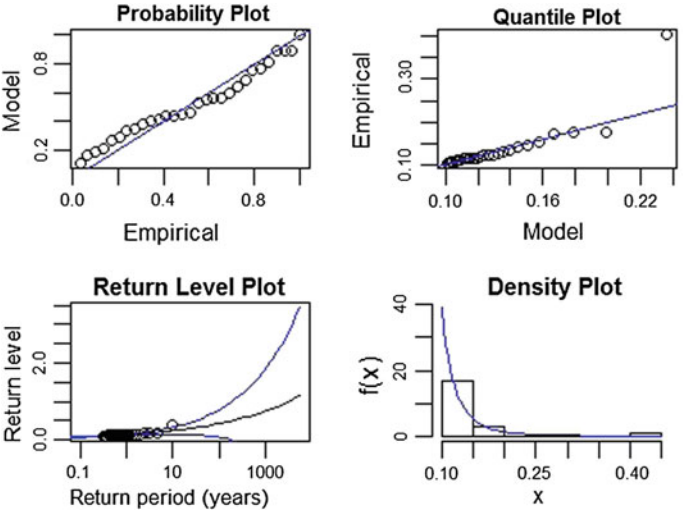


Fig. 4 Diagnostic plots for POT model fitted to negative return shock

VaR does not well represent the actual distribution. The most plausible *VaRs* at the 95 % level based on the statistical evidence, for positive and negative return shocks, are about 0.107 and 0.122, respectively.

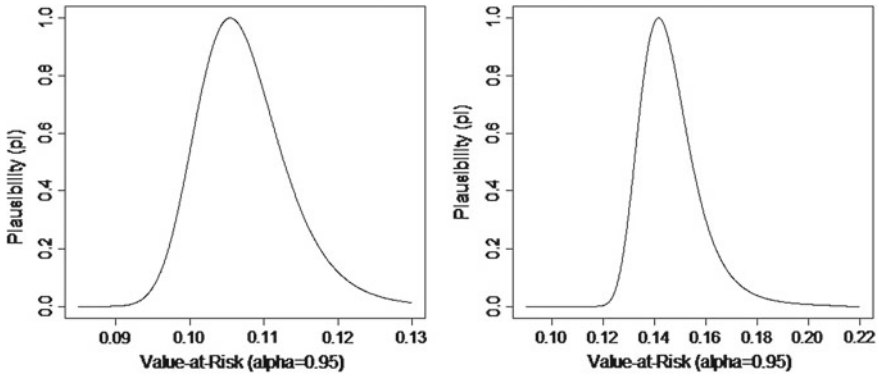


Fig. 5 Plausibility of for positive (*left*) and negative (*right*) return shocks

4 Conclusions

In this paper, we introduced a method to estimate the VaR of crude oil price. The method is based on three steps. First, the log-returns are fitted with an ARMA model. The extreme values of residuals computed from the POT method are then modeled by a GDP. Lastly, a belief function on the VaR is computed from the normalized likelihood function. This belief function quantifies estimation uncertainty and can be propagated in decision or forecasting procedures. In contrast with the Bayesian approach, the belief function method of inference does not require a prior probability distribution. However, the two methods coincide when a Bayesian prior is provided. The prediction of oil price could also be addressed using the same framework [13]. This topic is left for further research.

Acknowledgments The authors would like to thank Prof. T. Denœux, Prof. Dr. Hung T. Nguyen and Supanika Leurcharusmee for helpful comments and suggestions.

References

1. Balkema, A.A., De Haan, L.: Residual life time at great age. Ann. Probab. **2**, 792–804 (1974)
2. Ben Abdallah, N., Mouhous-Voyneau, N., Denoeux, T.: Combining statistical and expert evidence using belief functions: Application to centen nial sea level estimation taking into account climate change. Int. J. Approx. Reason. **55**(1), 341–354 (2014)
3. Birnbaum, A.: On the foundations of statistical inference. J. Am. Stat. Assoc. **57**(298), 269–306 (1962)
4. Brooks, C.: Introductory econometrics for finance. Cambridge University Press, Cambridge (2014)
5. Dempster, A.P.: Upper and lower probabilities induced by a multivalued mapping. Ann. Math. Stat. **38**(2), 325–339 (1967)
6. Dempster, A. P.:A generalization of bayesian inference. J. R. Stat. Soc. Ser. B (Methodol.) 205–247 (1968)

7. Denoe, T.: Likelihood-based belief function: Justification and some extensions to low-quality data. Int. J. Approx. Reason. **55**(7), 1535–1547 (2014)
8. Edwards, A.W.F.: Likelihood. Johns Hopkins University Press, Baltimore (1992) (expanded edition)
9. Embrechts, P., Klüppelberg, C., Mikosch, T.: Modelling extremal events: for insurance and finance. Springer, Berlin (1997)
10. Embrechts, P., Frey, R., McNeil, A.: Quantitative Risk Management, vol. 10. Princeton Series in Finance, Princeton (2005)
11. Enders, W.: Applied Econometric Time Series. Wiley, NewYork (2008)
12. Gilli, M.: An application of extreme value theory for measuring financial risk. Comput. Econ. **27**(2–3), 207–228 (2006)
13. Kanjanatarakul, O., Sriboonchitta, S., Denoeux, T.: Forecasting using belief functions: An application to marketing econometrics. Int. J. Approx. Reason. **55**(5), 1113–1128 (2014)
14. Krehbiel, T., Adkins, L.C.: Price risk in the nynex energy complex: an extreme value approach. J. Futur. Mark. **25**, 309–337 (2005)
15. Marimoutou, V., Raggad, B., Trabelsi, A.: Extreme value theory and value at risk: application to oil market. Energy Econ. **31**(4), 519–530 (2009)
16. McNeil, A.J.: Extreme value theory for risk managers. Working paper, ETH Zurich (1999)
17. Pickands, J.: Statistical inference using extreme order statistics. Ann. Stat. **3**, 119–131 (1975)
18. Reiss, R.D., Thomas, M.: Statistical Analysis of Extreme Values with Applications to Insurance Finance. Hydrology and Other Fields Birkhauser Verlag (1997)
19. Ren, F., Giles, D.E.: Extreme value analysis of daily canadian crude oil prices. Appl. Finan. Econ. **20**, 941–954 (2010)
20. Shafer, G.: A Mathematical Theory Of Evidence, vol. 1. Princeton University Press, Princeton (1976)
21. Shafer, G.: Allocations of probability. Ann. Prob. **7**, 827–839 (1979)
22. Tsay, R.S.: Analysis of Financial Time Series, vol. 543. Wiley, NewYork (2005)

Broad Monetary Condition Index: An Indicator for Short-Run Monetary Management in Vietnam

Pham Thi Tuyet Trinh and Nguyen Thien Kim

Abstract We construct broad monetary condition index (MCI) for monetary policy management in Vietnam. MCI is composed of key monetary transmission variables including interest rate, exchange rate, credit and stock market price. Weights of composite variables are derived from reduced form IS-PC framework and impulse response function based on vector autoregressive model with data in first difference form and difference-with-long-term-trend form. The best MCI is chosen based on three criteria: its causal relationship with output growth, its ability to explain output growth in short-run and its out-of-sample performance in forecasting output growth. Movement of chosen MCI indicates that the indicator has two essential characteristics of a supporting index for short-term monetary policy management, including quick responses to monetary policy changes and close relation with policy goal.

1 Introduction

To achieve policy goal of monetary policy, central banks construct some supporting targets in short run, known as operational target and intermediate target. In Vietnam, policy goal has been stated clearly to focus on inflation control under the Law on The State Bank of Vietnam 2010,[1] however, supporting targets for managing monetary to obtain policy goal have not been sufficient and transparent yet. Particularly, intermediate target is not officially announced though the State Bank of Vietnam (SBV) has relied mainly on annual objectives of money supply growth and credit growth as

[1] Previously, Vietnam has multiple policy goals including economic growth, price control, …, indicated in Law on The State Bank of Vietnam 1997.

P.T.T. Trinh (✉)
International Economics Faculty, Banking University, 39 Ham Nghi Street, District 1, Hochiminh, Vietnam
e-mail: trinhptt@buh.edu.vn

N.T. Kim
Faculty of Finance, Banking University, 39 Ham Nghi Street, District 1, Hochiminh, Vietnam
e-mail: kimnt@buh.edu.vn

V.-N. Huynh et al. (eds.), *Econometrics of Risk*, Studies in Computational Intelligence 583, DOI 10.1007/978-3-319-13449-9_27

Table 1 Target and actual money supply and credit growth

Year	Money supply growth (%)		Credit growth (%)	
	Target[a]	Actual[b]	Target[a]	Actual[b]
2000	–	24.50	–	34.19
2000	–	24.50	–	34.19
2001	–	21.1	–	30.4
2002	–	24	–	28
2003	–	20.6	–	26.2
2004	–	23.6	25	41.65
2005	22	29.65	25	31.1
2006	23–25	33.59	18–20	25.44
2007	23–25	46.12	17–21	53.89
2008	<32	20.31	<30	23.38
2009	25	28.99	21–23 25–27*	37.53
2010	25	33.3	25	31.19
2011	14–16	9.27	15–17	10.9
2012	14–16	18.46	15–17	8.85
2013	14–16	7.31**	12	4.72**

Note *Adjusted within fiscal year; **first 6 months
Source [a]SBV
[b]IFS

orientation for monetary policy management in fiscal year. Also, operational target is absolutely absent in monetary framework though short-run management aims to stabilize interest rate, exchange rate and ensure liquidity of the banking system, which is clearly indicated in monetary directives issued at the beginning of each fiscal year.[2] These limitations restrain the support and orientation of short-run target for monetary policy management, which have been revealed through the fact that: (i) Intermediate target does not quantitatively correlate with policy goal. SBV usually breaks the annual objectives of money supply and credit growth in other to achieve economic growth and inflation target (Table 1); (ii) Lack of operational target causes monetary policy slowly respond to market changes. Noticeably, interventions to control inflation have to wait signals from the response of money supply and credit or even price fluctuation. Therefore, the paper aims to compute Monetary Conditions Index (MCI) for Vietnam as a short-run indicator for SBV to partly fulfill the insufficiency of monetary framework.

MCI, formerly known as, is a weighted average of changes in values of interest rate and exchange rate to their relative values in a base period, which represents the impacts of these parameters to policy goal (usually output growth and inflation). In this paper, we follow the approach of MCI but modify some natures to make it

[2] Such as Directive 02/2010/CT-NHNN, Directive 01/2011/CT-NHNN, Directive 01/2012/CT-NHNN, Directive 01/2013/CT-NHNN.

appropriate with the characteristics of Vietnamese economy. In specific, beside interest rate and exchange rate, other key monetary transmission channels in Vietnam are incorporated, including real credit and stock price. Also, weights of composite variables are estimated by applying weighted-sum approach with two different methods: reduced form IS-PC framework and impulse response function (IRF) based on VAR.

2 Monetary Conditions Index (MCI)

MCI is firstly computed by Bank of Canada at the beginning of 1990s to reflect the state of monetary policy and used as a supportive indicator for central banks to manage monetary policy in short run. The idea for this index is based on transmission monetary mechanisms with two main channels including interest rate and exchange rate. MCI is defined as a weighted average of changes in values of interest rate (r) and exchange rate (e) to their relative values in a base period, in which, weights of interest rate and exchange rate express the impacts of these parameters on policy goal (usually output growth and inflation).[3]

$$MCI_t = \theta_r \cdot (r_t - r_0) + \theta_e \cdot (e_0 - e_t) \tag{1}$$

where MCI_t is MCI at time t, θ_r and θ_e are the relative weights of interest rate and exchange rate. The exchange rate is usually in logarithms or percent deviations from its base level while interest rate is at level. Both variables could be either in nominal (as MCI of Bank of Canada [3]) or real term (as MCI of Sveriges Riksbank [16]). In short-run, MCIs estimated from nominal and real terms have similar movement since relative prices and inflation rates are reasonably the same [10]. MCI decreases when interest rate decreases and exchange rate rises, reflecting contractionary monetary policy. MCI increases when interest rate increases and exchange rate falls, expressing expansionary monetary policy. Stable MCI addresses monetary policy is unchanged.

MCI can be used for various objectives including an operational target, an indicator and a monetary rule. In the first case, MCI is believed to be associated with long-run goals of monetary policy. For that reason, it is suggested that actual MCI should be brought in line with its desired value. This application is formerly employed by Bank of Canada from 1995 to 2006 and Reserve Bank of New Zealand from 1996 to 1999. As an indicator, MCI offers information about the level of monetary policy stance. In fact, when relatively calculated comparing to the previous period, MCI shows that policy has become tighter or looser. In other words, MCI acts as a leading indicator of policy stance. In the last case, MCI can be set to obtain a policy rule such as to correct deviations of inflation from objective and output from potential [2]. Norges Bank and Sveriges Riksbank have use MCI as an inflation target [10]. Apart from

[3] The formula is adjusted from the original formula (which is $MCI_t = \theta_r(r_t - r_0) + \theta_e(e_t - e_0)$) to appropriate with direct exchange rate quotation used in the paper.

central banks, MCI is also estimated by other organizations. For example, the index is constructed and reported in World Economic Outlook by International Monetary Fund (IMF) for policy evaluation of individual countries [18]. The application of MCI is also expanded into the business area. Financial Times showed MCIs' chart for four main EU countries accompanying with policy recommendations for Germany [7]. Goldman Sachs and JP Morgan discussed MCIs for a number of economies in their circulated publications [8, 31].

Since MCI is widely used, there have been many researches regarding the pros and cons of this index. Ericsson et al. considered MCI as an attractive indicator for two reasons. First, the concept of MCI is simple [11]. In fact, exchange rates and interest rates are two main transmission channels of monetary policy. Especially, in small open economies, exchange rate is one of the main factors influencing aggregate demand. Therefore, examining exchange rates and interest rates may be significant in policy making and economy's behavior understanding. The second attractive feature of this index is the simplicity in methodology of calculation. However, many researches criticize that MCI has limitations. Since the weights of variables included in MCI cannot be directly observed but empirically derived from a model, MCI is model dependent [10]. Similar to other empirical models, MCI's equation bears strong assumptions about parameter constancy, cointergration, dynamics, exogeneity and choices of variables. Moreover, MCI cannot capture the shocks affecting to the movements of monetary policy [4]. For these disadvantages, models employed to estimate MCI should be cautiously considered.

The robustness of this indicator is also suspected by its plainness. Since the index is constructed basing only on two transmission channels including interest rate and exchange rate, MCI does not fully reflect the effects of monetary policy on the economy. Recent theoretical and empirical research findings exhibit that property and equity prices play an important role in the transmission monetary mechanism through wealth effect [24] and credit channel [5]. For that reason, original MCI is modified by adding other parameters into calculation formula, representing for different transmission channels of monetary policy beside exchange rate and interest rate. The selecting parameters differ from country to country depending on the characteristics of particular economy. Goodhart and Hofmann applied coefficients summarized across lags to each contemporaneous component in constructing broad MCI for G7 countries including short term interest rate, real exchange rate, real property price and real share price [14]; Gauthier at el used the similar approach with housing price, share price and price differences among high yield bonds for estimating MCI of Canada [13]; Swiston employed a dynamic weight structure which accurately incorporates the timing of transmission from financial markets to real activity including bond price, stock price, exchange rate and credit availability [32]. In addition, original MCI is also modified by adjusting the formula to various form [14, 22, 23, 26]. These different ways of modification define broad version of MCI which is not unified as its original version but different to each other depending on specific conditions of each country.

Regarding to the methodology, literature addresses different approaches to estimate the weight of component variables of MCI. The most popular ones include

large-scale macro-econometric models, reduced-form of IS-PC model [1, 19, 27], impulse-response functions (IRFs) from a vector autoregression (VAR) [13, 32] and factor analysis [13]. The first approach is employed by national central banks and governmental institutions. According to Goodhart and Hofmann, large-scale macro-econometric models are obviously superior to reduced-form IS-PC model and VAR model since the structural features of the economy and the interaction among all variables are able to take into account [14]. However, due to data accessibility, this approach is considerably hard to apply by researchers. Whereas, reduced-form of IS-PC model and VAR model are widely used to estimate broad MCI. These two models also inherit potential limitations similar to the methodology applied to estimate original version of MCI such as model dependency, dynamics, parameter inconstancy and non-exogeneity of regressors [14]. However, some of these technical problems are recently defeated. For examples, Gauthier et al. employed the sum of coefficients on lags of variables and individual lags to take into account the dynamics of parameters [13]. In addition, generalized impulse response functions from a VAR and factor analysis are used to conquer the criticism of non-exogeneity of regressors and model dependency. Noticeably, because of not employing an empirical model, factor analysis approach does not face regarded limitations. Finally, high data frequency is adopted to avoid potential structural breaks and parameter inconstancy problem [13].

Although there are still limitations addressed in the estimated broad MCI, the index is still simultaneously studied and used for many different purposes including short term monetary policy. Besides, there is no perfect index or variable to support monetary policy operations in short term, MCI still be regarded as a considerable reference for central banks [12].

3 Methodology

3.1 Estimation Model

We employ formula (2) which is popularly used in recent researches (e.g. [22]) to calculate a broad MCI for Vietnam:

$$MCI_t = \sum_{i=1}^{n} w_i \cdot X_{i,t} \tag{2}$$

In which, X_i(i = 1, 2, 3 , . . .) includes composite variables of MCI, which also represent for key monetary transmission channels in Vietnam, including real interest rate (r), real exchange rate (e), real credit (cre), and stock price (vni) [6, 17, 21, 25, 29]; w_i is weight of variable X_i with sign (negative or positive) depending on impact direction of variable i on output.

Weights (w_i) of composite variables are estimated by applying weighted-sum approach with two different methods: reduced form IS-PC framework and impulse response function (IRF) based on VAR.

Following studies of Duguay [9], Goodhard and Hofmann [14] and Abdul Majid [1], reduced form IS-PC model includes an output equation (IS) and a Phillips curve equation (PC). The former has output (Y_t) as dependent variable and composite variables of MCI (X_i) as explanation variables. We also construct foreign output, world commodity price and fiscal policy as control variables which are represented by lag of output Y_{t-i} and dummy (Dum) representing for impact of global crisis from the third quarter of 2008. The latter describes relationship between output (Y_t) and price (π_t) with world commodity price (wcp_t) as control variable.

$$Y_t = \alpha_1 + \sum_{i=1}^{n} \beta_i \cdot Y_{t-i} + \sum_{j=1}^{n} \gamma_{i,j} \cdot X_{i,t-j} + Dum_t + \varepsilon_{1t} \tag{3}$$

$$\pi_t = \alpha_2 + \sum_{a=1}^{n} \sigma_{1a} \cdot \pi_{t-a} + \sum_{b=1}^{n} \sigma_{2b} \cdot Y_{t-b} + \sum_{c=0}^{n} \sigma_{3c} \cdot wpc_{t-c} + \varepsilon_{2t} \tag{4}$$

The above equations of IS-PC model are estimated separately by OLS. In IS equation, we apply 2 years as maximum lag length for all variables in the right hand-side because that is time horizon monetary policy is thought to have its full impact on output and inflation [13]. Lag length of each composite variable is specified by general-to-specific method. Weights of composite variables are calculated by adding all significant coefficients. We expect weights of composite variables as follows: (i) that of interest rate is negative (–), implying increase in interest rate as a result of tight monetary policy leads to decrease in output and vice versa; (ii) those of exchange rate, credit and stock price are positive (+), respectively implying increases in exchange rate, in credit and stock price as a result of expansionary monetary policy leads to increase in output and vice versa. In PC equation, maximum lag length is based on Akaike Information Criterion (AIC).[4]

Typical VAR model of Sims [30] is used to estimate weights of composite variable as follows:

$$X_t = C + \sum_{i=1}^{p} \Psi_i \cdot X_{t-i} + \theta \cdot Z_t + \varepsilon_t \tag{5}$$

In which, X_t represents for endogenous vector includes output, price, interest rate, exchange rate, credit and stock price $X_t = (Y, \pi, i, e, cre, vni)$; C is vector of constant; Ψ_i is matrix of auto-regressive coefficient; Z_t are exogeneous vector including US interest rate (ffr_t), world commodity price (wcp_t), world output ($wgdp_t$) represent for impact of world financial market, commodity market and foreign demand respectively; θ is coefficient matrix of exogenous variables. Lag length of model (5)

[4] We do not report estimation result of PC equation in the main contend of this paper.

is determined by AIC. Weight of each composite variable is calculated by accumulated response of output to its shock within two years. We use generalized impulse response function developed by Pesaran and Shin [28] since the result of estimated response is not sensitive to order of variables in the VAR system.

3.2 Variables Definition and Data Description

We use two forms of data including first difference specification and deviation-from-long-term-trend specification (hereafter gap specification) for estimation. With the two specifications of data, variables in reduced form IS-PC model and VAR model are defined as in Table 2.

Output (Y_t) is proxied by real gross domestic production (GDP) of Vietnam in index form from Data Stream. Real deposit interest rate is employed as a proxy of interest rate (r_t). This selection relies on four reasons. Firstly, deposit rate is one of the three interest rates (along with lending rate and interbank rate) expressed in the Annual Statement of SBV. Secondly, interbank rate is not a good indicator for money market in Vietnam [33]. In SBV's Statements, interbank rate is only mentioned from 2008. Thirdly, lending rate of commercial banks is determined mainly on deposit rate. Finally, although ceiling on deposit rate is applied from March 2011, it aims to control the race of raising deposit rate among commercial banks during the period but not to distort the market. Real interest rate is obtained by subtracting inflation rate from nominal interest rate taken from IFS. Exchange rate (e_t) is proxied by real effective exchange rate (REER) of Vietnam with 17 main trading partners[5] accounting for about 90 % annual total foreign trade of Vietnam. REER is calculated by geometric mean method[6] with data taken from various sources including: (i) nominal VND/USD rate is from SBV; (ii) nominal rate of USD against other foreign currency, price index of foreign country are from IFS; (iii) bilateral foreign trades of Vietnam and 17 main trading partners are from General Statistic Office of Vietnam. Credit (cre_t), proxied by real credit of the economy, price of Vietnam (π_t), proxied by consumer price index, world commodity price (wcp_t), proxied by world commodity price index and fed fund rate (ffr_t) are from IFS. Stock price (vni_t) is proxied by Stock price index of Hochiminh Stock Exchange (Vnindex). World output ($wgdp_t$) is calculated as weighted average of real GDP volume in index form of 17 main trading partners with Vietnam taken from IFS.

For data in gap specification, output, interest rate, exchange rate, credit, stock price are percentage deviation of their actual levels from their long-term-trend levels

[5] Including China, Singapore, Japan, Korea, Thailand, Malaysia, Hong Kong, The United State, Indonesia, Germany, Australia, UK, France, Russia, Philippines, Taiwan and Netherland.

[6] $REER_t = \prod_{i=1}^{k}(NER_{it} \cdot \frac{P_{it}^*}{P_n})^{w_{it}}$ in which NER_i is nominal exchange rate of currency i against VND, w_i represent for attached trade weigh of currency *i in currency basket*, P_i^* represents for producer price index or whole sale price index of country i; P_n represent for consumer price index (CPI) of Vietnam.

Table 2 Variable definitions and data sources

Variable		Proxy/calculation	Source	Data specification	
				First difference	Gap
Output	Y	Real GDP	Data stream	$Ln(Y_t) - Ln(Y_{t-1})$	$Ln(Y_t) - Ln(Y\_hp)$
Interest rate	r	Real deposit rate = nominal deposit rate inflation rate	IFS	$Ln(r) - Ln(r_{t-1})$	$r_t - r\_hp_t$
Exchange rate	e	REER calculated by geometric mean method	IFS, GSO, SBV	$Ln(e_t) - Ln(e_{t-1})$	$Ln(e_t) - Ln(e\_hp_t)$
Credit	cre	Real credit	IFS	$Ln(cre_t) - Ln(cre_{t-1})$	$Ln(cre_t) - Ln(cre\_hp_t)$
Stock price	vni	Vnindex	HOSE	$Ln(vni_t) - Ln(vni_{t-1})$	$Ln(vn_t) - Ln(vn\_hp_t)$
Vietnam price	π	Consumer price index	IFS	$Ln(\pi_t) - Ln(\pi_{t-1})$	Year-over-year inflation
Fed fund rate	ffr	3 month fed fund rate	FED	ffr (% PA)	ffr (% PA)
World output	$wgdp$	Weighted average of real GDP of 17 main trading partners with Vietnam	IFS, GSO	$Ln(wgdp_t) - Ln(wgdp_{t-1})$	Year-over-year growth rate
World commodity price	wcp	World commodity price index	IMF	$Ln(wcp_t) - Ln(wcp_{t-1})$	Year-over-year inflation

Note Ln denotes series data in logarithm; HP denotes series data derived by HP filter
Source Authors

which are calculated by Hodrick Prescott filter (HP) with smoothing parameter of 1600. Meanwhile, Vietnam price and world commodity price, world output are year-over-year inflation and growth rate.

We use quarterly data from the first quarter of 2000 to the second quarter of 2013 for estimation. Research period is chosen depending mainly on data availability. In addition, this is also strong integration period of Vietnam economy into the world economy through bilateral and multilateral trade agreements. Besides, SBV has issued new regulation for exchange rate quotation at commercial banks with Decision 64/1999/QD/NHNN and 65/1999/QD/NHNN since 1999. All data series (except for interest rate and fed fund rate) are in natural logarithm and seasonally adjusted by Census X12. Thus, we have two data sets (first difference specification and gap specification) depicted in Figs. 1 and 2.

Two sets of data series are also tested stationarity by Augmented Dickey-Fuller (ADF) v Phillips-Perron (PP). The results reported in Table 3 indicate that, in first difference specification, all series, except for output and credit, are stationary significantly at 1 and 5 % according to both ADF and PP. Output and credit are not stationary according to ADF but stationary significantly at 1 % according to PP. We accept these

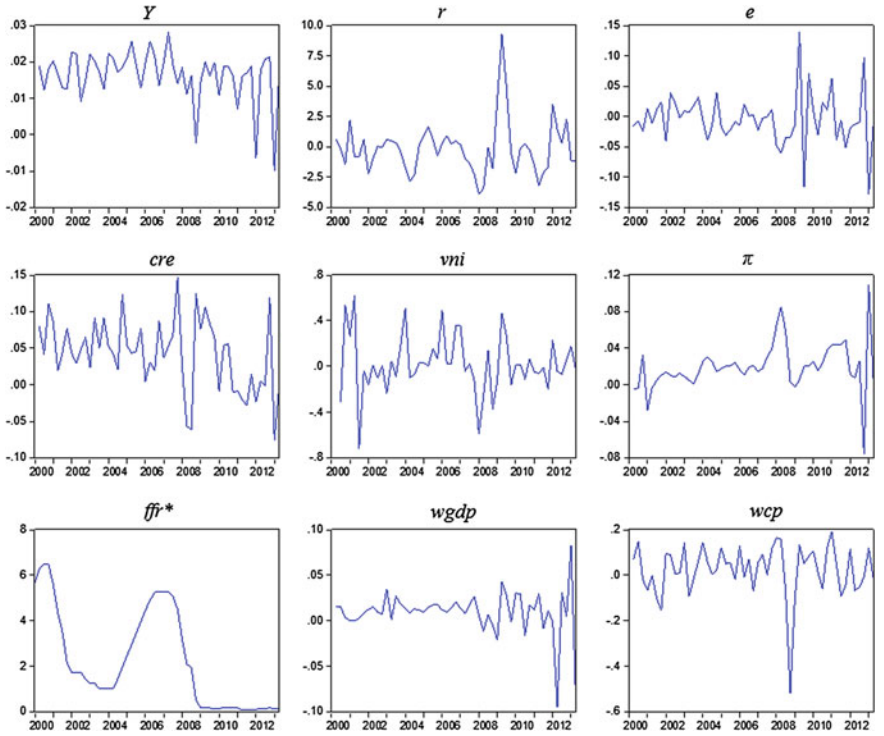


Fig. 1 Series data in first difference specification. *Note* * denotes data in percentage per annum. *Source* Authors' calculation

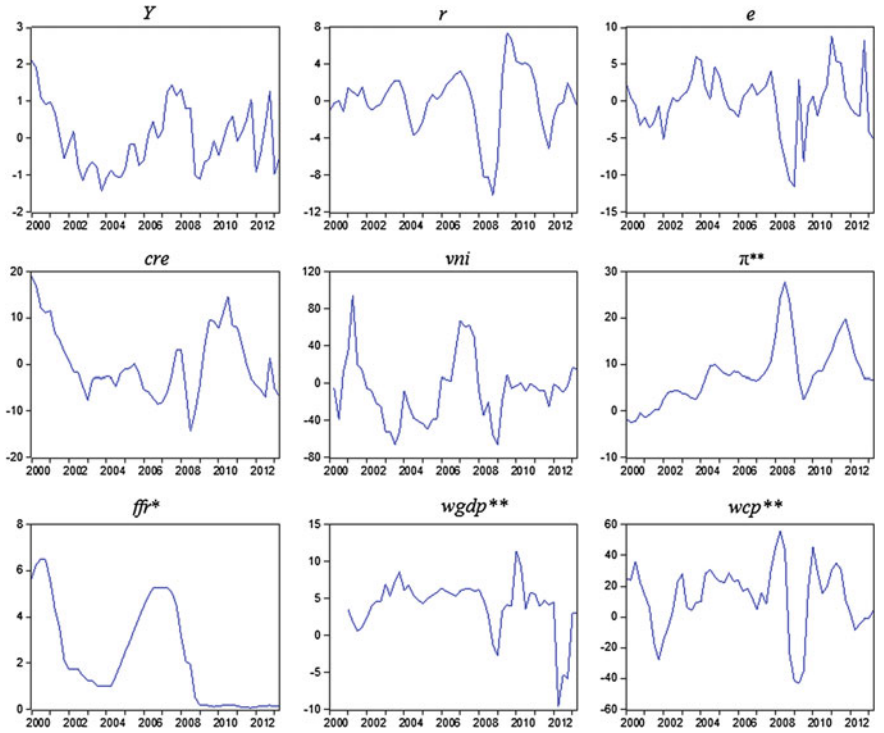


Fig. 2 Series data in gap specification. *Note* * and ** denote series data in percentage per annum and in year-over-year growth rate respectively. *Source* Authors' calculation

Table 3 Result of ADF and PP stationary test

Variable	First difference specification		Gap specification	
	t-stat, ADF test	t-sat, PP test	t-stat, ADF test	t-stat, PP test
Y	−2.111	−7.7917***	−3.0859**	−3.4659**
R	−4.6743***	−3.9213***	−4.7679***	−2.8368**
E	−9.9298***	−10.0537***	−4.2027***	−4.2875***
cre	−2.3111	−5.5715***	−3.1811**	−2.8096**
vni	−5.9697***	−5.9554***	−2.7229*	−2.7229*
π[b]	−6.9940***	−7.0403***	−2.4596	−1.8938
ffr[a]	−3.2625**	−3.4669**		
wcp[b]	−6.4737***	−6.4609***	−5.0722***	−3.0536**
wgdp[b]	−8.7765***	−8.8595***	−3.2581**	−3.2729**

Note ***, **, * indicate significance at 1, 5 and 10 % respectively;
[a] Data in percentage per annum %;
[b] Data in year-over-year growth rate in gap specification
Source Authors' calculation

two data series are stationary. In gap specification, all data series (except for price) are stationary significantly at 10 % and above. Price is not stationary according to both ADF and PP, therefore, we take first difference form of year-over-year inflation for estimation in gap specification.

4 Weight Estimation Results

With two estimation models and two data specifications, we generate four weight groups called as in Table 4.

4.1 Estimated Weight from Reduced Form IS-PC Model

Table 5 shows estimated results of IS equation with data in first difference specification (column 1 and 2) and in gap specification (column 3 and 4). Diagnostic tests also indicate the significance of estimated results (Residual has normal distribution and stability, no serial correlation, no heteroskedasticity).

With first difference specification, coefficients of output are significant at lag 1, 2, 3 and 4; those of interest rate are significant at lag 3, 4 and 5; those of exchange rate are significant at lag 2 and 6; those of credit are significant at lag 3 and 5; those of stock price is at lag 1, 3, 4 and 6; financial crisis has negative impact on output. Thus, the summations of significant coefficients of interest rate, exchange rate, credit and stock price are −0.035, 0.081, 0.086 and 0.035 respectively. All coefficient summations have expected signs indicating decrease in interest rate, increase in exchange rate, credit and stock price reflects monetary expansion and lead to increase in output while increase in interest rate, decrease in exchange rate, credit and stock price reflects monetary tightening and cause output to decrease. With gap specification, output has significant impact at lag 4 and 6; interest rate has significant impact at lag 4; exchange rate has significant impact at lag 2; credit has significant impact at lag 4 and 5; stock price has significant impact at lag 1 and 3; financial crisis also has significant negative impact on output. Summations of significant coefficient of interest rate, exchange rate, credit and stock price are −0.059, 0.023, 0.008 and 0.019 respectively and have expected sign.

Table 4 Weight groups

Variable	First difference specification	Gap specification
Specification	Reduce form IS-PC model	IRF based on VAR model
First difference	Group 1	Group 3
Gap	Group 2	Group 4

Source Authors

Table 5 Estimated result of IS equation

First difference specification		Gap specification	
Variables (1)	Coefficients (standard error) (2)	Variables (3)	Coefficients (standard error) (4)
C	0.0374*** (0.0089)	C	−0.0621 (0.0722)
Y_{t-1}	−0.4436*** (0.1486)	cre_{t-1}	0.0231** (0.0099)
vni_{t-1}	0.0121*** (0.0037)	vni_{t-1}	0.0114*** (0.0017)
Y_{t-2}	−0.5475*** (0.1542)	e_{t-2}	0.0235* (0.0151)
e_{t-2}	0.0418** (0.0208)	vni_{t-3}	0.0087*** (0.0022)
Y_{t-3}	−0.5745*** (0.1511)	Y_{t-4}	0.5057*** (0.0968)
i_{t-3}	0.1107** (0.0439)	i_{t-4}	−0.0590** (0.0255)
cre_{t-3}	0.0483** (0.0191)	cre_{t-4}	−0.0463** (0.0178)
vni_{t-3}	0.0111*** (0.0035)	cre_{t-5}	0.0317* (0.0159)
Y_{t-4}	0.2465* (0.1557)	Y_{t-6}	−0.7009*** (0.0913)
i_{t-4}	−0.1983*** (0.0541)	Dum	−0.3289*** (0.1178)
vni_{t-4}	0.0071** (0.0032)		
i_{t-5}	0.1356** (00511)		
cre_{t-5}	0.0380* (0.0209)		
i_{t-6}	−0.0044304		
e_{t-6}	0.0407* (0.0219)		
vni_{t-6}	0.0050* (0.0030)		
Dum	−0.0069** (0.0026)		
R^2	0.7959	R^2	0.8514
$Ad.R^2$	0.6719	$Ad.R^2$	0.8113
Loglikelihood	197.2829	*Loglikelihood*	−9.3392
F-statistic(*Prob*)	6.4213 (0.0000)	*F-statistic*(*Prob*)	21.2130 (0.0000)
AIC	−7.7949	*AIC*	0.8474
SBC	−7.0793	*SBC*	1.2762
DWstat	2.4717	*DWstat*	2.1863

Note ***, **, * indicate significance at 1, 5 and 10 % respectively
Source Authors' calculation

4.2 Estimated Weight from IRF Based on VAR Model

Table 6 shows accumulated impulse responses of output to shocks of interest rate, exchange rate, credit and stock price with data in first difference and gap specifications. Output has expected responses to shocks (negative response to interest rate shock and positive responses to shocks of exchange rate, credit and stock price).

Table 6 Accumulated impulse response of output to interest rate, exchange rate, credit and stock price shock

Period	R	e	cre	vni
		First difference	Specification	
1	−0.000562	0.002041	0.002688	−0.000375
2	−0.001609	0.001846	0.004072	0.000189
3	−0.000917	0.002899	0.005495	0.000497
4	−0.001312	0.005742	0.008229	0.000848
5	−0.001356	0.005329	0.009389	0.000546
6	−0.001366	0.004388	0.009254	0.004867
7	−0.001468	0.003569	0.008945	0.003172
8	−0.001475	0.002886	0.008316	0.00114
		Gap specification		
1	−0.143242	0.121949	0.159473	0.1581645
2	−0.316881	0.107188	0.272079	0.417559
3	−0.457483	0.172672	0.317857	0.554801
4	−0.57895	0.292024	0.409699	0.672792
5	−0.701296	0.312977	0.502208	0.79685
6	−0.817582	0.305276	0.550239	0.871414
7	−0.916458	0.325827	0.573775	0.900803
8	−0.977595	0.311965	0.588562	0.917666

Note Generalized impulse response function based on VAR estimation with 3 lag chosen by AIC
Source Authors' calculation

Table 7 Normalized weight groups

	Group 1	Group 2	Group 3	Group 4
r	−1	−7.38	−1	−3.14
e	2.31	2.88	2	1
cre	2.46	1	8	1.89
vni	1	2.38	2	2.95
MCI	MCI1	MCI2	MCI3	MCI4

Source Authors' calculation

After 8 quarter, accumulated responses of output to the four shocks in first difference specification are −0.001475, 0.002886, 0.008316 and 0.001140 respectively and in gap specification are −0.977595, 0.311965, 0.588562 and 0.917666 respectively. Diagnostic and stability tests of estimated VAR models also indicate estimated results are stable and significant.

With four estimated weight groups, we process to normalize and obtain results reported in Table 7. Four normalized weight groups are different with each other, reflecting differences in method and data specifications for estimation.

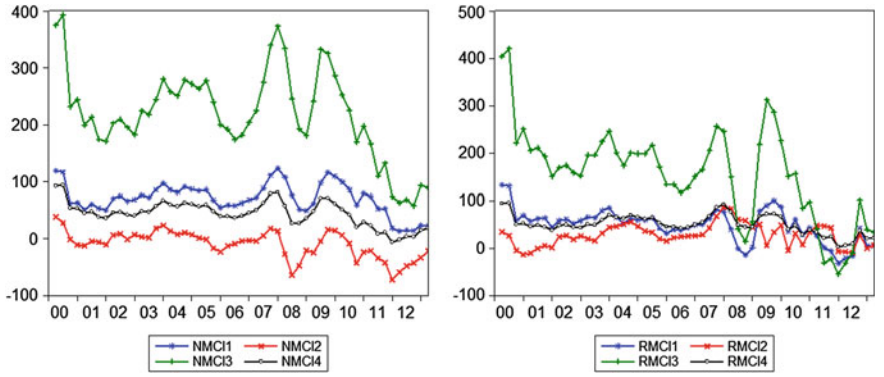


Fig. 3 Calculated nominal and real MCIs. *Source* Authors' calculation

Gauthier et al. [13][7] and Osorio et al. [26][8] also use various methodologies and data forms and find different weights of composite variables to calculate MCI. Among weight groups, group 1 and 3 has weight of credit as the largest, that of exchange rate as the smaller, and those of interest rate and stock price as the smallest; meanwhile, group 2 and 4 has weight of interest rate as the largest, that of exchange rate (of group 2) and stock price (of group 4) as the smaller, stock price and credit (of group 2) or credit and exchange rate (of group 4) as the largest.

5 MCI Calculation Results

With four normalized weight groups in Table 6, we calculate four nominal MCIs (NMCI) and four real MCIs (RMCI) by using formula 1. NMCIs are calculated by nominal data of interest rate, exchange rate, credit and stock price while RMCIs are calculated by their real data. We calculate both NMCIs and RMCIs because Vietnam is high inflation economy with annual average inflation rate at 7 % in the whole period (2000–2013), which can lead to different movements of nominal and real MCIs in both short and long term. Figure 3 depicts movements of calculated NMCIs and RMCIs indicating all NMCIs have rather similar movements. However, fluctuation of NMCI3 is much larger than that of the other NMCIs due to largest weight of credit, causing NMCI3 to separate from the rests. This is also the difference of RMCI3 with the other RMCIs. Figure 4 indicate that pairs of MCI1, MCI3, MCI4 are highly correlated while correlation of NMCI2 and RMCI2 is very low.

[7] The authors use three methods including: reduced-form IS-PC model, IRF based on VAR and factor analysis with data in first difference form and difference-with-long-term form.

[8] The authors use IRF based on VAR and Dynamic factor model to calculated MCIs for 13 economies including China, Australia, HongKong, Indonesia, India, Japan, Korea, Malaysia, New Zealand, Philippines, Singapore, Thailand and Taiwan.

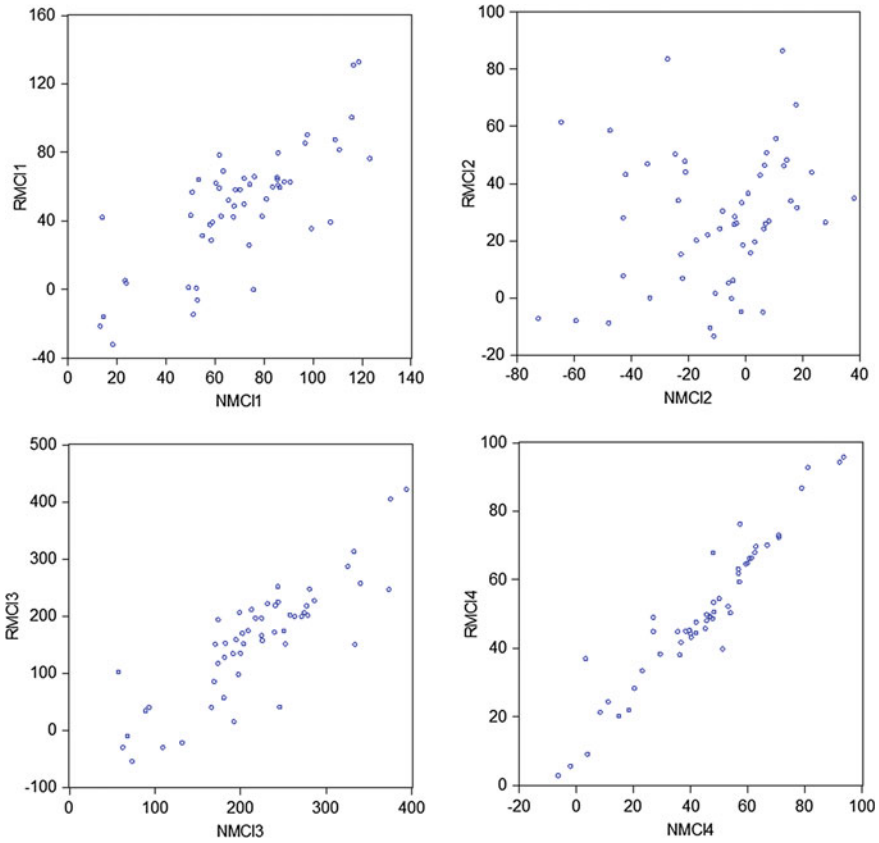


Fig. 4 Correlation between MCI pairs. *Source* Authors' calculation

6 MCIs Appraisal

6.1 Criteria of Appraisal

MCI can be used to support short-term monetary policy management, thus the index must have close relation with policy goals. We rely on three criteria which are also three various tests to appraise computed MCI and select the best MCI among various MCIs. The first is causal relationship of MCIs with output growth based on Granger Causality Test. MCI found to be no Granger cause with output growth is not appropriate for monetary policy management and will be eliminated in other method appraisal. The second is ability to explain for output growth in short run. We follow Guamata et al. [15], Osorio et al. [26] to use below simple Eq. (6).

$$growth_{t+h} = \alpha + \sum_{i=1}^{q} \beta_i \cdot growth_{t+1-i} + \delta \cdot INDEX_t + \varepsilon_t \quad (6)$$

In which, *growth* is real output growth of Vietnam economy; h is forecast horizon and is set at 1, 2, 3, and 4 quarters ahead; i is optimal lag length of growth based on AIC[9] which indicates 3 lag; *INDEX* represents for MCIs and individual monetary variables[10] including real interest rate, real exchange rate, real credit and stock price. Equation (6) implies that when coefficient δ is different from zero significantly and has expected sign, *INDEX* can explain for output growth at h period ahead. In order to compare explaining ability of MCIs and monetary variables, we rely on Sum Squared Residuals (SSR) of estimated equations. The third is ability of MCIs to predict output growth in the short-run by out-of-sample forecast. In this forecasting exercise, we use Eq. (6) again and perform 1, 2, 3 and 4 steps-ahead forecast for the period of 2010Q1–2013Q2. Forecast results are compared with actual output growth to judge predictive power of MCIs by Root Mean Square Error (RMSE).

6.2 Appraisal Results

Table 8 presents results of Granger causality test, indicating NMCI1, NMCI3, RMCI2, RMCI4 are not Granger cause of output growth while NMCI2, NMCI4, RMCI1, RMCI3 are Granger cause of output growth. Therefore, we eliminate NMCI1, NMCI3, RMCI2, RMCI4 in later tests. The results also show output growth is not causality of all calculated MCIs.

Table 9 indicates that estimated coefficients of all INDEX (except interest rate and credit) at 1, 2 and 3 step ahead have expected signs significantly. Coefficients of interest rate are significantly positive (which is contrary to theory) while coefficients of credit are not significant. Comparing SSRs of estimated equations, we find that: (i) between MCIs and monetary variables, SSRs of the former are smaller than those of the latter, implying MCIs, an indicator combining variables representing for transmission channels of monetary policy, have more powerful ability to explain for output growth than monetary variables; (ii) among MCIs, SSRs of RMCI are smaller than those of NMCIs. This result also reflects economic decisions are usually made based on real factors; and (iii) among RMCIs, SSR of RMCI1 is smaller than that of RMCI3, implying RMCI1 is the most powerful index in explaining output growth.

Figure 5 shows actual output growth and forecasted output growth at 1, 2, and 3 steps ahead[11] in period of 2010Q1–2013Q2. All forecasted output growths (growthf) have downward trend in the period after financial crisis, which is also key trend

[9] We determine lag of growth without *INDEX* to apply the same lag of growth into various equation estimation, therefore, SSRs of estimated equation can reflect explaining ability of *INDEX*.

[10] We also explore explanatory ability of other monetary variables to compare with that of calculated MCIs.

[11] We do not forecast at four steps ahead because all coefficient of MCIs at foure steps ahead are not significant as presented in Table 8.

Table 8 Granger causality test results

Null hypothesis	F-statstistic			
	Lag 1	Lag 2	Lag 3	Lag 4
NMCI1 does not Granger cause GROWTH	0.8103	1.1742	1.9047	2.4097*
GROWTH does not Granger cause NMCI1	0.4522	0.2778	0.1688	0.17
NMCI2 does not Granger cause GROWTH	8.2102***	4.6271**	3.3302**	5.0897***
GROWTH does not Granger cause NMCI2	0.0058	0.9293	0.6709	1.3511
NMCI3 does not Granger cause GROWTH	1.201	1.4065	1.7241	1.5624
GROWTH does not Granger cause NMCI3	0.3452	0.2028	0.1772	0.2024
NMCI4 does not Granger cause GROWTH	4.4717**	2.4074*	2.2278*	2.5219*
GROWTH does not Granger cause NMCI4	0.0209	0.3737	0.2954	0.3149
RMCI1 does not Granger cause GROWTH	4.5304**	2.5062*	3.5207**	3.7632**
GROWTH does not Granger cause RMCI1	0.0819	0.4555	0.5799	0.5137
RMCI2 does not Granger cause GROWTH	2.0193	1.7083	1.4895	1.7959
GROWTH does not Granger cause RMCI2	3.0869**	2.3470*	2.0461	1.3494
RMCI3 does not Granger cause GROWTH	5.9215**	3.0208*	3.1613**	3.1722**
GROWTH does not Granger cause RMCI3	0.4319	0.0848	0.2662	0.1972
RMCI4 does not Granger cause GROWTH	0.9994	0.5626	1.2579	1.1701
GROWTH does not Granger cause RMCI4	0.5742	0.6751	0.9488	0.9936

Note ***, **, * indicate significance at 1, 5 and 10 % respectively
Source Authors' calculation

Table 9 Ability to explain output growth of MCIs and monetary variables

INDEX	h = 1		h = 2		h = 3		h = 4	
	δ	SSR	δ	SSR	δ	SSR	δ	SSR
*NMCI*2	0.022*** (0.007)	35.938	0.034*** (0.008)	47.538	0.023** (0.009)	62.978	0.005 (0.010)	65.11
*NMCI*4	0.017** (0.008)	39.232	0.021** (0.010)	57.301	0.010 (0.011)	65.32	−0.003 (0.011)	65.474
*RMCI*1	0.013** (0.005)	35.122	0.022*** (0.005)	47.255	0.011* (0.006)	62.534	0.001 (0.007)	65.529
*RMCI*3	0.005*** (0.001)	36.428	0.007*** (0.002)	48.873	0.004* (0.002)	61.528	0.0009 (0.002)	65.402
r	0.114*** (0.034)	34.233	0.182*** (0.037)	40.534	0.130*** (0.044)	54.632	0.048 (0.048)	63.994
e	−3.616 (3.379)	42.329	7.791* (4.466)	59.409	9.256* (4.873)	61.115	1.089 (5.122)	65.534
cre	0.013 (0.014)	42.62	0.006 (0.017)	63.589	−0.011 (0.018)	66.031	−0.022 (0.019)	63.455
vni	0.694** (0.288)	38.213	1.171*** (0.331)	48.9096	1.031*** (0.357)	55.163	0.593 (0.383)	61.823

Note ***, **, * indicate significance at 1, 5 and 10 % respectively
Source Authors' calculation

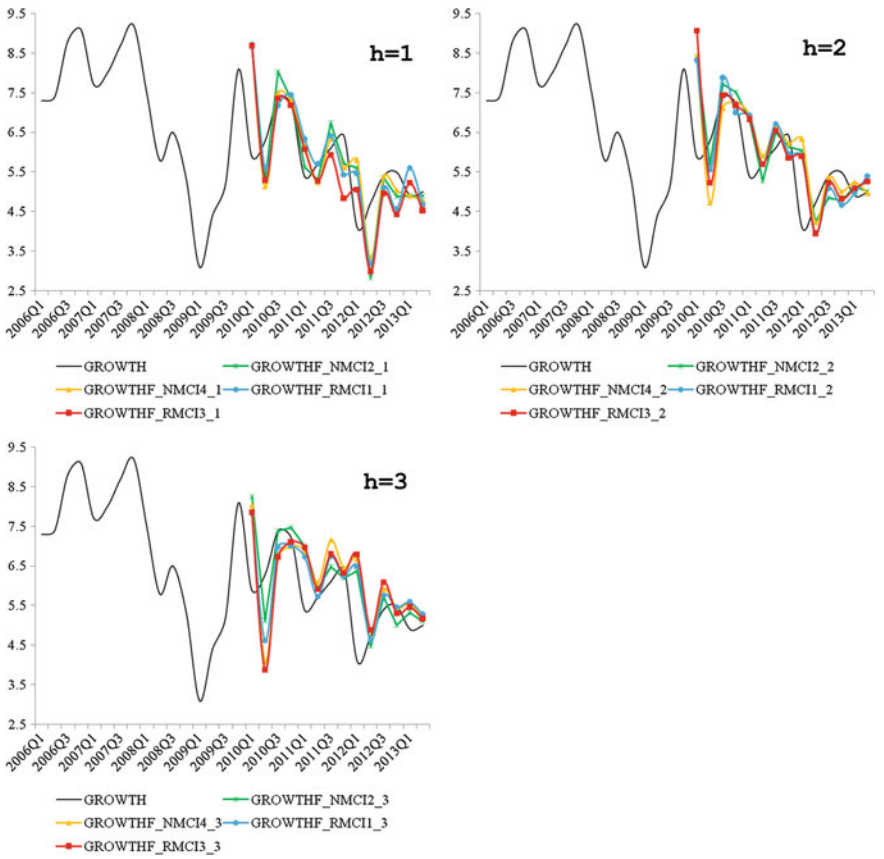


Fig. 5 Actual output growth and forecasted output growth at 1, 2, and 3 step-ahead. *Note* growthf denotes forecasted output growth. *Source* Authors' calculation

of actual growth output. Although forecast results are not quite coincident with actual performance, all forecasts have similar movements with actual output growth. Predictive powers of MCIs reported in Table 10 show RMCI1 has smallest RMSE, implying forecast results of model with RMCI1 is closer with actual output growth than those of the others.

In summary, results of three tests indicate RMCI1 outperforms the others in explaining and forecasting output growth in short run. Therefore, we choose RMCI1 to analyse for the purpose of exploring how the index performs in comparison with monetary policy management in studied period.

Table 10 Predictive power of MCIs

MCI	RMSE		
	h = 1	h = 2	h = 3
NMCI2	1.0821	1.0096	1.0547
NMCI4	1.0618	1.1104	1.2184
RMCI1	1.0673	1.0086	1.0494
RMCI3	1.1157	1.1373	1.2313

Source Authors' calculation

7 Analysis of MCI Evolution

In studied period, MCI evolution can be divided into five phases corresponding to five periods of monetary policy management performed by the changes in policy rates including rediscount rate, refinancing rate, base rate, and by movements of intermediate target variable (money supply growth), and leading to appropriate response of the economy through the changes of output growth and inflation (Figs. 6 and 7).

In 2000Q1–2001Q4, MCI fluctuated tremendously reflecting constant changes of monetary policy in very short time of the post-crisis period with volatile policy rates. Rediscount rate and refinancing rate declined to 4.2 % (from 4.8 % previously) and 4.8 % (from 5.4 % previously) in 8/2000, then rebounded to 5.4 and 6 % in 11/2000 respectively. Again, the rates respectively decreased by 4.8 and 5.4 % in 4/2001 while refinancing rate continued reducing to 4.8 % in 7/2001. Consequently, the economy spent hard time with low output growth and deflation. In 2002Q1–2007Q4, MCI tended to move upward, showing consistent expansion of monetary policy to pursue output growth target, which was mainly implemented by:

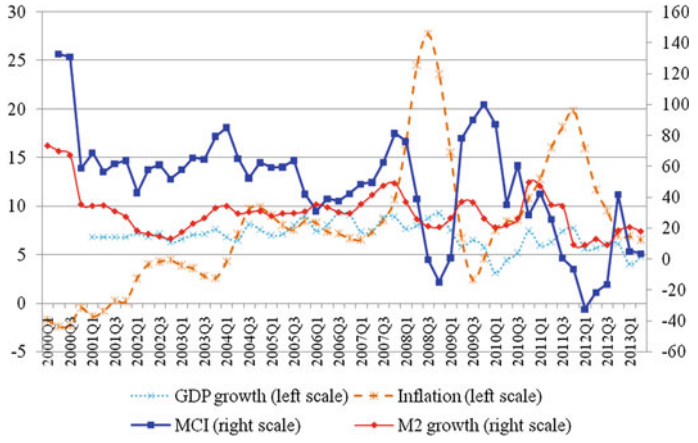


Fig. 6 Evolution of MCI, M2 growth and output growth, inflation. *Source* Authors' estimation (MCI); Data stream (GDP growth); IFS (Inflation, M2 growth)

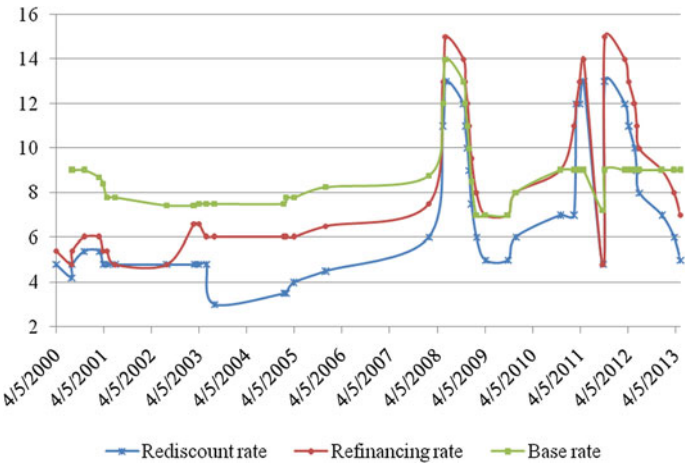


Fig. 7 Policy rates in 2000–2013. *Source* SBV

(i) keeping policy rates to be stable. Rediscount, refinancing and based rate were upward adjusted only three times within 2005 to 4.5, 7.5 and 8.25 % respectively to curb high inflation pressure in 2004 and stayed unchanged for the whole phase; (ii) keeping reserve requirement unchanged. SBV only raised reserve requirement twice times in 7/2004 and in 6/2007 to control inflation pressure also; (iii) maintaining the growth of money supply at over 30 % from 2003Q3 and over 40 % in the last three quarter in 2007. The expansion successfully achieved annual average output growth at 7.5 % but was undeniably one of the main causes of high inflation pressure at the end of 2007. In this period, MCI also appropriately declined in 2004Q2–2004Q3 and 2005Q4–2006Q1 to reflect tight monetary policy to curb inflation rising in the end of 2004. In 2008Q1–008Q4, MCI declined drastically when monetary policy turned to be tightened to control inflation. At the beginning of 2008, monetary policy was remarkably tightened by: (i) repeatedly adjusting policy rates upward. Rediscount rate and base rate were raised three times to 13 and 14 % respectively while refinancing rate was raised fourth times to 14 %; (ii) increasing reserve requirement by 1 % and applying this compulsory deposit for all terms (Decision 187/2008/QD-NHNN)[12]; (iii) issuing compulsory T-Bills worth 20,300 billion Vietnam dong to commercial banks (Decision 346/QD-NHNN); (iv) controlling total banking outstanding loans, discount financial instruments used to invest in the security market accounting for 20 % chartered capital of credit institutions, raising risk ratio for mortgage loans and security investment loans to 250 % (from 150 % previously) (Decision 03/2008/QD-NHNN). In 2009Q1–2009Q4, due to dramatic monetary expansion against the impact of global crisis with five times reduce all policy rates from 10/2008, MCI reversed to rise in the early of 2009. Monetary policy was remarkably expanded in

[12] Previously reserve requirements were only applied to current deposits and under 24 month deposits.

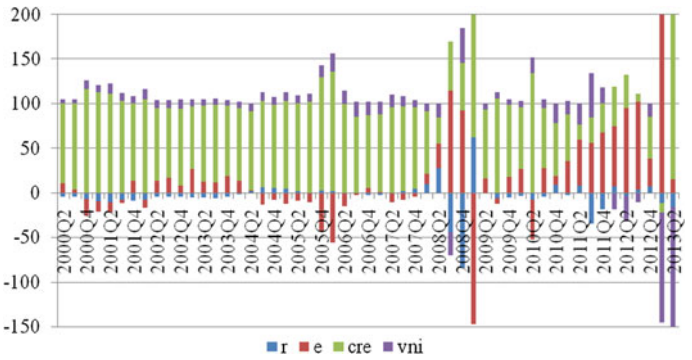


Fig. 8 Contribution of composite variables in MCI evolution. *Source* Authors' estimation

2009 through credit channel with two stimulus packages of the Government. This timely support made the economy successfully overcome recession with economic growth at 5.32 % and inflation at 6.5 %. In 2010Q1–2013Q2, after expansionary period to avoid recession, monetary policy was tightened again to control inflation pressure occurring at the end of 2009, which led to downward movement of MCI for the whole period. The tight monetary policy successfully controlled inflation in 2012 at 8 % and in the first half of 2013 at 2.4 % but traded off with low output growth. In the period of curbing inflation, MCI also reversed to rise in 2010Q3–2011Q1 and 2012Q4, reflecting monetary expansion to pursue output growth target in the second half of 2010 and the end of 2012, also indicating SBV still gives considerable weight for output growth. Figure 8 indicates that the evolution of MCI is mainly determined by movement of credit and exchange rate which are also key transmission channels of Vietnam founded in various empirical studied such as Guamata et al. [15] and Le [20]. Meanwhile, interest rate and stock price generally can explain little of MCI evolution, showing the limited transmission of these two channels also founded in studies of Le [21], Pham [29]. Besides, Fig. 8 also implies that all variables do not have the same fluctuations with each other and with MCI at all time. This means the variables have different responses to the dynamics of monetary policy. Thus, the parameters having the greatest response to monetary policy operations can determine the direction of MCI fluctuations as well as the economy response.

The analysis of innovation of MCI with monetary policy performance in studied period indicates that calculated MCI has very close relation with both monetary instruments (i.e. policy rates) and intermediate target variable (money supply) and policy goals (output growth and inflation). However, the calculated MCI also has some limitations. First, MCI does not have consistent lag response with changes of monetary instruments since composite variables of MCI have different lag response. For example, interest rate quickly reacts to policy rates changes (usually, in the same quarter); similarly, exchange rate also immediately responds to published interbank exchange rate adjustments; meanwhile, credit and stock price have slow reactions (usually from 1 to 2 quarter). Therefore, the variable having prominent reaction in

a specific period of time determines lag response of MCI. Second, in some period, SBV is forced to use its instruments to adjust official market variables under market pressure but has no intention to change the direction of monetary policy. However, changes of official market variable (e.g. exchange rate, interest rate) cause MCI to change reflecting change of monetary policy stance. For instant, due to significant increase of published interbank exchange rate in other to allow exchange rate quoted at commercial bank to follow market pressure in 2/2011,[13] MCI suddenly increases, reflecting monetary policy turn to expand though SBV is trying to control high inflation. These limitations of computed MCI can be explained by three reasons. The first one is that we calculate MCI based on weighted-sum approach with static weight but dynamic weight. The second limitation is intervention of SBV is to follow the market but not orient the market. For example, only when the deviation between exchange rate quoted at commercial banks and that used in the free market is considerably high, does SBV decide to raises published interbank exchange rate to cancel the gap. The third one comes from the nature of MCI while the weights included in the equation encompass the effects of other factors to aggregate demand rather than monetary policy operating only.

8 Conclusion and Recommendations

Although MCI constructed in this study are not appropriate with monetary management at all time, this indicator still have two essential characteristics of a supporting index for short term monetary policy management, including quick responses to monetary policy changes and close relation with policy goal. For that reason, SBV should consider to employ this indicator in short-run management of monetary policy.

References

1. Abdul Majid, M.Z.: Measuring monetary conditions in a small open economy: the case of Malaysia. J. Financ. Econ. Policy **4**(3), 218–231 (2012)
2. Ball, L.: Policy Rules for Open Economy. National Bureau of Economic Research. NBER Working Paper 6760 (1998)
3. Bank of Canada: Monetary Conditions. Bank of Canada Review Autumn, pp. 18–20 (1994)
4. Batini, N., Turnbull, K.: A dynamic monetary condition index for the UK. J. Policy Model. **24**, 257–281 (2002)
5. Bernanke, B.S., Gertler, M.: Inside the black box: the credit channel of monetary policy transmission. J. Econ. Perspect. **9**(4), 27–48 (1995)
6. Bhattacharya, R.: Inflation Dynamics and Monetary Policy Transmission in Vietnam and Emerging Asia. IMF working papers WP/13/155 (2013)
7. Chote, R: IMF urges rate cut by Germany. Financ. Times **1**, p. 1 (1996)

[13] Exchange rate quoted at commercial bank is determined by published interbank exchange rate from SBV and fluctuation band.

8. Davies, G., Simpson, J.: Summary. Int. Econ. Anal. (Goldman Sachs) **2**, 3–18 (1996)
9. Duguay, P.: Empirical evidence on the strength of monetary transmission mechanism in Canada: an aggregate approach. J. Monet. Econ. **33**(1), 39–61 (1994)
10. Eika, K., Ericsson, N., Nymoen, R.: Hazards in implementing a monetary conditions index. Oxf. Bull. Econ. Stat. **58**(4), 765–790 (1996)
11. Ericsson, N.R., Jansen, E.S., Kerbeshian, N.A., Nymoen, R.: Understanding a Monetary Conditions Index. Mimeo, Federal Reserve Board (1997)
12. Freedman, C.: The role of monetary conditions and the monetary condition index in the conduct of policy. Bank Canada Rev. (1995)
13. Gauthier, C., Graham, C., Liu, Y.: Financial Conditions Indexes for Canada. Bank of Canada Working Paper 22 (2004)
14. Goodhart, C., Hofmann, B.: Asset prices, financial conditions, and the transmission of monetary policy. Paper prepared for the conference on Asset Prices, Exchange Rates, and Monetary Policy, Stanford University (2001)
15. Guamata, N., Nir, K. and Eliphas, N.: A financial condition index for South Africa. IMF Working Paper WP/12/196 (2012)
16. Hansson, B., Lindberg, H.: Monetary conditions index—a monetary policy indicator. Q. Rev. **3**, 12–17 (1994). (Sveriges Riksbank—Swedish Central Bank)
17. Hoang, K.T.: Estimating the response of real output to monetary policy instruments shocks in Vietnam. University of East Anglia, Norwich Economic Papers (2009)
18. IMF: World Economic Outlook. International Monetary Fund, Washington (1996)
19. Khan, S., Qayyum, A.: Measures of Monetary Policy Stance: The Case of Pakistan. PIDE Working Papers 39 (2007)
20. Le, A.T.: Beyond inflation targeting: assessing the impacts and policy alternatives. In: Epstein, G.A., Yeldan, A.E. (eds.) Monetary Policy in Vietnam: Alternatives to Inflation Targeting, pp. 299–314. Edward Elgar Publishing Ltd, Cheltenham (2010)
21. Le, V.H.: VAR analysis of the monetary transmission mechanism in Vietnam. Appl. Econom. Int. Dev. **9**(1), 165–179 (2009)
22. Margarita, D., Maria, S.G.: Financial condition index for Asean economies. ADB economics Working paper series 333 (2013)
23. Mayes, D.G., Viren, M.: Financial Conditions Indexes. Economia Internazionale/International Economics. Camera di Commercio di Genova **55**(4), 521–550 (2002)
24. Modigliani, F.: Monetary policy and consumption: linkages via interest rate and wealth effects in the FMP model. In: Consumer Spending Money and Monetary Policy: The Linkages. Conference Series, pp. 9–84. Federal Reserve Bank, Boston (1971)
25. Nguyen, P. L.: Monetary transmission mechanism under quantitative analysis. Bank. Rev. **18**, 19–27 (2010)
26. Osorio, C., Pongsaparn, R., Unsal, D.F.: A quantitative assesment of financial conditions in Asia. IMF Working paper WP/11/170 (2011)
27. Peng, W., Leung, F.: A monetary conditions index for mainland China. Hong Kong Monet. Auth. Q. Bull. **1**, 5–14 (2005)
28. Pesaran, H., Shin, Y.: Generalized impulse response analysis in linear multivariate models. Econ. Lett. **58**, 17–29 (1998)
29. Pham, M.: A Structural Vector Autoregression Model of Monetary Policy in Vietnam. Available at SSRN http://ssrn.com/abstract=2272604 or http://dx.doi.org/10.2139/ssrn.2272604 (2013)
30. Sims, C.: Macroeconomics and reality. Econometrica **48**(1), 1–48 (1980)
31. Suttle, P.: Monetary Conditions Worldwide. World Financial Markets (JP Morgan), vol. 26 (1996)
32. Swiston, A.: A U.S. Financial conditions index: putting credit where credit is due. IMF Working Papers, No. WP/08/16 (2008)
33. Ta, Q.K.: Discussion on developing monetary market in Vietnam. Bank. Rev. **7**, 5–7 (2004)

Analysis of International Tourism Demand for Cambodia

Chantha Hor and Nalitra Thaiprasert

Abstract This study uses fixed–effect and random–effect models to analyze six sets of panel data over the period of 17 years (1996–2012) in order to investigate determinant factors that could affect international tourism demand for Cambodia. We find that the GDP per capita of the origin countries in the previous year has a significant and positive effect on international tourist arrivals to Cambodia, except for tourist arrivals from ASEAN and Europe. Higher relative prices in Cambodia have a significant and negative effect on tourist arrivals to Cambodia, except for tourist arrivals from Oceania and Europe. Appreciated Cambodian riel has a significant and negative effect on tourist arrivals from Oceania and North America, but has a significant and positive effect on tourist arrivals from Europe. Transportation cost has a significant and positive effect on tourist arrivals from Asia and Europe. This may suggest that tourists from these two regions are less sensitive to the transportation cost compared with other factors that drive them to visit Cambodia. The dummy variables which represent the significant events all yield the expected signs.

1 Introduction

Tourism is the world's expeditious growth industry which can contribute significantly to economic growth in both developed and developing countries. The industry directly employs more than 98 million people around the world, or around 3 % of all jobs. The job expansion in travel and tourism is forecasted to average about 1.9 % per year over the next decade, compared with the 1.2 % annual growth rate forecasted for the total number of jobs in the global economy [26]. Moreover, tourism is the world's largest export earner, generating around US$ 1.3 trillion and representing 6 % of the world's exports in 2013. International tourist arrivals climbed up 5 % to

C. Hor (✉) · N. Thaiprasert
Faculty of Economics, Chiang Mai University,
Chiang Mai 50200, Thailand
e-mail: chatha.hor@gmail.com

N. Thaiprasert
e-mail: nalitra@gmail.com

V.-N. Huynh et al. (eds.), *Econometrics of Risk*, Studies in Computational Intelligence 583, DOI 10.1007/978-3-319-13449-9_28

US$ 1.087 billion from US$ 1.035 billion in 2012. It is forecasted to grow up to US$ 1.8 billion in 2030 [27].

Tourism in Cambodia started to flourish in the 1960s, but it was seriously damaged in the 1970s and 1980s by the civil conflict and genocidal policies of the Pol Pot era, which destroyed all the related tourism systems in the country. The industry could turn around only after Cambodia had gained some peace through the 1991 Paris Peace Agreement [6]. Since then Cambodia's tourism industry has played an important role as an engine of its economic growth. It is the second largest profit-earning industry for the Cambodian national account after the garment industry, accounting for 12 % share of the whole economy in 2013 [31]. Tourist arrivals have increased dramatically to an average of 20 % annually during the period from 1993 to 2013. In 2004, international tourists contributed around 50.5 % to Cambodia's Gross Domestic Product (GDP) and around 24.4 % in 2012. The revenue from international tourism has increased from $2.21 billion in 2012 to $2.55 billion in 2013 [14]. Additionally, according to the World Travel and Tourism Council report in 2012, Cambodia's tourism industry has created around 1.45 million direct jobs in 2011, and the number is estimated to rise to 1.5 million in 2012 and 1.95 million in 2022 [31].

Cambodia has realized that the tourism industry is a key economic driver for helping Cambodia to achieve the United Nations Millennium Development Goals (MDG) and Cambodia's national development plan. The government has put much effort into building the available tourism system for attracting international tourists by developing the infrastructure, such as roads, bridges, airports, river and sea harbors, and power and water supply, and has undertaken the developing of more and more innovative tourist places for increasing the lengths of stay of international tourists. National tourism strategic plans and policies have been released to stimulate economic growth through tourism. Furthermore, the government has authorized the right to the ministry of tourism to cooperate with private sectors, NGOs, and international development partners in pushing the plans and policies into effect.

As the role of tourism industry in Cambodia has grown larger, however, there is still little attention paid to investigate factors affecting international tourists' decision to visit Cambodia. Since there has never been a research done using quantitative analysis for Cambodia's tourism industry, this paper is the first study to perform an econometric analysis on international tourism demand for Cambodia's tourism in order to understand factors influencing international tourists' decision-making in coming to Cambodia. Results from this study will help to develop the tourism industry in Cambodia.

2 Literature Review

Many analysis tools have been used in analyzing the tourism industry around the world, namely the classical multivariate regression, advanced modern econometric approaches (such as VAR model), vector autoregressive model, ARMAX model, system-of equation approach, autoregressive distributed lag model, co-integration

test, error-correction model, generalized method of moment (GMM) model, novel hybrid system, simple time-series models (such as naïve, simple autoregressive, smoothing exponential, and trend curve analysis), and advanced time-series models (such as seasonal ARIMA and conditional volatility models).

There are plenty of international tourism demand studies which use time series models in their analysis. For example, [18] use the ARMAX model to investigate the dynamic relationship between tourism demand and real income of Japan for New Zealand and Taiwan. The outcomes indicate that international travel is positively correlated to income of the origin country. Ouerfelli [22] uses co-integration analysis and error correction model to investigate European tourism demand in Tunisia and finds a large elasticity magnitude, which may be a reflection of tourism as a luxury goods bought by European countries, and the supply factor is a significant effect on the tourists' decision-making in visiting Tunisia. Sr and Croes [25] employ time series data using linear and double log-linear models to study the demand of the U.S. tourists to Aruba. The study reveals that the effects of income dominate those of prices and exchange rates. In general, the U.S. tourists appeared to be highly sensitive to the income variables and inelastic with respect to price. The exchange rate variable is not significant. Habibi et al. [9] use co-integration to find the UK and the U.S. tourism demand for Malaysia and find that the long-run equilibrium does exist among the variables selected (income in origin countries, tourism prices in Malaysia, and transportation cost between country of origin and destination country), and the tourists seem to be strongly sensitive to tourism prices. In addition to the studies mentioned above, there are plenty of international tourism demand studies using time series models, such as [1, 2, 5, 16, 17, 21, 28, 29].

There are also a handful of studies which use panel data in their econometric analysis. Serra et al. [24] use a dynamic panel model to estimate the international tourist overnight stays in Portugal from the six main countries (The UK, Germany, The Netherlands, Ireland, France, and Spain). Various controlled variables (income per capita, harmonized household consumption, unemployment rate, and final household consumption) are utilized in this study. Results from the study show that some tourist places in Portugal have high elasticity to the income per capita. [33] use panel three-stage least squares (3SLS) to investigate leading indicators influencing Australian domestic tourism demand by using -three dependent variables (numbers of nights stayed by holiday-makers, business travelers, and visitors who visit friends and relatives), and using the consumer sentiment index, household debt, and working hours of consumers as independent variables in the study. They find that the consumer sentiment index has significant impacts on visitors who visit friends and relatives, but not on holiday and business tourists. Also, household debt is increased because of domestic travels. Garin-Munoz and Montero-Martin [8] use panel data from 14 countries during the period of 1991–2003 and a dynamic model to study tourism in the Balearic Islands. Numerous explanatory variables are used to explain the number of tourist arrivals. They find that previous tourism consumption has a significant effect on consumers' willingness to visit the destination country. Moreover, the results suggest that demand is heavily dependent on the development of economic activity in the origin countries and on the cost of living in the destination country.

The authors suggest that tourism advertisement and high-quality services should be included in tourism policy. Ibrahim [12] uses panel data to investigate international tourism demand for Egypt and finds that all the explanatory variables except population are significant. Real gross domestic per capita, real exchange rate, and cost of living in Egypt are significant and inelastic, and tourism in Egypt is sensitive to relative prices. Apart from the studies mentioned above, there are many other studies that use panel data, such as [7, 15, 19, 20, 23].

3 Methodology

The panel data model is used in this study due to its two main advantages. Firstly, the use of annual data avoids the seasonality problem, which is dominant in this sector. Secondly, the utilization of panel data set involves relatively large numbers of observations and consequent increase in degrees of freedom, which reduces collinearity and improves the efficiency of the estimates [11].

The two main models, fixed effect and random effect, are originated from the panel data sets of 26 countries of origin during the period of 1996–2012 to explore the determinant factors that influence international tourism demand to Cambodia.

The estimation of international tourism demand to Cambodia from 26 countries of origin is presented in the following formula:

$$Q_{i,t} = f(GDPPC_{i,t-1}, RP_{i,t}, ER_{i,t}, TC_{i,t}, D_1, \ldots, D_{11}) \tag{1}$$

Or,

$$\begin{aligned} \ln Q_{i,t} = {} & \alpha_0 + \beta_1 \ln GDPPC_{i,t-1} + \beta_2 \ln RP_{i,t} + \beta_3 \ln ER_{i,t} + \beta_4 \ln TC_{i,t} \\ & + \theta_1 D_1 + \ldots + \theta_{11} D_{11} + \varepsilon_{i,t} \end{aligned} \tag{2}$$

where

$\ln Q_{i,t}$ = logarithm of number of tourist arrivals to Cambodia from country of origin i during year t, where t is the period of 1996–2012.

$\ln GDPPC_{i,t-1}$ = logarithm of GDP per capita of the origin country i at time $t-1$ in constant term.

$\ln PR_{i,t}$ = logarithm of the relative price level, using CPI in Cambodia over CPI in the origin country i at time t.

$\ln ER_{i,t}$ = logarithm of the exchange rate, using origin country's currency per USD i over Cambodian riel per USD at time t.

$\ln TC_{i,t}$ = logarithms of total cost of trip from country of origin i to Cambodia, which is measured by multiplying the distance between the origin country to Cambodia by the average annual price of crude oil per barrel in USD.

α_0 = the common value in the constant.

β = the parameter of independent variables (GDPPC, RP, ER, and TC).

θ = the parameter of dummy variables.

D_1 = the global financial crisis during 2008–2009, for which $D_1 = 1$ in the period of 2008–2009, and $D_1 = 0$ = otherwise.

D_2 = the financial crisis in Asia in the period of 1998–1999, for which $D_2 = 1$ during 1998–1999, and $D_2 = 0$ = otherwise.

D_3 = the September 11 attack in the U.S. during 2001–2002, for which $D_3 = 1$ in the period of 2001–2002, and $D_3 = 0$ = otherwise.

D_4 = the Thai military coup in 2006, for which $D_4 = 1$ in 2006, and $D_4 = 0$ = otherwise.

D_5 = the Cambodia-Thai border dispute during 2008–2011, for which $D_5 =$ 1during 2008–2011, and $D_5 = 0$ = otherwise.

D_6 = the SARS epidemic in East Asia in 2003, for which $D_6 =$ 1in 2003, and $D_6 = 0$ = otherwise.

D_7 = the political instability and political deadlock Cambodia in 1997–1998 and 2003, for which $D_7 = 1$ in 1997–1998 and 2003, and $D_7 = 0$ = otherwise.

D_8 = the tsunami in Japan during 2011–2012, for which $D_8 = 1$ in the period of 2011–2012, and $D_8 = 0$ = otherwise.

D_9 = the tsunami in Southeast Asia during 2004–2005, for which $D_9 = 1$ in the period of 2004–2005, and $D_9 = 0$ = otherwise.

D_{10} = the single visa entry scheme to enter five ASEAN countries (Cambodia, Laos, Myanmar, Thailand, and Vietnam), started in 2012, for which $D_{10} = 1$ in 2012, and $D_{10} = 0$ = otherwise.

D_{11} = the visa exemption among ASEAN countries in the period of 2006–2012, for which $D_{11} = 1$ during 2006–2012, and $D_{11} = 0$ = otherwise.

In exploring the determinants of the demand for tourism, we allow for the existence of individual effects which are potentially correlated with explanatory variables, such that

$$\varepsilon_{i,t} = \lambda_i + \gamma_{it} \tag{3}$$

Here, λ_i is unobserved country-specific effect that varies across countries but is invariant within a country over time, and γ is a white noise error term. To solve with the unobservable individual effect in a panel data model, using a within-panel estimator, fixed–effect or random–effect technique, omitting the individual effect is a standard estimation method. The fixed effect model assumes that each country has an individual unobserved country-specific effect and estimates the constant term (unobserved country-specific effect) for each country, while the random effect model estimates only one constant term by assuming that the unobserved country-specific effect follows a normal distribution [3, 4, 13, 30, 32]. Either the fixed–effect or the random–effect model is chosen through the Hausman test [10] for its results.

Table 1 List of regions used in panel data

Regions	Countries
ASEAN(8)	Brunei, Indonesia, Loa PDR, Malaysia, The Philippines, Singapore, Thailand, and Vietnam
East Asia (4)	China, Japan, South Korea, and Taiwan
South Asia (1)	India
Oceania (2)	Australia and New Zealand
Europe (9)	Belgium, Denmark, France, Germany, Italy, The Netherlands, Norway, The United Kingdom, and Switzerland
North America (2)	Canada and The U.S

4 Data

The dependent variable for this study is the number of tourist arrivals from 26 countries (origin countries), covering 6 main regions (see Table 1 for members in each region), over the period of 17 years (1996–2012). The number of tourist arrivals is used as a proxy for the international tourism demand. The independent variables are divided into two types, economic and non-economic variables. As for the economic variables, there are Gross Domestic Product per capita in constant term at time $t-1$ (GDPPC); relative price (RP) which is a ratio of consumer price index (CPI) of the destination country (Cambodia) over the CPI of each origin country; exchange rate (ER) which is the ratio of each origin country's currency per USD over Cambodia's riel per USD; and transportation cost (TC) which is measured by multiplying the distance from each origin country to Cambodia by the average annual price of a barrel of oil. The data for the economic variables are obtained from the CEIC database installed at the faculty of Economics, Chiang Mai University, the U.S. Energy Information Administration, and http://www.distancefromto.net. For the non-economic factors, dummy variables are used with 1 representing periods when each event occurs and (still a strong impact is maintained). The periods used in the dummy variables are from the authors' observation of each event.

5 Results and Discussion

Results either from the fixed–effect or the random–effect model, after being chosen according to the Hausman test, are reported in Table 2. The random–effect model is more appropriate for the 26-country group, ASEAN, Europe, and North America, while the fixed–effect model is more appropriate for Asia and Oceania.

Results from the analysis show that GDP per capita of the origin countries in the previous year has a significant and positive effect on international tourist arrivals to Cambodia, except for tourist arrivals from ASEAN and Europe. Higher relative prices in Cambodia have a significant and negative effect on international

Table 2 Regression Analysis Results, Using Number of International Tourist Arrivals as Dependent Variable

Variables	Total 26 countries	ASEAN	Asia (EA+SA)	Oceania	Europe	North America
	Random	Random	Fixed	Fixed	Random	Random
Constant	4.200*	9.949*	−13.203*	−21.65*	−2.791	−80.29***
	(0.026)	(0.026)	(0.012)	(0.013)	(0.661)	(0.000)
$\ln GDPPC_{(t-1)}$	0.502***	0.388	1.362***	1.707*	0.336	7.849***
	(0.001)	(0.378)	(0.002)	(0.086)	(0.583)	(0.000)
LnRP	−0.575***	−0.588**	−2.772***	−0.074	−0.450	−5.286***
	(0.000)	(0.041)	(0.004)	(0.914)	(0.430)	(0.000)
LnER	0.020	0.177	−0.126	−1.543***	0.041**	−0.695***
	(0.505)	(0.329)	(0.823)	(0.004)	(0.030)	(0.006)
LnTC	−0.035	−0.416	0.797**	0.082	0.623***	0.049
	(0.824)	(0.187)	(0.015)	(0.680)	(0.000)	(0.785)
D1	−0.088	−0.157	−0.063	0.173*	−0.051	−0.468**
	(0.384)	(0.555)	(0.803)	(0.092)	(0.593)	(0.000)
D2	−0.129	−0.452**	0.291	−0.187	0.248**	0.314
	(0.164)	(0.045)	(0.297)	(0.296)	(0.049)	(0.104)
D3	0.096	−0.123	0.109	0.239**	0.213**	−0.264***
	(0.235)	(0.565)	(0.592)	(0.025)	(0.014)	(0.007)
D4	−0.194*	−0.194	−0.239	0.016	−0.073	−0.263**
	(0.099)	(0.529)	(0.372)	(0.864)	(0.452)	(0.032)
D5	0.211*	0.287	0.378	−0.003	0.219	1.388***
	(0.076)	(0.355)	(0.231)	(0.982)	(0.153)	(0.000)
D6	−0.244***	−0.215	−0.103	−0.244***	−0.203***	−0.195**
	(0.005)	(0.325)	(0.616)	(0.007)	(0.008)	(0.030)
D7	−0.244***	−0.215	−0.103	−0.244***	−0.203***	−0.195**
	(0.005)	(0.325)	(0.616)	(0.007)	(0.008)	(0.030)
D8	1.473***	1.943***	1.178***	0.884***	1.290***	0.987***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
D9	1.337***	1.215***	0.933***	0.584***	1.186***	0.443***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
D10	0.387**	0.648	0.463**	0.1032	0.292	1.463***
	(0.020)	(0.138)	(0.012)	(0.567)	(0.118)	(0.000)
D11	0.668***	0.801*				
	(0.000)	(0.063)				
R-square	0.8106	0.7583	0.8190	0.9940	0.9398	0.9828
# of countries	26	8	5	2	9	2
# of obs	442	136	88	34	153	34

Note The numbers in parentheses are the p-values.
"*", "**" and "***" denote the statistical significance levels at 10 %, 5 %, and 1 %, respectively

tourist arrivals to Cambodia, except for tourist arrivals from Oceania and Europe. Appreciated Cambodian riel has a significant and negative effect on tourist arrivals from Oceania and North America, but has a significant and positive effect on tourist arrivals from Europe. The transportation cost has a significant and positive effect on tourist arrivals from Asia and Europe. The distance between Asia and Cambodia is logically not an obstacle for discouraging tourists from travelling to Cambodia, while tourists from Europe might take a trip to Cambodia via neighboring countries where it is convenient and easy to get into Cambodia.

The global financial crisis in the period of 2008-2009 (D1) has a significant positive and negative effect on tourist arrivals from Oceania and North America, respectively. The result is likely because of the hardship caused by the crisis that had more effect on tourist arrivals from North America, which discourages them from visiting Cambodia, while the event might not have impacted tourist arrivals from Oceania. The financial crisis in Asia in the period of 1998–1999 (D2) has a significant and negative effect on tourist arrivals from ASEAN, but has a significant and positive effect on tourist arrivals from Europe. This suggests that people from ASEAN were hard hit by the Asian financial crisis and were discouraged from visiting Cambodia, while the crisis may not have impacted people in Europe much as the results of the appreciated Cambodian riel also suggest that Europeans may be less sensitive to these two factors when compared with their motive to visit Cambodia. The September 11 attack in the U.S. during 2001–2002 (D3) has a significant and negative effect on tourist arrivals from North America, but has a significant and positive effect on tourist arrivals from Oceania and Europe. This result supports the fact that the American sentiment was low right after the September 11 attack and that the event probably discouraged them from traveling abroad.

The Thai military coup in 2006 (D4) has a significant and negative effect on tourist arrivals from the 26-country group and North America. This may suggest that international tourists, especially North American tourists, may plan to visit several ASEAN countries in one trip. Thus, turmoil in Cambodia's neighboring countries could also affect the tourism industry in Cambodia. The SARS epidemic in Asia in 2003 (D6) has a significant and negative effect on tourist arrivals from the 26-country group, Oceania, Europe, and North America. This result may suggest that SARS epidemic, though not so severe in Cambodia, could send out a very strong negative image of the situation that could discourage tourists from visiting Asia altogether. The political instability and political deadlock in Cambodia in the period of 1997-1998 and 2003 (D7) has a significant and negative effect on tourist arrivals from the 26-country group, Oceania, Europe, and North America. This suggests that internal conflicts in Cambodia could be very harmful to the tourism industry as they discourage tourists, especially from long-distance countries, from visiting Cambodia.

The tsunami in Japan in the period of 2011–2012 (D8) and the tsunami in Southeast Asia in the period of 2004–2005 (D9) have a significant and positive effect on international tourist arrivals to Cambodia from every region. This suggests that although the events are horrific and dismal to the affected countries, they benefit

Cambodias tourism industry as tourists may change their plan from traveling to the impacted countries to traveling to Cambodia instead.

The single visa entry scheme to enter five ASEAN countries (Cambodia, Laos, Myanmar, Thailand, and Vietnam), which was started in 2012 (D10), has a significant and positive effect on tourist arrivals from the 26-country group, Asia, and North America. This suggests that international tourists, especially from Asia and North America, may enjoy this benefit of the single visa as they may plan to visit several countries in ASEAN at the same time. The visa exemption among the ASEAN countries in the period of 2006–2012 (D11) has a significant and positive effect on tourist arrivals from ASEAN. This policy, no doubt, has created more tourist movements among the ASEAN countries.

6 Conclusions

This study explores determinant factors that influence international tourism demand in Cambodia over the period of 17 years (1996–2012) by employing panel data and analyzing with fixed–effect and random–effect models. We find that GDP per capita of the origin countries in the previous year has a significant and positive effect on international tourist arrivals to Cambodia, except for tourist arrivals from ASEAN and Europe. Higher relative prices in Cambodia have a significant and negative effect on tourist arrivals to Cambodia, except for tourist arrivals from Oceania and Europe. Appreciated Cambodian riel has a significant and negative effect on tourist arrivals from Oceania and North America, but has a significant and positive effect on tourist arrivals from Europe. Transportation cost has a significant and positive effect on tourist arrivals from Asia and Europe. This may suggest that tourists from these two regions are less sensitive to the transportation cost compared with the other factors that drive them to visit Cambodia. The global financial crisis in the period of 2008-2009 (D1) has a significant positive and negative effect on tourist arrivals from Oceania and North America, respectively. The financial crisis in Asia in the period of 1998–1999 (D2) has a significant and negative effect on tourist arrivals from ASEAN, but has a significant and positive effect on tourist arrivals from Europe. The September 11 attack in the U.S. during 2001–2002 (D3) has a significant and negative effect on tourist arrivals from North America, but has a significant and positive effect on tourist arrivals from Oceania and Europe. The Thai military coup in 2006 (D4) has a significant and negative effect on tourist arrivals from the 26-country group and North America, while the Cambodia-Thai border dispute in the period of 2008-2011 (D5) has a positive impact on the 26-country group and North America. The Cambodia-Thai border dispute in the period of 2008-2011 (D5) has a positive impact on the 26-country group and North America. The result may suggest that the dispute had become an unexpected publicity for the Cambodian tourism industry, which might have encouraged tourists to visit Cambodia. The SARS epidemic in Asia in 2003 (D6) has a significant and negative effect on tourist arrivals from the 26-country group, Oceania, Europe, and North America. The political instability and political

deadlock in Cambodia in 1998 and 2003 (D7) has a significant and negative effect on tourist arrivals from the 26-country group, Oceania, Europe, and North America. The tsunami in Japan during 2011–2012 (D8) and the tsunami in Southeast Asia during 2004–2005 (D9) have a significant and positive effect on international tourist arrivals to Cambodia from every region. The single visa entry scheme to enter five ASEAN countries (Cambodia, Laos, Myanmar, Thailand, and Vietnam), which was started in 2012 (D10), has a significant and positive effect on tourist arrivals from the 26-country group, Asia, and North America. In addition, the visa exemption among the ASEAN countries in the period of 2006–2012 (D11) has a significant and positive effect on tourist arrivals from ASEAN.

Results from the study suggest that the government of Cambodia should carefully monitor relative price changes in Cambodia. The government should also try to prevent any internal conflicts in the country since such conflicts could discourage tourists from visiting Cambodia.

Acknowledgments We would like to thank Associate Professor Dr. Komsan Suriya from the Faculty of Economics, Chiang Mai University for his valuable comments on the methodology of this study.

References

1. Akis, S.: A compact econometric model of tourism demand for Turkey. Tour. Manag. **19**(1), 99–102 (1998)
2. Algre, J., Pou, L.: The length of stay in the demand for tourism. Tour. Demand **27**, 1343–1355 (2006)
3. Baltagi, B., Bresson, G., Pirotte, A.: Fixed effects, random effects or Hausman-Taylor? Econ. Lett. **79**, 361–369 (2003)
4. Baltagi, Bresson, Priotte.: Panel data and unobservable indi- vidual effects. In: Hausman, J.A., Taylar, W.E. (eds.) Econometrica, **49**, 1377–1398 (1987)
5. Chan, F., Lim, C., McAleer, M.: Modelling multivariate international tourism demand volatility. Tour. Manage. **26**, 459–471 (2005)
6. Chheang, V.: Hun Sens Talks and Cambodias Tourism Development: the discourse of power. Ritsumeikan J. Asia Pac. Stud. **25**, 85–105 (2009)
7. Garin-Munioz, T., Amaral, T.: An econometric model for international tourism flows to Spain. Appl. Econ. **7**, 525–529 (2000)
8. Garin-Munoz, T., Montero-Maritin, L.: Tourism in the Balearic Islands:a dynamic model for international demand using panel data. Tour. Manag. **28**, 1224–1235 (2007)
9. Habibi, F., Rahim, K., Chin, L. . United Kingdom and United States tourism demand for Malaysia: a co-integration analysis. MPRA Paper No. 13590 (2008). http://mpra.ub.uni-muenchen.de/13590/
10. Hausman, J., W.E, T.: Panel data and unobservable individual. Econometrica **49**, 1377–1398 (1981)
11. Hsiao, C.: Analysis of panel data. Cambridge University Press, New York (2003)
12. Ibrahim, M.: The determinants of international tourism demand for Egypt: panel data evidence. Eur. J. Econ. **30**, 1450–2275 (2011)
13. Judge, G., Griffiths, W., Hill, R., Lukepohl, H., Lee, T.: Theory and Practice of Econometrics. Wiley, USA (1980)

14. Kong, S., Horth, V.: Tourism Statistics Report. Ministry of Tourism, Phnom Penh (2013) (in Cambodia)
15. Kusni, A., Kadir, N., Nayan, S.: International tourism demand in Malaysia by tourists from OECD countries: a panel data econometric analysis. Procedia Econ. Finan. **7**, 28–34 (2013)
16. Lim, C., McAleer, M.: Modelling international travel demand from Singapore to Australia. CIRJE-F- 214 (2003) (Unpublished Discussion Paper)
17. Lim, C., McAleer, M.: A co-integration analysis of annual tourism demand by Malaysis for Australia. Math. Comput. Simul. **59**, 197–205 (2001)
18. Lim, C., McAleer, M., Min, J.: ARMAX modelling of international tourism demand. Math. Comput Simul. **79**, 2879–2888 (2009)
19. Massidda, C., Etzo, I.: The determinants of Italian domestic tourism: a panel data analysis. Tour. Manag. **33**, 603–610 (2012)
20. Naude, W., Saayman, A.: The determinants of tourist arrivals in Africa: a panel data regression analysis. St. Catherine'College, Center for the study of African Economics (2004)
21. Nelson, A.L., Dickey, A.D., Smith, M.J.: Estimating time series and cross section tourism demand models: Mainland United States to Hawaii data. Tour. Manag. **32**, 28–38 (2011)
22. Ouerfelli, C.: Co-integration analysis of quarterly European tourism demand in Tunisia. Tour. Manag. **29**, 127–137 (2008)
23. Proenca, S., Elias, S.: Demand for tourism in Portugal: a panel data approach. Discussion Paper No. 29 (2005). www4.fe.uc.pt/ceue
24. Serra, J., Correia, A., Rodrigues, P.: A comparative analysis of tourism destination demand in Portugal. J. Destin. Mark. Mang. **2**, 21–227 (2014)
25. Sr, M., Croes, R.: Evaluation of demand US tourists to Aruba. Ann. Tour. Res. **4**, 946–963 (2000)
26. Turner, R.: Travel and tourism economic impact Cambodia. World Travel and Tourism Council (2013)
27. UNWTO.: UNWTO tourism highlights. World Tourism Organization (2013)
28. Uysal, M., Cromptop, C.: Determinants of demand for international tourist flows to Turkey. Tour. Manag. **11**(3), 288–297 (1984)
29. Var, T., Ico, O., Kozak, M.: Tourism demand in Turkey. Ann. Tour. Res. **25**(1), 236–240 (1998)
30. Wooldridge, J.: Introductory Econometrics. The MIT Press, Cambridge (2002)
31. World Travel and Tourism Council.: Travel and tourism economic impact on Cambodia. The Authority of World Travel and Tourism (2013)
32. Yang, C., Lin, H., Han, C.: Analysis of international tourist arrivals in China: the role of World Heritage Sites. Tour. Manag. **31**, 827–837 (2010)
33. Yap, G., Allen, D.: Investigating other leading indicators influencing Australian domestic tourism demand. Math. Comput. Simulat. pp. 1365–1374 (2011)

Modeling the Impact of Internet Broadband on e-Government Service Using Structural Equation Model

Sumate Pruekruedee, Komsan Suriya and Niwattisaiwong Seksiri

Abstract The aim of this study is to test the hypothesis whether the e-Government service is positively affected by the readiness of physical telecommunications infrastructure especially internet broadband and mobile broadband while negatively impacted by the inefficiency of bureaucratic system. It applies the structural equation model (SEM) with limited cross-sectional data of the Networked Readiness Index (NRI) from 140 countries. The style of the modelling is quite similar to the single-equation regression with the difference at the estimation method using latent variables. The results reveal that the e-Government service is ready to provide to people along with the readiness of the telecommunications infrastructure. Unfortunately, the service is pulled back and slowed down by the inefficiency in the bureaucratic system. It is a must for the government to improve the efficiency and incentives for staffs in related government agencies to make them proactive to catch up with updated situation among the competition in the digital world.

1 Introduction

Internet broadband empowers the government to deliver better services to people. It shortens the queuing for citizens in doing transactions with the government. It lessens the paper works and leads to paperless society. It reduces time to fill the same information in various forms. It speeds up the time to find essential information from government agencies.

S. Pruekruedee
Economics Department, School of Management,
Mae Fah Luang University, Chiang Rai, Thailand
e-mail: Sumate.pru@gmail.com

K. Suriya (✉) · N. Seksiri
Faculty of Economics, Chiang Mai University,
Chiang Mai, Thailand
e-mail: suriyakomsan@gmail.com

N. Seksiri
e-mail: s.niwattisaiwong@alumni.lse.ac.uk

V.-N. Huynh et al. (eds.), *Econometrics of Risk*, Studies in Computational Intelligence 583, DOI 10.1007/978-3-319-13449-9_29

The online government service or e-Government is flourishing in the era of the third generation of mobile phone (3G). Mobile broadband even catalyzes the benefits of internet broadband on the government service with its power to allow e-Government service to access anyone, anywhere and anytime. It is easier for people to pay tax via smart phone compared to traveling to the local revenue department. It also enhances scholars to access to government's data or announcements freely and rapidly on the website. It also encourages people to send complaints to government immediately right after their contacts to the incidents.

However, the benefits of internet broadband especially mobile broadband are unclear in current literatures. First, it may be to early too observe its impact in developing countries where internet broadband penetrates less than 20 % of the citizens. Eventhough the penetration ratio of mobile broadband is much higher and almost covers 90–95 % of population but the speed is limited to 2G rather than 3G. The 3G service in many countries are still emerging and covers less than half of their population. Second, the government in developing countries deliver low quality of e-Government services; In worse cases, some countries provide no e-Government service at all.

The quality of e-Government service then depends on two sides, the readiness of the physical telecommunications infrastructure and the readiness and effectiveness of government agencies to provide the e-Government service. The e-Government faces some uncertainties from both sides. On the readiness of the infrastructure, the service cannot control the stability of the connection provided by telecommunications operators. Drop calls and the weak signals will make the quality of e-Government service low. The narrow coverage of the mobile broadband over geographical areas restricts the e-Government service to reach just some big cities. Moreover, the subscriptions of fixed and mobile broadband limits the number of people that can access to the e-Government service.

On the readiness of the government agencies, the uncertainty is at the inefficiency of the bureaucratic system. The first inefficiency appears at the obsolete information provided online. Government officials are not active to update the information onto the servers. The second one disappoints people when the link in the website is broken and unavailable. The maintenance of the website takes too much time. The third one is at the reliability of the system especially when people deals with financial matters with the government such as paying tax online. These risks discourages people to use e-Government.

This paper will test the hypothesis whether the e-Government service is affected positively by the readiness of the physical telecommunications infrastructure and negatively by the inefficiency of bureaucratic system. The results will suggest the strategies to improve and upgrade the e-Government service by the application of internet broadband especially the mobile broadband.

2 Literature Review

There are hundred of literatures that evaluate the impact of internet broadband on e-Government in many countries. Among them, the work of Ferro et al. [1], Omar [2] and Trkman and Turk [3] may be relevant to the modeling in this study.

Ferro et al. [1] study the impact of internet broadband on e-Government in Italy. They find the positive relationship between the internet broadband and the supply of e-Government services. The crucial factor is the coverage of telecommunications network. Small town with less than 10,000 residents risks to be excluded from the e-Government service. This study emphasizes on the supply side especially at the importance of the telecommunication infrastructure as a foundation to supply e-Government services.

Omar [2] investigate the determinants of the adoption of e-Government service in Jordan. He discovers that the attitude of people toward the usage of e-Government affects the adoption more than the perceived benefits from the usage. This study switches the focus into the demand side. Regardless of the quality of the e-Government services, people in the Arab word rather choose to use the services by emotional and psychological reasons than the intrinsic benefits of the e-Government.

Trkman and Turk [3] include the analysis both on the demand and supply sides. They try to construct the framework to find the interaction between internet broadband and e-Government. They begin the development at the broadband development or e-readiness. Then, they link the internet broadband directly to the usage of e-Government services which produces both the government internal benefits and the benefits from serving people. This study categorizes e-Government services into four groups. They are the demand-side for economic usage, demand-side for social activities, supply-side for economic usage and supply-side for social activities. They give some examples of the e-Government services in each category.

From these literatures, modeling the impact of internet broadband on e-Government service should include both the supply and demand sides. On the supply side, the readiness of telecommunication infrastructure should be emphasized. On the demand side, the readiness of the user should be focused. However, to make this study different from previous literatures and contribute to the academic world, it models the demand side by the readiness of the government which is an important user of e-Government by itself as mentioned by Trkman and Turk [3] that government agencies are the biggest customers of the e-Government service. However, the readiness of government agencies can be viewed both as the users (demand side) and providers (supply side) of the e-Government service.

3 Methodology and Data

This study uses the structural equation model (SEM). Usually, the model is a multilayer analysis. The analysis is popular among social scientist and becomes more popular among economists to find the significant path that carries the effect of a

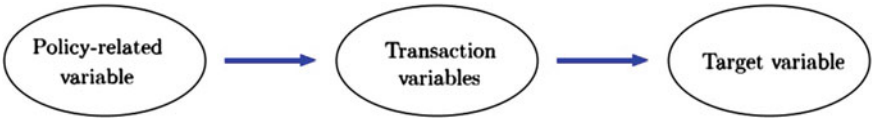


Fig. 1 The multi-layer analysis in the structural equation model (SEM)

policy-related variable to the target variable via some transaction variables in the middle (Fig. 1).

At the first place, this study aims at modeling the impact of mobile broadband on e-Government service in the multi-layer analysis style. However, by the limitation of the data from the networked readiness index (NRI) of 140 countries in 2013 provided by the World Economic Forum (WEF), the analysis needs to shorten the layers into two. The analysis is then compatible to the single equation regression. However, the estimation method is to make via latent variables that represent the readiness of the physical infrastructure, the readiness of government agencies and the quality of e-Government service.

The conceptual framework of the structural equation model in this study is as follows:

In Fig. 2, the quality of e-Government service depends on the physical readiness and government readiness. For the physical readiness, it is divided into the internet broadband and mobile broadband. Additionally, the model presents the effects of the readiness of both sides on government usage of internet too which may affect the e-Government service indirectly.

Fig. 2 Conceptual framework of structural equation model in this study

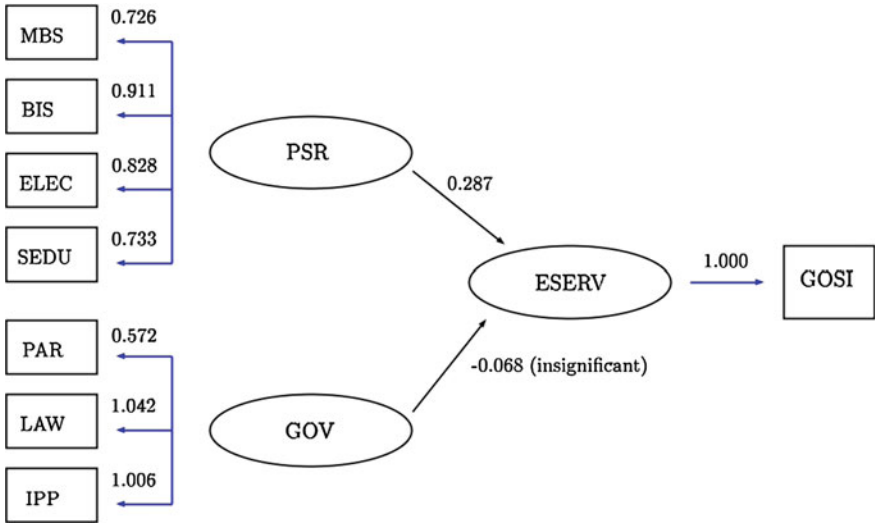


Fig. 3 The result of the impact of internet broadband on e-Government service

4 Results

It is shown by the results in Fig. 3, it is a conventional approach to present SEM result [4, 5]. The readiness of physical telecommunications infrastructures, both the internet broadband (BIS with loading factor of 0.911) and mobile broadband (MBS with loading factor of 0.726) are significant to the quality of e-Government service through the coefficient of 0.287 from the readiness of physical infrastructure (PSR) to the e-Government service (ESERV). The bureaucratic system almost affects the e-Government negatively (GOV with the coefficient of –0.068); however, the relationship is insignificant.

The explanations of all variables in the model are presented below:
The first layer: The readiness of the physical infrastructure and of the government agencies.

PSR: Latent variable of the readiness of physical infrastructure.
MBS: Mobile broadband subscriptions/100 pop.
BIS: Broadband Internet subscriptions/100 pop.
ELEC: Electricity production, kWh/capita.
SEDU: Secondary education gross enrollment rate.

GOV: Latent variable of the readiness of government agencies.
PAR: Effectiveness of law-making bodies.
LAW: Laws relating to ICT.
IPP: Intellectual property protection.

Table 1 Goodness of fit of the model

Chi-square (df)	p-value	SRMR	RMSEA	CFI	TLI
24.323 (16)	0.083	0.036**	0.068*	0.985**	0.974**

Remark * is acceptable fit level
** is good fit level

The second layer: The quality of e-Government service.

ESERV: Latent variable of the e-Government service.
GOSI: Government Online Service Index.

Table 1 provides conventional fit indices both absolute and incremental one. All incremental fit indices: Comparative Fit Index (CFI) and Tucker-Lewis Index (TLI) are at good fit level. Absolute fit indices: Standard Root Mean Square Residual (SRMR) is at good fit level, Root Mean Square Error of Approximation (RMSEA) is at acceptable fit level, while Chi-square and its p-value are not fit at significant level as usual when the number of observation is high [6].

5 Discussions

The availability of telecommunications infrastructure is ready for the government to provide a good quality of e-Government service but the service is not good enough because the bureaucratic system in the government. It is easy to change the infrastructure just by the investment in the hardware but it is hard to change the culture of government officials to be more proactive to provide a good service to people via internet broadband. The investment in human capital is not enough to help this. It crates the market for internet broadband and generates demand for the e-Government service but it cannot create smart persons in smart working environmental in government agencies to catch up with updated information and provide accurate information to people at the speed that allows the individuals and private firms to compete in the digital world.

6 Conclusions

This study uses structural equation model (SEM) to investigate the determinants of the quality of e-Government service with the data from 140 countries provided by the World Economic Forum (WEF). The model is limited by the number of observations so that its architecture is quite compatible to the single-equation regression. The results reveal that e-Government is ready to be served by the readiness of physical telecommunications infrastructures but restricted by the inefficiency of bureaucratic system. An only suggestion to the government is that it must pay a serious attention

to the "mobile integration" and make its e-Government service better for people. Otherwise the individuals and private firms will lose their competitiveness even though the investment in the telecommunications infrastructure goes beyond the sufficient level. The government should find ways to eradicate the inefficiency in the bureaucratic system and raise the incentives for staffs in government agencies to work more productively in order to catch up with the internet broadband and mobile broadband technologies in this digital world.

References

1. Ferro, E., De Leonardis, D., Dadayan L.: Broadband and e-Government Diffusion. Proceedings of Hawaii International Conference of System Sciences [online] http://www.researchgate.net/publication/232623594_Broadband_and_e-Government_Diffusion_(PDF) (2009)
2. Alhujran, O.: Determinants of e-Government services adoption in developing countries: a field survey and a case study. Doctoral Dissertation, School of Information Systems and Technology, Faculty of Informatics, University of Wollongong (2009) http://ieeexplore.ieee.org/Xplore/cookiedetectresponse.jsp
3. Peter, T., Turk, T.: A conceptual model for the development of broadband and e-Government. Gov. Inf. Q. **26**, 416–424 (2009)
4. Dyer, P., Gursoy, D., Sharma, B., Carter, J.: Structural modeling of resident perceptions of tourism and associated development on the Sunshine Coast, Australia. Tour. Manag. **28**, 409–422 (2007)
5. Chen, S., Raab, C.: Predicting resident intentions to support community tourism: toward an integration of two theories. J. Hosp. Mark. Manag. **21**, 270–294 (2012)
6. Kline, R.: Principles and Practice of Structural Equation Modeling, 3rd edn. The Guilford Press (2011)

Assessing Sectoral Risk Through Skew-Error Capital Asset Pricing Model: Empirical Evidence from Thai Stock Market

Nuttanan Wichitaksorn and S.T. Boris Choy

Abstract This paper presents a new approach to analyze sectoral risk premium using skew-error regression for capital asset pricing model (CAPM). We adopt the skew distributions proposed by Wichitaksorn et al. [16] to model the sectoral risk premium for the returns in Thai stock market. Applying these skew distributions to the CAPM allows us to (1) better assess the sectoral risk from the financial returns, which are usually slightly-skewed and heavy-tailed, and (2) efficiently implement the model using Bayesian Markov chain Monte Carlo methods. Results from an empirical study suggest that different sectors posses different risk levels.

1 Introduction

It has been well-known that many financial return data are slightly skewed and leptokurtic [3, 6, 8, 10]. Therefore, assuming a normal distribution to model the data is a misspecification and a biased parameter estimation is resulted. In the field of financial econometrics, the capital asset pricing model (CAPM) is mainly used to quantify the relationship between individual excess return (or risk premium) and market excess return and it has been widely used for assessing risk of financial returns. See, for example, Markowitz [11–13], Sharpe [15] and Lintner [9]. Over the past few decades, CAPM has been extended in various ways. But a major criticism is that, in parametric settings, the individual risk premium is always assumed to be normally distributed.

In this paper, we extend the analysis of CAPM by proposing a new approach to analyze sectoral risk premium using skew-error regression. The skew normal and skew Student-t error distributions that we adopt for statistical inference are presented in Wichitaksorn et al. [16]. The way that these skew distributions are constructed

N. Wichitaksorn (✉)
School of Mathematics and Statistics, University of Canterbury, Christchurch 8140, New Zealand
e-mail: nuttanan.wichitaksorn@canterbury.ac.nz

S.T.B Choy
Discipline of Business Analytics, University of Sydney, NSW 2006, Australia
e-mail: boris.choy@sydney.edu.au

V.-N. Huynh et al. (eds.), *Econometrics of Risk*, Studies in Computational Intelligence 583, DOI 10.1007/978-3-319-13449-9_30

simplifies the Gibbs sampling algorithm for Bayesian statistical inference. In addition, the skew Student-*t* error distribution provide a better assessment of the sectoral risk from the leptokurtic financial returns. In the empirical study of excess returns from 21 sectors in the Stock Exchange of Thailand, we implement the CAPM using three error distributions: normal, skew normal and skew Student-*t*. We show that the CAPM with skew Student-*t* error distribution performs significantly better than that of other two error distributions.

This paper is organized as follows. Section 2 presents the generic CAPM and regression model with skew error distributions. Section 3 provides the details of the MCMC algorithms for skew Student-*t* distribution in CAPM. Section 4 presents the results of an empirical study on Thailand sectoral return data. Finally, conclusions are described in Sect. 5.

2 Model

2.1 Capital Asset Pricing Model

In modern portfolio theory, the CAPM aims to quantify the risk of an individual stock (or portfolio) resulted from investor's utility maximizing behavior [9, 11, 15]. On the other hand, CAPM incorporates the mean and variance of the return of an associated stock to assess its individual risk toward the market risk. Let $R_t^{(i)}$ denote the return of stock or sector i at time t, R_t^f denote the risk-free rate, and R_t^m denote the market return. The CAPM is given by

$$E[R_t^{(i)}] - R_t^f = \beta^{(i)}(E[R_t^m] - R_t^f)$$

where $E[R_t^{(i)}] - R_t^f$ is the risk premium of sector i, $E[R_t^m] - R_t^f$ is the market premium and $\beta^{(i)}$ indicates the relative risk between the risk premium of sector i and market premium. For $\beta^{(i)} = 1$, sector i has the same risk level as that of the market. For $\beta^{(i)} > 1$, it is more risky than the market and investors expect to achieve higher return. On the contrary, it is less risky if $\beta^{(i)} < 1$. Since $\beta^{(i)}$ is the coefficient in the relationship between the sectorial risk premium and market premium, it is also a measure of the sensitivity of the sectorial risk premium to market premium. Hence, $\beta^{(i)}$ can be expressed as

$$\beta^{(i)} = \frac{Cov[R^{(i)}, R^m]}{Var[R^m]}$$

where $Cov[A, B]$ is the covariance of A and B and $Var[A]$ is the variance of A. As a result, the CAPM for sector i can be expressed by the following regression model

$$R_t^{(i)} - R_t^f = \alpha^{(i)} + \beta^{(i)}(R_t^m - R_t^f) + \varepsilon_t^{(i)}, \quad \text{for } t = 1, \ldots, T$$

where $\alpha^{(i)}$ is the additional intercept term and $\varepsilon_t^{(i)}$ is the error term. Other than being simply interpreted as an intercept, $\alpha^{(i)}$ measures the risk-adjusted performance of sector i. It indicates the level of excess return of the sector over the benchmark and is used as a technical indicator in portfolio theory. Therefore, if $\alpha^{(i)} = 0$, there is no risk-adjusted performance and if $\alpha^{(i)} > 0$ ($\alpha^{(i)} < 0$), there is a positive (negative) risk-adjusted performance and a higher (lower) return is expected for sector i. However, the method of least squares fails to provide efficient and unbiased estimators for $\alpha^{(i)}$ if the error distribution is asymmetric; see McDonald et al. [14]. Therefore, this paper adopts a skew error distribution to obtain an efficient and unbiased estimator for $\alpha^{(i)}$.

2.2 CAPM with Skew Error Distributions

Using normal scale mixtures [2, 5], Wichitaksorn et al. [16] propose a class of skew probability distributions whose pdf is given by

$$f_{skd}(\varepsilon|0,1,p) = 2\int_0^\infty \left[pN\left(\varepsilon\left|0,\frac{\lambda}{4(1-p)^2}\right.\right)I(\varepsilon \le 0) + (1-p)N\left(\varepsilon\left|0,\frac{\lambda}{4p^2}\right.\right)I(\varepsilon > 0)\right] f(\lambda)d\lambda,$$

where *skd* denotes a (generic) skew distribution, $p, 0 < p < 1$, is the skewness parameter, $N(x|a,b)$ is the normal pdf with mean a and variance b and $f(\lambda)$ is the pdf of the scale mixture variable λ. For the standard skew Student-t distribution, the pdf is given by

$$f_{st}(\varepsilon|0,1,\nu,p) = \frac{4p(1-p)\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\nu\pi}\,\Gamma\left(\frac{\nu}{2}\right)}\left(1+\frac{4\varepsilon^2}{\nu}(p-I(\varepsilon\le 0))^2\right)^{-\frac{\nu+1}{2}}$$

where ν is the degrees of freedom. The corresponding normal scale mixtures presentation is

$$f_{st}(\varepsilon|0,1,\nu,p) = 2\int_0^\infty \left[pN\left(\varepsilon\left|0,\frac{\lambda}{4(1-p)^2}\right.\right)I(\varepsilon \le 0) + (1-p)N\left(\varepsilon\left|0,\frac{\lambda}{4p^2}\right.\right)I(\varepsilon > 0)\right] \times IG\left(\lambda\left|\frac{\nu}{2},\frac{\nu}{2}\right.\right)d\lambda$$

where

$$IG(x|a,b) = \frac{1}{\Gamma(a)}b^a x^{-(a+1)}\exp\left(-bx^{-1}\right), \qquad x>0, a>0, b>0,$$

is the pdf of the inverse gamma $IG(a, b)$ distribution. With scale parameter σ, the standard deviation of the skew Student-t distribution is given by

$$\sigma_{st} = \frac{\sigma\sqrt{\nu[\pi^2(\nu-1)^2(1-3p+3p^2)-4(\nu-2)(1-2p)^2]}}{2\pi(\nu-1)p(1-p)\sqrt{\nu-2}}, \qquad \nu > 2.$$

If the scale mixture variable λ has a degenerate distribution at $\lambda = 1$, the normal scale mixtures representation of the skew normal distribution is

$$f_{sn}(\varepsilon|0,1,p) = 2\left(pN\left(\varepsilon\left|0, \frac{1}{4(1-p)^2}\right.\right) I(\varepsilon \le 0) + (1-p)N\left(\varepsilon\left|0, \frac{1}{4p^2}\right.\right) I(\varepsilon > 0)\right)$$

which after some algebra gives

$$f_{sn}(\varepsilon|0,1,p) = \frac{4p(1-p)}{\sqrt{2\pi}} \exp\left\{-2\varepsilon^2(p - I(\varepsilon \le 0))^2\right\}.$$

With scale parameter σ, the standard deviation of the skew normal distribution is given by

$$\sigma_{sn} = \frac{\sigma\sqrt{\pi(1-3p+3p^2)-2(1-2p)^2}}{2p(1-p)\sqrt{\pi}}.$$

The skew distributions, including the skew normal and skew Student-t, in Wichitaksorn et al. [16] are positively (negatively) skewed for $p < 0.5$ ($p > 0.5$) and are symmetric for $p = 0.5$.

Let $y_t^{(i)} = R_t^{(i)} - R_t^f$ and $x_t = R_t^m - R_t^f$. The CAPM regression for sector i is

$$y_t^{(i)} = \alpha^{(i)} + \beta^{(i)} x_t + \sigma^{(i)} \varepsilon_t^{(i)}$$

or

$$y_t^{(i)} = \mathbf{x}_t' \boldsymbol{\beta}^{(i)} + \sigma^{(i)} \varepsilon_t^{(i)},$$

where $\boldsymbol{\beta}^{(i)} = (\alpha^{(i)}, \beta^{(i)})'$, $\mathbf{x}_t = (1, x_t)'$, $\sigma^{(i)}$ is a scale parameter and $\varepsilon_t^{(i)}$ is the random error that follows a skew distribution with location 0 and scale 1.

3 Bayesian Inference

In this section, the implementation of CAPM relies on Bayesian computational approach with Gibbs sampling algorithm. To simplify the Gibbs sampler, we adopt the normal scale mixtures representation for the pdfs of the skew normal and skew Student-t error distributions.

Let $\boldsymbol{\theta}^{(i)} = (\boldsymbol{\beta}^{(i)}, \sigma^{(i)}, p^{(i)})'$, $\mathbf{y}^{(i)} = (y_1^{(i)}, \ldots, y_T^{(i)})'$, $\mathbf{X} = (\mathbf{x}_1, ..., \mathbf{x}_T)'$ is a $T \times 2$ matrix of covariates, and $f(\lambda_t^{(i)})$ are the pdf of the scale mixture variables. For sector i, the likelihood function is given by

$$L(\boldsymbol{\theta}^{(i)}|\mathbf{y}^{(i)}, \mathbf{X}) = \prod_{t=1}^{T} \left(\frac{2}{\pi}\right)^{1/2} p^{(i)} \left(1 - p^{(i)}\right) \int_0^{\infty} \frac{1}{\sqrt{\lambda_t^{(i)}}\sigma^{(i)}}$$
$$\exp\left\{-2\frac{(y_t^{(i)} - \mathbf{x}_t'\boldsymbol{\beta}^{(i)})^2}{\lambda_t^{(i)}\sigma^{2(i)}} \left(p^{(i)} - I(y_t^{(i)} \leq \mathbf{x}_t'\boldsymbol{\beta}^{(i)})\right)^2\right\} f(\lambda_t^{(i)}) d\lambda_t^{(i)}$$

for a generic skew error distribution. The likelihood functions for the skew normal and skew Student-t error distributions can be obtained accordingly. With independent prior distributions assigned to the parameters, we derive the full conditional posterior densities as follows. For simplicity, we omit the sector superscript (i).

Posterior Density for β

For the regression coefficients $\boldsymbol{\beta}$, we use a conjugate normal prior distribution with 2-dimensional mean vector $\boldsymbol{\beta}_0$ and 2×2 covariance matrix $\mathbf{B}_0$. The full conditional distribution of $\boldsymbol{\beta}$ can be easily shown to be a bivariate normal distribution, i.e.,

$$\boldsymbol{\beta}|\sigma^2, p, \boldsymbol{\lambda}, \mathbf{y}, \mathbf{X} \sim N_2\left(\bar{\boldsymbol{\beta}}, \sigma^2\mathbf{B}_1\right),$$

where $\boldsymbol{\lambda}$ is the T-dimensional vector of scale mixture variables, N_2 denotes the bivariate normal distribution,

$$\bar{\boldsymbol{\beta}} = \mathbf{B}_1\left[4\sum_{t=1}^{T} \lambda_t^{-1} y_t (p - I(y_t \leq \mathbf{x}_t'\boldsymbol{\beta}))^2 \mathbf{x}_t + \mathbf{B}_0^{-1}\boldsymbol{\beta}_0\right]$$

and

$$\mathbf{B}_1 = \left[4\sum_{t=1}^{T} \lambda_t^{-1}(p - I(y_t \leq \mathbf{x}_t'\boldsymbol{\beta}))^2 \mathbf{x}_t\mathbf{x}_t' + \mathbf{B}_0^{-1}\right]^{-1}.$$

Posterior Density for σ^2

A conjugate inverse gamma $IG(a_\sigma, b_\sigma)$ prior distribution is assumed for σ^2. As a result, we have the following inverse gamma full conditional distribution for σ^2.

$$\sigma^2|\boldsymbol{\beta}, p, \boldsymbol{\lambda}, \mathbf{y}, \mathbf{X} \sim IG\left(\frac{a_0 + T}{2}, \frac{b_0}{2} + 2\sum_{t=1}^{T} \lambda_t^{-1}(y_t - \mathbf{x}_t'\boldsymbol{\beta})^2 (p - I(y_t \leq \mathbf{x}_t'\boldsymbol{\beta}))^2\right).$$

Posterior Density for p
For skewness parameter p, we adopt a uniform $U(0, 1)$ distribution to express our ignorance about the parameter prior to data collection. Hence, it has a full conditional density given by

$$f(p|\boldsymbol{\beta}, \sigma^2, \boldsymbol{\lambda}, \mathbf{y}, \mathbf{X}) \propto p^T(1-p)^T \exp\left\{-\frac{2}{\sigma^2}\sum_{t=1}^{T}\lambda_t^{-1}(y_t - \mathbf{x}_t'\boldsymbol{\beta})^2(p - I(y_t \leq \mathbf{x}_t'\boldsymbol{\beta}))^2\right\}$$

for $0 < p < 1$. Since this posterior density is intractable, we use the independent kernel Metropolis-Hastings (M-H) algorithm to obtain a posterior realization from it. The proposal distribution is a truncated normal distribution to the $(0, 1)$ interval with mean and variance evaluated from maximizing the logarithm of the above full conditional density using a constrained numerical optimization (Newton's) method.

Posterior Density for λ_t
The full conditional distribution of the scale mixture variable λ_t depends on its scale mixture distribution. For the skew normal distribution, all λ_t degenerate to 1 and no MCMC updating step is required. For the skew Student-t distribution, the inverse gamma scale mixture distribution results in a conjugate full conditional distribution for λ_t, i.e. for $t = 1, ..., T$,

$$\lambda_t|\boldsymbol{\beta}, \sigma^2, \nu, p, \mathbf{y}, \mathbf{X} \sim IG\left(\frac{\nu+1}{2}, \frac{\nu}{2} + \frac{2(y_t - \mathbf{x}_t'\boldsymbol{\beta})^2(p - I(y_t \leq \mathbf{x}_t'\boldsymbol{\beta}))^2}{\sigma^2}\right).$$

Posterior Density for ν
For the degrees of freedom parameter ν, we adopt a uniform $U(2, 30)$ prior distribution and the full conditional density is given by

$$f(\nu|\boldsymbol{\beta}, \sigma^2, p, \boldsymbol{\lambda}, \mathbf{y}, \mathbf{X}) \propto \frac{\nu^{T\nu/2}}{(\Gamma(\nu/2))^T}\prod_{t=1}^{T}\lambda_t^{-\nu/2}\exp\left\{-\sum_{t=1}^{T}\frac{\nu}{2\lambda_t}\right\}, \quad 2 < \nu < 30.$$

Similar to p, simulation of posterior realizations from this non-standard density relies on the independent kernel M-H algorithm with a truncated normal density to the $(2, 30)$ interval as the proposal density. The mean and variance of the normal distribution are obtained from the constrained numerical maximization.

4 Data Analysis of Sectoral Index Returns of the Stock Exchange of Thailand

In this empirical study, the CAPM regression with various skew error distributions is used to analyze sectoral daily index returns of the Stock Exchange of Thailand (SET). Sectoral daily price indices are collected from Datastream. Upon data availability,

Table 1 Sector description and data availability

Sector	Description	Date
AGRI	Agribusiness	2/16/2001–8/31/2011
AUTO	Automotive	2/16/2001–8/31/2011
BANK	Banking	2/16/2001–8/31/2011
INFO	Information and communication technology	2/16/2001–8/31/2011
COMM	Commerce	2/16/2001–8/31/2011
CONM	Construction materials	2/16/2001–8/31/2011
ELEC	Electronic components	2/16/2001–8/31/2011
ENER	Energy and utilities	2/16/2001–8/31/2011
MEDIA	Media and publishing	2/16/2001–8/31/2011
FOOD	Food and beverage	2/16/2001–8/31/2011
FASH	Fashion	2/16/2001–8/31/2011
FINS	Finance and securities	2/16/2001–8/31/2011
HELTH	Health care services	2/16/2001–8/31/2011
HOME	Home and office products	2/16/2001–8/31/2011
TOUR	Tourism and leisure	2/16/2001–8/31/2011
MACH	Industrial materials and machinery	4/07/2006–8/31/2011
INSU	Insurance	2/16/2001–8/31/2011
PACK	Packaging	2/16/2001–8/31/2011
PROP	Property development	2/16/2001–8/31/2011
PETRO	Petrochemicals and chemicals	2/16/2001–8/31/2011
TRAN	Transportation and logistics	2/16/2001–8/31/2011

daily index return for 24 sectors are obtained for the period from February 2, 2001 to August 31, 2011. However, Mining, Professional Services, and Paper and Printing Materials are three sectors that are excluded from the analysis due to high volume of inactive price index. The remaining 21 sectors for analysis are presented in Table 1.

Let $r_t^{(i)}$ denote the price of sectoral index i at time t. The percentage return of index i is given by

$$R_t^{(i)} = 100 \times \left(\ln r_t^{(i)} - \ln r_{t-1}^{(i)} \right), \quad \text{for } i = 1, \ldots, I, \quad \text{and} \quad t = 1, \ldots, T.$$

The market rates, $R_t^m, t = 1, \ldots, T$ are the percentage returns of the SET index. For the risk-free rates, we use the yields of the 3-month treasury bill, which are deannualized and then converted to returns, $R_t^f, t = 1, \ldots, T$. Summary statistics of the risk-free rate, market rate and percentage returns of the 21 sectors are displayed in Table 2 and summary statistics of the individual risk premiums and market premium are given in Table 3. Obviously, risk premiums of all sectors are heavier-tailed than the normal distribution and are slightly skewed either positively or negatively. In particular, the Industrial Materials and Machinery sector

Table 2 Summary statistics of risk-free rate, market rate and percentage returns of 21 sectors

	Obs	Mean	S.D	Median	Skewness	Ex Kurtosis
R^f	2573	−0.005	0.345	0.000	−0.931	23.558
R^m	2573	0.046	1.584	0.046	−0.727	7.989
$R^{(i)}$						
AGRI	2573	0.065	1.341	0.038	−0.179	1.035
AUTO	2573	0.055	1.214	0.040	0.067	2.594
BANK	2573	0.032	1.977	−0.025	−0.407	5.304
INFO	2573	0.021	1.928	0.028	−0.549	9.135
COMM	2573	0.088	1.179	0.049	−0.041	2.085
CONM	2573	0.071	1.651	0.051	−0.149	1.988
ELEC	2573	0.006	1.646	−0.036	−0.094	3.005
ENER	2573	0.081	1.800	0.092	−0.373	5.405
MEDIA	2573	0.005	1.564	0.028	−0.508	8.116
FOOD	2573	0.078	1.137	0.072	−0.470	3.181
FASH	2573	0.038	0.875	0.030	0.021	10.688
FINS	2573	0.001	2.018	−0.055	−0.153	5.024
HELTH	2573	0.121	1.477	0.059	0.459	1.941
HOME	2573	0.048	1.346	0.048	−0.108	5.598
TOUR	2573	0.032	1.025	0.024	−0.225	5.120
MACH	1259	−0.014	1.981	−0.039	−2.161	30.191
INSU	2573	0.079	0.787	0.056	−0.002	4.825
PACK	2573	0.072	1.721	0.040	0.687	5.253
PROP	2573	0.062	1.857	0.080	−0.394	2.763
PETRO	2573	0.086	2.137	0.072	−0.015	0.988
TRAN	2573	0.021	2.017	0.015	−0.274	5.649

has the largest negative skewness and excess kurtosis amongst the 21 sectors. For these reasons, modeling risk premium using CAPM regression with a skew Student-t error distribution is a natural and better choice than using the normal error distribution.

In the Bayesian model implementation, we assign vague prior distributions to the two regression coefficients, an inverse gamma $IG(6, 0.4)$ prior distribution to σ^2, a uniform $U(0, 1)$ prior distribution to the skewness parameter p and a uniform $U(2, 30)$ prior distribution to the degrees of freedom ν. The Gibbs sampler was run for 5,000 iterations in the burn-in period, and then another 10,000 iterations to generate realizations which mimic a random sample of size 10,000 from the intractable joint posterior distribution for statistical inference. In the Gibbs sampling procedure, parameters with non-standard full conditional density function were simulated using Metropolis-Hastings algorithm and the convergence of the Markov chain is monitored using the trace plots.

Table 3 Summary statistics of market premium and 21 individual risk premiums

	Mean	S.D	Median	Skewness	Ex Kurtosis
$R^m - R^f$	0.051	1.620	0.061	−0.641	6.223
$R^{(i)} - R^f$					
AGRI	0.070	1.391	0.026	−0.036	0.451
AUTO	0.060	1.266	0.053	0.116	1.726
BANK	0.038	2.009	−0.011	−0.349	4.331
INFO	0.026	1.967	0.030	−0.450	7.411
COMM	0.093	1.236	0.057	−0.030	1.080
CONM	0.076	1.681	0.048	−0.135	1.187
ELEC	0.011	1.678	−0.001	−0.055	2.379
ENER	0.086	1.828	0.089	−0.365	4.408
MEDIA	0.010	1.600	0.022	−0.461	6.694
FOOD	0.083	1.187	0.065	−0.382	1.858
FASH	0.043	0.950	0.032	0.238	8.867
FINS	0.006	2.044	−0.029	−0.118	4.139
HELTH	0.126	1.518	0.054	0.453	1.527
HOME	0.053	1.400	0.038	−0.022	4.194
TOUR	0.038	1.090	0.029	−0.163	3.286
MACH	−0.021	1.995	−0.039	−2.150	28.555
INSU	0.084	0.857	0.058	0.055	3.499
PACK	0.078	1.761	0.041	0.616	4.647
PROP	0.067	1.884	0.080	−0.392	1.987
PETRO	0.091	2.164	0.082	−0.045	0.833
TRAN	0.026	2.038	0.038	−0.241	4.959

Table 4 reports the posterior means, standard deviations and 95 % credible intervals for the parameters α, β, ν, p and σ_{st} of the CAPM regression with skew Student-t error distribution for the 21 sectors. For the intercept α, it is significantly different from zero for the sectors AGRI, AUTO, BANK, CONM, ELEC, FINS, HOME, MACH, PACK, PETRO and TRAN and its estimated value is negative for 20 sectors.

For the slope β, BANK has an estimate which is significantly greater than one and therefore the banking section is considered to be more risky than the market. It is also the riskiest sector amongst all 21 sectors under study. A reason is that Thailand is still on the track of recovering from the 1997 financial crisis during the study period and the banking sector remains vulnerable. Though their β values are close to but less than 1, the sectors ENER, PROP, PETRO and FINS are riskier than other sectors. Note that the energy sector is risky due to the upsurge of the global fuel prices, the government's energy policy and the privatization of the nation's largest petroleum enterprise. Similar to the banking sector, the property sector and finance sector have not yet fully recovered from the 1997 financial crisis. The petrochemical sector is

Table 4 Estimation results of CAPM with skew Student-t error

	$\alpha^{(i)}$				$\beta^{(i)}$				$\nu^{(i)}$				$p^{(i)}$				$\sigma_{st}^{(i)}$			
	Mean	S.D	2.50%	97.50%	Mean	S.D	2.50%	97.50%	Mean	S.D	2.50%	97.50%	Mean	S.D	2.50%	97.50%	Mean	S.D	2.50%	97.50%
AGRI	−0.125	0.033	−0.190	−0.060	0.437	0.013	0.411	0.463	3.639	0.261	3.166	4.199	0.447	0.011	0.425	0.469	0.374	0.024	0.327	0.420
AUTO	−0.120	0.032	−0.184	−0.056	0.395	0.013	0.370	0.420	3.824	0.271	3.324	4.387	0.449	0.012	0.425	0.472	0.358	0.021	0.317	0.399
BANK	−0.127	0.030	−0.185	−0.068	1.089	0.011	1.066	1.110	5.090	0.474	4.266	6.106	0.461	0.012	0.438	0.484	0.378	0.019	0.341	0.415
INFO	−0.052	0.042	−0.134	0.030	0.876	0.016	0.844	0.908	4.784	0.415	4.049	5.662	0.490	0.012	0.467	0.513	0.527	0.026	0.475	0.578
COMM	−0.001	0.032	−0.063	0.062	0.451	0.012	0.427	0.475	4.925	0.432	4.162	5.836	0.476	0.012	0.454	0.499	0.403	0.020	0.363	0.441
CONM	−0.101	0.035	−0.169	−0.031	0.817	0.012	0.793	0.840	5.144	0.455	4.329	6.102	0.456	0.012	0.432	0.480	0.425	0.020	0.385	0.463
ELEC	−0.124	0.038	−0.196	−0.049	0.580	0.015	0.550	0.610	3.818	0.293	3.281	4.436	0.472	0.011	0.451	0.494	0.451	0.028	0.393	0.506
ENER	0.004	0.028	−0.051	0.058	0.998	0.011	0.976	1.019	3.827	0.289	3.307	4.442	0.487	0.011	0.464	0.510	0.323	0.020	0.283	0.361
MEDIA	−0.072	0.037	−0.144	0.002	0.587	0.015	0.558	0.617	4.120	0.326	3.569	4.818	0.484	0.012	0.461	0.507	0.441	0.025	0.394	0.489
FOOD	−0.040	0.030	−0.098	0.021	0.428	0.013	0.403	0.452	5.046	0.483	4.174	6.067	0.462	0.012	0.440	0.485	0.392	0.021	0.351	0.431
FASH	0.004	0.020	−0.035	0.043	0.230	0.009	0.213	0.248	3.005	0.188	2.659	3.394	0.489	0.011	0.469	0.510	0.208	0.017	0.174	0.241
FINS	−0.168	0.037	−0.243	−0.096	0.933	0.016	0.903	0.965	3.902	0.300	3.375	4.555	0.466	0.012	0.443	0.489	0.429	0.026	0.379	0.481
HELTH	−0.061	0.034	−0.128	0.005	0.329	0.015	0.299	0.358	2.845	0.163	2.551	3.180	0.459	0.011	0.438	0.481	0.322	0.028	0.265	0.375
HOME	−0.067	0.031	−0.126	−0.007	0.386	0.014	0.359	0.414	2.831	0.164	2.528	3.168	0.472	0.011	0.451	0.494	0.279	0.025	0.228	0.326
TOUR	−0.043	0.025	−0.092	0.006	0.222	0.010	0.201	0.243	2.846	0.168	2.540	3.194	0.473	0.011	0.451	0.495	0.233	0.021	0.191	0.273
MACH	−0.276	0.055	−0.383	−0.168	0.570	0.025	0.522	0.618	2.954	0.247	2.534	3.503	0.434	0.016	0.403	0.466	0.351	0.042	0.271	0.435
INSU	−0.012	0.019	−0.048	0.024	0.177	0.008	0.161	0.193	2.728	0.159	2.439	3.062	0.459	0.011	0.437	0.481	0.164	0.017	0.130	0.196
PACK	−0.129	0.034	−0.197	−0.063	0.407	0.014	0.380	0.435	2.466	0.135	2.216	2.750	0.457	0.011	0.436	0.478	0.258	0.037	0.180	0.325
PROP	−0.015	0.034	−0.082	0.052	0.960	0.012	0.935	0.984	4.719	0.427	3.999	5.684	0.490	0.011	0.468	0.512	0.434	0.023	0.391	0.481
PETRO	−0.099	0.046	−0.191	−0.009	0.934	0.018	0.899	0.969	4.115	0.333	3.535	4.815	0.469	0.012	0.446	0.492	0.541	0.032	0.479	0.602
TRAN	−0.145	0.039	−0.223	−0.069	0.837	0.016	0.805	0.869	3.506	0.229	3.099	4.015	0.470	0.011	0.447	0.492	0.434	0.027	0.382	0.487

also risky due to the drop in plastic price in the global market that has caused unsound financial conditions to some companies in this sector.

For the degrees of freedom ν, its estimate ranges from approximately 2.5–5.1 for the 21 sectors and this signifies that the error distribution is heavy-tailed than the normal distribution and the choice of a Student-t error distribution is reasonable. For the skewness parameter p, the estimate ranges from 0.43 to 0.49 for all 21 sectors and this means that the risk premiums are slightly positively skewed. However, the skewness parameter of five sectors namely INFO, ENER, MEDIA, FASH and PROP is not significantly different from 0.5 and perhaps a symmetric Student-t distribution will provide a better fit to the risk premium. The standard deviation of the skew Student-t distribution, σ_{st}, is estimated to be within the range from 0.16 for INSU and 0.54 for PETRO. Though not given here, the standard deviation of the skew normal distribution, σ_{sn}, is, on average, twice bigger than σ_{st}.

Finally, to compare the normal, skew normal and skew t distributions in fitting the risk premium for the 21 sectors, Bayesian information criterion (BIC) and deviance information criterion (DIC) are used. Table 5 presents the BIC and DIC values of the three error distributions for each of the 21 sectors. Obviously, the skew Student-t

Table 5 Model Comparisons using DIC and BIC

	DIC			BIC		
	Normal	Skew normal	Skew student-t	Normal	Skew normal	Skew student-t
AGRI	8,076	8,060	**7,749**	8,102	8,084	**7,779**
AUTO	7,693	7,654	**7,301**	7,718	7,679	**7,331**
BANK	6,762	6,732	**6,531**	6,788	6,757	**6,562**
INFO	8,623	8,622	**8,352**	8,648	8,647	**8,383**
COMM	7,102	7,098	**6,908**	7,127	7,123	**6,939**
CONM	7,345	7,316	**7,133**	7,370	7,341	**7,164**
ELEC	8,770	8,759	**8,401**	8,796	8,783	**8,432**
ENER	7,002	6,998	**6,635**	7,027	7,023	**6,666**
MEDIA	8,228	8,227	**7,946**	8,253	8,252	**7,976**
FOOD	6,888	6,883	**6,747**	6,913	6,907	**6,778**
FASH	6,418	6,411	**5,713**	6,443	6,436	**5,744**
FINS	8,403	8,382	**8,068**	8,428	8,406	**8,099**
HELTH	8,972	8,922	**8,435**	8,998	8,946	**8,466**
HOME	8,319	8,295	**7,693**	8,344	8,320	**7,724**
TOUR	7,294	7,293	**6,727**	7,320	7,318	**6,757**
MACH	4,687	4,688	**4,298**	4,710	4,710	**4,326**
INSU	5,964	5,949	**5,335**	5,990	5,974	**5,366**
PACK	9,642	9,590	**8,814**	9,667	9,615	**8,845**
PROP	7,579	7,578	**7,397**	7,604	7,602	**7,428**
PETRO	9,338	9,315	**9,043**	9,364	9,339	**9,074**
TRAN	9,129	9,102	**8,601**	9,155	9,127	**8,632**

CAPM outperforms the normal CAPM and skew normal CAPM in all cases. One can see that there is only a minor improvement from the normal distribution to the skew normal distribution but the improvement of using the skew Student-t distribution is remarkable.

5 Concluding Remarks

This paper extends the CAPM to the skew Student-t distribution for modeling individual risk premiums. By expressing the skew Student-t density function into a normal scale mixture representation, the computational burden of implementing MCMC algorithms can be significantly reduced. In addition, the skewness parameter p, which ranges from 0 to 1, provides a simple interpretation. The distribution is symmetric if $p = 0.5$ and it is positively (negatively) skewed if $p < 0.5$ ($p > 0.5$). In the empirical study of daily return data of sectoral indices of the Stock Exchange of Thailand using CAPM, it confirms that returns of all sectoral indices are heavy-tailed and slightly positively skewed, albeit some of them are insignificantly different from 0.5 at the 5 % significance level. Meanwhile, model comparisons based on BIC and DIC confirm that the skew t distribution is superior to the normal and skew normal distributions in CAPM of sectoral indices of Thailand.

The proposed CAPM with skew errors can be easily extended to CAPM with time varying market risk [7] and multi-regime CAPM model with GARCH-type volatility [4]. Moreover, the proposed MCMC algorithm can be easily modified for handling multivariate skew Student-t distribution in asset pricing and portfolio selection ([1]).

References

1. Adcock, C.J.: Asset pricing and portfolio selection based on the multivariate extended skew-Student-t distribution. Ann. Oper. Res. **176**, 221–234 (2010)
2. Andrews, D.F., Mallows, C.L.: Scale mixtures of normal distributions. J. R. Stat. Soc. Ser. B **36**, 99–102 (1974)
3. Bollerslev, T.: A conditionally heteroskedastic time series model for speculative prices and rates of return. Rev. Econ. Stat. **69**, 542–547 (1987)
4. Chen, C.W.S., Gerlach, R.H., Lin, A.M.H.: Multi-regime non-linear asset pricing models. Quant. Financ. **11**, 1421–1438 (2011)
5. Choy, S.T.B., Chan, J.S.K.: Scale mixtures distributions in statistical modelling. Aust. N. Z. J. Stat. **50**, 135–146 (2008)
6. Fridman, M., Harris, L.: A maximum likelihood approach for non-Gaussian stochastic volatility models. J. Bus. Econ. Stat. **16**, 284–291 (1998)
7. Jagganathan, R., Wang, Z.: The conditional CAPM and the cross-section of expected return. J. Financ. **51**, 3–53 (1996)
8. Jensen, J.L., Pedersen, J.: Ornstein-Uhlenbeck type process with non-normal distribution. J. Appl. Probab. **36**, 389–402 (1999)
9. Lintner, J.: The valuation of risk assets and the selection of risky investments in stock portfolios and capital budgets. Rev. Econ. Stat **47**, 13–37 (1965)

10. Mandelbrot, B., Hudson, R.L.: The (Mis)Behavior of Markets: A Fractal View of Financial Turbulence. Basic Books, New York (2004)
11. Markowitz, H.M.: Portfolio selection. J. Financ. **7**, 77–91 (1952)
12. Markowitz, H.M.: Foundations of portfolio theory. J. Financ. **46**, 469–477 (1991)
13. Markowitz, H.M.: The early history of portfolio theory: 1600–1960. Financ. Anal. J. **55**, 5–16 (1999)
14. McDonald, J.B., Michelfelder, R.A.: Robust regression estimation methods and intercept bias: a capital asset pricing model application. Multinatl. Financ. J. **13**, 293–321 (2009)
15. Sharpe, W.F.: Capital asset prices: a theory of market equilibrium under conditions of risk. J. Financ. **19**, 425–442 (1964)
16. Wichitaksorn, N., Choy, S.T.B., Gerlach, R.: A generalized class of skew distributions and associated robust quantile regression models. Can. J. Stat. **42**, 579–596 (2014)

Strategic Path to Enhance the Impact of Internet Broadband on the Creative Economy in Thailand: An Analysis with Structural Equation Model

Sumate Pruekruedee and Komsan Suriya

Abstract This paper aims at finding the impact of internet broadband on the creative economy in Thailand. It investigates the strength of linkages from the usage behavior of internet broadband to the goals of the production of creative products and the adjustment of organization that facilitates the production, via the applications of the making the digital database and analysing the data. It uses the structural equation model (SEM) to model this multi-layer relationship. The data are collected from field survey of 400 Small and Medium Enterprises (SMEs). The results reveal that the usage of internet broadband for communication is crucial for the production of creative products. It also discovers that the digital database is important for the modification of organization structure. The digital database is based on the usage of internet broadband in marketing research, searching for contents and collects transactional data of clients. These paths leads to a huge opportunity for mobile broadband technology to enhance the creative economy when the functions for communication, searching for contents and marketing research can be done conveniently by mobile devices.

1 The Creative Economy and Ways to Make It Successful

John Howkins [4] is the founder of the concept of the creative economy. He defines the creative product as an economic good or service that results from creativity and has economic value. Then he defines the creative economy as the transactions in these creative products. He includes four major industries into the creative economy which are the copyright industries, the patent industries, the trademark industries and the design industries.

The creative economy includes fifteen sectors by the definition of Howkins. They are advertising, architecture, art, crafts, design, fashion, film, music, performing arts,

S. Pruekruedee (✉)
Economics Department, School of Management, Mae Fah Luang University, Chiang Rai, Thailand
e-mail: Sumate.pru@gmail.com

K. Suriya
Faculty of Economics, Chiang Mai University, Chiang Mai, Thailand
e-mail: suriyakomsan@gmail.com

V.-N. Huynh et al. (eds.), *Econometrics of Risk*, Studies in Computational Intelligence 583, DOI 10.1007/978-3-319-13449-9_31

publishing, research and development, software, toys and games, TV and radio and video games. Most of the sectors are labeled as soft innovation by Stoneman [7] later in 2010.

It is interesting to find out the strategy to make the creative economy success. In this case, John Howkins [4] also develop ten rules for success in the creative economy. The rules are listed as follows:

Rule 1: Invent yourself
Rule 2: Put the priority on ideas not on data
Rule 3: Be nomadic
Rule 4: Define yourself by your own (thinking) activities
Rule 5: Learn endlessly
Rule 6: Exploit fame and celebrity
Rule 7: Treat the virtual as real and vice versa
Rule 8: Be kind
Rule 9: Admire success openly
Rule 10: Be ambitious

Moreover, in his book, Howkins constructs seven laws for the creative economy and writes them under the title of each chapter as follows:

Law 1: Create or die.
Law 2: Patents and copyright are the currency of the information age.
Law 3: $CE = CP \times T$ which means the value of the creative economy equals to the price of the creative subject times the rounds or copies of its sales.
Law 4: Creativity is a proper job.
Law 5: Stories come first.
Law 6: The new economy is creative plus electronics.
Law 7: Creative capital arises whenever someone holds back an idea, or part of an idea, for the future.

These suggestions of rules and laws are from the intuition of Howkins without any quantitative evidence supporting his idea. Therefore, it is risky for practitioners to follow his rules and laws then expects for the success in the creation of new products or innovation.

This paper tries to find the strategy bases on quantitative analysis to guide the practitioners to the success in the production of creative product and build a good organization to facilitate the creativity. It may also find some evidences to respond to Howkins upon his rules and laws for the success in the creative economy.

2 Methodology and Data

This paper uses the structural equation model (SEM) to analyze the data from the survey of 400 small and medium sized enterprises (SMEs) in Thailand. The data were collected by Poonchan [6] in 2012 in order to fulfill her thesis in 2013. The original

survey focuses on internet broadband without the separation between fixed and mobile broadband. The analysis in Poonchan [6] is based on single-layer analysis with seemingly unrelated regression (SURE).

The structural equation model (SEM) differs from SURE in that SEM is a multi-layer analysis [5]. It can show the paths that carry the effect of the policy variables on the target variables via the transaction variables. Therefore, the analyst can view not only the direct effect of the policy variables, as seen in the single-layer analysis, but also its indirect effect that acts upon some activities that lie in the middle of the process. This enables the analyst to choose appropriate strategies based on the significant path in the model to enhance the effects of the policy variables on the target variables.

In this study, there are three layers in the analysis. The first layer is on the usage behavior of internet broadband. This is to explore how staffs use the internet broadband and the functions that internet broadband can serve the organization. Most of them can be seen as routine procedures without specific focuses on the creative economy. The second layer is on the applications of internet broadband. This layer emphasizes two specific functions which are making the digital database and analysis the data. It focuses on the activities that support the creation of the creative product and the construction of good organization to facilitate the creativity. The third layer is the goals of the applications of the internet broadband. They focus specifically on the creative economy. The first goal is to produce the creative products and the second goal is to create a good environment in the organization that empowers and attracts staffs to create and deliver the creative products.

First layer: Usage behavior of internet broadband

MKT: The usage of internet broadband in market research
QO: Online questionnaire
SOCM: The creation of online society that use and give comments to the products
PRO: The distribution of testers on internet

SEAR: The usage of internet broadband in searching for information
TECH: Searching for new production techniques or new services
CREA: Searching for ideas to create new designs or new products
CRIT: Searching for customer's feedback

TRAN: The usage of internet broadband in doing transactions
PROD: Delivery of digital products
SERV: After sales service
ODR: Receiving purchasing orders via internet
PAY: Receiving payments via internet

COM: The usage of internet broadband in communication
INTER: Communication within organization
GOV: Communication between the organization and government agencies
SUPP: Communication between the organization and suppliers

Second layer: Applications of internet broadband

DB: The application of internet broadband in making digital database
SPEC: Digital database to search for specialists
EXPR: Digital database to store knowledge and experience of the organization
CONS: Digital database to store transactions with customers and other organizations

AN: The application of internet broadband in analyzing data
GOAL: The analysis to determine the targets of the organization
FIN: Financial analysis
SATIS: The analysis of customers satisfaction
GRW: The analysis of the market growth

Third layer: Goals of the applications of internet broadband

CO: Organizational management to facilitate the creativity
INC: Updating the latest creation in the organization to staffs
CONC: Updating the latest creation in the organization to customers
KM: Application of knowledge management in the organization

CP: Production of creative products
NEW: Production by the creation of new designs with new idea
CUL: Production by applied local intelligence
TRY: Production by trial and error

3 Result

The results of the model are presented in Fig. 1, it is a conventional approach to present SEM result [1, 2]. It is clear that there are strategic paths that carry the effect of the usage behavior of internet broadband to the goals of the creation of creative products via the activities of both the making of digital database and the analyzing the data.

The strongest path is the linkage between the usage of internet broadband for communication (COM in the first layer) to the production of the creative products (CP in the third layer). This is a direct link. The loading factor in the first layer are distributed quite equally among the communication within organization (INTER, 0.767), communication between the organization and government agencies (GOV, 0.773), and the communication between the organization and customers (SUPP, 0.741).

Table 1 provides conventional fit indices both absolute and incremental one. Incremental fit indices: Comparative Fit Index (CFI) and Tucker-Lewis Index (TLI) are at good fit level and acceptable fit level respectively. Absolute fit indices: Standard Root Mean Square Residual (SRMR) is at good fit level, Root Mean Square Error of Approximation (RMSEA) is at acceptable fit level, while Chi-square and its p-value are not fit at significant level as usual when the number of observation is high [5].

The usage of internet broadband for communication also plays a significant role via the analyzing of data (AN in the second layer) in order to affect the production

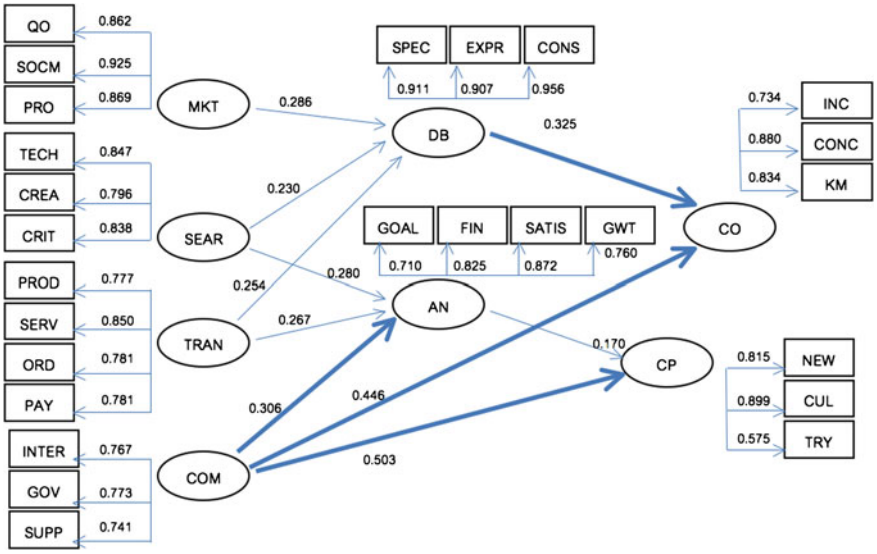


Fig. 1 The strategic path to enhance the impact of internet broadband on the creative economy

Table 1 Goodness of fit of the model

Chi-square (df)	p-value	SRMR	RMSEA	CFI	TLI
522.93 (277)	0.000	0.049**	0.050*	0.957**	0.950*

Remark ** is good fit level
* is acceptable fit level

of creative products. Even though this linkage is quite weaker than the direct effect, it is significant enough to ensure the presence of the indirect effect.

For the building of the creative organization which facilitate the production of creative products (CO in the third layer), it can be seen that both the direct effect from the usage of internet broadband for communication (COM in the first layer) and the indirect effect via the making of digital database (DB in the second layer) are significant. Interestingly, the making of digital database are the outputs of other usages of internet broadband (MKT, SEAR and TRAN in the first layer) but except for the communication. Moreover, the analyzing of data (AN in the second layer) does not affect the building of the creative organization.

4 Discussions

The results are consistent with the Thai culture in creating the creative economy. The Thai verify the data and information by asking peers rather than relying on data mining techniques. Therefore, the communication is crucial for the production

of creative products. In the era of internet broadband when data are easy to find but hard to believe, the peer reviewing by verbal communication is critical to filter the information in order to trust them. These validated information enters the production of creative products as inputs. It can be also viewed as the fifth kind of factors of production when information is another input for the production of creative products.

The building of organization relies on fact rather than just information reviewed by peers. The digital database collecting the transactions, market-research results, and searching results are crucial for the modification of the organizational structure and functions. This is relevant to the management by numbers suggested by Bill Gates [3]. The building of organization cannot rely on only the peer suggestions but must be based on facts and numbers. This is to ensure the efficiency of functions of the organization that can serves clients appropriately. Indeed, as Bill Gates mention, the digital database supports this function extremely well.

To respond to rules and laws proposed by Howkins [4], the study seems to support rule number 2, put the priority on ideas not on data, when the strongest linkage delivers the effect from the usage of internet broadband for communication to exchange idea and analyze the information to create the creative products. At the same time, the linkage from the digital database to the production of creative products is insignificant. Moreover, the study may support rule number 5, learn endlessly, when the usage of internet broadband for marketing research, searching for contents and collect the database of transaction are significant linkages to the success in the building good organization that facilitate the production of creative products.

The study may also convince readers that law number 5 of Howkins, stories come first, and law number 6, the new economy is creative plus electronics, are correct. The stories to create the creative products can mainly come from communication. Stories that are exchanged among peers within the organization may be the most important factor that deliver the success. The 6th law of Howkins are eventually confirmed by the significance of the usage and application of internet broadband on the production of creative products, building the organization that facilitate the creation, and the creation of the creative economy as appeared from the results of many significant linkages and paths in this structural equation model.

5 Conclusions

This paper analyses the paths that carry the effects of internet broadband on the production of creative products and the building of organization to facilitate the production of the creative products via the activities of making digital database and analyzing the data. It uses the structural equation model (SEM) to reflect the multilayer effects from the first layer of the usage behavior of internet broadband, the second layer of the applications of internet broadband, and the third layer of the goal to achieve the creative products and creative organization. It uses the data from 400 SMEs collected by field survey in 2012.

The results reveal that the usage of internet broadband for communication is the most important factor for the production of creative products. It also discovers that the making of digital database is crucial for the building the organization that supports the creation of creative products. These digital databases are based on the usage of internet broadband in doing marketing research, searching for contents and collect information on transactions of clients.

The strategic path to enhance the impact of internet broadband on the creative economy, thus, is to strengthen the communication at all levels and among all agents, and to manage the organization by facts and numbers. The combination of both strategies will ensure the success to achieve both the creative products and the creative organization that facilitate the production of these creative products. This strategy brings the mobile broadband technology to become an important catalyze when the technology suits: the activity of communication, searching for contents and knowledge, and assisting the marketing research, e.g. the filling the online questionnaire, so well and conveniently. Thus, it can be believed that the creative economy will be growing by the enhancement by both the internet broadband as a whole and especially the mobile broadband.

References

1. Chen, S., Raab, C.: Predicting resident intentions to support community tourism: toward an integration of two theories. J. Hospitality Marketing Manag. **21**, 270294 (2012)
2. Dyer, P., Gursoy, D., Sharma, B., Carter, J.: Structural modeling of resident perceptions of tourism and associated development on the Sunshine Coast. Aust. Tourism Manag. **28**, 409422 (2007)
3. Gates, B.: Business At the Speed of Thought. Grand Central Publishing, New York (1999)
4. Howkins, J.: The Creative Economy: How People Make Money From Ideas. Penguin Books, London (2007)
5. Kline, R.: Principles and Practice of Structural Equation Modeling, 3rd edn. The Guilford Press, New York (2011)
6. Poonchan, Thitimanan, Effects of Broadband Internet in Driving the Creative Economy of Small and Medium Sized Enterprises in Thailand. Master dissertation. Graduate School, Chiang Mai University, Thailand, 2013
7. Stoneman, P., Innovation, S.: Economics, Product Aesthetics and the Creative Industries. Oxford University Press, Oxford (2010)

Impact of Mobile Broadband on Non-life Insurance Industry in Thailand and Singapore

Niwattisaiwong Seksiri and Komsan Suriya

Abstract This study investigates whether the non-life insurance industries in Thailand and Singapore are growing by the usage of mobile broadband technology. It examines the impact of the third generation (3G) mobile phone in Thailand and compares the same to the impact in Singapore as a benchmark. It also figures out the impact of the fourth generation (4G) mobile phone in Singapore to learn the impact that may occur in Thailand after the country begins to offer the service later. It uses the piecewise regression to analyze the trend before and after those countries adopted the mobile technologies. The results show that the mobile broadband both in terms of 3G and 4G technologies does not significantly affect the growth of the non-life insurance industry in both countries. They reveal that the insurance companies are incapable of catching up with the "mobile integration" to such a degree that they can use the mobile broadband to create opportunity, boost sales, and make profit for their businesses. Insurance companies should emphasize their priorities in order to encourage their customers to access information regarding their products, purchase their products, and notify any accidents or incidents on a real-time basis via mobile broadband. These strategies will enhance the impact of mobile broadband technology on the non-life insurance industry and empower the industry to grow in the era of the mobile broadband.

1 Background and Rationale

The wireless and mobile broadband technology has brought about a lot of opportunities to the insurance business, such as the opportunity to directly communicate with existing and new customers, the opportunity to present new products faster, and the opportunity to enhance the efficiency of the insurers internal operations. A growing number of customers has turned to using insurance services through wireless electronic devices such as smartphones and tablets. The advantages of utilizing wireless and mobile broadband in the insurance business include the fact that customers gain

N. Seksiri(✉) · K. Suriya
Faculty of Economics, Chiang Mai University, Chiang Mai 52000, Thailand
e-mail: s.niwattisaiwong@alumni.lse.ac.uk

V.-N. Huynh et al. (eds.), *Econometrics of Risk*, Studies in Computational Intelligence 583, DOI 10.1007/978-3-319-13449-9_32

access to the services and insurers anywhere and anytime, while insurance companies could use the technology to improve their service efficiency. Customers usually access mobile webs to view product descriptions, report claims, and pay insurance premiums [7]. In general, insurance companies use the wireless and mobile broadband technology to connect with wireless electronic devices in order to (1) publicize their products, (2) improve customer services, and (3) develop new products [1].

The wireless and mobile broadband technology has high potential for bolstering insurance business operations, helping insurance companies maintain current customers, solving problems, and improving facilities by offering more convenient and faster services to customers. Insurers can also attract new customers as insurance information is being disseminated through the technology. Existing customers might as well recommend products and services to new customers. Moreover, the growing number of foreign insurance businesses tends to use wireless and mobile broadband in their operations. This study will examine whether the utilization of wireless and mobile broadband can propel the insurance industry in Thailand.

The aim of this research is to study the impact of mobile broadband on the insurance business in Thailand in contrast with that of Singapore. The study results will provide the basic data for the government to outline a proactive policy in developing the countrys mobile broadband infrastructure which draws level with international standards.

2 Non-life Insurance Business in Thailand

An initial study found that every insurance company in Thailand provides Internet-based services through their websites. Interestingly, every company's website can be accessed through wireless communications devices like smartphones and tablets. However, only some non-life insurance companies, such as Viriyah Insurance and Bangkok Insurance, are more interested in providing information through mobile devices by developing a mobile platform for their websites and have created mobile applications in order to serve customers faster. These applications let customers receive product information, file claims, view the insured's profiles, pay premiums, and contact the insurance companies. General insurance products mostly have common standards. They offer shorter period of coverage, mostly require a single-premium payment, and provide a shorter insurance term. Travel insurance is an example of general insurance, offering protection during the trip only and requiring a single premium payment.

Insurance companies are trying to improve their electronic services to meet customer demands, save expenses on insurance brokerage or agency, and enhance their competitiveness. The main advantage of electronic services is that the services are not limited by working hours or locations. Customers can browse through company websites anytime and from anywhere. However, customers are often found to be using online services only when they search for information, and they still opt to buy insurance or report claims via agents or brokers with whom they have direct contact.

Table 1 Number of policies sold through non-life insurance distribution channels (*unit* million)

Year	Agency	Broker	BA	Post	Telesale	DI	CS	Internet	Others	Total
2555	9.40	23.40	8.97	0.49	4.91	2.16	2.75	1.13	0.38	53.60
2556	10.71	36.62	7.52	0.002	7.69	1.61	2.81	1.74	0.56	69.25
Total	20.11	60.02	16.49	0.49	12.60	3.77	5.56	2.87	0.94	122.85

Source Office of the Insurance Commission (OIC) [8]
BA Banc-Assurance, *DI* Direct Contact with Insurers, *CS* Corporate Solutions

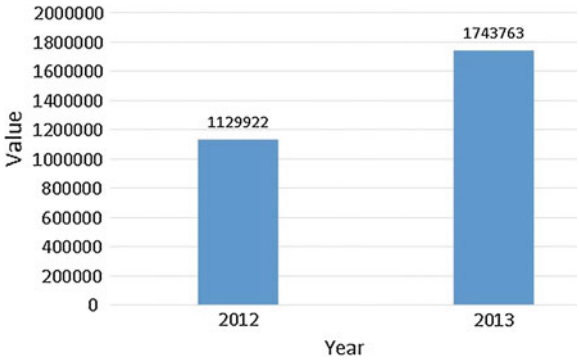


Fig. 1 The number of non-life insurance policies purchased via the Internet during 2012–2013. *Source* office of the insurance commission (OIC) [9]

Consider the statistics of distribution channels in non-life insurance, as shown in Table 1. The most common channel that the Thai people use to buy non-life insurance and process claims is brokerage, followed by agency.

The main reason customers still prefer using services through agents or brokers is because of the limitations of electronic services in their general characteristics. Generally, electronic systems require skillful users and high self-dependence because no staff will be there for customer support. Businesses that have complicated service platforms [10], such as the insurance business, find it difficult to provide electronic services as customers might not have adequate understanding of the procedures and other conditions, because of which the customers may be unable to buy products or services completely by themselves.

However, purchases of non-life insurance products via the Internet have increased by around 54 % during 2012–2013, as presented in Fig. 1. The statistics is apparently a sign which indicates that customers have begun to rely more on the electronic system when it comes to buying non-life insurance.

3 Non-life Insurance Business in Singapore

Singapore's insurance market is one of the most developed markets in Asia, with an impressive growth rate, to which can be attributed the increasing numbers of aging population and higher income. The life insurance business dominates Singapore's

insurance market, with it being the main driver of the market's growth. However, the non-life sector has been gaining momentum in recent years, with a large number of sub-sectors owning small shares in the overall industry.

The non-life insurance business in Singapore expanded strongly at the end of 2013, with total gross premiums amounting to 3.5 billion dollars, an increase of 4.54 %. Total net earned premiums increased by 5.46 % to 2.5 billion dollars. Such a growth was in line with that of the previous year when gross and net earned premiums grew by 5.41 and 6.47 %, respectively. However, underwriting profit decreased by 1.10 % to 285 million dollars, a huge decline from the 15.99 % growth to 288.21 million dollars in 2012.

As for the insurance distribution channels in Singapore [3], more insurers are using the multi-distribution channel strategy as they attempt to balance the needs of different consumer groups against the distribution costs. A certain distribution channel is not necessarily appropriate for every product and service. Distribution channels in Singapore are Insurance Agents, Trade Specific Agents (TSAs), Insurance Brokers, Online Internet Portals, and Direct Marketing. It is only the Online Internet Portals that will be discussed in this paper.

4 Online Internet Portals in Non-life Insurance Business in Singapore

4.1 Corporate Portals

People are spending more time on the Internet thanks to the growing network of online information. The Internet has also become a new market for online insurance products.

In recent years, numerous insurance companies have emerged in Singapore. These companies offer motor, travel, residence, and personal accident insurance products. Singapore's business platform leads to the possibility of offering insurance directly online, with availability of a 24 h customer support center as well as a full-fledged claim department.

In Singapore, most of the insurance products sold online are personal products covering home, motor, golf, travel, card protection, personal accident, income during hospitalization, and, even, maid packages. General insurers have to sell the products through their websites which provide quotations and accessibility to web brochures, proposal forms, and policy details.

4.2 Individual Portals

Apart from corporate portals, individual portals are also available. These portals are "passive", meaning they do not directly sell insurance plans and are not required to

be registered with the Agents Registration Board (ARB) of the General Insurance Association (GIA) of Singapore.

These passive portals are not involved with any sales distribution functions. For instance,

1. These portals do not sell or provide product advice, but offer product information without comments on product features, including premiums not being considered product advance;
2. They are not involved with any premium collections or insurance proposals;
3. They do not issue policies on behalf of insurers; and
4. The fees must not depend on the premiums.

It should be noted that the fee for online services must not be tied to the premiums of the products being sold. A premium-based fee paid will be considered sales commission which is involved in the distribution process.

Any online portals not meeting the above mentioned criteria will be deemed insurance intermediaries as defined by the Monetary Authority of Singapore (MAS). In that case, they will have to be registered with the ARB as agents, or licensed by the MAS as brokers.

5 Accessibility to Mobile Broadband in Singapore

The data from the Infocomm Development Authority of Singapore (IDA) [5] demonstrates accessibility to high-speed Internet through mobile broadband in Singapore during 1997–2014. Figure 2 shows that Singapores total number of mobile phone users has been increasing and tends to continue increasing.

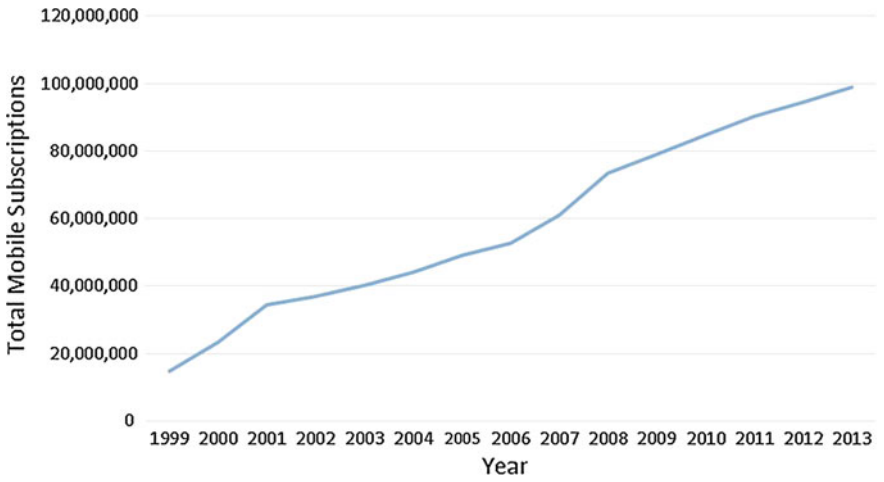


Fig. 2 The total number of mobile phone users [5]

The number of mobile users can be divided according to the mobile networks, namely 2G, 3G, and 4G. Note that Singapore has adopted the 3G technology since April 2001 and 4G since July 2012.

Figure 3a shows the total number of 2G users, which shows a declining trend, while the number of 3G users is obviously increasing, as presented in Fig. 3b. By comparing the numbers of 2G and 3G users, it can be found that, during 2005–2009 (Fig. 3c), the number of 2G users is stable, but that it clearly begins to decrease constantly from 2010. Meanwhile, 3G technology has been growing rapidly since 2005, and 3G users have outnumbered 2G users since 2010.

Figure 3d portrays the total number of 4G users. The 4G usage in Singapore is making a leap and tends to be growing. Figure 3e shows that the number of 3G users in Singapore continues in its downward trend, while the number of 4G users is on an increasing trend.

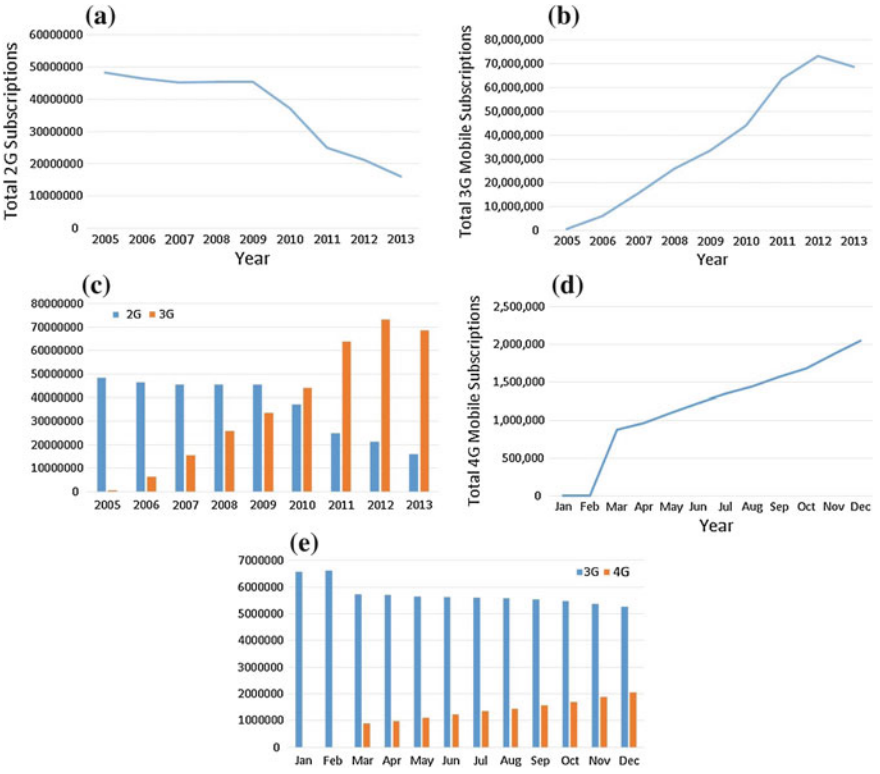


Fig. 3 Variation in volatility and auto-correlation plots [5]. **a** The total number of 2G users. **b** The total number of 3G users. **c** The comparison between the numbers of 2G and 3G users. **d** The total number of 4G users. **e** The comparison between the numbers of 2G and 3G users

6 Literature Review

Jarno Salonen et al. [6] conducted a research entitled "Exploring the Possibilities for Mobile Insurance Services", in which the development from web-based insurance services to mobile insurance services is discussed. The study finds that the development of either web-based insurance services or mobile insurance services was not an easy task because the insurance business is complicated and difficult to understand for most people. Therefore, people do not prefer buying insurance through the electronic channels and would rather depend on insurance agents or brokers. Moreover, the 3G (Third Generation Wireless Broadband Mobile Communications) and AIPN (All-IP Mobile Network) technologies, which offer high-speed Internet connections, will become one of the communication channels and support mobile services in the future. However, only a few mobile insurance applications are available at the moment. In the USA (2004), mobile applications represented only 10 % of all the distribution channels in the life and health insurance segments, and the proportion was only 3 % in the other non-life segments. The number of mobile application users is, however, increasing. One example of successful mobile insurance applications is "Pay As You Drive," developed by Norwich Group, the largest insurance group in the United Kingdom. Using this application, the company collects monthly motor insurance premiums calculated from the car usage. A telematics system installed in the car sends a satellite signal to identify the car's location whenever the car is in use. The system keeps track of the cars journey and reports it to the insurer in real-time. In case of a car theft, the telematics tool can also identify where the car is being taken to.

Frost and Sullivan [2] undertook a study entitled "Using Mobile Solutions to Improve Insurance Sector Performance Control costs, manage risk, and create new revenue" which proposes how smartphones and tablets should be adapted in insurance operations. The study finds that mobile technology will become a significant growth engine for insurance business in the future because it saves costs, reduces risks, and improves efficiency in insurance operations, helping insurance companies to attract new customers and maintain their existing customer base. Utilizing mobile technolgy along with data network, mobile devices, and mobile applications will make insurance products and services more interesting. Insurers will be able to respond faster to customers, thus offering greater value for money. Moreover, the use of wireless Internet and mobile broadband has substantially increased due to the inexpensive charges and user friendliness. Insurance companies know that it is important to use wireless and mobile broadband to improve and develop products, manage risks, and control expenses. The study also found that mobile technology could streamline services in two areas. First, in-vehicle telemetics systems can be developed. The M2M communication systems help collect and report driving data to insurers. The data received are highly accurate and can be transferred even when the car owners are far from the insurance companies locations. Telematics is cost-efficient and allows insurers to use the provided information to impose premiums required for the next payment more properly and accurately. Second, networks on

demand is possible through mobile technology. Field insurance agents often have to be at accident spots. Insurers can use wireless and mobile solutions to reach their agents anytime, and can support their customers immediately after the occurrence of any accidents, using the technology.

Hiwarkar and Khot [4] carried out a research on "E-Insurance: Analysis of the Collision and Allegation of E-Commerce on the Insurance and Banking" whose aim was to propose how the Internet should be utilized in the insurance and banking sector. The study finds that many insurance companies have introduced their products online, such as motor insurance products. However, a lot of customers are hesitant to accept such a policy to provide insurance online, which makes the development of online insurance services slower than that of other industries. There are also fewer life insurance transactions than general insurance transactions, as life insurance is naturally more complicated. Nevertheless, insurance companies are still developing online services because these services (1) are more cost-saving to manage; (2) help reduce brokerage fees, which keeps the premiums low and attracts more customers; (3) deliver a greater information management system; (4) shorten the insurance purchasing and claim handling processes; (5) allow customers to access insurance product information anytime and anywhere; and (6) save expenses on insurance office rentals.

7 Research Methodology

The study explores the impacts of mobile broadband on Thailands and Singapore's insurance businesses. To study the impacts on Thailands non-life sector, the direct gross non-life premiums the secondary data compiled by the Office of the Insurance Commission (OIC) are taken into account (the direct gross non-life premiums in Singapore, the secondary data compiled by the Monetary Authority of Singapore [MAS]) [8]. Piecewise regression is used to evaluate the structural changes in the value of direct gross non-life premiums after the 3G service is available in Thailand and Singapore, in this study.

The paper uses piecewise regression to evaluate the structural changes in the value of Thailands direct gross non-life premiums after the 3G service is available, as well as the value of Singapores gross premiums following the introduction of 3G and 4G networks.

Piecewise regression uses time-series data to find the structural changes that may divide a time horizon into two periods. Once the turning point is identified, that is, at the beginning of the 3G and 4G services in Thailand and Singapore, the regression measures the slopes of the time trend before and after that point. When the slopes are not significantly different, the effects from the turning point might be too small, or it may be too early to detect the impacts. On the other hand, when the slopes are significantly different, it means that structural changes might have occurred at the turning point. The differences between the projected time trend and the estimated result obtained from the piecewise regression at the point after the turning point measure the impacts of the 3G or 4G technologies.

The setting of the piecewise regression is as follows:

$$Y_i = \alpha + \beta_1 T_i + \beta_2 (T_i - T^*) D_i + u_i \tag{1}$$

where Y_i is the value of direct premiums in the period T_i, when $i = 0, 1, 2, \ldots, T$;
α, β are the parameters to be estimated; T_i is the time when $i = 0, 1, 2, \ldots, T$;
T^* is the time at the turning point where 3G and 4G services are launched;
D_i is the dummy variable whose value is 1 when T_i is after $T^* (T_i > T^*)$ and 0 when T_i is before $T^* (T_i < T^*)$; and
u_i is the error term.

Before the 3G or 4G availability, the model specification is defined by α and β_1 only; this is because D_i is equal to 0. Therefore, the Y-intercept is α and the slope is β_1:

$$Y_i = \alpha + \beta_1 T_i + u_i. \tag{2}$$

After 3G or 4G becomes available, Di becomes equal to 1. The Y-intercept is $\alpha - \beta 2T^*$, and the slope is $\beta_1 + \beta_2$. As a result, the specification of the model can be written as follows:

$$Y_i = (\alpha - \beta_2 T^*) + (\beta_1 + \beta_2) T_i + u_i. \tag{3}$$

When plotting a graph from both the models, the intersection will be at the turning point T^* which is the time at which the 3G or 4G networks are available in the respective country. Therefore, the impacts of 3G or 4G can be measured from the differences between Y_i, obtained from the time trend and the one estimated by the piecewise regression.

This paper measures the impacts of 3G technology using Thailand and Singapore as the case studies, while the impacts of 4G is evaluated based on the case of Singapore. When comparing the impacts in Thailand and Singapore, the results will be the ratio of the insurance sales volumes that occur after the turning point estimated by the piecewise regression to the volumes forecast by the time trend.

8 Models

This study measures the impacts of 3G technology on Thailand's and Singapore's direct gross non-life insurance premiums, and the impacts of 4G technology on the direct premiums in Singapores non-life insurance sector through an application of the following models:

Model 1: The impacts of a 3G service on Thailand's and Singapore's direct premiums of non-life insurance business:

$$\text{Thailand: } V_i^{TH} = \alpha_1 + \beta_{11} T_i + \beta_{21} (T_i - T_{3G}^{TH}) D_{3G,i}^{TH} + \beta_{31} C_1 + \beta_{41} C_2 + u_i \tag{4}$$

$$\text{Singapore: } V_i^{SG} = \alpha_2 + \beta_{12} T_i + \beta_{22} (T_i - T_{3G}^{SG}) D_{3G,i}^{SG} + \beta_{32} C_1 + \beta_{42} C_2 + \gamma_i \tag{5}$$

Model 2: The impacts of a 4G service on Thailand's and Singapore's direct premiums of non-life insurance business:

$$V_i^{SG} = \alpha_3 + \beta_{13} T_i + \beta_{23} (T_i - T_{4G}^{SG}) D_{4G,i}^{SG} + \beta_{33} C_1 + \beta_{43} C_2 + \varepsilon_i, \tag{6}$$

where V_i^{TH} is the decycled and seasonalized data of the value of Thailand's direct non-life insurance premiums at constant prices in 2003,

V_i^{SG} is the decycled data of the value of Singapore's direct non-life insurance premiums at constant prices in 1997,

α, β are the parameters to be estimated,

T_i is the time on a monthly basis for Thailand and an annual basis for Singapore,

T_{3G}^{TH} is the time a 3G service is launched in Thailand,

T_{3G}^{SG} is the time a 3G service is launched in Singapore,

T_{4G}^{TH} is the time a 4G service is launched in Singapore,

$D_{3G,i}^{TH}$ is the dummy variable indicating the 3G availability in Thailand after the bidding for the 2,100 MHz spectrum,

$D_{3G,i}^{SG}$ is the dummy variable indicating the 3G availability in Singapore,

$D_{4G,i}^{SG}$ is the dummy variable indicating the 4G availability in Singapore,

C_1 is the dummy variable indicating the subprime crisis in 2008,

C_2 is the dummy variable indicating the tsunami incident in 2011,

u_i, γ_i, ε_i are error terms.

The values of the direct non-life insurance premiums were constructed at constant prices in 2003 for Thailand and at constant prices in 1997 for Singapore to de-cycle the time series. To calculate the cyclical index, the average prices for each year were obtained by dividing the value of the annual direct gross premiums over the number of policies issued per year. Doing this makes the prices turn into indices starting at 100 in 2003 for Thailand and in 1997 for Singapore. Finally, the value at the constant prices is calculated by dividing the sales values on a monthly basis over the price indices in the related years.

The deseasonalization process begins with creating a seasonal index that compiles the data for every year for the period 2003–2013. The index is a ratio between the average monthly values of purchases to the mean of those averages. Note that the data for Singapore is based on an annual basis only, so it does not have to be deseasonalized.

The explanatory variables of the models are obtained from the following information. The 3G service was launched in May 2013 in Thailand. Singapore's 3G service was available from April 2001, while 4G was first launched in July 2012. The impacts of each crisis were set to last for 36 months, starting with the subprime crisis in September 2008, and then the tsunami crisis in March 2011. Rehabilitation from each of the crises is estimated to take at least three years.

9 Results

9.1 Case Study of Thailand

The study uses the direct premiums of non-life insurance business, compiled by the Office of the Insurance Commission (OIC) [9]. The methodology adopted is the piecewise regression which measures whether the provision of 3G mobile technology in Thailand draws positive, negative, or no impacts on the Thai general insurance business.

The results show that the sales of Thailand's general insurance tended to increase over the past decade, considering the positive trend (T_i) presented in Table 2. However, the positive value has no significance. The coefficient of $(T_i - T_{3G}^{TH})D_{3G,i}^{TH}$ is also negative. If the coefficient is significant, it means the 3G availability draws negative impacts on the insurance business, that is, the sales decrease. The table below, however, shows that such a conclusion is invalid because the coefficient of $(T_i - T_{3G}^{TH})D_{3G,i}^{TH}$ has no significance, meaning the application of 3G technology in the non-life insurance sector does not affect the business. Moreover, the table shows that what actually negatively affect Thailand's general insurance business are the subprime and the tsunami crises, although the impacts are not significant.

9.2 Case Study of Singapore

The study uses the direct premiums of non-life insurance business compiled by the Monetary Authority of Singapore (MAS). The methodology adopted is the piecewise regression which measures whether the provision of 3G mobile technology in Singapore draws positive, negative, or no impacts on Singapore's non-life insurance business.

The results show that the sales of Singapore's non-life insurance tended to increase over the past decade, considering the positive trend (Ti) presented in Table 3. However, the positive value has no significance. The coefficient of $(T_i - T_{3G}^{SG})D_{3G,i}^{SG}$ is also positive. If the coefficient is significant, it means the 3G availability draws positive

Table 2 Impacts of 3G service on Thailand's non-life insurance business

Dependent variable: V_i^{TH}					*Unit* Baht	
Method: OLS	Number of observations: 132				R-square: 0.0096	
Variables	Coeff.	S.D.	t	Prob.	95 % Confident interval	
T_i	2,058.696	2,692.329	0.76	0.446	−3,268.937	7,386.329
$(T_i - T_{3G}^{TH})D_{3G,i}^{TH}$	−59,720.22	67,609.42	−0.88	0.379	−193,507.1	74,066.62
C_1	−188,253.8	200,758.1	−0.94	0.350	−585,517.8	209,010.3
C_2	−148,901.1	248,913.8	−0.60	0.551	−641,456.6	343,654.4
Constant	5,853,400	145,066.6	40.35*	0.000	5,566,340	6,140,461

Source The authors own calculation

Table 3 Impacts of 3G service on Singapore's non-life insurance business

Dependent variable: V_i^{SG}					*Unit* USD	
Method: OLS	Number of observations: 16				R-square: 0.9743	
Variables	Coeff.	S.D.	t	Prob.	95 % Confident interval	
T_i	56.44977	46.09987	1.22	0.246	−45.01535	157.9149
$(T_i - T_{3G}^{SG})D_{3G,i}^{SG}$	91.85159	53.15268	1.73	0.112	−25.13668	208.8399
C_1	−157.123	86.8474	−1.81	0.098	−348.2728	34.02689
C_2	109.4349	128.8912	0.85	0.414	−174.2527	393.1226
Constant	1,484.45	147.1234	10.09*	0.000	1,160.634	1,808.267

Source The authors' own calculation
* 1 % Significant level

impacts on the insurance business, that is, the sales increase. Table 3, however, shows that the coefficient of $(T_i - T_{3G}^{SG})D_{3G,i}^{SG}$ has no significance, meaning the application of 3G technology in the non-life insurance sector does not affect the business. Moreover, it is found that the subprime crisis draws negative impacts, while the tsunami crisis's impacts are positive, but the coefficients of both the variables are found to be not significant, which means that both the crises do not draw any impacts on Singapore's non-life insurance business.

Furthermore, this paper studies the impacts of 4G, as well, on Singapore's non-life insurance sector. The study results, as displayed in Table 4, show that the people in Singapore tend to buy more non-life insurance products, given that Ti is positive and significant. Also, the coefficient of $(T_i - T_{4G}^{SG})D_{4G,i}^{SG}$ is positive, but not significant, so it is not conclusive whether the 4G availability does increase the sales of non-life insurance products. The subprime crisis affects the business negatively, while the tsunami draws positive impacts. However, the coefficients of both the incidents are not significant, meaning the crises have no impacts on Singapore's non-life insurance sector.

Table 4 Impacts of 4G service on Singapore's non-life insurance business

Dependent variable: V_i^{SG}					*Unit* USD	
Method: OLS	Number of observations: 16				R-square: 0.9743	
Variables	Coeff.	S.D.	t	Prob.	95 % Confident interval	
T_i	134.5205	9.789643	13.74*	0.000	112.9737	156.0674
$(T_i - T_{4G}^{SG})D_{4G,i}^{SG}$	68.57934	205.0733	0.33	0.744	−382.784	519.9427
C_1	−180.9149	96.17553	−1.88	0.087	−392.5959	30.76597
C_2	72.34941	333.7366	0.22	0.832	−662.2	806.8988
Constant	1,264.864	82.01233	15.42*	0.000	1,084.356	1,445.372

Source The authors' own calculation
* 1 % Significant level

10 Conclusion

This research studies the impacts of 3G technology on Thailand's and Singapore's non-life insurance business, using the direct premiums data from Thailand's Office of the Insurance Commission (OIC) and the Monetary Authority of Singapore (MAS). The piecewise regression method is applied in the study. The results show that 3G technology has no impacts on Thailand's non-life insurance sector, possibly because it is uncommon for the Thai people to buy insurance products through wireless electronic devices. Most of the people are unaware of the benefits of general insurance; to make matters worse, there are not enough personnel providing advices about purchasing general insurance through electronics channels. Therefore, the provision of the technology does not increase people's demand for general insurance products. Likewise, the 3G availability in Singapore does not impact the local non-life insurance business. Singapore has adopted a 4G network since 2012, so the study also explores the impacts of such technology. It finds that 4G technology does not affect Singapore's non-life insurance business, probably due to the complicated nature of general insurance. Buyers generally need advice before making a decision to buy, while many types of insurance also require long-term premium payments. Because of these reasons, people do not prefer buying general insurance through wireless devices. The availability of 3G and 4G services does not significantly increase the number of general insurance buyers through the electronic channels. It could be projected that, for Thailand, the arrival of 4G will not affect the non-life insurance business.

It is clear that the insurance industry has not caught up with the "mobile integration" and seems to be missing the tremendous opportunity to enhance its growth by applying the mobile broadband technology. The study points it out to the insurance companies that they should find some strategies to integrate the mobile broadband technology into their businesses; otherwise, they are bound to be left behind in the digital and connected world. Among the strategies, the priority should be placed on the provision of product information, online purchase, and real-time notification with location-based tracking system of any accidents or incidents such that the insurance companies can rush their rapid assistance to the customers.

References

1. Bertrand, D.: Mobile insurance: are you well positioned for this emerging channel? Company Report, Capgemini Consulting. 2 November 2012, http://www.capgemini-consulting.com/resource-file-access/resource/pdf/Mobile_Insurance__Are_You_Well_Positioned_for_this_Emerging_Channel_.pdf. Accessed 2 Apr 2014
2. Frost and Sullivan: Using mobile solutions to improve insurance sector performance control costs, manage risk, and create new revenue. Sprint, Mountain View, Canada (2011)
3. General Insurance Association of Singapore. Distribution channel. General insurance association of Singapore (2014), http://www.gia.org.sg/public_market_structure_distribution.php. Accessed 15 Mar 2014

4. Hiwarkar, T., Khot, P.G.: E-insurance: analysis of the collision and allegation of E-commerce on the insurance and banking. J. Bus. Manag. Soc. Sci. Res. **2**(6), 1–5 (2013)
5. Infocomm Development Authority of Singapore. Statistics on Telecom Services. Infocomm Development Authority of Singapore (2014), http://www.ida.gov.sg/Infocomm-Landscape/Facts-and-Figures/Telecommunications/Statistics-on-Telecom-Services. Accessed 29 Mar 2014
6. Jarno, S., Ahonen, A., Koskinen, H.: Exploring the possibilities for mobile insurance services. Frontier of E-business Research (2006)
7. Mitchell, C., Wannemacher, P., Ensor, B., Tincher, C.: Seizing insurance's mobile opportunity customer adoption and investment grow as US insurers play catch-up. Forrester Research. 22 October 2009
8. Monetary Authority of Singapore. Annual statistics. Monetary Authority of Singapore (2014), http://www.mas.gov.sg/Statistics/Insurance-Statistics/Annual-Statistics.aspx. Accessed 1 Apr 2014
9. Office of the Insurance Commission. Insurance statistics. Office of the Insurance Commission (2014), http://www.oic.or.th/en/statistics/index2.php. Accessed 1 Apr 2014
10. Vroomen, B., Donkers, B., Verhoef, P.C., Franses, P.H.: Selecting profitable customers for complex services on the internet. J. Serv. Res. **8**(1), 3747 (2005)

Using Conditional Copula to Estimate Value-at-Risk in Vietnam's Foreign Exchange Market

Vu-Linh Nguyen and Van-Nam Huynh

Abstract In this paper, we briefly review the basics of copula theory and the problem of estimating Value-at-Risk (VaR) of portfolio composed by several assets. We present two VaR estimation models in which each return series is assumed to follow AR(1)-GARCH(1, 1) model and the innovations are simultaneously generated using Gaussian copula and Student t copula. The presented models are applied to estimate VaR of a portfolio consisting of six currencies to VND. The results are compared with results from two VaR estimation models using AR(1)-GARCH(1, 1) model and the innovations are separately generated using univariate standard normal and Student t distribution.

1 Introduction

The theory of copula is a very powerful tool for modeling joint distributions because it does not require the assumption of joint normality which is rarely adequate in application [3, 15]. Applications based on copula theory center around Sklar theorem which allows to decompose any N-dimensional joint distribution into its N marginal distributions and a copula function which describes the dependence structure between the variables [3, 11, 15, 17]. Furthermore, the converse of Sklar theorem can be used to learn the dependence structure given prior information about distribution and copula.

During the last years, copula based models have been increasingly applied in finance and economics. Those models have shown advantages comparing with the traditional models, specially where dependency is non-linear and the involved random variables follow different univariate distributions. The book of Nelsen [12] provided a very good introduction about copulas including the basic of copula theory

V.-L. Nguyen (✉) · V.-N. Huynh
School of Knowledge Science, Japan Advanced Institute of Science and Technology,
1-1 Asahidai, Nomi, Ishikawa, Japan
e-mail: vulinh@jaist.ac.jp

V.-N. Huynh
e-mail: huynh@jaist.ac.jp

V.-N. Huynh et al. (eds.), *Econometrics of Risk*, Studies in Computational Intelligence 583, DOI 10.1007/978-3-319-13449-9_33

as well as advantages of using copulas to construct the joint distribution and learn the dependence [11]. Also, Cherubini et al. [3] provided a comprehensive guide for applying copulas in financial problems for example, asset pricing, risk management and credit risk analysis [3]. Bouye et al. [1] provided a statistical inference framework of copulas in the estimation problem [1]. Embrechts et al. [5] highlighted the pitfalls when finding the multivariate models and suggested simulation algorithms to avoid those problems [5]. Georges [8] used the normal copula to model options time of exercise and for derivative pricing [8]. Meneguzzo and Vecchiato [10] used copulas to model the risk of credit derivatives [10]. Cherubini and Luciano [2] proposed a VaR estimation model using the Archimedean copula family and the historical empirical marginal distribution [2]. Fortin and Kuzmics [7] used convex linear combinations of copulas to estimate the VaR of a portfolio consisting of the FSTE and DAX stock indices [7]. Embrechts et al. [6] used copula to learn the optimal bounds for risk measures of functions of dependent risks [6]. Rockinger and Jondeau [16] used the Plackett copula with GARCH process with innovations modeled by the Student-t asymmetrical distribution to learn the change of dependence through time of daily return of stock market indices [16]. Patton and Andrew [14] used conditional copula based models to explore the dependence structure of exchange rates [14]. Palaro and Hotta [15] used the SJC, Plackett and Student-t conditional copula to modeled the innovations of GARCH process and used simulation methods to estimate VaR of portfolio composed by Nasdaq and S&P500 stock indices [15].

The market of foreign exchange (forex) in Vietnam has remained relatively poorly developed despite more than two decades of general reform throughout the economy. Also, research on Vietnam's foreign exchange market is rather limited. In their studies, Nguyen et al. [12, 13] have pointed out that Vietnam's foreign exchange (forex) market has remained far less active and sophisticated than forex markets in many other countries.

In this paper, we apply conditional copula based models to estimate VaR of a portfolio composed by six currencies to VND namely VND/AUD, VND/EUR, VND/GBP, VND/JPY, VND/USD and VND/CNY. The paper is organized as follows. Section 2 summarizes the market risk problem and defines VaR measure. The basic of copulas is presented in Sect. 3. Section 4 presents two conditional copula based models for estimating VaR in which each return series is modeled using AR(1)-GARCH(1, 1) model and innovations are simultaneously modeled using the Gaussian copula and Student t copula. In Sect. 5, the presented models are applied to estimate VaR of a portfolio composed by six daily rate return and log return series of currencies. The results are compared with those obtained using simulation AR(1)-GARCH(1, 1) models where innovations are separately modeled using univariate standard normal and student t distributions. Finally, the conclusion is given in Sect. 6.

2 Market Risk Problem

Let us consider the problem of measuring the risk of holding an portfolio consists of N assets with returns at Tth day, denoted as $x_{n,T}$, given the historical data

$\{x_{n,t}|t = 1, 2, \ldots, T-1\}$, for $n = 1, 2, \ldots, N$ [15]. The portfolio return at tth day, denoted as x_t, is approximately equal to

$$x_t = \omega_1 x_{1,t} + \omega_2 x_{2,t} + \cdots + \omega_N x_{N,t}, \tag{1}$$

where ω_n is the portfolio weigh of asset n and $\sum_{n=1}^{N} \omega_{n,t} = 1$, for $t = 1, 2, \ldots, T$, $n = 1, 2, \ldots, N$.

In 1994, the American bank JP Morgan published a risk control method knows as Riskmetrics, based mainly on a parameter named *Value-at-Risk* (VaR). For a given time horizon T and confidence level p, the VaR is defined as the loss in market value over the time horizon T that is exceeded with probability $1 - p$. More precisely, VaR of a portfolio can be defined as follows.

Definition 1 Let $H_T(x_T \mid \Im)$ be the conditional distribution function of the returns of portfolio consisting of $x_1, x_2, \ldots, x_N$ at time T with conditional set $\Im$.

$$\Im = \{X_{n,t}|n = 1, 2, \ldots, N, t = 1, 2, \ldots, T-1\} \tag{2}$$

$\Im$ represents the past information from day 1 to day $T - 1$. Then the VaR of the portfolio at time T, with confidence level p, where $p \in (0, 1)$ is defined by

$$VaR_T(p) = \inf\{s =: H_T(s \mid \Im) \geq 1 - p\}. \tag{3}$$

Figure 1 illustrates VaR and p.

In this paper, VaR is approximated using simulation models. The exchange rate series are assumed to fit the AR(1)-GARCH(1, 1) models with standard normal and student t innovations. The historical data of innovations is used to fit the multivariate copula which then used to generated values of innovations simultaneously. The generated values of portfolio distribution obtained by substituting the values of innovation into AR(1)-GARCH(1, 1) models. Finally, VaR is approximated as the corresponding element of simulation series after increasingly ordering the simulated values of portfolio distribution.

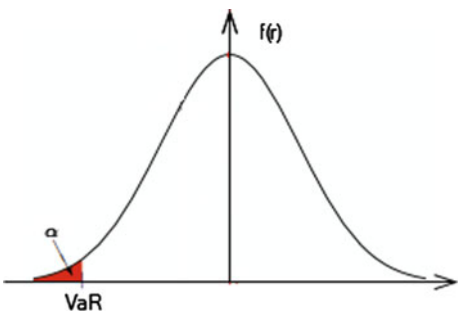


Fig. 1 The Value-at-Risk VaR and level $\alpha = 1 - p$

3 Copula

The concept of copulas was introduced by Sklar (1959), and has been recognized as a powerful tool for modeling dependence between random variables. Almost applications based on copula theory centralize around the Sklar theorem which ensures the relation between a N-dimensional distribution and a corresponding copula [3, 11].

A copula is a multivariate probability distribution for which the marginal probability distribution of each variable is uniform.

Definition 2 A N-dimensional copula (N-copula) is a function C, whose domain is $[0, 1]^N$ and whose range is $[0, 1]$ with the following properties:

1, For every $u \in [0, 1]^N$, $C(u) = 0$ if at least one coordinate of u is 0 and if all coordinates of u are 1 except u_n, then $C(u) = u_n$, $n = 1, 2, \ldots, N$.
2, For every $a, b \in [0, 1]^n$ such that $a \leq b$, $V_C([a, b]) \geq 0$.

Sklar theorem is perhaps the most important result regarding copulas [17]. It ensures the relation between a N-dimensional distribution function and a corresponding copula and is used in essentially all applications of copula.

Theorem 1 *Let H be a N-dimensional distribution function with 1 dimensional margins $F_1, F_2, \ldots, F_N$. Then there exists a N-copulas C such that for all x in $\mathbb{R}^N$,*

$$H(x_1, x_2, \ldots, x_N) = C(F_1(x_1), F_2(x_2), \ldots, F_N(x_N)). \tag{4}$$

If $F_1, F_2, \ldots, F_N$ are all continuous, then C is unique; Otherwise C is uniquely determined on $Ran F_1 \times Ran F_2 \times \cdots \times Ran F_N$.

Conversely, if C is a N-copula and $F_1, F_2, \ldots, F_N$ are distribution functions, then the function H defined by (4) is a N-distribution function with margins.

The following corollary is often known as the converse of Sklar theorem. We can use this corollary to find copula when the margins and joint distributions are given.

Corollary 1 *Let $H, C, F_1, \ldots, F_N$ be as in Theorem 1 and $F_1^{(-1)}, \ldots, F_N^{(-1)}$ be quasi-inverses of $F_1, \ldots, F_N$, respectively. Then, for any u in $[0, 1]^N$*

$$C(u_1, u_2, \ldots, u_N) = H(F_1^{(-1)}(u_1), F_2^{(-1)}(u_2), \ldots, F_N^{(-1)}(u_N)). \tag{5}$$

By applying Sklar theorem and exploiting the relation between the distribution and the density function, we can easily derive the multivariate copula density

$$c(F_1(x_1), F_2(x_2), \ldots, F_N(x_N))$$

associated with a copula function $C(F_1(x_1), F_2(x_2), \ldots, F_N(x_N))$:

$$h(x_1, x_2, \ldots, x_N) = \frac{\partial^N [C(F_1(x_1), F_2(x_2), \ldots, F_N(x_N))]}{\partial F_1(x_1) \cdots \partial F_N(x_N)} \prod_{n=1}^{N} f_n(x_n)$$
$$= c(F_1(x_1), F_2(x_2), \ldots, F_N(x_N)) \prod_{n=1}^{N} f_n(x_n), \quad (6)$$

where we define

$$c(F_1(x_1), F_2(x_2), \ldots, F_N(x_N)) = \frac{f(x_1, x_2, \ldots, x_N)}{\prod_{n=1}^{N} f_n(x_n)}. \quad (7)$$

The results of Sklar theorem and its corollary can be extended in conditional case as follows

$$H(x_1, x_2, \ldots, x_N \mid \Im) = C(F_1(x_1 \mid \Im), F_2(x_2 \mid \Im), \ldots F_N(x_N \mid \Im) \mid \Im), \quad (8)$$

and

$$C(u_1, u_2, \ldots, u_N \mid \Im) = H(F_1^{-1}(u_1 \mid \Im), F_2^{-1}(u_2 \mid \Im), \ldots, F_N^{-1}(u_N \mid \Im) \mid \Im)), \quad (9)$$

where $\Im$ is given the conditional set.

The Gaussian copula is a distribution over the unit cube $[0, 1]^N$. It is constructed from a multivariate normal distribution over $\mathbb{R}^N$ by using the probability integral transform. Formally, Gaussian copula is defined as follows

Definition 3 Let R be a symmetric, positive definite matrix with $\text{diag}(R) = 1$ and let Φ_R the standardized multivariate normal distribution with correlation matrix R. Then the multivariate Gaussian copula is defined by

$$C^{Gauss}(u_1, u_2, \ldots, u_N; R) = \Phi_R(\Phi^{-1}(u_1), \Phi^{-1}(u_2), \ldots, \Phi^{-1}(u_N)), \quad (10)$$

where Φ_R^{-1} denotes the inverse of the standard univariate normal distribution function Φ_R.

The associated multinormal copula density is

$$c^{Gauss}(\Phi(x_1), \Phi(x_2), \ldots, \Phi(x_N); R) = \frac{f^{Gauss}(x_1, x_2, \ldots, x_N)}{\prod_{n=1}^{N} f_n^{Gauss}(x_n)}$$
$$= \frac{\frac{1}{(2\pi)^{\frac{N}{2}} |R|^{\frac{1}{2}}} exp(-\frac{1}{2} x' R^{-1} x)}{\prod_{n=1}^{N} \frac{1}{\sqrt{2\pi}} exp(-\frac{1}{2} x_n^2)}, \quad (11)$$

and hence, fixing $u_n = \Phi(x_n)$, and denote

$$\zeta = (\Phi^{-1}(u_1), \Phi^{-1}(u_2), \ldots, \Phi^{-1}(u_N)^{'})$$

the vector of the Gaussian univariate distribution functions, we have

$$c(u_1, u_2, \ldots, u_N; R) = \frac{1}{|R|^{\frac{1}{2}}} exp[-\frac{1}{2}\zeta^{'}(R^{-1} - I)\zeta]. \tag{12}$$

The student t copula is defined as follows

Definition 4 Let R be a symmetric, positive definite matrix with $\text{diag}(R) = 1$ and let $T_{R,\nu}$ the standardized multivariate Student t distribution with correlation matrix R and ν degree of freedom. Then the multivariate Student t copula is defined as follows

$$C(u_1, u_2, \ldots, u_N; R, \nu) = T_{R,\nu}(t_\nu^{-1}(u_1), t_\nu^{-1}(u_2), \ldots, t_\nu^{-1}(u_N)), \tag{13}$$

where $t_\nu^{-1}(u_n)$ denotes the inverse of the Student t cumulative distribution function.

The associated Student t copula density is:

$$\begin{aligned} c(u_1, u_2, \ldots, u_N; R, \nu) &= \frac{f^{Student}(x_1, x_2, \ldots, x_N)}{\prod_{n=1}^{N} f_n^{Student}(x_n)} \\ &= |R|^{-\frac{1}{2}} \frac{\Gamma(\frac{\nu+N}{2})}{\Gamma(\frac{\nu}{2})} \left[\frac{\Gamma(\frac{\nu}{2})}{\Gamma(\frac{\nu+1}{2})}\right]^N \frac{(1 + \frac{\zeta^{'} R^{-1} \zeta}{\nu})^{-\frac{\nu+N}{2}}}{\prod_{n=1}^{N}(1 + \frac{\zeta_n^2}{\nu})^{-\frac{\nu+1}{2}}}, \end{aligned} \tag{14}$$

where $\zeta = (t_\nu^{-1}(u_1), t_\nu^{-1}(u_2), \ldots, t_\nu^{-1}(u_N))^{'}$.

4 Using Conditional Copula to Estimate VaR

This section presents two simulation models using conditional copulas to estimate VaR of a portfolio consists of several assets, namely AR(1)-GARCH(1, 1) + Gaussian copula and AR(1)-GARCH(1, 1) + Student t copula. In those models, each return series is assumed to follow AR(1)-GARCH(1, 1) models and the innovations are simultaneously generated using copulas.

4.1 Modeling the Marginal Distributions

Returns series has been successfully modeled by ARMA-GARCH models [4, 15]. In this paper, the AR(1)-GARCH(1, 1) models are used to model the margins as follows

$$\begin{aligned} x_{n,t} &= \mu_n + \phi_n x_{n,t-1} + \varepsilon_{n,t}; \\ \varepsilon_{n,t} &= \sigma_{n,t}\eta_{n,t}; \\ \sigma_{n,t}^2 &= \alpha_n + \beta_n \varepsilon_{n,t-1}^2 + \gamma_n \sigma_{n,t-1}^2; \end{aligned} \tag{15}$$

where $\{\eta_{n,t}\}$ is white noise process, α_n, β_n, γ_n satisfy the condition of GARCH model: $\beta_n + \gamma_n < 1$, for $n = 1, 2, \ldots, N$ and $t = 1, 2, \ldots, T$. The conditional distribution of the standardized innovations

$$\eta_{n,t} = \frac{\varepsilon_{n,t}}{\sigma_{n,t}} | \Im_{n,T}, n = 1, 2, \ldots, N,$$

was modeled by white noises and denoted by $F_{n,t}$ in general case (the marginal distributions). We consider the case that $\eta_{n,t}$ are standard normal distributions and student t distributions with the same degree of freedom, $n = 1, 2, \ldots, N$.

The joint distribution of innovation vector $\eta_t = (\eta_{1,t}, \eta_{2,t}, \ldots, \eta_{N,t})$ is model by conditional copula.

Let $u_{n,t} = F_{n,t}(\eta_{n,t}|\Im)$, $F_{1,t}$, $F_{2,t}$, ... and $F_{N,t}$ are marginal distributions conditioned to $\Im$, the information available up to time $T-1$. If the models were correctly specified then series $\{u_{n,t}|t = 1, 2, \ldots, T-1\}$ will be standard uniform series.

4.2 Modeling the Copula

We assume that $(\eta_{1,T}, \eta_{2,T}, \ldots, \eta_{N,T})$ has multivariate distribution function

$$H_T(\eta_{1,T}, \ldots, \eta_{n,T}; \theta_{1,T}, \theta_{2,T} \mid \Im) \tag{16}$$

and continuous univariate marginal distribution functions $F_{n,T}(\eta_{n,T}; \theta_{n,T} \mid \Im)$ where $\Im = \{\eta_{n,t}|n = 1, 2, \ldots, N, t = 1, 2, \ldots, T-1\}$.

Since the marginal distributions are continuous, the conditional copula C_T is uniquely defined according to Sklar theory. Furthermore, we have

$$\begin{aligned} &C_T(F_{1,T}(\eta_{1,T}; \theta_{1,T} \mid \Im), \ldots, F_{N,T}(\eta_{N,T}; \theta_{N,T} \mid \Im); \theta_{2,T} \mid \Im) \\ &\quad = H_T(\eta_{1,T}, \ldots, \eta_{N,T}; \theta_{1,T}, \theta_{2,T} \mid \Im), \end{aligned} \tag{17}$$

where $\theta_{1,T}$ is the margins' parameters and $\theta_{2,T}$ is copula's parameters of copula function C_T.

The parameters $\theta_{1,T}$, $\theta_{2,T}$ are estimated by using IFM (inference for the margins) method as follows

1. Firstly, we estimate the margin's parameters $\widehat{\theta}_{1,T}$ by performing the estimation of the univariate marginal distributions

$$\widehat{\theta}_{1,T} = \underset{\theta_{1,T}}{\operatorname{argmax}} \sum_{t=1}^{T-1} \sum_{n=1}^{N} \ln f_{n,T}(\eta_{n,t}; \theta_{1,T}). \tag{18}$$

2. Secondly, given $\widehat{\theta}_{1,T}$, we perform the estimation of the copula parameter $\widehat{\theta}_{2,T}$ as follows

$$\widehat{\theta}_{2,T} = \underset{\theta_{2,T}}{\operatorname{argmax}} \sum_{t=1}^{T-1} \ln c_T(F_{1,T}(\eta_{1,t}; \widehat{\theta}_{1,T}), \ldots, F_{n,T}(\eta_{n,t}; \widehat{\theta}_{1,T}); \theta_{2,T}). \quad (19)$$

If the marginal distributions $F_{n,T}$ are standard normal distributions then C_T is a multivariate Gaussian copula with correlation matrix $\theta_{2,T} = R_T$. And if the marginal distributions $F_{n,T}$ are Student's t distributions with same degree of freedom $\theta_{1,T} = \nu_T$, then C_T is a Student's t copula with parameter $\theta_{2,tT} = R_T$. In this case, N marginal distributions are assumed to have the same degree of freedom.

4.3 Monte Carlo Simulation

We use Gaussian and student t copula to simulate K vectors

$$\eta_{T,k} = (\eta_{1,T,k}, \eta_{2,T,k}, \ldots, \eta_{N,T,k}), \quad (20)$$

for $k = 1, 2, \ldots, K$.

In case of multivariate Gaussian copula, the Monte Carlo simulation can be processed as follows:

1. Find the Cholesky decomposition A of the linear correlation matrix R.
2. Simulate N i.i.d. $z = (z_1, z_2, \ldots, z_N)'$ from $N(0, 1)$
3. Set $\eta'_{T,k} = Az$

Similarly, we have the Monte Carlo simulation for multivariate student t copula

1. Find the Cholesky decomposition A of the linear correlation matrix R.
2. Simulate N i.i.d. $z = (z_1, z_2, \ldots, z_N)'$ from $N(0, 1)$
3. Simulate a random variate s from χ^2_ν independent of z
4. Set $y = Az$
5. Set $\eta'_{T,k} = \sqrt{(\nu/s)}y$

Then, we can simulate K vectors $(x_{1,T,k}, x_{2,T,k}, \ldots, x_{N,T,K})$ and K values of $x_{T,k}$ by using model (15), for $k = 1, 2, \ldots, K$. We order series $\{x_{T,k}\}$ in increasing order. Then we have the VaR of portfolio by $VaR_T(\alpha) = x_{T,K_p}$, equivalently, it is exactly the Kpth element of simulation series after ordering by increasing order.

5 Application

In this section, the presented copula based models are applied to estimate VaR of a portfolio composed by six currencies to VND namely VND/AUD, VND/EUR, VND/GBP, VND/JPY, VND/USD and VND/CNY. The results are compared with

results of with AR(1)-GARCH(1, 1) + N and AR(1)-GARCH(1, 1) + t in which each return series is assumed to follow AR(1)-GARCH(1, 1) models with the innovations are separately modeled by univariate standard normal and student t distribution.

5.1 Data Description

The database contains 1328 daily closing prices, from January 2nd 2007 to March 30th 2012. We denote the log-returns, of six exchange rates by variable $x_1, x_2, \ldots, x_6$, respectively. Note that for each exchange rate n, the log-return at day t is defined by $x_{n,t} = \ln(p_{n,t}) - \ln(p_{n,t-1})$, where $p_{n,t}$ is the closing price of currency n at day t, $n = 1, 2, \ldots, 6$ and $t = 1, 2, \ldots, 1328$. Figure 2 presents the plots of six series and Table 1 contains descriptive statistics.

In Fig. 2, we can see the evidence of volatility clustering which can be processed using GARCH models. Table 1 shows that all of six return series distributions have large kurtosis, especially VND/USD and VND/CNY have very large kurtosis

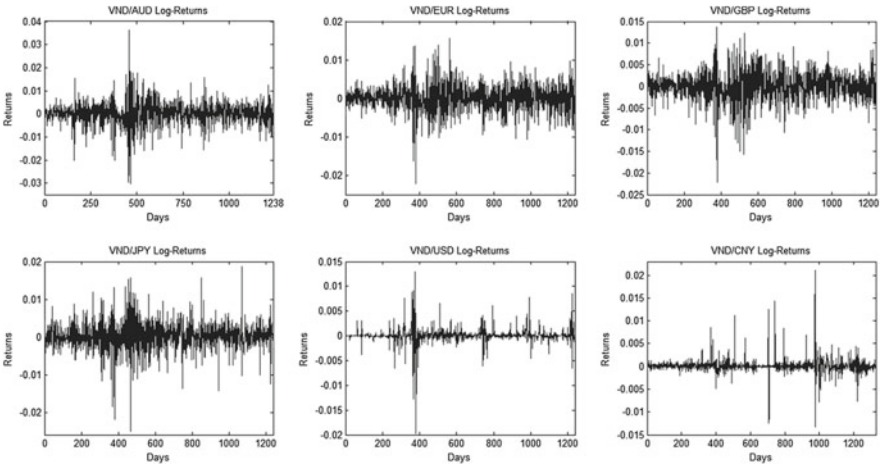


Fig. 2 Daily log returns of six currencies to VND

Table 1 Descriptive statistics of daily log-returns of six currencies to VND

Statistics	VND/AUD	VND/EUR	VND/GBP	VND/JPY	VND/USD	VND/CNY
Mean	17.16E–5	8.68E–05	1.72E–05	20.32E–5	8.53E–05	15.56E–5
Std	5.08E–3	3.50E–3	3.35E–3	3.66E–3	1.55E–3	1.56E–3
Minimum	−0.0303	−0.0222	−0.0221	−0.0250	−0.0198	−0.0133
Median	55.37E–5	19.14E–5	9.46E–05	10.47E–5	2.08E–05	5.94E–05
Maximum	0.0364	0.0157	0.0136	0.0186	0.0130	0.0212
Kurtosis	9.0186	5.5229	5.9954	7.8918	42.4722	59.0849
Asymmetry	−0.4711	−0.1548	−0.4275	−0.2536	−1.7482	3.1359

which make difficulty to capture its perturbation. The asymmetry of VND/EUR and VND/JPY are small implies that its distribution is nearly symmetric while other series have larger asymmetry.

5.2 Results and Evaluation

Let us consider a portfolio with equal weights for six indices, or in other word, the log return of portfolio at day tth is $x_t = \frac{1}{6}\sum_{n=1}^{6} x_{n,t}$.

In order to assess the accuracy of the estimated VaR we backtest the models at 95, 97.5, 99 and 99.5 % confidence level by the following procedure. For each day $T = 751, 752, \ldots, 1{,}327$, data in the 750 previous days are used to estimated VaR using AR(1)-GARCH(1, 1) + Gaussian copula and AR(1)-GARCH(1, 1) + Student t copula models. Since the dataset contains 1,327 observations, we have a total of 577 tests for VaR at each level α. We also do backtesting with AR(1)-GARCH(1, 1) + N and AR(1)-GARCH(1, 1) + t models in which each return series is assumed to follow AR(1)-GARCH(1, 1) model with the innovations are separately modeled using univariate standard normal and student t distribution. For each model, we repeat the test 10 times to access the robustness. To compare the performance of VaR estimation models, we compare the maximum, minimum and average of proportion of observations and number of proportion where the portfolio loss exceeded the estimated VaR among 10 testing times. The average number is average of proportion of observation (number of observations) in 10 testing times. The results are presented in Table 2.

In Table 2, the first two columns are corresponding to VaR estimation models using Gaussian and Student t copulas. Similarly, the last two columns are corresponding to VaR estimation models using AR(1)-GARCH(1, 1) and innovations are generated

Table 2 Proportion of observations (number of observations in brackets), for $t = 751$–$1{,}327$, where the portfolio loss exceeded the estimated VaR for $\alpha = 0.005, 0.01, 0.025$ and 0.05

Alpha (α)	Proportion	GARCH + Gaussian	GARCH + Student	GARCH + N	GARCH + t
$\alpha = 0.5\,\%$	Average	**0.0279(16.1)**	0.0307(17.7)	0.0808(46.6)	0.0858(49.5)
	Minimum	**0.0260(15)**	0.0277(16)	0.0780(45)	0.0832(48)
	Maximum	**0.0312(18)**	0.0329(19)	0.0832(48)	0.0858(50)
$\alpha = 1\,\%$	Average	**0.0387(22.3)**	0.0392(22.6)	0.0984(56.8)	0.1083(62.5)
	Minimum	**0.0364(21)**	0.0381(22)	0.0936(54)	0.1057(61)
	Maximum	**0.0416(24)**	**0.0416(24)**	0.1040(60)	0.1083(64)
$\alpha = 2.5\,\%$	Average	**0.0655(37.8)**	0.0747(43.1)	0.1322(76.3)	0.1336(77.1)
	Minimum	**0.0624(36)**	0.0728(42)	0.1300(75)	0.1300(75)
	Maximum	**0.0693(40)**	0.0780(45)	0.1352(78)	0.1369(79)
$\alpha = 5\,\%$	Average	**0.0924(53.3)**	0.0978(56.4)	0.1537(88.7)	0.1535(88.6)
	Minimum	**0.0936(52)**	0.0953(55)	0.1525(88)	0.1525(88)
	Maximum	**0.0936(54)**	0.0988(57)	0.1560(90)	0.1560(90)

Table 3 Proportion of observations (number of observations in brackets), for $t = 751$–1,327, where the portfolio loss exceeded the estimated VaR for $\alpha = 0.005, 0.01, 0.025$ and 0.05

Alpha (α)	Proportion	GARCH + Gaussian	GARCH + Student	GARCH + N	GARCH + t
$\alpha = 0.5\,\%$	Average	**0.0289(16.7)**	0.0302(17.4)	0.0801(46.2)	0.0854(49.3)
	Minimum	**0.0243(14)**	0.0260(15)	0.0780(45)	0.0832(48)
	Maximum	**0.0312(18)**	0.0329(19)	0.0832(48)	0.0884(51)
$\alpha = 1\,\%$	Average	0.0397(22.9)	**0.0395(22.8)**	0.1010(58.3)	0.1092(63.0)
	Minimum	**0.0364(21)**	**0.0364(21)**	0.0953(55)	0.1057(61)
	Maximum	**0.0416(24)**	**0.0416(24)**	0.1057(61)	0.1127(65)
$\alpha = 2.5\,\%$	Average	**0.0660(38.1)**	0.0768(44.3)	0.1314(75.8)	0.1335(77.0)
	Minimum	**0.0641(37)**	0.0745(43)	0.1300(75)	0.1317(76)
	Maximum	**0.0693(40)**	0.0797(46)	0.1335(77)	0.1352(78)
$\alpha = 5\,\%$	Average	**0.0946(54.6)**	0.1003(57.9)	0.1323(75.8)	0.1539(88.8)
	Minimum	**0.0919(53)**	0.0988(57)	0.1525(88)	0.1525(88)
	Maximum	**0.0971(56)**	0.1040(60)	0.1560(90)	0.1560(90)

using standard normal and Student t distribution. The results show that the AR(1)-GARCH(1, 1) + Gaussian copula model provided the best results for VaR estimation for all four levels of α. Two conditional copula based models provided better results comparing with two other models. Furthermore, the small difference between the minimum, maximum and average numbers of observations (proportion) among 10 repeated times shows that all four models are stable.

We also repeated the experiment for daily rate returns of six exchange rate with the rate return at day t is defined by $x_{n,t} = \frac{p_{n,t} - p_{n,t-1}}{p_{n,t-1}}$, where $t = 1, 2, \ldots, 1{,}328$ and $n = 1, 2, \ldots, 6$. The results are presented in Table 3.

Similar to the case of log return, all the experiment results of copula models are better than other models. The reason is that the copula could capture the dependence between series, which is then used to estimate portfolio distribution, while other models process without considering this dependency.

6 Conclusion

In this paper, we briefly review the basics of copula theory and two VaR estimation models namely AR(1)-GARCH(1, 1) + Gaussian copula and AR(1)-GARCH(1, 1) + Student t copula. Those models are applied to capture the dependency and estimate VaR of portfolio consists of six foreign exchange rate in Vietnam's market. The results of conditional copula based models are better than AR(1)-GARCH(1, 1) + N and AR(1)-GARCH(1, 1) + Student t models in which each return series is assumed to follow AR(1)-GARCH(1, 1) model and innovations are separately generated using standard normal and student t distribution. We repeat the estimation process 10 time and analyze the results to assess the stability of four models and make the conclusion that all considered models are quite stable.

Acknowledgments The authors are grateful to the anonymous reviewers and editors for their insightful and constructive comments that have helped to improve the presentation of this paper.

References

1. Bouye, E., Durrleman, V., Nikeghbali, A., Riboulet, G., Roncalli, T.: Copulas for Finance, A Reading Guide and Some Applications, Working paper, Financial Econometrics Research Center, City University, London (2000)
2. Cherubini, U., Luciano, E.: Value at risk trade-off and capital allocation with copulas. Econ. Notes **30**, 235–256 (2001)
3. Cherubini, U., Luciano, E., Vecchiato, W.: Copula Method in Finance. Wiley (2004)
4. Dias, A., Embrechts, P.: Dynamic copula models for multivariate high-frequency data in finance. Working Paper, ETH Zurich: Department of Mathematics (2003)
5. Embrechts, P., McNeil, A., Straumann, D.: Correlation and dependence in risk management: properties and pitfalls. Risk Manag. Value Risk Beyond, pp 176–223 (2002)
6. Embrechts, P., Hoing, A., Juri, A.: Using copulae to bound the value-at-risk for functions of dependent risks. Financ. stoch. **7**, 145–167 (2003)
7. Fortin, I., Kuzmics, C.: Tail dependence in stock return pairs. Int. J. Intell. Syst. Account., Financ. Manag. **11**, 89–107 (2002)
8. Georges, P., Lamy, A.G., Nicolas, E., Quibel, G., Roncalli, T.: Multivariate survival modelling: a unified approach with copulas. Working paper, Credit Lyonnais, Paris (2001)
9. Hansen, B.: Autoregressive conditional density estimation. Int. Econ. Rev. **35**, 705–730 (1994)
10. Meneguzzo, D., Vecchiato, W.: Copulas sensitivity in collaterized debt obligations and basket defaults swaps pricing and risk monitoring. Working paper, Veneto Banca (2002)
11. Nelsen, R.B.: An Introduction to Copulas. Springer (2005)
12. Nguyen, P.T., Nguyen, T.: A Market Microstructure Approach to the Foreign Exchange Market in Vietnam. In: 22nd Australasian Finance and Banking Conference 2009 (2009)
13. Nguyen, T.P., Nguyen, D.T.: Vietnam's exchange rate policy and implication for its foreign exchange market, 1986–2009 (2010)
14. Patton, Andrew J.: Modelling asymmetric exchange rate dependence. Int. Econ. Rev. **47**(2), 527–556 (2006)
15. Palaro, H.P., Hotta, L.K.: Using Conditional Copula to Estimate Value at Risk, State University of Campinas, Journal of Data Science (2006)
16. Rockinger, M., Jondeau, E.: Conditional Dependency of Financial Series: An Application of Copulas. Working paper NER # 82, Banque de France. Paris (2001)
17. Sklar, A.: Random Variables, Distribution Functions, and Copulas, A Personal Look Backward and Forward. In: Distributions with Fixed Marginals and Related Topics. IMS Lecture Notes—Monograph Series, vol. 28, p. 14 (1996)

The Effects of Foreign Direct Investment and Economic Development on Carbon Dioxide Emissions

Shu-Chen Chang and Wan-Tran Huang

Abstract This paper uses a threshold model to estimate the regime-specific marginal effect of foreign direct investment (FDI) and economic development on environmental carbon dioxide (CO_2) emissions within different regimes of population density. Our results demonstrate an asymmetrical nonlinear relationship between gross domestic product per capita and CO_2 emissions in different regimes of population density. In addition, our results reveal that CO_2 emissions decline significantly along with increasing FDI until a certain level of population density is reached. Our results also show that CO_2 emissions increase along with increasing value-added in industry during the early and growth stages of the industrial life cycle and decrease during its mature stage, when it has higher energy efficiency.

JEL Classification: C33 · G11 · Q53

1 Introduction

There is no uniform conclusion regarding whether foreign direct investment (FDI) and economic development are good or bad for the environment. There are two different findings on the issue of FDI and pollution. The first is that FDI decreases environmental pollution because foreign enterprises have clean technology, which improves energy efficiency and resource usage [3, 19, 30, 32, 38, 45]. The second is that FDI increases environmental pollution because pollution-intensive industries (such as chemicals, pesticides, oil refining, textiles, metal smelting, iron and steel, and food processing industries) flow from home countries to host countries [3, 10, 16, 24].

S.-C. Chang
Department of Business Administration, National Formosa University, Huwei, Taiwan
e-mail: shu-chen@nfu.edu.tw

W.-T. Huang (✉)
Department of Business Administration, Asia University, Taichung, Taiwan
e-mail: wthuang@asia.edu.tw

V.-N. Huynh et al. (eds.), *Econometrics of Risk*, Studies in Computational Intelligence 583, DOI 10.1007/978-3-319-13449-9_34

At the same time, there are two different findings on the issue of economic development and pollution. The first is that economic development decreases pollution because pollution is measured as an inferior good at high levels of income. In other words, pollution and economic development increase together up to a certain income level, after which the trend reverses. The second is that economic development increases pollution. At low levels of income, the "pollution is an inferior good" finding is not obtained by the effect of the increase in economic activities. Thus, the relationship between economic development and environmental quality is an inverted U-shape, called an environmental Kuznets curve (EKC) by Grossman and Krueger [21, 22]. As per the above discussion, the impact of FDI inflows and economic development on environmental pollution is nonlinear. Although many previous studies (e.g., [3, 10, 16, 24]) have discussed the effect of income on pollution (or the effect of FDI inflows on pollution), their models include income, its square and cubic term (or FDI and its square) to measure the nonlinear relationship between FDI and pollution or between income and pollution.

Based on neo-Malthusian theory, increased population density creates environmental degradation. In contrast, ecological evolutionary theory proposes that higher population density or urban agglomeration increases technological efficiency in the use of fossil fuels and reduces CO_2 emissions. To clearly define the impacts of both FDI and economic development on environmental degradation, this paper uses population density to be a threshold variable.

Some previous studies have suggested that the relationship between environmental degradation and FDI should introduce additional explanatory variables, such as the quality of political institutions [11] or population density [24]. Although some studies have shown that the environmental degradation effect of FDI depends on the quality of political institutions and used an interaction term to measure it, the "interaction term" approach did not allow the marginal impact of FDI and economic development to be regime specific. In other words, their estimation cannot capture that the effect of FDI and economic development on environmental degradation is different across regimes. Furthermore, their results may suffer from multicollinearity because the interaction term is highly correlated with original variables when an interaction term is derived by multiplying two predictor variables.

Hansens [23] threshold regression considered thresholds or asymmetry effects and split sample data into several distinct regimes according to a specific threshold variable. Thus, this paper uses Hansens threshold approach in regarding population density as a threshold variable, tests whether the threshold effect exists, and then estimates how marginal effects of FDI and economic development on CO_2 emissions differ across regimes that are identified by population density.

The remainder of this article is organized into six sections. Section 2 provides a review of the literature on FDI, economic development, and environmental degradation. Section 3 discusses the methodology. Section 4 presents the empirical results. Section 5 concludes and presents some implications of our findings.

2 Literature Review

2.1 Economic Development and Environmental Degradation

Greenhouse gases are produced by human activity primarily through the burning of fossil fuels. CO_2 emissions comprise the most important contributor to greenhouse gases and are directly related to energy use, which is an essential factor in economic activities such as industrial production and consumption. Although CO_2 emissions contribute to global warning and its social costs, Arrow et al. [1] and Friedl and Getzner [17] suggest that free-rider behavior in CO_2 emissions creates a close relationship between CO_2 emissions and income at all levels of income per capita. Thus, CO_2 emissions play an important role in the current debate on environmental degradation, so the relationship between CO_2 emissions and income is worthy of investigation.

In the past two decades, most previous studies used cross-section, panel data, and time series to investigate the EKC hypothesis on CO_2 emissions. In this case, the EKC hypothesis claims that the relationship between income per capita and CO_2 emissions has an inverted U-shape. In the CO_2-income framework, Coondoo and Dinda [9], Galeotti and Lanza [18], Moomaw and Unruh [35], and Shandra et al. [41] use a linear-quadratic model to support the EKC hypothesis on CO_2 emissions. When the linear-quadratic form is extended to a linear-cubic form for the CO_2 emissions model, Galeotti and Lanza [18] and Moomaw and Unruh [35] find an N-shaped relationship between CO_2 emissions and income, implying that as income grows over time, CO_2 emissions first increase, then decrease after the threshold income has been reached, and then increase again as income continues to grow that is, reducing CO_2 emissions exists in a narrow income range. Asici [2] find that positive effect of income on pollution is in middle income countries, but in high income countries, the effect is a statistically significant-negative result. Recently, Chang and Chang [7] use a panel data set from 1995 to 2005 that includes 57 countries, and also demonstrate the threshold effect of income on CO_2 emissions. In contrast, other studies [6, 8, 25, 40] found that CO_2 emissions increase monotonically with income per capita. Thus, the relationship between CO_2 emissions and income is mixed, and the validity of these studies has been questioned by Stern [44].

2.2 Population Density and Environmental Degradation

Although early studies, such as Grossman and Krueger [22], Moomaw and Unruh [35], and Shafik [40], concentrated on income as an explanatory variable and used ordinary least squares (OLS), later studies introduce additional explanatory variables to decompose factors of pollutant emissions. CO_2 emissions are attributed to direct and indirect effects. Direct effects occurred because of population density and

fossil-fuel use on CO_2 emissions while FDI is referred to as an indirect effect. Several studies assume a unitary elasticity of CO_2 emissions with respect to population [14, 46].

A few studies did not assume that the elasticity of CO_2 emissions with respect to population is unity but, on the contrary, they considered population an additional explanatory variable in their EKC model. However, their findings are inconsistent. For example, Bruvoll and Medin [6], Cropper and Griffiths [13], and Dietz and Rosa [14] find significant monotonically increasing relationships between population and CO_2 emissions, supporting the Malthusian assertion. Alternatively, Lantz and Feng [29] point out that the findings in Selden and Song [39] and Patel et al. [37] show a negative relationship between population density and CO_2 emissions, implying that population growth increases consideration of environmental degradation. Shi [42], using 93 countries over the period 1975–1996, found evidence of an out-of-sample EKC for CO_2 emissions per capita, in which the turning point is far beyond the maximum income in sample data and shows a monotonically increasing relationship between CO_2 emissions per capita and population. Lantz and Feng [29], using Canada over the period 1970–2000, did not support an EKC hypothesis for CO_2 emissions, but show an inverted-U-shaped relationship between CO_2 emissions and population as well as a U-shaped relationship between CO_2 emissions and technology, adding population and technology variables.

The indirect effects of CO_2 emissions can be classified in three categories: scale, composition, and technique effects. The scale effect relates to economic activity and is measured using income or economic growth variables. The composition effect relates to structural change in the economy, and the technique effect relates to technology adoption. Grossman [20] used economic growth, industrial composition, and environmental regulation to measure scale, composition, and technique effects, respectively. He found that these effects play an important role in pollution. Some studies include scale, composition, and technique variables in their EKC model, but their findings are inconsistent.

In the early 1990s, the International Bank for Reconstruction and Development [26] argued that an inverted-U-shaped relationship could be influenced by advanced techniques or structural changes due to globalization. This implies that economic growth leads to environmental degradation in the early stages of economic growth while it improves environmental quality in later stages due to environmentally friendly production techniques [5, 41]. For example, Neumayer [36] and Shafik [40], find evidence that the relationship between technology and CO_2 emissions is increasing, but Bruvoll and Medin [6], Shi [42], and Talukdar and Meisner [45] find a decreasing relationship between these two variables. The former finding implies that technological innovations and structural changes increase energy consumption while the latter finding shows that technological innovations and structural changes lead to production activities that are environmentally friendly. Lantz and Feng [29] found a U-shaped relationship between CO_2 emissions and technology in Canada and cited Shafiks [40] argument to explain that this effect is from "enhancing more environmentally friendly production techniques" to "encouraging CO_2 emissions enhancing production". These previous studies have in common that they find a

linear relationship between CO_2 emissions, income, and other variables, but their findings are not consistent.

2.3 FDI and Environmental Degradation

Some studies have attempted to investigate the ambiguous relationship between income, pollution, and FDI by applying a nonlinear model [11, 12, 16]. Their models suggest that the environmental pollution effect of FDI depends on socioeconomic conditions, such as population density [24, 28], national income [16], and political institutions [11, 12]. Among these socioeconomic conditions, population density is often used as a socioeconomic condition. With respect to population density, He [24] and Lan et al. [28] showed that pollution increases with high population density, given the same income and pollution levels. Shandra et al. [41] used population density to capture urbanization and examine the effect of population density on CO_2 emissions. In addition, regarding political institutions, Cole et al. [12] and Cole and Fredriksson [11] suggested that the degree of local government corruptibility plays an important role in the effect on environmental degradation of FDI because it affects the success of industrial lobbying.

3 Empirical Methodology

3.1 Theoretical Framework

FDI inflows provide high-technology or low-technology transfer and knowhow that is embodied in human capital. Technology can be expressed as a function of FDI. Previous studies also use FDI as a proxy for technology [4, 27, 33]. Regarding the relationship between technology, income and environmental pollution, Ehrlich and Holdrens [15] develop an IPAT model to discuss it, and provide a following simple theoretical framework.

$$I_{it} = f_1(P_{it}, A_{it}, T_{it}) \tag{1}$$

where I denotes environmental impact (e.g., environmental pollution); P denotes population size; A denotes affluence; and T denotes technology. Equation (1) shows these variables on the right-hand side as complex 8 interactions, which are therefore simplified in the model. Dividing by P, Eq. (1) can be rewritten as follows.

$$y_{it} = f_2(A_{it}, T_{it}) \tag{2}$$

where $y_{it} = \dfrac{I_{it}}{P_{it}}$ and y_{it} is per capita environmental pollution. The equation provides a way to analyze per capita environmental pollution and its distribution across countries and over time. Equation (2) is a modified version of the traditional EKC framework in which income is the only explanatory variable. The difference between Eq. (2)

and the traditional EKC framework is that Eq. (2) relates technology and income to per capita environmental pollution. That is, per capita environmental pollution in a country is the result of its affluence and the technology that it has implemented.

Because the IPAT framework is a useful instrument for illustrating environmental impact, previous empirical studies have tried to quantify the contribution of population, affluence, and technology to environmental deterioration. However, proxies for per capita affluence and technology cannot be readily identified. Thus, per capita real income is often used as a proxy for affluence. With respect to technology, that which results from research and development is transferred by FDI.

3.2 Empirical Model

This paper attempts to identify the driving forces behind environmental pollution on the basis of the IPAT framework and the EKC. Following previous studies in an IPAT framework, this paper uses per-capita GDP to proxy per-capita affluence and FDI as an indicator of technology to control for the spillover effects on the environment.

Based on the ecological-modernization perspective that the relationship between economic development and environmental degradation takes an inverted U-shape [22], here we add GDP per capita and its square to the environmental degradation model to capture this nonmonotonic relationship. The hypothesis of an inverted-U-shaped curve is proven when the coefficients of GDP per capita are positive and that of its square are negative. The inverted-U-shaped curve reflects a progression from relatively dirty technologies to cleaner technologies.

As described in the above discussion, in this paper the reduced-form relationship between FDI, economic development, and environmental degradation is modeled as follows,

$$PO_{it} = f(FDI_{it}, GDP_{it}, GDP_{it}^2, x_{it}) + u_{it} \tag{3}$$

where $f(\cdot)$ is a linear function. PO_{it} denotes environmental degradation. FDI_{it} is the ratio of net inflows of FDI in GDP; GDP_{it} represents economic development. GDP_{it}^2 is the square of GDP per capita. x_{it} is a set of control variables including composition changes (IV_{it}), energy efficiency (ENG_{it}) and government consumption (GOV_{it}).

To capture the nonlinear effect of FDI inflows and economic development on environmental pollution, there are two ways: regressions with interaction terms and threshold regressions. Because the regressions with interaction term might cause multicollinearity, this paper uses threshold regressions to avoid this problem. Given a threshold variable, the relationship between FDI, economic development, and CO_2 emissions per capita can be described as follows:

$$\begin{aligned} PO_{it} = \mu_i &+ \left(FDI_{it} GDP_{it} GDP_{it}^2\right) I(q_{it} \le \gamma) A_1 \\ &+ \left(FDI_{it} GDP_{it} GDP_{it}^2\right) I(q_{it} > \gamma) A_2 + B' X_{it} + \varepsilon_{it} \end{aligned} \tag{4}$$

where the subscript i refers to countries and t refers to years. A parameter q_{it} is an exogenous threshold variable, which is used by population density to capture urbanization-economies [43]. $I(\cdot)$ is an indicator function. Variable γ is the threshold value to be estimated; ε_{it} is an error term; A_1, A_2, and B are parameters to be estimated.

4 Empirical Results

4.1 Results of Panel Unit-Root Test

The sample data comprise ten-year balanced panel data from 1996 to 2005 covering 84 countries (see Appendix, Table 7). Data definitions and statistical descriptions of these variables are shown in Tables 1 and 2, respectively. This paper examines stationarity of the variables using Levin et al. [31] panel unit-root test. In applying this test, the optimal lag structure was determined by Schwarzs Information Criteria. Table 3 shows that p-values can reject the unit-root hypothesis at the 5 % significance level. The finding indicates that all data are stationary over time.

Table 1 Data definitions

Symbol	Variable	Definition and source
CO_{2it}	CO_2 emissions development	CO_2 emissions per capita. Source: World Bank, World Development Indicators (WDI) 2007. (Unit: metric tons per capita)
GDP_{it}	Economic	GDP per capita expressed in constant PPP (purchasing power parity), in constant 2005 US$. Source: World bank, WDI 2007 (Unit: U.S. dollars)
FDI_{it}	Foreign direct investment	Ratio of net inflows of foreign direct investment to GDP. Source: World bank, WDI 2007. (Unit: %)
IV_{it}	Composition changes	Ratio of value added of industry to GDP. WDI 2007. (Unit: %) Source: World bank,
ENG_{it}	Energy efficiency	GDP per unit of energy used based on 2005 US$. Source: World bank, WDI 2007. (Unit: U.S. dollars constant PPP, in constant per kg of oil equivalent)
q_{it}	Population density	Population density. Source: World bank, WDI 2007. (Unit: people per sq km)
GOV_{it}	Government consumption	Ratio of general government expenditure to GDP. Source: World bank, WDI 2007. (Unit: %)
EE_{it}	Economic freedom	The index is from 0 (most free) to 100 (least free). Source: Heritage foundation
CF_{it}	Corruption freedom	The index is from 0 (highest corruption) to 100 (lowest corruption) Source: Heritage foundation
ED_{it}	Degree of universal education	School enrollment, secondary. Source: World bank, WDI 2007 (Unit: %)

Table 2 Statistical description of all variables

	CO_{2it}	GDP_{it}	IV_{it}	ENG_{it}	q_{it}	GOV_{it}	FDI_{it}	EE_{it}	CF_{it}	ED_{it}
Mean	4.63	11,116.16	31.44	5.51	115.24	15.07	3.37	60.89	44.05	76.23
Median	3.51	7,764.88	29.74	5.07	73.31	14.38	2.43	60.88	40	81.16
Maximum	24.67	41,873.25	63.60	13.39	1176.32	28.39	45.14	82.33	100	130.95
Minimum	0.06	499.27	10.52	1.20	1.47	4.36	−15.10	30.02	4	5.76
Std. dev	4.40	10,253.11	9.46	2.48	149.05	5.07	4.24	8.95	22.81	27.50
Obs	840	840	840	840	840	840	840	840	840	840

Table 3 Result of LLC's panel unit root test

	Level	First difference
CO_{2it}	24.28(0.00)*	27.22(0.00)*
GDP_{it}	9.36(0.00)*	38.00(0.00)*
IV_{it}	12.42(0.00)*	34.11(0.00)*
ENG_{it}	12.08(0.00)*	38.46(0.00)*
q_{it}	32.07(0.00)*	32.07(0.00)*
GOV_{it}	21.19(0.00)*	28.60(0.00)*
FDI_{it}	21.83(0.00)*	15.31(0.00)*
EE_{it}	7.07(0.00)*	9.91(0.00)*
CF_{it}	13.19(0.00)*	13.57(0.00)*
ED_{it}	7.35(0.00)*	18.91(0.00)*

*statistically significant at the 5 % level. Numbers in parentheses are p-values

Table 4 Tests of threshold effects

Threshold variable: population density		
Null hypotheses	LR statistic (p-value)	R^2
No threshold effect	164.264* (0.05)	0.464
Single threshold effect	194.022** (0.00)	0.539
Double threshold effect	104.645 (0.13)	0.595

p-values are critical values of LR statistic, and are generated on the basis of 1000 iterations.
* and ** indicate statistically significant at the 5 and 1 % level, respectively

4.2 Results of Threshold Tests

This paper applies the LR statistic using 1,000 bootstrap replications to test whether the relationship between FDI, economic development, and CO_2 emissions has more threshold effects. The results are reported in Table 4. In Table 4, the p-values in tests of no threshold and single threshold are 0.05 and 0.00, respectively. These the p-values can be rejected at the 10 significance level. However, the p-values in the double-threshold test cannot be rejected at the same level. Hence, there is a double-threshold effect on the CO_2 emissions model, which uses population density

Table 5 Robust test

	Threshold variables					
	Corruption		Degree of universal education		Economic freedom	
Null hypothesis	LR statistic (p-value)	R^2	LR statistic (p-value)	R^2	LR statistic (p-value)	R^2
No threshold effect	20.852 (0.54)	0.365	65.891 (0.19)	0.400	28.345 (0.47)	0.371

p-values are critical values of LR statistic, and are generated on the basis of 1000 iterations

as a threshold variable. All observations are split into three regimes (most densely populated economies, least densely populated economies, and moderately densely populated economies) depending on population density.

4.3 Results of Robust Test

In choosing a threshold variable for robustness, this paper uses one of the three variables (i.e. economic freedom, freedom from corruption, and degree of universal education) in turn as a threshold variable to threshold effect. The results are presented in Table 5. Comparing Table 5 with Table 4, it shows that economic freedom, freedom from corruption, and degree of universal education are insignificant at the 10 % significance level while population density is significant at the same level and has a larger R^2. This implies that population density is a significant and better threshold variable than the others. In addition, consideration of threshold effect with population density can improve the model's fitness considerably.

4.4 Results of Estimated Effects on CO_2 Emissions

The results for marginal effects of FDI and economic development on CO_2 emissions are reported in Table 6. The three regimes are referred to as "most densely populated economies", "moderately densely populated economies", and "least densely populated economies", where the value relative to population density falls to $q_{it} > 252.935$, $10.330 < q_{it} \leq 252.935$, and $q_{it} \leq 10.330$, respectively. In Table 6, GDP per capita has a positive and significant effect on CO_2 emissions at the 5 % significance level, but the square of GDP per capita has a negative and significant effect on the first and second regimes at the same level. This finding supports the ecological modernization theory, which indicates that the relationship between economic development and CO_2 emissions takes an inverted-U shape. It also supports the findings of Shandra et al. [41]. However, when population density is viewed as a threshold variable in the pollution model, our results show that the size of coefficients for CO_2 emissions of GDP per capita differ across regimes. A 1 % increase in GDP per capita increases CO_2 emissions by 0.209 % in least densely populated economies, 0.468 % in moderately densely populated economies, and 0.808 % in most densely

Table 6 Estimation results of FDI and economic development on CO_2 emissions

Independent variables	Coefficients	Standard deviation
IV_{it}	0.019**	0.006
ENG_{it}	−0.614**	0.047
GOV_{it}	−0.011	0.010
$FDI_{it}(q_{it} \leq \gamma_1)$	−0.068**	0.018
$FDI_{it}(\gamma_1 \leq q_{it} \leq \gamma_2)$	−0.014**	0.005
$FDI_{it}(q_{it} \leq \gamma_2)$	−0.005	0.019
$GDP_{it}(q_{it} \leq \gamma_1)$	0.209**	0.052
$GDP_{it}(\gamma_1 \leq q_{it} \leq \gamma_2)$	0.468**	0.043
$GDP_{it}(q_{it} \geq \gamma_2)$	0.808**	0.049
$GDP_{it}^2(q_{it} \leq \gamma_1)$	0.001	0.001
$GDP_{it}^2(\gamma_1 < q_{it} < \gamma_2)$	−0.005**	0.001
$GDP_{it}^2(q_{it} \geq \gamma_2)$	−0.012**	0.002
R^2	0.539	
LR statistic (p-value)	194.022 (0.000)	

Bootstrap p-value is generated on the basis of 1,000 iterations. The * denotes statistically significant at the 5 % level

populated economies. Among these three regimes, our results show that the most densely populated economies show the largest effect for GDP per capita on CO_2 emissions, and the least densely populated economies show the least effect. This finding shows that CO_2 emissions increase along with increasing GDP. In addition, it provides an economic interpretation that increasing population and GDP, which both increase energy consumption, sharply raises CO_2 emissions. Therefore, asymmetric effects argue that the nonlinear relationship between GDP per capita and CO_2 emissions depends on densely populated economies.

Regarding the effect of FDI on CO_2 emissions, a 1 % increase in FDI would decrease CO_2 emissions by 0.068 % in least densely populated economies and by 0.014 % in moderately densely populated economies. Although its effect in the most densely populated economies is negative, it is insignificant at the 5 % significance level. This finding shows that CO_2 emissions decline along with increasing FDI up to a certain level of population density. Three factors explain this result. First, countries with the highest population density, where value-added by industry to GDP is between 23.07 and 59.34 %, tend to have pollution-intensive production. In addition, Moomaws [34] research has also shown that industries located in the most densely populated economies tend to have pollution-intensive production. Second, in moderately densely populated economies, the ratio of value-added by industry to GDP is between 4.77 and 36.4 % over the sample period. These countries tend to have less pollution-intensive production, comparing to the most densely populated economies. It is reasonable to conclude that decreases in pollution-intensive production will be the driving force behind decreases in future CO_2 emissions. Third, in the sample countries with the lowest population density, the value-added by services to GDP is

between 30.82 and 70.62 % over the sample period. This implies that these countries have more higher production from the service sector and thus lower CO_2 emissions.

In the same table, government consumption GOV_{it} has an insignificant effect on CO_2 emissions at the 5 % significance level. Composition changes IV_{it}, which are measured by industry's share of GDP, have a significant and positive effect (0.019) on CO_2 emissions at the 5 % significance level. This finding shows that increases in industry raise CO_2 emissions. However, energy efficiency ENG_{it}, which is related to energy consumption per GDP, has a significant and negative effect (0.614) on CO_2 emissions at the 5 % significance level. This finding implies that energy efficiency reduces pollution when countries replace low-quality with high-quality energy. Thus, as illustrated in the above discussion on the effects of composition changes and energy efficiency, CO_2 emissions increase substantially along with increasing value-added in industry, which depends significantly upon the use of fossil fuels when its life cycle is in the early and growth stages. However, industry adopts advanced energy efficiency and less energy intensity as its life cycle approaches maturity, which later reduces CO_2 emissions. Our findings are consistent with Boserup's [5] and Shandra et al.'s [41] argument that as industry approaches a mature stage it experiences improved energy efficiency and less energy intensity and thus produces lower emissions of CO_2.

5 Conclusion and Economic Implication

Although the effect of FDI on environmental pollution has been previously investigated, there is little research on the threshold relationship between FDI, economic development, and CO_2 emissions. This study uses the following procedures to investigate the marginal effects of FDI and economic development on CO_2 emissions under regimes with varying levels of population density: (1) modeling environmental CO_2 emissions, including FDI, GDP per capita, a threshold variable, and control variables; (2) testing 17 threshold effects by using the level of population density as a threshold variable; and (3) estimating the marginal pollution effects of FDI and economic development. This paper checks robustness using different threshold variables.

This study suggests several findings and economic implications. First, population density is a better choice of threshold variable than the others (such as political institutions variable, economic freedom, and degree of universal education). There is a double-threshold effect of FDI and economic development on CO_2 emissions using population density as the threshold variable, which allows our sample to split into three regimes identified as most, moderately, and least densely populated economies. Second, our results in the most and least densely populated economies support the EKC hypothesis and the ecological-modernization theory that an inverted-U-shaped curve exists in the relationship between income and CO_2 emissions. In addition, the marginal effect of economic development on CO_2 emissions is nonlinear effect.

Third, our findings show that CO_2 emissions decline significantly along with increasing FDI up to a certain level of population density. Based on Adam Smiths

theory of comparative advantage, densely populated countries have a labor-pool advantage. Therefore, these countries easily attract inflows of labor-intensive industries (e.g., primary metals, printing, chemicals, and fabricated metals), which increases pollution. In contrast, countries with low or moderate population density may have the advantage of capital intensity or land intensity. Such countries focus on enhancing energy efficiency and agricultural production.

Final, while increased composition changes increase CO_2 emissions, energy efficiency reduces pollution. It implies that increasing value-added of industry raises CO_2 emissions during the early and growth stages of the industry life cycle, but decreases CO_2 emissions when it approaches maturity.

Appendix

See Table 7

Table 7 Country sample

Belgium, Denmark, Finland, France, Germany, Greece, Ireland, Italy, Netherlands,
Spain, Sweden, Iceland, Norway, Switzerland, Albania, Bulgaria, Czech Republic,
Hungary, Poland, Romania, Slovak Republic, Canada, Mexico, USA, Australia,
Japan, New Zealand, South Korea, Turkey, Botswana, Côte d′Ivoire, Egypt,
Kenya, Madagascar, Malawi, Namibia, Senegal, South Africa, Tanzania, Tunisia,
Uganda, China, Hong Kong, India, Indonesia, Israel, Jordan, Malaysia, Philippines,
Singapore, Sri Lanka, Thailand, Vietnam, Argentina, Brazil, Chile, Colombia,
Costa Rica, Dominican Republic, El Salvador, Nicaragua, Panama, Paraguay, Peru,
Trinidad and Tobago, Uruguay, Venezuela

References

1. Arrow, K., Bolin, B., Costanza, R., Dasgupta, P., Folke, C., Holling, C.S., Janssen, B.-O., Levin, S., Mler, K., Perrings, C., Pimentel, D.: Economic growth, carrying capacity, and the environment. Science **268**, 520–521 (1995)
2. Asici, A.A.: Economic growth and its impact on environment: a panel data analysis. MPRA Paper No. 30238 (2011)
3. Baek, J., Koo, W.: A dynamic approach to the FDI-environment nexus: the case of China and India. American Agricultural Economics Association Annual Meeting, Orlando, 27–29 July 2008
4. Blomstrom, M., Kokko, A.: Foreign direct investment and spillovers of technology. Int. J. Technol. Manag. **22**(5–6), 435–454 (2001)

5. Boserup, E.: Population and Technological Change. University or Chicago Press, Chicago (1981)
6. Bruvoll, A., Medin, H.: Factors behind the environmental Kuznets curve: a decomposition of the changes in air pollution. Environ. Resour. Econ. **24**(1), 27–48 (2003)
7. Chang, S.-C., Chang, T.-Y.: Threshold effects of economic growth on air pollution under regimes of corruption. Econ. Bull. **32**(1), 1046–1059 (2012)
8. Coondoo, D., Dinda, S.: Causality between income and emission: a country-group-specific econometric analysis. Ecol. Econ. **40**(3), 351–367 (2002)
9. Coondoo, D., Dinda, S.: The carbon dioxide emission and income: a temporal analysis of cross-country distributional patterns. Ecol. Econ. **65**(2), 375–385 (2008)
10. Copeland, B., Taylor, S.: Trade, growth and the environment. J. Econ. Lit. **42**(1), 7–71 (2003)
11. Cole, M.A., Fredriksson, P.G.: Institutionalized pollution havens. Ecol. Econ. **68**(4), 1239–1256 (2009)
12. Cole, M.A., Elliott, R.J.R., Fredriksson, P.G.: Endogenous pollution havens: does FDI influence environmental regulations? Scand. J. Econ. **108**(1), 157–178 (2006)
13. Cropper, M., Griffiths, C.: The interaction of population growth and environmental quality. Am. Econ. Rev. **84**, 250–254 (1994)
14. Dietz, T., Rosa, E.A.: Effects of population and affluence on CO_2 emissions. Proc. Natl. Acad. Sci. USA **94**, 175–179 (1997)
15. Ehrlich, P., Holdren, J.: A bulletin dialogue on the closing circle critique: one dimensional ecology. Bull. At. Sci. **28**(5), 16–27 (1972)
16. Fredriksson, P.G., List, J.A., Millimet, D.L.: Corruption, environmental policy, and FDI: theory and evidence from the United States. J. Public Econ. **87**, 1407–1430 (2003)
17. Friedl, B., Getzner, M.: Determinants of CO_2 emissions in a small open economy. Ecol. Econ. **45**, 133–148 (2003)
18. Galeotti, M., Lanza, A.: Richer and cleaner? A study on carbon dioxide emissions in developing countries. FEEM Working Paper No. 87 (1999)
19. Gray, W.B., Shadbegian, R.J.: When do firms shift production across states to avoid environmental regulation? NBER Working Papers No. 8705 (2002)
20. Grossman, G.M.: Pollution and growth: what do we know? In: Goldin, I., Winters, L.A. (eds.) The Economics of Sustainable Development, pp. 19–46. Cambridge University Press, Cambridge (1995)
21. Grossman, G.M., Krueger, A.B.: 'Environmental impacts of a North American free trade agreement', in the U.S.-Mexico free trade agreement. In: Garber, P. (ed.) The U.S.-Mexico Free Trade Agreement, pp. 165–177. MIT Press, Cambridge (1993)
22. Grossman, G.M., Krueger, A.B.: Economic growth and the environment. Q. J. Econ. **110**(2), 353–377 (1995)
23. Hansen, B.E.: Threshold effects in non-dynamic panels: estimation, testing, and inference. J. Econom. **93**(2), 345–368 (1999)
24. He, J.: Pollution haven hypothesis and environmental impacts of foreign direct investment: the case of industrial emission of sulfur dioxide (SO_2) in Chinese province. Ecol. Econ. **60**, 228–245 (2006)
25. Heil, M.K., Selden, T.M.: Carbon emission and economic development: future trajectories based on historical experience. Environ. Dev. Econ. **6**, 63–83 (2001)
26. IBRD: Development and the Environment. World Development Report 1992. Oxford University Press, New York, pp. 38–39 (1992)
27. Kohpaiboon, A.: Foreign trade regime and FDI growth nexus: study of Thailand. J. Dev. Stud. **40**, 55–69 (2003)
28. Lan, J., Kakinaka, M., Huang, X.: Foreign direct investment, human capital and environmental pollution in China. Environ. Resour. Econ. **51**(2), 255–275 (2012)
29. Lantz, V., Feng, Q.: Assessing income, population, and technology impacts on CO_2 emissions in Canada: wheres the EKC? Ecol. Econ. **57**, 229–238 (2006)
30. Letchumanan, R., Kodama, F.: Reconciling the conflict between the "pollution-haven" hypothesis and an emerging trajectory of international technology transfer. Res. Policy **29**, 59–79 (2000)

31. Levin, A., Lin, C.F., Chu, C.: Unit root tests in panel data: asymptotic and finite-sample properties. J. Econom. **108**, 1–24 (2002)
32. Liang, F.: Does foreign direct investment harm the host countrys environment? Evidence from China, Haas School of Business, University of California, Berkeley, Working Paper, September (2006)
33. Madariaga, N., Poncet, S.: FDI in China: spillovers and impact on growth. World Econ. **30**, 837–862 (2007)
34. Moomaw, R.L.: Agglomeration economies: localization or urbanization? Urban Stud. **25**(1), 50–161 (1988)
35. Moomaw, W., Unruh, G.: Are environmental Kuznets curves misleading us? The case of CO_2 emissions. Environ. Dev. Econ. **2**, 451–464 (1997)
36. Neumayer, E.: Can natural factors explain any cross-country differences in carbon dioxide emissions? Energy Policy **30**, 7–12 (2002)
37. Patel, S., Pinckney, T., Jaeger, W.: Smallholder wood production and population pressure in East Africa: evidence of an environmental Kuznets curve? Land Econ. **71**(4), 516–533 (1995)
38. Porter, M., van der Linde, C.: Toward a new conception of the environment-competitiveness relationship. J. Econ. Perspect. **9**(4), 97–118 (1995)
39. Selden, T.M., Song, D.: Environmental quality and development: is there a Kuznets curve for air pollution? J. Environ. Econ. Environ. Manag. **27**, 147–162 (1994)
40. Shafik, N.: Economic development and environmental quality: an econometric analysis. Oxf. Econ. Pap. **46**, 757–773 (1994)
41. Shandra, J.M., London, B., Whooley, O.P., Williamson, J.B.: International nongovernmental organizations and carbon dioxide emissions in the developing world: a quantitative, cross-national analysis. Sociol. Inq. **74**(4), 520–545 (2004)
42. Shi, A.: The impact of population pressure on global carbon dioxide emissions, 1975–1996: evidence from pooled cross country data. Ecol. Econ. **44**, 29–42 (2003)
43. Smith, D.F., Florida, R.: Agglomeration and industrial location: an econometric analysis of Japanese-affiliated manufacturing establishments in automotive-related industries. J. Urban Econ. **36**, 23–41 (1994)
44. Stern, D.I.: The rise and fall of the environmental Kuznets curve. World Dev. **32**(8), 1419–1439 (2004)
45. Talukdar, D., Meisner, C.M.: Does the private sector help or hurt the environment? Evidence from carbon dioxide pollution in developing countries. World Dev. **29**(5), 827–840 (2001)
46. York, R., Rosa, E.A., Dietz, T.: STIRPAT, IPAT and IMPACT: analytic tools for unpacking the driving forces of environmental impacts. Ecol. Econ. **46**(3), 351–365 (2003)

Author Index

A
Autchariyapanitkul, Kittawit, 219, 233
Ayusuk, Apiwat, 319

B
Buthkhunthong, P., 161

C
Chanaim, Somsak, 219, 233
Chang, Shu-Chen, 483
Chen, Juei Chue, 273
Cheng, Hsiao, 3
Choy, S.T. Boris, 435

D
Darolles, Serge, 85
Denœux, Thierry, 171, 185

G
Ghazali, Mohd Fahmi, 203
Gong, Xue, 305
Gouriéroux, Christian, 17, 85

H
Harnpornchai, Napat, 245
Hencic, Andrew, 17
Hor, Chantha, 415
Hu, Liangjian, 135
Huang, Wan-Tran, 483
Hunzinger, Chadd B., 41
Huynh, Van-Nam., 471

J
Junchuay, A., 161

K
Kaewsompong, Nachatchapong, 171
Kanjanatarakul, Orakanya, 171
Kreinovich, Vladik, 53, 63

L
Labuschagne, Coenraad C.A., 41
Lean, Hooi Hooi, 203
Leurcharusmee, Supanika, 185
Li, Baokun, 115
Liu, Jianxu, 287

M
Min, Nyo, 343

N
Nguyen, Hung T., 53, 63
Nguyen, Phuong Anh, 115, 149
Nguyen, Thien Kim, 391
Niwattisaiwong, Seksiri, 245, 427, 457
Nguyen, Vu-Linh, 471

O
Ongeera, I., 161
Ouncharoen, Rujira, 63

P
Pham, Thi Tuyet Trinh, 391

V.-N. Huynh et al. (eds.), *Econometrics of Risk*, Studies in Computational Intelligence 583, DOI 10.1007/978-3-319-13449-9

Phochanachan, Panisara, 377
Piamsuwannakit, Sutthiporn, 259
Pruekruedee, Sumate, 427, 449

R
Roland-Holst, David, 287

S
Santiwipanont, T., 161
Schoch, Daniel, 75
Sha, Wei-Shun, 273
Sirisrisakulchai, Jirakom, 185, 343, 359, 367
Sriboonchitta, Songsak, 53, 171, 185, 219, 233, 259, 287, 305, 319, 329, 343, 359, 377
Sumetkijakan, S., 161
Komsan, Suriya, 427, 449, 457

T
Tang, Jiechen, 329
Teiletche, Jérome, 85
Thaiprasert, Nalitra, 415
Tian, Weizhong, 135
Tran, Hien D., 135, 149

W
Wang, Tonghui, 115, 135
Wei, Zheng, 115
Wiboonpongse, Aree, 287
Wichian, Anyarat, 359
Wichitaksorn, Nuttanan, 435
Wu, Berlin, 273

Y
Yuan, Xinyu, 329

Z
Zilberman, David, 287