Marco Corazza
Florence Legros
Cira Perna
Marilena Sibillo
Editors

Mathematical and Statistical Methods for Actuarial Sciences and Finance

MAF 2016

Springer

Mathematical and Statistical Methods for Actuarial Sciences and Finance

Marco Corazza • Florence Legros • Cira Perna •
Marilena Sibillo
Editors

Mathematical and Statistical Methods for Actuarial Sciences and Finance

MAF 2016

Springer

Editors

Marco Corazza
Department of Economics
Ca' Foscari University of Venice
Venezia, Italy

Florence Legros
Département Finance, Audit, Comptabilité et Contrôle
ICN Business School
Nancy, France

Cira Perna
Department of Economics and Statistics
University of Salerno
Fisciano (SA), Italy

Marilena Sibillo
Department of Economics and Statistics
University of Salerno
Fisciano (SA), Italy

ISBN 978-3-319-50233-5 ISBN 978-3-319-50234-2 (eBook)
https://doi.org/10.1007/978-3-319-50234-2

Library of Congress Control Number: 2017962872

Printed on acid-free paper

This Springer imprint is published by Springer Nature
The registered company is Springer International Publishing AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Preface

This volume is a collection of referred papers from the several tens that were presented at the international conference *MAF 2016—Mathematical and Statistical Methods for Actuarial Sciences and Finance*.

The conference was held in Paris (France), from March 30 to April 1, 2016, at the Université Paris-Dauphine. It was organized by the Executive MBA CHEA—Centre des Hautes Etudes d'Assurances and by the Department of Economics of the Ca' Foscari University of Venice (Italy), with the collaboration of the Department of Economics and Statistics of the University of Salerno (Italy).

The conference was the seventh of an international biennial series which began in 2004. It was born out of the idea that enhanced cooperation between mathematicians and statisticians working in actuarial sciences, in insurance, and in finance can result in improved research in these fields.

The wide participation at all the conferences in the series constitutes proof of the merits of this idea.

The papers published in this volume present theoretical and methodological contributions and their applications to real contexts.

Of course, the success of *MAF 2016* would not have been possible without the valuable help of all members of the Scientific Committee, the Organizing Committee, and the Local Committee.

Finally, we are pleased to inform readers that the organizing machine for the next edition is already in action: the conference *MAF 2018* will be held in Madrid (Spain), from April 4 to 6, 2018 (for details visit the website http://www.est-econ.uc3m.es/maf2018/).

We look forward to seeing you there.

Venezia, Italy — Marco Corazza
Nancy, France — Florence Legros
Fisciano (SA), Italy — Cira Perna
Fisciano (SA), Italy — Marilena Sibillo
August 2017

Contents

The Effects of Credit Rating Announcements on Bond Liquidity: An Event Study

Pilar Abad, Antonio Diaz, Ana Escribano, and M. Dolores Robles

Abstract This paper investigates liquidity shocks on the US corporate bond market around credit rating change announcements. These shocks may be induced by the information content of the announcement itself, and abnormal trading activity can be triggered by the release of information after any upgrade or downgrade. Our findings show that: (1) the market anticipates rating changes, since trends liquidity proxies prelude the event, and additionally, large volume transactions are detected the day before the downgrade; (2) the concrete materialization of the announcement is not fully anticipated, since we only observe price overreaction immediately after downgrades; (3) a clear asymmetric reaction to positive and negative rating events is observed; (4) different agency-specific and rating-specific features are able to explain liquidity behavior around rating events; (5) financial distress periods exacerbate liquidity responses derived from downgrades and upgrades.

1 Introduction

Information on rating actions has been a permanent subject of debate. Credit rating agencies (CRAs) state that they consider insider information when assigning and revising ratings, without disclosing specific details to the public at large. The literature examines prices and/or returns responses to rating events. However, the information about the creditworthiness of issuers disclosed by rating actions can not only affect prices. Besides this, it can induce specific market dynamics concerning the liquidity of the re-rated bonds. One important role of ratings is to reduce the information asymmetry between lenders and borrowers. As this asymmetry is

P. Abad
Universidad Rey Juan Carlos de Madrid, Madrid, Spain
e-mail: pilar.abad@urjc.es

A. Diaz • A. Escribano (✉)
Universidad de Castilla-La Mancha, Albacete, Spain
e-mail: antonio.diaz@uclm.es; ana.escribano@uclm.es

M.D. Robles
Universidad Complutense de Madrid, Madrid, Spain
e-mail: mdrobles@ccee.ucm.es

M. Corazza et al. (eds.), *Mathematical and Statistical Methods for Actuarial Sciences and Finance*, https://doi.org/10.1007/978-3-319-50234-2_1

inversely related to liquidity, if credit rating changes (CRCs) release specific news about the financial situation of firms, they will affect firms' bond liquidity.

In order to analyze this question, we go beyond the traditional price analysis by analyzing corporate bond liquidity patterns around CRC announcements. We examine different dimensions of corporate bond liquidity, and compute different adaptations of traditional microstructure-based liquid measures on stock markets to bond markets, as well as other traditional bond market liquidity measures.[1]

Our paper relates to several strands of the literature. First, we contribute to research that seeks to better understand the liquidity of corporate bond markets (e.g. [1, 8] and Chen et al. [3]). Second, to the literature that studies the information content of rating announcements and their impact on bond market (e.g. Steiner and Heinke [18] and May [15]) and on stock markets (e.g. Norden and Weber [16]).

We study a comprehensive sample of 2727 CRCs in the whole US corporate bond market, using TRACE transaction data from 2002 to 2010. We also study the impact of the recent global financial crisis on the response of the different liquidity aspects to CRC announcements. We consider the default of Lehman Brothers in mid-2008 to be the starting point of the financial turmoil.

Our results indicate three clear patterns in liquidity and prices, depending on the time period around the announcement when we consider the whole sample period. First, we observe trends in prices and liquidity deterioration before the announcement. Additionally, nervousness emerges in the market the day before downgrades. Second, there is price pressure for a few days after the downgrades. This fact could imply transaction prices below fundamental values. Third, we observe that prices converge to the correct value and the level of trading activity clearly rises during the second fortnight. In the case of upgrades, there is no price impact.

Aside from analyzing impacts on liquidity derived from credit rating migrations, we examine the determinants of abnormal liquidity observed before and after the announcements. To find the drivers of abnormal liquidity, we carry out a cross-sectional analysis including as key factors different characteristics of the rating event, the issue, and the issuer, to explain liquidity responses to rating changes. Our premise is that rating changes that provide more relevant information to the market must cause stronger impacts on liquidity.

Our results should enable market participants to manage portfolios, given that they need to have an understanding of the way in which the liquidity and the liquidity premium on prices behave around CRC.

The remainder of the paper is arranged as follows: Sect. 2 explains the hypotheses to be tested. Section 3 presents the data description. Section 4 examines different measures of abnormal liquidity. The main results are presented in Sect. 5. Section 6 includes the cross-sectional analysis. Finally, Sect. 7 concludes.

[1]Recent papers using some of these measures corroborate the liquidity effects on prices (see, e.g., Bao et al. [1], Dick-Nielsen et al. [8] and Friewald et al. [10]).

2 The Expected Response of Liquidity to Rating Actions

We consider that the effects of rating changes can be explained by different possible hypotheses. The main hypothesis states that CRAs are supplied with considerable non-public information about firms, such as information about the total firm value and its organizational effectiveness. In the liquidity literature, the market microstructure models indicate that trading activity responses to news releases are related to the existence of asymmetric information among informed traders, uninformed traders, and market-makers. Kim and Verrecchia [14] state that the fact that some traders are able to make better decisions than others, based on the same information, leads to information asymmetry and positive abnormal trading volume, despite a reduction in liquidity after the release of new information about the firm. A rating revision may provide additional information about the firm. Different investors' risk perception can induce portfolio rebalancing processes. In this context, higher trading activity after CRCs will be expected.

The second theory we analyze is the reputation hypothesis. This hypothesis (Holthausen and Leftwich [11]) states that rating agencies face asymmetric loss functions, and that they allocate more resources to revealing negative credit information than positive information, because the loss of reputation is more severe when a false rating is too high than when it is too low. Reputation costs create an incentive for CRAs to truthfully reveal the investment quality, since investors can eventually learn and punish the agency.

The last hypothesis points out that the usual liquidity premium on prices widens after a CRC. We examine whether the liquidity impact on prices is exacerbated after a CRC. If rating announcements disclose new and relevant information about the default risk of a bond, then prices should immediately incorporate this information. Independently of price adjustment, liquidity may drive additional price changes. The traditional literature considers liquidity to be a key component of corporate bond prices. Recent papers corroborate this result. For instance, [1] conclude that illiquidity explains a substantial part of the yield spreads of high-rated bonds, overshadowing the credit risk component. Chen et al. [3] and Friewald et al. [10] also observe that the economic impact of liquidity is significantly larger for speculative-grade bonds.

3 Data Description

We use two main sources of data in our analysis: the NASD's Trade Reporting and Compliance Engine (TRACE) transactions data for corporate bonds and the Mergent Fixed Income Securities Database (FISD), with complete information on the characteristics of each bond.[2] According with FISD bond information, we limit the sample to straight corporate bonds.[3] FISD data set also provides rating information per bond from the main three CRAs, i.e. Moody's, Standard and Poor's, and Fitch, from which we compute more than 225,000 CR changes.

Matching both data sets, we select those CR announcements which meet certain criteria. First, a minimum trading activity level. The bond should be traded at least once in the 20 working days before the event, and once in a similar period after the event. Second, bonds must be traded for at least 20% of the trading days in the period of 1 month starting 2 months prior the CRC. Third, events preceded by other rating announcements in the previous 61 working days, are also ignored.

The final data sample consists of 2620 unique CRC, involving 1342 bonds from 286 issuers; and with nearly 4.5 million trades, it forms approximately 10.6% of the total number of trades reported by TRACE during the period July 1, 2002 to March 31, 2010.[4] It comprises 907 upgrades and 1713 downgrades.

4 Proxies of Liquidity

We consider liquidity proxies that focus on three aspects of liquidity, i.e. price impact, market impact, and trading frequency, and select some of the most used measures in the existing empirical literature on liquidity. As price impact measures, we compute the 'Amivest' ratio (*AV*), the 'Bao' et al. measure (*Bao*), the imputed roundtrip cost measure (*IRC*) and the 'price dispersion' measure (*PD*). As proxies of market impact, we consider both the 'market share' (*MS*) and the 'trading volume' (*TV*). Finally, we compute the 'number of trades' (*NT*) as trading frequency measure.

---

[2]The use of the TRACE data set for research purposes requires previous filtering. Edwards et al. [9] and Dick-Nielsen [7] propose using algorithms to filter out the reporting errors, and we apply these, introducing minor variations. From the original 45 million transactions, we include in our sample about 4.5 million.

[3]We exclude zero or variable coupon bonds, TIPS, STRIPS, perpetual bonds and bonds with embedded options, such as putable, callable, tendered, preferred, convertible or exchangeable bonds. Additionally, we ignore municipal bonds, international bonds and eurobonds. We also eliminate those bonds that are part of a unit deal.

[4]Bonds double- or triple-rated by more than one CRAs, are set as unique events. We compute the final rating as the average rating using the numeric value assigned by the long-term debt rating equivalences, with values from AAA = 1 to D = 25. CRCs in the opposite direction are ignored.

4.1 Amivest

$$AV_{i,t} = \frac{1}{D_{i,t}} \sum_{k=1}^{D_{i,t}} \frac{TV_{i,k}}{|r_{i,k}|} \tag{1}$$

The Amivest liquidity ratio proposed by Cooper et al. [4] is a liquidity proxy. Following this ratio, larger liquidity implies lower price impact of new transactions. Bonds with high levels of this ratio are the most liquid bonds, i.e. those bonds with less price impact.

4.2 Bao

$$Bao_{i,t} = -Cov(\triangle p_t, \triangle p_{t+1}) \tag{2}$$

The [1] measure is a illiquidity measure. Following this measure, an illiquid bond is traded with a large bid-ask spread, which implies highly negatively correlated consecutive prices and hence a high positive value of the *Bao* measure.

4.3 Imputed Roundtrip Cost

$$IRC_{i,t} = \frac{P_{i,t}^{max} - P_{i,t}^{min}}{P_{i,t}^{max}} \tag{3}$$

Dick-Nielsen et al. [8] proposes the *IRC* measure. According to this illiquidity measure, higher values represent large differences between the maximum and minimum prices, which can be interpreted as large transaction costs. Therefore, bonds with high *IRC* values are less liquid bonds.

4.4 Price Dispersion

$$PD_{i,t} = \sqrt{\frac{1}{\sum_{k=1}^{N_{i,t}} TV_{i,k,t}} \sum_{k=1}^{N_{i,t}} (p_{i,k,t} - \overline{p}_{i,t})^2 \cdot TV_{i,k,t}} \tag{4}$$

The price dispersion measure proposed by Jankowitsch et al. [12] is an illiquidity measure. It can be interpreted as the volatility of the price dispersion, hence a low level of *PD* measure indicates high liquidity, i.e. the bond can be traded close to its fair value.

4.5 Market Share

$$MS_{i,t} = \frac{\sum_{k=1}^{N_{i,t}} TV_{i,k,t}}{TTV_k} \tag{5}$$

Following Diaz et al. [6] we compute market share as the ratio of the trading volume of a bond in a day to the total trading volume in the whole market, including any transaction involving any outstanding issue. The larger the market share, the higher the bond liquidity.

Finally, we also include the trading volume (*TV*) as a market activity measure and the number of trades (*NT*) as a trading frequency measure. A larger volume as well as a large number of trades indicates large liquidity.

5 Effects of Rating Change Announcements on Liquidity

5.1 Methodology

To analyze the effects of CRC announcements we carry out an event study. We compare liquidity around the rating-change days to normal liquidity days. We define the date of the announcement as day $t = 0$, and compute the observed liquidity in a window around the CRC, from day $t = t_1$ to day $t = t_2$ (CRC window) for each CRC in the sample as the averaged liquidity proxy:

$$OL_{i(t_1,t_2)} = \frac{\sum_{t=t_1}^{t_2} l_{i,t}}{T_O} \tag{6}$$

where $OL_{i(t_1,t_2)}$ is the observed liquidity for bond i in the window (t_1, t_2), T_O is the number of days in (t_1, t_2) and $l_{i,t}$ is the liquidity proxy computed in a day-by-day basis for the CRC i.

We compare this observed liquidity to the expected liquidity in "stable-rating times".[5] We consider 2 months prior the CRC announcement without any other

[5]Following Corwin and Lipson [5] we use the firm-specific past history to compute the expected liquidity in a period of typical liquidity. Other alternative is considering an appropriate matching portfolio of similar bonds with stable ratings. However, the lack of liquidity in corporate bond markets makes this approach unsuitable. Bessembinder et al. [2] indicate that construct abnormal bond returns using the matching portfolio models (the benchmark is a matching portfolio designed

credit event. Then, we define the stability-rating window from day $t = -41$ to day $t = -21$. The window ends 21 trading days prior to the announcement day, in order to avoid possible price lead-up preceding the shocks. The expected liquidity, EL_i, for CRC i is computed as the averaged liquidity proxy in this benchmark window:

$$EL_{i(t_1,t_2)} = \frac{1}{N} \sum_{t=-41}^{-21} l_{i,t} \tag{7}$$

To taking into account the different scales of liquidity proxies across different bonds, we compare the liquidity around the rating change and the expected liquidity in logarithms. Then, the abnormal liquidity in a specific event window (t_1, t_2) for the CRC i is obtained as the difference between observed liquidity in that rating-change window and the expected liquidity both in logs:

$$AL_{i(t_1,t_2)} = ln(OL_{i(t_1,t_2)}) - ln(EL_{i(t_1,t_2)}) \tag{8}$$

where $ln(.)$ indicates the natural logarithm.

In order to test the null hypothesis of no effects on liquidity due to rating changes, we compute the Averaged Abnormal Liquidity as:

$$AAL_{(t_1,t_2)} = \frac{1}{N} \sum_{i=1}^{N} AL_{i(t_1,t_2)} \tag{9}$$

where N is the number of rating changes in the considered subsample of events (upgrades or downgrades). Under the null, the expected value of AAL must be zero.

To test the statistical significance we first compute the well-known t-ratio test, asymptotically normally distributed under the null hypothesis. Second, we compute two non-parametric tests (the Fisher sign test and Wilcoxon rank test) that are robust to non-normality, skewness and other statistical characteristics of liquidity data that may affect the t-ratio properties. The Fisher sign test calculates the number of times abnormal liquidity is positive. The Wilcoxon rank test accounts for information of both magnitudes and signs. We report p-values for the asymptotic normal approximation to these tests.[6]

We study different width windows to analyze the behavior of abnormal liquidity before and after the release date: $[-10, -6]$, $[-5, -1]$, $[0, 5]$, $[6, 10]$, and $[11, 20]$. Besides, we examine separately the effects from both, downgrades and upgrades.

to adjust risk and time-to-maturity) or the mean-adjusted model (the benchmark is Treasury security with the most similar maturity date) induces a bias.

[6] See Sheskin [17] for details.

5.2 Event Study Results for Downgrades and Upgrades

The left-hand side of Table 1 presents results for rating downgrades.[7] In the case of price impact proxies (Panel A), results suggest that liquidity becomes scarce both before and after the event. In the case of the *AV* liquidity measure, the mean abnormal value estimated is negative and statistically significant in all the pre- and post-announcement day event windows. This result suggests an increase of the cost of demanding liquidity around a downgrade event. This result is robust to the method that we use (mean and median abnormal liquidity, parametric and non-parametric tests). The abnormal value of the *Bao* illiquidity measure is only statistically significant for the last two windows, with positive estimated coefficients. This result suggests larger effective bid-ask spreads, i.e. a liquidity deterioration, during the period $[6, 20]$. The analysis for the *IRC* illiquidity proxy depicts significant positive abnormal values for all the windows, with higher mean and median values after the event. Results indicate a fall of liquidity around the event, i.e. transaction costs higher than usual. Finally, the *PD* illiquidity measure shows positive and statistically significant abnormal values after the event. The price dispersion increases after a downgrade. These results are in line with those in [18], that find significant bond price reactions after downgrades.

Panel B shows the market impact measures. *MS* and *TV* show a decrease in abnormal trading activity in the previous days to the downgrade and in the second post-announcement week, while these measures are positive and significant in the first week after. We highlight that mean and median abnormal liquidity are negative and particularly low in the weeks previous to the event. Trading activity drops again during the second post-announcement week $[6, 10]$ and converges to regular values from the second fortnight after the downgrade $[11, 20]$. Abnormal TV is insignificant for the first week $[0, 5]$. These findings are consistent with results in [15], that find increasing trading activity after CR downgrade announcements.

Finally, Panel C of Table 1 displays the results for the trading frequency measure, *NT*. In this case, there is a significantly abnormal number of trades after the event, in the $[0, 20]$ window. The estimated coefficient is especially high immediately after the event $[0, 5]$. These results are also robust to the tests applied.

The right-hand side of Table 1 presents the results for upgrades. Panel A shows the effects on liquidity, measured by the price impact measures. In general, mean and median abnormal values of the four proxies are statistically significant, but their signs are a little confusing. Liquidity level becomes abnormally high during all the periods $[-10, 20]$ according to three proxies (*Bao*, *IRC* and *PD*), but we observe the opposite result in the case of Amivest. This suggests shorter effective bid-ask spreads, lower transaction costs, and lesser price dispersion during all the periods, but a higher price impact of new transactions. Results show inappreciable differences of these proxies between each week around the event.

[7] Other results for different subsamples and periods are available upon request.

Table 1 Event study results for Downgrades and Upgrades

		Downgrades (N = 1713)					Upgrades (N = 907)				
Event window		[−10, −6]	[−5, −1]	[0, 5]	[6, 10]	[11, 20]	[−10, −6]	[−5, −1]	[0, 5]	[6, 10]	[11, 20]
Panel A: Price impact measures. H_0: Abnormal liquidity = 0											
AV	Mean	−0.737*	−0.666*	−0.426*	−0.671*	−0.332*	−0.766*	−0.699*	−0.419*	−0.807*	−0.282*
	Median	−0.732$^{+\diamond}$	−0.685$^{+\diamond}$	−0.446$^{+\diamond}$	−0.619$^{+\diamond}$	−0.329$^{+\diamond}$	−0.716$^{+\diamond}$	−0.601$^{+\diamond}$	−0.414$^{+\diamond}$	−0.750$^{+\diamond}$	−0.211$^{+\diamond}$
	% > 0	35.0%	35.5%	40.8%	36.6%	42.8%	35.5%	37.3%	41.8%	34.5%	45.9%
Bao	Mean	0.031	0.008	0.027	0.048*	0.107*	−0.019	−0.025	−0.078*	−0.070*	−0.115*
	Median	−0.008	−0.015	0.024	0.044	0.095$^{+\diamond}$	−0.010	−0.027	−0.099$^{+\diamond}$	−0.086$^{\diamond}$	−0.122$^{+\diamond}$
	% > 0	49.4%	49.2%	50.9%	51.2%	53.7%	48.8%	49.4%	46.1%	47.3%	46.6%
IRC	Mean	0.001*	0.003*	0.009*	0.005*	0.004*	−0.000*	−0.001*	−0.001*	−0.001*	−0.001*
	Median	−0.001^{+}	−0.000$^{\diamond}$	0.002$^{\diamond}$	0.001$^{\diamond}$	0.001$^{\diamond}$	−0.001$^{\diamond}$	−0.001$^{\diamond}$	−0.000$^{\diamond}$	−0.001$^{\diamond}$	−0.001$^{\diamond}$
	% > 0	45.6%	48.5%	57.3%	53.8%	55.4%	42.3%	41.9%	45.4%	40.4%	42.4%
PD	Mean	−0.083*	−0.003	0.180*	−0.022	0.144*	−0.103*	−0.204*	−0.180*	−0.225*	−0.274*
	Median	0.015	0.048$^{\diamond}$	0.188$^{\diamond}$	0.159$^{\diamond}$	0.144$^{\diamond}$	−0.084$^{\diamond}$	−0.110$^{\diamond}$	−0.129$^{\diamond}$	−0.092$^{\diamond}$	−0.141$^{\diamond}$
	% > 0	50.9%	52.2%	58.8%	57.5%	56.9%	44.4%	45.7%	42.9%	45.2%	42.5%
Panel B: Market impact measures. H_0: Abnormal liquidity = 0											
MS	Mean	−0.367*	−0.253*	0.102*	−0.112*	−0.034	−0.505*	−0.381*	−0.170*	−0.405*	−0.144*
	Median	−0.292$^{+\diamond}$	−0.188$^{+\diamond}$	0.154$^{+\diamond}$	−0.051$^{\diamond}$	0.004	−0.377$^{+\diamond}$	−0.231$^{+\diamond}$	−0.120$^{+\diamond}$	−0.294$^{+\diamond}$	−0.101$^{+\diamond}$
	% > 0	39.7%	43.7%	55.0%	48.5%	50.2%	36.5%	43.0%	45.4%	40.5%	45.0%
TV	Mean	−0.395*	−0.272*	0.019	−0.195*	−0.061*	−0.472*	−0.407*	−0.131*	−0.371*	−0.173*
	Median	−0.312$^{+\diamond}$	−0.205$^{+\diamond}$	0.050	−0.119$^{+\diamond}$	−0.014	−0.333$^{+\diamond}$	−0.267$^{+\diamond}$	−0.109$^{+\diamond}$	−0.247$^{+\diamond}$	−0.115$^{+\diamond}$
	% > 0	39.2%	41.7%	51.5%	45.6%	49.6%	37.8%	40.0%	46.0%	40.5%	43.4%

(continued)

Table 1 (continued)

		Downgrades (N = 1713)					Upgrades (N = 907)				
Event window		[−10, −6]	[−5, −1]	[0, 5]	[6, 10]	[11, 20]	[−10, −6]	[−5, −1]	[0, 5]	[6, 10]	[11, 20]
Panel C: Trading frequency measures. H_0: Abnormal liquidity = 0											
NT	Mean	−0.048*	0.001	0.141*	0.027*	0.032*	−0.035*	−0.044*	0.016*	−0.056*	−0.045*
	Median	−0.050+◇	−0.018	0.128+◇	0.020	0.034+◇	−0.039+◇	−0.057+◇	−0.022	−0.033+◇	−0.049+◇
	% > 0	45.9%	47.9%	58.0%	50.4%	52.3%	44.8%	44.9%	48.1%	45.8%	43.3%

This table shows T-ratio and non-parametric tests, for the Average Abnormal Liquidity (*AAL*). Results for downgrades are on the left side and for upgrades on the right side. *AV*, *Bao*, *IRC*, *PD*, *MS*, *TV* and *NT* represent the Amivest, Bao, Imputed Roundtrip Costs, Price Dispersion, Market Share, Trading Volume and Number of Trades. The time period covered is July 2002 to March 2010. The subsample for downgrade events involves 1050 different bonds from 223 different issuers, and for upgrades, 621 different bonds issued by 150 different issuers. *, + and ◇ indicates significance at 10% or lower level for the t-ratio, Sign and Rank tests respectively

Panels B and C show the results for the market impact and trading frequency measures. Results for the abnormal liquidity behavior of *MS*, *TV* and *NT* indicate a statistically significant drop in the trading activity around the upgrade, except for an increase in the average number of trades immediately after the announcement.

6 Determinants of the Liquidity Response to Credit Rating Changes

In this section we identify the drivers of the abnormal liquidity levels that we found in the previous sections. We expect to find stronger reactions of abnormal liquidity in the case of events that provide more information to the market.

First, we consider rating-specific features. As Jorion and Zhang [13] state, we expect that the liquidity reaction to rating changes would depend on the prior rating level. In addition, the liquidity response to the final CRC can be affected by the opening of a rating review process. The size of the rating change, i.e. the number of notches, must be also important in determining abnormal liquidity. We hypothesize that the number of notches downgraded (or upgraded) acts as a signal of the amount of information that this rating change conveys.

Besides, we analyse whether there is an agency-specific effect on the liquidity response to rating actions. We also consider information derived from rating agencies collected through dummy variables, such as, different initial ratings; simultaneous announcements; agencies disagreements; agencies consensus; rating trends or rating break trends.

To test these hypotheses, we run different regressions of the abnormal liquidity against a set of variables to control for relevant bonds' characteristics and dummy variables.[8] We estimate all econometric models by OLS and compute standard errors by using the White heteroskedasticity-consistent covariance matrix. We consider a 10% or lower significance level for the tests.

Table 2 shows the results for the impact of rating changes in liquidity during the 4 weeks after the rating change announcement.[9] Panel A displays the results for the pre-Lehman Brothers' Default (LBD) period. On the left-hand side, the results for downgrades indicates that variables related to rating-specific and agency-specific features explain abnormal liquidity levels after the disclosure of new information. Variables such as the prior rating, the existence of a previous review process, or the jump size explain abnormal liquidity after downgrades. On the right-hand side we present the results for upgrades, where we do not find a large number of variables that can explain the abnormal liquidity detected in previous sections. Panel B displays the results during the post-LBD period. On the left-hand side we

[8]As control variables we include the bond's age, the relative offering amount, and two dummy variables for firms in the industrial and financial industries.

[9]Other results for the pre-event window $[-5, -1]$ are available upon request.

present results for downgrades and the drivers of abnormal liquidity seem to be the same as for the pre-LBD period, but sometimes with different signs. Finally, on the right-hand side we present the results for upgrades. In this case, few from the proposed variables are able to explain abnormal liquidity after the credit event. We only observe a positive response in abnormal liquidity in the post-LBD period, when the upgrade implies a break in the rating history trend. These results are in line with results in [16], that find differences in the market reaction to rating announcements by Moody's and S&P.

7 Concluding Remarks

This paper examines liquidity performance around CR changes in the US corporate bond market. Our sample involves 1342 straight bonds from 286 issuers that are affected by 2620 rating changes over the period from July 2002 to March 2010.

Our results indicate shocks in liquidity around downgrades with three clear patterns: before, immediately after, and during 1 month after the announcement. First, liquidity proxies before downgrades suggest that the market anticipates the deterioration of credit quality. Second, there is price pressure and abnormally high trading volumes during the first days after the downgrades. Third, prices converge to a stable level, with low-impact price liquidity and normal levels of trading activity, during the second fortnight. In the case of upgrades, proxies of price impact show liquidity levels to be higher than usual, whereas trading activity remains below the common level. Our results are similar than those in [15], that show different intensity responses to CRC depending on whether it deals with a downgrade or an upgrade.

To conclude, the cross-sectional analyses show different variables that may affect abnormal liquidity levels. Distinguishing the different responses depending on pre- and post-LBD periods, the results show that abnormal liquidity before the announcement is affected by rating-change specific features, such as the size of the jump, or the category, for both upgrades and downgrades. Abnormal liquidity immediately after the announcement can be explained by rating change characteristics, principally by the size of the jump and the fallen angel/rising star condition, for both upgrades and downgrades. These findings are in line than those reported by Jorion and Zhang [13], where the bond reaction to credit rating changes depend on the prior rating level.

Finally, our findings have important implications for corporate bond pricing, as long as prices around credit events may have information. Credit risk managers can also anticipate trading and search for better positions.

Table 2 Determinants of abnormal post-event [0, 20] liquidity

	Panel A. Pre-LBD period										Panel B. Post-LBD period									
	Downgrades (N = 1237)					Upgrades (N = 817)					Downgrades (N = 476)					Upgrades (N=90)				
	IRC	PD	MS	TV	NT	IRC	PD	MS	TV	NT	IRC	PD	MS	TV	NT	IRC	PD	MS	TV	NT
C	−0.01	−0.81	−0.49*	−0.50*	−0.63*	0.00	1.66	−0.51	−0.58	−0.76*	0.00	0.22	0.19	0.14	−0.52*	0.02	−0.95	−0.04	0.26	−0.35
PRI	0.00*	0.08*	0.06*	0.07*	0.06*	0.00	−0.04	0.01	0.01	0.01	−0.01*	−0.05*	−0.02	−0.02	0.01	0.00	0.04	0.03	0.00	−0.02
WN/WP	−0.00*	0.05	−0.10*	−0.09*	0.02	0.00	0.04	0.19*	0.15*	−0.03	0.02*	0.14*	0.24*	0.26*	0.09	0.00	−0.22	−0.05	0.07	−0.05
JUM	0.01*	0.11*	0.09*	0.08*	0.09*	0.00	−0.08	0.06	0.09*	0.03	−0.02*	−0.16*	−0.17*	−0.18*	−0.04	−0.01	−0.39*	−0.05	−0.05	−0.06
SPE	−0.03*	−0.52*	−0.44*	−0.53*	−0.55*	0.00	0.23	0.04	0.1	0.01	0.04*	0.32*	−0.20	−0.21	−0.21	0.02	−0.29	−0.62	−0.43	−0.47
LOW	−0.01*	−0.25	−0.07	−0.15*	−0.13*	0.00	0.27	−0.02	−0.05	0.00	0.03*	0.30*	0.05	0.10	0.06	0.01	0.41	0.35	0.20	0.74*
FA/RI	−0.02*	−0.33*	−0.13*	−0.20*	−0.23*	0.00	−0.09	0.25*	0.20	0.02	−0.02*	−0.12	−0.14	−0.46*	−0.26*	0.04	4.28*	0.47	0.18	1.23
BSJ	−0.01*	−0.28	−0.05	−0.07	−0.11*	0.00	0.09	0.04	0.05	0.02	0.01*	0.03	−0.08	−0.08	−0.01	0.01	0.41*	0.23	0.06	−0.23
SIM	0.00	0.20	0.09	0.12	0.20*	0.00	0.21	0.17	0.10	0.16	0.07*	0.34*	0.37	0.52*	0.10					
SEC						−0.00*	−0.24*	0.12*	0.03	−0.15*										
TRE	−0.01*	−0.10	0.02	0.02	−0.04	0.00	−0.02	0.10	0.11*	0.00	−0.02*	−0.18*	0.03	−0.08	−0.02	0.01	0.04	0.20	0.18	−0.13
BRT	−0.00*	−0.15	−0.02	0.00	0.04	0.00	−0.09	0.01	0.00	0.03	−0.01*	−0.12*	0.03	−0.1	−0.11	0.00	−0.02	0.20*	0.20*	0.01
S&P	0.01*	0.02	0.10*	0.14*	−0.07*	0.00	0.11	0.03	0.08	0.04	−0.01	0.00	−0.18*	−0.20*	−0.06	0.01	0.47	−0.16	−0.01	0.34
MOO	0.00	−0.1	−0.04	−0.02	−0.13*	0.00	0.14	0.01	0.05	0.02	0.02*	0.17*	0.00	0.03	0.02	0.00	0.08	−0.25	−0.15	0.09
FIN	0.00	0.75	0.14	0.09	−0.08	0.00	−1.64	0.26	0.25	0.06	0.00	−0.19	−0.43*	−0.38*	−0.31*	0.00	0.49	−0.28	−0.29	−0.14
IND	0.00	0.42	0.02	0.02	−0.16*	0.01*	−1.93	0.24	0.24	0.07	0.01	−0.11	−0.36*	−0.22	−0.24*					

(continued)

Table 2 (continued)

	Panel A. Pre-LBD period										Panel B. Post-LBD period									
	Downgrades (N = 1237)					Upgrades (N = 817)					Downgrades (N = 476)					Upgrades (N=90)				
	IRC	PD	MS	TV	NT	IRC	PD	MS	TV	NT	IRC	PD	MS	TV	NT	IRC	PD	MS	TV	NT
AGE	0.00	0.15	0.11	0.07	0.12*	0.00	0.06	−0.05	−0.04	−0.02	0.01	−0.03	0.14	0.17	0.13*	−0.01*	−0.44*	−0.28*	−0.23	−0.34*
WEI	0.00	−0.03	0.02	0.01	−0.02*	0.00	−0.05*	0.00	0.00	0.01	0.00	0.01	0.01	0.01	0.02*	0.00	0.32	0.22	0.18	0.07
R^2	15.8%	5.5%	8.1%	7.1%	14.4%	2.8%	5.1%	3.1%	3.6%	3.3%	45.8%	27.2%	10.8%	17.2%	11.3%	29%	22.8%	40.3%	39%	43.3%

Panel A shows results before Leman's default, and *Panel B* after its collapse. All models are estimated by OLS, with the robust covariance matrix adjusted for heteroskedasticity. *indicates significance at 10% level or lower. *IRC, PD, MS, TV* and *NT* represent the Imputed Roundtrip Costs, Price Dispersion, Market Share, Trading Volume and Number of Trades. *PRI* is a variable that indicates the previous rating of the firm, encoded by assigning value 1 to AAA, … and 25 to D. *WN/WP* is a dummy variable that takes value 1 if the bond is in a negative (N) or positive (P) watching list before the announcement, and zero otherwise. *JUM* indicates the numerical value of the rating change, following the previous codes. *SPE* is a dummy variable equal to one if the prior rating of the firm is on speculative level and zero otherwise. *LOW* is a dummy variable that takes value one if the bond has a rating below A1 or A+. *FA/RI* is a dummy variable equal to one if the rating change is a fallen angel (downgrades) or if it is a rising star (upgrades) and zero otherwise. *BSJ* is a dummy variable equal to one if the downgrade involves a change in the rating tranches established in Basel II, and zero otherwise. *SIM* is a dummy variable equal to one if the rating action has been announced by two or more agencies the same day, and zero otherwise. *SEC*, i.e. second mover, is a dummy variable equal to one if the rating change has been preceded by other change from a different agency and in the same direction within 90 days, and zero otherwise. *TRE* is a dummy variable that takes value one when the rating change has been preceded by two changes in the same direction, and zero otherwise. *BRT* is a dummy variable equal to one if the rating action has been preceded by two rating announcements in the opposite direction and zero otherwise. *S&P* (*MOO*) is a dummy variable equal to one if the rating action is announced by S&P (Moody's) and zero otherwise. *FIN* (*IND*) is a dummy variable equal to one if the rating action involves financial (industrial) sector firms, and 0 otherwise. *AGE* is the age of the bond at the announcement date.*WEI* represent the weight of the average dollar offering amount by bond's issuer over the average dollar offering amount by bond's principal issuer

References

1. Bao, J., Pan, J., Wang, J.: The illiquidity of corporate bonds. J. Financ. **66**(3), 911–946 (2011)
2. Bessembinder, H., Kahle, K.M., Maxwell, W.F., Xu, D.: Measuring abnormal bond performance. Rev. Financ. Stud. **22**(10), 4219–4258 (2008)
3. Chen, L., Lesmond, D.A., Wei, J.: Corporate yield spreads and bond liquidity. J. Financ. **62**(1), 119–149 (2007)
4. Cooper, S.K., Groth, J.C., Avera, W.E.: Liquidity, exchange listing, and common stock performance. J. Econ. Bus. **37**(1), 19–33 (1985)
5. Corwin, S.A., Lipson, M.L.: Order flow and liquidity around NYSE trading halts. J. Financ. **55**(4), 1771–1805 (2000)
6. Diaz, A., Merrick, J., Navarro, E.: Spanish Treasury bond market liquidity and volatility pre- and post-European Monetary Union. J. Bank. Financ. **30**, 1309–1332 (2006)
7. Dick-Nielsen, J.: Liquidity biases in TRACE. J. Fixed Income **19**(2), 43–55 (2009)
8. Dick-Nielsen, J., Feldütter, P., Lando, D.: Corporate bond liquidity before and after the onset of the subprime crisis. J. Financ. Econ. **103**(3), 471–492 (2012)
9. Edwards, A.K., Harris, L.E., Piwowar, M.S.: Corporate bond market transaction costs and transparency. J. Financ. **62**(3), 1421–1451 (2007)
10. Friewald, N., Jankowitsch, R., Subrahmanyam, M.: Illiquidity or credit deterioration: a study of liquidity in the US corporate bond market during financial crises. J. Financ. Econ. **105**, 18–36 (2012)
11. Holthausen, R.W., Leftwich, R.W.: The effect of bond rating changes on common stock prices. J. Financ. Econ. **17**(1), 57–89 (1986)
12. Jankowitsch, R., Nashikkar, A., Subrahmanyam, M.G.: Price dispersion in OTC markets: a new measure of liquidity. J. Bank. Financ. **35**, 343–357 (2011)
13. Jorion, P., Zhang, G.: Good and bad credit contagion: evidence from credit default swaps. J. Financ. Econ. **84**, 860–883 (2007)
14. Kim, O., Verrecchia, R.E.: Market liquidity and volume around earnings announcements. J. Account. Econ. **17**, 41–67 (1994)
15. May, A.D.: The impact of bond rating changes on corporate bond prices: new evidence from the over-the-counter market. J. Bank. Financ. **34**(11), 2822–2836 (2010)
16. Norden, L., Weber, M.: Informational efficiency of credit default swap and stock markets: the impact of credit rating announcements. J. Bank. Financ. **28**, 2813–2843 (2004)
17. Sheskin, D.J.: Handbook of Parametric and Nonparametric Statistical Procedures. CRC Press, New York (1997)
18. Steiner, M., Heinke, V.G.: Event study concerning international bond price effects of credit rating actions. Int. J. Financ. Econ. **6**, 139–157 (2001)

The Effect of Credit Rating Events on the Emerging CDS Market

Laura Ballester and Ana González-Urteaga

Abstract We document the cross-border spillover impact of S&P sovereign credit rating events on sovereign CDS using an extensive sample of emerging economies. First, we find on average a competition (imitation) effect of downgrades (upgrades) among emerging portfolios. Results confirms that non-event portfolios responds positively to credit deteriorations in terms of an improvement in sovereign credit risk. Second, the sovereign credit risk of non-event countries within the same portfolio benefit (suffer) from downgrades (upgrades). As expected, this implies a competition effect in terms of sovereign credit risk. Moreover, we find that downgrades are more likely to spill over into other emerging markets than upgrades, and they do so with a greater impact. Finally, there is enough evidence of cross-over effects to support the importance of this study.

1 Introduction

During the last decade, sovereign credit ratings and their impact on sovereign debt have received considerable attention, playing a pivotal role especially for emerging market investments, given the expansion of these economies over recent years. The latest literature confirms that sovereign ratings serve the function of enhancing the transparency of the emerging market's credit risk profile and therefore can significantly influence its national stock and bond market investment flows [5]. Kim and Wu [13] hypothesize that rating changes within emerging markets have significant information value to improve institutional quality for facilitating long-run financial and economic development. In short, sovereign ratings represent valuations of governments' capacity to deal with their financial obligations, as well as their capacity to obtain better financial conditions.

L. Ballester
University of Valencia, Avda. Los Naranjos s/n, 46022 Valencia, Spain
e-mail: laura.ballester@uv.es

A. González-Urteaga (✉)
Public University of Navarre, Arrosadia Campus, 31006 Pamplona, Spain
e-mail: ana.gonzalezu@unavarra.es

M. Corazza et al. (eds.), *Mathematical and Statistical Methods for Actuarial Sciences and Finance*, https://doi.org/10.1007/978-3-319-50234-2_2

Nowadays, emerging sovereigns are among the largest high-yield borrowers in the world; however, their nature is different from other high-yield obligors. Since rating agencies usually assign them the non-investment grade status, they are considered to be more likely to default. However, emerging countries in financial distress generally do not enter bankruptcy proceedings or ever liquidate their assets, but rather they go through debt restructuring mechanisms that allow them to exchange defaulted bonds for new longer maturity, lower yield debt instruments.

This paper extends the literature related to the effect of credit rating announcements on emerging markets, providing new analyses untested to date. Initial studies have investigated the reaction of international sovereign and/or corporate CDS spreads to credit events, [8, 9, 11, 15], among others. Overall, they find that the CDS market anticipates ratings announcements, especially for downgrades. However, they all focus exclusively on the effect on the rerated firm or country. A growing strand of the literature focuses on cross-border spillover effects, measuring whether the impact of rating events also extends to economies beyond the respective country. In this line, [7, 10] and [1] examine the cross-border effect of sovereign credit ratings on international sovereign bond spreads, stocks and European Union sovereign bond and CDS spreads, respectively. They all find the existence of asymmetric spillovers, with the negative effect of downgrades being the most pronounced. More recently, [4] confirm previous results, studying the impact of sovereign rating events on the international sovereign bond market. In addition, their results suggest that the effect is more pronounced for countries within the same region. By contrast, at the international CDS corporate level, [17] find the existence of spillover effects on competitors. This means that non-event firms benefit (suffer) from downgrades (upgrades) in a given firm.

In this paper, we focus exclusively on emerging markets, given the significant growth that their credit market has experienced in recent years. Following [17], we argue that an analysis of the reaction on the country where the event occurs is incomplete, because it does not reveal how much of the rating announcement's information is country-specific and how much is market-wide. Given this result, we focus on the cross-border analysis, that allows us to investigate if non-event countries (seen as competitors) benefit or not from the rating event in a given country. We distinguish between positive and negative rating events to test the potential asymmetry of events. Additionally, we also examine the effect in different time windows, differentiating between periods surrounding the event, as well as before and after the event. Moreover, given the advantages of CDS spreads over bond spreads, we use them as a proxy of the sovereign credit risk.

In this sense, this paper is closely related with [12] who investigate the cross-border spillover impact of sovereign credit events on sovereign CDS spreads during the period 2001–2009. However, we differ from their paper in several ways. We have doubled the number of emerging markets analysed, and the sample period includes the more recent period, from 2004 to 2015. Following related literature (e.g., [8, 11, 15, 17]) we use the event study methodology to test a great variety of different spillover analyses. In a first analysis, we test the spillover effect on average through all the countries and events considered in the sample. As a novel contribution of the paper, we provide a better understanding of how the spillover works in the sovereign CDS emerging market. In a second analysis, we study the effect at portfolio level with the aim to examine whether the CDS market reaction to rating announcements are different across representative portfolios. More concretely, at this point, we perform two different analysis. We examine whether a rating event in a given portfolio have significant effects in the other portfolios considered all together. Afterwards, we study whether there is any portfolio that leads the spillover effect, with the purpose of isolating each transmitting portfolio. Finally, we repeat the same steps to analyse the relationships at the country level inside each portfolio. We argue that is seems more likely to find significant effects among the countries belonging to the same portfolio, since they are more likely to be seen as competitors. Our research not only complements but also deepens the literature on international information transmission across emerging countries by examining the impact of sovereign rating changes on the sovereign CDS.

The remaining part of this study is organized as follows. Sections 2 and 3 describe the data and the cross-border estimation methodology, respectively. Section 4 presents our empirical results and discusses their interpretation and we end with a brief conclusion in Sect. 5.

2 Data

We use two major datasets. One consists of daily sovereign CDS spreads, collected from Datastream, for 45 emerging countries. We consider US dollar denominated, senior tier and 5-year CDS quotes, since these contracts are the most liquid and the largest of the segment of the emerging economies' CDS market ([12] and [6], among others). The sample comprises a wide period from January 1, 2004 to March 4, 2015, with 114,587 unbalanced panel observations for 2915 days. Our interest in sovereign emerging markets is twofold. Firstly, since rating agencies usually assign them the non-investment grade status, they are considered more likely to default. However, they do not fall into default in classical terms due to the special nature of their default risk.

The 45 emerging countries[1] have been classified into seven representative portfolios (BRIC, CIVEST, Eastern Europe, Asia, Middle East Asia, America, Africa).[2] Figure 1 illustrates the daily time evolution of the mean CDS spreads

---

[1]We retain all the emerging countries with available data in our sample period. Furthermore, we exclude countries for which no S&P rating history is available.

[2]The first portfolio is the well-known BRIC portfolio, which is comprised by Brazil, China, India and Russia. This is a sub-group of emerging countries with a remarkable strong development over the recent years. Secondly, CIVEST portfolio is constituted by Colombia, Egypt, Indonesia, South Africa, Turkey and Vietnam. These economies are considered very promising and they have been called the new BRICs. The remaining five portfolios are formed by geographical zone. Eastern Europe (portfolio that is made up of 9 countries), Asia (that contains 7 countries), Middle East

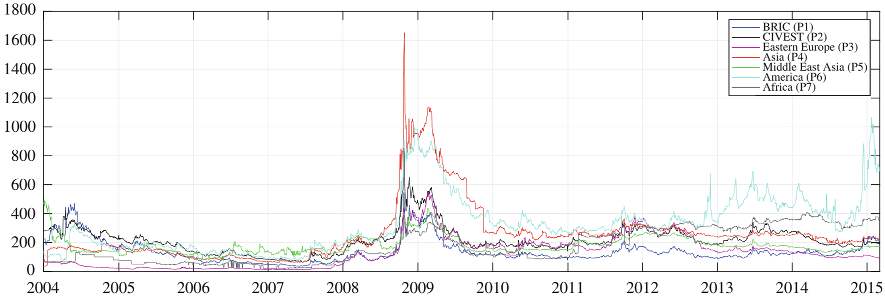

Fig. 1 Emerging portfolios CDS spreads

through all the seven portfolios. CDS spreads differ substantially by country and portfolio.[3] The mean of CDS spreads range from 130.06 bps for Eastern Europe (with Czech Republic presenting the minimum mean of the sample) to 324.01 bps for America (with Argentina and Venezuela displaying the maximum means of the sample). The sharp increase in the CDS premiums during 2008 is notable, corresponding to the global financial crisis. After 2009 they strongly decrease, but still exceed the values they had before the crisis.

Our second data set contains credit rating events that occur for all the emerging countries considered and for the same period as the CDS data. We collect rating announcement events from S&P's Sovereign Rating and Country Transfer and Convertibility Assessment Histories. Previous studies show that S&P rating changes occur more frequently, are less predictable by markets, and antecedent those of other rating agencies [10, 16]. In this study a credit rating event consists of a change in a country's actual rating, a change in its review for a rating change or its entrance onto the watch-list for a possible rating change. Positive (negative) events are upgrades (downgrades) of S&P's credit ratings or revisions in the sovereign's outlook, as well as on the watch-list.

Table 1 displays the distribution of credit rating events per year (Panel A) and per portfolio (Panel B). We observe a total of 373 credit rating announcements for the 45 emerging countries in our sample, where positive and negative rating events

Asia (which encompasses 5 countries), America (that includes 11 countries), and Africa (which is formed of 3 countries). Using CDS spreads portfolios rather than individual country CDS allows us to achieve two objectives. On the one hand, to summarize the information content of the individual CDS series eliminating possible idiosyncratic country effects and, on the other hand, to obtain an average credit risk measure for each one of the portfolios that allow us to obtain a better interpretation of the results [2].

[3]Descriptive statistics on the CDS data for each country and portfolio are not shown but available upon request.

Table 1 The distribution of sovereign credit rating events (CREs)

	Upgrades	Downgrades	Total
Panel A: CREs per year			
2004	26	2	30
2005	39	4	43
2006	30	7	37
2007	22	7	29
2008	12	39	51
2009	6	18	24
2010	18	7	25
2011	19	27	46
2012	8	24	32
2013	10	21	31
2014	6	11	17
2015	1	9	10
Panel B: CREs per portfolio			
BRIC (P1)	29	10	39
CIVEST(P2)	29	26	55
Eastern Europe (P3)	45	53	98
Asia (P4)	24	22	46
Middle East Asia (P5)	16	16	32
America (P6)	51	32	83
Africa (P7)	3	17	20
Total	197	176	373

are slightly asymmetrical, with 199 upgrades in contrast to the 174 downgrades.[4] This is also the case at the portfolio level for CIVEST, Asia and Middle East Asia, which have almost the same number of upgrades as downgrades. In the other four portfolios this relationship is asymmetrical. BRIC and America display more positive credit rating events, in contrast to Eastern Europe and Africa that show more negative credit rating events. Finally, until 2008, positive events clearly dominate negative ones. However, during 2008, in the global financial crisis context, the tendency changes and negative events become most numerous. In addition, 2008 is the year that has more rating events and, concretely, more negative ones. This large number of downgrades in 2008 is directly related to the credit quality of the countries, displayed as a rise in the CDS.

[4]A similar pattern is observed if we look at the type of event, with more positive credit rating changes (110 versus 61) and positive outlooks (97 versus 84). The opposite is given in the credit watch-list, which presents only downgrades (21). These results are not shown but are available upon request.

3 Methodology

We employ the standard event study methodology [8, 11, 14, 15], but we apply it to test the cross-border effects. In particular, the cumulative abnormal returns (CARs) of a country (portfolio) around the credit rating event date of a different country (portfolio) will be tested. The aim is to investigate whether the credit rating announcement in a given emerging country (portfolio) has any impact on the sovereign CDS spreads of cross-border emerging economies.

More concretely, in a first analysis we will test the spillover effect on average through all the countries and events considered in the sample, distinguishing between rating upgrades and downgrades. That way, we will be able to analyse whether the cross-border reaction of sovereign CDS is symmetric to the responses of positive and negative rating news in a given country. Secondly, we will repeat the previous analysis, employing in this case the seven representative portfolios considered, in order to study the effect at the portfolio level. Afterwards, we will study whether there is any portfolio that leads the spillover effect among portfolios, with the purpose of isolating each transmitting portfolio. Finally, we will repeat the same steps to analyse the relationships at the country level inside each portfolio. We argue that it seems more likely to find significant spillovers among the countries belonging to the same portfolio, since they are more likely to be seen as competitors.

The methodology follows a two-stage empirical procedure. The first step consists of calculating at each day t the abnormal return of each CDS series i by applying the following formula:

$$AR_{it} = \Delta CDS_{it} - \Delta Index\, CDS_t$$

where ΔCDS_{it} represents the increment in the credit spread for country or portfolio i at time t, and the $Index\, CDS_t$ is a benchmark that represents the market factor. Following [12], this index is calculated as the average of all the CDS considered in the analysis. Therefore, it consists in measuring the adjusted increment in the CDS spread by taking away the increment in a benchmark CDS spread from the absolute increment in the sovereign CDS spread for country or portfolio i to control for changes in sovereign CDS emerging market conditions.

The second step consists of using the abnormal returns to calculate the CAR, which is given by the following equation:

$$CAR_{i(t_1,t_2)} = \sum_{t=t_1}^{t_2} AR_{it}$$

where (t_1, t_2) is the window where we analyse whether the sovereign credit rating event in a given emerging economy has any impact on the CDS spreads of cross-border emerging economies. Following previous literature, we first consider a window around the event date $[-1, 1]$, where the credit event date is considered the day zero. To analyse the spillover effects prior to and after the occurrence of the event, we consider the windows $[-30, -2]$, $[-60, -31]$ and $[2, 30]$, $[31, 60]$, respectively. To test the absence of effects we use the standard t-test following [3].[5]

A significant prior-effect suggests that before the event occurs in a given country (portfolio), the sovereign CDS of the others incorporate the rating information. On the other hand, a significant post-effect suggests that rating news in a particular country (portfolio) contains new information that has a significant impact on the sovereign CDSs of the other bordering countries (portfolios).

The use of positive and negative credit rating events separately allows us to distinguish two types of effects among countries/portfolios: the competition effect and the imitation effect. If downgrades in a given country/portfolio lead a significant and negative (positive) CAR mean, this indicates a decrease (increase) on average of sovereign CDS increments of the rest of the countries/portfolios, which means an improvement (worsening) in their sovereign credit risk. Hence, the investors see the rest of the countries/portfolios as non-substitute (substitute) assets, and thus there exists a competitive (imitation) effect. Similarly, if upgrades in a given country/portfolio lead a significant and negative (positive) CAR mean, this indicates an improvement (worsening) in the sovereign credit risk of the rest of the countries/portfolios. Hence, the investors see them as substitute (non-substitute) assets, and therefore an imitation (competitive) effect exists.

If we assume that given a rating announcement in a given country/portfolio, the rest of the emerging economies are competitors, we expect that changes in the credit quality would have an impact on refinancing conditions of cross-border economies. More specifically, we expect that competitors benefit (suffer) from downgrades (upgrades) in terms of decreasing (increasing) sovereign credit risk. This signifies a competitive effect for both negative and positive events, which will be reflected in a negative and positive significant CAR, respectively.

4 Results

First of all we analyse the possible spillover effect on average through all the countries and all the events considered in the sample. We observe that there are not significant values in any of the cases. Certainly, there is a notable heterogeneity among the 45 emerging countries considered, hence the absence of cross-over effects between credit rating events and sovereign CDS is not surprising when considering these markets all together, which does not mean that there are not effects among some countries.[6]

[5]As a robustness test we use the non-parametric Wilcoxon signed-rank test. Overall, we observe more significant cross-border spillover effects, but the general conclusions hold. The results are not shown, but are available upon request.

[6]The results are not shown due to their non-significance, but they are available upon request.

Table 2 Credit rating events cross-border effect on emerging sovereign CDS at the portfolio level

	Post-effect		Around-effect	Prior-effect	
	[31, 60]	[2, 30]	[−1, 1]	[−30, −2]	[−60, −31]
Panel A: Downgrades					
Average				−3.87*	−4.26**
Eastern Europe					−4.26**
Asia	−6.51*		−3.89**		−9.12**
America				−11.61***	
Panel B: Upgrades					
Average				−2.13**	
America				−2.42**	

In view of the results it seems more convenient to carry out the study in terms of portfolios. The purpose is to isolate each transmitting portfolio and investigate whether its events have on average spillover effects on the remaining portfolios, the latter considered all together. We repeat the significance test considering the seven CDS spreads emerging portfolios previously constructed. Table 2 exhibits the results for downgrades (Panel A) and upgrades (Panel B).[7] As we expected, at the portfolio level we do find significant spillover effects on average among portfolios. It is notable that significant values (on average) are observed exclusively prior to the rating announcement. The negative CAR values for downgrades and upgrades indicate that non-event portfolios benefit in terms of an improvement in sovereign credit risk. There is an asymmetric response to negative and positive events, with a competitive and imitation effect, respectively.

Next, we study the spillover effects of the rating announcements in a given portfolio on the rest, to examine if any portfolio leads the prior-effect previously found in terms of average, or if there are also other effects depending on the analysed portfolio. In short, the results in Table 2 show that emerging portfolios benefit from sovereign downgrades, in Eastern Europe, Asia and America (prior, post or around the event depending on the portfolio), and upgrades in America (prior to the event). Thus, there is evidence of an asymmetric spillover effect of sovereign credit announcements, with a competition (imitation) effect of downgrades (upgrades) among portfolios. It means that, all the emerging portfolios responds on average positively to credit deteriorations in Eastern Europe, Asia and America portfolios, whereas credit improvements in America also affect positively to the rest of emerging areas. It is noticeable the greater cross-border impact of downgrades, which are not associated (at least at portfolio level) to a negative spillover. This

[7]Henceforth tables present the credit rating events prior, around and post effect on average through all the portfolios and all the events, and from each portfolio to the rest of the portfolios, distinguishing between positive and negative events. The average CAR are shown. The tables only report the windows that are significant, using the standard t-test at the 10%(*), at the 5% (**) or 1% (***).

finding differ from those of [10] and [4] who report that negative events are associated with a significant increase in bond spreads. We attribute the discrepancy to institutional and liquidity differences between CDS and bond markets and to the sample period. Besides, it is quite reasonable to find different spillover relationships between portfolios and between countries.

Finally, we measure the cross-border effect of sovereign rating announcements inside each portfolio at the country level. We argue that it seems more likely to find significant spillovers among the countries belonging to the same portfolio, since they are more likely to be seen as competitors. Thus, we perform an intra-portfolio analysis. Table 3 presents the results.

In general, for negative rating events we observe the same pattern as we found at portfolio level. Downgrades in certain event country cause a decrease in CDS spreads of the others non-event countries included in the same portfolio (negative and significant CAR). As a result, a competition effect is found for rating downgrades. At international industry level, [17] found the same positive spillover between rating changes and the CDS market. This effect is mostly observed around and prior the event occurs. Our analysis allows us to identify which countries inside each portfolios are the leaders in the spillover transmission process regarding this competition effect associated to downgrades. These are the cases of Russia (BRIC portfolio), Egypt and Vietnam (CIVEST portfolio), Pakistan and to a lesser extent Kazakhstan (Asia portfolio), Bahrain and Lebanon (Middle East portfolio) and Ghana (Africa portfolio). The case of the American portfolio is slightly different. Argentina and Venezuela are the only two cases among the eleven American countries presenting a competitive effect due to downgrades. However, the effect is not clear, because it changes depending on the window. In addition to that, there are also some other countries showing the opposite effect (an imitation effect) related to downgrades. The case of Philippines is worth mentioning. The positive CAR indicates negative spillovers that should be closely monitored, since downgrades in Philippines imply a credit deterioration of non-event Asian countries.

Related to upgrades, there are some important results. First, effects in upgrades are less significant than effects in downgrades, since the latter are more frequent and have more impact. Second, despite upgrades displayed an imitation effect between portfolios, at country level we find clear evidence of significant competitive effect due to upgrades. Thus, competitors suffer from upgrades in terms of credit risk. Non-event countries response negatively to positive events in the event countries included in the same portfolio. This finding reflects how international and local contagion (the portfolio versus the intra-portfolio analysis) are due to different spillovers and effects. Moreover, as in the downgrades results, prior and around effects are predominant in the intra-portfolio analysis. Third, the results enables us to identify the transmitters of this competitive effect related to upgrades. Brazil and China (BRIC portfolio), Egypt (CIVEST portfolio), Lithuania (Eastern Europe portfolio) Sri Lanka (Asia portfolio), Lebanon (Middle East portfolio) and Argentina (American portfolio). Credit enhancements in these countries worsen the credit risk of cross-border economies that belong to the same region.

Table 3 Credit rating events cross-border effect on emerging sovereign CDS inside each portfolio at the country level

	Post-effect		Around-effect	Prior-effect	
	[31, 60]	[2, 30]	[−1, 1]	[−30, −2]	[−60, −31]
Panel A: Downgrades					
BRIC-*Average*			−9.56*		−50.44***
Russia				−15.90*	−83.89***
CIVEST-*Average*				−4.52*	−5.40***
Egypt				−7.21**	−8.24***
Vietnam	8.18**		−2.11**		−4.25*
Asia-*Average*					
Kazakhstan		−20.34**			
Pakistan		−48.32***	−33.28***	−57.19**	−43.25***
Philippines			10.16***	1.82***	9.02***
Middle East Asia-*Average*		2.97**	−5.07**	−5.60***	
Bahrain		3.69*	−1.84**	−9.07***	
Lebanon	−12.89*		−11.33*		
America-*Average*			6.34**	−41.31***	
Argentina		18.39***		−62.81***	21.61***
Venezuela			16.63**	−66.58***	
Africa-*Average*					
Ghana	−21.61*				
Panel B: Upgrades					
BRIC-*Average*			1.17*	4.97*	
Brazil			2.13**		6.00**
China				8.38**	
CIVEST-*Average*					
Egypt	14.47**				
Eastern Europe-*Average*					
Lithuania			1.44**		
Asia-Average					
Pakistan				−22.99**	
Sri Lanka					49.48**
Middle East Asia-*Average*				6.28*	
Lebanon				22.50**	
America-*Average*					
Argentina					13.93*

To sum up, our findings provide evidence of asymmetric market reaction around sovereign credit announcements at the portfolio level, in which competitors gain advantage in terms of decreasing their sovereign credit risk from upgrades and downgrades in the event portfolio. This means an imitation (competition) effect for upgrades (downgrades). The market reaction is more pronounced among countries within the same portfolio. Overall, the results reveal that both downgrades and upgrades have a competition effect on non-event countries. Thus, competitors benefit (suffer) from downgrades (upgrades) when they belong to the same portfolio. Accordingly, rating announcements contain information that is both country-specific and market-wide. Setting aside the particular case of upgrades in the portfolio analysis, rating announcements are generally related to a competition effect, supporting the results of [17] for international corporate CDSs. Furthermore, the asymmetry in the results related to positive and negative events are also notable since the latter are more frequent and have more impact.[8]

5 Conclusions

We study the cross-border spillover impact of sovereign credit rating announcements on the sovereign CDS using an extensive sample of emerging economies and covering a large period from 2004 to 2015. Traditionally, the literature has focused on the analysis of the reaction on the country in which the event occurs. However, we argue that this study is incomplete, because it does not reveal how much of the rating announcement's information is country-specific and how much is market-wide. Cross-border analysis allows us to investigate if non-event emerging economies (seen as competitors) benefit or not from the sovereign rating event in a given country.

Our empirical analysis show that, at the portfolio level, both positive and negative rating events positively affect non-event portfolios in terms of decreasing their sovereign credit risk. This implies an imitation (competition) effect for upgrades (downgrades). Nevertheless, among countries within the same portfolio the market reaction is more pronounced. As we expected, rating events are generally related to a competition effect (which supports the results of [17]). In this way, sovereign credit risk of non-events countries within the same portfolio benefit (suffer) from downgrades (upgrades). In this way, we have found sufficient evidence of cross-over effects to support the importance of studying not only the impact of credit rating events on the event country, but also on the non-event cross-border economies through a spillover analysis.

These findings may have practical applications. Investors could evaluate industry models and hedge against the effect of future credit rating announcements in one country on the non-event bordering economies. This information is crucial in order to appropriately construct portfolios sensitive to sovereign credit risk. Moreover, it permits the identification of the competition effect produced by negative and positive rating events in cross-border emerging economies. Additionally, given the

[8]This result is consistent with previous related literature ([1, 7, 10], among others). By contrast, [12] find that upgrades are more likely to spill over to the emerging economies. At this point, it is important to take into account that they results are obtained with a more limited number of countries, an smaller sample period and a different methodology.

importance and the growth in the CDS market, which is considered a reasonable proxy for credit risk, these results may also be helpful for future regulators when implementing new capital adequacy frameworks for individual countries and portfolios in the sovereign credit risk market.

Acknowledgements The authors would like to express their gratitude for the grant received from the Fundación Ramón Areces. A. González-Urteaga acknowledges financial support from ECO2015-67035-P and ECO2016-77631-R (AEI/FEDER,UE). We thank the participants at the Mathematical and Statistical Methods for Actuarial Sciences and Finance (MAF) Conference for stimulating discussions on earlier drafts of this paper.

References

1. Afonso, A., Furceri, D., Gomes, P.: Sovereign credit ratings and financial markets linkages: application to European data. J. Int. Money Financ. **31**, 606–638 (2012)
2. Ballester, L., Casu, B., González-Urteaga, A.: Bank fragility and contagion: evidence from the CDS market. J. Empir. Financ. **38**, 394–416 (2016)
3. Boehmer, E., Masumeci, J., Poulsen, A.B.: Event-study methodology under conditions of event-induced variance. J. Financ. Econ. **30**, 253–272 (1991)
4. Böninghausen, B., Zabel, M.: Credit ratings and cross-border bond market spillovers. J. Int. Money Financ. **53**, 115–136 (2015)
5. Christopher, R., Kim, S.J., Wu, E.: Do sovereign credit ratings influence regional stock and bond market interdependencies in emerging countries? J. Int. Financ. Mark. **22**, 1070–1089 (2012)
6. Eichengreen, B., Mody, A., Nedeljkovic, M., Sarno, L.: How the subprime crisis went global: evidence from bank credit default swap spreads. J. Int. Money Financ. **31**, 1299–1318 (2012)
7. Ferreira, M.A., Gama, P.M.: Does sovereign debt ratings news spill over to international stock markets? J. Bank. Financ. **31**, 3162–3182 (2007)
8. Finnerty, J.D., Miller, C.D., Chen, R.R.: The impact of credit rating announcements on credit default swap spreads. J. Bank. Financ. **37**, 2011–2030 (2013)
9. Galil, K., Soffer, G.: Good news, bad news and rating announcements: an empirical investigation. J. Bank. Financ. **35**, 3101–3119 (2011)
10. Gande, A., Parsley, D.C.: News spillovers in the sovereign debt market. J. Financ. Econ. **75**, 691–734 (2005)
11. Hull, J., Predescu, M., White, A.: The relationship between credit default swap spreads, bond yields, and credit rating announcements. J. Bank. Financ. **28**, 2789–2811 (2004)
12. Ismailescu, I., Kazemi, H.: The reaction of emerging market credit default swap spreads to sovereign credit rating changes. J. Bank. Financ. **34**, 2861–2873 (2010)
13. Kim, S.J., Wu, E.: Sovereign credit ratings, capital flows and financial sector development in emerging markets. Emerg. Mark. Rev. **9**, 17–39 (2008)
14. Kothari, S.P., Warner, J.B.: Measuring long-horizon security price performance. J. Financ. Econ. **43**, 301–339 (1997)
15. Norden, L., Weber, M.: Informational efficiency of credit default swaps and stock markets: the impact of credit rating announcements. J. Bank. Financ. **28**, 2813–2843 (2004)
16. Reisen, H., Von Maltzan, J.: Boom and bust and sovereign ratings. Int. Financ. **2**, 273–293 (1999)
17. Wengner, A., Burghof, H.P., Schneider, J.: The impact of credit rating announcements on corporate CDS markets-Are intra-industry effects observable? J. Econ. Bus. **78**, 79–91 (2015)

A Generalised Linear Model Approach to Predict the Result of Research Evaluation

Antonella Basso and Giacomo di Tollo

Abstract Peer review is still used as the main tool for research evaluation, but its costly and time-consuming nature triggers a debate about the necessity to use, alternatively or jointly with it, bibliometric indicators. In this contribution we introduce an approach based on generalised linear models that jointly uses former peer-review and bibliometric indicators to predict the outcome of UK's *Research Excellence Framework* (REF) 2014. We use the outcomes of the *Research Assessment Exercise* (RAE) 2008 as peer-review indicators and the departmental *h*-indices for the period 2008–2014 as bibliometric indicators. The results show that a joint use of bibliometric and peer-review indicators can be an effective tool to predict the research evaluation made by REF.

1 Introduction

Peer-review is often considered as the most reliable tool to evaluate the research output: it relies on experts assessments, and it is based on independent examination performed by scholars [2, 8]. Though it presents some drawbacks [4], it has extensively been used by national agencies to assess the research performed by Universities. Every country has its own rules to provide this assessment, but they mostly rely on peer-review. Hence, the overall process takes several months in order to produce the eventual assessment, and the period between two consecutive assessments is significant.

The assessment can be aimed either to compute a *score*, which is a single number s that represents an overall measure of the quality of the group [11], or to assign Universities (or departments) a rating class [7].

A. Basso • G. di Tollo (✉)
Department of Economics, Ca' Foscari University of Venice, Cannaregio 873, 30121 Venice, Italy
e-mail: basso@unive.it; giacomo.ditollo@unive.it

M. Corazza et al. (eds.), *Mathematical and Statistical Methods for Actuarial Sciences and Finance*, https://doi.org/10.1007/978-3-319-50234-2_3

In both cases, scholars have started to investigate whether faster automated scientometric or bibliometric analysis may be used to replace or to predict peer-review based evaluations [15]: a timely prediction could lead universities to better allocate their resources and to develop operational strategies in order to improve their results. However, there is evidence against using bibliometrics as stand-alone indicators for several reasons: they come from different databases, and their use may lead to different results when comparing researchers in different areas; furthermore citations accumulate too slowly to be used in a short-term research assessment [5, 9, 10]. Hence, some scholars have suggested to use bibliometrics jointly with peer-review [6, 8] in order to perform research evaluation: Mrygold et al. [11] show that bibliometrics may be used as a proxy for peer-review analysis, but not as a measure of research quality. In later works, they find a significant correlation between the departmental *h*-indices and the outcome of the 2008 Research Assessment Exercise (RAE) in the UK, and propose the *h*-index as a predictor for the outcome in the next research assessment [12]. However, this approach failed to anticipate the outcome of the 2014 Research Excellence Framework (REF) [13].

In our work we try to predict the outcome of the next national research assessment exercise by using an automated approach that resorts to bibliometric indicators and the output of the previous national research assessment. We apply our approach to the Higher Education System in the United Kingdom, aiming to predict the *Research Excellence Framework* (REF) 2014, i.e., the score computed by the HEFCE (*Higher Education Funding Council for England*) and the rating class for each department. As previous assessment indicators, we use the outcome of the previous research assessment; as for bibliometric indicators we use the departmental *h*-indices (a departmental *h*-index of *n* means that *n* papers, authored by staff from a given department, and in a given subject area, were cited *n* times or more in a given time period).

With regard to the automated approach, we use a Generalised Linear Model (GLM), that could be useful since GLMs represent an extension of ordinary linear regression models which introduces a link function to create a relationship between a linear prediction and the model output.

The paper is organised as follows: Sect. 2 introduces GLMs; Sect. 3 presents the models proposed to predict the results of the research evaluation in UK; Sect. 4 details the experimental phase to fit the model, that will be used in Sect. 5 to predict the outcome of the next research assessment. Section 6 concludes the paper.

2 Two Generalised Linear Models for Research Evaluation

Generalised Linear Models [14] are a class of regression models in which the variable of interest y is assumed to follow a member of the exponential family of probability distributions (Gaussian, inverse Gaussian, Poisson, Gamma, Binomial ...), i.e., a distribution whose density function can be reconducted to

$$f_y(y;\theta,\phi) = \exp\left(\frac{y\theta - b(\theta)}{a(\phi)} + c(y,\phi)\right)$$

where θ is the canonical parameter and ϕ is the scale parameter.

GLMs relate the parameter θ to a set of predictors by defining a differentiable one-to-one *link function* $g(\mu)$ that transforms the mean of the response variable $\mu = E(y)$ into the linear combination of predictors $\eta = \sum_i \beta_i x_i$. In this way we allow a linear model to be related to the response variable.

For example, in the *Gamma* family the random component that specifies the conditional distribution of the variable of interest follows a Gamma distribution, that can be expressed in exponential form as follows:

$$f(y;\mu,\phi) = \exp\left(\frac{\frac{y}{\mu} - (-\ln\mu)}{-\phi} + \frac{1-\phi}{\phi}\ln(y) - \frac{ln(\phi)}{\phi} - \ln\Gamma(\frac{1}{\phi})\right)$$

In this case the link function is given by $g(\mu) = \frac{1}{\mu}$ so that $\eta = \frac{1}{\mu}$. The resulting model will be referred to as the *Gamma model* in what follows.

In the *Gaussian* family, instead, the random component that specifies the conditional distribution of the variable of interest follows a Gaussian distribution with mean μ and variance σ^2, with density function

$$f(y;\mu,\phi) = \exp\left(\frac{y\mu - \frac{\mu^2}{2}}{\sigma^2} - \frac{y^2}{2\sigma^2} - \frac{1}{2}ln(2\pi\sigma^2)\right)$$

In this case the link function is given by $g(\mu) = \mu$ and thus $\eta = \mu$. Please notice that in this formulation, that will be referred to as the *Gaussian model* in what follows, the GLM corresponds to the standard linear model, which can therefore be considered as a special case of GLMs.

The Gamma distributions are well-suited when dealing with strictly positive continuously distributed data; furthermore, differently from other distributions, they are easier to interpret [1]. On the other hand, the Gaussian distributions are well-known and widespread, and their use is well suited for dealing with unbounded real values.

In this contribution we use GLMs to study the relationship between the predictors and the *score* computed by HEFCE, the mean rating class and the mode of the rating class. Both in the case of the score and in that of the mean rating class we deal with bounded continuous random variables; the mode of the rating class, instead, is a bounded integer random variable (see Sect. 3).

3 Using GLMs to Predict the Results of REF Evaluation

The research assessment procedures devised to allocate public research funds in the UK divide academic disciplines into units of assessment (*UoA*s: sociology, biology etc.). Universities are invited to submit research outputs produced on these *UoA*s to be evaluated by expert panels: every institution has to submit four outputs for each staff member, that have to be evaluated w.r.t. originality, significance and rigour. Panel experts assess department's submissions by assigning each research product to one out of five rating class (4*, 3*, 2*, 1* and *Unclassified*, in decreasing order of relevance), and quantifying the percent value that falls into each rating class. Starting from these percent values, HEFCE computes a score for each university by averaging out these percent values. This is made by using a formula that has been modified over time. For the REF outcomes (2014) the formula used was

$$s_{REF} = p_{4,REF} + \frac{1}{3} p_{3,REF} \tag{1}$$

where $p_{i,RAE}$ $(i \in \{1 \dots 4\})$ denotes the percent value of the outputs attributed to the rating class i^*. In order to compare the results obtained in the 2008 and 2014 evaluation exercises, we have applied Eq. (1) not only to the results (percent values) of the REF 2014 exercise but also to the results of the RAE 2008 evaluation exercise, thus obtaining two comparable scores, denoted by s_{RAE} and s_{REF}, respectively. Along with the results from RAE and REF, for the period taken into account we have considered the departmental h-index of the universities and HEIs in the sample.

Our goal is to define an appropriate relationship between the set of predictors (h-indices, $p^*_{i,RAE}$ and s_{RAE}) and:

- the rating class c_{REF} to which a department/university belongs to;
- the score s_{REF}.

As for the rating class, we either assign every department to the class showing the maximum output research percentage (referred to as $c_{REF,mode}$ [3]), or we consider the mean value of the rating class which is obtained by computing the weighted average of the class (referred to as $c_{REF,mean}$), with the weights given by the percentage of research outputs in each class.

In order to reduce the number of predictors, necessary when using small-samples, we remark that empirical investigations led some authors to report that departmental h-indices tend to follow a linearly increasing trend over time [12]. Hence, we have computed the mean yearly h-index variation over the time period considered, i.e. $\Delta h = \frac{h_{2014} - h_{2008}}{6}$, and we have focused on it as a predictor to be used in conjunction with the initial h-value (h_{2008}) (experiments with the final h-value h_{2014} led to the same results).

As for the $p_{i,RAE}$ values, in order to avoid multicollinearity in the data we have removed the *Unclassified* and 1^*_{RAE} classes from the predictors set: as a matter of fact, not only the percent values in the 5 classes sum to 100 for all HEIs, but also in most cases the *Unclassified* class has a zero percent value. Furthermore, when predicting s_{REF} we have removed the 2^*_{RAE} class, too, since Eq. (1) takes into account only the classes 4^*_{REF} and 3^*_{REF} in the computation of s_{REF}.

In a nutshell, for predicting $c_{REF,mode}$ and $c_{REF,mean}$ we use the following predictors: h-2008, Δh, $p_{4,RAE}$, $p_{3,RAE}$, $p_{2,RAE}$. The linear predictor η is defined as

$$\eta_{i,c} = \alpha + \beta_{h2008} h_{2008i} + \beta_{\Delta h} \Delta h_i + \beta_{4^*_{RAE}} p_{4,RAE} + \beta_{3^*_{RAE}} p_{3,RAE} + \beta_{2^*_{RAE}} p_{2,RAE}. \quad (2)$$

For predicting s_{REF} we use the following predictors: h-2008, Δh, $p_{4,RAE}$, $p_{3,RAE}$, and the linear predictor η is defined as:

$$\eta_{i,s} = \alpha + \beta_{h2008} h_{2008i} + \beta_{\Delta h} \Delta h_i + \beta_{4^*_{RAE}} p_{4,RAE} + \beta_{3^*_{RAE}} p_{3,RAE}. \quad (3)$$

These predictors are used to estimate a GLM. The resulting model is then applied to predict s_{REF}, $c_{REF,mode}$ and $c_{REF,mean}$. We have tested several GLMs in order to determine which one is the most suitable for the problem at hand. In Sect. 4 we report the outcomes obtained with the models that produced the best results: the *Gamma* model, and the *Gaussian* model.

The other distributions tried show sometimes good results on specific instances but lack generality: this is the case for the *ordinal logistic* distribution, that is well suited for predicting the rating class, but needs non null values in at least three classes for the dependent variable (and this is not the case in some instances), and the binomial GLM with the *logit* link function, which considers only two possible values for the dependent variable (and this occurs in some instances but is far from being the general case). On the contrary, the models we have chosen are apt to be applied to a wide range of instances and generally provide a good performance. The results of the experiments performed by using these models will be presented in the next sections.

4 An Empirical Investigation on REF Data

We have run experiments on four *UoA*s which were included in the list of UoAs for both RAE and REF: biology, sociology, chemistry and physics.[1] The data are publicly available. To this aim we have applied the two GLM models proposed in the previous section.

For the computations we have used the R package *glm*, that allows the user to estimate the aforementioned models with the most common link functions. From an operational point of view, the estimations were computed on a cluster with AMD Opteron 2216 dual core CPUs running at 2.4 GHz with 2×1 MB L2 cache and 4 GB of RAM under Cluster Rocks distribution built on CentOS 5.3 Linux. The average execution time was comprised between 2 and 3 s for each run.

[1] HEIs with missing values have been removed from the data set.

In what follows, we discuss the experiments performed to estimate the GLMs presented in Sect. 3; these models will be used in Sect. 5 to make some predictions for the next REF, that will be held in 2020.

First of all, we have estimated the GLM models presented in Sect. 3 on the data considered.

As for s_{REF}, it represents the score assigned to each HEI, hence no further preprocessing is necessary. As for c_{REF}, instead, it represents the *rating class*, and we have performed some experiments with two different definitions that can be used to obtain its value:

- the mean rating class $c_{REF,mean}$, obtained by computing the weighted average of the class with the weights given by the percentage of research outputs in each class: $\sum_{i=1}^{4} i^{*}_{REF} p_{i,REF}$;
- the sample mode $c_{REF,mode}$ of the percent distribution of the research outputs in the rating classes 4^{*}_{REF}, 3^{*}_{REF}, 2^{*}_{REF}, 1^{*}_{REF}, i.e. the class showing the maximum research percentage [3]; it will be referred to as $c_{REF,mode}$.

In Table 1 we report the results of the parameter estimation for the models obtained for the score s_{REF}, while Tables 2 and 3 display the results of the parameter estimation for the models obtained for the rating class $c_{REF,mean}$ and $c_{REF,mode}$, respectively. For each model, Tables 1, 2, and 3 report the parameter estimates obtained, along with their p-values, and three statistics useful to assess the regression significance: R^2, the p-value of the F-test of the regression (the probability to obtain a value for the F statistic greater than the F-value of the model, under the null hypothesis that the regression model is not significant), and the Akaike Information Criterion (referred to as AIC: a measure for choosing between competing statistical models by assessing the information lost in the model).

By observing Tables 1, 2, and 3, we may note that our approach shows quite good performances, with a good significance of the estimated models, though in several cases the estimates of some parameters show little significance and the R^2 statistics varies considerably according to the sample considered.

With regard to the specific form of the GLM model, a comparison of the results of the *Gamma* and *Gaussian* models does not allow to determine which is the best model conclusively, since neither of the two models always offers the best results: this can be seen by analysing the three statistics reported. At first, we remark that AIC values are highly (negatively) correlated with R^2 over s_{REF} and $c_{REF,mean}$, and by observing these two measures we remark that in some cases the best estimation is given by the *Gamma* model, in other cases it is given by the Gaussian model. However, the overall results seem to indicate that the *Gamma* model is more robust than the *Gaussian* model, since by observing the p-value of the F-test we remark that the estimated models of Tables 1, 2, and 3 are highly significant in all cases for the *Gamma* model (the p-value of the F statistics is always lower than 0.001 but in one case, in which it is anyway lower than 0.01) while in some cases (5 out of 24) they are not significant at a 5% level for the *Gaussian* model.

Table 1 Generalised linear models for s_{REF}: estimated parameters, *p*-values and significance statistics for the *Gamma* and *Gaussian* models

	Gamma model				Gaussian model			
UoA	Biology	Chemistry	Physics	Sociology	Biology	Chemistry	Physics	Sociology
Sample size	29	8	33	21	29	8	33	21
Parameter values								
Intercept	0.028	0.038	0.030	0.029	31.488	10.321	27.058	31.795
h-2008	−9.33E−05	−4.99E−05	2.68E−05	−0.0004	0.211	−0.001	−0.042	1.015
Δh	0.001	−0.002	−0.001	0.0005	−4.710	5.615	2.487	−1.450
4^*_{RAE}	−0.0002	0.0004	3.62E−05	−0.0001	0.672	−1.048	−0.107	0.264
3^*_{RAE}	1.54E−05	−0.0001	−0.0001	−9.46E−06	−0.113	0.372	0.226	−0.087
Parameter p-values								
Intercept	1.63E−05	0.010	0.0001	4.73E−06	0.012	0.573	0.069	0.005
h-2008	0.023	0.938	0.793	0.029	0.046	0.999	0.838	0.046
Δh	0.354	0.550	0.132	0.474	0.276	0.525	0.101	0.508
4^*_{RAE}	0.033	0.308	0.802	0.126	0.016	0.361	0.712	0.194
3^*_{RAE}	0.912	0.396	0.556	0.935	0.709	0.511	0.534	0.755
R^2	0.523	0.809	0.207	0.634	0.588	0.771	0.267	0.619
p-value of the F-test	<0.001	<0.001	<0.001	<0.001	<0.001	0.235	0.060	0.002
AIC	213.8	57.89	243.1	141.9	209.3	59.74	235.1	142.7

Table 2 Generalised linear models for $c_{REF,mean}$: estimated parameters, *p*-values and significance statistics for the *Gamma* and *Gaussian* models

	Gamma model				Gaussian model			
UoA	Biology	Chemistry	Physics	Sociology	Biology	Chemistry	Physics	Sociology
Sample size	29	8	33	21	29	8	33	21
Parameter values								
Intercept	0.342	0.389	0.415	0.464	2.954	2.584	2.187	1.917
h-2008	0	0	0	−0.003	0.003	0.002	−0.001	0.024
Δh	0.010	−0.009	−0.005	0.003	−0.105	0.094	0.049	−0.028
4^*_{RAE}	−0.001	0.002	0.000	−0.001	0.012	−0.019	0.002	0.011
3^*_{RAE}	0	−0.001	−0.001	−0.001	−0.003	0.005	0.011	0.004
2^*_{RAE}	0	0	−0.001	−0.001	−0.001	0	0.006	0.009
Parameter p-values								
Intercept	0	0.182	0	0	0.001	0.330	0.007	0.003
h-2008	0.091	0.958	0.871	0.040	0.115	0.957	0.878	0.049
Δh	0.229	0.597	0.134	0.583	0.211	0.598	0.120	0.600
4^*_{RAE}	0.222	0.595	0.831	0.083	0.185	0.581	0.838	0.104
3^*_{RAE}	0.749	0.808	0.291	0.512	0.676	0.870	0.278	0.604
2^*_{RAE}	0.949	0.934	0.494	0.319	0.919	0.989	0.479	0.326
R^2	0.558	0.795	0.242	0.679	0.573	0.782	0.259	0.670
p-value of the F-test	<0.001	0.048	<0.001	<0.001	0.001	0.459	0.130	<0.001
AIC	−20.06	−5.89	−16.83	−13.54	−21.03	−4.835	−19.23	−13.16

Table 3 Generalised linear models for $c_{REF,mode}$: estimated parameters, *p*-values and significance statistics for the *Gamma* and *Gaussian* models

	Gamma model				Gaussian model			
UoA	Biology	Chemistry	Physics	Sociology	Biology	Chemistry	Physics	Sociology
Sample size	29	8	33	21	29	8	33	21
Parameter values								
Intercept	1.086	1.755	0.074	−0.210	−1.916	−5.593	4.831	7.254
h-2008	−0.002	−0.009	−5.93E−04	0.001	0.019	0.055	0.004	−0.020
Δh	0.013	0.045	−0.001	−0.023	−0.100	−0.273	0.009	0.203
4^*_{RAE}	−0.005	−0.013	0.002	0.004	0.031	0.080	−0.021	−0.037
3^*_{RAE}	−0.009	−0.013	8.38E−05	8.10E−04	0.064	0.083	−4.45E−04	−0.014
2^*_{RAE}	−0.007	−0.015	0.007	0.017	0.042	0.093	−0.052	−0.120
Parameter p-values								
Intercept	4.99E−05	0.007	0.688	0.590	0.157	0.016	0.001	0.010
h-2008	0.092	0.042	0.594	0.866	0.081	0.038	0.574	0.695
Δh	0.195	0.042	0.869	0.502	0.169	0.037	0.849	0.407
4^*_{RAE}	0.041	0.021	0.354	0.332	0.049	0.017	0.281	0.220
3^*_{RAE}	3.46E−04	0.014	0.970	0.888	9.39E−05	0.011	0.978	0.668
2^*_{RAE}	0.031	0.013	0.004	0.010	0.039	0.009	0.003	0.010
R^2	0.585	0.988	0.617	0.434	0.656	0.993	0.662	0.456
p-value of the F-test	<0.001	0.004	<0.001	<0.001	<0.001	0.016	<0.001	0.075
AIC	23.33	11.78	28.29	58.66	12.81	8.345	19.06	57.18

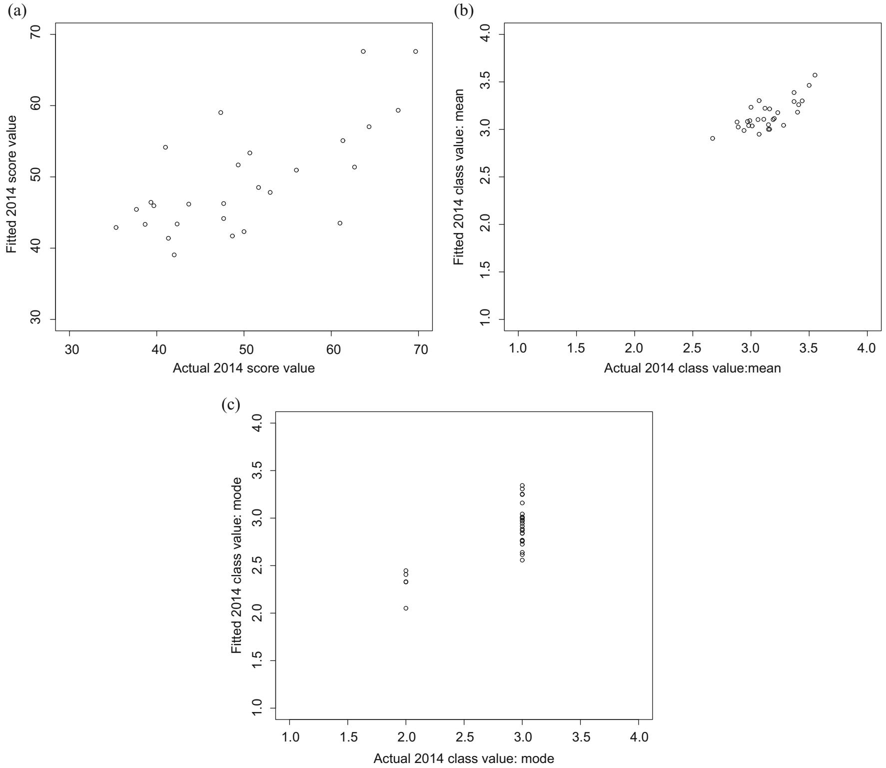

Fig. 1 Results of the experiments with the *Gamma* model: values of s_{REF}, $c_{REF,mean}$, and $c_{REF,mode}$ vs their fitted values for the UoA *Biology*. (**a**) UoA *Biology*: s_{REF} vs fitted s_{REF}. (**b**) UoA *Biology*: $c_{REF,mean}$ vs fitted $c_{REF,mean}$. (**c**) UoA *Biology*: $c_{REF,mode}$ vs fitted $c_{REF,mode}$

The adequate goodness of fit of the models is confirmed by the plot of the actual versus the fitted values; Fig. 1 shows the actual versus the fitted values obtained with the *Gamma* model for s_{REF} (a), $c_{REF,mean}$ (b), and $c_{REF,mode}$ (c), for the UoA *Biology* (the graphs of the other UoAs are similar).

5 Predicting the Results of Next REF

In this section we use the models estimated in Sect. 4 to predict the outcome of the next REF, that will be held in 2020. The basic idea is to use $p_{i,REF}$ $(i \in \{1 \dots 4\})$ instead of $p_{i,RAE}$ $(i \in \{1 \dots 4\})$ and *h*-2014 instead of *h*-2008, by weighting them in the model with the values of the parameters estimated in Sect. 4.

However, since some of the variables used in the estimated models refer to a future period of time (2014–2020), we have to determine an estimate for their (future) value.

Moreover, some pre-processing operations are necessary to take into account the fact that the h-index is cumulative, hence it follows an increasing trend over time [12]. This means that if the same HEI shows a h-2014 value bigger than that of h-2008, this is only partly attributable to the research carried out in the last 6 years. If it is not properly taken into account, this effect could lead to incoherent results when these data are used to predict the next research evaluation outcome.

This issue does not affect the variable Δh: since h-indices on average generally follow an increasing linear trend over time [12], we may use the same value of Δh used to estimate the model also as a prediction of the variable over the next period, without introducing a bias in the results. Clearly, this corresponds to assume that the number and the quality of the research outputs in the next 6-year period will be analogous to those of the previous 6 years. As for $p_{i,REF}$, they are expressed in percent values, hence their use in place of $p_{i,RAE}$ appears to be justified, too.

On the other hand, if we want to use the value of h-2014 in our models in place of the value of h-2008, we need a way to neutralise the average cumulative effect of time. There could be different ways to obtain this effect; we chose to subtract the average increase observed in the 6-year period 2008–2014 (the average being computed over all the HEIs of the *UoA* under examination) to the value of h-2014. Hence, we subtract the quantity $6 \cdot \overline{\Delta h}$, where $\overline{\Delta h}$ represents the mean value of Δh, to h-2014; the resulting normalised variable will be denoted by h^*-2014:

$$h^*_{2014} = h_{2014} - 6 \cdot \overline{\Delta h}.$$

In a nutshell, in order to predict the results of the assessment of the next 2020 REF we have applied both the *Gamma* and the *Gaussian* models estimated in the previous section with the following input variables: h^*-2014, Δh, and $p_{i,REF}$ $(i \in \{1 \ldots 4\})$.

Although it is difficult to assess the significance of such a prediction, it is interesting to compare the results provided by the different models. To this aim, in Table 4 we report the results of a correlation analysis that we have performed for s_{REF}, $c_{REF,mean}$ and $c_{REF,mode}$ between:

1. the actual REF 2014 values and the REF 2020 values predicted with the *Gamma* model (labelled with the "*hat*" symbol);
2. the actual REF 2014 values and the REF 2020 values predicted with the *Gaussian* model (labelled with the "*hat*" symbol);
3. the REF 2020 predictions obtained with the *Gamma* and the *Gaussian* models.

The correlations reported in Table 4 are measured by the Pearson correlation coefficient (denoted by ρ) for s_{REF} and by the rank-based Spearman correlation coefficient (denoted by r_s) for $c_{REF,mean}$ and $c_{REF,mode}$. Moreover, the rank correlation between $c_{REF,mean}$ and $c_{REF,mode}$ is also reported.

We may notice the high correlation values between $\hat{s}_{REF,gamma}$ and $\hat{s}_{REF,gauss}$, i.e. the estimates obtained for s_{REF} with the Gamma and Gaussian model, for 3 out of 4 *UoAs*, showing that the two models often lead to highly correlated predictions. A similar remark can be made also for the prediction of $c_{REF,mode}$ and $c_{REF,mean}$

Table 4 Correlation analysis of the predictions obtained with the models proposed

	Biology	Chemistry	Physics	Sociology
$\rho(s_{REF}, \hat{s}_{REF,gamma})$	0.87	−0.58	0.29	0.21
$\rho(s_{REF} / \hat{s}_{REF,gauss})$	0.95	−0.72	0.28	0.80
$\rho(\hat{s}_{REF,gamma}, \hat{s}_{REF,gauss})$	0.84	0.96	0.99	0.17
$r_s(c_{REF,mean}, \hat{c}_{REF,mean-gamma})$	0.89	−0.51	0.14	0.79
$r_s(c_{REF,mean}, \hat{c}_{REF,mean-gauss})$	0.90	−0.51	0.14	0.76
$r_s(\hat{c}_{REF,mean-gamma}, \hat{c}_{REF,mean-gauss})$	1	1	0.99	0.99
$r_s(c_{REF,mode}, \hat{c}_{REF,mode-gamma})$	0.33	−0.41	−0.01	−0.07
$r_s(c_{REF,mode}, \hat{c}_{REF,mode-gauss})$	0.34	0.58	0.01	−0.06
$r_s(\hat{c}_{REF,mode-gamma}, \hat{c}_{REF,mode-gauss})$	0.99	−0.14	0.98	0.91
$r_s(c_{REF,mode}, c_{REF,mean})$	0.32	0.50	−0.01	−0.05

obtained with the two models. On the contrary, the rank correlation between the two ways to define the ranking class, $r_s(c_{REF,mode}, c_{REF,mean})$, is very low, denoting that the information contained in the two aggregated variables is different. Hence, it makes little sense to compare the *mean* and *mode* results.

As for the correlation between the actual REF 2014 values and the REF 2020 predicted values, their value varies very much between the *UoAs*; in many cases, however, the predictions of the score (s_{REF}) seem a little closer to the actual 2014 values than those of the ranking class, suggesting that it may be easier for the HEIs to change their relative rating class over consecutive exercises, rather than change substantially their relative score. Assessing the ranking class by the *mean* reduces this phenomenon, but no clear conclusions can be drawn.

6 Concluding Remarks

The costly and time-consuming nature of peer-review procedures for research evaluation has led to an ongoing debate about the use of bibliometric indicators, either as alternative procedures or as complementary tools. Evaluations based on bibliometrics indicators are thought to be possible alternative candidates, but they also present shortcomings such as different citation magnitude over various disciplines, self-citing, bias, and technical errors related to data retrieval.

In this paper we have proposed a quantitative approach based on generalised linear models to predict the result of the national Research Excellence Framework (REF) in the UK.

We have used as predictors both a bibliometric indicator and the result of the previous national assessment procedure (RAE), performing experiments to predict both the overall score assigned to a university/department and its rating class.

The results show that generalised linear models provide a flexible tool that can be used in order to combine the information contained in both peer-review procedures and bibliometric analysis for predicting the outcomes of the research assessment.

Future activities, aimed to improve the model, include a deeper investigation of the dynamics of the departmental h-index. The idea, left for future research, is to try to model the dynamics of the departmental *h*-index in order to improve both the explanatory power of the model and its predictions.

Furthermore, it might be interesting to see if this model can be used also to investigate and predict the results of the research evaluation carried out in countries different from the UK. For example, the results of the Italian *eValuation of the Quality of Research VQR 2011–2014* have recently been made public, and it can be interesting to see if a similar model applies also to Italy.

References

1. Abramowitz, M., Stegun, I.A.: Gamma and related functions (Chapter 6). In: Handbook of Mathematical Functions. Dover, New York (1972)
2. Bertocchi, G., Gambardella, A., Jappelli, T., Nappi, C.A., Peracchi, F.: Bibliometric evaluation vs. informed peer-review: evidence from Italy. Res. Policy **44**(2), 451–466 (2011)
3. Bezdek, J.C., Ehrlich, R., Full, W.: FCM: the fuzzy c-means clustering algorithm. Comput. Geol. **10**(2–3), 191–203 (1984)
4. Bornmann, L.: The hawthorne effect in journal peer review. Scientometrics **91**, 857–862 (2012)
5. Bornmann, L., Leydesdorff, I.: Scientometrics in a changing research landscape. EMBO Rep. **15**(12), 1228–1232 (2014)
6. De Bellis, N.: Bibliometrics and Citation Analysis: From the Science Citation Index to Cybermetrics. The Scarecrow Press, Lanham (2009)
7. Engels, T.C.E., Goos, P., Dexters, N., Spruyt, E.H.J.: Group size, h-index, and efficiency in publishing in top journals explain expert panel assessments of research group quality and productivity. Res. Eval. **22**(4), 224–236 (2013)
8. Evidence: The future of the UK university research base. Technical report, a Thomson Reuters business, Universities UK (2010)
9. Lawrance, P.: The politics of publications. Nature **422**, 259–261 (2003)
10. MacRoberts, M.H., MacRoberts, B.R.: Problems of citation analysis: a critical review. J. Am. Soc. Inf. Sci. **40**(5), 342–349 (1989)
11. Mryglod, O., Kenna, R., Holovatch, Y., Berche, B.: Absolute and specific measures of research group excellence. Scientometrics **95**(1), 115–127 (2013)
12. Mryglod, O., Kenna, R., Holovatch, Yu., Berche, B.: Predicting results of the research excellence framework using departmental h-index. Scientometrics **102**(3), 2165–2180 (2015)
13. Mryglod, O., Kenna, R., Holovatch, Yu., Berche, B.: Predicting results of the research excellence framework using departmental h-index: revisited. Scientometrics **104**(3), 1013–1017 (2015)
14. Nelder, J., Wedderburn, R.: Generalized linear models. J. R. Stat. Soc. Ser. A **135**(3), 370–384 (1972)
15. Wouters, P., Thelwall, M., Kousha, K., Waltman, L., de Rijcke, S., Rushforth, A., Franssen, T.: The metric tide: literature review. Technical report, Higher Education Funding Council for England (2015)

Projecting Dynamic Life Tables Using Data Cloning

Andrés Benchimol, Irene Albarrán, Juan Miguel Marín, and Pablo Alonso-González

Abstract In this paper we introduce a hierarchical Lee-Carter model (LC) specification to forecast the death rates of a set of demographically related countries. We assume that the latent mortality factor of LC is common for all of them, given the linkage among them. On the other hand, hierarchical modeling is usually conducted by Bayesian approach, which has the disadvantage that assumptions on the prior distributions are needed, which are not usually known or obtainable, introducing thus subjectivity in the model when setting these prior distributions. An option to overcome this limitation is provided by Data Cloning, a novel technique raised in the Ecology field that allows approximating maximum likelihood estimates in hierarchical settings. Even though this technique works with MCMC algorithms, it constitutes a frequentist approach, and the results are invariant to the prior distributions. Finally, we apply the methodology to a set of linked countries, getting a very satisfactory forecasting, concluding that it can be used in both private insurance companies and public pensions systems in order to forecast mortality and mitigate longevity risk.

1 Introduction

Actuaries in life insurance business usually work with life tables in order to compute death and survival probabilities for pricing, calculating annuities, pension benefits and reserves.

During the second part of the twentieth century, life actuaries faced the situation that the average lifetime of annuitants had been systematically above what it had been expected according to these life tables.

This underestimation of survival probabilities has implied great losses for the insurance industry, since companies were forced to allocate more capital to sustain their annuity business. Therefore, it is crucial to rightly calibrate projected life

A. Benchimol (✉) • I. Albarrán • J.M. Marín • P. Alonso-González
Department of Statistics & Actuarial Science, The University of Hong Kong, Pok Fu Lam, Hong Kong
e-mail: benchi@hku.hk; irene.albarran@est-econ.uc3m.es; jmmarin@est-econ.uc3m.es; pablo.alonsog@uah.es

M. Corazza et al. (eds.), *Mathematical and Statistical Methods for Actuarial Sciences and Finance*, https://doi.org/10.1007/978-3-319-50234-2_4

tables. Otherwise the consequences are transferred to the financial statements of the insurance companies, which result in lower profits, losses or even bankruptcy.

Thus, a projected life table must account for the improvement in mortality in order to prevent from adverse effects on reserves and profitability, and therefore to look after insurance companies solvency in the long run. This situation is what is known as *longevity risk*.

The paper is structured as follows. Section 2 briefly summarizes the Lee-Carter model and introduces a Bayesian approach to it. Section 3 introduces data cloning, a novel methodology to approximate maximum likelihood estimators in hierarchical structures. Section 4 explains the validation of the estimation procedure of our proposed hierarchical Lee-Carter model with data cloning. Section 5 shows the application of our model to real data. Finally, Sect. 6 presents some conclusions.

2 Bayesian Approach to LC Model

Lee and Carter [4] proposed a model (*LC* from now on) to forecast mortality as a function of a time-varying index. This paper was the seminal work for further developments in the estimation of future mortality, such as [1, 3, 7] and [8]. The *LC* model can be expressed as:

$$\log\left[m_x(t)\right] = \alpha_x + \beta_x \kappa_t + \varepsilon_{x.t}, \tag{1}$$

where α_x's parameters describe the pattern of the average mortality at each age, while β_x's parameters describe deviations from this average pattern when κ_t varies. Both sets of parameters are independent of time. The variable, κ_t, is an unobservable index. It is a time series that describes the change in the level of mortality over time. Finally, $\varepsilon_{x.t}$, the error term, denotes random deviations unexplained by the model, and it is assumed to have mean 0 and constant variance σ_ε^2.

This section introduces a Bayesian methodology for making inferences about the parameters of a *LC* model for a set of related populations by means of a hierarchical Bayesian model. Originally, the Bayesian approach of the *LC* model applied to one population was studied by Pedroza [8] who analysed U.S. mortality data.

Hierarchical modeling is used when information about observational units is available on different levels. The hierarchical form of analysis and organization is quite useful regarding multi-parameter problems (see [2]). This paper is concerned to a case that considers several populations with some social-economic characteristics in common. For this reason, the parameters of the *LC* model can be treated as connected in some way, implying that the dependence among them can be reflected with a joint probability model.

Let be $j = 1.\ldots, J$ populations, in such a way that $m_x^{(j)}(t)$ is the central death rate for an age x at time t and population j, such that the corresponding *LC* model can be expressed as:

$$\log\left[m_x^{(j)}(t)\right] = \alpha_x^{(j)} + \beta_x^{(j)}\kappa_t + \varepsilon_{x.t}^{(j)},$$
$$\kappa_t = \kappa_{t-1} + \theta + \omega_t,$$

where $\varepsilon_{x.t}^{(j)} \sim N\left(0; \sigma_\varepsilon^{2(j)}\right)$ and $\omega_t \sim N\left(0; \sigma_\omega^2\right)$.

As usual, we observe $m_x^{(j)}(t)$ whereas κ_t is unobserved. The goal is to estimate the set of parameters $\alpha_x^{(j)}$, $\beta_x^{(j)}$ and κ_t and use them to forecast the central death rates at each age for each generation in each population.

We assume proper prior distributions with a hierarchical structure. We consider normal distributions for $\alpha_x^{(j)}$ parameters and Dirichlet distributions for $\beta_x^{(j)}$ ones in order to include the restriction $\sum_x \beta_x = 1$, we assume inverse-gamma distributions for variances as it is a conjugate distribution in a normal model, allowing to deal with most of real cases about prior information regarding the parameters of the model.

In particular, we assume as prior distributions for the parameters (for $j = 1, \ldots, J$):

$$\alpha_x^{(j)} \sim N(\mu_x^{(j)}, \sigma_x^{2(j)}),\ \beta_x^{(j)} \sim \text{Dirichlet}(1, 1, \ldots, 1),\ \sigma_\varepsilon^{2(j)} \sim InvGamma(\gamma_1,\ \gamma_2),$$
$$\kappa_1 \sim N(0,\ \sigma_\omega^2),\ \theta \sim N(0,\ 100) \text{ and } \sigma_\omega^2 \sim InvGamma(1,\ 1). \qquad (2)$$

And the hyperprior distributions for the hyperparameters:

$$\mu_x^{(j)} \sim N(0,\ 10),\ \sigma_x^{2(j)} \sim InvGamma(1,\ 1),$$
$$\gamma_1 \sim Gamma(1,\ 1) \text{ and } \gamma_2 \sim Gamma(1,\ 1).$$

On the other hand, the likelihood function for this model is:

$$L\left(m_x^{(j)}(t); \Theta\right) = \prod_{j=1}^{J}\prod_{t=1}^{n}\prod_{x=1}^{\omega} \frac{1}{\sqrt{2\pi}\sigma_\varepsilon^{(j)}} \exp\left[-\frac{1}{2}\left(\frac{\log\left[m_x^{(j)}(t)\right] - \alpha_x^{(j)} - \beta_x^{(j)}\kappa_t}{\sigma_\varepsilon^{(j)}}\right)^2\right],$$

where $\Theta = \left(\alpha_x^{(j)}, \beta_x^{(j)}, \kappa_1, \mu_x^{(j)}, \sigma_x^{2(j)}, \sigma_\varepsilon^{2(j)}, \theta, \sigma_\omega^2, \gamma_1, \gamma_2\right)$ for $j = 1, \ldots, J$ and $x = 1, \ldots, \omega$.

The joint posterior distribution for all the parameters is obtained by multiplying the likelihood function by the corresponding prior distributions (2). In general, the full set of conditional distributions is required to implement an MCMC algorithm. Then, the conditional posterior distributions of each parameter is easily obtained from the joint posterior distribution, considering only the proportional terms to each parameter.

3 The Data Cloning Methodology

The data cloning method is a simulation technique to compute maximum likelihood estimates of parameters along with their asymptotic variances, by using a MCMC methodology (see [5] and [6]). It uses the simplicity of Monte Carlo algorithms to calculate maximum likelihood estimations in models whose complexity make necessary the use of high-dimensional integration to obtain them. The methodology is based on the basic idea of repeating an experiment several times conditioned to obtain always the same data.

Let us denote $\log[m_x(t)]$ as y_t for $(t = 1, \ldots, n)$ in such a way that $\mathbf{y} = (y_1, \ldots, y_n)$. In a MCMC procedure, once data $\mathbf{y}$ has been observed, the samples from the posterior distribution $\pi(\Theta|\mathbf{y})$ are generated. This posterior distribution is proportional to the product of the likelihood function $L(\boldsymbol{\Theta}|\mathbf{y})$ and a given proper prior distribution $\pi(\boldsymbol{\Theta})$. Then, in data cloning, samples are generated from the posterior distribution, $\pi^{(K)}(\boldsymbol{\Theta}|\mathbf{y})$, that is proportional to the Kth power of the likelihood, $[L(\Theta|\mathbf{y})]^K$, multiplied by a proper prior distribution, $\pi(\boldsymbol{\Theta})$.

The expression $[L(\Theta|\mathbf{y})]^K$ is the likelihood for K copies of the original data and, for K large enough, $\pi^{(K)}(\boldsymbol{\Theta}|\mathbf{y})$ converges to a multivariate normal distribution with mean equal to the maximum likelihood estimates of the parameters, and covariance matrix equal to $1/K$ times the inverse of the Fisher information matrix for the maximum likelihood estimates (see [5]).

Once the samples have been obtained with a MCMC procedure, sample means are computed based on the posterior distributions of the parameters. They provide an approximation of the maximum likelihood estimates of the parameters. As a summary, the data cloning algorithm follows the next steps:

Step 1: Create K-cloned data set $\mathbf{y}^{(K)} = (\mathbf{y}, \mathbf{y}, \ldots, \mathbf{y})$, where the observed data vector is repeated K times.

Step 2: Using an MCMC algorithm, generate random numbers from the posterior distribution that is based on a prior $\pi(\boldsymbol{\Theta})$ and the cloned data vector $\mathbf{y}^{(K)} = (\mathbf{y}, \mathbf{y}, \ldots, \mathbf{y})$, where the K copies of $\mathbf{y}$ are assumed to be independent of each other. In practice, any proper prior distribution can be used.

Step 3: Compute the sample mean and variances of each of the individual values of the vector of parameters $\boldsymbol{\Theta}$ (for M iterations of the MCMC algorithm) generated from the marginal posterior distribution. The maximum likelihood estimates of $\boldsymbol{\Theta}$ and the approximate variances of the maximum likelihood estimates are those referred to the posterior mean values and to K times the posterior variances respectively.

The algorithm has been programmed using the package `dclone` [10] from the `R` project (R Core Team. R Foundation for Statistical Computing (2012)). The optimal number of clones has been established considering some statistics computed in the package `dclone` [10], such as the maximum eigenvalue of the posterior covariance matrix, the minimum squared error and the squared error (see [6]).

4 Validating the Estimation Procedure

Before applying the procedure to real data, we check it by means of a validation study. The validation has been done by simulating an array of log-central death rates, assuming 4 populations, 41 consecutive ages and 50 consecutive calendar years each.

In order to simulate data we took as a starting point the estimated parameters for male French data taken from the *Human Mortality Database* (see www.mortality.org) for ages between 60 and 100, and for calendar years between 1960 and 2009, including both ends in both cases. The estimation was undertaken by the standard frequentist procedure based on the singular value decomposition raised by Lee and Carter [4] using `R`.

For population 1, we generated each parameter $\alpha_x^{(1)}$ as the respective French $\hat{\alpha}_x$ parameter plus a random component following a $\mathrm{N}(\mu = 0; \sigma = 0.001)$ distribution. For populations 2, 3 and 4, $\alpha_x^{(j)}$ parameters were generated taking 95%, 90%, and 105% of French $\hat{\alpha}_x$ parameters, plus a random component following $\mathrm{N}(\mu = 0; \sigma = 0.002)$. $\mathrm{N}(\mu = 0; \sigma = 0.003)$ and $\mathrm{N}(\mu = 0; \sigma = 0.004)$ distributions, respectively. Regarding $\beta_x^{(j)}$ parameters for each population, they were simulated by four vectors of dimension 41, each of them following a Dirichlet distribution with all its parameters equal to 1, in order to fulfill the constraint $\sum_x \beta_x = 1$.

On the other hand, regarding the mortality index κ_t, adjusting for France estimates, parameters were assumed to be as $\hat{\kappa}_0 = -11.2075$, $\hat{\theta} = -0.5728$ and $\hat{\sigma}_\omega^2 = 5$. Thus, κ_t was fixed as a vector of 50 values from $t = 1960$ up to $t = 2009$, where each κ_t was a random value generated from a normal distribution with mean $\kappa_{t-1} + \hat{\theta}$ and variance $\hat{\sigma}_\omega^2$.

Finally, the data set of log-central death rates $\log\left[m_x^{(j)}(t)\right]$ (four matrices of dimensions 50×41, one for each population) was simulated by means of four normal distributions with means $\alpha_x^{(j)} + \beta_x^{(j)}\kappa_t$ and variances $\sigma_y^{2(1)} = 0.0010$, $\sigma_y^{2(2)} = 0.0015$, $\sigma_y^{2(3)} = 0.0020$ and $\sigma_y^{2(4)} = 0.0025$, respectively.

We used the package `dclone` [10] from the `R` project [9] in order to program the algorithm. We checked the optimal number of clones by means of some statistics computed in the package `dclone` [10], such as the maximum eigenvalue of the posterior variance, the minimum squared error and the squared error (see [6]). There were not relevant improvements in their values when the number of clones was larger than 5, and therefore we worked with 5 clones to estimate the parameters. Besides, we used the same prior distributions as in Sect. 2.

Then, the parameters of this hierarchical model were estimated applying our algorithm with the data cloning technique, on the same simulated data set 100 times, generating thus 100 replicates, and the results were assessed.

Let us consider the Pearson's coefficient of variation (CV) and the relative mean squared error ($RMSE$) respectively as $CV(\hat{\theta}) = \sigma_{\hat{\theta}}/|\bar{\hat{\theta}}|$ and $RMSE(\hat{\theta}) = \sqrt{E[(\hat{\theta}-\theta)^2]}/|\theta|$.

We calculated both relative dispersion measures for the 8200 (50 calendar years × 41 ages × 4 populations) estimators of the log-central death rates, $\log\left[\widehat{m}_x^{(j)}(t)\right]$. The mean CV for the whole sample was 0.048, and the mean RMSE was 0.047. In the case of the CV, 154 out of 8200 were larger than 0.2. while in the case of the RMSE, 171 out of 8200 were larger than this threshold (1.87% and 2.08% of the sample, respectively). These results suggest that the procedure seems to be unbiased and stable.

5 Application to Real Data

The data cloning methodology has been applied to a set of European countries male mortality data. We have taken the central death rates from France, Italy, Portugal and Spain from the *Human Mortality Database* (see www.mortality.org).

We have selected a time span between years 1960 and 2009 and ages between 60 and 100 years old. These countries present a similar social development and their populations enjoy parallel welfare states. More specifically, they show similar standard demographic indices such as:

- *Life expectancy at birth* (LEB).
- *Life expectancy at age 65* (LE65)
- *Median age of the population* (MAP)
- *Old-dependency ratio* (ODR): it is the proportion between the number of people over 65 years old and the number of people between 16 and 64 years old.

Table 1 Demographic indices

	LEB		LE65		MAP		ODR
	Male	Female	Male	Female	Male	Female	General
France	78.7	85.4	19.1	23.4	38.2	41.2	27.5
Italy	79.8	84.8	18.5	22.1	43.0	45.6	32.7
Portugal	77.3	83.6	17.6	21.3	37.6	41.9	29.4
Spain	79.5	85.8	18.7	22.8	40.1	42.9	26.3

Source: Eurostat 2012, 2013

The corresponding values of these indices for years 2012–2013 for each country are shown in Table 1.

In order to validate the predictive performance of the model, the data set was split into two groups: the first one (training sample) includes data from 1960 to 1999, whereas the second one (validation sample) includes data from 2000 to 2009.

We complete the analysis of the hierarchical model by applying the data cloning technique. We have programmed the algorithm using package `dclone` [10] from the `R` project [9]. We check what is the optimal number of clones, regarding some statistics computed in the package `dclone` [10], such as the maximum eigenvalue of the posterior variance, the minimum squared error and the squared error (see [6]). There are not relevant improvements in their values when the number of clones is larger than 5, therefore we use this number of clones to analyse the data. We have used the same prior distributions as in Sect. 2.

Mean and standard deviation of parameters were estimated for the training sample using 5 clones after 50,000 iterations of the MCMC algorithm.

Using the estimations of $\alpha_x^{(j)}$. $\beta_x^{(j)}$ for each age x and each country, along with the forecasted values κ_t, the log-central death rates $\log\left[m_x^{(j)}(t)\right]$ were estimated for the period 2000–2009.

The 95% prediction intervals for the predicted values, based on the Wald approximation, and the actual values of the log-central death rates are shown in Tables 2, 3, 4, and 5 for France, Italy, Portugal and Spain, respectively, for ages $x = 60, 70, 80, 90$ and 100, and for a prediction horizon $t = 2000, \ldots, 2009$. Notice that all intervals include the actual values of the log-central death rates.

Table 2 France

t	x = 60			x = 70			x = 80		
	95% PI			95% PI			95% PI		
	Up B.	Low B.	Actual value	Up B.	Low B.	Actual value	Up B.	Low B.	Actual value
2000	−4.7128	−3.9458	−4.4295	−3.9740	−3.1368	−3.6005	−2.9710	−2.2088	−2.5806
2001	−4.8500	−3.8354	−4.5117	−4.1305	−3.0090	−3.6439	−3.1051	−2.1003	−2.6520
2002	−4.9636	−3.7463	−4.4671	−4.2609	−2.9074	−3.6412	−3.2169	−2.0137	−2.6533
2003	−5.0685	−3.6671	−4.4559	−4.3786	−2.8188	−3.6893	−3.3199	−1.9357	−2.6526
2004	−5.1645	−3.5965	−4.5236	−4.4888	−2.7362	−3.7474	−3.4149	−1.8647	−2.7465
2005	−5.2567	−3.5288	−4.5153	−4.5919	−2.6616	−3.7872	−3.5042	−1.8004	−2.7589
2006	−5.3435	−3.4674	−4.5394	−4.6924	−2.5902	−3.8231	−3.5915	−1.7383	−2.7926
2007	−5.4298	−3.4073	−4.5413	−4.7886	−2.5226	−3.8467	−3.6755	−1.6796	−2.8166
2008	−5.5138	−3.3487	−4.5631	−4.8817	−2.4572	−3.8839	−3.7573	−1.6232	−2.8341
2009	−5.5926	−3.2943	−4.5516	−4.9730	−2.3955	−3.9001	−3.8348	−1.5706	−2.8810

t	x = 90			x = 100		
	95% PI			95% PI		
	Up B.	Low B.	Actual value	Up B.	Low B.	Actual value
2000	−1.8009	−1.2453	−1.5477	−0.9304	−0.4895	−0.7227
2001	−1.8655	−1.1954	−1.5622	−0.9413	−0.4828	−0.7344
2002	−1.9237	−1.1512	−1.5732	−0.9512	−0.4780	−0.7267
2003	−1.9790	−1.1110	−1.5299	−0.9617	−0.4719	−0.6907
2004	−2.0279	−1.0758	−1.6495	−0.9730	−0.4645	−0.7840
2005	−2.0791	−1.0391	−1.5507	−0.9839	−0.4583	−0.7511
2006	−2.1249	−1.0080	−1.7740	−0.9956	−0.4512	−0.7869
2007	−2.1721	−0.9752	−1.6663	−1.0048	−0.4465	−0.7803
2008	−2.2179	−0.9427	−1.6750	−1.0183	−0.4371	−0.7791
2009	−2.2609	−0.9147	−1.7800	−1.0291	−0.4317	−0.7737

Table 3 Italy

	x = 60			x = 70			x = 80		
	95% PI			95% PI			95% PI		
t	Up B.	Low B.	Actual value	Up B.	Low B.	Actual value	Up B.	Low B.	Actual value
2000	−4.8909	−4.0362	−4.6628	−3.8806	−3.1145	−3.5972	−2.8838	−2.1691	−2.4969
2001	−5.0499	−3.9076	−4.6356	−4.0097	−3.0106	−3.6276	−3.0013	−2.0753	−2.6571
2002	−5.1816	−3.8045	−4.7128	−4.1192	−2.9252	−3.6861	−3.0990	−1.9990	−2.6116
2003	−5.3034	−3.7119	−4.6450	−4.2206	−2.8484	−3.7132	−3.1909	−1.9292	−2.5919
2004	−5.4134	−3.6305	−4.7028	−4.3126	−2.7804	−3.7491	−3.2751	−1.8678	−2.6863
2005	−5.5203	−3.5526	−4.7689	−4.4024	−2.7153	−3.7828	−3.3551	−1.8101	−2.6757
2006	−5.6238	−3.4781	−4.7879	−4.4877	−2.6554	−3.8509	−3.4315	−1.7571	−2.7243
2007	−5.7188	−3.4124	−4.7855	−4.5708	−2.5979	−3.8849	−3.5070	−1.7033	−2.7186
2008	−5.8175	−3.3427	−4.8258	−4.6518	−2.5407	−3.8709	−3.5804	−1.6518	−2.7292
2009	−5.9069	−3.2828	−4.8757	−4.7274	−2.4889	−3.9394	−3.6488	−1.6054	−2.7952

	x = 90			x = 100		
	95% PI			95% PI		
t	Up B.	Low B.	Actual value	Up B.	Low B.	Actual value
2000	−1.8463	−1.2411	−1.5906	−1.0218	−0.5293	−0.7425
2001	−1.9261	−1.1781	−1.5580	−1.0580	−0.5028	−0.7473
2002	−1.9971	−1.1236	−1.5955	−1.0905	−0.4797	−0.7606
2003	−2.0629	−1.0755	−1.5318	−1.1223	−0.4577	−0.7106
2004	−2.1228	−1.0319	−1.6455	−1.1541	−0.4348	−0.8115
2005	−2.1829	−0.9880	−1.6189	−1.1834	−0.4151	−0.7651
2006	−2.2381	−0.9500	−1.6573	−1.2137	−0.3945	−0.8031
2007	−2.2939	−0.9112	−1.6724	−1.2431	−0.3753	−0.7857
2008	−2.3487	−0.8733	−1.6676	−1.2702	−0.3571	−0.7717
2009	−2.3990	−0.8390	−1.7730	−1.2988	−0.3380	−0.7741

Table 4 Portugal

t	x = 60			x = 70			x = 80		
	95% PI			95% PI			95% PI		
	Up B.	Low B.	Actual value	Up B.	Low B.	Actual value	Up B.	Low B.	Actual value
2000	−4.7066	−3.9522	−4.4006	−3.8358	−3.0136	−3.4477	−2.7732	−2.0580	−2.4174
2001	−4.8325	−3.8508	−4.4030	−3.9827	−2.8943	−3.4950	−2.8888	−1.9652	−2.4902
2002	−4.9381	−3.7692	−4.5044	−4.1050	−2.7993	−3.5153	−2.9855	−1.8910	−2.4653
2003	−5.0367	−3.6946	−4.4533	−4.2187	−2.7128	−3.4940	−3.0754	−1.8231	−2.5002
2004	−5.1270	−3.6272	−4.5181	−4.3214	−2.6370	−3.6001	−3.1572	−1.7632	−2.5275
2005	−5.2140	−3.5650	−4.4892	−4.4198	−2.5651	−3.6227	−3.2372	−1.7058	−2.5157
2006	−5.2962	−3.5075	−4.4954	−4.5145	−2.4988	−3.6885	−3.3123	−1.6530	−2.5828
2007	−5.3785	−3.4490	−4.4733	−4.6057	−2.4354	−3.7206	−3.3865	−1.6010	−2.5627
2008	−5.4562	−3.3954	−4.5403	−4.6952	−2.3725	−3.7285	−3.4590	−1.5502	−2.5732
2009	−5.5290	−3.3465	−4.5708	−4.7800	−2.3146	−3.7243	−3.5266	−1.5045	−2.6045

t	x = 90			x = 100		
	95% PI			95% PI		
	Up B.	Low B.	Actual value	Up B.	Low B.	Actual value
2000	−1.7562	−1.1486	−1.4443	−0.9781	−0.4296	−0.6758
2001	−1.8262	−1.0944	−1.4653	−0.9989	−0.4157	−0.6769
2002	−1.8884	−1.0482	−1.4877	−1.0189	−0.4023	−0.6744
2003	−1.9470	−1.0049	−1.4203	−1.0408	−0.3875	−0.6642
2004	−2.0037	−0.9639	−1.5946	−1.0614	−0.3738	−0.7622
2005	−2.0542	−0.9286	−1.4342	−1.0813	−0.3609	−0.6674
2006	−2.1069	−0.8914	−1.5836	−1.1012	−0.3485	−0.7414
2007	−2.1592	−0.8559	−1.5311	−1.1219	−0.3347	−0.7582
2008	−2.2075	−0.8224	−1.4946	−1.1418	−0.3217	−0.7629
2009	−2.2542	−0.7905	−1.6661	−1.1615	−0.3099	−0.7741

Table 5 Spain

t	x = 60			x = 70			x = 80		
	95% PI			95% PI			95% PI		
	Up B.	Low B.	Actual value	Up B.	Low B.	Actual value	Up B.	Low B.	Actual value
2000	−4.8579	−4.1486	−4.5375	−4.0072	−3.2149	−3.6268	−2.9560	−2.2346	−2.5579
2001	−4.9948	−4.0364	−4.5612	−4.1679	−3.0828	−3.6153	−3.0980	−2.1186	−2.6364
2002	−5.1104	−3.9457	−4.5602	−4.3009	−2.9785	−3.6766	−3.2134	−2.0281	−2.6282
2003	−5.2139	−3.8671	−4.5422	−4.4202	−2.8872	−3.6679	−3.3210	−1.9464	−2.6194
2004	−5.3089	−3.7964	−4.6032	−4.5296	−2.8057	−3.7297	−3.4180	−1.8737	−2.6573
2005	−5.4001	−3.7303	−4.5863	−4.6345	−2.7290	−3.7503	−3.5114	−1.8063	−2.6116
2006	−5.4855	−3.6698	−4.5824	−4.7335	−2.6593	−3.8027	−3.5991	−1.7439	−2.7440
2007	−5.5687	−3.6114	−4.6295	−4.8296	−2.5917	−3.7810	−3.6848	−1.6836	−2.7161
2008	−5.6523	−3.5528	−4.6183	−4.9237	−2.5258	−3.8295	−3.7681	−1.6255	−2.7657
2009	−5.7290	−3.5002	−4.7161	−5.0119	−2.4652	−3.8685	−3.8470	−1.5721	−2.7853

t	x = 90			x = 100		
	95% PI			95% PI		
	Up B.	Low B.	Actual value	Up B.	Low B.	Actual value
2000	−1.8425	−1.3199	−1.5932	−0.9796	−0.6158	−0.7771
2001	−1.9231	−1.2565	−1.5937	−0.9986	−0.6030	−0.7818
2002	−1.9922	−1.2027	−1.6224	−1.0165	−0.5899	−0.7872
2003	−2.0574	−1.1537	−1.5202	−1.0340	−0.5783	−0.7302
2004	−2.1149	−1.1115	−1.6183	−1.0512	−0.5671	−0.7838
2005	−2.1721	−1.0694	−1.5957	−1.0676	−0.5569	−0.7632
2006	−2.2258	−1.0328	−1.6630	−1.0863	−0.5434	−0.8045
2007	−2.2808	−0.9931	−1.6424	−1.1022	−0.5334	−0.7979
2008	−2.3306	−0.9589	−1.6481	−1.1183	−0.5232	−0.7949
2009	−2.3799	−0.9258	−1.6944	−1.1350	−0.5122	−0.8178

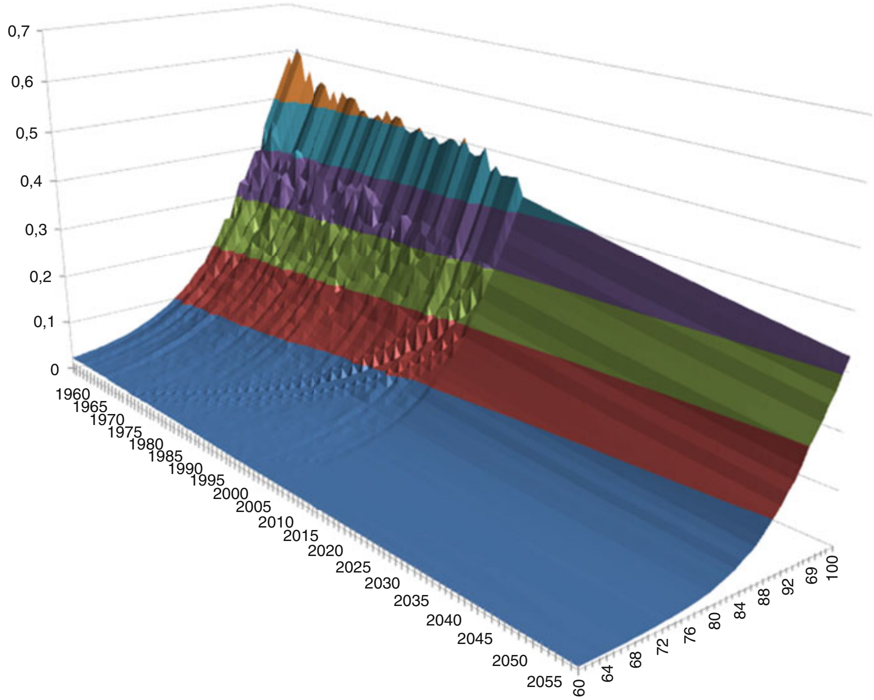

Fig. 1 France

Finally, once the predictive capability of the proposed model and methodology has been validated, we may consider projecting mortality for a long time horizon, e.g. for the following 50 years. In Figs. 1, 2, 3, and 4, projected mortality surfaces for France, Italy, Portugal and Spain, respectively, are shown for the following 50 years. In the right hand side of each figure, the observed central death rates $m_x^{(j)}(t)$ are represented by the height of the surface for $t = 1960, \ldots, 2009$, while in the right hand side, the projected central death rates for $t = 2010, \ldots, 2059$ are represented. Each change of color in the surfaces shows a change in tenth of the death rates. Notice that a substantial improvement in mortality rates is expected for the four populations.

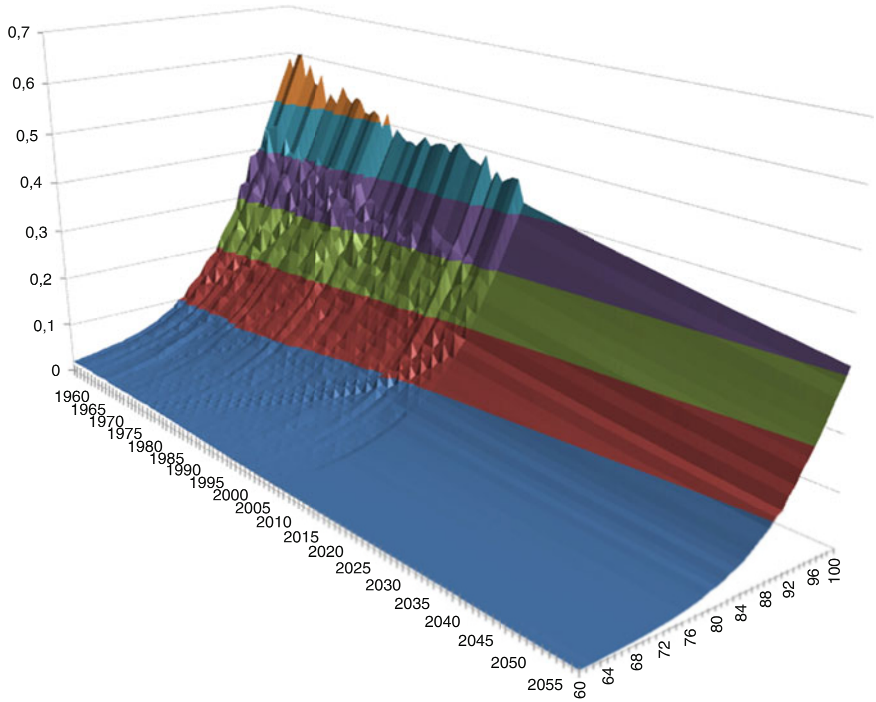

Fig. 2 Italy

6 Conclusions

We have introduced a hierarchical Lee-Carter specification to forecast the death rates of a set of demographically linked countries. This hierarchical scheme is based on the assumption of a common and latent mortality factor. This proposal is useful to estimate the parameters of the model because it allows actuaries to take advantage of the whole set of information, this is, the forecasts of a particular country are based not only on its death rates but also in those of the rest of the considered linked countries.

Bayesian methodology is the usual way to deal with hierarchical models. However, this approach has the limitation that it is necessary to determine the prior distributions for all parameters and hyperparameters of the model. Thus, data cloning provides a tool to overcome the previous limitation and it permits to approximate maximum likelihood estimates. In this approach, the role of the prior distributions is indifferent, since the results are invariant to these prior distributions.

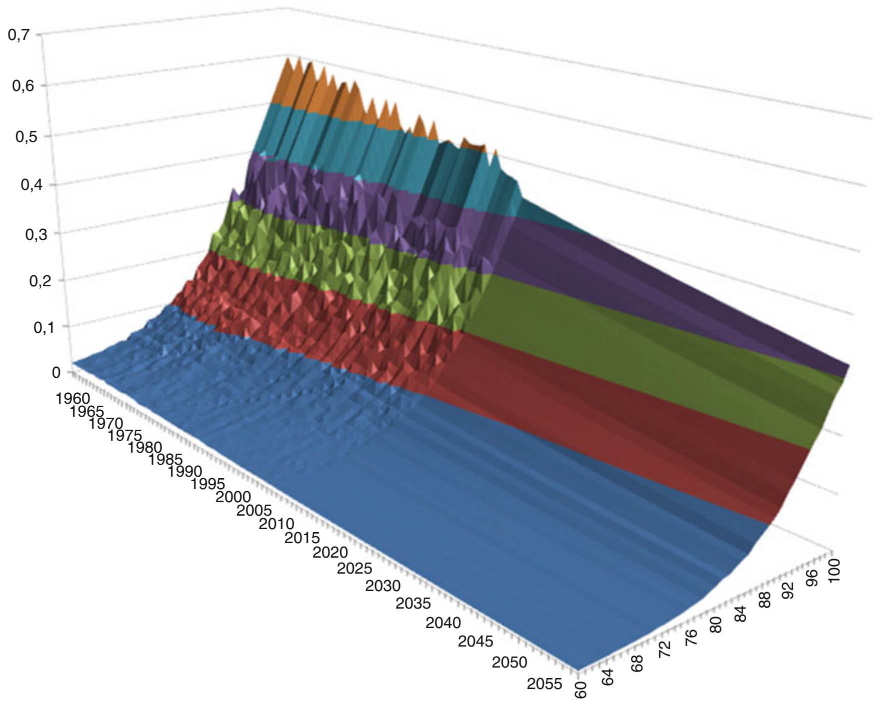

Fig. 3 Portugal

So far, Data Cloning has been applied to the Ecology field and recently in Finance, and it is the first time that this methodology is used in the Actuarial field, to the best of our knowledge.

The set of information includes the central death rates of France, Italy, Portugal and Spain. In order to check the validity of the forecasts, the sample has been split into two sets. The first one is devoted to estimate the parameters, whereas the second one is used to contrast the accuracy of the results. It was shown that the model is able to rightly predict the central death rates in all cases, using 95% approximated prediction intervals. This model can be directly applied to annuities business and/or public pension systems, to obtain accurate death rates forecastings, mitigating longevity risk.

Future research will involve the implementation of this methodology to other kind of models used to forecast death rates, such as Cairns-Blake-Dowd (*CBD*) stochastic mortality models, or those based on P-splines.

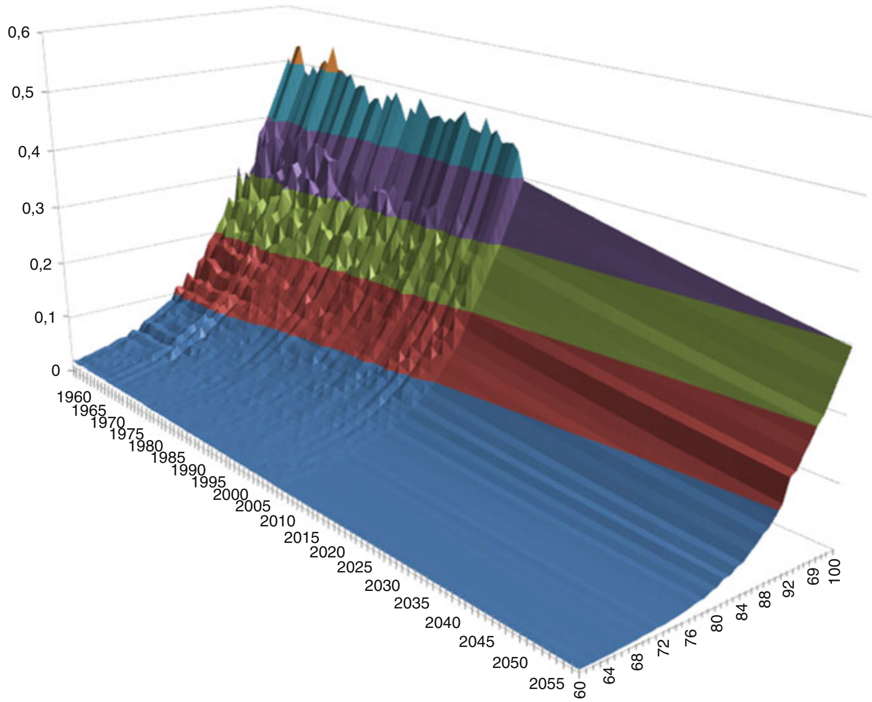

Fig. 4 Spain

References

1. Booth, H., Maindonald, J., Smith, L.: Applying Lee-Carter under conditions of variable mortality decline. Popul. Stud. **56**(3), 325–336 (2002)
2. Gelman, A., Carlin, J.B., Stern, H.S., Rubin, D.B.: Bayesian Data Analysis. Taylor & Francis, Abingdon (2014)
3. Lee, R.D.: The Lee-Carter method for forecasting mortality, with various extensions and applications. N. Am. Actuar. J. **4**(1), 80–93 (2000)
4. Lee, R.D., Carter, L.R.: Modeling and forecasting U.S. mortality. J. Am. Stat. Soc. **87**, 659–675 (1992)
5. Lele, S.R., Dennis, B., Lutsche, F.: Data cloning: easy maximum likelihood estimation for complex ecological models using Bayesian Markov Chain Monte Carlo methods. Ecol. Lett. **10**, 551–563 (2007)
6. Lele, S.R., Nadeem, K., Schmuland, B.: Estimability and likelihood inference for generalized linear mixed models using data cloning. J. Am. Stat. Soc. **105**, 1617–1625 (2010)
7. Li, N., Lee, R.D.: Coherent mortality forecasts for a group of populations and extension of the Lee-Carter method. Demography **42**, 575–594 (2005)
8. Pedroza, C.: A Bayesian forecasting model: predicting U.S. male Mortality. Biostatistics **7**(4), 530–550 (2006)
9. R Core Team: R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing. http://www.R-project.org/ (2015)
10. Sólymos, P.: dclone: Data Cloning and MCMC Tools for Maximum Likelihood Methods, R package version (2009)

Markov Switching GARCH Models: Filtering, Approximations and Duality

Monica Billio and Maddalena Cavicchioli

Abstract This paper is devoted to show duality in the estimation of Markov Switching (MS) GARCH processes. It is well-known that MS GARCH models suffer of path dependence which makes the estimation step unfeasible with usual Maximum Likelihood procedure. However, by rewriting the model in a suitable state space representation, we are able to give a unique framework to reconcile the estimation obtained by filtering procedure with that coming from some auxiliary models proposed in the literature. Estimation on short-term interest rates shows the feasibility of the proposed approach.

1 Introduction

Time varying volatility is one of the main property of many financial time series. Moreover, describing and, where possible, forecasting volatility is a key aspect in financial economics and econometrics. A popular class of models which describe time-varying volatility are Generalized Autoregressive Conditional Heteroschedasticity (GARCH) models. GARCH models [8, 21, 22] describe the variance as a linear function of the squares of past observations, so that one type of shock alone drives both the series itself and its volatility. One potential source of misspecification derives from the fact that structural forms of conditional means and variances are relatively inflexible and held fixed throughout the sample period. In this sense, they are called *single-regime* models since a single structure for the conditional mean and variance is assumed.

M. Billio (✉)
Department of Economics, Ca' Foscari University of Venice, Cannaregio 873, 30121 Venice, Italy
e-mail: billio@unive.it

M. Cavicchioli
Department of Economics, University of Verona, Via Cantarane 24, 37129 Verona, Italy
e-mail: maddalena.cavicchioli@univr.it

M. Corazza et al. (eds.), *Mathematical and Statistical Methods for Actuarial Sciences and Finance*, https://doi.org/10.1007/978-3-319-50234-2_5

To allow more flexibility, the assumption of a single regime could be relaxed in favour of a *regime-switching* model. The coefficients of this model are different in each regime to account for the possibility that the economic mechanism generating the financial series undergoes a finite number of changes over the sample period. These coefficients are unknown and must be estimated, and, although the regimes are never observed, probabilistic statements can be made about the relative likelihood of their occurrence, conditional on an information set. A well-known problem to face when dealing with the estimation of Markov Switching (MS) GARCH models is the path dependence. Cai [9] and Hamilton and Susmel [16] have argued that MS GARCH models are essentially intractable and impossible to estimate since the conditional variance depends on the entire path history of the data. That is, the distribution at time t, conditional on the current state and on available information, is directly dependent on the current state but also indirectly dependent on all past states due to the path dependence inherent in MS GARCH models. This is because the conditional variance at time t depends upon the conditional variance at time $t-1$, which depends upon the regime at time $t-1$ and on the conditional variance at time $t-2$, and so on. Hence, the conditional variance at time t depends on the entire sequence of regimes up to time t.

Some methods are proposed in the literature to overcome the problem of path dependence present in MS GARCH. The trick is mainly found in adopting different specifications of the original MS GARCH model. Some authors propose Quasi Maximum Likelihood (QML) procedures of a model which allows similar effects of the original one. Models which elude in this way the path dependence problem are proposed by Gray [14], Dueker [10] and Klaassen [18], among others, and are known as collapsing procedures. Gray [14] proposes a model in which path dependence is removed by aggregating the conditional variances from the regimes at each step. This aggregated conditional variance (conditional on available information, but aggregated over the regimes) is then all that is required to compute the conditional variance at the next step. The same starting idea is used in [10], with a slightly different approach. The author extends the information set including also current information on the considered series. Furthermore, Klaassen [18] puts further this idea. Particularly, when integrating out the unobserved regimes, all the available information is used, whereas [14] uses only part of it. Another method to deal with MS GARCH models has been proposed by Haas et al. [15] for which the variance is disaggregated in independent processes; this is a simple generalization of the GARCH process to a multi-regime setting. Furthermore, Bayesian approaches based on Markov Chain Monte Carlo Gibbs technique for estimating MS GARCH can be found in [3, 4, 17] or [7]. Other works based on both Monte Carlo methods combined with expectation-maximization algorithm and importance sampling to evaluate Maximum Likelihood (ML) estimators are conducted by Augustyniak [1] and Billio et al. [5, 6]. Finally, a recent paper by Augustyniak et al. [2] proposes estimation of MS GARCH models with a deterministic particle filter.

The contribution of our paper is to give a unique framework to reconcile MS GARCH estimation obtained by the above auxiliary models from one side, and a filtering algorithm from the other. This relationship provides the missing link to justify the validity of approximations in estimating MS GARCH models. The use of filtering is a flexible approach and it allows the estimation of a broad class of models that can be put in a switching state space form. However, to make the filter operable, at each iteration we need to collapse M^2 posteriors (where M is the number of switching regimes) in M of it, employing an approximation as suggested by Kim [19]. Then, QML estimation of the model recovers the unknown parameters. Our algorithm is readily programmable and with a very limited computational cost. An empirical application shows the feasibility of this approach.

The paper is structured as follows. Section 2 introduces the MS GARCH model of interest and reviews the main auxiliary models proposed in the literature to overpass the path dependence problem. In Sect. 3 we present a filtering algorithm for MS GARCH models which serves to prove our duality results. In Sect. 4 we compare estimation of the parameters using different approximations in the proposed filters for financial data. Section 5 concludes. Finally, derivations of some formulae are given in the Appendix.

2 Markov Switching GARCH and Its Auxiliary Models

Let ϵ_t be the observed univariate time series variable (as for instance, returns on a financial asset) centered on its mean. The univariate MS GARCH(1,1) model is defined as

$$\begin{cases} \epsilon_t = \sigma_t(\Psi_{t-1}, s_t)u_t \\ \sigma_t^2(\Psi_{t-1}, s_t) = \omega_{s_t} + \alpha_{s_t}\epsilon_{t-1}^2 + \beta_{s_t}\sigma_{t-1}^2(\Psi_{t-2}, s_{t-1}) \end{cases} \tag{1}$$

where $u_t \sim IID(0, 1)$, $\omega_{s_t} > 0$, $\alpha_{s_t}, \beta_{s_t} \geq 0$. The state s_t is a discrete, unobserved variable following a first order Markov chain with M regimes and (time invariant) transition probabilities $\pi_{ij} = p(s_t = j|s_{t-1} = i)$, where $\sum_{j=1}^{M} \pi_{ij} = 1$, for every $i = 1, \ldots, M$. We assume that (s_t) is independent of (u_t). For necessary and sufficient stationarity conditions related with MS GARCH(p, q) models, we refer to [11], Theorems 1 and 2. Consistency of maximum likelihood estimates, L^2 structure and inference for univariate MS GARCH models have been investigated by Francq et al. [11, 12] and Bauwens et al. [3]. Here the common approach to eliminate path dependence is to replace the lagged conditional variance derived from the original MS GARCH model with a proxy. Various authors have proposed different auxiliary models which differ only by the content of the information used to define such a proxy. In general, different auxiliary models can be obtained by approximating the conditional variance of process in (1) with

$$\sigma_t^2(\Psi_{t-1}, s_t) = \omega_{s_t} + \alpha_{s_t}{}^{(SP)}\epsilon_{t-1}^2 + \beta_{s_t}{}^{(SP)}\sigma_{t-1}^2. \tag{2}$$

In the literature there are different specifications (in short, SP) of ${}^{(SP)}\epsilon_{t-1}^2$ and ${}^{(SP)}\sigma_{t-1}^2$ which in turn define different approximations of the original process. These collapsing procedures are illustrated below. First, we introduce some concepts and notations: $p(s_t = j|\Psi_{t-1}) = p_{j,t|t-1}$ are prediction probabilities; $p(s_t = j|\Psi_t) = p_{j,t|t}$ are filtered probabilities and from these we can compute augmented filtered probabilities as

$$p(s_{t-1} = i|s_t = j, \Psi_{t-1}) = \frac{\pi_{ij}\ p_{i,t-1|t-1}}{p_{j,t|t-1}} = p_{i,t-1|t,t-1}.$$

Here Ψ_t denotes the information set of observations available up to time $t-1$. Note that the filtering algorithm computes $p_{t|t-1,t} = p(s_t|s_{t-1}, \Psi_t)$ in terms of $p_{t|t-1,t-1}$ and the conditional density of ϵ_t which depends on the current regime s_t and all past regimes, i.e, $f(\epsilon_t|s_1, \ldots, s_t, \Psi_{t-1})$. Computation details are shown in Appendix A1.

2a. Gray's Model The first attempt to eliminate the path dependence is proposed by Gray [14]. He approximates the original model by replacing the lagged conditional variance σ_{t-1}^2 with a proxy ${}^{(G)}\sigma_{t-1}^2$ as follows:

$$\begin{aligned} {}^{(G)}\sigma_{t-1}^2 &= E[\sigma_{t-1}^2(\Psi_{t-2}, s_{t-1})|\Psi_{t-2}] \\ &= \sum_{i=1}^{M} \sigma_{t-1}^2(\Psi_{t-2}, s_{t-1} = i)\ p(s_{t-1} = i|\Psi_{t-2}) = \sum_{i=1}^{M} {}^{(G)}\sigma_{i,t-1|t-2}^2\ p_{i,t-1|t-2} \end{aligned} \tag{3}$$

where, according to the model, ${}^{(G)}\sigma_{t-1|t-2}^2$ turns out to be a function of Ψ_{t-2} and $s_{t-1} = i$. Note that the model originally proposed by Gray is not centered as in our case, but this can always be assumed without loss of generality.

2b. Dueker's Model In the previous approximation, the information coming from ϵ_{t-1} is not used. Dueker [10] proposes to change the conditioning scheme including ϵ_{t-1} while assuming that σ_{t-1}^2 is a function of Ψ_{t-2} and s_{t-2}. Hence

$$\begin{aligned} {}^{(D)}\sigma_{t-1}^2 &= E[\sigma_{t-1}^2(\Psi_{t-2}, s_{t-2})|\Psi_{t-1}] \\ &= \sum_{k=1}^{M} \sigma_{t-1}^2(\Psi_{t-2}, s_{t-2} = k)\ p(s_{t-2} = k|\Psi_{t-1}) = \sum_{k=1}^{M} {}^{(D)}\sigma_{k,t-1|t-2}^2\ p_{k,t-2|t-1} \end{aligned} \tag{4}$$

so that ${}^{(D)}\sigma_{t-1|t-2}^2$ is a function of Ψ_{t-2} and $s_{t-2} = k$, and $p_{k,t-2|t-1}$ is one-period ahead smoothed probability which, shifting one period, can be computed as

$$p_{i,t-1|t} = p(s_{t-1} = i|\Psi_t) = p_{i,t-1|t-1} \sum_{j=1}^{M} \frac{\pi_{ij}\ p_{j,t|t}}{p_{j,t|t-1}}.$$

2c. Simplified Klaassen's Model The approximation proposed by Klaassen [18] is similar to that from [10] but it assumes that σ_{t-1}^2 is a function of Ψ_{t-2} and s_{t-1}. So it results computationally simpler. In fact, we have

$$
\begin{aligned}
{}^{(SK)}\sigma^2_{t-1} &= E[\sigma^2_{t-1}(\Psi_{t-2}, s_{t-1})|\Psi_{t-1}] \\
&= \sum_{i=1}^{M} \sigma^2_{t-1}(\Psi_{t-2}, s_{t-1} = i)\ p(s_{t-1} = i|\Psi_{t-1}) = \sum_{i=1}^{M} {}^{(SK)}\sigma^2_{i,t-1|t-2}\ p_{i,t-1|t-1}.
\end{aligned}
\tag{5}
$$

Then from the considered model, ${}^{(SK)}\sigma^2_{t-1|t-2}$ results to be a function of Ψ_{t-2} and $s_{t-1} = i$.

2d. Klaassen's Model Finally, Klaassen [18] generalizes the previous auxiliary model including in the conditioning set the information also coming from the current regime s_t. So σ^2_{t-1} turns out to be approximated as

$$
\begin{aligned}
{}^{(K)}\sigma^2_{t-1} &= E[\sigma^2_{t-1}(\Psi_{t-2}, s_{t-1})|\Psi_{t-1}, s_t = j] \\
&= \sum_{i=1}^{M} \sigma^2_{t-1}(\Psi_{t-2}, s_{t-1} = i)\ p(s_{t-1} = i|s_t = j, \Psi_{t-1}) = \sum_{i=1}^{M} {}^{(K)}\sigma^2_{i,t-1|t-2}\ p_{i,t-1|t,t-1}
\end{aligned}
\tag{6}
$$

where $p_{i,t-1|t,t-1}$ is the augmented filtered probability as defined above. Consequently, here ${}^{(K)}\sigma^2_{t-1|t-2}$ becomes a function of Ψ_{t-2} and $s_{t-1} = i$.

3 Filtering and Duality

In order to develop a theory of linear filtering for MS GARCH models, we need to link the model with some state space representations. Here we propose a state space representation and write the associated filter. Consider the model as in (1). For every $s_t = j$ and $s_{t-1} = i$, let us define $\epsilon^2_t = \sigma^2_{j,t} + v_t$, where $\sigma^2_{j,t} = \sigma^2_t(\Psi_{t-1}, s_t = j)$ and $v_t = \sigma^2_{j,t}(u^2_t - 1)$. Then (v_t) is zero mean serially uncorrelated. However, (v_t) is not independent over time since it does not have a constant variance in time (i.e., it is not a homoskedastic process). Now we have

$$
\epsilon^2_t = \sigma^2_{j,t} + v_t = \omega_j + \alpha_j\epsilon^2_{t-1} + \beta_j\sigma^2_{i,t-1} + v_t = \omega_j + \alpha_j\epsilon^2_{t-1} + \beta_j(\epsilon^2_{t-1} - v_{t-1}) + v_t
$$

where ω_j, α_j and β_j are the elements obtained by replacing s_t by j in ω_{s_t}, α_{s_t} and β_{s_t}, respectively. So we can write a new representation of model (1) as

$$
(1 - \delta_j L)\epsilon^2_t = \omega_j + (1 - \beta_j L)v_t \tag{7}
$$

where $\delta_j = \alpha_j + \beta_j$ for $j = 1, \dots, M$. Now, setting $B_t = (\epsilon^2_t \quad v_t)'$, we get

$$
\epsilon^2_t = \omega_j + (\delta_j \quad -\beta_j)\begin{pmatrix}\epsilon^2_{t-1} \\ v_{t-1}\end{pmatrix} + v_t = \omega_j + (\delta_j \quad -\beta_j)B_{t-1} + v_t
$$

for every $j = 1, \dots, M$. In order to simplify notations, let us define

$$y_t = \epsilon_t^2, \quad H = (1 \quad 0), \quad F_{s_t} = \begin{pmatrix} \delta_{s_t} & -\beta_{s_t} \\ 0 & 0 \end{pmatrix}, \quad G = \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \quad \mu_{s_t} = \begin{pmatrix} \omega_{s_t} \\ 0 \end{pmatrix}.$$

Then, for every s_t, we obtain the following switching state space representation:

$$\begin{cases} y_t = HB_t \\ B_t = \mu_{s_t} + F_{s_t} B_{t-1} + Gv_t \end{cases} \tag{8}$$

Representation (8) is similar but different to the switching dynamic model of [19] and [20]. The state vector B_t is called the state (of the system) at time t; in line with [13, p. 576], it can be partially unobservable as it is in this case. In fact, it includes the squared observed returns. Then, conditional on $s_{t-1} = i$ and $s_t = j$, we obtain the following filter:

Prediction

$$\begin{aligned} B_{t|t-1}^{(i,j)} &= \mu_j + F_j B_{t-1|t-1}^{i} \\ P_{t|t-1}^{(i,j)} &= F_j P_{t-1|t-1}^{i} F_j^{'} + GG^{'} \sigma_{vj}^2 \\ \eta_{t|t-1}^{(i,j)} &= y_t - y_{t|t-1}^{(i,j)} = y_t - HB_{t-1|t-1}^{(i,j)} \\ f_{t|t-1}^{(i,j)} &= HP_{t-1|t-1}^{(i,j)} H^{'} \end{aligned}$$

Updating

$$\begin{aligned} B_{t|t}^{(i,j)} &= B_{t|t-1}^{(i,j)} + K_t^{(i,j)} \eta_{t|t-1}^{(i,j)} \\ P_{t|t}^{(i,j)} &= P_{t|t-1}^{(i,j)} - K_t^{(i,j)} HP_{t|t-1}^{(i,j)} \end{aligned}$$

where $\sigma_{vj}^2 = var(v_t|\Psi_{t-1}, s_t = j)$ and $K_t^{(i,j)} = P_{t|t-1}^{(i,j)} H^{'} [f_{t|t-1}^{(i,j)}]^{-1}$ is the Kalman gain

Initial Conditions

$$B_{0|0}^{j} = (I_2 - F_j)^{-1} \mu_j = \begin{pmatrix} (1-\delta_j)^{-1} \omega_j \\ 0 \end{pmatrix}$$

$$\text{vec}(P_{0|0}^{j}) = \sigma_{vj}^2 (I_4 - F_j \otimes F_j)^{-1} \text{vec}(GG^{'}) = \sigma_{vj}^2 \begin{pmatrix} (1-\delta_j^2)^{-1}(1 - 2\delta_j\beta_j + \beta_j^2) \\ 1 \\ 1 \\ 1 \end{pmatrix}$$

$p(s_0 = i) = \pi_i$(steady-state probability).

Explicit derivation of the above filtering procedure is detailed in Appendix A2. Here $Y_{t-1} = \{y_{t-1}, \ldots, y_1\}$ is the information set up to time $t-1$, $B^i_{t|t-1} = E(B_t|Y_{t-1}, s_{t-1} = i)$ is an inference on B_t based on Y_{t-1} given $s_{t-1} = i$; $B^{(i,j)}_{t|t-1} = E(B_t|Y_{t-1}, s_t = j, s_{t-1} = i)$ is an inference on B_t based on Y_{t-1}, given $s_t = j$ and $s_{t-1} = i$; $P^i_{t-1|t-1}$ is the mean squared error matrix of $B^i_{t-1|t-1}$ conditional on $s_{t-1} = i$; $P^{(i,j)}_{t|t-1}$ is the mean squared error matrix of $B^{(i,j)}_{t|t-1}$ conditional on $s_t = j$ and $s_{t-1} = i$; $\eta^{(i,j)}_{t|t-1}$ is the conditional forecast error of y_t based on information up to time $t-1$, given $s_t = j$ and $s_{t-1} = i$; and $f^{(i,j)}_{t|t-1}$ is the conditional variance of forecast error $\eta^{(i,j)}_{t|t-1}$. Each iteration of the Kalman Filter produces an M-fold increase in the number of cases to consider. It is necessary to introduce some approximations to make the filter operable. The key is to collapse the $(M \times M)$ posteriors $B^{(i,j)}_{t|t}$ and $P^{(i,j)}_{t|t}$ into M posteriors $B^j_{t|t}$ and $P^j_{t|t}$. Hence, we mimic the approximation proposed by Kim and Nelson [20] and Kim [19] applied to this state space representation (see Appendix A3). Let $B^j_{t|t}$ be the expectation based not only on Y_t but also conditional on the random variable s_t taking on the value j. Then the approximation results to be

$$B^j_{t|t} = \sum_{i=1}^{M} p_{i,t-1|t,t}\; B^{(i,j)}_{t|t}. \tag{9}$$

To justify the use of this approximation, note that the Kalman filter would give the conditional expectation if, conditional on Ψ_{t-1} and on $s_t = j$, $s_{t-1} = i$, the distribution of B_t is normal. However, the distribution of B_t, conditional on Ψ_{t-1}, $s_t = j$ and $s_{t-1} = i$, is a mixture of normal for $t > 2$. Hence, Kim proposes an approximation in which the exponential Gaussian mixture is collapsed down to M Gaussians at each step. This is a natural proxy for such a process. Having such a convenient switching state space form associated to the initial MS GARCH, gives us the possibility to reconcile in a unique framework the estimation through linear filter or via auxiliary models. Duality exists when we modify the approximation described in (9) with different conditioning sets. The existence of this relationship between the approximated filter and collapsing procedures gives a theoretical ground in using the latter approaches. In fact, these approximated models of the MS GARCH were historically given based only on intuitive arguments. Now, from the measurement equation in (8) and using (9), we have

$$\begin{aligned} y^j_{t|t-1} &= E(y_t|s_t = j, Y_{t-1}) = HE(B_t|s_t = j, Y_{t-1}) \\ &= H\sum_{i=1}^{M} p_{i,t-1|t,t-1}\; B^{(i,j)}_{t|t-1} = \sum_{i=1}^{M} p_{i,t-1|t,t-1}\; y^{(i,j)}_{t|t-1} \end{aligned}$$

and

$$\begin{aligned} y^j_{t-1|t-1} &= E(y_{t-1}|s_t = j, Y_{t-1}) = E(\sigma^2_{t-1}|Y_{t-1}, s_t = j) \\ &= \sigma^2_{j,t-1|t-1} = \sum_{i=1}^{M} p_{i,t-1|t,t-1}\ \sigma^2_{ij,t-1|t-2}. \end{aligned}$$

In particular, if the conditional variance is not a function of $s_t = j$, we get

$$\begin{aligned} y_{t-1|t-1} &= E(\epsilon^2_{t-1}|Y_{t-1}) = E(\sigma^2_{t-1}|Y_{t-1}) \\ &= \sigma^2_{t-1|t-1} = \sum_{i=1}^{M} p_{i,t-1|t,t-1}\ \sigma^2_{i,t-1|t-2} \end{aligned} \tag{10}$$

which coincides with ${}^{(K)}\sigma^2_{t-1}$ in Formula (6). Here ${}^{(K)}\sigma^2_{t-1}$ is only a function of $s_{t-1} = i$. Thus the approximation of the filter is dual to the one used as auxiliary model in [18]. This also means that if we change the conditioning scheme in (10), we obtain other auxiliary models. In fact, if we assume probabilities to be only function of $s_{t-1} = i$ and if still σ^2_{t-1} is a function of s_{t-1}, we have the simplified Klaassen model. This gives the expression in (5), in fact:

$${}^{(SK)}\sigma^2_{t-1} = \sum_{i=1}^{M} {}^{(SK)}\sigma^2_{i,t-1|t-2}\ p_{i,t-1|t-1}\ .$$

Moreover, if we assume instead that σ^2_{t-1} is a function of $s_{t-2} = k$ and also considering prediction probabilities of $s_{t-2} = k$, we get the auxiliary model proposed by Dueker [10]:

$${}^{(D)}\sigma^2_{t-1} = \sum_{i=1}^{M} {}^{(D)}\sigma^2_{k,t-1|t-2}\ p_{k,t-2|t-1}$$

which is Eq. (4). Finally, if we consider the conditioning set up to Y_{t-2} rather than Y_{t-1}, we obtain

$${}^{(G)}\sigma^2_{t-1} = \sum_{i=1}^{M} {}^{(G)}\sigma^2_{i,t-1|t-2}\ p_{i,t-1|t-2}$$

which is Eq. (3) and corresponds to Gray's model. Hence, if we slightly change the conditioning set, we can obtain different specifications of the auxiliary models, moving from the state space form in (8). The usefulness of linking the filter to the four approximations is not only theoretical but also practical, providing a relatively easy method to conduct estimation. In particular, we have showed a direct connection between the filter and Klaassen collapsing procedure. However, an exercise to empirically prove the accuracy of the approximations can be found in [23], where it has been empirically established that Klaassen model is the most effective one in generating consistent estimates for the path-dependent MS GARCH models.

4 An Application on US Treasury Bill Rates

We consider 1-month US Treasury bill rates obtained from FRED for the period January 1970 through April 1994 as in [14]. Figure 1 plots the data in level and in first difference. It is immediate the dramatic increase in interest rates that occurred during the Fed experiment and the OPEC oil crisis, which leads us to consider a 2-regime model. Then we fit a 2-state MS GARCH model as in (1) with both changes in regimes in the intercept term and in the persistence parameters of the volatility process. The values of the estimation are reported in Table 1 where estimated values along with robust standard errors are reported. In particular, the model estimated by filtering mimic with Kim and Nelson's (KN) approximation is labelled with Approximation 1 and it is dual to Klaassen (K). Note that Approximation 2 reproduces the auxiliary model of Gray (G) and values are in fact very close (see [14, Table 3, p. 44]). Finally, Approximation 3 and 4 are respectively dual to Dueker (D) and the simplified Klaassen (SK) models. The high-volatility regime is characterized by more sensitivity to recent shocks ($\hat{\alpha}_2 > \hat{\alpha}_1$) and less persistence ($\hat{\beta}_2 < \hat{\beta}_1$) than the low-volatility regime. Within each regime, the GARCH processes are stationary ($\hat{\alpha}_i + \hat{\beta}_i < 1$) and the parameter estimates suggest that the regimes are very persistent, so the source of volatility persistence will be important. In

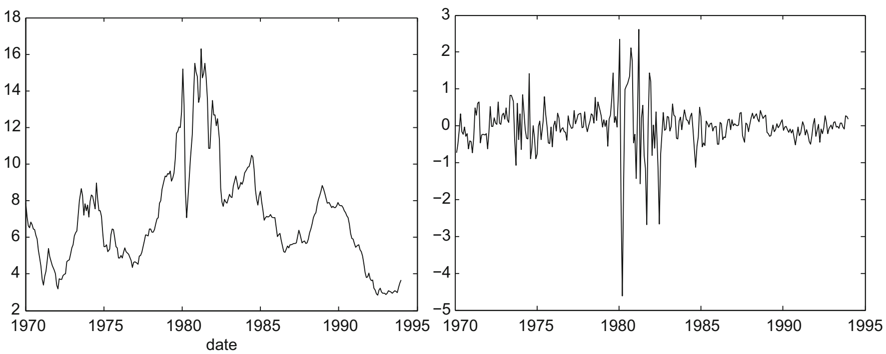

Fig. 1 The left panel contains a time series plot of 1-month US Treasury bill rates (in annualized percentage term). The sample period is from January 1970 to April 1994; a total of 1267 observations. First differences of the series are shown in the right panel. The data are obtained from FRED database

Table 1 Estimation of the parameters for 2-state MS GARCH model as in (1)

Approximations	$\hat{p}$	$\hat{q}$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\alpha}_1$	$\hat{\alpha}_2$	$\hat{\omega}_1$	$\hat{\omega}_2$
Approximation 1 (KN = K)	0.8467 (0.1402)	0.9982 (0.0334)	0.4113 (0.0407)	0.0085 (0.0927)	0.0184 (0.0338)	0.4967 (0.0331)	0.023 (0.2626)	0.068 (0.0908)
Approximation 2 (=G)	0.8018 (0.1146)	0.9157 (0.0157)	0.391 (0.1624)	0.0062 (0.2609)	0.0203 (0.0640)	0.4801 (0.1089)	0.045 (0.3591)	0.0713 (0.2853)
Approximation 3 (=D)	0.8467 (0.1406)	0.9983 (0.0336)	0.4112 (0.0420)	0.0086 (0.0923)	0.0184 (0.0342)	0.4967 (0.0337)	0.023 (0.2651)	0.068 (0.0799)
Approximation 4 (=SK)	0.8119 (0.1140)	0.9160 (0.0157)	0.397 (0.1629)	0.0064 (0.2618)	0.0212 (0.0642)	0.4831 (0.1092)	0.039 (0.3603)	0.0692 (0.2867)

Robust standard errors in parenthesis. $\hat{p}$ and $\hat{q}$ are the elements on the main diagonal of the transition probability matrix. The observables are 1-month US Treasury bill rates (in annualized percentage term). The sample period is from January 1970 to April 1994; a total of 1267 observations. The data are obtained from FRED database.

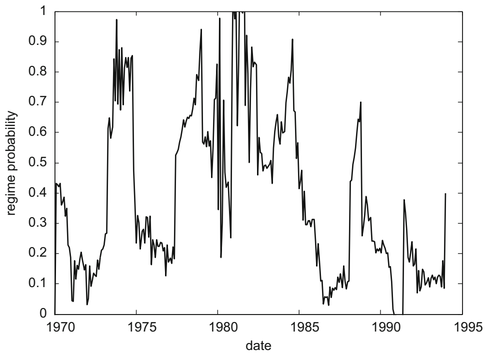

Fig. 2 The panel shows MS GARCH smoothed probabilities of being in the high-volatility regime. Parameter estimates are based on a dataset of 1-month Treasury Bill rates, reported in annualized percentage terms. The sample period is from January 1970 to April 1994; a total of 1267 observations. The data are obtained from FRED database

general, the four approximations are not very dissimilar to the others and the filtering algorithm has a very limited computational cost. Figure 2 plots smoothed probabilities $Pr(s_t = 1|\Psi_T)$ which are of interest to determine if and when the regime switching occurs. The plot of smoothed probabilities manages to identify crises periods that affected market indices. In particular, we could recognize three periods of high-variance. The first (1973–1975) corresponds to the OPEC oil crisis. The second is shorter and corresponds to the Fed experiment (1979–1983). The third is a short period around 1987, after the stock market crash.

5 Conclusions

We deal with estimation of Markov Switching GARCH models. It is well-known that these models suffer of path-dependence, i.e., dependence of the entire path history of the data which makes Maximum Likelihood procedures unfeasible to apply. Therefore, we introduce filtering procedures based on approximated algorithms for switching state space representations. We show duality results in the estimation by approximated filtering method and auxiliary models proposed in the literature. Our filtering algorithm is readily programmable and with a very limited computational cost. An application on financial data shows the feasibility of the proposed approach.

Appendix

A1 We show that $p_{t|t-1,t} = p(s_t|s_{t-1}, \Psi_t)$ can be expressed in terms of $p_{t|t-1,t-1}$ and the conditional density of ϵ_t which depends on the current regime s_t and the past regimes, i.e, $f(\epsilon_t|s_1, \ldots, s_t, \Psi_{t-1})$. In fact,

$$\begin{aligned} p_{t|t-1,t} &= p(s_t|s_{t-1}, \Psi_t) = p(s_t|s_1, \ldots, s_{t-1}, \Psi_t) = p(s_t|s_1, \ldots, s_{t-1}, \epsilon_t, \Psi_{t-1}) \\ &= \frac{f(\epsilon_t|s_1, \ldots, s_t, \Psi_{t-1}) p(s_t|s_1, \ldots, s_{t-1}, \Psi_{t-1})}{f(\epsilon_t|s_1, \ldots, s_{t-1}, \Psi_{t-1})} \\ &= \frac{f(\epsilon_t|s_1, \ldots, s_t, \Psi_{t-1}) p(s_t|s_{t-1}, \Psi_{t-1})}{f(\epsilon_t|s_1, \ldots, s_{t-1}, \Psi_{t-1})} = \frac{f(\epsilon_t|s_1, \ldots, s_t, \Psi_{t-1}) p_{t|t-1,t-1}}{f(\epsilon_t|s_1, \ldots, s_{t-1}, \Psi_{t-1})} \end{aligned}$$

where $f(\epsilon_t|s_1, \ldots, s_{t-1}, \Psi_{t-1}) = \sum_{s_t=1}^{M} f(\epsilon_t|s_1, \ldots, s_t, \Psi_{t-1}) p(s_t|s_{t-1}, \Psi_{t-1})$ $= \sum_{s_t=1}^{M} f(\epsilon_t|s_1, \ldots, s_t, \Psi_{t-1})\, p_{t|t-1,t-1}$.

A2 We present explicit derivation of the filter for MS GARCH as given in Sect. 3. The prediction step is obtained as follows

$$\begin{aligned} B^{(i,j)}_{t|t-1} &= E(B_t|\Psi_{t-1}, s_t = j, s_{t-1} = i) = E(\mu_{s_t} + F_{s_t} B_{t-1} + G v_t|\Psi_{t-1}, s_t = j, s_{t-1} = i) \\ &= \mu_j + F_j E(B_{t-1}|\Psi_{t-1}, s_t = j, s_{t-1} = i) = \mu_j + F_j B^i_{t-1|t-1}. \end{aligned}$$

In particular, we have $B_t - B^{(i,j)}_{t|t-1}|_{s_t=j} = F_j(B_{t-1} - B^i_{t-1|t-1}) + G v_t$. Then

$$\begin{aligned} P^{(i,j)}_{t|t-1} &= E[(B_t - B^{(i,j)}_{t|t-1})(B_t - B^{(i,j)}_{t|t-1})^{'}|\Psi_{t-1}, s_t = j, s_{t-1} = i] \\ &= E[(F_j(B_{t-1} - B^i_{t-1|t-1}) + G v_t)(F_j(B_{t-1} - B^i_{t-1|t-1}) + G v_t)^{'}|\Psi_{t-1}, s_t = j, s_{t-1} = i] \\ &= F_j E[(B_{t-1} - B^i_{t-1|t-1})(B_{t-1} - B^i_{t-1|t-1})^{'}|\Psi_{t-1}, s_{t-1} = i] F_j^{'} + G E(v_t^2|s_t = j) G^{'} \\ &= F_j P^i_{t-1|t-1} F_j^{'} + G G^{'} \sigma^2_{vj} \end{aligned}$$

and

$$\begin{aligned} \eta^{(i,j)}_{t|t-1} &= y_t - y^{(i,j)}_{t|t-1} = y_t - E(y_t|\Psi_{t-1}, s_t = j, s_{t-1} = i) = y_t - E(H B_t|\Psi_{t-1}, s_t = j, s_{t-1} = i) \\ &= y_t - H E(B_t|\Psi_{t-1}, s_t = j, s_{t-1} = i) = y_t - H B^{(i,j)}_{t|t-1}. \end{aligned}$$

Hence, $\eta^{(i,j)}_{t|t-1}|_{\Psi_{t-1}, s_t=j, s_{t-1}=i} = H(B_t - B^{(i,j)}_{t|t-1}) + v_t$ and

$$\begin{aligned} f^{(i,j)}_{t|t-1} &= E[(\eta^{(i,j)}_{t|t-1})^2|\Psi_{t-1}, s_t = j, s_{t-1} = i)] \\ &= E[(H(B_t - B^{(i,j)}_{t|t-1}))(H(B_t - B^{(i,j)}_{t|t-1}))^{'}|\Psi_{t-1}, s_t = j, s_{t-1} = i] \\ &= H E[(B_t - B^{(i,j)}_{t|t-1})(B_t - B^{(i,j)}_{t|t-1})^{'}|\Psi_{t-1}, s_t = j, s_{t-1} = i] H^{'} = H P^{(i,j)}_{t|t-1} H^{'}. \end{aligned}$$

Furthermore, the updating step is derived as follows. Define $Z_1 = B_t$ and $Z_2 = \eta^{(i,j)}_{t|t-1} = y_t - y^{(i,j)}_{t|t-1}$. Then $\mu_1 = E[Z_1|\Psi_{t-1}, s_t = j, s_{t-1} = i] = B^{(i,j)}_{t|t-1}$, $\mu_2 = E[Z_2|\Psi_{t-1}, s_t = j, s_{t-1} = i] = 0$, $\Sigma_{11} = P^{(i,j)}_{t|t-1}$ and $\Sigma_{22} = f^{(i,j)}_{t|t-1}$. We have

$$\begin{aligned}\Sigma_{12} &= cov(Z_1, Z_2) = E[(B_t - B^{(i,j)}_{t|t-1})\eta^{(i,j)}_{t|t-1}|\Psi_{t-1}, s_t = j, s_{t-1} = i] \\ &= E[(B_t - B^{(i,j)}_{t|t-1})(B_t - B^{(i,j)}_{t|t-1})' H' |\Psi_{t-1}, s_t = j, s_{t-1} = i] = P^{(i,j)}_{t|t-1} H' = \Sigma_{21}'.\end{aligned}$$

Thus $Z_1|_{Z_2,\Psi_{t-1},s_t=j,s_{t-1}=i}$ is given by $\mu_{1|2} = \mu_1 + \Sigma_{12}\Sigma_{22}^{-1}(Z_2 - \mu_2)$, that is, $B^{(i,j)}_{t|t} = B^{(i,j)}_{t|t-1} + P^{(i,j)}_{t|t-1} H' [f^{(i,j)}_{t|t-1}]^{-1} \eta^{(i,j)}_{t|t-1}$. Further, we have $\Sigma_{11|2} = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}$, hence $P^{(i,j)}_{t|t} = P^{(i,j)}_{t|t-1} - K^{(i,j)}_t H P^{(i,j)}_{t|t-1}$, where $K^{(i,j)}_t = P^{(i,j)}_{t|t-1} H' [f^{(i,j)}_{t|t-1}]^{-1}$ is the Kalman gain.

A3 Here we derive the approximation on the line of [20] applied to model in (8), which is Eq. (9):

$$\begin{aligned} B^j_{t|t} &= \frac{\sum_{i=1}^M B^{(i,j)}_{t|t} p(s_{t-1} = i, s_t = j|Y_t)}{p(s_t = j|Y_t)} = \sum_{i=1}^M \frac{p(s_{t-1} = i, s_t = j|Y_t)}{p(s_t = j|Y_t)} B^{(i,j)}_{t|t} \\ &= \sum_{i=1}^M p(s_{t-1} = i|s_t = j, Y_t)\ B^{(i,j)}_{t|t} = \sum_{i=1}^M p_{i,t-1|t,t}\ B^{(i,j)}_{t|t}. \end{aligned}$$

References

1. Augustyniak, M.: Maximum likelihood estimation of the Markov-switching GARCH model. Comput. Stat. Data Anal. **76**, 61–75 (2014)
2. Augustyniak, M., Boudreault, M., Morales, M.: Maximum likelihood estimation of the Markov-switching GARCH model based on a general collapsing procedure. Methodol. Comput. Appl. Probab. 1–24 (2017). https://doi.org/doi:10.1007/s11009-016-9541-4
3. Bauwens, L., Preminger, A., Rombouts, J.: Theory and inference for a Markov switching GARCH model. Economet. J. **13**(2), 218–244 (2010)
4. Bauwens, L., Dufays, A., Rombouts, J.V.: Marginal likelihood for Markov-switching and change-point GARCH models. J. Economet. **178**, 508–522 (2014)
5. Billio, M., Monfort, A., Robert, C.P.: A MCMC approach to maximum likelihood estimation. In: Huskova, M., Lachout, P., Visek, J.A. (eds.) Prague Stochastics '98, vol. 1, pp. 49–54. Union of Czach Mathematicians and Physicists, Prague (1998)
6. Billio, M., Monfort, A., Robert, C.P.: The simulated likelihood ratio (SLR) method. Document de Travail du CREST 9828, Paris (1998)
7. Billio, M., Casarin, R., Osuntuyi, A.: Efficient Gibbs sampling for Markov switching GARCH models. Comput. Stat. Data Anal. **100**, 37–57 (2016)
8. Bollerslev, T.: Generalized autoregressive conditional heteroskedasticity. J. Economet. **52**, 5–59 (1986)
9. Cai, J.: A Markov model of unconditional variance in ARCH. J. Bus. Econ. Stat. **12**, 309–316 (1994)

10. Dueker, M.J.: Markov switching in GARCH processes and mean reverting stock market volatility. J. Bus. Econ. Stat. **15**(1), 26–34 (1997)
11. Francq, C., Zakoïan, J.M.: The L^2-structures of standard and switching-regime GARCH models. Stoch. Process. Appl. **115**, 1557–1582 (2005)
12. Francq, C., Roussignol, M., Zakoïan, J.M.: Conditional heteroskedasticity driven by hidden Markov chains. J. Time Ser. Anal. **22**(2), 197–220 (2001)
13. Gourieroux, C., Monfort, A.: Time Series and Dynamic Models. Cambridge University Press, Cambridge (1997)
14. Gray, S.F.: Modelling the conditional distribution of interest rates as a regime-switching process. J. Financ. Econ. **42**, 27–62 (1996)
15. Haas, M., Mittnik, S., Paolella, M.S.: A new approach to Markov-switching GARCH models. J. Financ. Economet. **2**, 493–530 (2004)
16. Hamilton, J.D., Susmel, R.: Autoregressive conditional heteroskedasticity and changes in regime. J. Economet. **64**, 307–333 (1994)
17. Henneke, J.S., Rachev, S.T., Fabozzi, F.J., Metodi, N.: MCMC-based estimation of Markov switching ARMA–GARCH models. Appl. Econ. **43**(3), 259–271 (2011)
18. Klaassen, F.: Improving GARCH volatility forecasts with regime-switching GARCH. Empir. Econ. **27**(2), 363–394 (2002)
19. Kim, C.J.: Dynamic linear models with Markov switching. J. Economet. **64**, 1–22 (1994)
20. Kim, C.J., Nelson, C.R.: State-Space Models with Regime Switching: Classical and Gibbs-Sampling Approaches with Applications. The MIT Press, Cambridge (1999)
21. Lamoureux, C.G., Lastrapes, W.D.: Persistence in variance, structural change, and the GARCH model. J. Bus. Econ. Stat. **8**, 225–234 (1990)
22. Nelson, D.B.: Stationarity and persistence in the GARCH(1,1) model. Economet. Theory **6**, 318–344 (1990)
23. Osuntuy, A.A.: Essays on bayesian inference with financial applications. PhD thesis, University Ca' Foscari of Venice (2014)

A Network Approach to Risk Theory and Portfolio Selection

Roy Cerqueti and Claudio Lupi

Abstract In the context of portfolio theory, the evaluation of risk is of paramount relevance. In this respect, the connections among the risky assets of the portfolio should be carefully explored. This paper elaborates on this topic. We define a portfolio through a network, whose nodes are the assets composing it. The weights on the nodes and the arcs represent the share of capital invested on the assets and the dependence among them, respectively. The risk profile of the portfolio will be given through a suitably defined risk measure on the portfolio-network. The standard Markowitz theory will be rewritten in this particular setting. Surprisingly, we will note that the resulting decision problem is not consistent with an adapted version of the axiomatization of the standard expected utility theory.

1 Introduction

Since its inception in [7], portfolio theory has attracted the academic debate not only for its potential applications, but also for the space left for theoretical improvements. The original model is grounded on several restrictive assumptions such as uniperiodal setting, Gaussian returns, risks measured by variances and expected returns represented by historical means, absence of transaction costs and of default risks. In the following years, Markowitz's framework has been extended under several respects. One of the most notable improvements of the original formulation can be found in [9, 10], where continuous time has been introduced, and in [12], where the model has been extended by introducing a multiperiodal setting. Among recent extensions, it is interesting for this paper to mention [5], that provides the reinterpretation of the mean-variance model in the context of semi-copulas; [11] adopts a topological perspective for highlighting the relationship between portfolio

R. Cerqueti (✉)
Department of Economics and Law, University of Macerata, Via Crescimbeni, 20, I-62100 Macerata, Italy
e-mail: roy.cerqueti@unimc.it

C. Lupi
Department of Economics, University of Molise, Via F. De Sanctis, I-86100 Campobasso, Italy
e-mail: lupi@unimol.it

M. Corazza et al. (eds.), *Mathematical and Statistical Methods for Actuarial Sciences and Finance*, https://doi.org/10.1007/978-3-319-50234-2_6

and Euclidean distance in the Markowitz model; [2, 3] present a mixed discrete-continuous time mean-variance model in the presence of infrequently traded assets. An extensive survey on 60 years of research on the mean-variance model has been outlined by its pioneer (see [8]).

In this paper we further proceed along this research path and provide an extension of the Markowitz model by including in the model formulation the structure of the connections among the assets. For this purpose, we move from [4] and identify portfolios with networks, whose nodes are the assets.[1]

Furthermore, we discuss the consistency of the Markowitz theory with the axiomatization of the expected utility. We adopt the theoretical framework developed in [4], and argue that the goodness of a risk-minimization criterion for ordering preferences should be not necessarily in accord to expected utility. In so doing, we are exactly in line with the philosophical proposal of [1], who show that commonly used risk acceptance criteria do not agree with expected utility.

The original contributions of the paper are the following: first, we present a novel formulation of a classical portfolio problem in the language of networks, hence including explicitly the connections among the risky assets; second, we adopt a recent axiomatization of the expected utility theory in the language of the networks (see [4]) and show that there exists a parametrization of the mean-variance utility function which is not consistent with such a parametrization.

In general, we feel in line with [6], who discuss risk minimization problems by explicitly modelling also the topological and the stochastic dependence structure of the involved decision variables. In particular, to the best of our knowledge, this is the first attempt to model Markowitz theory in the language of networks.

The rest of the paper is organized as follows: the next section introduces the portfolios we deal with. Section 3 outlines the concept of risk measure, along with a needed equivalence relation over the set of portfolios. Section 4 contains the reinterpretation of Markowitz's mean-variance utility theory in our specific setting, and the analysis of the consistency of the associated preference criterion with expected utility. The last section offers some concluding remarks. To be as self-contained as possible, we have reformulated the expected utility axiomatization presented in [4] in our framework. Such a reformulation is contained in the Appendix.

2 The Set of Portfolios

We propose here a very general setting of a portfolio model: such a general setting will then be adapted to the Markowitz model.

First, we introduce a probability space $(\Omega, \mathscr{F}, \mathbb{P})$ for all the random variables used in this paper. These random variables are collected in the set $\mathscr{A}$.

[1] For a survey on networks, see [13, 14].

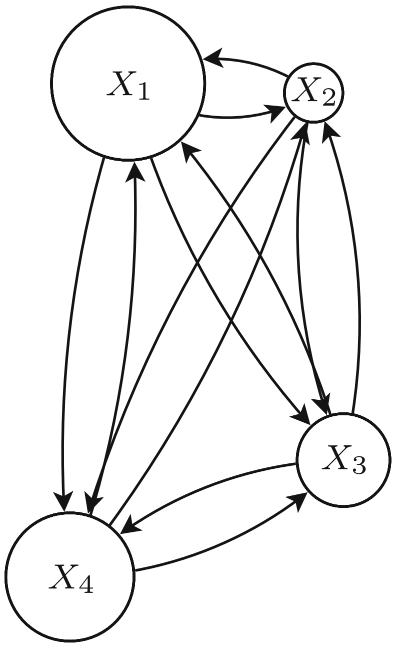

Fig. 1 Network representation of a four-asset portfolio. The size of each node is proportional to ρ_j; the length ($\delta_{i \to j}$) of each arc is inversely proportional to the dependence among the assets; $v(i,j) = 1 \ \forall i \neq j$

Portfolios are assumed to be formed by n assets, whose random returns are the elements of a subset $\mathscr{S}$ of $\mathscr{A}$ as follows:

$$\mathscr{S} = \{X_1, X_2, \ldots, X_n\} \subseteq \mathscr{A}. \tag{1}$$

The set $\mathscr{S}$ is identified as the set of the nodes (vertices) of a weighted oriented graph (see Fig. 1).

Weights of nodes and arcs are formalized as follows:

- we define a function $\rho : \mathscr{S} \to \mathbb{R}$ such that $\rho(X_j) = \rho_j$ represents the weight of X_j, for each $j = 1, 2, \ldots, n$;
- we introduce a binary variable which certifies the existence of an arc:

$$v(i,j) = \begin{cases} 1, & \text{if there exists the oriented arc connecting } i \text{ and } j, \\ 0, & \text{otherwise;} \end{cases}$$

  for each $i, j = 1, 2, \ldots, n$;
- we introduce a real number $\delta_{i \to j} \in \mathbb{R}$ representing the weight of the oriented arc from X_i to X_j, for each $i, j = 1, 2, \ldots, n$, $i \neq j$.

A portfolio $\boldsymbol{\pi}$ is then formalized by the knowledge of its nodes and of functions ρ's, v's and δ's, so that:

$$\boldsymbol{\pi} = (\mathscr{S}, \rho, v, \delta), \tag{2}$$

where $\rho = \{\rho_j\}_{j=1,2,\ldots,n}$, $v = \{v(i,j)\}_{i,j=1,2,\ldots,n}$ and $\delta = \{\delta_{i\to j}\}_{i,j=1,2,\ldots,n}$.

The space of the portfolios is then given by

$$\mathbf{P} = \left\{(\mathscr{S}, \rho, v, \delta) \in \mathscr{P}(\mathscr{A}) \times \mathbb{R}^{|\mathscr{S}|} \times \{0,1\}^{|\mathscr{S}|^2} \times \mathbb{R}^{|\mathscr{S}|^2}\right\},$$

where $\mathscr{P}(\star)$ is the usual notation for the set of the parts of $\star$.

The presence of the shares of capital invested in the assets of $\mathscr{S}$ is implicitly included in the definition of the weights ρ's and δ's. We will see this in details while discussing the Markowitz model.

3 Risk Measure of Portfolio

In order to define risk measures, we need to introduce the concept of equivalence of portfolios. Indeed, as we will see, a consistent risk measure should assign the same value to portfolios with some specific characteristics.

We adopt the framework of [4], here adapted to our specific context.

Definition 1 Consider two portfolios

$$\boldsymbol{\pi}_k = \left(\mathscr{S}^{(k)}, \rho^{(k)}, v^{(k)}, \delta^{(k)}\right) \in \mathbf{P}, \qquad k = 1, 2.$$

We say that portfolio $\boldsymbol{\pi}_1$ is equivalent to portfolio $\boldsymbol{\pi}_2$—i.e.: $\boldsymbol{\pi}_1 \equiv \boldsymbol{\pi}_2$—when one of the following conditions is satisfied:

(i) $\left(\mathscr{S}^{(1)}, \rho^{(1)}, v^{(1)}, \delta^{(1)}\right) = \left(\mathscr{S}^{(2)}, \rho^{(2)}, v^{(2)}, \delta^{(2)}\right)$;

(ii) if $1 = v^{(k_1)}(i,j) \neq v^{(k_2)}(i,j) = 0$, then $\delta^{(k_1)}_{i\to j} = 0$, for each i,j indices of the nodes of $\mathscr{S}^{(k_1)}$ and $k_1, k_2 = 1, 2$ with $k_1 \neq k_2$, all other things being equal;

(iii) if $X_j \in \mathscr{S}^{(k_1)} \setminus \mathscr{S}^{(k_2)}$, then $\rho^{(k_1)}_j = 0$ and $\delta^{(k_1)}_{i\to j} = \delta^{(k_1)}_{j\to i} = 0$, for each i index of the nodes in $\mathscr{S}^{(k_1)} \cup \mathscr{S}^{(k_2)}$ and $k_1, k_2 = 1, 2$ with $k_1 \neq k_2$, all other things being equal.

Definition 1 has a very relevant interpretation. In fact, once a portfolio $\boldsymbol{\pi}$ is selected, then it is possible to create another portfolio $\widetilde{\boldsymbol{\pi}}$ which is equivalent to $\boldsymbol{\pi}$. To do this, one can add other assets, labeled with $j_1, \ldots, j_k$. Any new asset must be associated to null weights, so that $\rho_j = 0$ and $\delta_{i\to j} = \delta_{j\to i} = 0$, for each $j = j_1, \ldots, j_k$ and for each i-th node of $\widetilde{\boldsymbol{\pi}}$. The equivalence is then of paramount relevance, because it allows us to consider homogeneous portfolios sharing the same set of nodes. Furthermore, under the condition of imposing some further null weights, it is also not restrictive to assume that all the couples of assets of the portfolio are associated to two oriented arcs. Indeed, it is equivalent to consider $v(i,j) = 0$ or, alternatively, $v(i,j) = 1$ and $\delta_{i\to j} = \delta_{j\to i} = 0$.

In order to deal with the elements of the set $\mathbf{P}$, we need to introduce the operators allowing to combine portfolios. In particular, the following definitions formalize the direct sum of two portfolios and the product of a portfolio for a scalar.

Definition 2 Consider $\boldsymbol{\pi}_1, \boldsymbol{\pi}_2 \in \mathbf{P}$ such that $\boldsymbol{\pi}_k \equiv \left(\mathscr{S}^{(k)}, \rho^{(k)}, v^{(k)}, \delta^{(k)}\right)$ for $k = 1, 2$. The direct sum of $\boldsymbol{\pi}_1$ and $\boldsymbol{\pi}_2$ is

$$\boldsymbol{\pi} = \boldsymbol{\pi}_1 \oplus \boldsymbol{\pi}_2 \in \mathbf{P}, \tag{3}$$

where $\boldsymbol{\pi} \equiv (\mathscr{S}, \rho, v, \delta)$ with $\mathscr{S} = \mathscr{S}^{(1)} \cup \mathscr{S}^{(2)}$ and, for each $i,j = 1, \dots, n$, it is $\rho_j = \rho_j^{(1)} + \rho_j^{(2)}$, $\delta_{i\to j} = \delta_{i\to j}^{(1)} + \delta_{i\to j}^{(2)}$ and $v(i,j) = \max\{v^{(1)}(i,j), v^{(2)}(i,j)\}$.

Definition 3 Consider $\boldsymbol{\pi} \equiv (\mathscr{S}, \rho, v, \delta) \in \mathbf{P}$ and a scalar $\alpha \in \mathbb{R}$.

The product scalar-network $\alpha \cdot \boldsymbol{\pi}$ is a new network $\boldsymbol{\pi}_\alpha \equiv \left(\mathscr{S}, \rho^{(\alpha)}, v, \delta^{(\alpha)}\right)$ such that $\rho_j^{(\alpha)} = \alpha \cdot \rho_j$ and $\delta_{i\to j}^{(\alpha)} = \alpha \cdot \delta_{i\to j}$, for each $i,j = 1, \dots, n$.
The topological structure of the set $\mathbf{P}$—endowed with the binary operator $\oplus$—assures that the set of portfolios is quite rich. We refer to [4] for an extensive analysis of this aspect.

We maintain the definition of risk measure of a portfolio as general as possible.

Definition 4 A risk measure of a portfolio is a function $\mu : \mathbf{P} \to \mathbb{R}$ such that:

- If $\boldsymbol{\pi}_1 \equiv \boldsymbol{\pi}_2$, then $\mu(\boldsymbol{\pi}_1) = \mu(\boldsymbol{\pi}_2)$;
- μ provides a total order $\preceq_\mu$ over the set $\mathbf{P}$ as follows:

$$\boldsymbol{\pi}_1 \preceq_\mu \boldsymbol{\pi}_2 \qquad \text{if and only if} \qquad \mu(\boldsymbol{\pi}_1) \le \mu(\boldsymbol{\pi}_2),$$

  for each $\boldsymbol{\pi}_1, \boldsymbol{\pi}_2 \in \mathbf{P}$;
- $\mu(\boldsymbol{\pi}_1) < \mu(\boldsymbol{\pi}_2)$ means that $\boldsymbol{\pi}_1$ is less risky—in the sense of $\preceq_\mu$—than $\boldsymbol{\pi}_2$.

In [4] the interested reader can find a detailed discussion of this type of risk measures, along with some examples in the general context of networks. Moreover, the quoted paper contains also the analysis of the consistency of this kind of risk measures with respect to a reformulation of the expected utility theory: for completeness, we offer in the Appendix a brief account of the main results.

It is important to note that the risk measures of a portfolio could be viewed as equivalence classes rather than single elements. This aspect will be of remarkable relevance in the discussion of our variant of the Markowitz model and the consistency of the associated preference order with the expected utility axioms (see the next section). Thus, we here present a concept of equivalence for risk measures of portfolios, in agreement with the general scheme proposed in [4]:

Definition 5 Consider two risk measures of portfolios $\mu_k : \mathbf{P} \to \mathbb{R}$, with $k = 1, 2$. μ_1 is equivalent to μ_2—and we indicate $\mu_1 \equiv \mu_2$—if and only if

$$\mu_1(\boldsymbol{\pi}_1) \leq \mu_1(\boldsymbol{\pi}_2) \qquad \text{if and only if} \qquad \mu_2(\boldsymbol{\pi}_1) \leq \mu_2(\boldsymbol{\pi}_2),$$

for each $\boldsymbol{\pi}_1, \boldsymbol{\pi}_2 \in \mathbf{P}$.
Notice that the equivalence instituted by Definition 5 involves how portfolios are ranked by the risk measure, and not how they are objectively evaluated. Substantially, equivalence is associated to the order of the portfolios, even in presence of very different scales of evaluation.

4 Rewriting the Markowitz Model

This section is devoted to the proposal of a rewriting of the standard Markowitz model. The aim is twofold: first, we pursue the scope of applying the network theoretical framework presented above to the classical portfolio theory; second, we show that risk measures of portfolios in Markowitz theory could also be not consistent with the expected utility axiomatization. More than this, we show that two equivalent measures could behave differently in this respect. Such a finding suggests a substantial independence of the family of risk acceptance criteria with the axiomatization of Von Neumann and Morgerstern.

Suppose that the shares of capital invested on $X_1, X_2, \ldots, X_n$ are $x_1, x_2, \ldots, x_n$, respectively.

The general formulation of the optimal portfolio problem is:

$$\begin{cases} \min\limits_{x_1,\ldots,x_n} & \mu(\boldsymbol{\pi}); \\ s.t., & \sum_{j=1}^n x_j = 1, \end{cases}$$

where μ is the risk measures of portfolios. See below for two proposals for μ.

4.1 First Model

Moving from the definition of portfolios provided in Sect. 2, we identify the weights associated to Markowitz's mean-variance utility case.

The weights of the (oriented) arcs are given by:

$$\delta_{i \to j} = \delta_{j \to i} = x_i x_j \mathbb{C}\left[X_i, X_j\right], \qquad \forall\, i, j = 1, \ldots, n,\ i \neq j,$$

where $\mathbb{C}$ is the covariance operator. If $i = j$, then

$$\delta_{j\to j} = x_j \sqrt{\mathbb{V}\left[X_j\right]}.$$

The binary variables are redundant, in the sense that $v(i,j) = 1$ for each $i,j = 1,\ldots,n$.

Functions ρ's are given by expected values of the returns of the assets weighted for the related shares of the portfolio, so that:

$$\rho_j = x_j \mathbb{E}\left[X_j\right], \qquad \forall\ j = 1,\ldots,n.$$

The mean-variance utility can be written as (is equivalent to) $\mu : \mathbf{P} \to \mathbb{R}$ such that:

$$\mu_1(\boldsymbol{\pi}) \equiv -\sum_{j=1}^{n} \rho_j + a \cdot \left\{ \sum_{j=1}^{n} v(j,j) \left[\delta_{j\to j}\right]^2 + 2 \sum_{i<j} v(i,j) \delta_{i\to j} \right\}. \tag{4}$$

4.2 Second Model

In this case, the weights of the (oriented) arcs are given by:

$$\delta_{i\to j} = \delta_{j\to i} = x_i x_j \mathbb{C}\left[X_i, X_j\right], \qquad \forall\ i,j = 1,\ldots,n,\ i \neq j,$$

where $\mathbb{C}$ is the covariance operator. If $i = j$, then

$$\delta_{j\to j} = x_j^2 \mathbb{V}\left[X_j\right].$$

As in the previous case, $v(i,j) = 1$ for each $i,j = 1,\ldots,n$.

Functions ρ's are given by expected values of the returns of the assets weighted for the related shares of the portfolio, so that:

$$\rho_j = x_j \mathbb{E}\left[X_j\right], \qquad \forall\ j = 1,\ldots,n.$$

The mean-variance utility can be written as (is equivalent to) $\mu : \mathbf{P} \to \mathbb{R}$ such that:

$$\mu_2(\boldsymbol{\pi}) \equiv -\sum_{j=1}^{n} \rho_j + a \cdot \left\{ \sum_{j=1}^{n} v(j,j) \delta_{j\to j} + 2 \sum_{i<j} v(i,j) \delta_{i\to j} \right\}. \tag{5}$$

4.3 Some Results

It is easy to check that the following proposition is true:

Proposition 1 *The risk measures of portfolios μ_1 and μ_2 are equivalent.*

Observe that the preference order induced by $\mu_1 : \mathbf{P} \to \mathbb{R}$ in (4) is not consistent with the expected utility theory. This is true by Cerqueti and Lupi [4, Proposition 3], since μ_1 is nonlinear with respect to $\delta_{i \to j}$ and depends also on ρ_j.

On the contrary, a simple computation gives that $\mu_2 : \mathbf{P} \to \mathbb{R}$ in (5) is consistent with the expected utility axiomatization.

5 Conclusions

In this paper we present a rewriting of the classical mean-variance portfolio theory in the context of networks. With this aim, we have proposed a definition of the set of portfolios in the context of networks. The operators acting on such a set have been introduced, to allow the analysis of the portfolio theory in this specific context. Moreover, we have developed two portfolio models and created a suitable risk measure on portfolios. One of the considered risk measure is not consistent with the expected utility axiomatization, as reinterpreted by Cerqueti and Lupi [4]. This outcome is in line with a recent strand of literature (see [1, 4]), and suggests further developments to classify in this respect the main risk measures used in portfolio theory.

Appendix: A Reformulation of the Expected utility Axiomatization

Cerqueti and Lupi [4] offer a reformulation of the standard expected utility axiomatization as proposed, e.g., by Abrahamsen and Aven [1]. In this Appendix we collect for completeness the expected utility axioms expressed in terms of networks—which are now our portfolios—as proposed in [4, Section 4].

1. **Weak order**—Preferences are: (1) *complete*, i.e. the decider can state whether two portfolios are equivalent or whether one is preferred to the other; (2) *transitive*, i.e. given three portfolios $\boldsymbol{\pi}_1, \boldsymbol{\pi}_2, \boldsymbol{\pi}_3$, if $\boldsymbol{\pi}_1$ is preferred to $\boldsymbol{\pi}_2$ and $\boldsymbol{\pi}_2$ is preferred to $\boldsymbol{\pi}_3$, then $\boldsymbol{\pi}_1$ is preferred to $\mathbf{N}_3$; (3) *reflexive*, i.e. the decision-maker is indifferent between two equivalent portfolios.
2. **Continuity**—Given three different portfolios $\boldsymbol{\pi}_1, \boldsymbol{\pi}_2, \boldsymbol{\pi}_3$ such that $\boldsymbol{\pi}_1$ is preferred to $\boldsymbol{\pi}_2$ and $\boldsymbol{\pi}_2$ is preferred to $\boldsymbol{\pi}_3$, then there exists a number $p \in (0, 1]$ such that the decider is indifferent between the compounded portfolio $p \cdot \boldsymbol{\pi}_1 \oplus (1-p) \cdot \boldsymbol{\pi}_3$ and $\boldsymbol{\pi}_2$.

3. **Preference increasing with probabilities** (and with connections and weights)—Consider $\mathscr{S}^{(1)} = \{X_1^{(1)}, X_2\}$ and $\mathscr{S}^{(2)} = \{X_1^{(2)}, X_2\}$, where $X_1^{(1)}, X_1^{(2)}, X_2 \in \mathscr{A}$ and $X_1^{(1)} >_{SD1} X_2^{(2)}$, where $>_{SD1}$ denotes stochastic dominance of order 1.

   Moreover, consider two portfolios $\boldsymbol{\pi}_k \equiv \left(\mathscr{S}^{(k)}, \rho^{(k)}, v, \delta^{(k)}\right)$, for $k = 1, 2$, such that:

$$\begin{cases} \rho_1^{(1)} \geq \rho_1^{(2)}; \\ \rho_2^{(1)} = \rho_2^{(2)}; \\ \delta_{i\to j}^{(1)} \geq \delta_{i\to j}^{(2)}, \text{ for } i = 1, j = 2; \\ \delta_{i\to j}^{(1)} = \delta_{i\to j}^{(2)}, \text{ otherwise.} \end{cases} \tag{6}$$

   Then the decision maker prefers portfolio $\boldsymbol{\pi}_1$ to $\boldsymbol{\pi}_2$.
4. **Compound portfolios**—Fix $n \in \mathbb{N}$. Consider n portfolios $\boldsymbol{\pi}_k \equiv \left(\mathscr{S}^{(k)}, \rho^{(k)}, v^{(k)}, \delta^{(k)}\right)$, for $k = 1, 2, \ldots, n$, and a compound portfolio $\boldsymbol{\pi}^\star$ having the n portfolios as nodes as follows:

$$\boldsymbol{\pi}^\star \equiv (\{\boldsymbol{\pi}_1, \boldsymbol{\pi}_2, \ldots, \boldsymbol{\pi}_n\}, \rho^\star, v^\star, \delta^\star)\,.$$

   Furthermore, define $\widetilde{\boldsymbol{\pi}} \in \mathbf{P}$ as $\widetilde{\boldsymbol{\pi}} \equiv \left(\widetilde{\mathscr{S}}, \tilde{\rho}, \tilde{v}, \tilde{\delta}\right)$, where

$$\widetilde{\mathscr{S}} = \bigcup_{k=1}^{n} \mathscr{S}^{(k)},$$

   two transformations ϕ and ψ can be identified such that:

$$\begin{cases} \tilde{\rho} = \phi\left(\rho^\star, \rho^{(1)}, \rho^{(2)}, \ldots, \rho^{(n)}\right); \\ \tilde{\delta} = \psi\left(\delta^\star, \delta^{(1)}, \delta^{(2)}, \ldots, \delta^{(n)}\right) \end{cases} \tag{7}$$

   and, for each $(i,j) \in \mathscr{S}^{(k_i)} \times \mathscr{S}^{(k_j)}$, we have

$$\begin{cases} \tilde{v}(i,j) = 1 \text{ if } k_i \neq k_j \text{ and } v^\star(\mathbf{N}_i, \mathbf{N}_j) = 1; \\ \tilde{v}(i,j) = 1 \text{ if } k_i = k_j = k \text{ and } v^{(k)}(i,j) = 1; \\ \tilde{v}(i,j) = 0 \text{ otherwise.} \end{cases}$$

   Then, two transformations ϕ and ψ as in (7) exist such that $\widetilde{\boldsymbol{\pi}}$ and $\boldsymbol{\pi}^\star$ are equivalent for the preference order.
5. **Independence**—Consider two portfolios $\boldsymbol{\pi}_k \equiv \left(\mathscr{S}^{(k)}, \rho^{(k)}, v^{(k)}, \delta^{(k)}\right)$, for $k = 1, 2$. Assume that $\mathscr{S}^{(1)} = \{X_1^{(1)}, \ldots, X_n^{(1)}\}$ and $\mathscr{S}^{(2)} = \{X_1^{(2)}, \ldots, X_n^{(2)}\}$, with $X_i^{(1)} \neq X_i^{(2)}$ for each $i = 1, \ldots, n$. Suppose that $X_1^{(1)} >_{SD1} X_1^{(2)}$ and suppose also

that $\boldsymbol{\pi}_1$ is preferred to $\boldsymbol{\pi}_2$. Let us now consider two portfolios $\widetilde{\boldsymbol{\pi}}_1, \widetilde{\boldsymbol{\pi}}_2$ defined as follows:

$$\widetilde{\boldsymbol{\pi}}_k \equiv \left(\tilde{\mathscr{S}}^{(k)}, \rho^{(k)}, v, \delta^{(k)}\right), \qquad k = 1, 2,$$

where $\tilde{\mathscr{S}}^{(1)} = \{\tilde{X}_1^{(1)}, X_2^{(1)}, \dots, \tilde{X}_n\}$ and $\tilde{\mathscr{S}}^{(2)} = \{\tilde{X}_1^{(2)}, X_2^{(2)}, \dots, \tilde{X}_n\}$, being $\tilde{X}_1^{(1)} >_{SD1} \tilde{X}_1^{(2)}$.

Then $\widetilde{\boldsymbol{\pi}}_1$ is preferred to $\widetilde{\boldsymbol{\pi}}_2$.

References

1. Abrahamsen, E.B., Aven, T.: On the consistency of risk acceptance criteria with normative theories for decision-making. Reliab. Eng. Syst. Saf. **93**(12), 1906–1910 (2008)
2. Castellano, R., Cerqueti, R.: Optimal consumption/investment problem with light stocks: a mixed continuous-discrete time approach. Appl. Math. Comput. **218**(12), 6887–6898 (2012)
3. Castellano, R., Cerqueti, R.: Mean-variance portfolio selection in presence of infrequently traded stocks. Eur. J. Oper. Res. **234**(2), 442–449 (2014)
4. Cerqueti, R., Lupi, C.: Risk measures on networks and expected utility. Reliab. Eng. Syst. Saf. **155**, 1–8 (2016)
5. Cerqueti, R., Spizzichino, F.: Extension of dependence properties to semi-copulas and applications to the mean-variance model. Fuzzy Sets Syst. **220**, 99–108 (2013)
6. Leippold, M., Trojani, F., Vanini, P.: A geometric approach to multiperiod mean variance optimization of assets and liabilities. J. Econ. Dyn. Control. **28**(6), 1079–1113 (2004)
7. Markowitz, H.: Portfolio selection. J. Financ. **7**(1), 77–91 (1952)
8. Markowitz, H.: Mean-variance approximations to expected utility. Eur. J. Oper. Res. **234**(2), 346–355 (2014)
9. Merton, R.C.: Lifetime portfolio selection under uncertainty: the continuous-time case. Rev. Econ. Stat. **51**, 247–257 (1969)
10. Merton, R.C.: Optimal consumption and portfolio rules in a continuous time model. J. Econ. Theory **3**, 373–413 (1971)
11. Rambaud, S.C., Perez, J.G., Sanchez Granero, M.A., Trinidad Segovia, J.E.: Markowitz's model with Euclidean vector spaces. Eur. J. Oper. Res. **196**, 1245–1248 (2009)
12. Samuelson, P.A.: Lifetime portfolio selection by dynamic stochastic programming. Rev. Econ. Stat. **51**, 239–246 (1969)
13. Scott J.: Social Network Analysis, 3rd edn. Sage, London (2013)
14. Wasserman, S., Faust, K.: Social Network Analysis: Methods and Applications. Cambridge University Press, Cambridge (1994)

An Evolutionary Approach to Improve a Simple Trading System

Marco Corazza, Francesca Parpinel, and Claudio Pizzi

Abstract In this paper we consider a simple trading system (TS) based on a set of Technical Analysis (TA) indicators. Their peculiarity is the dependence on the time-window widths used to calculate them. To attempt to improve the performances of the TS, we optimize these parameters (that is the time-window widths) by the Particle Swarm Optimization (PSO), which is a metaheuristic used to solve global optimization problems. The use of PSO is necessary since the involved optimization problem is nonlinear, nondifferentiable and integer: in summary, it is complex. In such a case, the use of exact solution methods would be excessively time-consuming, in particular for practical purposes. The proposed TS is tested using the daily closing prices from January 2, 2001, to June 30, 2016, of eight Italian stocks of different economic sectors. As benchmark, we consider the same TS but with standard time-window lengths. Irrespective of their signs, both in-sample and out-of-sample performances achieved by the TS with optimized parameters are better than those achieved by the benchmark, highlighting that parameter optimization can play an important role in TA-based TSs.

1 Introduction

The massive amount of data available in the financial markets, also free downloadable, allows us to strategically process and to convert them into useful information about the future trend of the prices of one or more assets.

Hu et al. [5] identify three different analysis methods: technical, fundamental and blending. The main aim of Technical Analysis (TA) is to compute technical indicators using the time series of stock prices and volumes and, starting from these indicators, to generate a Buy, Hold or Sell (BHS) signal [8]. On the other hand, Fundamental Analysis generates trading rules when the stock is undervalued or overvalued with respect to its fundamental value. The last analysis, the Blending one, combines both of them.

M. Corazza (✉) • F. Parpinel • C. Pizzi
Department of Economics, Ca' Foscari University of Venice, Cannaregio 873, 30121 Venice, Italy
e-mail: corazza@unive.it; parpinel@unive.it; pizzic@unive.it

M. Corazza et al. (eds.), *Mathematical and Statistical Methods for Actuarial Sciences and Finance*, https://doi.org/10.1007/978-3-319-50234-2_7

As underlined by Wu et al. [11], to elaborate a set of profitable trading rules we have to forecast the direction of the asset prices, so an effective information extraction is needed.

In this work, we consider the formulation of a trading system (TS) from the perspective of TA, so that the analysis of the patterns of price and/or of volume sequences can enables us to generate trading rules [3]. The literature on TA has proposed several indicators, to each of which is associated a trading rule generating BHS signal. The indicators, and consequently the signals generated, depend on some parameters, the most important of which is the time-window width. Thus, the selection of the time-window width is crucial. According to the latest state of our knowledge, there is not an estimation methodology for these parameters, furthermore the rule of thumb used by many practitioners appears to us inappropriate and risky. The BHS signals generated by an indicator results in profits or losses. So, we need a procedure to select the time-window widths in a such way to maximize the profit or to minimize the losses. According to this, we look for an optimization algorithm allowing the estimation of the time-window widths.

Several Authors, [1, 5, 9, 12], present interesting applications of evolutionary optimization approaches implemented in the trading framework. For instance, the Ant Colony Optimization (ACO), as well as Particle Swarm Optimization (PSO), that maximizes/minimizes a fitness function by mimicking the behaviour of a group of insects or animals. Moreover, designing a TS based on several indicators, we have to consider possible conflicting BHS signals, so we have also to tackle the issue of combining these signals. It follows that there are two aspects to face when we examine this problem. On one hand we have to determine the parameter values (that is the time-window widths), on the other we have to combine the different signals in a single one.

In this paper we focus on the first problem, since the combination of the signals can be solved considering the simple unweighted sum of each signal. In particular, in this paper we propose a TS based on four indicators: two momentum indicators, typically used for trend detection, such as the Exponential Moving Average (EMA) and the Bollinger Bands (BB), and two oscillator indicators such as the Relative Strength Index (RSI) and the Moving Average Convergence/Divergence (MACD). All these indices depend on the time-window widths that we can specify using the swarm intelligence approach.

The paper is organized as follows. The next section will introduce the methodology we will use. In Sect. 3 we will show some in-sample and out-of-sample results of the implemented methodology. In the last section we will give some final remarks.

2 Methodology

As we said in the previous section, effective TA tools settings are crucial to reach good performance by a TS. Nevertheless, the evidence is that many trading practitioners follow a rule of thumb, meaning that parameters have a default. For example, several financial web-sites, like for instance `finance.yahoo.com` or, in an Italian context, `borsaitaliana.it`, provide indicators and oscillators with identical or very similar settings. In order to improve the performances of TA indicators, we propose a more objective method to optimize their parameter settings. In particular, we use a metaheuristic known as Particle Swarm Optimization (PSO) to estimate the best time-window widths. The use of this solver is justified by the complexity of the involved optimization problem for which the exact solution methods could be extremely time-consuming for practical purposes.

In the next subsection we will describe the adopted TS. Then, we will introduce the optimization problem. At last, in Sect. 2.3, the PSO and its specific implementation will be presented in the considered context.

2.1 A Simple TS

As previously highlighted, we propose a combination of four classical and widely used indicators, two momentum indicators and two oscillators, to define our TS. The main aim of this work is to verify if using PSO allow us to improve the performance in terms of rate of return with respect to the same TS with default settings of the indicators. The involved TA indicators are: Exponential Moving Average (EMA), Bollinger Bands (BB), Moving Average Convergence/Divergence (MACD), and Relative Strength Index (RSI). For the description of such indicators we refer to [7].

For each indicator we define a decisional rule that provides a trading signal. More precisely the signal may be: "-1", namely "*Sell or stay short in the market*"; "0", namely "*Stay out from the market*"; "$+1$", namely "*Buy or stay long in the market*". The four trading signals are then aggregated in order to obtain a single signal. As far as the four decisional rules are concerned, let us specify as trading period the discrete time interval $t = 1, \dots, T > 1$, and let us assume that at time $t = 1$ each of the four trading signals is equal to 0. From $t = 2$ to $t = T$, the four decisional rules are:

- the one based on *EMA*, with $EMA_f(\cdot)$ a fast *EMA* and $EMA_s(\cdot)$ a slow *EMA*:

$$signal_{EMA}(t) = \begin{cases} -1 \text{ if } EMA_f(t) < EMA_s(t) \wedge EMA_f(t-1) \geq EMA_s(t-1) \\ +1 \text{ if } EMA_f(t) > EMA_s(t) \wedge EMA_f(t-1) \leq EMA_s(t-1); \\ signal_{EMA}(t-1) \text{ otherwise} \end{cases}$$

- the one based on *RSI*:

$$signal_{RSI}(t) = \begin{cases} -1 \text{ if } RSI(t) > 70 \wedge RSI(t-1) \leq 70 \\ +1 \text{ if } RSI(t) < 30 \wedge RSI(t-1) \geq 30; \\ signal_{RSI}(t-1) \text{ otherwise} \end{cases}$$

- the one based on *MACD*, where $DL(\cdot)$ and $SL(\cdot)$ are respectively the so-called differential line and signal line:

$$signal_{MACD}(t) = \begin{cases} -1 \text{ if } DL(t) < SL_{sl}(t) \wedge DL(t-1) \geq SL_{sl}(t-1) \\ +1 \text{ if } DL(t) > SL_{sl}(t) \wedge DL(t-1) \leq SL_{sl}(t-1); \\ signal_{MACD}(t-1) \text{ otherwise} \end{cases}$$

- the one based on *BB*, where $P(\cdot)$ is the price of the considered financial asset, $BB_L(\cdot)$ and $BB_U(\cdot)$ are respectively the lower band and the upper band:

$$signal_{BB}(t) = \begin{cases} -1 \text{ if } P(t) < BB_U(t) \wedge P(t-1) \geq BB_U(t-1) \\ +1 \text{ if } P(t) > BB_L(t) \wedge P(t-1) \leq BB_L(t-1). \\ signal_{BB}(t-1) \text{ otherwise} \end{cases}$$

We must keep in mind that any decisional rule will depends on the values of the indicators which, in their turn, will depends on a given parametrization. In the following, we will denote by $\mathbf{w}$ the vector of the following parameters: w_f and w_s are the time-window widths related to the fast *EMA* and to the slow *EMA*, respectively; w_{RSI} is the time-window width related to *RSI*; $w_{MACD,1}$, $w_{MACD,2}$ and w_{sl} are the three time-window widths related to *MACD*; w_{LB} and w_{UB} are the two time-window widths related to *BB*.

To define only one operational trading signal, we propose to aggregate the trading signals in the following way:

$$signal(t) = \text{sign}(signal_{EMA}(t) + signal_{RSI}(t) + signal_{MACD}(t) + signal_{BB}(t)),$$

where $\text{sign}(\cdot)$ is the signum function.

Note that if three or four out of four decisional rules give the same trading signal then the single operational trading signal is equal to it. Moreover, it is easy to prove that if two decisional rules provide the same trading signal and the other two decisional rules provide different trading signals, also between them, then the single operational trading signal is equal to the one of the two former decisional rules.

2.2 The Optimization Problem

In this paper we measure the performance of a TS in a quite simple way, although there are other ways to do this. An intuitive measure is the net capital at the end of the trading period, $C(T)$, where "net" means that we explicitly take into accounts of the transaction costs. In detail, let δ be the transaction costs expressed in percentage; we define the net rate of return, $e(t)$, obtained by the TS from $t-1$ to t, as follows:

$$e(t) = signal(t-1) \ln\left(\frac{P(t)}{P(t-1)}\right) - \delta \left|signal(t) - signal(t-1)\right|, t = 2, \ldots, T,$$

and, fixing $C(1)$, the equity line produced by the TS is:

$$C(t) = C(t-1)[1 + e(t)], t = 2, \ldots, T.$$

At this stage, we can formulate the constrained optimization problem as:

$$\max_{\mathbf{w}} \; C(T) \\ \text{s.t.} \; \begin{cases} w_{EMA_f} < w_{EMA_s} \\ w_{sl} < w_{MACD,1} < w_{MACD,2} \\ w_{EMA_f}, w_{EMA_s}, w_{RSI}, w_{sl}, w_{MACD,1}, w_{MACD,2}, w_{BB} \in \mathbb{N}^+ \end{cases} . \quad (1)$$

This constrained maximization problem is "complex" because the objective function $C(\cdot)$ is nonlinear a non differentiable, moreover it is formulated in terms of variables that must assume integer values. So, as underlined above, we need to use as a solver a metaheuristic like PSO.

2.3 Particle Swarm Optimization

PSO is an iterative metaheuristic for the solution of global unconstrained continuous optimization problems [6, 10], that may be adapted for solving also constrained ones. The basic idea of PSO is to mimic the social behaviour of swarm of bees or of flocks of birds cooperating for searching for food. For this purpose, each particle, or member, of the swarm moves in the search area. The direction and the velocity of the movement depend on its best position reached so far, and on the exchanges of information with the neighboring particles in the swarm. The behaviour of each particle allows to the whole swarm to converge towards the best global position. From a mathematical point of view, the paradigm of a flying swarm may be formulated as follows: given an optimization problem, each particle of the swarm represents a possible solution; its starting position $\mathbf{x}_j^1$ and starting velocity $\mathbf{v}_j^1$ are randomly assigned, so determining its initial direction and velocity of movement.

Let us consider the global unconstrained optimization problem $\min_{\mathbf{x} \in \mathbb{R}^d} f(\mathbf{x})$, where $f : \mathbb{R}^d \mapsto \mathbb{R}$ is the objective function. Suppose to solve it using PSO and considering M particles. At the k-th iteration of the algorithm, three vectors are associated to the j-th particle, with $j = 1, \ldots, M$:

- the position $\mathbf{x}_j^k \in \mathbb{R}^d$;
- the velocity $\mathbf{v}_j^k \in \mathbb{R}^d$;
- the best position visited so far $\mathbf{p}_j \in \mathbb{R}^d$.

Moreover, $pbest_j = f(\mathbf{p}_j)$ is the value of the objective function in position $\mathbf{p}_j$, while $\mathbf{p}_g$ and *gbest* are respectively the best position reached by the swarm and the value of the objective function in such a position.

The steps of the algorithm are the following:

1. Randomly assign the starting position $\mathbf{x}_1^j$ and the starting velocity $\mathbf{v}_1^j$ for $j = 1, \ldots, M$.
2. Set $pbest_j = +\infty$ for $j = 1, \ldots, M$, set $gbest = +\infty$, and set $k = 1$.
3. Evaluate $f(\mathbf{x}_j^k)$ for $j = 1, \ldots, M$.
4. If $f(\mathbf{x}_j^k) < pbest_j$ then set $\mathbf{p}_j = \mathbf{x}_j^k$ and $pbest_j = f(\mathbf{x}_j^k)$ for $j = 1, \ldots, M$.
5. If $f(\mathbf{x}_j^k) < gbest$ then set $\mathbf{p}_g = \mathbf{x}_j^k$ and $gbest_j = f(\mathbf{x}_j^k)$ for $j = 1, \ldots, M$.
6. Update position and velocity of the j-th particle for $j = 1, \ldots, M$ as:

$$\mathbf{v}_j^{k+1} = w^{k+1}\mathbf{v}_j^k + \mathbf{U}_{\phi_1} \otimes (\mathbf{p}_j - \mathbf{x}_j^k) + \mathbf{U}_{\phi_2} \otimes (\mathbf{p}_g - \mathbf{x}_j^k) \tag{2}$$

$$\mathbf{x}_j^{k+1} = \mathbf{x}_j^k + \mathbf{v}_j^{k+1} \tag{3}$$

   where $\mathbf{U}_{\phi_1}, \mathbf{U}_{\phi_2} \in \mathbb{R}^d$ and their components are uniformly randomly distributed in $[0, \phi_1]$ and $[0, \phi_2]$ respectively, and denoting $\otimes$ the component-wise product.
7. If a convergence criterion is not satisfied then set $k = k + 1$ and go to step 3.

The values of ϕ_1 and ϕ_2 strongly affect the strength of the attractive forces towards the personal and the swarm best positions explored so far by the particles. Thus, in order to get the convergence of the swarm, they have to be set carefully in accordance with the value of the inertia weight w^k. The parameter w^k generally decreases linearly with the number of steps, as:

$$w^k = w_{max} + \frac{w_{min} - w_{max}}{K} k$$

where K is usually the maximum number of iterations allowed. The values for w_{max} and w_{min} are typically 0.9 and 0.4.

We previously said that our optimization problem is a constrained integer one and, for this reason, we have to adapt the standard PSO algorithm to deal with these peculiarities.

As concern the presence of integer variables, we apply the approach suggested in [7] and ≪[e]*ach particle of the swarm* [is] *truncated to the closest integer, after the determination of its new position* [by (3)]≫ [7, p. 1584]. As pointed out in [7] ≪[t]*he truncation of real values to integers seems not to affect significantly the performance of the method, as the experimental results indicate. Moreover, PSO outperforms the* [Branch and Bound] *technique for most test problems*≫ [7, p. 1583]. So, with this approach we can manage the constraints w_{EMA_f}, w_{EMA_s}, w_{RSI}, w_{sl}, $w_{MACD,1}$, $w_{MACD,2}$, $w_{BB} \in \mathbb{N}^+$ of our optimization problem.

As regards the other constraints, different strategies are proposed in literature. In this paper we use PSO as a tool for the solution of unconstrained optimization problems according to its original aim, so we have to reformulated our problem into an unconstrained one. To this purpose, we use an approach described in [4] and recently applied in the financial context [2]. Such an approach uses a nondifferentiable ℓ_1 penalty function to reformulated our problem into an unconstrained one and is known as *exact penalty method*. The term "exact" refers to the correspondence between the minimizers of the original constrained problem and the minimizers of the unconstrained (penalized) one.

Let ϵ be the penalty parameter, then the reformulated version of our optimization problem is:

$$\max_{\mathbf{w}} C(T) - \frac{1}{\epsilon}\Big[\max\{0, w_{EMA_f} - w_{EMA_s}\} + \max\{0, w_{sl} - w_{MACD,1}\} + \max\{0, w_{MACD,1} - w_{MACD,2}\}\Big]. \tag{4}$$

A correct choice of the value of the penalty parameter ensures the correspondence between the solutions of the original constrained problem and of the reformulated unconstrained one. Note that in the (4), the constraints w_{EMA_f}, w_{EMA_s}, w_{RSI}, w_{sl}, $w_{MACD,1}$, $w_{MACD,2}$, $w_{BB} \in \mathbb{N}^+$ do not appear as they are taken into account by the truncation of real values described above.

3 Applications

As stated above, in this paper we want to study the performances of a simple TS based on indicators coming from TA, finding the optimal values of the time-window widths associated to such indicators by using PSO.

To assess the actual improvement of the applied procedure, we have carried out both an in-sample and an out-of-sample analysis. In particular, we have determined the optimized parameter values of the considered TS solving the optimization problem (4) by the version of the PSO described in Sect. 2.3.

We have considered time series of eight stocks in the period from January 2, 2001 to June 30, 2016 (3932 prices). More specifically, we have tested the TS on the following stocks: BUZZI UNICEM S.p.A. (BU), ENEL S.p.A. (EE), ENI S.p.A. (EI), Generali S.p.A. (GE), INTESA SANPAOLO S.p.A. (IS), LUXOTTICA GROUP S.p.A. (LG), STMICROELECTRONICS S.p.A. (ST) and TELECOM ITALIA S.p.A. (TI). We have choosen the stocks considering their importance in the Italian stock market. In fact, all of them are components of the Italian stock index FTSE MIB and they represent some meaningful sectors of the Italian economy.

We have performed the in-sample analysis in the period going from June 2, 2001 to June 30, 2016 (3880 prices[1]), whereas the out-of-sample analysis has been carried out in the period from November 1, 2011 to June 30, 2016 (1180 prices), using the period from January 2, 2001 to October 31, 2011 (2752 prices) to evaluate the optimal values of the time-window widths.

All the applications have used $\delta = 0.15\%$ (percentage transaction cost currently applied by several Italian brokerage companies) and $C(1) = 100$.

As regards the PSO algorithm, we have used the following setting: $M = 10$ as number of particles; $K = 100$ as maximum number of iterations; $\phi_1 = \phi_2 = 1.49618$ as coefficients; $\epsilon = 0.0001$ as penalty parameter. The first two values have been determined by a trial-and error procedure, the last three values are commonly suggested in the literature. Considering that the methodology proposed in Sect. 2.3 is stochastic in reason of the random initialization (of both position and velocity of the particles) and of the random disturbance in the updating equation of the velocity (see step 6 of the algorithm presented in Sect. 2.3), we have applied 100 times our methodology to each stock, then we have calculated the mean values and other statistics.

As far as the TS with standard setting is concerned, following the relevant professional literature, we have used the following values for the parameters: $w_{EMA_f} = 12$, $w_{EMA_s} = 26$, $w_{RSI} = 26$, $w_{sl} = 9$, $w_{MACD,1} = 12$, $w_{MACD,2} = 26$ and $w_{BB} = 26$.

3.1 In-Sample Analysis

The in-sample performances achieved by the two TSs (with default and optimized settings) for the eight stocks are presented in Table 1. More specifically, in column 2 we report the annualized rate of return performed by the TS with standard setting (r); in columns 3 and 4 we respectively report the average annualized rate of return performed by the TS with optimized parameter values ($\overline{r}$) and the associated standard deviation (s_r); column 5 shows the 95% confidence interval calculated using $\overline{r}$ and s_r ($[\cdot, \cdot]_{95\%,r}$); in columns 6 and 7 we have the minimum of r ($r_{\min}$) and the maximum of r ($r_{\max}$), respectively, over the 100 applications of our methodology; in column 8 and 9 we respectively report the average percentages of times in which, during the trading period, the value of the equity line produced by the TS with optimized parameter values has been greater than, and not less than, the value of the equity line produced by the TS with standard setting ($\overline{\% >}$ and $\overline{\% \geq}$).

[1]The first 52 prices need to calculate the starting values of indicators.

Table 1 In-sample performances achieved by the various TSs

Stock	r	$\overline{r}$	s_r	$[\cdot,\cdot]_{95\%,r}$	$r_{\min}$	$r_{\max}$	$\overline{\%>}$	$\overline{\%\geq}$
BU	−2.86%	16.25%	4.26%	[7.89%, 24.60%]	1.14%	23.75%	90.18%	91.66%
EE	−17.12%	4.31%	2.60%	[−0.79%, 9.41%]	−3.17%	9.75%	87.89%	89.30%
EI	−16.77%	0.46%	2.84%	[−5.11%, 6.034%]	−9.20%	10.63%	98.66%	99.98%
GE	−4.92%	9.00%	3.03%	[3.07%, 14.93%]	0.00%	13.65%	95.90%	97.72%
IS	−9.69%	18.08%	6.79%	[4.77%, 31.38%]	0.00%	33.47%	93.16%	94.71%
LG	−5.88%	13.20%	3.04%	[7.24%, 19.15%]	0.00%	17.30%	74.97%	76.43%
ST	−1.00%	16.01%	5.99%	[4.26%, 27.76%]	−1.52%	24.49%	95.24%	96.63%
TI	−9.56%	10.75%	4.88%	[1.19%, 20.30%]	−3.29%	18.51%	98.11%	99.59%

We point out that all the annualized rates of return achieved by the TS with standard setting are negative, whereas all the average annualized rates of return performed by the TS with optimized parameter values are far greater than the former ones and are all positive. It indicates that also in the case of simple TA-based TS, like the one considered in this paper, the parameters optimization can play an important role. Then, no average annualized rate of return achieved by the TS with standard setting belongs to the 95% confidence interval calculated using $\overline{r}$ and s_r. This indicates that, for all the investigated stocks, $\overline{r}$ is statistically different from r at the 5% significance level. Moreover, note also that, with the only exception of the stock asset ST, all rs are lower than the corresponding $r_{\min}$ ($r_{\min}$s are in column 6), which means that the worst results obtained with optimized parameters is generally better than those obtained with standard settings.

All the previous remarks concern with the performances achieved by the various TSs in the final time instant $t = T$ of the trading period, that is June 30, 2016. But the results in column 8 and 9 well put in evidence that for very large part of the trading period (never lower than 74.97%) the TSs with optimized parameter values perform better than the TS with standard setting.

As example, in Fig. 1 we show the in-sample performance related to the stock asset TI. In particular, on the top, the closing price time series is shown; in the second panel, the operational trading signal is reported; on the bottom, the time series of the gross equity line produced by the TS with optimized parameter values (dotted curve), of the net equity line produced by the same TS (bold curve), and of the net equity line produced by the TS with standard setting (continuous curve) are shown.

3.2 *Out-of-Sample Analysis*

In Table 2 we present the out-of-sample performances attained by the various TSs. Its columns are the same of those of Table 1.

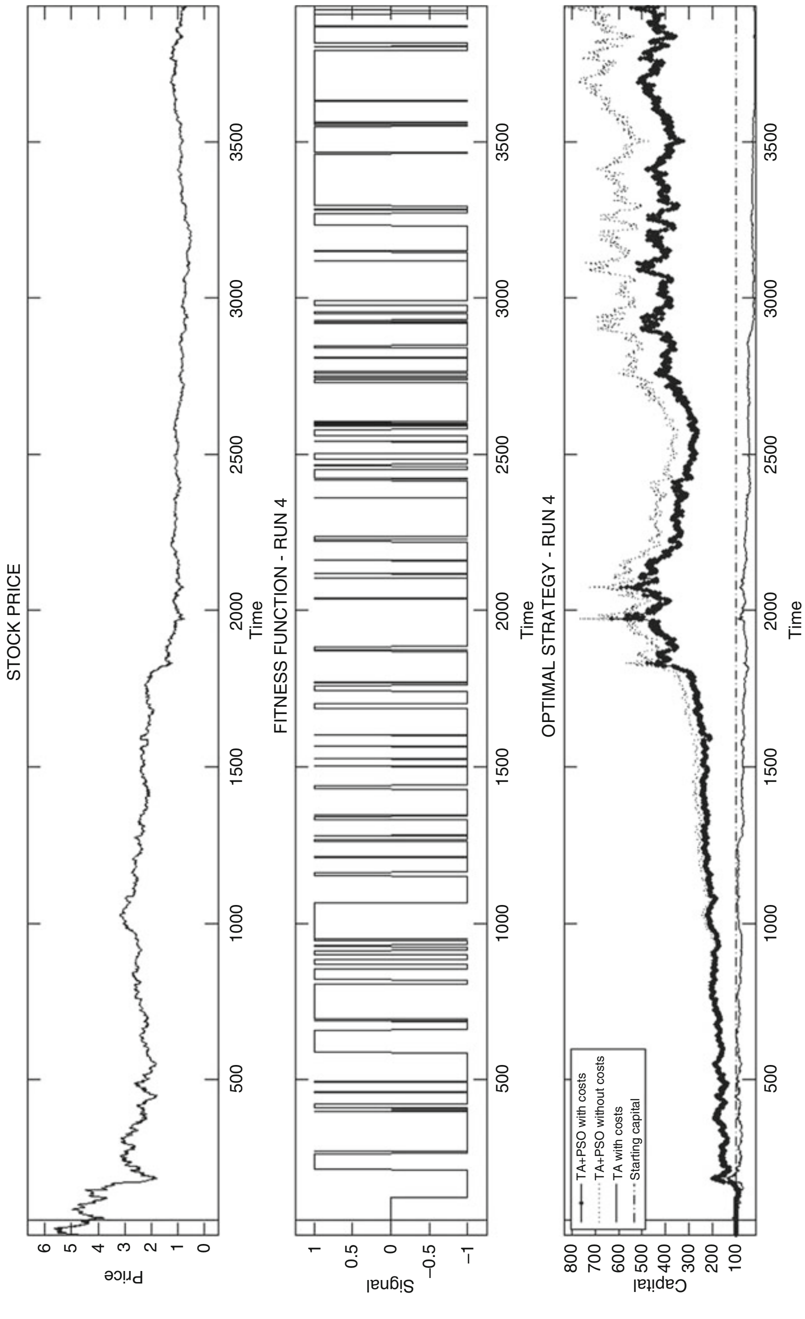

Fig. 1 In-sample performance related to the stock asset TI. Data to the left of the vertical line are related to the first 52 prices used for calculating the starting values of indicators and oscillators

Table 2 Out-of-sample performances achieved by the various TSs

Stock	r	$\overline{r}$	$\overline{s}_r$	$[\cdot, \cdot]_{95\%,r}$	$r_{\min}$	$r_{\max}$	$\overline{\% >}$	$\overline{\% \geq}$
BU	−27.07%	−9.27%	9.34%	[−27.57%, 9.04%]	−34.87%	10.93%	79.65%	79.75%
EE	−21.11%	−7.85%	8.68%	[−24.86%, 9.16%]	−28.57%	8.12%	81.92%	82.02%
EI	−19.04%	−8.89%	9.70%	[−27.90%, 10.12%]	−31.89%	14.74%	86.23%	86.31%
GE	−9.99%	−9.31%	6.10%	[−21.27%, 2.65%]	−27.86%	4.90%	28.12%	28.21%
IS	−18.67%	−7.36%	11.88%	[−30.65%, 15.93%]	−25.63%	29.91%	54.89%	55.00%
LG	−13.89%	−3.04%	9.82%	[−22.29%, 16.20%]	−31.91%	18.10%	59.49%	59.57%
ST	−21.18%	−16.92%	9.38%	[−35.30%, 1.47%]	−39.35%	1.60%	70.08%	70.17%
TI	−15.23%	−12.14%	8.24%	[−28.28%, 4.00%]	−32.02%	1.21%	86.41%	86.50%

Similarly to what highlighted by the in-sample analysis, even in the out-of-sample one all the annualized rates of return achieved by the TS with standard setting (column 2) are negative and all the average annualized rates of return performed by the TS with optimized parameter values (column 3) are greater than the former. But conversely, now all $\overline{r}$s are negative. It indicates that the parameter optimization play a positive role also in the out-of-sample applications, although its importance is significantly diminished with respect to the in-sample ones. As a further confirmation of such an importance reduction, note that now all the annualized rates of return achieved by the TS with standard setting (column 2) belong to the 95% confidence intervals calculated using $\overline{r}$ and s_r (column 5), and that each r is greater than the corresponding $r_{\min}$ ($r_{\min}$s are in column 6). This may suggest that in the out-of-sample applications the standard setting does not appear so far from the optimal setting.

We recall that all the previous remarks concern with the performances obtained by the various TSs in the final time instant $t = T$, that is June 30, 2016. But the results in column 8 and 9 point out that, with the only exception of the stock asset GE, for large part of the trading period the TSs with optimized parameter values perform better than the TS with standard setting.

As example, in Fig. 2 we show the out-of-sample performances related to the stock asset TI. Its panels are the same of those of Fig. 1.

4 Conclusions

In this paper we have considered a simple TS based on four indicators, two momentum and two oscillators, coming from TA. Instead of using a rule of thumb to select the time-window widths, we have proposed to apply an adapted version of PSO in order to determine the optimal values of these parameters. The results we have obtained, summarized in Table 1 for the in-sample analysis and in Table 2 for the out-of-sample analysis, show that parameter optimization can play an important role as the results of our proposal, expressed in terms of annualized rate of return, are always better than the classical TA-based ones.

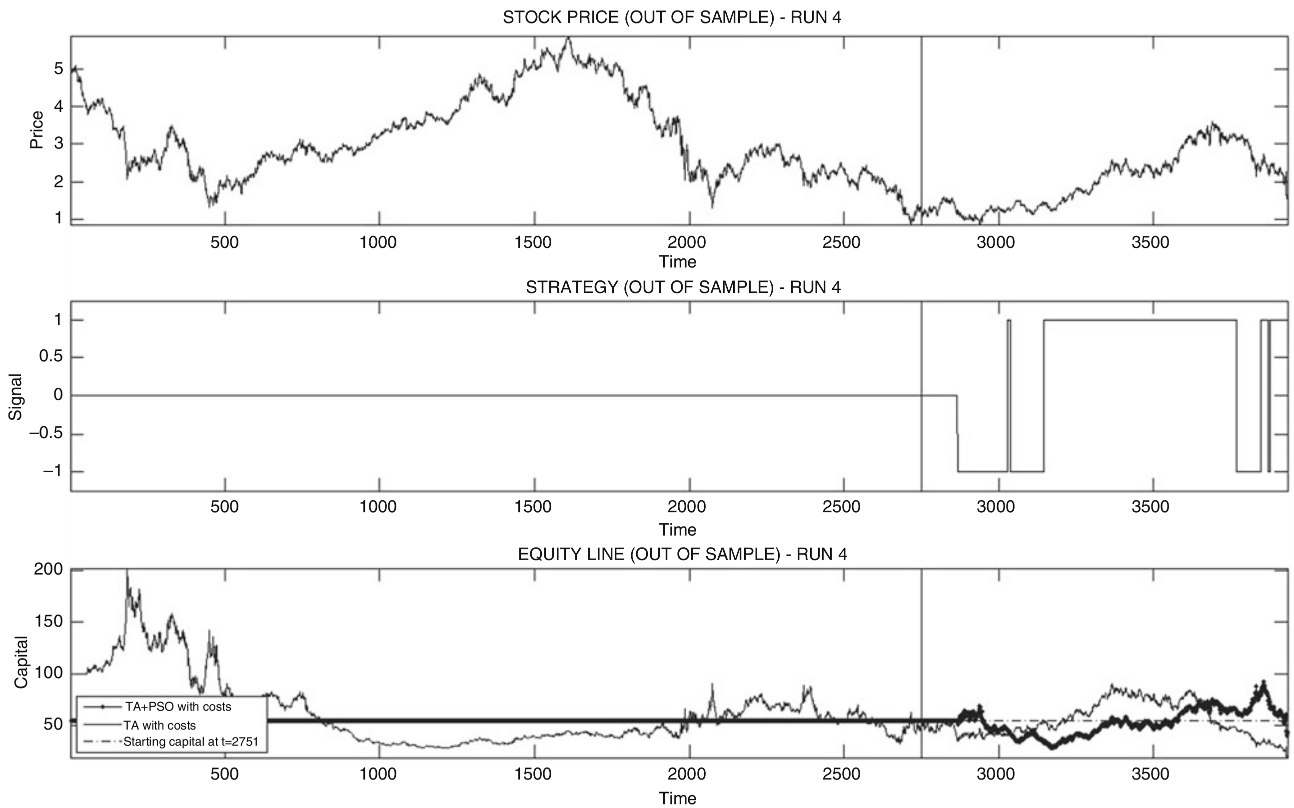

Fig. 2 Out-of-sample performances related to the stock asset IS. Data to the right of the vertical line are related to the out-of-sample period

Another advantage of the procedure is that the parameter setting may vary according to the stock, so avoiding the use of the same parameter values in possible different contexts.

Our future goals are: to check the proposed technique with different price time series coming from other financial markets; to improve our methodology in order to anticipate the market signals using observational data; to generalize the proposed procedure in order to generate trading rule.

References

1. Allen, F., Karjalainen, R.: Using genetic algorithms to find technical trading rules. J. Financ. Econ. **51**, 245–271 (1999)
2. Corazza, M., Fasano, G., Gusso R.: Particle Swarm Optimization with no-smooth penalty reformulation, for a complex portfolio selection problem. Appl. Math. Comput. **224**, 611–624 (2013)
3. Dash, R., Dash, P.K.: A hybrid stock trading framework integrating technical analysis with machine learning techniques. J. Finance Data Sci. **2**, 42–57 (2016)
4. Fletcher, R.: Practical Methods of Optimization. Wiley, Glichester (1991)
5. Hu, Y., Liu, K., Zhang, X., Su, L., Ngai, E.W.T., Liu, M.: Application of evolutionary computation for rule discovery in stock algorithmic trading: a literature review. Appl. Soft Comput. **36**, 534–551 (2015)
6. Kennedy, J., Eberhart, R.C.: Particle Swarm Optimization. In: Proceedings of the 1995 IEEE International Conference on Neural Networks, vol. IV, pp. 1942–1948 (1995)
7. Laskari, E.C., Parsopoulos, K., Vrahatis, M.N.: Particle swarm optimization for integer programming. In: Proceedings of the 2002 Congress on Evolutionary Computation, vol. 2, pp. 1582–1587 (2002)
8. Murphy, J.J.: Technical Analysis of the Financial Markets. A Comprehensive Guide to Trading Methods and Applications. New York Institute of Finance, New York (1999)
9. Ni, H., Yin, H.: Exchange rate prediction using hybrid neural networks and trading indicators. Neurocomputing **72**, 2815–2823 (2009)
10. Poli, R., Kennedy, J., Blackwell, T.: Particle swarm optimisation: an overview. Swarm Intell. J. **1**, 33–57 (2007)
11. Wu, J., Yu, L., Chang, P.: An intelligent stock trading system using comprehensive features. Appl. Soft Comput. **23**, 39–50 (2014)
12. Yao, J.T., Tan, C.L.: A case study on using neural networks to perform technical forecasting of forex. Neurocomputing **34**, 79–98 (2000)

Provisions for Outstanding Claims with Distance-Based Generalized Linear Models

Teresa Costa and Eva Boj

Abstract In previous works we developed the formulas of the prediction error in generalized linear model (GLM) for the future payments by calendar years assuming the logarithmic link and the parametric family of error distributions named power family. In the particular case of assuming (overdispersed) Poisson and logarithmic link the GLM gives the same provision estimations as those of the Chain-Ladder deterministic method. Now, we are studying the possibility to use distance-based generalized linear models (DB-GLM) to solve the problem of claim reserving in the same way as GLM is used in this context. DB-GLM can be fitted by using the function `dbglm` of the `dbstats` package for `R`. In this study we calculate the prediction error associated to the accident years future payments and total payment, and also to the calendar years future payments using DB-GLM in the general case of the power families of error distributions and link functions. We make an application with the well known run-off triangle of Taylor and Ashe.

1 Introduction

In this paper we propose the use of DB-GLM to solve the claim reserving problem in the context of Solvency II (Directive 2009/138/EC of the European Parliament and of the Council on the taking-up and pursuit of the business of Insurance and Reinsurance).

In Solvency II it is indicated that the best estimate of technical provisions must be calculated taking account of the time value of money, that is, the expected present value of the future payments. The payments by calendar years allow us to work in a financial environment because each amount is situated in the corresponding future calendar year.

The paper is organized as follows. In Sect. 2 we describe the DB-GLM model and we show the expression of the prediction error for DB-GLM in the general parametric families of error distributions and link functions named power families.

T. Costa (✉) • E. Boj
Facultat d'Economia i Empresa, Universitat de Barcelona, Avinguda Diagonal 690, 08034 Barcelona, Spain
e-mail: tcosta@ub.edu; evaboj@ub.edu

M. Corazza et al. (eds.), *Mathematical and Statistical Methods for Actuarial Sciences and Finance*, https://doi.org/10.1007/978-3-319-50234-2_8

In Sect. 3 we develop the expressions of the prediction error for the total future payment, for the accident years future payments and for the calendar years future payments. In Sect. 4 we illustrate calculations with the well known data set of Taylor and Ashe. In Sect. 5 we present the main conclusions.

2 Distance-Based Generalized Linear Model

The DB-GLM, defined in [7], extends the ordinary GLM allowing information on predictors to be entered as interdistances between observation pairs instead of as individual coordinates. The estimation process of a DB-GLM is schematically as follows: a Euclidean configuration, X_w, is obtained by a metric multidimensional scaling-like procedure, then the linear predictor of the underlying GLM is a linear combination of the resulting Euclidean coordinates, latent variables in the model.

Let $\Omega = (\Omega_1, \ldots, \Omega_n)$ be a population of n individuals; let $Y : (Y_1, \ldots, Y_n)^T$ be the random response variable, and $(y_1, \ldots, y_n)^T$ be the observed response variable of size $n \times 1$; let $(w_1, \ldots, w_n)^T$ be a priori weights of individuals of size $n \times 1$ with $w_i \in (0, 1)$; let $\left(F_1, \ldots, F_p\right)$ be the set of p observed mixed predictors; and let Δ be an $n \times n$ matrix, whose entries are the squared distances $\delta^2 \left(\Omega_i, \Omega_j\right)$. For the sake of expediency, henceforth we will refer to Δ as the 'distance matrix'.

The distance matrix Δ is calculated from the observed predictors by means of a distance function with the Euclidean property. It contains predictor's information and it is the only information entered in the model in the predictor's space $\wp$. Therefore, distance-based prediction can be applied to mixed (qualitative and quantitative) explanatory variables or when the regressor is of functional type.

In DB-GLM we assume that the response distribution is in an exponential dispersion family, as in any GLM. Besides, the relation between the linear predictor, η, and the expected response, μ, is given by a link function: $\mu = g(\eta)$. We calculate the inner products matrix $G_w = -\frac{1}{2} J_w \cdot \Delta \cdot J_w$ where $J_w = I - 1 \cdot w^T$ is the w-centering matrix. A DB-GLM consists of random variables $(Y_1, \ldots, Y_n)^T$ whose expectation, $(\mu_1, \ldots, \mu_n)^T$, transformed by the link function and w-centered, is a vector in the column space of G_w. This space coincides with the column space of any Euclidean configuration X_w of Δ, by definition any matrix such that $G_w = X_w \cdot X_w^T$.

DB-GLM contains GLM as a particular case: if we start from a w-centered matrix X_w of continuous predictors and we define Δ as the matrix of squared Euclidean l^2 distances between rows of X_w, then X_w is trivially a Euclidean configuration of X_w, hence hat matrix, response and predictions with DB-GLM and ordinary GLM are the same. DB-GLM can be fitted using the `dbglm` function of the `dbstats` package for `R` (see [5]).

2.1 Prediction Error

Assume a DB-GLM and the power family for the variance function:

$$V(\mu_i) = \mu_i^{\xi}, \tag{1}$$

for $i = 1, \ldots, n$. Mean and variance of the random response variable are then:

$$E[y_i] = \mu_i,$$

$$Var[y_i] = (\phi/w_i)\, V(\mu_i) = (\phi/w_i)\, \mu_i^{\xi},$$

where ϕ is the dispersion parameter and w_i are a priori weights of the data. Particular cases in this family are: Poisson distribution when $\xi = 1$; Gamma distribution when $\xi = 2$; and Inverse Gaussian distribution when $\xi = 3$.

Throughout this section we follow formulations from [11]. We define the mean squared error (MSE) of prediction as:

$$E\left[(y_i - \hat{\mu}_i)^2\right]. \tag{2}$$

To estimate the MSE for the original data, $i = 1, \ldots, n$, we only need to compare the original response values, y_i, with the corresponding fitted values, $\hat{\mu}_i$, for $i = 1, \ldots, n$. But we are interested in calculating the prediction error for a new observation: $i = n + 1$ (see [6] e.g.). Prediction error, by definition, is the square root of the MSE. For new observations, when the observed predictor values do not coincide with those of the original sample we need an estimated MSE.

The MSE of prediction can be approximated by the sum of two components, the process variance and the estimation variance:

$$E\left[(y_i - \hat{\mu}_i)^2\right] \cong Var[y_i] + Var[\hat{\mu}_i]. \tag{3}$$

Using the delta method we can derive that:

$$Var[\hat{\mu}_i] \cong \left|\frac{\partial \mu_i}{\partial \eta_i}\right|^2 Var[\eta_i],$$

and then, the MSE in (2) can be approximated as:

$$E\left[(y_i - \hat{\mu}_i)^2\right] \cong (\phi/w_i)\, \mu_i^{\xi} + \left|\frac{\partial \mu_i}{\partial \eta_i}\right|^2 Var[\eta_i]. \tag{4}$$

Finally, the expression of the prediction error is given by the square root of the MSE (4):

$$\sqrt{E\left[(y_i - \hat{\mu}_i)^2\right]} \cong \sqrt{(\phi/w_i)\,\mu_i^{\xi} + \left|\frac{\partial \mu_i}{\partial \eta_i}\right|^2 Var\,[\eta_i]}. \tag{5}$$

Consider the power family of link functions as in formula (2.10) of [13]:

$$\eta_i = g\,(\mu_i) = \begin{cases} \mu_i^{\lambda}, \ \lambda \neq 0, \\ \log\,(\mu_i)\,,\ \lambda = 0. \end{cases} \tag{6}$$

The particular expressions of prediction error (5) are:

- If $\mu_i^{\lambda} = \eta_i$:

$$\sqrt{(\phi/w_i)\,\mu_i^{\xi} + \lambda^2 \mu_i^{(2\lambda+2)} Var\,[\eta_i]}.$$

- If $\log \mu_i = \eta_i$:

$$\sqrt{(\phi/w_i)\,\mu_i^{\xi} + \mu_i^2 Var\,[\eta_i]}. \tag{7}$$

To estimate the scale parameter $\hat{\phi}$ from the sample data we can use the mean of the Pearson residuals $\hat{\phi}^P$.

An alternative way to obtain the MSE (3) is to estimate $Var\,[\hat{\mu}_i]$ by bootstrap. We propose to use the resampling technique of pairs bootstrap in which each bootstrap sample consists of n response-predictor pairs from the original data (see, e.g., [9]). This technique is adequate for distance-based models as is shown in [4] where an F-test was defined for the DB-LM, and in [8] where a Wald test was defined for influence coefficients in DB-GLM.

We can estimate the predictive distribution of $\hat{\mu}_i$, called $\hat{\mu}_i^{boot}$. Then the estimation variance, the second component of formula (3), can be approximated by the variance of the predictive distribution: $Var\left[\hat{\mu}_i^{boot}\right]$. And finally the bootstrap estimation of prediction error is:

$$\sqrt{\left(\hat{\phi}^P \Big/ w_i\right)\hat{\mu}_i^{\xi} + Var\left[\hat{\mu}_i^{boot}\right]}. \tag{8}$$

3 Claim Reserving

Consider a portfolio of risks and assume that each claim is settled either in the accident year or in the following k development years. Consider a family of random variables $\left\{c_{ij}\right\}_{i,j\in\{0,1,\ldots k\}}$, where c_{ij} is the amount of claim losses of accident year

i which is paid with a delay of j years and hence in development year j and in calendar year $i+j$. Assume that the incremental losses c_{ij} are observable for calendar years $i + j \leq k$ and that they are collected in a run-off triangle, where the rows correspond to the accident years, the columns to the development years and the against-diagonals to the calendar years.

The future payments for the different accident years $i = 1, \ldots, k$ are obtained by adding the predicted future incremental losses in the corresponding row of the square:

$$FP_i = \sum_{j=k-i+1}^{k} \hat{c}_{ij}. \tag{9}$$

The total future payment is calculated by adding all the predicted future incremental losses in the bottom-right part of the run-off triangle:

$$FP = \sum_{i=1}^{k} \sum_{j=k-i+1}^{k} \hat{c}_{ij}. \tag{10}$$

And the future payments for the different calendar years $t = k + 1, \ldots, 2k$ are obtained by adding the incremental losses that were made in the future calendar years, i.e., the values of the same against-diagonal t:

$$FP_t = \sum_{j=t-k}^{k} \hat{c}_{t-j,j}. \tag{11}$$

It is well known that several often used methods to complete a run-off triangle can be described by a GLM: the Chain-Ladder method, the arithmetic and geometric separation methods and the de Vylder's least squares method. In particular, the classic Chain-Ladder deterministic method (see [17]) can be derived from a GLM by assuming over-dispersed Poisson distribution, the logarithmic link function and by including the dummies corresponding to the accident and development years in the linear predictor, with prior weights equal to one (see, e.g., [3, 10, 11] and [12]). In that case, the provision estimations coincide with both methods, the classic Chain-Ladder and the GLM. For this reason GLM can be considered a stochastic Chain-Ladder method of claim reserving. In this paper we propose the use of DB-GLM. Because DB-GLM contains ordinary GLM as a particular instance when we assume the Euclidean l^2 metric between factors, DB-GLM generalizes the same claim reserving methods as those generalized by GLM.

In this study, we develop the formulas of the prediction error for DB-GLM when we assume the power family of distributions (1) and the power family of links (6). For these families the Chain-Ladder method is reproduced when we choose $\xi = 1$ and $\lambda = 0$ and we use the Euclidean l^2 metric. To illustrate computations, in the next section we assume $\xi = 1$ and $\xi = 2$, the Poisson and the Gamma distributions as in [3, 10, 11] and [12].

Applying (4), and assuming the logarithmic link as in (7), the MSEs for the future payments by accident years, for $i = 1, \ldots, k$ are:

$$\begin{aligned} E\left[\left(FP_i - \widehat{FP_i}\right)^2\right] &\cong \sum_{\substack{j=1,..,k\\ i+j>k}} \phi\mu_{ij}^{\xi} + \mu_i^T Var\left[\eta_i\right]\mu_i \\ &= \sum_{\substack{j=1,..,k\\ i+j>k}} \phi\mu_{ij}^{\xi} + \sum_{\substack{j=1,..,k\\ i+j>k}} \mu_{ij}^2 Var\left[\eta_{ij}\right] + 2 \sum_{\substack{j_1,j_2=1,..,k\\ j_2>j_1\\ i+j_1>k, i+j_2>k}} \mu_{ij_1}\mu_{ij_2} Cov\left[\eta_{ij_1}, \eta_{ij_2}\right] \end{aligned} \tag{12}$$

The MSE for the total future payment is:

$$\begin{aligned} E\left[\left(FP - \widehat{FP}\right)^2\right] &\cong \sum_{\substack{i,j=1,...,k\\ i+j>k}} \phi\mu_{ij}^{\xi} + \mu^T Var\left[\eta\right]\mu \\ &= \sum_{\substack{i,j=1,...,k\\ i+j>k}} \phi\mu_{ij}^{\xi} + \sum_{\substack{i,j=1,...,k\\ i+j>k}} \mu_{ij}^2 Var\left[\eta_{ij}\right] + 2 \sum_{\substack{i_1,j_1,i_2,j_2=1,...,k\\ i_1+j_1>k, i_2+j_2>k\\ i_1j_1\neq i_2j_2}} \mu_{i_1j_1}\mu_{i_2j_2} Cov\left[\eta_{i_1j_1}, \eta_{i_2j_2}\right] \end{aligned} \tag{13}$$

And the MSEs for the future payments by calendar years, for $t = k+1, \ldots, 2k$ are:

$$\begin{aligned} E\left[\left(FP_t - \widehat{FP_t}\right)^2\right] &\cong \sum_{\substack{i,j=1,..,k\\ i+j=t}} \phi\mu_{ij}^{\xi} + \mu_t^T Var\left[\eta_t\right]\mu_t \\ &= \sum_{\substack{i,j=1,..,k\\ i+j=t}} \phi\mu_{ij}^{\xi} + \sum_{\substack{i,j=1,..,k\\ i+j=t}} \mu_{ij}^2 Var\left[\eta_{ij}\right] + 2 \sum_{\substack{i_1,i_2,j_1,j_2=1,..,k\\ i_1j_1\neq i_2j_2\\ i_1+j_1=t, i_2+j_2=t}} \mu_{i_1j_1}\mu_{i_2j_2} Cov\left[\eta_{i_1j_1}, \eta_{i_2j_2}\right]. \end{aligned} \tag{14}$$

The bootstrap estimations (8) of the prediction error for the future payments by accident years (9), $PE^{boot}(FP_i)$ for $i = 1, \ldots, k$, are:

$$PE^{boot}(FP_i) \cong \sqrt{\sum_{\substack{j=1,..,k\\ i+j>k}} \hat{\phi}^P \hat{c}_{ij}^{\xi} + Var\left[\widehat{FP}_i^{boot}\right]}. \tag{15}$$

The bootstrap estimation of the prediction error for the total future payment (10), $PE^{boot}(FP)$, is:

$$PE^{boot}(FP) \cong \sqrt{\sum_{\substack{i,j=1,...,k\\ i+j>k}} \hat{\phi}^P \hat{c}_{ij}^{\xi} + Var\left[\widehat{FP}^{boot}\right]}. \tag{16}$$

And the bootstrap estimations of the prediction errors for the future payments by calendar years (11), $PE^{boot}(FP_t)$ for $t = k+1, \ldots, 2k$, are:

$$PE^{boot}(FP_t) \cong \sqrt{\sum_{\substack{i,j=1,..,k \\ i+j=t}} \hat{\phi}^P \hat{c}_{ij}^{\xi} + Var\left[\widehat{FP}_t^{boot}\right]}. \quad (17)$$

4 Numerical Example

In Table 1 we show the triangle of Taylor and Ashe (see [16]) with incremental losses to illustrate the calculus of the best estimate of provisions (without risk margin). This dataset is used in many texts on claim reserving as are [1–3, 10, 11, 14] and [15] to illustrate the use of the GLM and other techniques.

In Tables 2 and 3 we show the estimations of the future incremental losses for the Poisson, $\xi = 1$, and the Gamma, $\xi = 2$, distributions of the power family (1). We assume the logarithmic link, $\lambda = 0$ in the power family of links (6), and the l^2 metric between accident and development factors.

Table 1 Run-off triangle of [16] with 55 incremental losses

	0	1	2	3	4	5	6	7	8	9
0	357,848	766,940	610,542	482,940	527,326	574,398	146,342	139,950	227,229	67,948
1	352,118	884,021	933,894	1,183,289	445,745	320,996	52,7804	266,172	425,046	
2	290,507	1,001,799	926,219	1,016,654	750,816	146,923	495,992	280,405		
3	310,608	1,108,250	776,189	1,562,400	272,482	352,053	206,286			
4	443,160	693,190	991,983	769,488	504,851	470,639				
5	396,132	937,085	847,498	805,037	705,960					
6	440,832	847,361	1,131,398	1,063,269						
7	359,480	1,061,648	1,443,370							
8	376,686	986,608								
9	344,014									

Table 2 Fitted values with DB-GLM assuming the over-dispersed Poisson distribution, the logarithmic link and the l^2 metric for the run-off triangle of Table 1

	1	2	3	4	5	6	7	8	9
1									94,634
2								375,833	93,678
3							247,190	370,179	92,268
4						334,148	226,674	339,456	84,611
5					383,287	351,548	238,477	357,132	89,016
6				605,548	414,501	389,349	264,121	395,534	98,588
7			1,310,258	725,788	508,792	466,660	316,566	474,073	118,164
8		1,018,834	1,089,616	603,569	423,113	388,076	263,257	394,241	98,266
9	856,804	897,410	959,756	531,636	372,687	348,126	231,182	347,255	86,555

Table 3 Fitted values with DB-GLM assuming the Gamma distribution, the logarithmic link and the l^2 metric for the run-off triangle of Table 1

	1	2	3	4	5	6	7	8	9
1									93,316
2								356,295	90,212
3							214,335	316,641	80,172
4						339,733	228,765	337,960	85,569
5					418,059	354,459	238,681	352,609	89,278
6				613,160	452,559	383,710	258,379	381,708	96,646
7			1,225,659	684,190	504,985	428,160	288,310	425,926	107,842
8		982,882	1,049,906	586,081	432,573	366,765	246,968	364,851	92,378
9	853,417	873,269	932,819	520,720	384,331	325,862	219,426	324,162	82,076

If we put by rows in `cij` the data of Table 1, the instructions with the function `dbglm` of the `dbstats` package for `R` to fit the distance-based models are:

```
R> n <- length(cij)
R> k <- trunc(sqrt(2*n))
R> i <- rep(1:k, k:1); i <- as.factor(i)
R> j <- sequence(k:1); j <- as.factor(j)
R> dbglm.Poisson <- dbglm(cij ~ i + j, family = quasipoisson,
 metric = "euclidean", method = "rel.gvar", rel.gvar = 1)
R> dbglm.Gamma <- dbglm(cij ~ i + j, family = Gamma(link =
 "log"), metric = "euclidean", method = "rel.gvar",
 rel.gvar = 1)
```

In Tables 4 and 5 we show the future payments by calendar years, the prediction errors and the coefficients of variation for the Poisson and Gamma models in the

Table 4 Future payments by calendar years, prediction errors and coefficients of variation for DB-GLM assuming over-dispersed Poisson and Gamma distributions, the logarithmic link and the l^2 metric, using the formula (14) for the run-off triangle of Table 1

Calendar year	Payment (Poisson)	Prediction error (Poisson)	Coefficient of variation (Poisson)	Payment (Gamma)	Prediction error (Gamma)	Coefficient of variation (Gamma)
10	5,226,536	747,370	14.30%	5,096,855	847,282	16.62%
11	4,179,394	710,144	16.99%	4,050,002	749,550	18.51%
12	3,131,668	644,140	20.57%	3,064,408	628,141	20.50%
13	2,127,272	479,126	22.52%	12,078,011	431,886	20.78%
14	1,561,879	404,968	25.93%	1,510,393	345,881	22.90%
15	1,177,744	364,295	30.93%	1,095,403	292,256	26.68%
16	744,287	294,425	39.56%	692,118	220,058	31.79%
17	445,521	250,987	56.34%	416,540	181,227	43.51%
18	86,555	108,269	125.09%	82,076	47,918.1	58.38%
Total	18,680,856	2,945,659	15.77%	18,085,805	2,702,710	14.94%

Table 5 Future mean payments by calendar years, prediction errors and coefficients of variation for DB-GLM assuming over-dispersed Poisson and Gamma distributions, the logarithmic link and the l^2 metric, using pairs bootstrap with size 1000 to estimate (17) for the run-off triangle of Table 1

Calendar year	Mean payment (Poisson)	Prediction error (Poisson)	Coefficient of variation (Poisson)	Mean payment (Gamma)	Prediction error (Gamma)	Coefficient of variation (Gamma)
10	5,322,048	651,320	12.46%	5,237,837	760,410	14.92%
11	4,277,227	595,875	14.25%	4,185,515	659,058	16.27%
12	3,250,817	522,637	16.68%	3,199,644	547,190	17.86%
13	2,240,502	436,147	20.50%	2,211,180	407,490	19.61%
14	1,671,934	374,469	23.97%	1,641,125	338,333	22.40%
15	1,273,321	327,034	27.76%	1,229,713	290,089	26.48%
16	861,295	276,203	37.10%	829,033	238,753	34.50%
17	543,109	217,778	48.88%	534,532	190,842	45.82%
18	174,216	144,268	166.67%	179,832	132,739	161.73%
Total	19,554,135	2,231,054	11.94%	19,248,389	2,270,695	12.55%

Table 6 Present values of the future payments by calendar years for DB-GLM assuming over-dispersed Poisson and Gamma distributions, the logarithmic link and the l^2 metric, assuming the risk free curves that must be used at 31-01-2016 and published in the web of EIOPA, for the run-off triangle of Table 1

Calendar year	Deferral (in years)	Term structure	Payment (Poisson)	Payment (Gamma)
10	1	0.036%	5,226,536	5,096,855
11	2	0.015%	4,179,394	4,050,002
12	3	0.067%	3,131,668	3,064,408
13	4	0.155%	2,127,272	2,078,011
14	5	0.267%	1,561,879	1,510,393
15	6	0.389%	1,177,744	1,095,403
16	7	0.514%	744,287	692,118
17	8	0.641%	445,521	416,540
18	9	0.760%	86,555	82,076
Present value			18,556,354	17,967,985

case of using the asymptotic formulation of (14) and in the case of estimating by bootstrapping the estimation variance as in formula (17).

In Table 6 we calculate the present value of the future payments by calendar years for the Poisson and Gamma models using the risk free curves at 31-01-2016 of https://eiopa.europa.eu/regulation-supervision/insurance/solvency-ii-technical-information/risk-free-interest-rate-term-structures. The risk free interest rates correspond to annual zero-coupon spot rates. In particular, for the calculations of the present values in the paper we choose the *basic risk free curves with volatility adjustment* for the euro, included in the file *EIOPA_RFR_20160131_Term_Structures.xlsx*.

Table 7 Standard deviation, $VaR_{77.5}$ and $VaR_{99.5}$ of the predictive distributions of the future payments by calendar years for DB-GLM assuming over-dispersed Poisson and Gamma distributions, the logarithmic link and the l^2 metric, using pairs bootstrap with size 1000, for the run-off triangle of Table 1

Calendar year	Std Dev. (Poisson)	$VaR_{77.5}$ (Poisson)	$VaR_{99.5}$ (Poisson)	Std Dev. (Gamma)	$VaR_{77.5}$ (Gamma)	$VaR_{99.5}$ (Gamma)
10	656,450	5,838,751	7,154,049	389,687	5,524,047	6,323,486
11	619,952	4,681,521	6,154,360	369,609	4,457,042	5,350,769
12	531,939	3,629,494	4,734,649	332,197	3,435,365	4,230,436
13	447,469	2,577,467	3,576,893	278,398	2,418,214	2,963,346
14	386,091	1,946,250	2,788,398	244,434	1,817,224	2,337,500
15	333,985	1,525,439	2,262,122	215,448	1,396,467	1,815,355
16	285,609	1,052,027	1,683,507	191,483	1,011,477	1,280,757
17	230,971	736,419	1,209,831	156,331	706,661	880,765
18	153,242	263,007	631,216	130,036	344,003	425,190

In Table 7 we include the standard deviation of the predictive distributions of the future payments by calendar years, and the value at risk (VaR) for the confidence levels 77.5% and 99.5% for DB-GLM assuming over-dispersed Poisson and Gamma distributions, the logarithmic link and the l^2 metric, using pairs bootstrap with size 1000. We use a sample size of 1000 for the bootstrap because the coefficients of variation have four decimals of precision (a sample size of 500 is enough for two decimals).

5 Conclusions

We propose the use of DB-GLM as a methodology to solve the problem of claim reserving. To complete the tool we design a bootstrapping pairs procedure which allows us to estimate the predictive distribution of the future payments for reserving.

DB-GLM has GLM as a particular case when we use the Euclidean l^2 metric, i.e., we obtain the same provision estimations. As it is well known, the Chain-Ladder classic method is reproduced when we assume (over-dispersed) Poisson distribution and logarithmic link function. Then, DB-GLM can be considered a stochastic Chain-Ladder method.

In this study we show the formulas of the prediction errors for the accident years future payments, (12) and (15), for the total future payment, (13) and (16), and for the calendar years future payments, (14) and (17), when we assume the power family of distributions (1) and the power family of links (6) in DB-GLM.

It is of relevance to calculate provisions taking into account the Solvency II Directive. We apply an interest term structure published by EIOPA. It is of interest to estimate the predictive distributions of the best estimates of the future payments by calendar years if we want to calculate the reserve risk in the own risk & solvency assessment, ORSA, and to estimate the undertaking specific parameters, USP, of the company.

We illustrate the distance-based method of claim reserving with the run-off triangle of Taylor and Ashe used in many actuarial papers. We assume in DB-GLM the Poisson, $\xi = 1$, and the Gamma, $\xi = 2$, distributions of the power family (1), the logarithmic link, $\lambda = 0$ of the power family of links (6), and the Euclidean l^2 metric between accident and development factors. We use the function `dbglm` of the `dbstats` package for `R` to fit the distance-based models.

Acknowledgements Work supported by the Spanish Ministerio de Educación y Ciencia, grant MTM2014-56535-R.

References

1. Boj, E., Costa, T.: Claim reserving: calendar year reserves for the GLM. In: Guillén, M., et al. (eds.) Current Topics on Risk Analysis: ICRA6 and RISK2015 Conference. Cuadernos de la Fundación MAPFRE, vol. 205, pp. 169–177. Fundación MAPFRE, Madrid (2015)
2. Boj, E., Costa, T.: Claim reserving using distance-based generalized linear models. In: Cao, R., González-Manteiga, W., Romo, J. (eds.) Nonparametric Statistics. Springer Proceedings in Mathematics & Statistics, vol. 175, pp. 135–148. Springer, Cham (2016)
3. Boj, E., Costa, T.: Provisions for claims outstanding, incurred but not reported, with generalized linear models: prediction error formulation by calendar years. Cuadernos de Gestión **17**(2), 157–174 (2017)
4. Boj, E., Claramunt, M.M., Fortiana, J.: Selection of predictors in distance-based regression. Commun. Stat. Theory Methods **36**, 87–98 (2007)
5. Boj, E., Caballé, A., Delicado, P., Fortiana, J.: dbstats: distance-based statistics (dbstats). R Package Version 1.4 (2014). http://CRAN.R-project.org/package=dbstats
6. Boj, E., Costa, T., Fortiana, J.: Prediction error in distance-based generalized linear models. Presented in the Conference of the International Federation of Classification Societies, IFCS2015, Bologna, Italy (2015)
7. Boj, E., Delicado, P., Fortiana, J., Esteve A., Caballé, A.: Global and local distance-based generalized linear models. TEST **25**, 170–195 (2015)
8. Boj, E., Costa, T., Fortiana, J., Esteve, A.: Assessing the importance of risk factors in distance-based generalized linear models. Methodol. Comput. Appl. Probab. **17**, 951–962 (2015)
9. Efron, B., Tibshirani, J.: An Introduction to the Bootstrap. Chapman and Hall, New York (1998)
10. England, P.D.: Addendum to 'Analytic and bootstrap estimates of prediction errors in claim reserving'. Insur. Math. Econ. **31**, 461–466 (2002)
11. England, P.D., Verrall, R.J.: Analytic and bootstrap estimates of prediction errors in claims reserving. Insur. Math. Econ. **25**, 281–293 (1999)
12. England, P.D., Verrall, R.J.: Predictive distributions of outstanding liabilities in general insurance. Ann. Actuarial Sci. **1**(2), 221–270 (2006)
13. McCullagh, P., Nelder, J.A.: Generalized Linear Models, 2nd edn. Chapman and Hall, London (1989)
14. Renshaw, A.E.: Chain ladder and interactive modelling (claims reserving and GLIM). J. Inst. Actuaries **116**(3), 559–587 (1989)

15. Renshaw, A.E.: On the second moment properties and the implementation of certain GLIM based stochastic claims reserving models. Actuarial Research Paper 65. Department of Actuarial Science and Statistics, City University, London (1994)
16. Taylor, G., Ashe, F.R.: Second moments of estimates of outstanding claims. J. Econ. Bus. **23**, 37–61 (1983)
17. van Eeghen, J., Greup, E.K., Nijssen, J.A.: Loss Reserving Methods. Surveys of Actuarial Studies, vol. 1. National Nederlanden, Rotterdam (1981)

Profitability vs. Attractiveness Within a Performance Analysis of a Life Annuity Business

Emilia Di Lorenzo, Albina Orlando, and Marilena Sibillo

Abstract Combining insurer's profitability with products' attractiveness in terms of marketing competitiveness is a critical issue within the risk/profit management of an insurance business. In particular life insurance products are characterized by the presence of financial and demographic risk sources, whose combined effect requires suitable management strategies. This paper deals with the impact of the load factor on life annuity portfolio performance from the insurers point of view. The aim is to build a performance indicator that clearly points out the role of the load factor in the performance making, giving to it a central role in the company management strategy. Such index is characterized by a simple mathematical structure and fits to the purpose: in fact it provides clear indications to the manager about the influence of the load factor on the performance of the life annuity business line.

1 Life Annuity Portfolios and Profitability Indexes: A Tangled Issue

Life insurance business is moving towards contractual models increasingly tailor-made following the purpose of paying attention to both the contracting parties' characteristics and taking into account the dynamics of the financial and the demographic contexts in which the inflows and the outflows are going to be valued. The management activity is continuously engaged in monitoring the business performances and in controlling that the risks impacting on them have been correctly

E. Di Lorenzo
Department of Economic and Statistical Sciences, University of Naples Federico II – Via Cinthia, Complesso Monte Sant'Angelo, 80126 Naples, Italy
e-mail: diloremi@unina.it

A. Orlando
National Research Council – Via P. Castellino, 80128 Naples, Italy
e-mail: A.Orlando@iac.cnr.it

M. Sibillo (✉)
Department of Economics and Statistics, University of Salerno – Campus Universitario, 84084 Fisciano (SA), Italy
e-mail: msibillo@unisa.it

M. Corazza et al. (eds.), *Mathematical and Statistical Methods for Actuarial Sciences and Finance*, https://doi.org/10.1007/978-3-319-50234-2_9

managed. The valuation of the insurance product performance at the issue time holds a meaningful task: it contains several information of exploratory and guiding nature. As long ago as 2007, Easton and Harris [8] stated that company's performance and its efficiency are well described by the profitability ratios, referred both to the company as a whole and to a specific business line as life annuities. The topic is deepened in (1), in which the performance of pension plans is studied. Understanding the concept of profitability and exploring the efficient way for representing it by means of a synthetic index, is a significant topic; the management can make use of efficient indicators for internal control aims and for communicating outside (to the policyholders and the stakeholders) the health of the company or of a specific portfolio, in other words if and how they give rise to value and profits. Staying in the case of life annuity business, mainly saving products, profitability has to be valued in the long-term perspective, as it is implicit in the contract structure, and concerning the specific risks impacting on the product under consideration. As clearly explained by Swiss Re 2012 [13], the high number of years during which the policy remains in force in the portfolio and the high number of payments in and out of the portfolio, make the performance valuation very difficult.

The present analysis is based on the actuarial control of the payments [1], [2] valued on assumptions about the future. The length of the future to take into account in the valuations is the aspect making definitely different the performance measures used in the non life sector, mainly short-term pointed, and the life sector, with a long-term perspective. These different perspectives make complex the performance valuations at a group level. De Mey [6] reasons on the meaning of the performance measures in life and non-life insurance and concludes that an integration between Economic Value Added (EVA) and market-consistent embedded value (MCEV) seems to be the best way to manage the performance measure of an insurance company. Kraus [11] highlights the two most popular performance metrics within the two sections: EVA and risk-adjusted return on capital (RAROC) for non-life and MCEV for life and proposes a unified approach for measuring the performance of the company as a whole.

The analysis we present is framed specifically in the life annuity portfolios and in their performance valuation, aiming at getting actuarial quantities easy to calculate and concise enough to produce indices clearly interpretable and readily communicable, as introduced by Coppola et al. in [3] and [4]. In this sense, one of the most expressive first indices proposed in life annuities performance analysis can be found in [10], then regained in [5]: it is the Funding Ratio, defined as the ratio between the market value of assets and liabilities, projected from the issue time 0 with stochastic assumptions for the main risk drivers, and valued at a certain time t. It can usefully be interpreted as the amount of assets for each unit of liability, and, in this sense, it is a useful solvency measure. Still assuming stochastic hypotheses for financial and demographic variables, in [7] the Authors propose profitability indices for variable life annuity portfolio founded on the surplus value, explained as the difference between retrospective gain and prospective loss. In particular they introduce actuarial restyled versions for two popular profitability indexes and discuss their behavior as function of the time of valuation.

This paper deepens the topic considering the influence of the loading factors on the life annuity portfolio performance. This aspect is crucial in the correct assessment of the equilibrium between insurer and insured and impacts on the appeal of the life annuities as well as pension annuities contracts with the potential insureds. The issue is already treated in [9]. Here Friedman and Warshawsky study the annuity prices in the United States, deepening the differences between the fair value of the contracts and their loaded prices. Interesting tables show the impact of the load factors on the will of subscribe them and set if the values are or are not actuarially fair. In [12] it is shown that the load factor can be considered among the main factors explaining the lack of participation in the voluntary annuity market. The analysis is performed from the insureds point of view; the Author obtains the threshold load factor for defined levels of financial wealth at retirement, setting the limit beyond which the potential insured will consider disadvantageous the life annuity contract, deciding to remain outside the market.

Moving the lens to the business strategy, the paper we present deals with the impact from the insurers point of view. The aim is to build a performance indicator that clearly points out the role of the load factor in the performance making, giving to it a central role in the company management strategy. The new index we propose is characterized by a simple mathematical structure that at the same time is able to transfer clear indications to the manager about the influence of the load factor on the performance of the life annuity business line.

In Sect. 2 financial quantities and formulas are presented. The performance ratio expressed as function of the load factor is introduced and its behavior is studied. In Sect. 3 an empirical application is presented. The stochastic scenario on which the results have been obtained is outlined and several numerical results are collected in tables referred to a specific life annuity contract at different times of valuations. Section 4 also concludes the paper with some final observations.

2 The Impact of the Load Factors on the Life Annuity Business Performance

Let us consider a portfolio of homogeneous deferred life annuities, where each policy is issued to each of N_0 lives aged x, with constant installment R payable at the end of each year while (x) survives and τ the deferment period.

The insured pays periodic constant premiums P, during all the deferment period, with load factor θ.

Aim of the paper is to structure an index able to synthesize the portfolio performance as function of the load factors. The index we propose is based on S_t, the portfolio surplus at time t and on A_t, the *unconstrained assets* at the same time. The portfolio surplus is given by the difference between the accumulated value of past premiums collected net of past benefits paid and the discounted value of future obligations net of future premiums. The *unconstrained assets* are the revenues net of the residual debt. The mathematical expressions are respectively the following:

$$S_t = \sum_j N_j((P + P\theta)\mathbf{1}_{j<\tau} - R\mathbf{1}_{j>\tau})v(t,j) \tag{1}$$

$$A_t = \sum_{j\leq t} N_j(P + P\theta)\mathbf{1}_{j<\tau}v(t,j) - \sum_{j>t} N_j(R\mathbf{1}_{j>\tau} - (P + P\theta)\mathbf{1}_{j<\tau})v(t,j) \tag{2}$$

$v(t,j)$ being the stochastic value at time t of one monetary unit at time j and N_j the number of survivors at time j, belonging to the initial cohort of N_0 lives at time 0. It is straightforward to verify that the difference between A_t and S_t gives the value in t of the residual insurer's obligations:

$$A_t - S_t = R\sum_{j\leq t} N_j\mathbf{1}_{j>\tau}v(t,j) \tag{3}$$

The influence of the load factors on the life annuity portfolio performance is a crucial issue. Within this context, the ratio $\frac{S_t}{A_t}$ provides interesting suggestions concerning the correct assessment of the equilibrium between insurer and insured, as well as the appeal of life annuity contracts with the potential insureds. Such a ratio is a measure of the financial health of the portfolio, quantified by the portfolio surplus, measured per unit of unconstrained assets. The ratio can be expressed by means of the following formula:

$$\frac{S_t}{A_t} = 1 - \frac{R\sum_{j\leq t} N_j\mathbf{1}_{j>\tau}v(t,j)}{A_t} = \tag{4}$$

$$= 1 - \frac{R\sum_{j\leq t} N_j\mathbf{1}_{j>\tau}v(t,j)}{\sum_{j\leq t} N_j(P + P\theta)\mathbf{1}_{j<\tau}v(t,j) + \sum_{j>t} N_j((P + P\theta)\mathbf{1}_{j<\tau} - R\mathbf{1}_{j>\tau})v(t,j)}$$

We assume that during the portfolio life the stochastic demographic and financial variables are independent and that their descriptions are assumed at the issue time. The performance ratio we introduce, is $\Psi_t^{(x)}(\theta) = \frac{E[S_t]}{E[A_t]}$, is *the Loading Performance Ratio* (LPR therein):

$$\Psi_t^{(x)}(\theta) = 1 - \frac{R\sum_{j\leq t} E[N_j]\mathbf{1}_{j>\tau}E[v(t,j)]}{K_t(\theta) + H_t(\theta)}$$

with

$$K_t(\theta) = \sum_{j\leq t} E[N_j](P + P\theta)\mathbf{1}_{j<\tau}E[v(t,j)]$$

and

$$H_t(\theta) = \sum_{j>t} E[N_j]((P + P\theta)\mathbf{1}_{j<\tau} - R\mathbf{1}_{j>\tau})E[v(t,j)]$$

It holds:

$$\lim_{\theta\to\infty} \Psi_t^{(x)}(\theta) = 1 \tag{5}$$

This means that when the load factor tends to infinity, the whole of the free assets flows into the surplus of the insurer and discloses for him the most desirable, even if unreasonable, situation. The more the LPR is close to 1, that is the more free assets converges with the surplus of the business at that time, the best the performance of the product.

On the other hand, we have:

$$\lim_{\theta\to 0} \Psi_t^{(x)}(\theta) = \tag{6}$$

$$= 1 - \frac{R\sum_j E[N_j]\mathbf{1}_{j>\tau}E[v(t,j)] - R\sum_{j>t} E[N_j]\mathbf{1}_{j>\tau}E[v(t,j)]}{\sum_j E[N_j]P\mathbf{1}_{j<\tau}E[v(t,j)] - R\sum_{j>t} E[N_j]\mathbf{1}_{j>\tau}E[v(t,j)]}$$

On the basis of the equity principle, the premium calculations and the valuation process are performed using the same financial and demographic bases. From formula 6 we can write:

$$\lim_{\theta\to 0} \Psi_t^{(x)}(\theta) = 0$$

Summarizing:

$$0 \le \Psi_t^{(x)}(\theta) \le 1$$

and the higher the LPR, the better life product performance. This is confirmed by the constant value 1 the LPR assumes during the deferment period, in which only inflows come true.

3 Numerical Examples

In this section, referring to a specific pension annuity portfolio, we develop an application for getting some evidence of the behaviour of the performance index $\Psi_t^{(x)}$ and reflect upon its practical implications.

The contract we are taking into account is characterized by a deferment period of 10 years, during which the insurer only perceives premiums and accumulates them. During this time interval no payment has be done yet: the insurer picks up inflows due to the premium payments made by the insureds, no outflow has ate away the funding amount he is building up and these circumstances involve that this is the most healthy period for the insurer.

The situation is perfectly photographed by the index LPR, constantly taking value 1, pointing out the extremely profitable situation the insurer would get if the contract would end exactly during this period.

The LPR index produces a sort of snapshot of the financial situation of the business line in the specific time in which the valuation is done and marks, broadly speaking, the financial status if the portfolio or the single contract (the index is independent on the number of contracts) would be excluded within the deferment. An eventual discharge of the contract within 10 years would be mostly profitable for the insurer. Due to this reason, the index LPR takes sense strictly during the decumulation/annuitization period, that is after the deferment.

We study the LPR in the case of a portfolio of deferred lifelong life annuities, with deferment period $\tau = 10$, each one issued on a person aged x. Within this application, the constant premiums, paid during all the deferment period, are calculated by means of the technical rate 0.02 and the Italian male survival probabilities inferred from the Lee-Carter model.

The stochastic interest rate environment consists of a CIR process, for which we assume—by way of an example—the long term mean, the volatility and the drift equal to 0.025, 0.0062 and 0.02962, respectively.

Recalling what above observed, the tables are referred to times of valuations belonging to the annuitization period; in our example $t \geq 10$.

In Tables 1, 2 and 3 we report the values of $\Psi_t^{(x)}(\theta)$ at the valuation times $t = 15, 20, 30$, as function of $\theta = 0.05h$, with $h = 0, 1, \ldots, 6$, considering the ages 40, 45, 50, 60.

We observe an increasing behaviour of $\Psi_t^{(x)}(\theta)$, when θ increases and, in general, higher values of $\Psi_t^{(x)}(\theta)$ for simultaneous higher values of ages at issue and θ. For a fixed value of θ, $\Psi_t^{(x)}(\theta)$ increases with the ages at issue, by virtue of the smaller time duration where the insurer's obligations come true. When the valuation time increases, ceteris paribus, $\Psi_t^{(x)}(\theta)$ decreases and the weight of the installments already carried out is increasingly heavy.

Table 1 $\Psi_t^{(x)}(\theta)$, age from 40 to 60, $t = 15$, loading factor from 0 to 0.30

θ	$x = 40$	$x = 45$	$x = 50$	$x = 60$
0	0.0053	0.047	0.1412	0.1758
0.05	0.1369	0.2060	0.2922	0.3285
0.10	0.2377	0.3217	0.3980	0.4335
0.15	0.3177	0.4080	0.4763	0.5101
0.20	0.3821	0.4748	0.5366	0.5684
0.25	0.4356	0.5281	0.5844	0.6144
0.30	0.5805	0.5715	0.6233	0.6514

Table 2 $\Psi_t^{(x)}(\theta)$, age from 40 to 60, $t = 20$, loading factor from 0 to 0.30

θ	$x = 40$	$x = 45$	$x = 50$	$x = 60$
0	0.0043	0.0395	0.1187	0.148
0.05	0.0695	0.1492	0.2119	0.2398
0.10	0.1369	0.2258	0.2872	0.3171
0.15	0.1952	0.2897	0.3494	0.3801
0.20	0.2461	0.3439	0.4016	0.4324
0.25	0.2909	0.3904	0.4460	0.4767
0.30	0.3307	0.4307	0.4843	0.5145

Table 3 $\Psi_t^{(x)}(\theta)$, age from 40 to 60, $t = 30$, loading factor from 0 to 0.30

θ	$x = 40$	$x = 45$	$x = 50$	$x = 60$
0	0.0032	0.0433	0.0979	0.1222
0.05	0.0561	0.1012	0.1564	0.1831
0.10	0.0247	0.1524	0.2078	0.2361
0.15	0.0706	0.1981	0.2532	0.2826
0.20	0.1125	0.2392	0.2938	0.3238
0.25	0.1507	0.2762	0.3301	0.3605
0.30	0.1875	0.3098	0.3630	0.3935

The ratio $\Psi_t^{(x)}(\theta)$ always decreases when t increases (fixing the age and the load value). The best financial situation of the insurer happens when he collects and this is photographed by $\Psi_t^{(x)}(\theta) = 1$ during the deferment. The more the insurer pays, the worse the product performance. The ratio increases with x (fixing the load factor and the time of valuation), and this can easily be explained observing that, being in all likelihood (thanks to the longevity) the same the number of collected premiums, the number of payments most likely decreases. Of course the ratio increases with θ, for each age at issue and time of valuation.

The corresponding Figs. 1, 2 and 3 clearly mark the afore explained behaviour of the LPR and highlight its trend in relation to possible couples of values of the insureds' age and load factors.

Even more explicitly, in Figs. 1, 2 and 3 we can observe the trends followed by the index LPR when age and load factor vary, in the three illustrative cases chosen for developing the preceding numerical application.

The figures refer respectively to the valuation times $t = 15$, 20 and 30. Among these three cases, the best financial status, pointed out by high values of LPR, happens when $t = 15$, the age at issue is 60 and of course theta is 0.30. This suggests that the future benefit payments are well counterbalanced by the almost certain premium payments of a 60 years old contractor, in particular if the time of valuation is close to the end of the deferment period. The more the contract lives beyond the end of the deferment period, the less the product performance, fixing all the other variables. Moreover, the more the time of valuation is far from the issue time, the stronger the increasing trend of the index with the age at issue is, in particular for high values of the load factor.

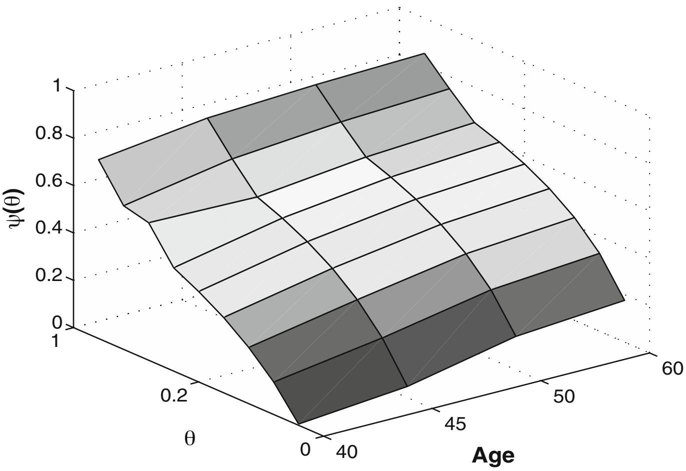

Fig. 1 LPR(15)

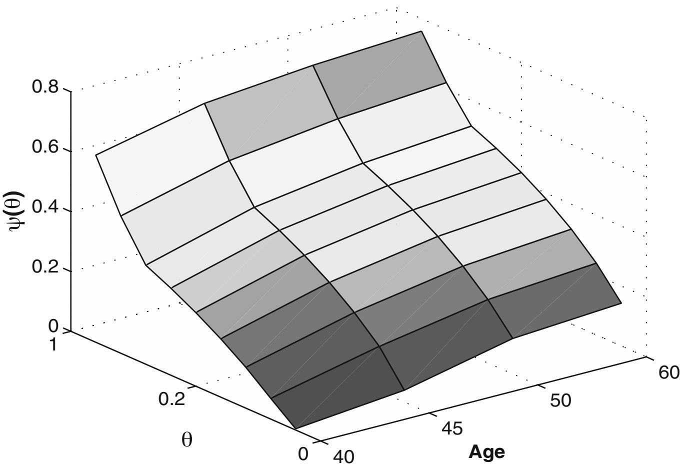

Fig. 2 LPR(20)

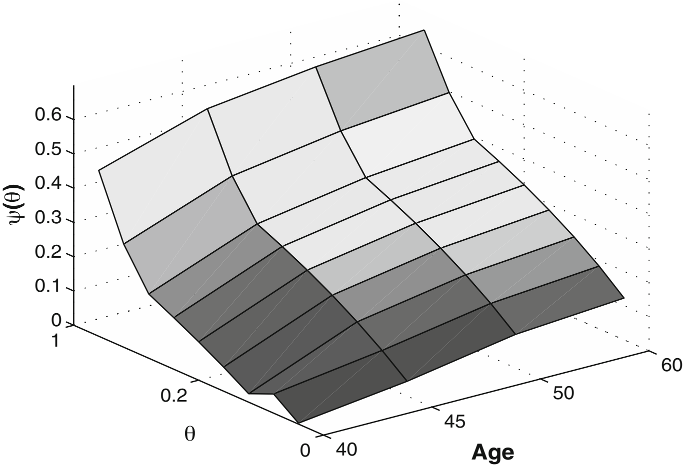

Fig. 3 LPR(30)

4 Conclusions

To conclude, the index we propose constitutes a meaningful quantification of a life annuity performance reaction to the load factor change. We developed the analysis in the forward perspective with respect of the issue time, considering the insurers purpose of analysing the product performance at the moment of its launching. The study can be managed year by year updating the stochastic financial and demographic scenarios by the opportune re-calibration of the processes involved in the calculus.

The ratio we introduce, the Loading Performance Ratio (LPR), therefore measures the level of surplus on assets, depending on premium load, time horizon, age of the insured. It provides guidance about the *rating* of the insurance business framed within the market conditions, which take into account also the product attractiveness, in term of costs for the insured.

References

1. Andreu, L., Swinkels, L.A.P.: Performance evaluation of balanced pension plans. Quant. Finan. **12**(5), 819–830 (2012)
2. Beal, R.W.: Bridging the gap between ROE and IRR. North Am. Actuarial J. **4**, 1–11 (2002)
3. Coppola, M., Di Lorenzo, E., Orlando, A., Sibillo, M.: Measuring demographic uncertainty via actuarial indexes. In: Skiadas, C.H. (ed.) Recent Advances in Stochastic Modelling and Data Analysis, pp. 122–129. World Scientific, Singapore (2007)
4. Coppola, M., Di Lorenzo, E., Orlando, A., Sibillo, M.: Solvency analysis and demographic risk measures. J. Risk Finan. **12**(4), 252–269 (2011)

5. D'Amato, V., Di Lorenzo, E., Haberman, S., Russolillo, M., Sibillo, M.: The Poisson log-bilinear Lee Carter model: applications of efficient bootstrap methods to annuity analyses. North Am. Actuarial J. **15**(2), 315–333 (2011)
6. De Mey, J.: Reporting on the financial performance of life insurers. Geneva Pap. Risk Insur. **34**, 228–241 (2009)
7. Di Lorenzo, E., Orlando, A., Sibillo, M.: Measuring risk-adjusted performance and product attractiveness of a life annuity portfolio. J. Math. Finan. (special issue on Actuarial Science and Quantitative Finance) **7**(1), 83–101 (2017)
8. Easton, A.E., Harris J.F.: Actuarial Aspects of Individual and Annuity Contracts, 2nd edn. Actex Publication Inc., Winsted (2007)
9. Friedman, B.M., Warshawsky, M.: Annuity Prices and Saving Behavior in the United States. NBER Working Paper No. 1683 (Also Reprint No. r1025) (1985)
10. Hari, N., De Waegenaere, A., Melenbergb, B., Nijman, T.E.: Longevity risk in portfolios of pension annuities. Insur. Math. Econ. **42**, 505–519 (2008)
11. Kraus, C.: EVA/RAROC vs. MCEV Earnings: A Unification Approach. The Geneva Papers on Risk and Insurance - Issues and Practice, December 2012, pp. 113–136 (2012)
12. Lopes, P.: The Effects of Load Factors and Minimum Size Restrictions on Annuity Market Participation (2009). http://www.lse.ac.uk/finance/people/Teaching%20Fellows/pdfs%20and%20images/lopesAnnuities.pdf%20and%20images/lopesAnnuities.pdf
13. Swiss Re: Understanding Profitability in Life Insurance (2012). http://www.swissre.com/-media-/news_releases/nr_20120207_measuring_life

Uncertainty in Historical Value-at-Risk: An Alternative Quantile-Based Risk Measure

Dominique Guégan, Bertrand Hassani, and Kehan Li

Abstract The financial industry has extensively used quantile-based risk measures relying on the Value-at-Risk (*VaR*). They need to be estimated from relevant historical data sets. Consequently, they contain uncertainty due to the finiteness of observations in practice. We propose an alternative quantile-based risk measure (the Spectrum Stress *VaR*) to capture the uncertainty in the historical *VaR* approach. This one provides flexibility to the risk manager to implement prudential regulatory framework. It can be a *VaR* based stressed risk measure. In the end we propose a stress testing application for it.

1 Introduction

The financial industry has extensively used quantile-based risk measures based on the Value-at-Risk (*VaR*). In statistical terms, the *VaR* is a quantile reserve, often using the pth ($p \in [0, 1]$) percentile of the loss distribution. Typically, the *VaR* is not known with certainty and needs to be estimated from sample estimators of relevant observations. Bignozzi and Tsanakas [6] point out that the observations are often very small creating statistical errors, which means that the values of sample estimators can diverge substantially from the true values. Jorion [11] calls it the risk in Value-at-Risk itself. Pérignon and Smith [13] find that historical *VaR* is the most

D. Guégan
University Paris 1 Panthéon -Sorbonne, LabEx ReFi, **IPAG**, Paris, France

Department of Economics, Ca'Foscari University of Venezia, Venezia, Italy
e-mail: dguegan@univ-paris1.fr

K. Li (✉)
Université Paris 1 Panthéon-Sorbonne, CES UMR 8174, 106 bd l'Hopital 75013, Labex ReFi, Paris, France
e-mail: Kehan.Li@malix.univ-paris1.fr

B. Hassani
Grupo Santander and Université Paris 1 Panthéon-Sorbonne, CES UMR 8174, Labex ReFi, Paris, France
e-mail: bertrand.hassani@malix.univ-paris1.fr

M. Corazza et al. (eds.), *Mathematical and Statistical Methods for Actuarial Sciences and Finance*, https://doi.org/10.1007/978-3-319-50234-2_10

popular *VaR* method, as 73% of the banks report their *VaR* estimation methodologies using historical *VaR*.

Our paper proposes an alternative risk measure based on the historical *VaR*. A confidence interval (CI) is considered to integrate the uncertainty contained in the historical *VaR*. It is a tail risk measure at multiple confidence levels (Alexander et al. [2]). It provides the flexibility to the risk manager to implement a prudential regulatory framework (Basel Committee on Banking Supervision (BCBS) [3] and Acharya [1]). Additionally, it can be a *VaR* based stressed risk measure relying on a continuous 12-month period of significant financial stress following the requirement of the Basel Committee (BCBS [5]). We propose a stress testing application of this risk measure.

There are several pertinent articles which investigate the use of stressed *VaR*. For instance, Santos et al. [16] develop an approach to identify optimal portfolios with minimum regulatory capital based on a formula that involves both *VaR* and stressed *VaR*. Colletaz et al. [8] provides a method to validate estimates of various measures of downside risk including stressed *VaR*. Alexander et al. [2] find that minimum regulatory capital based on both *VaR* and stressed *VaR* is ineffective in preventing banks from taking substantive tail risk in their trading books without capital requirement penalties. However, it is important to point out that these stressed *VaR* implications do not integrate the model uncertainty. Furthermore, they only use single confidence level p when they build their stressed *VaR*. It leads to ignoring the information contained the tail.

Some papers have discussed the confidence interval of the *VaR*. For example, Pritsker [14] computes a nonparametric CI to evaluate the accuracy of different *VaR* approaches. Christoffersen and Gonçalves [7] assess the precision of *VaR* forecast by using bootstrap prediction intervals. Jorion [11] provides the asymptotic standard error and confidence bands for a sample quantile, assuming the loss distribution is known. All these approaches mainly use their CI (provided by asymptotic result or bootstrap) as a complementary tool to assess the quality of the *VaR*. In our work we consider another approach to building the CI (we do not assume that the loss distribution is known and we do not use simulation). We use an asymptotic result and a parametric approach. We focus on a fat-tailed distribution[1] to capture historical stress information, in order to build a stressed risk measure. Finally we use the lower (or upper) bound of CI directly as one boundary of our risk measure.

This paper is organised as follows. Section 2 describes our risk measure. Section 3 proposes a stress testing application for the risk measure. Section 4 concludes.

[1]A fat-tailed distribution has the property that exhibits large kurtosis or has power law decay in the tail of the distribution.

2 The Spectrum Stress *VaR* Measure

Consider a random variable (r.v.) X (for example the return of a portfolio, the return of a risk factor or an operational loss), with a cdf $F_{\boldsymbol{\theta}}$ ($f_{\boldsymbol{\theta}}$ is the associated probability density function (pdf) and $\boldsymbol{\theta}$ are the parameters). Let $X_1, \ldots, X_n$ be the information set of X with length n. We assume they are independent and identically distributed (i.i.d).[2] We sort them and obtain $X_{(1)} \leq \ldots \leq X_{(n)}$. Given $0 < p < 1$, we define the *hVaR* as $X_{(m)}$, where $m = np$ if np is an integer and $m = [np] + 1$ otherwise.[3] Rao [15] provides an asymptotic normality (AN) approximation for the distribution of $X_{(m)}$.

Theorem 1 (Asymptotic Normality Approximation (Rao [15])) *Assume $F_{\boldsymbol{\theta}}$ is continuous and differentiable and $f_{\boldsymbol{\theta}}$ is strictly positive at $F_{\boldsymbol{\theta}}^{-1}(p)$, then*

$$\sqrt{n}(X_{(m)} - F_{\boldsymbol{\theta}}^{-1}(p)) \rightarrow_{(d)} N(0, V), \quad as \quad n \longrightarrow \infty \tag{1}$$

where $\rightarrow_{(d)}$ means convergence in distribution, $V = \frac{p(1-p)}{f_{\boldsymbol{\theta}}(F_{\boldsymbol{\theta}}^{-1}(p))^2 n}$. $N(F_{\boldsymbol{\theta}}^{-1}(p), V)$ represents the Gaussian distribution with mean $F_{\boldsymbol{\theta}}^{-1}(p)$ and variance V.

Notice that expression (1) depends on the values of $F_{\boldsymbol{\theta}}$ and $f_{\boldsymbol{\theta}}$, which are unknown in most cases. Therefore density estimation is necessary. One possible way is to use the Siddiqui-Bloch-Gastwirth estimator, whose construction crucially depends on the choice of a smoothing parameter (see Hall and Sheather [10]). Instead of using such nonparametric estimator which suffers from difficulty of smoothing parameter choice, we fit a panel of distributions using $X_1, \ldots, X_n$ to compute the estimators of $\boldsymbol{\theta}$, denoted $\hat{\boldsymbol{\theta}}$. Then $F_{\hat{\boldsymbol{\theta}}}$ and $f_{\hat{\boldsymbol{\theta}}}$ are the estimators of $F_{\boldsymbol{\theta}}$ and $f_{\boldsymbol{\theta}}$. By plugging $F_{\hat{\boldsymbol{\theta}}}$ and $f_{\hat{\boldsymbol{\theta}}}$ in expression (1), we provide a corollary of Theorem 1.

Corollary 1 (Plug-in AN Approximation) *Assume $F_{\boldsymbol{\theta}}$ and $f_{\boldsymbol{\theta}}$ are continuous functions with respect to (w.r.t) $\boldsymbol{\theta}$, and $\hat{\boldsymbol{\theta}}$ is an asymptotically consistent estimator of $\boldsymbol{\theta}$.*[4] *Then we have*

$$\sqrt{n}(X_{(m)} - F_{\hat{\boldsymbol{\theta}}}^{-1}(p)) \rightarrow_{(d)} N(0, \widehat{V}), \quad as \quad n \longrightarrow \infty \tag{2}$$

where $\widehat{V} = \frac{p(1-p)}{[f_{\hat{\boldsymbol{\theta}}}(F_{\hat{\boldsymbol{\theta}}}^{-1}(p))]^2 n}$.

The proof is presented in Appendix. Given confidence level $0 < q < 1$,[5] we build a confidence interval $CI_{p,q}$ around $X_{(m)}$ from Corollary 1:

$$X_{(m)} \in \left[F_{\hat{\boldsymbol{\theta}}}^{-1}(p) - z_{\frac{1+q}{2}} \sqrt{\widehat{V}}, \quad F_{\hat{\boldsymbol{\theta}}}^{-1}(p) + z_{\frac{1+q}{2}} \sqrt{\widehat{V}} \right] \tag{3}$$

[2] Or if they are not, we assume that we can transform them to an i.i.d set by filtering.

[3] $[x]$ denotes the largest integer less than or equal to x.

[4] Asymptotically consistent estimator means $\hat{\boldsymbol{\theta}} \rightarrow_{(P)} \boldsymbol{\theta}$, where $\rightarrow_{(P)}$ represents convergence in probability.

[5] p is the confidence level of historical *VaR* and q is the confidence level of its confidence interval.

where

$$\widehat{V} = \frac{p(1-p)}{[f_{\hat{\theta}}(F_{\hat{\theta}}^{-1}(p))]^2 n} \quad (4)$$

and $z_{\frac{1+q}{2}}$ is the $\frac{1+q}{2}th$ quantile of standard Gaussian distribution. According to the expression (3), $CI_{p,q}$ depends on $n, f_{\hat{\theta}}, p$ and q.

In practice for a sequence $p_1 < p_2 < \ldots < p_k$, given $\{q_i\}_{i=1,\ldots,k}$, we compute the sequences $\{F_{\hat{\theta}}^{-1}(p_i)\}$ and CI_{p_i,q_i} for $i = 1, \ldots, k$. We define an area delineated by $F_{\hat{\theta}}^{-1}(p_i)$ and the lower (or upper) bound of CI_{p_i,q_i} for $i = 1, \ldots, k$. We call this area the Spectrum Stress *VaR* measure (SSVaR). Figure 1 provides a graph of the SSVaR. The lower (green) and upper (red) curves correspond to the boundaries of CI_{p_i,q_i} for $\{p_i\}$ and $\{q_i\}$, $i = 1, \ldots, k$. The black curve in the middle is associated with the sequence of $\{\{F_{\hat{\theta}}^{-1}(p_i)\}$ for $i = 1, \ldots, k$. The black shadow area is the SSVaR.

In Fig. 1, we observe that when the confidence level p increases, for example from 95% or 99% (traditional confidence levels proposed by the Basel committee) to 99.9%, the wideness of bounds of the SSVaR increases also. It means that the historical VaR estimates with extreme confidence levels may contain more uncertainty than those estimates with ordinary confidence levels. Thus, it is necessary to use the SSVaR to quantify potentially large violations of the extreme historical VaR

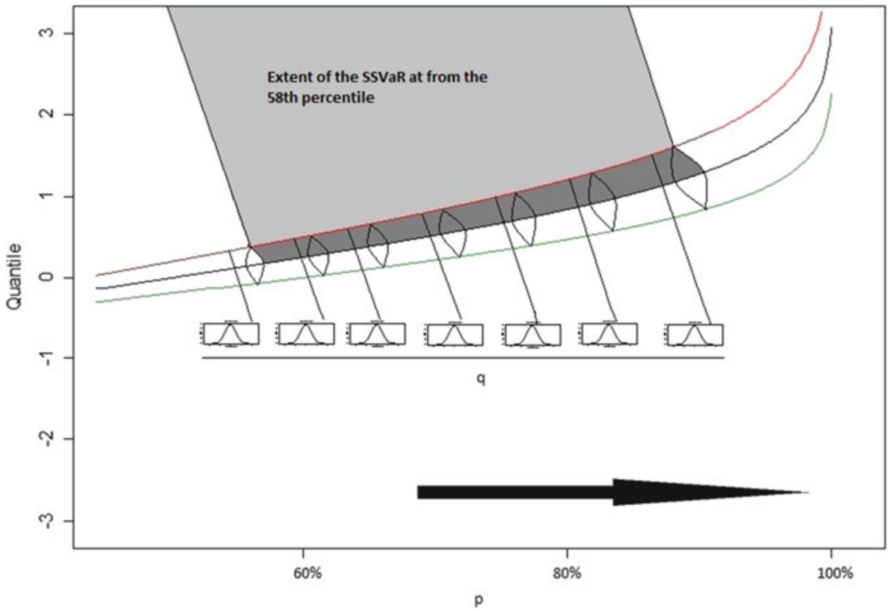

Fig. 1 The lower (green) and upper (red) curves correspond to the boundaries of CI_{p_i,q_i} for $\{p_i\}$ and $\{q_i\}$, $i = 1, \ldots, k$. The black curve in the middle is associated to the sequence of $\{\{F_{\hat{\theta}}^{-1}(p_i)\}$ for $i = 1, \ldots, k$. The black shadow area is the SSVaR. When the values of q_i change, SSVaR can shift to the grey area

estimates. These outliers are crucial because they concern the large moves in the financial market. It is also important to point out that when the risk manager has to work within the prudential regulatory framework, he can choose higher q_i so that the SSVaR can shift to the grey area. Also, he can shift the SSVaR to the grey area by choosing a fat-tailed $\hat{f}$. In fact a fat-tailed $\hat{f}$ can take more stress information from a period of significant financial turmoil than a thin tail fit. Consequently the SSVaR is a stressed risk measure in essence.

3 A Stress Testing Application of the SSVaR

During a recent crisis, some investors have suffered considerable losses due to extreme events. Consequently there has been a growing literature on stress testing. In particular, banks that use the *VaR* approach must have in place a rigorous stress testing program (BCBS [4]). In response, we propose a SSVaR measure applicable to stress testing. The result of the stress testing is also a criterion to choosing a reasonable $\hat{f}$ to build the SSVaR, which we can use first as an alert indicator.

To explain our purpose, we consider a fictive financial institution. This one holds a Chinese market portfolio (that is, the same stock components and weights as the Shanghai Stock Exchange Composite Index (SHCOMP)). We compute the SSVaR using the daily return of SHCOMP from 29/06/2007 to 20/06/2008 (it contains 246 points and we call it Ω_1). The historical *VaR* of Ω_1 are computed. For the stress testing, we compute the empirical quantiles on the daily return of SHCOMP from 01/12/2014 to 09/11/2015 (it contains 241 points and we call it Ω_2).[6] Table 1 provides the empirical statistics of the data sets. It shows that these two data sets are left skewed and leptokurtic (Kurtosis > 3). The distributions which characterise these two data sets need to have these properties. In the following we build SSVaR using Ω_1, with Gaussian distribution as a benchmark and Normal-inverse Gaussian distribution (NIG, Godin [9]). We provide the estimates of Gaussian and NIG distributions for Ω_1 in Table 2.

Table 1 Empirical statistics of SHCOMP daily returns from 29/06/2007 to 20/06/2008 (Ω_1) and from 01/12/2014 to 09/11/2015 (Ω_2)

	Mean	Variance	Skewness	Kurtosis
Ω_1 ($n = 246$)	−0.0017	0.0007	−0.3796	3.7876
Ω_2 ($n = 241$)	0.0010	0.0007	−1.0509	5.0698

Table 2 We provide the estimates of Gaussian and NIG distributions for Ω_1 in Table 2

	α	β	μ	δ
Gaussian	−0.0017	0.027		
NIG	90.63	−25.74	0.016	0.058

[6]The data sets are downloaded from Bloomberg.

To take into account the left tail market risk, we use $0.01 \le p_i \le 0.1$ and fixed $q = 0.95$. The stress testing exercise is following: we build the SSVaR for Ω_1 using Gaussian distribution and NIG distribution, in order to quantify the uncertainty of the historical *VaR*. We use the historical *VaR* estimates of Ω_1 and the empirical quantiles of Ω_2 to check the robustness of the SSVaR. The results are provided in Fig. 2: on the left graph the dashed (blue and green) lines are the upper and lower bounds of the SSVaR corresponding to the Gaussian distribution. On the right graph the dashed (blue and green) lines are the upper and lower bounds of the SSVaR corresponding to the NIG distribution. In these two graphs, the solid (red) lines are the historical *VaR* and the solid-dot (brown) lines are the empirical quantiles for Ω_2.

The left graph of Fig. 2 suggests that the SSVaR based on a Gaussian distribution underestimates the risk computed using Ω_1 and Ω_2, because the left part of the historical *VaR* and the empirical quantiles are outside the SSVaR. The right graph of Fig. 2 shows that the SSVaR built using a NIG distribution, enables the risk to be controlled more efficiently since they are almost within the SSVaR. Additionally, ignoring the uncertainty in the historical *VaR* (that is, using the empirical quantiles directly as the risk measure) underestimates the risk computed using Ω_2, because the left part of the empirical quantiles is lower than the historical *VaR*.

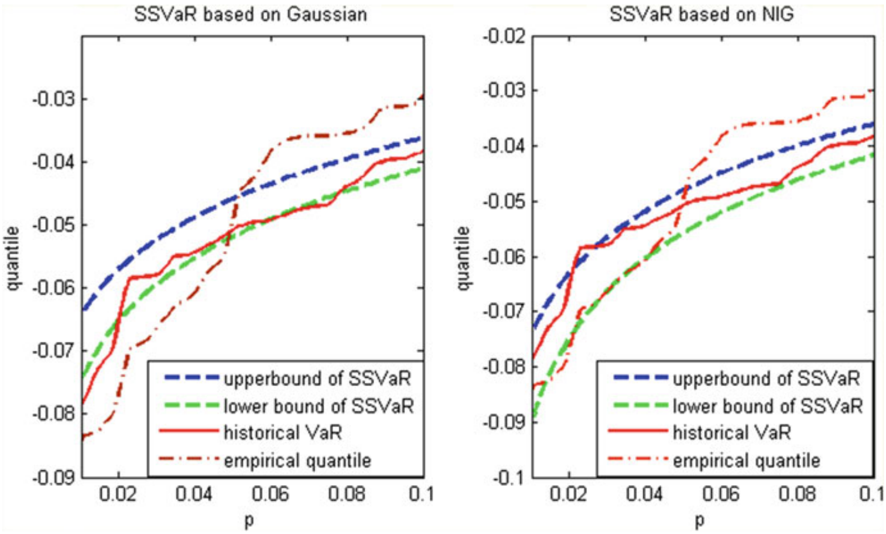

Fig. 2 We use $0.01 \le p_i \le 0.1$ and fixed $q = 0.95$ and build the SSVaR for Ω_1 using Gaussian distribution (mean -0.0017 and variance 0.0007) and NIG (with tail parameter equalling 90.63, skewness parameter equalling -25.73, location parameter equalling 0.0155 and scale parameter equalling 0.058). In this figure, on the left graph the dashed (blue and green) lines are the upper and lower bounds of the SSVaR corresponding to the Gaussian distribution. On the right graph the dashed (blue and green) lines are the upper and lower bounds of the SSVaR corresponding to the NIG distribution. In these two graphs, the solid (red) lines are the historical *VaR* and the solid-dot (brown) lines are the empirical quantiles for Ω_2

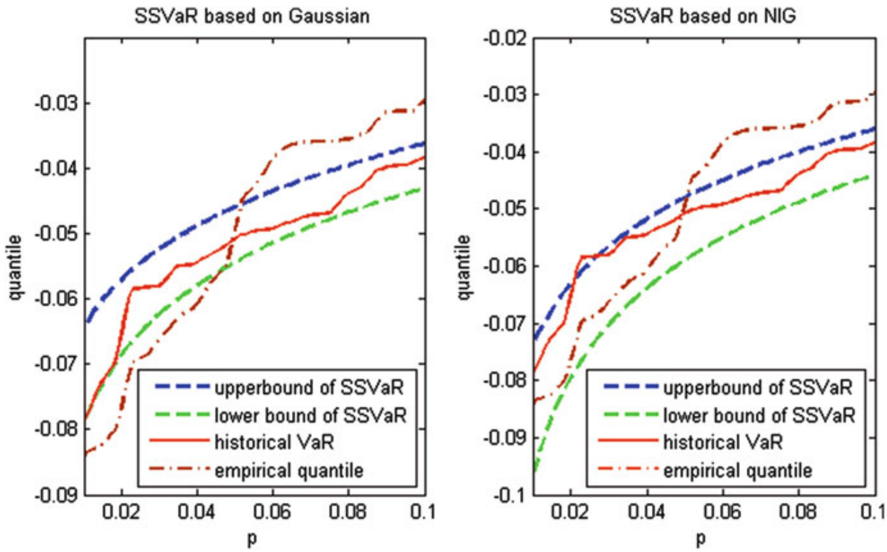

Fig. 3 We use $0.01 \leq p_i \leq 0.1$ and fixed $q = 0.99$ and build the SSVaR for Ω_1 using Gaussian distribution (mean -0.0017 and variance 0.0007) and NIG (with tail parameter equalling 90.63, skewness parameter equalling -25.73, location parameter equalling 0.0155 and scale parameter equalling 0.058). In Fig. 2, on the left graph the dashed (blue and green) lines are the upper and lower bounds of the SSVaR corresponding to the Gaussian distribution. On the right graph the dashed (blue and green) lines are the upper and lower bounds of the SSVaR corresponding to the NIG distribution. In these two graphs, the solid (red) lines are the historical *VaR* and the solid-dot (brown) lines are the empirical quantiles for Ω_2

For robustness, we change q from 0.95 to 0.99, and perform the same stress testing exercise in Fig. 2 again. The results are provided in Fig. 3: we observe that in the left graph, for $0.05 \leq p \leq 0.01$, the empirical quantiles of Ω_2 are below the lower bound of the SSVaR. It means that the SSVaR based on the Gaussian fit still underestimates the risks, even if we increase the q to be prudential. However, in the right graph, the empirical quantiles of Ω_2 are totally above the lower bound of the SSVaR. That means for these data, 0.99 may be an appropriate value of q to obtain a robust SSVaR. In practice, the SSVaR is an improvement risk measure of historical *VaR*. The risk manager can use it directly to allocate capital reserve to the risk of measurement uncertainty. Additionally, it can be a stressed and tail risk measure providing flexibility to the risk manager to work within the prudential regulatory framework.

4 Conclusion

In this article, we propose an alternative quantile-based risk measure SSVaR, to integrate the uncertainty from the historical *VaR*. Additionally, it is a tail risk measure. In addition it provides flexibility to the risk manager to implement the prudential regulatory framework. It can be a *VaR* based stressed risk measure.

Additionally, we propose a stress testing application for the SSVaR, by illustrating the magnitude of the exceptions based on the empirical quantile of two data sets from SHCOMP. The results suggest that ignoring the uncertainty in the historical *VaR* underestimates risks. Also, we observe that when the data sets are skewed and leptokurtic, risk manager needs to fit a skewed and leptokurtic distribution to build SSVaR. It helps control the risk efficiently.

As the purpose of a forthcoming paper, some improvements to this approach could be done. Indeed, the expression (1) relies on the assumption of independence for $X_1, \ldots, X_n$ (Rao [15]). Nevertheless, we can extend the results in the case of $\alpha-mixing$ (Leadbetter et al. [12]) data sets. Also, the SSVaR can be used directly for the operational risks which are mainly independent. For other risks we can calibrate dynamics on $X_1, \ldots, X_n$, like $X_t = f(X_{t-1}) + \epsilon_t$ where ϵ_t is a white noise. Then we build the SSVaR using the residuals $\{\epsilon_t\}$, and the time series modelling can be used to introduce dynamics within the SSVaR.

Appendix

In appendix, first we provide the proof of Corollary 1. Second, we present some well-known facts about the NIG distribution.

Proof of Corollary 1

At first, we introduce the Slutsky's theorem

Theorem 2 (Slutsky's Theorem) *Let $\{X_n\}$, $\{Y_n\}$ be sequences of r.v., If $\{X_n\}$ converges in distribution ($\rightarrow_{(d)}$) to a r.v. X and $\{Y_n\}$ converges in probability ($\rightarrow_{(p)}$) to a constant c, then*

$$X_n Y_n \rightarrow_{(d)} cX \tag{5}$$

Proof To prove Corollary 1, we begin with

$$\begin{aligned}\frac{X_{(m)} - F_{\hat{\boldsymbol{\theta}}}^{-1}(p)}{\sqrt{\widehat{V}}} &= \frac{X_{(m)} - F_{\boldsymbol{\theta}}^{-1}(p)}{\sqrt{\widehat{V}}} + \frac{F_{\boldsymbol{\theta}}^{-1}(p) - F_{\hat{\boldsymbol{\theta}}}^{-1}(p)}{\sqrt{\widehat{V}}} \\ &= \sqrt{\frac{V}{\widehat{V}}}\frac{X_{(m)} - F_{\boldsymbol{\theta}}^{-1}(p)}{\sqrt{V}} + \frac{F_{\boldsymbol{\theta}}^{-1}(p) - F_{\hat{\boldsymbol{\theta}}}^{-1}(p)}{\sqrt{\widehat{V}}} \\ &= \frac{f_{\hat{\boldsymbol{\theta}}}(F_{\hat{\boldsymbol{\theta}}}^{-1}(p))}{f_{\boldsymbol{\theta}}(F_{\boldsymbol{\theta}}^{-1}(p)}\frac{X_{(m)} - F_{\boldsymbol{\theta}}^{-1}(p))}{\sqrt{V}} + \frac{F_{\boldsymbol{\theta}}^{-1}(p) - F_{\hat{\boldsymbol{\theta}}}^{-1}(p)}{\sqrt{\widehat{V}}}\end{aligned} \tag{6}$$

Since convergence in probability is preserved under continuous transformations, from $\hat{\boldsymbol{\theta}} \rightarrow_{(P)} \boldsymbol{\theta}$ we have

$$F_{\boldsymbol{\theta}}^{-1}(p) - F_{\hat{\boldsymbol{\theta}}}^{-1}(p) \rightarrow_{(P)} 0 \tag{7}$$

$$f_{\hat{\boldsymbol{\theta}}}(F_{\hat{\boldsymbol{\theta}}}^{-1})(p) - f_{\boldsymbol{\theta}}(F_{\boldsymbol{\theta}}^{-1}(p)) \rightarrow_{(P)} 0 \tag{8}$$

From Theorem 1 we know that $\frac{X_{(m)} - F_{\boldsymbol{\theta}}^{-1}(p))}{\sqrt{V}} \rightarrow_{(d)} N(0, 1)$, then

$$\frac{X_{(m)} - F_{\hat{\boldsymbol{\theta}}}^{-1}(p)}{\sqrt{\widehat{V}}} \rightarrow_{(d)} N(0, 1) \tag{9}$$

NIG Distribution

The Normal Inverse Gaussian (NIG) distribution is a member of the wider class of generalized hyperbolic distributions. It is well-known fact that the returns of most financial portfolios have fat tails and the actual kurtosis is higher than 3. The NIG distribution can be fat-tailed and asymmetrical. Therefore, it allows to take into account the fat-tailed and asymmetrical properties of the financial data sets. The density of the NIG distribution is given by:

$$nig(x; \alpha, \beta, \delta, \mu) = \frac{\alpha\delta}{\pi} exp(\delta\sqrt{\alpha^2 - \beta^2} + \beta(x - \mu)) \frac{K_1(\alpha\sqrt{\delta^2 + (x - \mu)^2})}{\sqrt{\delta^2 + (x - \mu)^2}} \tag{10}$$

where $x, \mu \in \Re$, $0 \leq \delta$ and $0 \leq |\ \beta\ | \leq \alpha$. α is the tail parameter; β is the asymmetry parameter; μ is the location parameter; δ is the scale parameter.

Acknowledgements This work was achieved through the Laboratory of Excellence on Financial Regulation (Labex ReFi) supported by PRES heSam under the reference ANR10LABX0095. It benefited from a French government support managed by the National Research Agency (ANR) within the project Investissements d'Avenir Paris Nouveaux Mondes (investments for the future Paris New Worlds) under the reference ANR11IDEX000602.

References

1. Acharya, V.V.: A theory of systemic risk and design of prudential bank regulation. J. Financ. Stab. **5**(3), 224–255 (2009)
2. Alexander, G.J., Baptista, A.M., Yan S.: On regulatory responses to the recent crisis: an assessment of the basel market risk framework and the Volcker rule. Finan. Markets Inst. Instrum. **24**(2–3), 87–125 (2015)
3. Basel Committee on Banking Supervision: A New Capital Adequacy Framework. Technical Report, Bank for International Settlements, Basel, Switzerland (1999)
4. Basel Committee on Banking Supervision: Amendment to the Capital Accord to incorporate market risks. Working Paper, Bank for International Settlements, Basel, Switzerland (2005)
5. Basel Committee on Banking Supervision: Basel III: A global regulatory framework for more resilient banks and banking systems. Working Paper, Bank for International Settlements, Basel, Switzerland (2011)
6. Bignozzi, V., Tsanakas A.: Parameter uncertainty and residual estimation risk. J. Risk Insur. **83**(4), 949–978 (2016). https://doi.org/doi:10.1111/jori.12075
7. Christoersen, P., Goncalves, S.: Estimation risk in financial risk management. J. Risk **7**(3), 1–28 (2005)
8. Colletaz, G., Hurlin C., Pérignon, C.: The risk map: a new tool for validating risk models. J. Bank. Financ. **37**(10), 3843–3854 (2013)
9. Godin, F., Mayoral, S., Morales, M.: Contingent claim pricing using a normal inverse gaussian probability distortion operator. J. Risk Insur. **79**(3), 841–866 (2012)
10. Hall, P., Sheather, S.J.: On the distribution of a studentized quantile. J. R. Stat. Soc. Ser. B **50**(3), 381–391 (1988)
11. Jorion, P.: Risk2: measuring the risk in value at risk. Financ. Anal. J. **52**(6), 47–56 (1996)
12. Leadbetter, M.R., Lindgren, G., Rootzen, H.: Extremes and Related Properties of Random Sequences and Processes. Springer, New York (1983)
13. Pérignon, C., Smith, D.R.: The level and quality of value-at-risk disclosure by commercial banks. J. Bank. Financ. **34**(2), 362–377 (2010)
14. Pritsker, M.: Evaluating value at risk methodologies: accuracy versus computational time. J. Financ. Serv. Res. **12**(2–3), 201–242 (1997)
15. Rao, C.R.: Linear Statistical Inference and Its Applications. Wiley, Hoboken (2002)
16. Santos, A.A.P., Nogales, F.J., Dijk, D.V.: Optimal portfolios with minimum capital requirements. J. Bank. Financ. **36**(7), 1928–1942 (2012)

Modeling Variance Risk Premium

Kossi Gnameho, Juho Kanniainen, and Ye Yue

Abstract The bias between the expected realized variance under the historical measure and the risk neutral probability introduces the concept of the variance risk premium (VRP). Our work introduced a probabilistic modeling of the VRP via a parametric class of stochastic volatility models which incorporates the nonlinear class.

1 Introduction

Modeling the variance or volatility of asset prices has been a paramount topic for asset pricing and risk management. The family of Generalized Autoregressive Conditional Heteroskedasticity (GARCH, for short) times series models (see e.g. Engle [8], Nelson [12], Chorro et al. [4]) are popular to estimate and forecast the volatility or variance. Additionally, derivative products written on realized variance have become increasingly popular (e.g. variance swap, VIX options). As the volatility is latent, investors are interested to seek a proper model for asset variance in both physical and risk neutral worlds. Furthermore, the premium of variance is crucial for investors to make trading decisions and manage the variance risk. Many empirical studies have demonstrated that the cornerstone premise for studying VRP is that return variance is stochastic. This paper develops a probabilistic modeling of the VRP in a particular class of parametric stochastic volatility models. The general definition of the variance risk premium at the time instance t with an horizon T is given by $VRP_t^T = \mathbb{E}(RV_{t,T} \big| \mathscr{F}_t) - \mathbb{E}^*(RV_{t,T} \big| \mathscr{F}_t)$, $0 \leq t \leq T$. The quantity $RV_{t,T}$ represents the realized variance over the period $[t, T]$ of a determined risky asset S^x. The existent literature provides strong evidence that many popular affine variance models are empirically outperformed by non-affine models. Our framework deals

K. Gnameho (✉)
Department of Quantitative Economics, Maastricht University, Maastricht, The Netherlands
e-mail: kossi.gnameho@gmail.com

J. Kanniainen • Y. Yue
Laboratory of Industrial and Engineering Management, Tampere University of Technology, Tampere, Finland
e-mail: juho.kanniainen@tut.fi; ye.yue@tut.fi

M. Corazza et al. (eds.), *Mathematical and Statistical Methods for Actuarial Sciences and Finance*, https://doi.org/10.1007/978-3-319-50234-2_11

with this important class of non-affine continuous time diffusions of the spot-variance process. We give a general backward stochastic representation of the VRP via some basis notion of Malliavin Calculus and provide two applications.

1.1 Notations and Assumptions

All activity occurs on a filtered probability space $(\Omega, \mathscr{F}, \mathbb{P}, \mathbb{F})$ where $\mathscr{F} = \mathscr{F}_T$. The filtration $\mathbb{F} = (\mathscr{F}_t)_{0 \le t \le T}$ denotes a complete natural filtration generated by a one-dimensional Brownian motion W and T a finite horizon.

- All the equalities and the inequalities between random variables are understood in almost sure sense unless explicitly stated otherwise.
- $\mathbb{P}$ is the objective or the historical probability measure and $\mathbb{P}^*$ denotes the risk neutral measure according to the literature of pricing derivatives.
- The operator $\mathbb{E}_t[.] := \mathbb{E}[.\big|\mathscr{F}_t]$ denotes the time-t conditional expectation and D denotes the "Malliavin derivative" operator (cf. Ocone [14]) associated with the measure $\mathbb{P}$ with respect to the one-dimensional Brownian motion W. The same definition is applied to the conditional operator $\mathbb{E}^*$ and the Malliavin derivative operator D^* associated with the equivalent measure $\mathbb{P}^*$ with respect to another one-dimensional Brownian motion W^*. The constant r denotes the daily interest rate. The risk-free rate is the minimum an investor will require if the investment doesn't carry any risk.
- Unless explicitly stated otherwise, (S^x, V) denotes a couple of continuous stochastic processes. S^x describes the dynamic of an risky underlying asset and V its instantaneous variance. The real value x denotes the initial value of S^x.

2 Estimation of the Variance Risk Premium

Variance Swap is a forward contract on realized variance. It is commonly used by companies or investors to trade a future realized variance of a given underlying asset or an index return. With the maturity time T, the payoff of the variance swap contract over the period $[t, T]$ is given by $(V_{t,T} - K)N$, where:

- N denotes the notional amount associated to the variance swap contract,
- K is a constant parameter called the fixed leg of the contract,
- $V_{t,T}$ is the realized annualized variance of the underlying asset over $[t, T]$.

The realized annualized variance of the underlying asset can be evaluated in discrete or continuous time framework. In the case of discrete time evaluation, we build an uniform partition π of the interval $[t, T]$ defined as follows: $t < t_1 < \ldots < t_m = T$,

$\Delta := t_{i+1} - t_i$ with $i \in \{0, 1, .., m-1\}$. The common definition of the discrete realized annualized variance $V^d_{t,T}$ over the period $[t, T]$, is given by

$$V^d_{t,T} = \frac{C_t}{m-1} \sum_{i=0}^{m-1} \left(\ln \left(\frac{S^x_{t_{i+1}}}{S^x_{t_i}} \right) \right)^2 ,$$

where C_t is the annualized factor, $S^x_{t_i}$ is the ith observation of the price of S^x. The continuous version of the realized annualized variance is considered as the limit in probability of the discrete version. From the fact the net present value at the initialization of the variance swap contract is zero, the fair conditional variance strike $K_{t,T} := \mathbb{E}^*(V_{t,T} \big| \mathscr{F}_t)$. For the sake of clarity, we consider a continuous -path diffusion of the underlying asset S^x where its realized variance over the period $[0, T]$ is given by the annualized integrated variance. Following Demeterfi et al. [5] or Carr et al. [3], the fair variance strike of the variance swap contract is the value which makes the contract net present value equal to zero. This value is equal to:

$$K_{t,T} = \frac{2}{T-t} \int_0^\infty \frac{O_t(\kappa, T)}{\kappa^2 B^{1.\$}_{t,T}} d\kappa, \qquad \text{where} \tag{1}$$

- $B^{1.\$}_{t,T}$ is the time-t price of a Bond paying one dollar at the maturity T,
- $O_t(\kappa, T)$ is the time-t price of OTM (out of the money) vanilla options prices with characteristics (κ, T). The parameters k is the associated strike and T is the maturity date of the OTM contracts.

By decomposing the term $O_t(\kappa, T)$ in (2), the expression of $K_{t,T}$ becomes

$$K_{t,T} = \frac{2}{T-t} \int_0^{F_{t,T}} \frac{1}{k^2} P_t(k, T) dk + \int_{F_{t,T}}^\infty \frac{1}{k^2} C_t(k, T) dk, \qquad \text{where} \tag{2}$$

- $F_{t,T}$ is the future value contracted at t with maturity T of the underlying asset,
- $P_t(k, T)$ is the price at t of an OTM European put option and $C_t(k, T)$ is the time-t price of an OTM European call option with characteristics (k, T) as above.

The natural estimator $\hat{K}_{t,T}$ of the fair variance strike $K_{t,T}$ can be defined as

$$\hat{K}_{t,T} = \frac{2}{T-t} \sum_{k_i \in \mathbb{K}_P} \frac{1}{k_i^2} P_t(k_i, T) \Delta k + \sum_{k_i \in \mathbb{K}_C} \frac{1}{k_i^2} C_t(k_i, T) \Delta k, \qquad \text{where} \tag{3}$$

- $\mathbb{K}_C$ is the strip of strikes associated with the European OTM call options where the strikes varies from $F_{t,T}$ to infinity with a fixed maturity date T,
- $\mathbb{K}_P$ is the strip of strikes associated with the European OTM put options where the strikes varies from zero $F_{t,T}$ with a fixed maturity date T,
- Δk is a constant mesh associated to the discretization of the interval $[0, \infty[$.

Without loss of generality in the above decompositions, we have assumed that the risk free rate is zero. From the relation (2), we can approach the price of a variance swap contract with standard European contracts where the strip of strikes varies from zero to infinity. Due to the lack of data on some markets, the application of the formula (2) could introduce a large bias in the estimation of the variance swap rate according to the Euler approximation above.

3 Modeling of the Variance Risk Premium

We provide two important results which are one of the cornerstones of our probabilistic modeling. The basic theory of risk management in pricing models assumes in general a martingale condition under the risk neutral measure. In this section, we will recall the martingale representation theorem and show further its connection with the VRP in continuous time framework via the Clark-Ocone formula.

Theorem 1 *We consider the previous filtered probability space* $(\Omega, \mathscr{F}, \mathbb{P}, \mathbb{F})$. *For every square integrable martingale* $(M_t)_{0\leq t\leq T}$ *in the Brownian filtration* $\mathbb{F}$, *there exists a progressively measurable process* $H \in \mathscr{H}^2(\mathbb{R})$ *such that*

$$M_t = M_0 + \int_0^t H_s dW_s.$$

Corollary 1 *Considering the same filtered probability space, we can represent every square integrable martingale* $(M_t)_{0\leq t\leq T}$ *in the Brownian filtration by*

$$M_t = M_0 + \int_0^t \mathbb{E}(D_s M_t \big| \mathscr{F}_s) dW_s. \tag{4}$$

The proof of Corollary 1 can be found in Nunno et al. [6]. The representation (4) is a useful tool in mathematical finance. For instance it can be used to find hedging portfolios strategy and to compute some trading parameters. The previous two results are very important for the rest of our work regarding the backward representation of the VRP. Under the historical probability $\mathbb{P}$, we assume that the variance process of the underlying asset S^x is described by an Itô process. We specify the dynamic of the instantaneous variance process $(V_t)_{0\leq t\leq T}$ as

$$\begin{cases} dV_t = \beta(V_t)dt + \sigma(V_t)dW_t, & 0 < t \leq T, \\ V_0 = v_0 > 0, \quad \text{where} \quad \beta(x) = b(x) - ax \quad \text{and} \quad a \neq 0. \end{cases} \tag{5}$$

In derivatives market, from Harrison and Kreps [9], the existence of an equivalent martingale measure $\mathbb{P}^*$ guarantees the absence of arbitrage among a broad class of admissible trading strategy. In our work, we assume the existence of a unique risk neutral probability $\mathbb{P}^*$ under which the risk neutral dynamics of V is assumed to be driven by another standard Brownian motion W^*;

$$\begin{cases} dV_t = \beta^*(V_t)dt + \sigma^*(V_t)dW_t^*; \qquad 0 < t \leq T, \\ V_0 = v_0 > 0, \end{cases} \tag{6}$$

where $\beta^*(x) = b^*(x) - a^*x$ and $a^* \neq 0$. We assume that the solutions of (5) and (6) exist and are unique. We assume the following general assumptions:

$$(H) \begin{cases} (H1): \ b, b^*\sigma, \sigma^* \text{ are continuously differentiable almost elsewhere and Lipschitz.} \\ (H2): \ \text{the variance process } V \text{ admits a unique Malliavin derivative.} \\ (H3): \ \text{there exists } k_\sigma, K_\sigma > 0 \text{ such that } \ k_\sigma \leq \sigma^2(x) + [\sigma^*(x)]^2 \leq K_\sigma. \end{cases}$$

The assumption $(H1)$ can be relaxed in some particular case of stochastic volatility model. The hypothesis $(H3)$ ensures the non-degeneracy property of the volatility. For a purpose of clarity, the dynamic of the underlying asset S is constrained in such way that, its realized variance $RV_{t,T}$ over the period $[0, T]$, is given by the annualized integrated variance $RV_{t,T} = \dfrac{1}{T-t}\displaystyle\int_t^T V_u du$.

3.1 Backward Representation

As highlighted in the introduction, the probabilistic estimation of the VRP is useful due to the problem of lack of data often occurred in many markets. The next section is devoted to provide a backward representation variance risk premium under some regularity conditions of the coefficients of variance process.

3.1.1 Model-Free Modeling

Before describing our probabilistic modeling, we provide below a linear representation of the variance risk premium where the realized variance $RV_{t,T}$ is given by the annualized integrated variance of S^x. We remark that this representation of variance risk premium is linear in the drift term and similar to the representation of the Brownian bridge in the literature.

Lemma 1 *We have* $dVRP_t^T = \dfrac{1}{T-t}VRP_t^T dt - Z_t dB_t, \quad 0 \leq t < T$ *and* $VRP_T^T = 0$, *where* $Z_t = \frac{1}{T-t}\begin{pmatrix} \gamma_t \\ \gamma_t^* \end{pmatrix}$ *and* $B_t = \begin{pmatrix} W_t \\ W_t^* \end{pmatrix}$. *The processes* γ *and* γ^* *are defined by*

$$\gamma_t = -\mathbb{E}_t\big[D_t\mathbb{E}_t(\int_0^T V_s ds)\big] \quad \text{and} \quad \gamma_t^* = \mathbb{E}_t^*\big[D_t^*\mathbb{E}_t^*(\int_0^T V_s ds)\big].$$

Proof We first prove that the terminal condition of the variance risk premium is zero and show in the second step the backward representation. The limit

$$\lim_{h \mapsto 0} VRP_t^{t+h} = \lim_{h \mapsto 0} \frac{1}{h} \left(\mathbb{E} \left(\int_t^{t+h} V_r dr \big| \mathscr{F}_t \right) - \mathbb{E}^* \left(\int_t^{t+h} V_r dr \big| \mathscr{F}_t \right) \right) = 0.$$

The backward representation is a direct consequence of Corollary 1. By noticing that $\int_0^T V_u du - \int_0^t V_u du = \int_t^T V_u du$, we obtain the following representation

$$(T-t) VRP_t^T = \mathbb{E} \left(\int_0^T V_u du \big| \mathscr{F}_t \right) - \mathbb{E}^* \left(\int_0^T V_u du \big| \mathscr{F}_t \right). \tag{7}$$

The process M defined by $M_t = \mathbb{E} \left(\int_0^T V_u du \big| \mathscr{F}_t \right)$, $0 \leq t \leq T$, is a martingale under $\mathbb{P}$. Correspondingly, $M_t^* = \mathbb{E}^* \left(\int_0^T V_u du \big| \mathscr{F}_t \right)$ is also a martingale under the measure $\mathbb{P}^*$. The result of the lemma is then a direct consequence of the Corollary 1.

□

3.1.2 Forward Modeling

Considering the dynamics (5) and (6), we provide below a backward representation variance risk premium dynamic in our class of stochastic volatility models.

Proposition 1 *Under the assumptions* (*H*)*, the variance risk premium* (VRP^T) *admits the following stochastic backward representation*

$$\begin{cases} VRP_t^T = \dfrac{1}{aa^*(T-t)} \left[\displaystyle\int_t^T a^* \mathbb{E}_s \Big(D_s \big[\mathbb{E}_t (V_T) \big] \Big) dW_s - \int_t^T a \mathbb{E}_s^* \Big(D_s^* \big[\mathbb{E}_t^* (V_T) \big] \Big) dW_s^* \right] \\ \qquad\qquad + \dfrac{1}{a^*(T-t)} \displaystyle\int_t^T \sigma^*(V_s) dW_s^* - \dfrac{1}{a(T-t)} \int_t^T \sigma(V_s) dW_s - B_{T,t} + B_{T,t}^*, \\ VRP_T^T = 0, \quad \textit{and} \quad t \in [0, T), \quad \textit{where} \end{cases}$$

$$a(T-t) B_{T,t} = \int_t^T \mathbb{E}_s \left(\int_s^T D_s b(V_u) du \right) dW_s, \quad B_{T,T} = 0,$$
$$a^*(T-t) B_{T,t}^* = \int_t^T \mathbb{E}_s^* \left(\int_s^T D_s^* b^*(V_u) du \right) dW_s^*, \quad B_{T,T}^* = 0.$$

Proof By symmetry, we will detail the proof only for the objective term under the probability measure $\mathbb{P}$. We decompose the variance risk premium as follows

$$(T-t) VRP_t^T = \mathbb{E} \left(\int_0^T V_u du \big| \mathscr{F}_t \right) - \mathbb{E}^* \left(\int_0^T V_u du \big| \mathscr{F}_t \right). \tag{8}$$

From Eq. (5), we know that $V_T - v_0 = \int_0^T b(V_t)dt - a\int_0^T V_t dt + \int_0^T \sigma(V_t)dW_t$. By taking the conditional expectation with respect to $\mathscr{F}_t$

$$a\mathbb{E}\left(\int_0^T V_s ds \big| \mathscr{F}_t\right) = v_0 + \mathbb{E}\left(\int_0^T b(V_u)du \big| \mathscr{F}_t\right) - \mathbb{E}\left(V_T \big| \mathscr{F}_t\right) + \mathbb{E}\left(\int_0^T \sigma(V_u)dW_u \big| \mathscr{F}_t\right).$$

By the rule of Chasles applied to the Lebesgue and the stochastic integral,

$$\begin{aligned} a\mathbb{E}\left(\int_0^T V_s ds \big| \mathscr{F}_t\right) &= v_0 + \int_0^t b(V_u)du + \mathbb{E}\left(\int_t^T b(V_u)du \big| \mathscr{F}_t\right) \\ &\quad -\mathbb{E}\left(V_T \big| \mathscr{F}_t\right) + \int_0^t \sigma(V_u)dW_u. \end{aligned}$$

Taking the Malliavin derivative of both side of the above equality and applying the chain rule result (see Nualart [13]) to the stochastic integral term, we obtain

$$\begin{aligned} aD_s\mathbb{E}\left(\int_0^T V_s ds \big| \mathscr{F}_t\right) &= \int_s^t D_s b(V_u)du + D_s\mathbb{E}\left(\int_t^T b(V_u)du \big| \mathscr{F}_t\right) \\ &\quad - D_s\left(\mathbb{E}\left(V_T \big| \mathscr{F}_t\right)\right) + \sigma(V_s) + \int_s^t \sigma'(V_u)D_s V_u dW_u. \end{aligned} \tag{9}$$

Taking the conditional expectation with respect to $\mathscr{F}_s$ and Fubini's theorem,

$$a\mathbb{E}\Big(D_s M_t \big| \mathscr{F}_s\Big) = \mathbb{E}\left(\int_s^T D_s b(V_u)du \big| \mathscr{F}_s\right) - \mathbb{E}[D_s\mathbb{E}\left(V_T \big| \mathscr{F}_t\right) \big| \mathscr{F}_s] + \sigma(V_s)$$

From Corollary 1, we can finally represent the martingale $(M_t)_{0\le t\le T}$ as

$$M_t = \int_0^T V_t dt - \frac{1}{a}\int_t^T \left[\mathbb{E}_s\left(\int_s^T D_s b(V_u)du\right) - \mathbb{E}_s[D_s\mathbb{E}\left(V_T \big| \mathscr{F}_t\right)] + \sigma(V_s)\right] dW_s.$$

Correspondingly, we obtain the similar representation in the filtration generated by the Brownian motion $(W_t^*)_{0\le t\le T}$ for the process $(M_t^*)_{0\le t\le T}$ under the measure $\mathbb{P}^*$.

$$M_t^* = \int_0^T V_t dt - \frac{1}{a^*}\int_t^T \left[\mathbb{E}_s^*\left(\int_s^T D_s^* b^*(V_u)du\right) - \mathbb{E}_s^*[D_s^*\mathbb{E}\left(V_T \big| \mathscr{F}_t\right)] + \sigma^*(V_s)\right] W_s^*.$$

Plugging the previous two equations in the equality (7), the result follows.
In some cases, the expression of the variance risk premium can be simplified by evaluating the Malliavin derivative arising in the backward representation. For instance, in the affine case of stochastic volatility model, the backward representation of the variance risk premium representation becomes very simple and easy to compute.

Corollary 2 (Affine Case) *We assume that $\beta(x) = \mu - ax$ and $\beta^*(x) = \mu^* - a^*x$, the variance risk premium (VRP^T) admits the following backward representation. For $t \in [0, T]$,*

$$\begin{cases} VRP_t^T = \dfrac{1}{a(T-t)} \left(\displaystyle\int_t^T e^{-a(T-t)} \mathbb{E}_s\left(D_s V_t\right) dW_s - \int_t^T \sigma(V_s) dW_s \right) \\ \qquad\qquad - \dfrac{1}{a^*(T-t)} \left(\displaystyle\int_t^T e^{-a^*(T-t)} \mathbb{E}_s^*\left(D_s^* V_t\right) dW_s^* - \int_t^T \sigma_s^*(V_s) dW_s^* \right) \\ VRP_T^T = 0 \end{cases} \tag{10}$$

Proof The proof is a direct consequence of Proposition 1. We first remark that $(B_{T,t})_{0 \le t \le T}$ and $(B^*_{T,t})_{0 \le t \le T}$ defined in Proposition 1 are identically zero. It is known that under the probability measure $\mathbb{P}$, $V_T - V_0 = \mu T - a \int_0^T V_t dt + \int_0^T \sigma(V_t) dW_t$. Conditionally to $\mathscr{F}_s$, the deterministic function $t \mapsto \mathbb{E}\left(V_t \middle| \mathscr{F}_s\right)$ follows

$$\begin{cases} \dfrac{d}{dt}\mathbb{E}\left(V_t \middle| \mathscr{F}_s\right) = \mu - a\mathbb{E}\left(V_t \middle| \mathscr{F}_s\right)), \\ V_s \quad \textit{the initial condition.} \end{cases}$$

Solving the above linear ordinary differential equation, we obtain $\mathbb{E}\left(V_T \middle| \mathscr{F}_t\right) = \frac{\mu}{a} + (V_t - \frac{\mu}{a}) \exp(-a(T-t))$. Taking the Malliavin derivative of both side of the above equality, $D_s\mathbb{E}\left(V_T \middle| \mathscr{F}_t\right) = \exp(-a(T-t))D_s V_t$. We obtain a similar equality under the risk neutral probability $\mathbb{P}^*$, $D_s^*\mathbb{E}^*\left(V_T \middle| \mathscr{F}_t\right) = \exp(-a^*(T-t))D_s^* V_t$. The last two results and Proposition 1 conclude.

4 Applications: Affine Case and Non-affine Case

Often used in pricing model, several stochastic volatility models are based on an affine structure. In this section, we will apply the results of Sect. 3.1 to two cases of stochastic volatility models: an affine and a non-affine case. Heston in [10] introduces a two factor stochastic volatility model where the first component describes the fluctuation of the underlying asset S^x and the second component describes its instantaneous variance process V. The variance process V follows a Cox-Ingersoll-Ross (CIR) process. Under the objective probability $\mathbb{P}$,

$$S_t^x = x + \mu \int_0^t S_s^x ds + \int_0^t \sqrt{V_s} S_s^x dW_s^1, \qquad 0 \le t \le T, \tag{11}$$

$$V_t = v_0 + \int_0^t \kappa(\theta - V_s) ds + \int_0^t \sigma \sqrt{V_s} dW_s^2, \qquad 0 \le t \le T, \tag{12}$$

where $v_0, \kappa, \theta, \sigma > 0$ and μ a constant drift coefficient that represents the expected rate of return of S^x. The couple of processes $(W_t^1, W_t^2)_{t\in[0,T]}$ denotes two correlated standard Brownian motion where the correlation coefficient $\rho \in [-1, 0]$. Also known as a Bessel square process, the existence and uniqueness of the solution of a CIR process is not a direct consequence of the standard results because the square root function is not Lipschitz. Under the risk neutral probability $\mathbb{P}^*$,

$$S_t^x = x + \int_0^t rS_u^x du + \int_0^t \sqrt{V_t} S_u^x dW_u^{1^*}, \qquad 0 \le t \le T, \tag{13}$$

$$V_t = v_0 + \int_0^t \kappa^*(\theta^* - V_u)du + \int_0^t \sigma\sqrt{V_u} dW_u^{2^*}, \qquad 0 \le t \le T, \tag{14}$$

where $\kappa^* = \kappa + \lambda, \quad \theta^* = \frac{\kappa\theta}{\kappa+\lambda}$. The couple of processes (W^{1^*}, W^{2^*}) denotes two correlated standard Brownian motions where the correlation coefficient $\rho \in [-1, 0]$.

Proposition 2 *Under the Feller condition $2\kappa\theta \ge \sigma^2$, the variance risk premium admits the following backward representation*

$$\begin{cases} VRP_t^T = \dfrac{\sigma}{T-t}\displaystyle\int_t^T f_T^{\kappa}(s)\sqrt{V_s} dW_s^2 - \dfrac{\sigma}{T-t}\displaystyle\int_t^T f_T^{\kappa^*}(s)\sqrt{V_s} dW_s^{2^*}, \\ VRP_T^T = 0, \qquad 0 \le t < T, \ \textit{where } f_T^{\rho}(t) := \dfrac{1}{\rho}(\exp\big\{-\rho(T-t)\big\} - 1). \end{cases}$$

In order to prove the result of the above proposition, the following lemma provides the differentiability result for the variance process.

Lemma 2 *In one-dimensional setting and under the Feller condition $2\kappa\theta \ge \sigma^2$,*

i) $\mathbb{P}\big(\inf\{t \ge 0, V_t = 0\} = \infty\big) = 1, \quad \textit{for} \quad v_0 > 0.$
ii) $\mathbb{E}\left(D_s V_t \big| \mathscr{F}_s\right) = \sigma e^{-\kappa(t-s)}\sqrt{V_s}, \quad s \le t \le T.$

where D denotes the Malliavin derivative with respect to the Brownian motion W^2. The result *i*) comes from Proposition 6.2.4 in Lamberton et al. [11]. The second point *ii*) comes from Lemma 4.1 in Alòs et al. [1].

Proof (Proposition 2) The result is direct consequence of Corollary 2. It will be enough to compute the conditional expectation in Eq. (10) to obtain the result. It is a direct consequence of the result *ii*) of Lemma 2.

Remark 1 By Itô's formula, we recover the classical linear term structure in the Heston framework.

Lemma 3 (Classical Result)

$$\begin{cases} VRP_t^T = \dfrac{1}{T-t}F_T(\kappa,\theta,T-t,V_t) - \dfrac{1}{T-t}F_T(\kappa^*,\theta^*,T-t,V_t), \quad 0 \le t < T, \\ VRP_T^T = 0 \quad where \quad F_T(x,y,s,v) := -f_T^x(T-s)(v-y) + ys. \end{cases} \tag{15}$$

Next, a second application that represents the variance risk premium under the non-affine stochastic volatility model entitled $3/2$ model is introduced. We suppose that the couple (S, V) is now governed by the system

$$S_t^x = x + \mu \int_0^t S_s^x ds + \int_0^t \sqrt{V_s} S_s^x dW_s^1, \qquad 0 \le t \le T, \tag{16}$$

$$V_t = v_0 + \int_0^t \eta V_s - b_\epsilon(V_s) ds + \int_0^t \sigma V_s^{3/2} dW_s^2 \qquad 0 \le t \le T, \tag{17}$$

where $v_0, x, \sigma > 0,\ b_\epsilon(x) = \epsilon x^2$. The assumption $\epsilon + \frac{1}{2}\sigma^2 > 0$ insures the strict positiveness of the variance process. In Carr et al. [2], a risk neutral dynamic of the variance process of the $3/2$ model is derived under some plausible assumptions. As in the Heston case, the $3/2$ model preserves its structure. Therefore, we specify the following risk-neutral dynamic of the couple (S, V),

$$S_t^x = x + \int_0^t r S_s^x ds + \int_0^t \sqrt{V_s} S_s^x W_s^{1^*}, \qquad 0 \le t \le T, \tag{18}$$

$$V_t = v_0 + \int_0^t \eta^* V_s - b_\epsilon^*(V_s) ds + \int_0^t \sigma V_s^{3/2} dW_s^{2^*}, \qquad 0 \le t \le T, \tag{19}$$

where $\epsilon^* = \epsilon - \sigma\rho,\ b_\epsilon^*(x) = \epsilon^* x^2$, the constant r denotes the daily interest rate and η^* denotes a constant risk neutral parameter. Due to the negative correlation between the underlying asset and its instantaneous variance dynamic, the assumption $\epsilon^* + \frac{1}{2}\sigma^2 > 0$ hold true from the assumption $\epsilon + \frac{1}{2}\sigma^2 > 0$ under the probability measure $\mathbb{P}$. The following proposition give a backward representation of the variance risk premium (VRP^T) in the framework of the $3/2$ stochastic volatility model.

Proposition 3 *If $\epsilon + \frac{1}{2}\sigma^2 > 0$, then the variance risk premium (VRP^T) follows*

$$\begin{aligned} VRP_t^T &= \frac{1}{\eta\eta^*(T-t)}\Big[(\eta - \eta^*)v_T + \eta^*\mathbb{E}_t(v_T) - \eta\mathbb{E}_t^*(v_T)\Big] \\ &- \frac{\sigma}{\eta^*(T-t)}\int_t^T V_s^{3/2} dW_s^{2^*} + \frac{\sigma}{\eta(T-t)}\int_t^T V_s^{3/2} dW_s^2 - B_{T,t}^{\sigma,\eta,\epsilon} + B_{T,t}^{\sigma,\eta^*,\epsilon^*}, \end{aligned} \tag{20}$$

with $VRP_T^T = 0$ *and* $t \in [0, T)$. $B_{T,t}^{\sigma,\eta,\epsilon}$ *and* $B_{T,t}^{\sigma,\eta^*,\epsilon^*}$ *are defined by*

$$\eta(t-T)B_{T,t}^{\sigma,\eta,\epsilon} = \int_t^T H_s^T(\sigma,\eta,\epsilon)dW_s^2, \quad B_{T,T}^{\sigma,\eta,\epsilon} = 0,$$

$$\eta^*(t-T)B_{T,t}^{\sigma,\eta^*,\epsilon^*} = \int_t^T H_s^T(\sigma,\eta^*,\epsilon^*)dW_s^{2^*}, \quad B_{T,T}^{\sigma,\eta^*,\epsilon^*} = 0,$$

$$\text{where} \quad H_s^T(x,y,z) = \int_s^T e^{y(s-u)}F_{x,y,z}^u(I_u)du, \quad C_{x,y}(t) = \frac{2y}{x^2(1-e^{-yt})}1_{\{x>0, y\in\mathbb{R}^*\}},$$

$$F_{x,y,z}^u(v) = \sqrt{v}\; C_{x,y}(u)\int_0^1 \theta^2\Big(1-\theta\Big)^{\frac{2z}{x^2}-3}\exp\Big\{-\theta\exp(-yu)C_{x,y}(u)v\theta\Big\}d\theta.$$

Proof The result is a direct consequence of Proposition 1. The first term of the dynamic of the variance risk premium is straight forward. The second term is more challenging to achieve. Under $\mathbb{P}$, $V_t = v_0 + \int_0^t \eta V_s - \epsilon V_s^2 ds + \int_0^t \sigma V_s^{3/2} dW_s^2$. By Itô's formula $I_t = \frac{1}{v_0} + \int_0^t (\epsilon + \sigma^2 - \eta I_s)ds - \int_0^t \sigma I_s^{1/2} dW_s^2$. By the Chain rule (See Nualart [13]), $\mathbb{E}_s\left(\int_s^T D_s b_\epsilon(V_t)dt\right) = \mathbb{E}_s\left(\int_s^T b'_\epsilon(V_t)D_sV_tdt\right)$ where $b_\epsilon(x) = -\epsilon x^2$. By the Chain rule, we evaluate the Malliavin derivative of I_t,

$$\mathbb{E}_s\left(\int_s^T D_s b_\epsilon(V_t)dt\right) = 2\epsilon\mathbb{E}_s\left(\int_s^T V_t^3 D_s I_t dt\right).$$

With Fubini's theorem and Theorem 2.2.2 in Nualart [13],

$$\mathbb{E}_s\left(\int_s^T D_s b_\epsilon(V_t)dt\right) = 2\epsilon\sigma\int_s^T \mathbb{E}_s\left(\frac{V_t^3\; Y_t}{Y_s}\right)dt,$$

where $\frac{Y_t}{Y_s} = \exp\{-\eta(t-s)\}\frac{M_t}{M_s}$ and $M_t = \exp\big(\frac{\sigma}{2}\int_0^t \sqrt{V_u}dW_u^2 - \frac{\sigma^2}{8}\int_0^t V_u du\big)$. It will be enough to compute the conditional expectation above to complete the proof. Let us consider the probability $\mathbb{P}_M$ defined with the process M under the assumption $\epsilon + \frac{1}{2}\sigma^2 > 0$. The corresponding Radon-Nikodym densities is given by:

$$M_t = \frac{d\mathbb{P}_M}{d\mathbb{P}}\bigg|_{\mathscr{F}_t}, \quad M_t = \exp\big(\frac{\sigma}{2}\int_0^t \sqrt{V_u}dW_u^2 - \frac{\sigma^2}{8}\int_0^t V_u du\big).$$

We have by change of measure, $\mathbb{E}_s\left(\int_s^T D_s b_\epsilon(V_t)dt\right) = 2\epsilon\sigma \int_s^T \mathbb{E}_s^M\left(V_t^3\right)dt$, where $\mathbb{E}^M$ is the expectation operator associated with the probability measure $\mathbb{P}_M$. Under the probability measure $\mathbb{P}_M$, the process I follows a CIR dynamic. We have

$$I_t = \frac{1}{v_0} + \int_0^t (\epsilon + \frac{1}{2}\sigma^2 - \eta I_s)ds - \int_0^t \sigma I_s^{\frac{1}{2}} dB_s^M, \qquad 0 \leq t \leq T.$$

where B^M is a Brownian motion under $\mathbb{P}_M$. The process B^M is defined by $dB_t^M = dW_t^2 - \frac{1}{2}\sigma V_t^{\frac{1}{2}} dt$. From Lemma 1.1.5 in Diop et al. [7], $\mathbb{E}_s^M\left(V_t^3\right) = \frac{1}{2\sqrt{I_s}} F_{\sigma,\eta,\epsilon}^t(I_s)$ where $F_{x,y,z}^u(v) = \sqrt{v}\ C_{x,y}(u) \int_0^1 \theta^2\left(1-\theta\right)^{\frac{2z}{x^2}-3} \exp\left\{-\theta \exp(-yu)C_{x,y}(u)v\theta\right\} d\theta$. From Proposition 1,

$$\eta(t-T)B_{T,t}^{\sigma,\eta,\epsilon} = \int_t^T \mathbb{E}_s\left(\int_s^T D_s b_\epsilon(V_u)du\right) dW_s^2 = \int_t^T H_s^T(\sigma,\eta,\epsilon)dW_s^2,$$

with $H_s^T(\sigma,\eta,\epsilon) = \int_s^T e^{\eta(s-u)} F_{\sigma,\eta,\epsilon}^u(I_u)du$. We obtain the same structure result for the risk neutral part $B_{T,t}^{\sigma,\eta^*,\epsilon^*}$. Proposition 1 and the two previous equalities conclude. The risk neutral part of (20) is obtained analogously.

5 Conclusion

A new approach to construct the variance risk premium is proposed in our work. We derive this by modeling the variance process as a continuous semi-martingale. We provide two applications: the first with the popular linear Heston stochastic model and the second under the non-affine stochastic 3/2 model. One advantage is that our approach incorporates many standard stochastic volatility models including non-affine specifications. Our probabilistic modeling is essentially devoted to illiquid markets and the result can also be applied to price and hedge many volatility derivatives. It is well known that the times series in financial market are generally discontinuous. A possible extension of this work is to take into account this jump behavior in our modeling. Moreover the incompleteness of many markets generates the non-existence of a unique risk neutral measure. This remark introduces also an interesting problem to investigate in the future development of our work.

Acknowledgements This research received funding from the European Union Seventh Framework Programme (FP7/2007–2013) under grant agreement No. 289032 (HPCFinance).

References

1. Alòs, E., Ewald, C.-O.: Malliavin differentiability of the Heston volatility and applications to option pricing. Available at SSRN 986316 (2007)
2. Carr, P., Sun, J.: A new approach for option pricing under stochastic volatility. Rev. Deriv. Res. **10**, 87–150 (2007)
3. Carr, P., Wu, L.: Variance risk premiums. Rev. Financ. Studies **22**, 1311–1341 (2009)
4. Chorro, C., Guégan, D., Ielpo, F.: A Time Series Approach to Option Pricing. Springer, Berlin (2014)
5. Demeterfi, K., Derman, E., Kamal, M., Zou, J.: A guide to volatility and variance swaps. J. Deriv. **6**, 9–32 (1999)
6. Di Nunno, G., Øksendal, B., Proske, F.: Malliavin Calculus for Lévy Processes with Applications to Finance. Springer, Berlin (2009)
7. Diop, A.: Sur la discrétisation et le comportement a petit bruit d'eds unidimensionnelles dont les coefficients sont a dérivées singulieres, These Université de Nice Sophia-Antipolis (2003)
8. Engle, R.F.: Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflation. Econometrica: J. Econ. Soc. **50**(4), 987–1007 (1982)
9. Harrison, J.M., Kreps, D.M.: Martingales and arbitrage in multiperiod securities markets. J. Econ. Theory **20**, 381–408 (1979)
10. Heston, S.L.: A closed-form solution for options with stochastic volatility with applications to bond and currency options. Rev. Financ. Studies **6**, 327–343 (1993)
11. Lamberton, D., Lapeyre, B.: Introduction to Stochastic Calculus Applied to Finance. CRC Press, Boca Raton (2007)
12. Nelson, D.B.: Conditional heteroskedasticity in asset returns: a new approach. Econometrica: J. Econ. Soc. **59**(2), 347–370 (1991)
13. Nualart, D.: The Malliavin Calculus and Related Topics, vol. 1995. Springer, Berlin (2006)
14. Ocone, D.: Malliavin's calculus and stochastic integral representations of functional of diffusion processes. Stochastics: Int. J. Prob. Stoch. Processes **12**, 161–185 (1984)

Covered Call Writing and Framing: A Cumulative Prospect Theory Approach

Martina Nardon and Paolo Pianca

Abstract The covered call writing, which entails selling a call option on one's underlying stock holdings, is perceived by investors as a strategy with limited risk. It is a very popular strategy used by individual, professional and institutional investors. Previous studies analyze behavioral aspects of the covered call strategy, indicating that hedonic framing and risk aversion may explain the preference of such a strategy with respect to other designs. In this contribution, following this line of research, we extend the analysis and apply Cumulative Prospect Theory in its continuous version to the evaluation of the covered call strategy and study the effects of alternative framing.

1 Introduction

The covered call (CC) writing (or buy-write) is a popular strategy, used both by experienced investors and non-professional traders who are not so familiar with derivatives, but perceive the strategy as risk-reducing with respect to investing on a single stock or a stocks' portfolio. In May 2002 the CBOE released the Buy Write Monthly Index, later called Buy Write Index (BXM), which tracks the performance of a synthetic CC strategy on the S&P 500 Index (SPX). The methodology and performance of the BXM are described by Whaley in [13]; the BXM is a passive total return index formed with a long position on an S&P 500 portfolio and a sequence of short positions on one-month at-the-money (or just out-of-the-money) CCs on the S&P 500 Index. The BXM has become a benchmark for measuring the performance of buy-write implementations.

The seller of the call option owns the underlying asset and her/his risk is limited, but this is not sufficient to explain the success of such a strategy amongst investors, or the preference for CC despite several alternative and less known strategies with similar profit profiles, which register significantly lower trading volume. Using

M. Nardon (✉) • P. Pianca
Department of Economics, Ca' Foscari University of Venice, Cannaregio 873, 30121 Venice, Italy
e-mail: mnardon@unive.it; pianca@unive.it

M. Corazza et al. (eds.), *Mathematical and Statistical Methods for Actuarial Sciences and Finance*, https://doi.org/10.1007/978-3-319-50234-2_12

modern Prospect Theory's arguments, we are able to analyze some aspects that characterize the behavior and choices of the decision makers.

Shefrin and Statman [9] were the first to suggest hedonic framing [10] and risk aversion in the domain of gains as main reasons for departure from standard financial theory: writers of CC prefer this strategy to a stock-only position and are loath to repurchase the call when this entails a realization of a loss, out-of-the money calls are preferred to in-the-money calls in the strategy, fully covered positions are preferred to partially covered ones, CC is preferred to alternative strategies. Recently, Hoffmann and Fischer [5] test empirically all these hypothesis: 60.1% of the respondents prefer the stock-only position with respect to the CC strategy when the profits and losses are described graphically, whereas the CC preference is 60.6% when the profits and losses are described by text; a fully covered position is preferred to a partially CC in 69.7% of the cases; out-of-the-money calls are preferred (87.4%) to in-the-money calls when forming the strategy; at-the-money CCs are preferred (79.3%) to at-the-money naked puts. Such results highlight that decision frames and also the way alternatives are presented to investors do influence actual investment choices.

In this contribution, we extend the analysis of [9] based on a simple one-period binomial model, by evaluating the CC strategy under cumulative prospect theory [11] in a continuous framework [3], focusing in particular on the effects of alternative framing of the results.

The remainder of the paper is organized as follows: Sect. 2 summarizes the main concepts of prospect theory; Sect. 3 introduces the notions of mental accounting and hedonic framing; Sects. 4–7 describe the evaluation of the CC portfolio under different frames; Sect. 8 concludes.

2 Prospect Theory

According to Prospect Theory (PT) risk attitude, loss aversion and probability perception are described by two functions: a value function v and a weighting function w. Outcomes are evaluated relative to a certain *reference point* instead of in terms of final wealth. The shape of the value and weighting functions describe actual investors behaviors. Function v is typically convex in the range of losses (risk-seeking) and concave in the range of gains (risk aversion), it is steeper for losses (loss aversion). Decision makers have also biased probability estimates: they tend to underweight high probabilities and overweight low probabilities.

Let us denote with x_i, for $-m \leq i < 0$ negative outcomes and for $0 < i \leq n$ positive outcomes, with $x_i \leq x_j$ for $i < j$. The prospect value is displayed as follows

$$V = \sum_{i=-m}^{n} \pi_i \cdot v(x_i)\,, \tag{1}$$

with decision weights π_i and values $v(x_i)$ based on relative outcomes.

Specific parametric forms have been suggested in the literature for the value function. A function which is used in many empirical studies is

$$\begin{cases} v^- = -\lambda(-x)^b & x < 0, \\ v^+ = x^a & x \geq 0, \end{cases} \tag{2}$$

with positive parameters that control risk attitude ($0 < a \leq 1$ and $0 < b \leq 1$) and loss aversion ($\lambda \geq 1$); v^- and v^+ denote the value function for losses and gains, respectively. Function (2) has zero as reference point; it is concave for positive outcomes and convex for negative outcomes, it is steeper for losses. Parameters values equal to one imply risk and loss neutrality.

In Cumulative Prospect Theory (CPT) [11] decision weights π_i are differences in transformed cumulative probabilities of gains or losses. Formally:

$$\pi_i = \begin{cases} w^-(p_{-m}) & i = -m \\ w^-\left(\sum_{j=-m}^{i} p_j\right) - w^-\left(\sum_{j=-m}^{i-1} p_j\right) & i = -m+1, \ldots, -1 \\ w^+\left(\sum_{j=i}^{n} p_j\right) - w^+\left(\sum_{j=i+1}^{n} p_j\right) & i = 0, \ldots, n-1 \\ w^+(p_n) & i = n, \end{cases} \tag{3}$$

with w^- for losses and w^+ for gains, respectively.

In financial applications, in particular in the evaluation of options, prospects may involve a continuum of values; hence, Prospect Theory cannot be applied directly in its original or cumulative versions. Davies and Satchell [3] provide the continuous cumulative prospect value:

$$V = \int_{-\infty}^{0} \psi^-(F(x)) f(x)\, v^-(x)\, dx + \int_{0}^{+\infty} \psi^+(1 - F(x)) f(x)\, v^+(x)\, dx, \tag{4}$$

where $\psi = \frac{dw(p)}{dp}$ is the derivative of the weighting function w with respect to the probability variable, F is the cumulative distribution function and f is the probability density function of the outcomes.

Prosect theory involves a probability weighting function which models probabilistic risk behavior. A weighting function w is uniquely determined, it maps the probability interval [0, 1] into [0, 1], and is strictly increasing, with $w(0) = 0$ and $w(1) = 1$. In this work we will assume continuity of w on [0, 1], even thought in the literature discontinuous weighting functions are also considered. The *curvature* of the weighting function is related to the risk attitude towards probabilities. Empirical evidence suggests a particular shape of probability weighting functions: small probabilities are overweighted $w(p) > p$, whereas individuals tend to underestimate large probabilities $w(p) < p$. This turns out in a typical *inverse-S shaped* weighting function: the function is initially concave (probabilistic risk-seeking) for probabilities in the interval $(0, p^*)$, and convex (probabilistic risk aversion) in the interval $(p^*, 1)$, for a certain value of p^*. A linear weighting function describes probabilistic risk neutrality or objective sensitivity towards probabilities, which characterizes Expected Utility. Empirical findings indicate that the intersection between the weighting function and the 45$^{\text{o}}$ line, $w(p) = p$, is for p^* in the interval $(0.3, 0.4)$.

Different parametric forms for the weighting function with the above mentioned features have been proposed in the literature, and their parameters have been estimated in many empirical studies. In this work we applied, in particular, the function suggested by [11]:

$$w(p) = \frac{p^\gamma}{(p^\gamma + (1-p)^\gamma)^{1/\gamma}}, \qquad (5)$$

where γ is a positive constant (with some constraint in order to have an increasing function), with $w(0) = 0$ and $w(1) = 1$. The parameter γ captures the degree of sensitivity toward changes in probabilities from impossibility (zero probability) to certainty. When $\gamma < 1$, one obtains the typical inverse-S shaped form; the lower the parameter, the higher is the curvature of the function.

3 Hedonic Framing

Standard finance and, in particular, option pricing assume frame invariance, where investors are indifferent among frames of cash flows, but many empirical studies highlight that the framing of alternatives exerts a crucial effect on investment choices. Moreover, investors might not be aware of the frames that affect their actual decisions. Prospect theory posits that individuals evaluate outcomes with respect to deviations from a reference point rather than with respect to net final wealth. Decision frames are influenced by the way in which alternatives are presented to investors. People may keep different *mental accounts* for different types of outcomes. When combining these accounts to obtain overall result, typically they do not simply sum up all monetary amounts, but intentionally use *hedonic frame* (see Thaler [10]) such that the combination of the outcomes appears more favorable and increases their utility.

Consider the simple case of two sure outcomes x and y, the hedonic optimizer would combine the results according to the following rule:

$$V = \max\{v(x+y),\ v(x) + v(y)\}\,. \qquad (6)$$

Outcomes are aggregated or segregated depending on what leads to the highest possible prospect value: multiple gains are preferred to be segregated (*narrow framing*) to enjoy the gains separately, losses are preferred to be integrated with other losses (or large gains) in order to ease the pain of the loss. Considering mixed outcomes, these would be integrated in order to cancel out losses when there is a net gain or a small loss; in case of large losses and a small gain, they usually are segregated in order to preserve the *silver lining*. This is due to the shape of the value function in PT, characterized by risk-seeking or risk aversion, diminishing sensitivity and loss aversion. Loss aversion implies also that the impact of losses is more important than that of gains of same amount.

An investment in a portfolio can be isolated (narrow framing) or integrated, which reflects a broader decision frame of the investors (see [7]). Framing can be intertemporal: gains and losses can be time segregated or aggregated, depending on how the perception of the results is affected by the evaluation period. Invertors are influenced by narrow framing and regret when they sell portfolio "winners" and keep portfolio "losers", in order to avoid the realization of a loss (*disposition effect*), but they integrate outcomes through simultaneous trades when they sell a loser together with a winner to reduce their regret (see [7]). When we consider a financial option, the premium is cashed in advance with respect to the realization of the payoff, hence these outcomes could be evaluated into separate mental accounts or jointly. The option premium itself might be perceived as an income, cash that can be used immediately for consumption (according to the *life cycle theory* [9]).

In the following sections we will evaluate the results of a CC portfolio under alternative frames, analyzing the effects of the hedonic and time framing. Covered calls are frequently promoted as *"an investment strategy that can make you extra money"* (see the discussion in [9]) and proposed to the investors as the sum of three possible sources of profit (three mental accounts): the call premium (considered as an extra dividend), regular dividends, the capital gain on the stock (when the strategy is formed with out-of-the-money calls, and in case of exercise).

4 Covered Call Writing versus Stock Only Position

In this and the following sections, we consider strategies which include options on non-dividend paying assets for which early exercise is not convenient or possible (American or European calls on single stock or indexes). We retain the usual assumptions of frictionless markets; in particular, we do not consider the effects of taxes and transaction costs, which may lead to different choices. We assume that the dynamics of the underlying price process S is governed by a geometric Brownian motion as in the model [1].

Shefrin and Statman [9] evaluate the CC strategy in a simple one-period binomial model under PT in its original version [6]; they use only a value function, do not consider probability weighting, and assume a zero risk-free interest rate r. We extend the analysis and model the problem under continuous CPT.

Let $S_0 > 0$ be the current stock price, it is the amount paid to buy one share of stock at time $t = 0$; consider a time horizon of T years, the buyer of the stock registers a profit if $S_T > S_0 e^{rT}$ (a loss otherwise).

The prospect value of the stock-only position V^s is displayed as

$$V^s = \int_0^{S_0 e^{rT}} \psi^-(F(x)) f(x) v^-(x - S_0 e^{rT}) dx + \int_{S_0 e^{rT}}^{+\infty} \psi^+(1 - F(x)) f(x) v^+(x - S_0 e^{rT}) dx \,, \tag{7}$$

where f and F are the probability density function and cumulative probability function of S_T.

In the definition of the prospect value V^s, we have assumed that the price S_0 could have been invested at the risk-free interest rate r. Under this hypothesis, a risk-neutral investor (setting the parameters $a = b = 1$, $\lambda = 1$, $\gamma^+ = \gamma^- = 1$, and taking risk-neutral dynamics for the process S) has a value V^s equal to zero, equivalent to the status quo.

Assume now an out-of-the-money[1] call option written on the same stock, with strike price X (with $X > S_0 e^{rT}$) and maturity T. Let c be the option premium (computed with the Black-Scholes formula). The prospect value of a CC position, when the option premium and the CC result are segregated into two mental accounts, is given by

$$\begin{aligned} V^{cc} = v^+(c\, e^{rT}) &+ \int_0^{S_0 e^{rT}} \psi^-(F(x)) f(x) v^-(x - S_0 e^{rT}) dx \\ &+ \int_{S_0 e^{rT}}^{X} \psi^+(1 - F(x)) f(x) v^+(x - S_0 e^{rT}) dx \\ &+ \int_X^{+\infty} \psi^+(1 - F(x)) f(x) v^+(X - S_0 e^{rT}) dx \,. \end{aligned} \tag{8}$$

It is worth noting that in the first term in (8), we have assumed that the call premium is not used for consumption, but invested at the risk-free interest rate (otherwise one has to replace the term with $v^+(c)$).

Shefrin and Statman argue that "the PT expected value of the CC position exceeds the PT value of the stock-only position for investors who are sufficiently risk-averse in the domain of gains". This hypothesis seems not be confirmed by Hoffmann and Fischer [5], whereas the authors find strong evidence for framing effects. de Groot and Dijkstra [4] argue that a PT investor with above average risk aversion for gains prefers the CC.

[1]Here we take an out-of-the-money forward option; alternatively, the strategy can be built with out-of-the-money options with $X > S_0$.

The CC position (8) is preferred to the stock only position (7) if

$$v^{+}(c\,e^{rT}) + \int_{X}^{+\infty} \psi^{+}(1 - F(x))f(x)v^{+}(X - S_0 e^{rT})dx$$
$$> \int_{X}^{+\infty} \psi^{+}(1 - F(x))f(x)v^{+}(x - S_0 e^{rT})dx\,. \quad (9)$$

Note that inequality (9) depends only on the value function in the domain of gains (when applying function (2), it depends only on the value of the parameter a), and on the weighting function w^{+}. Considering a concave value function v^{+}, one obtains a higher value on the right hand side of (9) when the option premium is segregated from the (positive) result of selling the stock at $X > S_0 e^{rT}$. Numerical results confirm a framing effects.

5 Fully Covered versus Partially Covered Call Position

The CC strategy discussed in the previous section is a fully covered position, which entails buying one share and selling one call option. Empirical evidence suggests that most CC traded strategies are fully covered. Hoffmann and Fischer [5], in their investigation, find significant preference for fully covered over partially covered positions. Shefrin and Statman [9] compare in their one-period binomial model a fully covered position with a partially covered one composed by 1.25 shares and 2.5 calls. Both positions have the same profit and loss profiles. The authors examine the prospect values of two possible frames with identical cash flows, consisting of: (1) a fully covered position of 1.25 shares and a short position of 1.25 (naked) calls; (2) a stock only position of 1.25 shares and a short position of 2.5 (naked) calls. They conclude that the prospect value in the first frame exceeds the one in the second frame for investors who are highly risk-averse in the domain of gains and highly risk-seeking in the domain of losses. Moreover, the fully covered position is preferred to the partially covered one when the concavity and convexity of the value function for gains and losses, respectively, are sufficiently pronounced.

We can generalize the analysis of [9] in a continuous setting, comparing the following two frames:

1. a fully covered position of $\alpha\,\beta$ shares (with $0 < \alpha < 1$) and a short position of $(1-\alpha)\beta$ (naked) calls ($\beta > 0$);
2. a stock only position of $\alpha\,\beta$ shares and a short position of β (naked) calls.

As limit cases, when $\alpha = 0$ we have a short position of β (naked) calls, and when $\alpha = 1$ the short position of β calls is fully covered.

The prospect value in the first frame, when the option premium and the CC net result are segregated into separate mental accounts, is given by

$$
\begin{aligned}
V^{pcc1} &= V^{cc}(\alpha\beta) + V^{sc}((1-\alpha)\beta) \\
&= v^{+}(\alpha\beta\, c\, e^{rT}) + \int_{0}^{S_0 e^{rT}} \psi^{-}(F(x)) f(x)\, v^{-}\left(\alpha\beta\,(x - S_0 e^{rT})\right) dx \\
&\quad + \int_{S_0 e^{rT}}^{X} \psi^{+}(1-F(x)) f(x)\, v^{+}\left(\alpha\beta\,(x - S_0 e^{rT})\right) dx \\
&\quad + \int_{X}^{+\infty} \psi^{+}(1-F(x)) f(x)\, v^{+}\left(\alpha\beta\,(X - S_0 e^{rT})\right) dx \\
&\quad + v^{+}\left((1-\alpha)\beta\, c\, e^{rT}\right) \\
&\quad + \int_{X}^{+\infty} \psi^{-}(1-F(x)) f(x)\, v^{-}\left((1-\alpha)\beta\,(X - x)\right) dx\,.
\end{aligned}
\tag{10}
$$

The last two terms in (10) represent the prospect value of the writer position of $(1-\alpha)\beta$ calls in the time segregated frame V^{sc} (see the Appendix), when the option premium received at time $t = 0$ is evaluated in a separate mental account. We assume that the premium is reinvested at the risk free interest rate r.

Alternatively, the naked position could be evaluated in a time aggregated mental account V^{ac} (see the Appendix) as follows:

$$
V^{pcc1} = V^{cc}(\alpha\beta) + V^{ac}((1-\alpha)\beta),
$$

where

$$
\begin{aligned}
V^{ac}((1-\alpha)\beta) &= w^{+}\,(F(X))\, v^{+}\left((1-\alpha)\beta\, c\, e^{rT}\right) \\
&\quad + \int_{X}^{X + c\, e^{rT}} \psi^{+}\,(F(x)) f(x)\, v^{+}\left((1-\alpha)\beta\left(c\, e^{rT} - (x - X)\right)\right) dx \\
&\quad + \int_{X + c\, e^{rT}}^{+\infty} \psi^{-}\,(1-F(x)) f(x)\, v^{-}\left((1-\alpha)\beta\left(c\, e^{rT} - (x - X)\right)\right) dx.
\end{aligned}
\tag{11}
$$

Note also that the CC strategy in (10) is built with out-of-the money (forward) calls; other assumptions are also possible and will be discussed in the next section.

The prospect value in the second frame is the sum of a stock only position of $\alpha\beta$ shares and a short position of β naked calls; when the option premium and the option payoff are segregated into separate mental accounts, one obtains

$$
\begin{aligned}
V^{pcc2} &= V^{s}(\alpha\beta) + V^{sc}(\beta) \\
&= \int_{0}^{S_0 e^{rT}} \psi^{-}(F(x))f(x)\, v^{-}\left(\alpha\beta(x - S_0 e^{rT})\right) dx \\
&\quad + \int_{S_0 e^{rT}}^{+\infty} \psi^{+}(1 - F(x))f(x)\, v^{+}\left(\alpha\beta(x - S_0 e^{rT})\right) dx \\
&\quad + v^{+}\left(\beta\, c\, e^{rT}\right) + \int_{X}^{+\infty} \psi^{-}(1 - F(x))f(x)\, v^{-}\left(\beta\,(X - x)\right) dx\,. \qquad (12)
\end{aligned}
$$

The last two terms in (12) represent the prospect value of the writer position in the time segregated frame; as an alternative, such a position can be evaluated in the time aggregated case.

6 Covered Call Writing with Out-of-the-Money versus In-the-Money Options

In the previous sections, CC strategy was formed with out-of-the-money options. This is the way in which more often the strategy has been proposed to the investors, in order to obtain a profit in case of exercise. Due to the shape of the value function and hedonic framing, prospect investors will tend to segregate such a gain.

We consider a call option which is in-the-money forward with $X < S_0 e^{rT}$ (the case $X < S_0$ can be treated analogously). When $S_T \geq X$ the option is exercised and the profit and loss profile for the CC investor is: $ce^{rT} - (S_T - X) + (S_T - S_0 e^{rT})$, assuming that both the call premium and the amount S_0 could be invested at the risk free rate.[2] By construction of the strategy, the difference $X - S_0 e^{rT}$ is negative. On the other hand, when $S_T < X$ the option is not exercised and the result of the long position on the underlying, $S_T - S_0 e^{rT}$, is negative as well. The CC strategy gives rise to a profit only if the call premium is sufficiently higher than the loss on the combined positions on the call and the underlying.

An investor who applies hedonic framing will segregate the call premium from other losses in order to obtain a better prospect value and, when such a value is negative, to preserve the silver lining. Formally we have:

$$
\begin{aligned}
V^{cc-itm} &= v^{+}(c\, e^{rT}) + \int_{0}^{X} \psi^{-}(F(x))f(x)v^{-}(x - S_0 e^{rT})dx \\
&\quad + \int_{X}^{+\infty} \psi^{-}(1 - F(x))f(x)v^{-}(X - S_0 e^{rT})dx\,.
\end{aligned}
\qquad (13)
$$

[2] This hypothesis can be easily relaxed, considering as in [9] that the cashed premium could be used for consumption or disregarding the effects of a non zero interest rate.

Shefrin and Statman [9] compare two portfolios formed with stocks, short positions in different proportions of out-of-the-money or in-the-money call options and borrowed cash (when the strategy is formed with in-the-money options), in a one-period binomial framework. In their particular example, both portfolios entail the same initial cash outflow and the same gains or losses in the up and down states. When comparing the prospect value of the two strategies, the authors conclude that CC with out-of-the-money options is preferred when investors are sufficiently risk-averse in the domain of gains and risk-seeking in the domain of losses.

In our analysis, we consider continuous variables, and the comparison of portfolios with identical cash outflows and results is no longer possible. Nevertheless, we observe that both the integrals in (13) result in a negative value, which has to be compensated by the positive value of the call premium in order to obtain a positive prospect value. This requires that the option is over-evaluated by the prospect investors, in order to prefer the strategy to the status quo. It is difficult to separate the different effects of framing, subjective evaluation of gains and losses, and the over- and under-weighting of probabilities, which requires a large numerical analysis which is not included in this contribution. Numerical experiments based on Tversky and Kahneman [11] sentiment parameters yield negative prospect values. In particular, when the parameter λ, which models loss aversion in (2), departs from the value 1, one obtains negative prospect values (excluding the effects of the probability weighting function). When we include in the analysis the weighting function, low probabilities of extreme events are overweighted and this may lead to different decisions about the preferred frame.

7 Selling the Underlying Asset versus Buying Back the Option

Shefrin and Statman [9] argue that a PT investor will not repurchase the call option when it is likely to be exercised, because this would imply the realization of a loss. They also explain that this happens when investors are sufficiently risk-averse in the domain of gains. Hoffmann and Fischer [5] tested this hypothesis, finding no statistical evidence for the preference of repurchasing the call or selling the stock.

We compare two different frames: first we consider an investor who segregates the call premium from the combined result of the call and stock positions; second, the investor separates the result of the call position from the gains or losses on the stock. In both cases let us assume $X > S_0 e^{rT}$, hence the CC strategy is formed with out-of-the-money calls.

In the first frame the prospect value V^{cc} is given as in (8), whereas the prospect value of the CC strategy when the call is repurchased is:

$$
\begin{aligned}
V^{ccr} = & \, w^{+}(F(X))v^{+}\left(ce^{rT}\right) \\
& + \int_{X}^{X+ce^{rT}} \psi^{+}(F(x))f(x)v^{+}\left(ce^{rT}-(x-X)\right)dx \\
& + \int_{X+ce^{rT}}^{+\infty} \psi^{-}(1-F(x))f(x)v^{-}\left(ce^{rT}-(x-X)\right)dx \\
& + \int_{0}^{S_0e^{rT}} \psi^{-}(F(x))f(x)v^{-}\left(x-S_0e^{rT}\right)dx \\
& + \int_{S_0e^{rT}}^{+\infty} \psi^{+}(1-F(x))f(x)v^{+}\left(x-S_0e^{rT}\right)dx\,.
\end{aligned}
\tag{14}
$$

In this case, the prospect value of the call position is given by the first three terms, and the gains and losses on the stock are evaluated through the last two integrals.

If V^{ccr} is lower than V^{cc} as defined in (8), the investor would prefer selling the stock (hence realizing a gain), rather than repurchasing the option which entails the realization of a loss. It is worth noting that the preference depends not only on the shape of the value function in the region of gains, but also on the risk-seeking behavior of the investor in the domain of losses, and his or her loss aversion.

In some numerical experiments, first neglecting the effects of the probability weighting function (taking $\gamma^{+} = \gamma^{-} = 1$), we found that due to the shape of the value function the frame defined by (8) is preferred, yielding to a higher prospect value, even with values of the parameters a and b close to unit. As in the comparisons with other frames in the previous sections, the loss aversion parameter plays an important role, because it enhances the negative valuation of losses. When also the probability perception through the weighting function is considered, this can lead to different results, due to the fact that low probabilities of extreme events are overweighted.

8 Alternative Trading Strategies and Concluding Remarks

In this paper we have analyzed the effects of hedonic framing and PT on the evaluation of one of the most popular trading strategies. The CC writing is perceived as a strategy with limited risk, but this is not sufficient to explain its success amongst investors, or the preference for CC with respect to alternative strategies with similar profit profiles. For example, the CC profit and loss profile can be compared with that of a naked put; [5] in their empirical study find that investors have a preference for the first strategy.

Under the CPT framework applied in this contribution, other structured financial products can be evaluated (see [2]).

Numerical results (which are reported in a separate paper) suggest that different frames may lead to different preference for alternative designs of the CC strategy.

Appendix: European Options Valuation

Versluis et al. [12] provide the prospect value of writing call options, considering different intertemporal frames. The option premium represents a sure gain for the writer, whereas the negative payoff is a potential loss; the writer may aggregate or segregate such results in different ways. Nardon and Pianca [8] extend the model of [12] to the case of put options, considering the problem both from the writer's and holder's perspectives, and use alternative weighting functions.

Let S_t be the price at time $t \in [0, T]$ of the underlying asset of a European option with maturity T. Let c be the call option premium with strike price X. At time $t = 0$, the option's writer receives c and can invest the premium at the risk-free rate r, obtaining $c\, e^{rT}$. At maturity, the amount $S_T - X$ is paid to the holder if the option expires in-the-money.

In the *time segregated* case the option premium is evaluated separately (through the value function) from the option payoff. Considering zero as a reference point (*status quo*), the prospect value of the writer's position is

$$V^{sc} = v^+ \left(c\, e^{rT}\right) + \int_X^{+\infty} \psi^- \left(1 - F(x)\right) f(x) v^- \left(X - x\right)\, dx\,, \tag{15}$$

with f and F being, respectively, the probability density function and the cumulative distribution function of the future underlying price S_T, and v is defined as in (2). One equates V^{sc} at zero and solves for the price c.

In the *time aggregated* frame, gains and losses are integrated in a unique mental account, then one obtains the prospect value

$$\begin{aligned} V^{ac} &= w^+ \left(F(X)\right) v^+ \left(c\, e^{rT}\right) \\ &+ \int_X^{X + c\, e^{rT}} \psi^+ \left(F(x)\right) f(x) v^+ \left(c\, e^{rT} - (x - X)\right)\, dx \\ &+ \int_{X + c\, e^{rT}}^{+\infty} \psi^- \left(1 - F(x)\right) f(x) v^- \left(c\, e^{rT} - (x - X)\right)\, dx. \end{aligned} \tag{16}$$

In this case, the option price evaluated by a PT investor is implicitly defined by the equation $V_a = 0$ and has to be determined numerically.

References

1. Black, F., Scholes, M.: The pricing of options and corporate liabilities. J. Polit. Econ. **81**, 637–654 (1973)
2. Breuer, W., Perst, A.: Retail banking and behavioral financial engineering: the case of structured products. J. Bank. Financ. **31**, 827–844 (2007)
3. Davies, G.B., Satchell, S.E.: The behavioural components of risk aversion. J. Math. Psychol. **51**, 1–13 (2007)
4. de Groot, J.S., Dijkstra, T.K.: Writing covered calls: should be done? Gröninger SOM Research School, SOM Research Reports Series (1996)
5. Hoffmann, A.O.I., Fischer, E.T.: Behavioral aspects of covered call writing: an empirical investigation. J. Behav. Financ. **13**, 66–79 (2012)
6. Kahneman, D., Tversky, A.: Prospect theory: an analysis of decision under risk. Econometrica **47**, 263–291 (1979)
7. Kumar, A., Lim, S.S.: How do decision frames influence the stock investment choices of individual investors? Manag. Sci. **54**, 1052–1064 (2008)
8. Nardon, M., Pianca, P.: A behavioural approach to the pricing of European options. In: Corazza, M., Pizzi, C. (eds.) Mathematical and Statistical Methods for Actuarial Sciences and Finance, pp. 217–228. Springer, Milano (2014)
9. Shefrin, H., Statman, M.: Behavioral aspects of the design and marketing of financial products. Financ. Manage. **22**, 123–134 (1993)
10. Thaler, R.H.: Mental accounting and consumer choice. Mark. Sci. **4**(3), 199–214 (1985)
11. Tversky, A., Kahneman, D.: Advances in prospect theory: cumulative representation of the uncertainty. J. Risk Uncertain. **5**, 297–323 (1992)
12. Versluis, C., Lehnert, T., Wolff, C.C.P.: A cumulative prospect theory approach to option pricing. Working Paper. LSF Research Working Paper Series 09–03, Luxembourg School of Finance (2010)
13. Whaley, R.E.: Return and risk of CBOE buy write monthly index. J. Deriv. **10**, 35–42 (2002)

Optimal Portfolio Selection for an Investor with Asymmetric Attitude to Gains and Losses

Sergei Sidorov, Andrew Khomchenko, and Sergei Mironov

Abstract The description of Cumulative Prospect Theory (CPT) includes three important parts: a value function over outcomes, $v(\cdot)$; a weighting function over *cumulative* probabilities, $w(\cdot)$; CPT-utility as unconditional expectation of the value function v under probability distortion w. In this paper we consider the problem of choosing an CPT-investor's portfolio in the case of complete market. The problem of finding the optimal portfolio for CPT-investor is to maximize the unconditional expectation of the value function v under probability distortion w over terminal consumption, subject to budget constraint on initial wealth. We find the optimal payoffs for CPT-investor for the classic Black-Scholes environment assuming that there are a single lognormally distributed stock and a risk free bond. We compare the optimal payoffs of CPT-investor with the optimal payoffs of the investor that maximizes expected power utility over terminal payoffs, subject to budget constraint on initial wealth.

1 Introduction

One of the classic problems of the portfolio investment theory is the following one: to find an optimal portfolio for a given set of assets with the known prices and distribution function of returns.

It is well-known [4] that in the case of the complete market and the absence of arbitrage there is a unique positive discount factor m, such that the fair price of the asset $p = E_0(mx)$, where E_0 denotes the conditional expectation at the initial time $t = 0$, and x is the flow of future payments.

Let x_T be the (random) price of the portfolio at time $t = T$, and w be the wealth of the investor at time $t = 0$. Then the problem of finding the optimal portfolio can be represented as follows:

$$E_0(u(x_T)) \to \max_{\{x_T\}}, \quad s.t. \; E_0(mx_T) = w, \tag{1}$$

S. Sidorov • A. Khomchenko • S. Mironov (✉)
Saratov State University, 83, Astrakhanskaya, 410012 Saratov, Russia
e-mail: sidorovsp@info.sgu.ru; aahomchenko@gmail.com; mironovsv@info.sgu.ru

M. Corazza et al. (eds.), *Mathematical and Statistical Methods for Actuarial Sciences and Finance*, https://doi.org/10.1007/978-3-319-50234-2_13

where $E_0(u(x_T))$ is the expected value (at the time $t = 0$) of utility $u(x_T)$, $\max_{\{x_T\}}$ means that the investor finds the optimal payoff in every state of nature at time T, m is a stochastic discount factor.

The classical theory of portfolio investment considers an investor with a concave utility function. However, Tversky and Kahneman [10] provide examples and demonstrations showing that under the conditions of laboratory experiments the predictions of expected utility theory are regularly disturbed. Furthermore, a new theory was proposed—the prospect theory which accounts for people's behavior in decision-making under risk in the experiments where the traditional theory of expected utility failed.

Cumulative prospect theory (CPT) proposed in [17] is the further development of prospect theory, the difference being that cumulative probabilities undergo transformation, rather than the probabilities themselves. Current economic literature views the cumulative prospect theory as one of the best models explaining the behavior of the players, as well as the behavior of investors in the experiment and in decision-making under risk.

The work [1] demonstrates that a number of decision making paradoxes can be resolved with the help of the prospect theory, but it is not a ready-made model for economic applications as the author notes. However, in the recent years we can see growing interest in the issues lying in the intersection of prospect theory and portfolio optimization theory.

It should be mentioned, that because of the computational burden connected to the complexity of the numerical evaluation of the CPT-utility, there are not so many research papers on the portfolio optimization problem under the framework of both prospect theory [6, 11] and cumulative prospect theory [2, 3, 8, 9, 14, 15, 18]. Most of the research is based on the supposition that testing data are normally allocated. However, many asset allocation problems include non-normally distributed returns as far as commodities generally have fat tails and are skewed [7].

Levy and Levy [11] attempt to find the portfolio with the highest prospect theory utility among the other portfolios in the mean variance efficient frontier. Elaborating on this idea, the work of Pirvu and Schulze [14] proves that an analytical solution of the problem is mostly equivalent to maximising the CPT-utility function along the mean-variance efficient frontier.

Among recent works we should mention the paper of Nardon and Pianca [12], which evaluates European options within the continuous cumulative prospect theory.

The paper [16] is based on the ideas of [3] and deals with the issue of finding the optimal portfolio for an investor with asymmetric attitudes to gains and losses described in Tversky and Kahneman's prospect theory. It examined the portfolio optimization problem for an investor who follows the assumptions of the prospect theory and the cumulative prospect theory under conditions on the stochastic behavior both of the portfolio price and the discount factor.

In this paper we consider the problem of choosing an CPT-investor's portfolio in the case of complete market. The problem of finding the optimal portfolio for CPT-investor is to maximize the unconditional expectation of the value function v under probability distortion w over terminal consumption, subject to budget constraint on initial wealth. In the beginning, we briefly present the summary of main ideas of the cumulative prospect theory, and then we proceed to the problem of finding the optimal payoffs for CPT-investor for the classic Black-Scholes environment assuming that there is a single lognormally distributed stock and a risk free bond. We compare the optimal payoffs of CPT-investor with the optimal payoffs of the investor that maximizes expected power utility over terminal payoffs, subject to budget constraint on initial wealth.

2 EU-, PT- and CPT- Investors

2.1 Expected Utility Theory

The most popular approach to the problem of portfolio choice under risk and uncertainty is the expected utility hypothesis. For an introduction to utility theory, see [5]. Von Neumann and Morgenstern [13] formulated the utility in terms of a function. Let X is the set of all possible payoffs of a portfolio at time $t = T$, x refers to the element of X, and let $u : X \to \mathbb{R}$ denote a utility function such that the value of $u(x)$ is a measure of the decision maker's preference derived from the payoff x: $x \succeq y \;\Leftrightarrow\; u(x) \geq u(y)$, where $x \succeq y$ means the outcome x is preferred at least as much as the outcome y. Thus, the relationship between wealth and the utility of consuming this wealth is described by a utility function, $u(\cdot)$. Let $f_\xi(x)$ be the probability density function of a random variable ξ.

Definition 1 The expected utility (EU) of a portfolio G is the expected value of the utility functions of possible payoffs of the portfolio at time $t = T$ weighted by the corresponding probabilities:

$$U_{EU}(G) = \int_X u(x) f_\xi(x) dx.$$

The expected utility hypothesis states that the individual (EU-investor) will make decisions following the principle of maximizing the value of his expected utility.

The most exploited type of utility functions is the power utility function defined by $u(x) = \frac{x^{1-\gamma}}{1-\gamma}$, where $\gamma \in (0, 1)$. Marginal utility is $u'(x) = x^{-\gamma} > 0$ for all $x > 0$. We have $u''(x) = -\gamma x^{-\gamma-1} < 0$ for all $x > 0$.

2.2 Prospect Theory

Prospect theory (PT) has three essential distinctions from Expected Utility Theory:

- investor makes investment decisions based on deviation of his/her final wealth from a reference point and not according to his/her final wealth, i.e. PT-investor concerned with deviation of his/her final wealth from a reference level, whereas Expected Utility maximizing investor takes into account only the final value of his/her wealth.

- utility function is S-shaped with turning point in the origin, i.e. investor reacts asymmetrical towards gains and losses; moreover, he/she dislikes losses with a factor of $\lambda > 1$ as compared to his/hers liking of gains.
- investor evaluates gains and losses not according to the real probability distribution per ce but on the basis of the transformation of this real probability distribution, so that investor's estimates of probability are transformed in the way that small probability (close to 0) is overvalued and high probability (close to 1) is undervalued.

PT includes three important parts:

- a value function over outcomes, $v(\cdot)$;
- a weighting function over probabilities, $\omega(\cdot)$;
- PT-utility as unconditional expectation of the value function v under probability distortion ω.

Definition 2 The value function derives utility from gains and losses and is defined as follows [17]:

$$v(x) = \begin{cases} x^{\alpha}, & \text{if } x \geq 0, \\ -\lambda(-x)^{\beta}, & \text{if } x < 0. \end{cases} \tag{2}$$

Note that the value function is convex over losses if $0 \leq \beta \leq 1$ and it is strictly convex if $0 < \beta < 1$. Moreover, the value function reflects loss aversion when $\lambda > 1$. It follows from the fact that individual investors are more sensitive to losses than to gains. Kahneman and Tversky estimated the parameters of the value function $\alpha = \beta = 0.88$, $\lambda = 2.25$ based on experiments with gamblers [10].

Definition 3 Let $f_{\xi}(x)$ be the probability density function of a random variable ξ. The PT-probability weighting function $w : [0, 1] \to [0, 1]$ is defined by

$$w\left(f_{\xi}(x)\right) = \frac{\left(f_{\xi}(x)\right)^{\delta}}{\left((f_{\xi}(x))^{\delta} + (1 - f_{\xi}(x))^{\delta}\right)^{1/\delta}}, \ \delta \leq 1 \tag{3}$$

It is easy to verify that

1. $w : [0, 1] \to [0, 1]$ is differentiable on [0,1];
2. $w(0) = 0$, $w(1) = 1$;
3. if $\delta > 0.28$ then w is increasing on [0,1];
4. if $\delta = 1$ then $w(f_{\xi}(x)) = f_{\xi}(x)$.

The function $w(\cdot)$ is well-defined when α, β are less than 2δ. In the following we will assume that $0.28 < \delta \leq 1$ and $\alpha < 2\delta$, $\beta < 2\delta$.

Definition 4 The PT-utility of a portfolio G with stochastic return ξ is defined as [10]

$$U_{PT}(G) = \int_{-\infty}^{\infty} v(x) w(f_\xi(x)) dx, \quad (4)$$

where $f_\xi(x)$ is the probability density function of ξ.

2.3 Cumulative Prospect Theory

We will consider the development of the prospect theory, Cumulative Prospect Theory, published in 1992 [17]. The description of CPT includes three important parts:

- a value function over outcomes, $v(\cdot)$;
- a weighting function over *cumulative* probabilities, $w(\cdot)$;
- CPT-utility as unconditional expectation of the value function v under probability distortion w.

Definition 5 Let $F_\xi(x)$ be cumulative distribution function (cdf) of a random variable ξ. The probability weighting function $w : [0, 1] \to [0, 1]$ is defined by

$$w\left(F_\xi(x)\right) = \frac{\left(F_\xi(x)\right)^\delta}{\left((F_\xi(x))^\delta + (1 - F_\xi(x))^\delta\right)^{1/\delta}}, \ \delta \leq 1 \quad (5)$$

Definition 6 The CPT-utility of a gamble G with stochastic return ξ is defined as in [1]

$$U_{CPT}(G) = \int_{-\infty}^{0} v(x) dw(F_\xi(x)) - \int_{0}^{\infty} v(x) dw(1 - F_\xi(x)), \quad (6)$$

where $F_\xi(x)$ is cumulative distribution function of ξ.

If we apply integration by part, then CPT-utility of G defined in (6) can be rewritten as

$$U_{CPT}(G) = \int_{0}^{\infty} w(1 - F_\xi(x)) dv(x) - \int_{-\infty}^{0} w(F_\xi(x)) dv(x). \quad (7)$$

3 Portfolio Optimization Problem

Let us consider the problem of choosing a CPT-investor's portfolio in the case of complete market. It is well-known [4] that given the absence of arbitrage opportunities, there is a unique positive stochastic discount factor m, such that the fair price of asset $p = E(mx)$ for any future payoff x, E is the conditional expectation at the initial time $t = 0$. Let W_0 denote the initial wealth (portfolio) of the investor at time $t = 0$. Let x_T denote the payoff of the investor's portfolio at time $t = T$. Then the price of the portfolio at time $t = 0$ is $p(x_T) = E(mx_T)$. The problem of finding the optimal portfolio for PT- and CPT-investors can be written as

$$E^{\omega}(v(x_T - X)) \to \max_{x_T}, \quad s.t. \;\; E(mx_T) = W_0, \tag{8}$$

where X is a reference point, $E^{\omega}(\cdot)$ denotes the transformed expected value of $(\cdot)$ under the probability transformation ω, defined in Definition 5 (for CPT-investor) and in Definition 3 (for PT-investor), the maximization means to choose the optimal payoff in every state of nature x_T at time T.

The first order condition at the state x_T is $v'(x_T - X) = c_{\omega}\theta m$, where θ is the Lagrange multiplier and c_{ω} is the ratio of the real probability of the state x_T to

1. the transformed probability under the probability transformation ω defined by (5), for CPT-investor;
2. the transformed probability under the probability transformation ω defined by (3), for PT-investor.

In this paper we will assume that $\delta = 1$, i.e. transformation ω is the identity operator for both PT- and CPT- investors. Then $c_{\omega} = 1$, $E^{\omega}(\cdot) = E(\cdot)$ and the first order condition at the state x_T is

$$v'(x_T - X) = \theta m, \tag{9}$$

where θ is the Lagrange multiplier. The solution of the problem (8) is

$$x_T = v'^{-1}(\theta m) + X.$$

To solve Eq. (9) we need examine cases $x_T \in (-\infty, X)$ and $x_T \in (X, \infty)$ separately.

1. Let $x_T > X$. Then $v(x_T - X) = (x_T - X)^{\alpha}$ and $x_T = \left(\frac{\theta m}{\alpha}\right)^{\frac{1}{\alpha-1}} + X$.

   Using the budget constraint $W_0 = E(mx_T)$ we can find the Lagrange multiplier θ:

$$W_0 = E\left[m\left(\left(\frac{\theta m}{\alpha}\right)^{\frac{1}{\alpha-1}} + X\right)\right] = \left(\frac{\theta}{\alpha}\right)^{\frac{1}{\alpha-1}} E(m^{1+\frac{1}{\alpha-1}}) + E(mX) =$$

$$\left(\frac{\theta}{\alpha}\right)^{\frac{1}{\alpha-1}} E(m^{1+\frac{1}{\alpha-1}}) + Xe^{-rT},$$

and we get

$$\left(\frac{\theta}{\alpha}\right)^{\frac{1}{\alpha-1}} = \frac{W_0 - Xe^{-rT}}{E(m^{1+\frac{1}{\alpha-1}})}.$$

Consequently,

$$x_T = (W_0 - Xe^{-rT})\frac{m^{\frac{1}{\alpha-1}}}{E(m^{1+\frac{1}{\alpha-1}})} + X. \tag{10}$$

Let us suppose that the portfolio, bond and discount factor follow the classic Black-Scholes environment:

$$\frac{dS}{S} = \mu dt + \sigma dz, \tag{11}$$

$$\frac{dB}{B} = rdt, \tag{12}$$

$$\frac{d\Lambda}{\Lambda} = -rdt - \frac{\mu - r}{\sigma}dz, \tag{13}$$

where S is the price of portfolio, Λ is discount factor, r is risk-free rate, with initial conditions $S(0) = S_0$, $\Lambda(0) = \Lambda_0$, $B(0) = B_0$.

The solutions of (11)–(13) are

$$\ln S_T = \ln S_0 + \left(\mu - \frac{\sigma^2}{2}\right)T + \sigma\sqrt{T}\varepsilon, \tag{14}$$

$$\ln \Lambda_T = \ln \Lambda_0 - \left(r + \frac{1}{2}\left(\frac{\mu - r}{\sigma}\right)^2\right)T - \frac{\mu - r}{\sigma}\sqrt{T}\varepsilon, \tag{15}$$

where $\varepsilon \sim N(0, 1)$, and S_0 is the price of the portfolio at $t = 0$. It follows from (15) that

$$m_T = \frac{\Lambda_T}{\Lambda} = \exp\left[-\left(r + \frac{1}{2}\left(\frac{\mu - r}{\sigma}\right)^2\right)T - \frac{\mu - r}{\sigma}\sqrt{T}\varepsilon\right]$$

and we can obtain

$$E(m_T^{1+\frac{1}{\alpha-1}}) = \exp\left[-\left(1 + \frac{1}{\alpha - 1}\right)\left(r + \frac{1}{2(1-\alpha)}\left(\frac{\mu - r}{\sigma}\right)^2\right)T\right]. \tag{16}$$

Denote $R_T = S_T/S_0$. Then

$$m_T^{-\frac{1}{\alpha-1}} = \exp\left[\frac{1}{1-\alpha}\left(r - \frac{1}{2}\left(\frac{\mu - r}{\sigma}\right)^2 - \frac{\mu - r}{\sigma^2}(r - \sigma^2/2)\right)T + \frac{1}{1-\alpha}\frac{\mu - r}{\sigma^2}\ln R_T\right]. \quad (17)$$

Let $\gamma = \frac{1}{1-\alpha}\frac{\mu-r}{\sigma^2}$. It follows from (10), (16), (17) that

$$x_T = (W_0 - Xe^{-rT})e^{(1-\gamma)(r+\frac{1}{2}\gamma\sigma^2)T}R_T^{\gamma} + X. \quad (18)$$

2. Let $x_T < X$. Then $v(x_T - X) = -\lambda(X - x_T)^{\beta}$ and

$$x_T = X - \left(\frac{\theta m}{\lambda\beta}\right)^{\frac{1}{\beta-1}}.$$

Using the budget constraint $W_0 = E(mx_T)$ we can find the Lagrange multiplier θ:

$$W_0 = E\left[m\left(X - \left(\frac{\theta m}{\lambda\beta}\right)^{\frac{1}{\beta-1}}\right)\right] = E(mX) - \left(\frac{\theta}{\lambda\beta}\right)^{\frac{1}{\beta-1}} E(m^{1+\frac{1}{\beta-1}}) + = Xe^{-rT} - \left(\frac{\theta}{\lambda\beta}\right)^{\frac{1}{\beta-1}} E(m^{1+\frac{1}{\beta-1}}),$$

and we get

$$\left(\frac{\theta}{\lambda\beta}\right)^{\frac{1}{\beta-1}} = \frac{Xe^{-rT} - W_0}{E(m^{1+\frac{1}{\beta-1}})}.$$

Consequently,

$$x_T = X - (Xe^{-rT} - W_0)\frac{m^{\frac{1}{\beta-1}}}{E(m^{1+\frac{1}{\beta-1}})}. \quad (19)$$

We have

$$E(m_T^{1+\frac{1}{\beta-1}}) = \exp\left[-\left(1 + \frac{1}{\beta-1}\right)\left(r + \frac{1}{2(1-\beta)}\left(\frac{\mu - r}{\sigma}\right)^2\right)T\right] \quad (20)$$

and

$$m_T^{-\frac{1}{\beta-1}} = \exp\left[\frac{1}{1-\beta}\left(r - \frac{1}{2}\left(\frac{\mu-r}{\sigma}\right)^2 - \frac{\mu-r}{\sigma^2}(r-\sigma^2/2)\right)T + \frac{1}{1-\beta}\frac{\mu-r}{\sigma^2}\ln R_T\right]. \qquad (21)$$

Let $\nu = \frac{1}{1-\beta}\frac{\mu-r}{\sigma^2}$. It follows from (19)–(21) that

$$x_T = X - (Xe^{-rT} - W_0)e^{(1-\nu)(r+\frac{1}{2}\nu\sigma^2)T}R_T^{\nu}. \qquad (22)$$

We have proved the following proposition.

Theorem 1 *Let stock, bond and discount factor follow* (11)–(13)*. If* $\delta = 1$ *then the solution* x_T *of the problem* (8) *is unique and defined by*

$$x_T = \begin{cases} X - (Xe^{-rT} - W_0)e^{(1-\nu)(r+\frac{1}{2}\nu\sigma^2)T}R_T^{\nu}, \; x_T < X, \\ (W_0 - Xe^{-rT})e^{(1-\gamma)(r+\frac{1}{2}\gamma\sigma^2)T}R_T^{\gamma} + X, \; x_T \geq X, \end{cases}$$

where $R_T = S_T/S_0$, $\nu = \frac{1}{1-\beta}\frac{\mu-r}{\sigma^2}$ *and* $\gamma = \frac{1}{1-\alpha}\frac{\mu-r}{\sigma^2}$.

Equation (18) says that the optimal portfolio consists of two parts: the bond that guarantees the payoff X at time $t = T$, and the wealth $(W_0 - Xe^{-rT})$ invested under power utility maximization.

Corollary 1 *If* $\delta = 1$ *and* $X = W_0e^{rT}$ *then the optimal portfolio is* $x_T = W_0e^{rT}$.

4 Comparison Analysis with EU-Investor

The cumulative prospect theory argues that the investor evaluates the expected deviation of final value of wealth from the reference value X at the time moment T and makes an investment decision on the basis of this assessment. Let $0 \leq \rho \leq 1$. For both PT-investor and CPT-investor we will assume that reference level of wealth at the time moment $t = T$ is $X = W_0(\rho e^{rT} + (1-\rho)R_T)$, i.e. X is the amount of wealth the investor would have received on the date $t = T$ after investing ρW_0 with the continuously compounding rate r and $(1-\rho)W_0$ in the stock. Then the deviation from reference point X on the date $t = T$ is equal to $x_T - X$ (investor compares the portfolio price x_T with X at the moment $t = T$ and if $x_T > X$, he/she considers $x_T - X$ as a gain; in the case $x_T < X$, investor thinks it to be a loss $X - x_T$). The following proposition is a simple corollary of Theorem 1 and describes the optimal payoffs of CPT-investor with the reference point $X = W_0(\rho e^{rT} + (1-\rho)R_T)$.

Corollary 2 *Let* $0 \le \rho \le 1$. *If* $\delta = 1$ *and* $X = W_0(\rho e^{rT} + (1-\rho)R_T)$ *then the return of the optimal portfolio* $\hat{R}$ *is*

$$\hat{R} = \begin{cases} \rho e^{rT} + (1-\rho)R_T - (1-\rho)(R_T e^{-rT} - 1)e^{(1-\nu)(r+\frac{1}{2}\nu\sigma^2)T} R_T^{\nu}, \ R_T > e^{rT}, \\ \rho e^{rT} + (1-\rho)R_T + (1-\rho)(1 - R_T e^{-rT})e^{(1-\gamma)(r+\frac{1}{2}\gamma\sigma^2)T} R_T^{\gamma}, \ R_T \le e^{rT}. \end{cases}$$

On the other hand, EU-investor maximizes expected utility over terminal consumption $E(u(x_T))$, subject to budget constraint on initial wealth W_0, x_T is the payoff of his/her portfolio at terminal time T:

$$E_0(u(x_T)) \to \max_{\{x_T\}}, \quad s.t. \ E_0(m x_T) = w, \tag{23}$$

where $\max_{\{x_T\}}$ means that investor should find the optimal payoff in every state of nature. Given the power utility

$$u(x) = \frac{x^{1-\phi}}{1-\phi}, \tag{24}$$

the return on the optimal portfolio is $\hat{R} = m^{-\frac{1}{\phi}} / E(m^{1-\frac{1}{\phi}})$ (see [4]). It was shown in [4] that if the stock, bond and discount factor follow (11)–(13) then the return of the payoff of optimal portfolio is $\hat{R} = e^{(1-\zeta)(r+\frac{1}{2}\zeta\sigma^2)T} R_T^{\zeta}$, where $R_T = S_T/S_0$ denotes the stock return, and $\zeta = \frac{1}{\phi}\frac{\mu - r}{\sigma^2}$.

Thus, the optimal payoff of EU-investor with power utility (24) is the power function of the stock return. Figure 1 shows the optimal payoffs for EU-investor using quite realistic $\mu = 0.09$, $r = 0.01$, $\sigma = 0.16$ and for two different $\phi = 0.9$ and $\phi = 8$ (using two different values of parameter ϕ we would like to illustrate two different types of risk aversion).

Moreover, Figs. 1 and 2 plot the optimal payoffs for CPT-investor with $\rho = 0$ and $\rho = 0.5$ respectively using the same values of μ, r, σ and for two different sets of $\alpha = 0.7$, $\beta = 0.8$ and $\alpha = 0.1$, $\beta = 0.85$. The first set of parameters is close to the values obtained in [10].

If α and β are approaching to 1 then the optimal payoffs for CPT-investor with $\rho = 0$ is converging to R_T, so the optimal strategy is to hold the stock.

For all $0 < \alpha, \beta < 1$ the payoff of CPT-investor is bigger than $X = W_0(\rho e^{rT} + (1-\rho)R_T)$ if $0 \le R_T < e^{rT}$, and is less than X otherwise. In other words, CPT-investor accepts lower payoffs in the good states of nature ($R_T > e^{rT}$) in order to get slightly better payoffs in the bad states of nature on the left ($R_T < e^{rT}$).

For both EU- and CPT-investors, the optimal payoffs are nonlinear. Figures 1 and 2 show that both types of investors buy a complex set of contingent claims and they trade dynamically. To get the payoffs shown in Fig. 1, one could buy a set of options. For example, the payoff of EU-investor with $\phi = 0.9$ can be approximated by buying a set of call options (or by holding the stock and writing put options). The left side of the PT-investor's payoffs can be replicated by writing call options (or by holding the stock and buying put options).

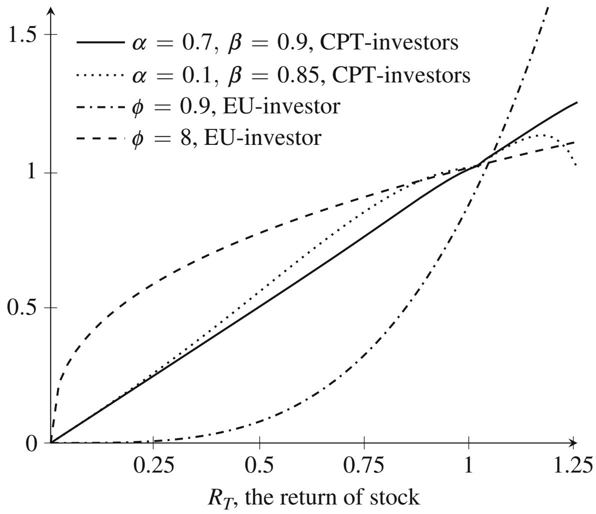

Fig. 1 The dependance of R_T on $\hat{R}$ for EU- and CPT-investors with $\rho = 0$ and different parameters

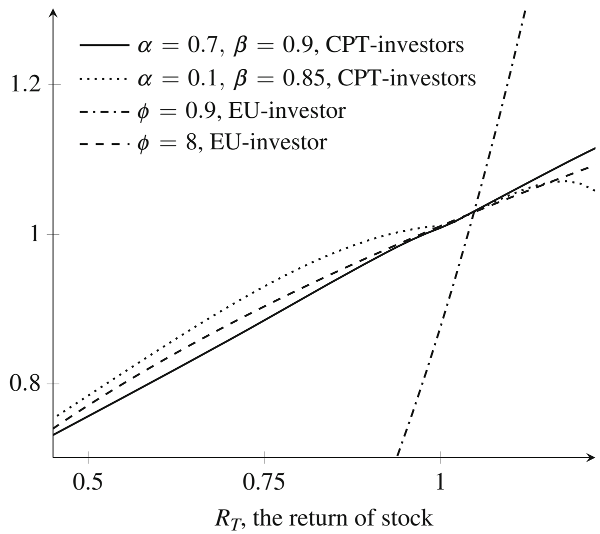

Fig. 2 The dependance of R_T on $\hat{R}$ for EU- and CPT-investors and different parameters, $\rho = 0.5$

5 Conclusion and Future Work

In the classical portfolio theory, a portfolio is defined to be a set of assets with weights, the sum of which is equal to 1 (the budget constraint) and the problem is to find the weights of the specific assets in portfolio to maximize the expected utility (rather than final payoff). Therefore, the next step in implementation of the results of this paper is to find optimal portfolios based on characterizations of the optimal payoffs. In the case of complete market Corollary 2 describes the number of contingent claims to every state of nature at time T that the CPT-investor should buy.

We remark that the transformation of the optimal payoffs into the trading strategy is mainly a technical problem.

Our main assumption is that the market is complete. However, the real world is different. It is well-known that if the market is not complete, then there are (infinitely) many discount factors. Moreover, we should be sure that payoffs are in the space of all available payoffs. Therefore, it should be noted that the problem of extending in some way the results for complete market to the case of incomplete ones can be addressed in future work.

This paper examines the one-period problem. Another problem we would like to address is the problem of extending the results to dynamic setup with long-lived investors and time-varying moments of assets returns. For example, it would be interesting to find the solution of the optimal dynamic portfolio problem for PT-investor in the lognormal iid setup with infinite horizon.

Acknowledgements This work was financially supported by the Russian Fund for Basic Research under Grant 16-01-00507.

References

1. Barberis, N.C.: Thirty years of prospect theory in economics: a review and assessment. J. Econ. Perspect. **27**(1), 173–196 (2013)
2. Barberis, N.C., Huang, M.: Stocks as lotteries: the implications of probability weighting for security prices. Am. Econ. Rev. **98**, 2066–2100 (2008)
3. Bernard, C., Ghossoub, M.: Static portfolio choice under cumulative prospect theory. Math. Finan. Econ. **2**, 277–306 (2010)
4. Cochrane, J.H.: Asset Pricing. Princeton University Press, Princeton (2005)
5. Eeckhoudt, L., Gollier, C., Schlesinger, H.: Economic and Financial Decisions Under Risk. Princeton University Press, Princeton (2005)
6. Gomes, F.J.: Portfolio choice and trading volume with loss-averse investors. J. Bus. **78**, 675–706 (2005)
7. Grishina, N., Lucas, C., Date, P.: Prospect theory based portfolio optimisation: an empirical study and analysis. Quant. Finan. **17**, 353–367 (2017). https://doi.org/doi:10.1080/1497688.2016.1149611
8. He, X.D., Zhou, X.Y.: Portfolio choice under cumulative prospect theory: an analytical treatment. Manag. Sci. **57**(2), 315–331 (2011)
9. Hitaj, A., Mastrogiacomo, E.: Portfolio Choice Under Cumulative Prospect Theory: Sensitivity Analysis and an Empirical Study. Working Paper (2015). http://ssrn.com/abstract=2598186
10. Kahneman, D., Tversky, A.: Prospect theory: an analysis of decision under risk. Econometrica **62**, 1291–1326 (1979)
11. Levy, H., Levy, M.: Prospect theory and mean-variance analysis. Rev. Financ. Stud. **17**, 1015–1041 (2004)
12. Nardon, M., Pianca, P.: A behavioural approach to the pricing of European options. In: Mathematical and Statistical Methods for Actuarial Sciences and Finance, pp. 219–230. Springer, Cham (2014)
13. von Neumann, J., Morgenstern, O.: Theory of Games and Economic Behavior. Princeton University Press, Princeton (1944)
14. Pirvu, T.A., Schulze, K.: Multi-stock portfolio optimization under prospect theory. Math. Finan. Econ. **6**(4), 337–362 (2012)

15. Shi, Y., Cui, X., Li, D.: Discrete-time behavioral portfolio selection under cumulative prospect theory. J. Econ. Dyn. Control **61**, 283–302 (2015)
16. Sidorov, S.P., Homchenko, A., Barabash, V.: Stochastic models for assets allocation under the framework of prospect and cumulative prospect theory. In: Lecture Notes in Engineering and Computer Science: Proceedings of The World Congress on Engineering 2015 (WCE 2015), 1–3 July, 2015, London, pp. 704–709 (2015)
17. Tversky, A., Kahneman D.: Advances in prospect theory: cumulative representation of uncertainty. J. Risk Uncertain. **5**, 297–323 (1992)
18. Zakamouline, V., Koekebakker, S.: A generalisation of the mean-variance analysis. Eur. Financ. Manag. **15**, 934–970 (2009)