Marco Corazza
Cira Perna
Claudio Pizzi
Marilena Sibillo
Editors

Mathematical and Statistical Methods for Actuarial Sciences and Finance

MAF

2022

Springer

Mathematical and Statistical Methods for Actuarial Sciences and Finance

Marco Corazza · Cira Perna ·
Claudio Pizzi · Marilena Sibillo
Editors

Mathematical and Statistical Methods for Actuarial Sciences and Finance

MAF 2022

Springer

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

Claudio Pizzi
Department of Economics
Ca' Foscari University of Venice
Venice, Italy

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

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

Preface

This volume is a collection of papers selected and peer reviewed from the more than 100 presented at the International Conference on Mathematical and Statistical Methods for Actuarial Sciences and Finance–MAF2022, held at the University of Salerno from 20 to 22 April 2022.

In its organizational phase, the course of the COVID-19 pandemic was still unpredictable, and the MAF2022 steering committee made the decision to hold the event in a hybrid form, online or in-person, leaving each participant free to choose the most appropriate mode of participation. Nevertheless, we have always hoped to have the widest possible participation in presence, both as a desired sign of normality and as a return to the tradition of the conference, always characterized by cultural and human exchanges that only in presence can be fully realized.

This year's conference, organized by the Department of Economics and Statistics of the University of Salerno with the collaboration of the Department of Economics of the University of Venice Cà Foscari, is the tenth in a two-year series that began in 2004.

It was in fact in 2003 that the mathematicians and statisticians of the Department of Economics and Statistics of the University of Salerno, colleagues and friends among them, conceived and grew the purpose of developing through scientific meetings the cooperation and exchange of ideas among those who, like them, were engaged in research in actuarial science and finance. The enthusiasm about the initiative was always based on the deep conviction that this interaction would surely bear good fruit.

And so, the initiative has followed regularly, availing since 2008 of the valuable collaboration of the Department of Economics of the University of Venice Cà Foscari.

The first six editions were held in Italy, namely in 2004 and 2006 in Salerno, in 2008 in Venice, in 2010 in Ravello (Salerno), in 2012 again in Venice and in 2014 in Vietri sul Mare (Salerno). The international dimension of the conference has grown over time, attracting a wider and wider audience. Thus, in 2016 the MAF was held in Paris and in 2018 in Madrid. The 2020 edition, already suffering from the COVID-19 pandemic, was held in a fully online version from Venice.

This tenth edition confirms the growing interest of the international scientific community towards the initiative, with about 200 participants, more than 170 scientific contributions proposed in the form of abstracts or papers and four prestigious plenary speakers, namely

Prof. Elsa Fornero, Honorary Professor, University of Turin, who presents an invited talk entitled: "Reform, Inform, Educate": a new paradigm for the sustainability of pension system;
Prof. Massimiliano Caporin, University of Padua, who presents an invited talk entitled: Realized Covariance Modelling, Forecast Error Variance Decompositions and a Model-Based Diebold-Yilmaz Index;
Prof. Marcello Galeotti, University of Florence, who presents an invited talk entitled: Applications of Game Theory to Risk Models: Evolutionary and Cooperative Approaches;
Dr. Michel Dacorogna, Prime Re Solutions, Zug, Switzerland, who presents an invited talk entitled: Pro-Cyclicality Beyond Business Cycles: The Case of Traditional Risk Measurements.

Since 2006, all editions of the conference have been accompanied by a book published by Springer, a product that has often been counted among the most downloaded on the platform. Also, this tenth edition proposes the associated book, with the aim of offering the selected scientific contributions in a concise form of maximum 6 pages, in which the authors present their idea and the methodology behind its development, providing, when possible, an illustrative application.

The goal is to create a forum for comparison of ideas, topics and research perspectives, which embodies and represents at best the soul of MAF as a place of meeting and scientific exchange.

Several are the research areas to which the papers are dedicated with a focus on applicability and/or applications of the results:

Actuarial models, analysis of high-frequency financial data, behavioural finance, carbon and green finance, credit risk methods and models, dynamic optimization in finance, financial econometrics, forecasting of dynamical actuarial and financial phenomena, fund performance evaluation, insurance portfolio risk analysis, interest rate models, longevity risk, machine learning and soft computing in finance, management in insurance business, models and methods for financial time series analysis, models for financial derivatives, multivariate techniques for financial markets analysis, neural networks in insurance, optimization in insurance, pricing, probability in actuarial sciences, insurance and finance, real-world finance, risk management, solvency analysis, sovereign risk, static and dynamic portfolio selection and management, trading systems.

In its almost twenty years, the initiative has always availed itself of the support of the Departments of Economics and Statistics of the University of Salerno (Italy) and of the Department of Economics of the University Ca' Foscari of Venice (Italy) and nonetheless of the scientific associations:

- AMASES—Association for Mathematics Applied to Social and Economic Sciences
- SIS—Italian Statistical Society.

Further, we would also like to express our deep gratitude to the members of the scientific and organizing committees and to all the people whose collaboration contributed to the success of the MAF2022 conference. In particular, our heartfelt thanks go to Giovanna Bimonte and Antonio Naimoli, who have worked unstintingly with great enthusiasm and efficiency, continually showing with their work the sharing of the aims of the initiative. We would also like to thank all the participants for their precious and indispensable contribution.

Finally, we are pleased to inform you that the organizational machine is already at work, looking forward to the MAF2024 edition.

April 2022

Marco Corazza
Cira Perna
Claudio Pizzi
Marilena Sibillo

Contents

Absolute and Relative Gender Gap in Pensions: The Impact of the Transition from DB to NDC in Italy

Antonio Abatemarco[1]([✉]) [iD] and Maria Russolillo[2] [iD]

[1] Department of Economics and Statistics and Centre for Economic and Labour Policy, Evaluation (CELPE), University of Salerno, Via Giovanni Paolo II, 132, 84084 Fisciano, SA, Italy
aabatemarco@unisa.it

[2] Department of Economics and Statistics, University of Salerno, Via Giovanni Paolo II, 132, 84084 Fisciano, SA, Italy
mrussolillo@unisa.it

Abstract. In this paper, we analyze how the progressive transition from the DB to the NDC scheme has affected both the absolute and the relative gender gap at retirement in Italy for individuals retiring from 1980 to 2027 by using data from SHARELIFE (Wave 7). With this purpose in mind, the two opposite effects originating respectively from (i) improving labor market conditions for women and (ii) increasing actuarial fairness of the pension plan are simultaneously considered. We observe a U-shaped pattern since the gender gap in pensions is found to be decreasing up to 2020 but increasing afterward. By using both absolute and relative gender-gap indicators, we show that the increasing pattern for the gender gap at retirement after 2020 is driven by (i) decreasing redistributive impact of the pension scheme, and (ii) women's penalization in the pro-rata mechanism due to lower contributions paid in the early working life. Specifically, due to the transition from a very generous to an actuarially fair pension scheme, gender disparities are found to be slightly different when comparing absolute and relative indicators.

Keywords: Gender gap · Pension · Redistribution · Actuarial fairness

1 Introduction

The Gender Gap in Pensions (GGP) indicates the percentage by which women's average pension is lower than men's. The GGP is observed in all countries with different magnitude and reveals the weaker pension position many women find themselves compared to men.

During the last decades, a decreasing trend has been observed for the GGP in most of the European countries, which is still going on in the recent times. As for the population

aged more than 65 years old in EU-SILC, the GGP is found, on average, 39.1% in 2009 [3], whereas the same indicator for the population of retirees aged 65–79 years old in 2017 is found sensibly lower at 35.7% [5]. Nevertheless, the transition from a Defined-Benefit (DB) and redistributive pension plan to a Defined-Contribution (DC) and actuarially fair scheme, foreseen by many pension reforms, is expected to increase this gap by introducing a stronger link between earnings and pensions [7], especially in Italy.

Basically, during the last decades two driving forces impacting on the GGP have been observed in Italy: on the one hand, since the Sixties, significant improvement of labor market conditions for women in terms of both average labor market earnings and participation in the labor market, have narrowed the gap [8]. On the other hand, pension reforms introduced from the Nineties – promoting the transition from a redistributive DB to an actuarially fair NDC scheme – have reduced generic rich-to-poor redistribution in old-age and seniority pensions [1], and are then expected to open the GGP [6]. The present work is tailored to study, by using a counterfactual approach (in the absence of behavioural responses), the dynamics of the gender gap at retirement by cohorts and by year of retirement, while comparing absolute and relative indicators in order to capture possible gender disparities.

2   Data and Methodology

This paper uses data from SHARE (The Survey of Health, Aging and Retirement in Europe) Wave 7 (https://doi.org/10.6103/SHARE.w7.711), see [4] for methodological details.[1]

2.1   Data

The study focuses on a population of individuals (employees and self-employed) with a public pension plan, born from 1940 up to 1969, and retired from 1980 up to 2016 (retrospective analysis), or expected to meet minimum eligibility requirements within 2027 (prospective analysis). Data are taken from SHARELIFE [4] questionnaire (wave 7) and are collected by taking into account, for each job position lasting six months at least, the following features: employment spells, employment status, job characteristics, income, retirement benefits, typology of contribution plans, and the type of public pension benefits. To focus on the insurance mechanism of the pension scheme (independently from

social assistance), we consider the sole respondents receiving an old-age or seniority benefit from contributions to a public pension plan (so excluding first-tier pension benefits). The analysis here carried out considers real data for the reconstruction of the contribution career of retirees (last chronological year 2016) divided by three consecutive intervals for the year of birth (cohort), respectively 1940–49, 1950–59, and 1960–69, and three consecutive intervals for the year of retirement, respectively 1980–09, 2010–19, and 2020–27 and it is based on the following assumptions.

(a) Salary progression within each job spell is simulated according to the earnings growth rate differentiated by macro sector from official reports of the Italian National Institute of Statistics (ISTAT).

(b) Gross earnings are obtained from net earnings applying tax brackets and marginal tax rates from three different tax systems (1974, 1992, 2007).

(c) Transformation Coefficients for the bienniums 2023–24, 2025–26 and 2027 are obtained by projecting mortality rates according to the Lee-Carter model [9, 10] by using data from the Human Mortality Database [2].

(d) As far as our analysis is of the retrospective kind for people retiring before 2017 and prospective with respect to still-in-job individuals who are expected to retire by 31st December 2027, we assume that still-in-job individuals in 2016 preserve their working conditions ad retire as soon as they reach minimum eligibility requirements according to pension rules in 2022.

(e) Since official capitalization rates to be applied in the NDC scheme are not available before 1962 and after 2021, we reconstruct the missing values by using the 5 yearly moving average of the GDP growth rate (ISTAT).

2.2 Methodology

Given the information on the working career of individuals retired from 1980 to 2016, as well as on the population of still-in-job individuals expected to retire within 2027, let $b_i := \{b_{1i}, \ldots, b_{n_i i}\} \in \mathfrak{R}^{n_i}_+$ be the *actual* distribution of first pension benefits in the i_{th} cell and $b_i^v := \{b_{1i}^v, \ldots, b_{n_i i}^v\} \in \mathfrak{R}^{n_i}_+$, be the *virtual* distribution of first pension benefits calculated under the hypothesis of a fully NDC (actuarially fair) scheme.

Following the existing literature on the gender gap [3, 5], the actual and the virtual GGPs in the i_{th} cell are, respectively,

$$GGP_i^a = 1 - \frac{\overline{b}_i^a(F)}{\overline{b}_i^a(M)}$$

$$GGP_i^v = 1 - \frac{\overline{b}_i^v(F)}{\overline{b}_i^v(M)}$$

$$(1)$$

with $\overline{b}_i(F)$ and $\overline{b}_i(M)$ indicating the average pension benefit in the i_{th} cell of females and males respectively. Notably, the money (absolute) gap between average pension benefits is taken in relative terms with respect to the average pension of males.

3   Preliminary Results

In the analysis we carry on, following the reforms of the Italian pension system, in particular Dini (Law 335/1995) and Monti-Fornero (Decree-Law 201/2011) Reform, we consider three different cells. Cell 1 consists of workers with 18 years of contribution at least in 1995 and retiring before 2012, for whom a fully DB benefit formula applies in the computation of the first pension. Cell 2 consists of workers with 18 years of contribution at least in 1995 but retiring after 2012; for these workers the old DB scheme applies for years of contribution until 2011, whereas the NDC scheme applies afterward. Finally, Cell 3 is composed by workers with less than 18 years of contribution in 1995, for whom the old DB scheme applies until 1995 whereas the NDC scheme applies afterward.

The dynamics of the composition of the population in each cell with respect to the pension scheme in force is as follows: in Cell 1 the 98.2% of retirees obtain a fully DB pension; in Cell 2, only the 10.5% of retirees obtain a fully DB pension, while NDC from 2012 for 80.7% of retirees applies; finally, in Cell 3 NDC applies from 1996 for 85.8% of retirees.

The actual and the virtual GGPs calculated according to formula (1) are illustrated in Table 1.

Table 1. Gender Gap in Pensions by cohort and by year of retirement

Gender Gap	Cell 1*	Cell 2*	Cell 3*
GGP^a	0.289	0.183	0.190
GGP^v	0.318	0.221	0.171

*Cell 1: cohort 1940–49, retirement 1980–09. Cell 2: cohort 1950–59, retirement 2010–19. Cell 3: cohort 1960–69, retirement 2020–27.

From Table 1, it appears that the GGP^a reduces significantly, in percentage terms, with the transition from Cell 1 to Cell 2. This virtuous gender gap reduction trend interrupts with the transition from Cell 2 to Cell 3, thus determining a U-shaped pattern for the GGP^a. This trend clearly highlights the redistributive impact of the DB pension scheme in Cell 1 and 2, compared to the actuarially fair NDC scheme, oddly absent in Cell 3. The second raw in Table 1 shows on the contrary a decreasing pattern for the GGP^v for all the considered cohorts. While this result is expected, we cannot say the same for the first one. In order to better explain the U-shaped GGP^a, in Table 2 we report the actual mean pension benefit for male and female and the Absolute Actual Gender Gap in Pensions (Abs. GGP^a) by cohort and year of retirement.

As concerns the dynamics of the actual pension benefits across cells, it is worth observing that, for both males and females, actual pension benefits are first increasing from Cell 1 to Cell 2, then decreasing from Cell 2 to Cell 3. This is mostly the result of the two driving forces above described (Sect. 1) moving in opposite directions, where the first effect due to increasing labor market earnings from the eighties dominates from Cell 1 to Cell 2, whereas the lower rate of return of the NDC scheme dominates from Cell 2 to Cell 3, causing the reduction of mean pension benefits for both males and

Table 2. Absolute Actual Gender Gap in Pensions by cohort and by year of retirement

Gender Gap (€)	Cell 1*	Cell 2*	Cell 3*
$\overline{b}_i^{\,a}(M)$	19878.2	25724.0	22859.8
$\overline{b}_i^{\,a}(F)$	14123.2	21014.7	18489.2
Abs.GGP^a	5755.0	4709.3	4370.6

*Cell 1: cohort 1940–49, retirement 1980–09. Cell 2: cohort 1950–59, retirement 2010–19. Cell 3: cohort 1960–69, retirement 2020–27.

females. With the progressive transition from DB to NDC scheme the GGP^a has started to growing again in percentage terms, while slowing down sharply in absolute terms.

From these results it might appear that women would make better with a fully NDC in Cell 3. This might be due to the Pro-rata mechanism insidious for late entering in the labor market and discontinuous early working career of the women compared to men. To explore this result further, in Table 3, the actual and virtual women's average pension benefit by cohort and year of retirement are compared.

Table 3. Women's average pension benefit by cohort and by year of retirement

Gender Gap (€)	Cell 1*	Cell 2*	Cell 3*
$\overline{b}_i^{\,a}(F)$	14123.2	21014.7	18489.2
$\overline{b}_i^{\,v}(F)$	11634.9	16091.6	17545.3

*Cell 1: cohort 1940–49, retirement 1980–09. Cell 2: cohort 1950–59, retirement 2010–19. Cell 3: cohort 1960–69, retirement 2020–27.

From Table 3, the gross actual pension benefit for females is found to be higher than the virtual one for all of the three Cells, even if the gap is sensibly lower in Cell 3. Hence, even if women would do better in terms of GGP under a fully NDC scheme, in absolute terms the pro-rata mechanism is still to be preferred to the fully NDC due to a more generous pension benefit formula.

4   Remarks

Our analysis proves that the progressive transition from the DB to the NDC scheme has stopped the decreasing pattern of the gender gap in pensions in Italy (in terms of old-age and seniority pensions), even if a decreasing pattern is still observed in money (absolute) terms. To our knowledge, this is the first analysis investigating the dynamics of the gender gap in pensions by cohort and by year of retirement, which is crucial to capture the effect of pension reforms over time. Future research efforts will be devoted to the identification of alternative methodologies to be used to investigate the robustness of the results obtained from the counterfactual GGP analysis implemented in our analysis.

References

1. Abatemarco, A.: On the measurement of intra-generational lifetime redistribution in pension systems. CERP Working Paper Series, vol. 55 (2006)
2. Berkeley Human Mortality Database. University of California, Berkeley (USA), and Max Planck institute for demographic research (Germany). www.mortality.org/ or www.humanmortality.de. Accessed Dec 2020
3. Bettio, F, Tinios, P, Betti, G.: The Gender Gap in Pensions in the EU. European Commission, publications office of the European Union, Luxembourg (2013)
4. Börsch-Supan, A., et al.: Data resource profile: the survey of health, ageing and retirement in Europe (SHARE). Int. J. Epidemiol. **150**, 549–568 (2013)
5. Dessimirova, D., Bustamante, M.A.: The gender gap in pensions in the EU. Policy Department for Economic, Scientific and Quality of Life Policies (2019)
6. Leombruni, R., Mosca, M.: Le système de retraite italien compense-t-il les inégalités hommes-femmes sur le marché du travail? Retraite et Société **2**(63), 139–163 (2012)
7. Mazzaferro, C.: The transition to NDC in Italy: assessing distributive and financial effects. WorkINPS Papers, vol. 25 (2019)
8. Mussida, C., Picchio, M.: The trend over time of the gender wage gap in Italy, Empir. Econ. **46**(3), 1081–1110 (2014)
9. Lee, R.D., Carter, L.R.: Modelling and forecasting U.S. Mortality. J. Am. Statist. Assoc. **87**, 659–671 (1992)
10. Russolillo, M., Giordano, G., Haberman, S.: Extending the Lee-Carter model: a three-way decomposition. Scand. Actuar. J. **2**, 96–117 (2011)

TPPI: Textual Political Polarity Indices.
The Case of Italian GDP

Alessandra Amendola[1($\boxtimes$)], Walter Distaso[2], and Alessandro Grimaldi[1]

[1] University of Salerno, Salerno, Italy
{alamendola,algrimaldi}@unisa.it
[2] Imperial College Business School, London, UK
w.distaso@imperial.ac.uk

Abstract. In this work, we propose a data-driven approach to derive a Textual Political Polarity Index (*TPPI*) based on the verbatim reports of the Italian "*Senate of the Republic*". Our procedure allows us to build a set of polarity indices reflecting the impact of political debate and (dis)agreement within parties' groups on a chosen economic variable - the Italian GDP growth rate - over time. Results point to a nontrivial predictive power of the proposed indices, which (importantly) do not rely on a subjective choice of an affective lexicon.

Keywords: NLP · Sentiment analysis · Text as data · Parliamentary debate · Time series

1 Introduction

Recently the interest in "*text as data*" among scholars of political science, economics, and finance has dramatically increased [2], due to the advantages of textual analysis in terms of costs [11] and timeliness [7].

In this work we propose new text-based indices to improve predictions of traditional economic indicators [7] using parliamentary debates [4,9]. Texts *sentiment polarities* are derived without relying on a subjective choice of an *affective lexicon* [8] in a completely data driven manner, following the intuition of [6].

2 Data

We work with two different sets of data: a corpus of parliamentary debates and the time series of a macroeconomic variable.

Let $t = 1, \ldots, T = 73$ be the sample (S) years corresponding to the calendar years 1948 to 2020. We use years $t = 1, \ldots, M = 43$ as *training sample* (S_1) for words polarities computation and years $t = M + 1, \ldots, T$ as *test sample* (S_2) for indices time series construction and evaluation.

2.1   The Italian Senate Verbatim Reports

The corpus we use is the collection of the *Italian Senate of the Republic* parliamentary verbatim reports. It consists of more than 11,500 transcripts from 08 May 1948 to 31 December 2020 which we convert from raw *.pdf* format into a *Document Term Matrix* (DTM).

In order to do so, we perform optical character recognition (OCR) when needed, then apply standard pre-processing and cleaning [1], *stop-words* and rare terms removal [4,9] and *stemming* to reduce vocabulary size.

After splitting the texts into speeches, we filter out orators identified by their office only [3].

Then we associate each orator speech to a set of metadata used to identify four *Political Groups* each orator belonged to the day they spoke. Namely, the *Government* group includes all Government members. *Majority* and *Opposition* group members either support the Government or not, respectively. We disregard the residual *Mixed* group as it aggregates the (few) orators not belonging to any previous group.

We call $\boldsymbol{C} = (c_{d_t,v})$ the original $D \times V$ DTM where: $D = \sum_{t=1}^{M} D_t$ is the *total number of documents in* S_1; D_t is the *number of documents at each time* t; V is the *total number of words in* S_1; $c_{d_t,v} \geq 0$ is the *count of word* v *in document* d_t *at time* t; $d_t = 1, \ldots, D_t$; $t = 1, \ldots, M$; $v = 1, \ldots, V$.

To align the sample frequencies of corpus texts and economic variable we create the $M \times V$ matrix $\widetilde{\boldsymbol{C}}$ whose elements are given by $\widetilde{c}_{t,v} = \sum_{d_t=1}^{D_t} c_{d_t,v}$.

In order to put more weight on the most relevant words at each time, we apply the *tf-idf* transformation[1] on $\widetilde{\boldsymbol{C}}$ and obtain the $M \times V$ matrix $\boldsymbol{F} = (f_{t,v})$ where $f_{t,v}$ is the *tf-idf weighted frequency of word* v *in document* t. Finally, such frequencies are normalised as in [10] via the *L2-norm* so to obtain the elements $\widetilde{f}_{t,v}$ with $t = 1, \ldots, M$ of the matrix $\widetilde{\boldsymbol{F}}$.

2.2   The Italian Yearly GDP Time Series

The macroeconomic variable we use is the Italian gross domestic product (GDP). To cover our sample period we create a custom time series from two already available series. Specifically, we used the 1948–1995 values of the GDP time series re-constructed by the Italian institute of statistics (*ISTAT*) and *Bank of Italy* and the 2017–2020 values of the GDP time series as made available by *ISTAT* in its National Accounts database. For years 1995–2017 a weighted average of the two series is calculated at each time. Both series are yearly based, measured in millions of euros at current market prices, and constructed by the original sources in accordance with the *European National Accounts 2010* (ESA 2010). The series obtained is converted in a fixed 2000 market prices series using specific multipliers made available by *ISTAT*.

---

[1] Let x be a word in a corpus, its *term frequency-inverse document frequency* is given by: $tf\text{-}idf(x) = tf(x) \times idf(x)$ where $tf(x)$ is the *relative frequency of* x and $idf(x) = \log \frac{total\ number\ of\ documents}{number\ of\ documents\ containing\ x}$ is the *inverse document frequency of* x.

Let $\{y_t\}_{t=0}^{M}$ be the time series of interest and let $\boldsymbol{y} = (y_0, \ldots, y_M)$ be the $(M+1) \times 1$ vector associated with it. As we are mostly interested in the economy temporal dynamic, and to have a stationary series, we compute the *year-on-year* growth rates series $\{\varDelta y_t\}_{t=1}^{M}$, where $\varDelta y_t = \frac{y_t - y_{t-1}}{y_{t-1}}$.

3   Determining Words Sentiment Polarities

To determine the words sentiment polarities, each normalised weighted word frequency $\widetilde{f}_{t,v}$ is assigned the corresponding growth rates at each time $t = 1, \ldots, M$.

To do so, we need to choose a "*textual window*" - i.e. the portion of texts to be used in calculations. Let $B_{k,h}$ be the generic "*textual window configuration*" where h and k are, respectively, its starting and ending time points. Textual windows control for possible delayed impacts of Parliamentary discussion on the Country economy.

Assigning the growth rates signs results in the construction of the "*polarised textual windows*" matrices $\boldsymbol{sgnW}_t^{(*)}$ where $* = \{B_{0,0}, \ldots, B_{1,4}\}$.

Finally, we obtain each "*word sentiment polarity*", $\widetilde{p}_v^{(y)}$, by summing $\boldsymbol{sgnW}_t$ elements across rows and re-scaling the computed values.

4   Polarity Indices Time Series

After determining each word polarity on the training sample S_1, we construct time series of polarity indices on the test sample S_2.

4.1   Total Textual Political Polarity Index (*TPPI-T*)

To construct the polarity indices time series we start from the (test sample) normalised *tf-idf* weighted frequencies $\widetilde{f}_{t,v}$ and use them to obtain the "*polarised words frequencies*" $g_{t,v}^{(y)} = \widetilde{f}_{t,v} \times \widetilde{p}_v^{(y)}$ where $t = M+1, \ldots, T$ and $v = 1, \ldots, V$.

We create time series of *positive*, *negative* and *total* polarities whose elements are, respectively, given by: $pos_t^{(y)} = \sum_{v=1}^{V} \left(g_{t,v}^{(y)} \mid g_{t,v}^{(y)} > 0 \right)$ while $neg_t^{(y)} = \sum_{v=1}^{V} \left(g_{t,v}^{(y)} \mid g_{t,v}^{(y)} < 0 \right)$ and $tot_t^{(y)} = \sum_{v=1}^{V} |g_{t,v}^{(y)}|$.

Finally, we construct the "*Total Textual Political Polarity Index*" (*TPPI-T*) time series whose elements are given by:

$$TPPI\text{-}T_t^{(y)} = \frac{pos_t^{(y)} + neg_t^{(y)}}{tot_t^{(y)}} \qquad t = M+1, \ldots, T \qquad (1)$$

We call such index "*Total*" as it is constructed without conditioning on the political group belonging of each orator at the time they were speaking.

4.2   Group Specific Textual Political Polarity Indices (*TPPI-GS*)

With the procedure previously detailed, we construct analogous indices conditioning on orators' political group: *Government*, *Majority* and *Opposition*.

Therefore, the following *three* "*Group Specific Textual Political Polarity Indices*", *TPPI-GS*, are built - one for each political group:

$$TPPI\text{-}GS_{i,t}^{(y)} \qquad i = \{Gov, Maj, Opp\} \qquad t = \{M+1, \dots, T\} \qquad (2)$$

4.3   Polarity Divergence Indices (*TPPI-D*)

The *TPPI-T* and the *TPPI-GS* aim at modelling the *sentiment polarity* of debate during Assembly sessions over time.

Another interesting feature is the *degree of (dis)agreement* between one political group and the others. Therefore we also propose "*Group Polarity Divergence Indices*", *TPPI-D*, derived directly from the *TPPI-GS* basing on the squared differences between each of the three *TPPI-GS* at each time.

Let *TPPI-GS*$_{i,t}^{(y)}$ be the t-th time series point of the i-th *TPPI-GS* with $i = \{Gov, Maj, Opp\}$. The t-th time series point of the ij-th *TPPI-D* is:

$$TPPI\text{-}D_{ij,t}^{(y)} = \left(TPPI\text{-}GS_{i,t}^{(y)} - TPPI\text{-}GS_{j,t}^{(y)} \right)^2 \qquad (3)$$

where $i, j = \{Gov, Maj, Opp\}$, $i \neq j$ and $t = M+1, \dots, T$.

In addition, we construct an *Average Group Divergence Index*, *TPPI-D$_{Avg}$*, by averaging the values of the *three* polarity divergence series at each time:

$$TPPI\text{-}D_{Avg,t}^{(y)} = \frac{1}{C} \sum_{ij} TPPI\text{-}D_{ij,t}^{(y)} \qquad (4)$$

where $t = M+1, \dots, T$ and C is the *number of possible couples of i and j* with $i, j = \{Gov, Maj, Opp\}$, $i \neq j$. Hence, we obtained *three* *TPPI-D*.

5   Evaluating Indices Configurations

Each index is constructed in 8 versions depending on the textual window configuration. In total we had $N = 32$ "*primary*" indices - i.e. not deriving from other indices: 8 *TPPI-T* and 24 *TPPI-GS*.

To choose the best configurations, we used the model confidence set (MCS) introduced by [5]. As loss function for the procedure, we used the squared error function: $SE_t^{(*)} = \left(y_t - TPPI_t^{(*)} \right)^2$ where $t = M+1, \dots, T$ and $TPPI^{(*)}$ is a *TPPI-T* or *TPPI-GS* under textual window $* = \{B_{0,0}, \dots, B_{1,4}\}$.

We performed *four* MCS - one for each of the indices *TPPI-T*, *TPPI-GS$_i$* with $i = \{Gov, Maj, Opp\}$ - and select the best $N^* = 9$ textual window configurations: $B_{0,0}$ and $B_{0,2}$ for the *TPPI-T*; $B_{0,1}$, $B_{0,0}$ and $B_{1,1}$ for the *TPPI-GS$_{Gov}$*; $B_{0,0}$ and $B_{0,2}$ for the *TPPI-GS$_{Maj}$*; $B_{0,0}$ and $B_{1,1}$ for the *TPPI-GS$_{Opp}$*.
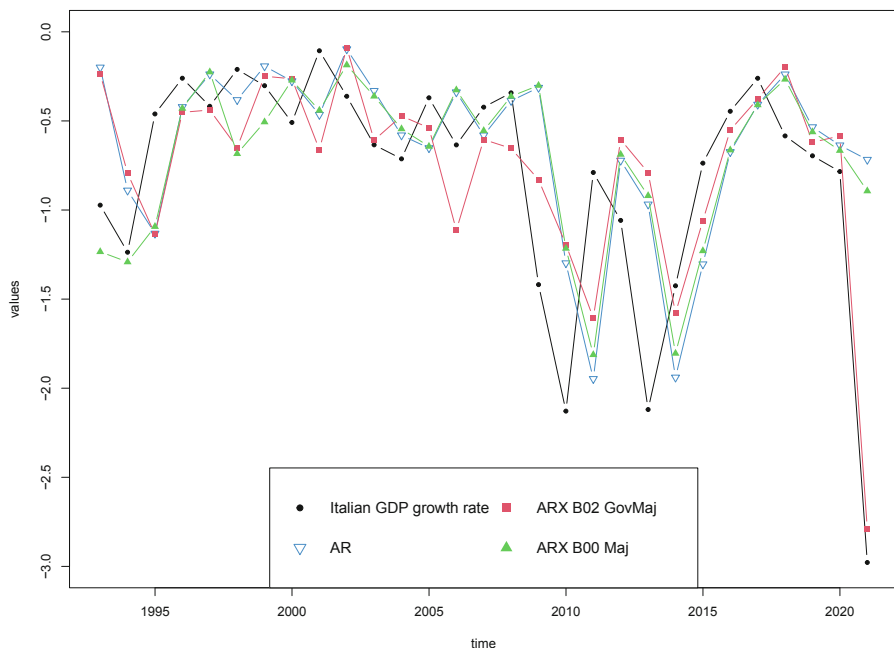
Fig. 1. Annual growth rate of the Italian yearly GDP at 2000 constant prices: best $AR\text{-}X$ models via MCS vs AR model

For the 24 "*secondary*" indices *TPPI-D* the choice is based on the selection of the *TPPI-GS* they derive from. For each best *TPPI-GS* version, the corresponding *TPPI-D* with the same window configuration are selected. Hence, 10 *TPPI-D* versions are selected: $B_{0,0}$, $B_{1,1}$, $B_{0,1}$, $B_{0,2}$ for the *TPPI-D*$_{GovMaj}$; $B_{0,0}$, $B_{1,1}$, $B_{0,1}$ for the *TPPI-D*$_{GovOpp}$; $B_{0,0}$, $B_{1,1}$, $B_{0,2}$ for the *TPPI-D*$_{MajOpp}$. Finally, for the *TPPI-D*$_{Avg}$ only configurations $B_{0,0}$ and $B_{1,1}$ are considered.

To test their ability to improve the forecast accuracy of Italian GDP growth rate, following [7], each of the 21 selected indices is employed as *exogenous variable* (X) in an autoregressive model ($AR\text{-}X$) and compared against the simpler autoregressive model (AR) of order one.

Then, we identify the best performing models via a MCS, hence selecting the $AR\text{-}X$ including the indices $TPPI\text{-}D_{GovMaj}^{(B_{0,2})}$ and $TPPI\text{-}GS_{Maj}^{(B_{0,0})}$ as the two best models with a confidence level of 99% and a mean squared error (MSE) of 0.476 and 0.614, respectively. In all cases $AR\text{-}X$ models outperform the classic AR model providing evidence in favour of the informative content of the constructed indices. Specifically, the $AR\text{-}X$ regressions including the $TPPI\text{-}D_{GovMaj}^{(B_{0,2})}$ and $TPPI\text{-}GS_{Maj}^{(B_{0,0})}$ show an R-squared (R^2) of 0.466 and 0.111, respectively, against an $R^2 = -0.013$ of the AR model. Figure 1 shows the GDP predictions obtained by the two best performing $AR\text{-}X$ models according to the MCS and the benchmark AR model compared against the observed data.

6   Conclusion

In this work, we derive two classes of textual indices linking political debate and economic dynamic. The *TPPI-T* and *TPPI-GS* measure sentiment polarity and aim at replicating directly the economic variable of interest; on the other hand, the *TPPI-D* incorporate information about political (dis)agreement.

Results are achieved by analysing the full corpus of Italian Senate debates and creating a database ready to use for textual analysis.

Importantly, words sentiment polarities are derived via a completely data-driven approach without relying on the choice of an affective lexicon which is a more common practice.

The proposed indices are used as additional predictors in regression models. Results highlight a significant negative linear relation between the dynamic of the Italian GDP growth rate and the tone divergence in the discussion between Government and Majority members.

References

1. Denny, M.J., Spirling, A.: Text preprocessing for unsupervised learning: Why it matters, when it misleads, and what to do about it. Political Anal. **26**(2), 168–189 (2018). https://doi.org/10.1017/pan.2017.44
2. Gentzkow, M., Kelly, B., Taddy, M.: Text as data. J. Econ. Lit. **57**(3), 535–74 (2019). https://doi.org/10.1257/jel.20181020
3. Gentzkow, M., Shapiro, J.M., Taddy, M.: Measuring group differences in high-dimensional choices: method and application to congressional speech. Econometrica **87**(4), 1307–1340 (2019). https://doi.org/10.3982/ECTA16566
4. Grimmer, J.: A bayesian hierarchical topic model for political texts: Measuring expressed agendas in senate press releases. Political Analysis **18**(1), 1–35 (2010). https://doi.org/10.1093/pan/mpp034
5. Hansen, P.R., Lunde, A., Nason, J.M.: The model confidence set. Econometrica **79**(2), 453–497 (2011). https://doi.org/10.3982/ECTA5771
6. Ke, Z., Kelly, B.T., Xiu, D.: Predicting returns with text data. University of Chicago, Becker Friedman Institute for Economics Working Paper No. 2019-69, Yale ICF Working Paper No. 2019-10, Chicago Booth Research Paper No. 20-37 (2021). https://doi.org/10.2139/ssrn.3389884
7. Larsen, V.H., Thorsrud, L.A.: The value of news for economic developments. J. Econ. **210**(1), 203–218 (2019). https://doi.org/10.1016/j.jeconom.2018.11.013
8. Loughran, T., McDonald, B.: When is a liability not a liability? textual analysis, dictionaries, and 10-ks. J. Financ. **66**(1), 35–65 (2011). https://doi.org/10.1111/j.1540-6261.2010.01625.x
9. Quinn, K.M., Monroe, B.L., Colaresi, M., Crespin, M.H., Radev, D.R.: How to analyze political attention with minimal assumptions and costs. Am. J. Political Sci. **54**(1), 209–228 (2010). https://doi.org/10.1111/j.1540-5907.2009.00427.x
10. Sebastiani, F.: Machine learning in automated text categorization. ACM Comput. Surv. **34**(1), 1–47 (2002). https://doi.org/10.1145/505282.505283
11. Shapiro, A.H., Sudhof, M., Wilson, D.J.: Measuring news sentiment. J. Econ. (2020). https://doi.org/10.1016/j.jeconom.2020.07.053

Quantile Regression Forest
for Value-at-Risk Forecasting
Via Mixed-Frequency Data

Mila Andreani[1(✉)], Vincenzo Candila[2], and Lea Petrella[2]

[1] Scuola Normale Superiore, Pisa, Italy
`mila.andreani@sns.it`
[2] MEMOTEF Depart., Sapienza University of Rome, Rome, Italy
{`vincenzo.candila,lea.petrella`}`@uniroma1.it`

Abstract. In this paper we introduce the use of mixed-frequency variables in a quantile regression framework to compute high-frequency conditional quantiles by means of low-frequency variables. We merge the well-known Quantile Regression Forest algorithm and the recently proposed Mixed-Data-Sampling model to build a comprehensive methodology to jointly model complexity, non-linearity and mixed-frequencies. Due to the link between quantile and the Value-at-Risk (VaR) measure, we compare our novel methodology with the most popular ones in VaR forecasting.

Keywords: Value-at-risk · Quantile regression · Random Forests · Mixed data sampling

1 Introduction

Quantile regression [10] is a popular technique that offers a more complete picture of the conditional distribution of the response variable with respect to standard linear regression. It also allows to make inference on processes violating the Gaussianity assumptions of many econometric models, and for this reason it is widely popular among researchers and practitioners in a variety of fields, such as finance, economics and machine learning.

In the latter context, [9,13,14] applied quantile regression to a non-parametric framework by means of neural networks and support vector machines, whereas [1,2,12] extend Random Forests [3] to a quantile regression setting. Here we use the Random Forests approach, which represents a machine learning algorithm based on training a multitude of decision trees and computing the conditional expected value of the response variable by averaging the forecasts of each individual tree. In a quantile regression framework, the natural extension of Random Forests proposed by [12], denoted as Quantile Regression Forest (QRF), estimates the whole conditional distribution of the response variable and then computes the quantile at a probability level τ.

These models are particularly appealing since they mitigate the standard assumptions presented in traditional quantile regression models.

Given the relation between the concept of VaR and quantile, quantile regression has been successfully implemented also for financial risk management purposes. However, financial and economics data are usually observed at different frequencies. This feature does not allow to include potentially useful variables in many standard econometric models, which usually account for variables sampled at the same frequency. For instance, in daily VaR forecasting it might be useful to include monthly or quarterly economic variables (for instance, GDP or inflation) that serve as a proxy for the general state of the economy. For this reason, the Mixed Data Sampling (MIDAS) model [7] has been specifically developed to consider mixed frequency variables in linear models and applied to quantile regression and VaR forecasting by [4]. In spite of the relevance of the quantile regression technique in VaR forecasting, the QRF and the MIDAS models have never been merged together in the literature. Thus, the aim of our paper is to combine the MIDAS framework with the QRF one to forecast daily VaR by considering monthly macroeconomic variables as covariates. Up to our knowledge, this represents the first attempt in the literature to merge the quantile regression approach, Random Forests and mixed-frequency data models.

The resulting model, called MIDAS-QRF, allows to detect non-linear relations among variables without specifying a-priori any functional form. Moreover, the use of Random Forests allows to build an interpretable and computationally efficient algorithm. As a matter of fact, Random Forests have a higher grade of intepretability and require a lower computation effort to be trained with respect to other machine learning algorithms, such as neural networks.

The empirical application on a real financial dataset shows that our approach outperforms existing models in forecasting daily VaRs of the S&P 500 index in terms of backtesting procedure and by means of [8] quantile loss function. The rest of the paper is organised as follows: Sect. 2 concerns the methodology used to develop our model, Sect. 3 presents the empirical results in terms of statistical adequacy and forecast accuracy and Sect. 4 concludes.

2   Methodology

In this section the methodology used to build the MIDAS-QRF, based on the MIDAS [7] and the QRF [12] models, is presented.

Let $\{(Y_{i,t}, \mathbf{X}_{i,t})\}_{i=1}^T \in \mathbb{R} \times \mathbb{R}^P$ be the sample of random variables drawn from the unknown joint distribution of the random variables $(Y, \mathbf{X})$. In particular, $Y_{i,t}$ is the response variable sampled at time i of the $t-th$ period of the year and $\mathbf{X}_{i,t} = (X_{i,t}^1 \dots, X_{i,t}^P)'$ is the P-vector at time i of covariates sampled at the same frequency of $Y_{i,t}$. We also define $\mathbf{Z}_t = (Z_t^1, \dots, Z_t^N)'$ to be the vector of N low-frequency variables sampled at time t. In our framework, the dependent variable $Y_{i,t}$ is represented by daily financial returns sampled at day i of the $t-th$ month of the year, and the low-frequency variable might be a monthly economic variable representing the general state of the economy sampled at the

$t - th$ month of the year. The VaR at day i at probability level $\tau \in (0, 1)$, denoted with $VaR_{i,t}$, is defined as $Pr(Y_{i,t} < VaR_{i,t}|F_{i-1,t}) = \tau$, i.e. the τ-th conditional quantile of $Y_{i,t}$, given the filtration $F_{i-1,t}$. The aim of this paper is to extend the QRF framework building a MIDAS-QRF model introducing mixed-frequency data in the QRF algorithm exploiting the approach of [7] and [4], and is specified as follows:

$$Y_{i,t} = f(\mathbf{X}_{i-1,t}, \mathbf{MC}_{i-1,t}) + \varepsilon_{i,t} \tag{1}$$

where $\mathbf{MC}_{i-1,t} = (MC^1_{i-1,t}, \dots, MC^N_{i-1,t})'$ is the set containing the MIDAS component for each low-frequency variable in $\mathbf{Z}_t$, with $MC_{i-1,t} = \sum_{k=1}^{K} \phi_k(\omega) Z_{t-k}$. The MIDAS component is a filter of the last K observations of the low-frequency covariate Z_t up to the last day of the $t - th$ period. The values of $MC_{i-1,t}$ are computed through a weighting function, that in our paper is the Beta one, as [4]:

$$\phi_k(\omega) = \frac{(k/K)^{\omega_1-1}(1 - k/K)^{\omega_2-1}}{\sum_{k=1}^{K}(k/K)^{\omega_1-1}(1 - k/K)^{\omega_2-1}}. \tag{2}$$

By setting $\omega_1 = 1$ and $\omega_2 > 1$, a greater weight is imputed to more recent observations and the only parameter of the $MC_{i-1,t}$ component is ω_2. In our algorithm, the values of each MIDAS component are obtained by optimising ω_2. We denote with $\boldsymbol{\omega_2^*} = (\omega_2^1, \dots, \omega_2^N)'$ the vector containing the optimal values ω_2^*. The goal of the MIDAS-QRF is to compute the conditional quantile $Q_\tau(Y_{i,t}|\mathbf{X}_{i-1,t}, \mathbf{MC}_{i-1,t})$ at level τ by estimating $f(\cdot)$ in Eq. (1) non-parametrically.

In particular, the MIDAS-QRF algorithm consists in computing quantiles by training a QRF with high-frequency variables $\{(Y_{i,t}, \mathbf{X}_{i,t})\}_{i=1}^{T}$ and with $\mathbf{MC}_{i-1,t}$ as additional set of covariates. Each MIDAS component is computed as in Eq. (2) by optimising ω_2^* via grid search. The grid search consists in computing different sets $\mathbf{MC}_{i-1,t}$ for different values of $\boldsymbol{\omega_2}$. Then, each set $\mathbf{MC}_{i-1,t}$ and $\{(Y_{i,t}, \mathbf{X}_{i,t})\}_{i=1}^{T}$ are used to train a QRF and a vector of out-of-sample quantile forecasts is computed. The optimal $\boldsymbol{\omega_2^*}$ is the one corresponding to the QRF delivering the lowest quantile loss.

3   Empirical Application

We apply our model on a dataset containing as $Y_{i,t}$ the daily log-returns of the S&P 500 index (SPX) from December 2013 to February 2020. The covariates observed at the same daily frequency are the $Y_{i-1,t}$ and the log-difference of VIX index considered at time $i - 1, t$. The low-frequency variables introduced in the model are the American Industrial Production (MV_1) and the Consumer Price Index (MV_2) observed at monthly time, that are going to be included in the MIDAS part of the model. Data are differentiated when necessary. The MIDAS-QRF model has been trained with an expanding window approach and has been used to forecast the VaR at probability levels $\tau = 0.05, 0.01$. At each

level τ, we use 30% of the full set as set of out-of-sample observations to compute one-step-ahead VaR forecasts.

In order to evaluate the performance of the MIDAS-QRF, we compare the results with several traditional models used to forecast VaR. The first group of models comprehends the standard Quantile Regression Forest, the Linear Quantile Regression and the GARCH model with two different distributions of the errors, Normal and Student's t distributions. None of them contemplate the MIDAS component. The second group of models includes the GARCH-MIDAS, the Double Asymmetric GARCH-MIDAS and the MIDAS-QRF with one low-frequency MIDAS covariate included. In the latter cases, for each model we use first the Consumer Price Index (MV_1) as low-frequency variable and then the Industrial Production (MV_2) one.

The statistical adequacy of the VaR forecasts obtained with our model is assessed with the quantile loss results computed as in [8] and the three backtesting procedures: Unconditional Coverage Test (UC), Conditional Coverage Test (CC) and Dynamic Quantile (DQ) test [5,6,11]. We report also the values of the Actual Exceedances (AE) and the ratio of the quantile loss of the MIDAS-QRF and the loss of the benchmark model (%Loss). For the former, for $\tau = 0.05$ are expected 35 exceedances, whereas for $\tau = 0.01$ are expected 7 exceedances. For the latter, values smaller than 100 indicate that the MIDAS-QRF model deliver a smaller quantile loss with respect to the benchmark model. The results of the AE and %Loss measures are reported in Table 1 along with the p-values of the three above-mentioned backtesting procedures.

Table 1. Results of the Backtesting procedures and quantile loss for 5% and 1% VaR levels.

VaR level	5%						1%					
S&P 500	Loss	% Loss	AE	UC	CC	DQ	Loss	% Loss	AE	UC	CC	DQ
QRF-MIDAS (MV_1, MV_2)	7.62	-	30	0.41	0.67	0.24	2.57	-	9	0.50	0.71	0.11
QRF-MIDAS (MV_1)	8.12	93.90	38	0.69	0.93	0.93	2.64	97.31	8	0.75	0.87	0.07
QRF-MIDAS (MV_2)	8.29	91.97	37	0.90	0.82	0.14	2.71	94.94	9	0.50	0.71	0.10
QRF	9.31	81.88	41	0.29	0.54	0.14	2.89	88.93	10	0.31	0.52	0.12
QR	9.95	76.65	37	0.67	0.05	0.00	3.29	77.99	9	0.50	0.71	0.00
GARCH-norm	8.47	90.06	39	0.47	0.40	0.28	3.05	84.15	22	0.00	0 .00	0.00
GARCH-std	8.55	89.18	51	0.00	0.02	0.00	2.75	93.41	17	0.00	0.00	0.00
GM (MV_1)	8.35	91.31	39	0.58	0.567	0.921	2.91	88.19	21	0.00	0.00	0.00
GM (MV_2)	8.38	90.97	42	0.22	0.46	0.14	3.04	84.50	25	0.00	0.00	0.00
DAGM (MV_1)	8.35	91.31	37	0.67	0.93	0.97	2.95	86.94	23	0.00	0.00	0.00
DAGM (MV_2)	8.40	90.76	43	0.17	0.38	0.14	3.06	84.09	25	0.00	0.00	0.00

The average number of degrees of freedom for the GARCH-std is 4.5.

Empirical results for VaR at 1% and 5% indicate that the MIDAS-QRF model delivers statistically adequate forecasts, as represented by the backtesting results, and produces more accurate forecasts in terms of quantile loss and AE with respect to any other model considered in our analysis.

In particular, for VaR at 5% level, each model passes all three tests at 95% confidence level, except for the Linear Quantile Regression model and the

GARCH model with Student's t distribution of errors. In terms of quantile loss, the MIDAS-QRF delivers the smallest quantile loss, as represented by the values of %Loss, which is well bellow the threshold of 100. The two specifications of the MIDAS-QRF also deliver a smallest quantile loss and a smaller AE value with respect to the GM and DAGM models, indicating a higher risk assessment capability of our model.

At 1% level, only the three specifications of our model and the QRF are successfully backtested, whereas the rest of the models underestimate the market risk by delivering very high values for the AE results. Again, the MIDAS-QRF delivers the smallest quantile loss and produces an AE value very similar to the threshold, which is equal to 7.

The empirical results show that the MIDAS-QRF delivers the most accurate forecasts in terms of quantile loss and AE, allowing a more accurate risk assessment. This might be due to several reasons: first, differently from traditional econometric models, the MIDAS-QRF offers a non-parametric setting to compute quantiles, which does not need any hypothesis regarding the distribution of the variables. Second, machine learning algorithms are well-known for their ability to model non-linear relationships among variables, that characterise relations among economic and financial variables and that are not considered by classical econometric models such as the GARCH and Linear Quantile Regression models. Third, our model exploits information from variables sampled at different frequencies, improving the forecast accuracy by a maximum of 23% for VaR at both 5% and 1% levels.

4   Conclusions

This paper proposes a new methodology to compute conditional quantiles in a machine learning framework to jointly account for complexity, non-linearity and mixed-frequencies in data. We propose a novel algorithm called MIDAS-QRF to exploit information coming from low-frequency variables in order to compute quantiles by merging the QRF and the MIDAS models. The empirical application on a real financial dataset shows that our model delivers adequate VaR forecasts and that it outperforms popular existing models used in VaR forecasting in terms of quantile loss.

References

1. Andreani, M., Petrella, L.: Dynamic quantile regression forest. In: Book of Short Papers, Proceedings of the Italian Statistical Society 2020 (2020)
2. Athey, S., Tibshirani, J., Wager, S., et al.: Generalized random forests. Ann. Statist. **47**(2), 1148–1178 (2019)
3. Breiman, L.: Random forests. Mach. Learn. **45**(1), 5–32 (2001)
4. Candila, V., Gallo, G. M., Petrella, L.: Using mixed-frequency and realized measures in quantile regression (2020). SSRN 3722927
5. Christoffersen, P.F.: Evaluating interval forecasts. Int. Econ. Rev. **39**(4), 841–862 (1998)

6. Engle, R.F., Manganelli, S.: CAViaR: conditional autoregressive value-at-risk by regression quantiles. J. Bus. Econ. Statist. **22**(4), 367–381 (2004)
7. Ghysels, E., Sinko, A., Valkanov, R.: MIDAS regressions: further results and new directions. Econ. Rev. **26**(1), 53–90 (2007)
8. Gonzàlez-Rivera, G., Lee, T.-H., Mishra, S.: Forecasting volatility: a reality check based on option pricing, utility function, value-at-Risk, and predictive likelihood. Int. J. Forecast. **20**(4), 629–645 (2004)
9. Hwang, J., Shim, J.: A simple quantile regression via support vector machine. In: Wang, L., Chen, K., Ong, Y.S. (eds.) Advances in Natural Computation. ICNC 2005. Lecture Notes in Computer Science, vol. 3610, pp. 512–520. Springer, Berlin (2005). https://doi.org/10.1007/11539087_66
10. Koenker, R., Bassett, G.: Regression quantiles. Econometrica. **46**(1), 33–50 (1978)
11. Kupiec, P.: Techniques for verifying the accuracy of risk measurement models. J. Deriv. **3**(2), 1–10 (1995)
12. Meinshausen, N.: Quantile regression forests. J. Mach. Learn. Res. **7**(Jun), 983–999 (2006)
13. White, H.: Nonparametric estimation of conditional quantiles using neural network. In: Page, C., LePage, R. (eds.) Computing Science and Statistics, pp. 190–199. Springer, New York (1992). https://doi.org/10.1007/978-1-4612-2856-1_25
14. Xu, Q., Zhang, J., Jiang, C., Huang, X., He, Y.: Weighted quantile regression via support vector machine. Expert Syst. App. **42**(13), 5441–5451 (2015)

Gender Attitudes Toward Longevity and Retirement Planning: Theory and Evidence

Giovanna Apicella[(⊠)] and Enrico De Giorgi

Department of Economics, University of St. Gallen, 9000 St. Gallen, Switzerland
{giovanna.apicella,enrico.degiorgi}@unisg.ch

Abstract. This paper fosters discussion about the gender pension gap. We propose a research framework in financial economics, centered on the role of gender in longevity risk perception. Our approach is essentially made by three steps, aiming at the: (i) identification of drivers of subjective longevity assessment (e.g., biases), (ii) the measurement of the economic significance of longevity (mis)-perception in relation to saving and investment behaviors, (iii) the design of strategies to help women understand the opportunities behind long-term planning for retirement.

Keywords: Gender · Retirement · Longevity perception · Nudging

1 Introduction

The Agenda of the United Nations (see [22]) sets the goal of achieving gender equality and the empowerment of all women and girls, at all levels of decision-making in political, economic, and public life. Nevertheless, the EU is at least sixty years beyond from reaching gender equality (see [7]). Women aged 65+ receive, on average, 26% less income than men from the pension system in the OECD (cf. [17]). Policy makers at national and international levels recognize that the gender pension gap is a problem with serious consequences on the financial and social well-being of women in old age.

This paper intends to foster the discussion on individuals' decision-making as a driver of the gender pension gap. Such a gap represents a complex issue that goes much beyond the obvious explanation that refers to the wage gender gap, as it also relates to the outcome of individuals' forward-looking behaviour when planning for their retirement (e.g., investment, saving and annuitization choices). We advocate the use of a multidisciplinary approach to address the pension gap. We propose a research framework entailing insights and methodologies from behavioural finance, household finance, and actuarial sciences. We illustrate where the state-of-the-art stands so far in the explanation of retirement behaviour, with a focus on those driving factors that are documented to reverberate gender differences (i.e., financial literacy, risk attitudes, and beliefs). We emphasize what is not yet thoroughly addressed in the literature. Although there is consensus on the fact

that women and men form different beliefs about their survival, there are not well-established results about the nature of this discrepancy (e.g., if it arises from more marked behavioural heuristics in women) and about its impact on financial preparedness for retirement. Accordingly, no approaches have been envisaged yet to increase women's awareness of their longevity and nudge them towards better long-term decisions. We describe how longevity perception could play a role in future research on the gender pension gap.

2   Drivers of Retirement Behaviour: the State-of-the-Art

Well-being in old age depends on the ability to generate retirement savings. Financial literacy and behavioural heuristics that could lead to misjudgements are identified in the literature as drivers of individuals' retirement planning behaviour and thus offer possible explanations for the pension gap.

The standard microeconomic approach assumes that fully rational and well-informed individuals arrange their optimal saving and decumulation patterns in such a way to obtain smooth consumption over the life cycle. However, empirical evidence shows that individuals' behaviour can be inconsistent with conventional economic models. According to a growing body of economic research, a lack of adequate financial literacy can cause sub-optimal choices about financial planning and pensions. [14] found that, in the U.S., women who understand the concept of risk diversification are more likely to plan for retirement. [12] found, only among women, a statistically significant and positive relationship between engagement in retirement planning and an extended measure of financial literacy (that included investment risks). The chance to develop a retirement plan, especially among women, thus increases with a deeper financial knowledge than the basic one, i.e., financial sophistication embracing proper risks understanding and perception. A subjective dimension comes thus into question. Subjective assessment of risks and uncertainty is, indeed, found to be a determinant of retirement planning behaviour, also in connection to risks that are not merely related to the financial markets, such as longevity risk. In particular, the role of such determinant surpasses that of financial literacy, as traditionally measured and interpreted, as shown in [9], where individuals who expected to live longer were found to be more likely to be planners. The literature shows that retirement planning can be affected by the way individuals deal with uncertainty and assess the economic risks arising from it. Longevity is an important source of uncertainty, that can potentially undermine old-age financial well-being. In the next Section, we illustrate how subjective survival beliefs are proven to affect savings, investments and annuitization. Such decisions play the major role in determining economic preparation for retirement (cf. [16]) and can reverberate gender differences in decision outcomes.

3   Subjective Longevity, Gender and Economic Choices

Individuals' residual life span, that is inherently random, defines the time horizon over which a retirement income will be required. Research contributions in

economics and finance focus their attention on the effects that individuals' expectations about future events have on intertemporal decision-making. Survival expectations are documented to affect individual saving decisions (see, e.g., [19] and [11]). Under a gender perspective, literature addresses the impact of risk tolerance on women's engagement in saving (cf. [8]), but there is little research about the effects of gender differences in survival expectations. Saving for retirement is increasingly taking the form of investing in financial assets. [4] found that women invest less in stocks because of their attitudes toward risk. Longevity risk affects portfolio choice, as shown by the theoretical models predicting that investors with longer horizons should allocate more wealth to risky assets, especially under loss aversion (cf. [3]). [20] found that, after appropriate controls, investors with longer subjective time horizons hold a larger share of risky assets. [18] showed that the riskiness of household portfolios increases only with the husband's subjective time horizon. To our knowledge, this is the only research contribution addressing how differing subjective survival beliefs between females and males affect their investment decisions. A further crucial financial decision for individuals at retirement relates to the purchase of annuities. Such products provide retirement benefits, while insuring individuals against the risk of longevity and of old-age poverty. Nevertheless, a few consumers actually allocate a substantial fraction of their retirement savings to the private annuity market, in contrast with the prediction of standard economic theory ("annuity puzzle", cf. [15]). Influential empirical and experimental literature documents gender differences in annuitization decisions, that are also analysed through the lenses of behavioural finance. Women and men show a different sensitivity to the framing that is used to present risk information relative to the annuities. [1] showed that emphasizing the consequences of longevity risk (e.g. the chance to outlive the own assets) did not significantly enhance the likelihood to choose an annuity for women, but it did for men. This sheds light on the issue, still poorly addressed in the literature, that women and men could have different longevity awareness.

4   Our Research Framework and Directions

Our line of research is centered on longevity risk and on its crucial importance for long-term planning and thus for economic preparedness for retirement. In this Section, we highlight our research outcomes relative to the identification of the driving factors of longevity perception and its gender connotation. As a novel contribution of this paper, we shed light on our empirical findings about the effects that longevity (mis-)perception has on financial risk taking. We then describe how these findings pave the way for further important research on the economic implications of subjective longevity and the use of nudging in the context of retirement choices.

As a first step in the research process we have investigated how individuals assess their longevity. It is well-established in the empirical literature that subjective survival expectations fail to align with actuarial probabilities. Indeed, young individuals tend to under-estimate their survival, while the opposite holds for

older individuals (i.e., aged more than seventy). It is far from straightforward to explain both survival under-estimation at younger ages and survival over-estimation at old ages (the so-called "survival gap") in a purely rational setting (e.g., in the framework of the rational Bayesian learning paradigm). Indeed, previous literature detects behavioural heuristics behind the survival gap (cf. [10,11,13]). Nevertheless, to understand the economic consequences of individuals' beliefs about survival, the identification of judgemental biases should be accompanied by their measurement. The state-of-the art focusing on behavioural explanations tends to neglect that subjective survival beliefs also include private information that is relevant to assess survival probabilities but standard actuarial probabilities (reflecting only the effects of age, period and gender on mortality) do not incorporate (e.g., health information). In [2], we theoretically explain and empirically assess the survival gap. We propose a parsimonious model of sentiment, that is based on the assumption that subjective survival probabilities evolve according to sentiment-driven expectations about the future health status. Our model predicts that the survival gap is positive when individuals preserve their optimism in the face of health shocks. Such prediction finds significant empirical support in survey data (from the U.S. Health and Retirement Study (HRS)[1] and from the Survey of Health, Ageing, and Retirement in Europe (SHARE) (cf. [6]). Our empirical measure of the survival gap is based on health-dependent actuarial probabilities, this allowing a better quantification of the bias affecting their subjective counterpart. Besides sentiment, gender is a significant determinant of the survival gap: women are more likely to under-estimate the own survival than men, although, on average, they experience higher longevity.

As an original contribution of this paper, we discuss our preliminary empirical evidence on the relation between longevity risk perception, as measured by the survival gap, and individuals' willingness to engage in financial risks. This aspect is not addressed in [2], but seems crucial to understand household decisions, from asset allocation to insurance purchase, that is a target of our future research. Specifically, given the empirical evidence reported in [2] on the role of gender in relation to the magnitude of the survival gap, analysing how longevity risk perception affects financial risk taking represents an important step to address the pension gap. Our econometric analyses, based on data from SHARE, suggest that, after appropriate controls, those individuals who express a more precise estimation of their survival prospects are more likely to accept at least some level of financial risk when saving and investing. For these individuals, the likelihood of being financial risk-takers are up to 6% higher than for the other respondents. Furthermore, the odds of taking some financial risk is, for females, 11% to 13% lower than for males. Our research also shows that changes in subjective survival beliefs are statistically significant drivers of financial risk tolerance updating. These results shed light on the impact that the age-dependent pattern of subjective survival expectations can have on life-cycle behaviour. For instance, rational life-cycle financial planners should adopt investment strategies designed as follows: during the accumulation phase, people should gradually switch from

---

[1] https://hrs.isr.umich.edu/about.

equity-type to bond type-investments as the retirement date approaches, while, during the decumulation phase, they should exchange bonds for life annuities (e.g. the phased annuitization in [5]). A lack of longevity awareness in younger (female) people can lead them to conservative investment strategies during their working years. This would imply lower expected returns over the long term and smaller chances to obtain a high retirement income and high annuity payments. This paper aims also to provide innovative directions for further research on retirement planning. Accounting for age-dependent patterns of longevity risk perception and financial risk attitudes in models of economic behaviour would be beneficial to quantify deviations from the predictions of standard economic theory. We also advocate the use of behavioural economics for the design of strategic solutions to help the people being at disadvantage in retirement provision, as women, to make better choices (e.g. the prescriptive programs for saving decisions in [21]). In this respect, our future research targets the choice architecture where individuals make their economic decisions relevant to retirement. Indeed, individuals should be nudged towards balanced-risk investment strategies and annuitization already at earlier stages in their life when they underestimate their survival. A target for intervention is represented, for instance, by default investment options within defined contribution retirement savings arrangements.

References

1. Agnew, J.R., Anderson, L.R., Gerlach, J.R., Szykman, L.R.: Who chooses annuities? An experimental investigation of the role of gender, framing and defaults. Am. Econ. Rev. **98**(2), 418–422 (2008)
2. Apicella, G., De Giorgi, E.G.: A Behavioural Gap in Survival Beliefs (2021). SSRN. https://ssrn.com/abstract=3821595

3. Berkelaar, A.B., Kouwenberg, R., Post, T.: Optimal portfolio choice under loss aversion. Rev. Econ. Stat. **86**(4), 973–987 (2004)
4. Bernasek, A., Shwiff, S.: Gender, risk, and retirement. J. Econ. Issues **35**(2), 345–356 (2001)
5. Blake, D., Wright, D., Zhang, Y.: Age dependent investing: optimal funding and investment strategies in defined contribution pension plans when members are rational life cycle financial planners. J. Econ. Dyn. Control **38**, 105–124 (2014)
6. Börsch-Supan, A., et al.: Data resource profile: The Survey of Health, Ageing and Retirement in Europe (SHARE). Int. J. Epidemiol. (2013). https://doi.org/10.1093/ije/dyt088
7. EIGE: Gender Equality Index 2020. Digitalisation and the future of work 2020. https://eige.europa.eu/publications/gender-equality-index-2020-report
8. Fisher, P.J.: Gender difference in personal saving behaviors. J. Financial Couns. Plan. **21**(1) (2010)
9. Farrar, S., Moizer, J., Lean, J., Hyde, M.: Gender, financial literacy, and preretirement planning in the UK. J. Women Aging **31**(4), 319–339 (2019)
10. Grevenbrock, N., Groneck, M., Ludwig, A., Zimper, A.: Cognition, optimism, and the formation of age-dependent survival beliefs. Int. Econ. Rev. **62**(2), 887–918 (2021)
11. Heimer, R.Z., Myrseth, K.O.R., Schoenle, R.S.: YOLO: mortality beliefs and household finance puzzles. J. Finance **74**(6), 2957–2996 (2019)
12. Kalmi, P., Ruuskanen, O.-P.: Financial literacy and retirement planning in Finland. J. Pension Econ. Finance **17**(3), 335–362 (2018)
13. Ludwig, A., Zimper, A.: A parsimonious model of subjective life expectancy. Theor. Decis. **75**(4), 519–541 (2013)
14. Lusardi, A., Mitchell, O.S.: Planning and financial literacy: how do women fare? Am. Econ. Rev. **98**(2), 413–417 (2008)
15. Modigliani, F.: Life cycle, individual thrift, and the wealth of nations. Am. Econ. Rev. **76**(3), 297–313 (1986)
16. OECD: Improving retirement incomes considering behavioural biases and limited financial knowledge, OECD Pensions Outlook 2018, OECD Publishing Paris (2018)
17. OECD: Towards Improved Retirement Savings Outcomes for Women. OECD Publishing, Paris (2021)
18. Pak, T.-Y., Babiarz, P.: Asset allocation of two-person households under different longevity expectations. J. Consum. Aff. **53**(3), 1234–1254 (2019)
19. Post, T., Hanewald, K.: Longevity risk, subjective survival expectations, and individual saving behavior. J. Econ. Behav. Organ. **86**, 200–220 (2013)
20. Spaenjers, C., Spira, S.M.: Subjective life horizon and portfolio choice. J. Econ. Behav. Organ. **116**, 94–106 (2015)
21. Thaler, R.H., Benartzi, S.: Save More Tomorrow^{TM}: using behavioural economics to increase employee saving. J Polit. Econ. **112**(S1). Papers in Honor of Sherwin Rosen: A Supplement to **112**, S164–S187 (2004)
22. United Nations: Transforming Our World: The 2030 Agenda for Sustainable Development, A/RES/70/1 (2015). https://sdgs.un.org/publications/transforming-our-world-2030-agenda-sustainable-development-17981

Semiclassical Pricing of Variance Swaps in the CEV Model

Axel A. Araneda[1(✉)] and Marcelo J. Villena[2]

[1] Institute of Financial Complex Systems, Faculty of Economics and Administration, Masaryk University, 602 00 Brno, Czech Republic
axelaraneda@mail.muni.cz

[2] Department of Commercial Engineering, Universidad Técnica Federico Santa María, 7660251 Santiago, Chile

Abstract. Path integrals are a well-known tool in quantum mechanics and statistical physics. They could be used to derive the propagator or kernel of stochastic processes, analogous to solving the Fokker-Planck equation. In finance, they become an alternative tool to address the valuation of derivatives. Here, taking advantage of the hedging formula of the realized variance by means of the log contract, we use path integrals for the pricing of variance swaps under the Constant Elasticity of Variance (CEV) model, approximating analytically the propagator for the log contract by semiclassical arguments. Our results demonstrate that the semiclassical method provides an alternative and efficient computation which shows a high level of accuracy but at the same time lower execution times.

Keywords: Path integral · CEV model · Variance swap

1 Introduction

The constant elasticity of variance (CEV) model [1,2] is featured by its capacity to address some phenomena as heteroskedasticity, implied volatility skew, bankruptcy, and leverage effect, but also by its analytical tractability [3].

The variance swap (VS) pricing, when the underlying asset rules under the CEV assumption, has been addressed by means of the valuation of the log contract [4] or using an approximation via small disturbance asymptotic expansion of the CEV variance [5]. Moreover, Carr & Sun [6] provide closed-form pricing formulas by means of Laplace inversion to the variance 3/2 model. It should be noted that the CEV model implies that instantaneous variance follows a 3/2 process, which has received a lot of empirical support [6,7].

Our idea here is to provide an alternative approach to the VS valuation, based on the path integral framework [8]. Using the replication of the realized variance by means of the log contract [9], we approximate the transition density for the price process (a.k.a. propagator) using the well-known semiclassical approach for the path integral, a functional analog of the stationary phase approximation [10].

This method was used recently in the CEV option pricing problem [11], showing be an efficient approach and competitive or even superior performance with other numerical schemes, namely Monte Carlo simulation, binomial tree, and finite difference method. Here, we will study its application in the VS valuation in terms of the accuracy and the capacity to provide significantly lower running times than the valuation using the exact analytical propagator.

2   The Model

In the CEV environment, the price S is ruled by the following stochastic differential equation:

$$dS_t = rS_t + \delta S_t^{\alpha+1}dW_t \tag{1}$$

where W is a standard Brownian motion and r, $\sigma > 0$, and $\alpha \in [-1,0[$ the constant parameters of the model.

The transition probability density function of the price at time T, a.k.a propagator, is given by [11]:

$$P\left(S_T, T \mid S_0, 0\right) = (2-\alpha)\, k^{\frac{1}{2-\alpha}} \left(yw^{1-2\alpha}\right)^{\frac{1}{2(2-\alpha)}} e^{-y-w} I_{1/(2-\alpha)}\left(2\sqrt{yw}\right) \tag{2}$$

where $I_\lambda(x)$ is the modified Bessel function of the first kind of order λ and:

$$k = y\frac{2r}{\sigma^2\,(2-\alpha)\left[e^{r(2-\alpha)T}-1\right]}, \; y = kS_0^{2-\alpha}e^{r(2-\alpha)T}, \; w = kS_T^{2-\alpha}$$

On the other hand, the instantaneous variance of the model is given by:

$$V_t = \delta^2 S_t^{2\alpha} \tag{3}$$

In addition, by Itó's lemma, is clear that the variance process follows a $3/2$ model:

$$dV_t = \alpha V_t\left[(2\alpha-1)V_t + 2r\right]dt + 2\alpha V_t^{\frac{3}{2}}dB_t$$

2.1   Variance Swap Pricing

The VS is a derivative based on the difference between the annualized realized variance over the expected one, in a defined time span. In continuous-time, it's defined as:

$$VS = \left(\frac{1}{T}\int_0^T V_t dt - E_V\right) \times N \tag{4}$$

being V_t the instantaneous variance of the underlying asset and N the variance notional amount. At the initial time $(t = 0)$, the entry cost of the contract is zero. It means that the variance strike E_V is fixed to be equal to the realized variance in the interval $[0, T]$. Then, in a risk-neutral world, the pricing of a variance swap is reduced to compute the expected value of the realized variance.

3   Realized Variance Replication

The realized variance could be replicated by means of the log contract following the approach of Demeterfi et al. [9]:

$$\mathbb{E}\left[\frac{1}{T}\int_0^T V_t\right] = 2r + \frac{2}{T}\log\left(S_0\right) - \frac{2}{T}\mathbb{E}\left[\log\left(S_T\right)\right] \tag{5}$$

Then, according to Eq. (5), the variance swap pricing is reduced to compute the expected value for $\log\left(S_T\right)$, which is estimated by:

$$\mathbb{E}\left[\log\left(S_T\right)\right] = \int_{-\infty}^{\infty}\log\left(S_T\right)\cdot P\left(S_T,T|\,S_0,0\right)\mathrm{d}S_T \tag{6}$$

The previous integral can be addressed numerically but its computation is especially costly under i) short maturities, ii) short volatilities, and iii) when elasticity tends to zero [4].

3.1   The Semiclassical Approximation for the Log Contract

First, by means of the transformation $x = -\left(\alpha\delta S^\alpha\right)^{-1}$, we obtain a new process with constant (unitary) diffusion coefficient:

$$\mathrm{d}x_t = \left(\frac{\alpha+1}{\alpha x} - r\alpha x\right)\mathrm{d}t + \mathrm{d}W_t \tag{7}$$

Following the methodology given by Bennati et al. [12], the propagator can be obtained using the following path integral representation:

$$P\left(x_T,T|\,x_0,0\right) = \int\mathcal{D}x(t)\mathrm{e}^{-\int_o^T\mathcal{L}_x\mathrm{d}t}$$

being $\mathcal{D}x(t)$ the measure of the integration which considers all the trajectories from $x(t=0)=x_0$ to $x(t=T)=x_T$.

$\mathcal{L}_x$ is called the Lagrangian and is given by [12]:

$$\mathcal{L}_x = \frac{1}{2}\left(\dot{x} + r\alpha x - \frac{\alpha+1}{2\alpha x}\right)^2 - \frac{1}{2}\left(\frac{\alpha+1}{2\alpha x^2} + r\alpha\right)$$
$$= \mathcal{L}'_x + \mathcal{G}_x$$

where

$$\mathcal{L}'_x = \frac{\dot{x}^2}{2} + \frac{r^2\alpha^2 x^2}{2} + \frac{1-\alpha^2}{8\alpha^2 x^2} \tag{8}$$
$$\mathcal{G}_x = \dot{x}\left(r\alpha x - \frac{\alpha+1}{2\alpha x}\right) + \frac{(2\alpha+1)\,r}{2}$$

Before starting, we develop the time integral of $\mathcal{G}_x$ in the following way:

$$\int_0^T \mathcal{G}_x \mathrm{d}t = \frac{r\alpha}{2}\left(x_T^2 - x_0^2\right) - \frac{(\alpha+1)}{\alpha}\left(\ln x_T - \ln x_0\right) - \frac{(2\alpha+1)\,rT}{2}$$

and since the above result is path-independent (only depends on the x_T and x_0):

$$P\left(x_T, T \middle| x_0, 0\right) = \left(\frac{x_T}{x_0}\right)^{\frac{(\alpha+1)}{\alpha}} \mathrm{e}^{\frac{(2-\alpha)rT}{2} - \frac{r\alpha}{2}\left(x_T^2 - x_0^2\right)} \int \mathcal{D}x(t)\mathrm{e}^{-\int_0^T (\mathcal{L}_x')\mathrm{d}t} \quad (9)$$

Instead of computing analytically the path integral inside (9), we could approximate it by semiclassical arguments by means of Pauli's formula [13]:

$$\int \mathcal{D}x(t)\mathrm{e}^{-\int_o^T \mathcal{L}_x'\mathrm{d}t} \approx \mathrm{e}^{-A_{\bar{y}}}\sqrt{-\frac{\mathcal{M}}{2\pi}} \quad (10)$$

being $A_{\bar{x}}'$ the Action functional (time integral of the Lagrangian) evaluated at the classical path $\bar{x}$; i.e., classical action. The classical path is who links the initial and final points and obeys the Euler-Lagrange equation. The term $\mathcal{M}$ is called the Van-Vleck-Morette determinant [10] and is computed as:

$$\mathcal{M} = \frac{\partial^2 A_{\bar{x}}}{\partial x_0 \partial x_T}$$

Directly from Eq. (8), the classical path related to $\mathcal{L}_x'$ (which solve the associated E-L equation) is given by:

$$x(t) = \sqrt{\frac{4\omega^2\left(\lambda - \frac{1}{4}\right) - C_1^2\left(2\omega C_2 + \mathrm{e}^{-2\omega t}\right)^2}{4\omega^2 C_1 \mathrm{e}^{-2wt}}}$$

where C_1 and C_2 are constant of integration obtained by means of the extremality conditions $(x(0) = x_0$ and $x(T) = x_T)$.

After replacing the classical path into Eq. (8) and its time integral (i.e., $A_{\bar{x}}'$), the Van-Vleck-Morette determinant can be obtained and, in consequence, the semiclassical approximation is achieved, and then, the propagator could be approximated analytically.

4   Numerical Results

By means of a standard numeric integration scheme, the computation of (6) is performed. We can observe, in Table 1, the successfulness of the semiclassical approach for the pricing of the log contract, and in consequence the variance swap pricing, compared with the benchmark (i.e., using the analytical and exact propagator). The semiclassical way offers both i) very low computational times (a difference of two orders of magnitude) and ii) high accuracy contrasted to the standard way of computation. Through the relative error under different parameter sets, it's clear that the better performance of the semiclassical approximation occurs for elasticities near zero, short maturities, and small volatilities.

Table 1. Comparison for the semiclassical approximation and benchmark using $S_0 = 100$ and $r = 5\%$.

T	δ	α	Semiclassical		Benchmark		Relative error (%)
			$\mathbb{E}\left[S_T\right]$	Time (s)	$\mathbb{E}\left[S_T\right]$	Time (s)	
0.25	0.05	-0.02	4.6162	0.0002	4.6174	0.0773	0.03
		-0.2	4.6061	0.0002	4.6061	0.0793	0.25
	0.5	-0.02	4.5895	0.0002	4.5917	0.0778	0.05
		-0.2	4.5990	0.0002	4.6127	0.0780	0.30
1	0.05	-0.02	4.6427	0.0002	4.6541	0.0777	0.25
		-0.2	4.6086	0.0002	4.6550	0.0801	1.00
	0.5	-0.02	4.5413	0.0002	4.5504	0.0817	0.20
		-0.2	4.6086	0.0002	4.6354	0.0779	1.18

References

1. Cox, J.C.: Notes on option pricing I: constant elasticity of variance diffusions. Working paper, Stanford University (1975)
2. Cox, J.C.: The constant elasticity of variance option pricing model. J. Portfolio Manag. **23**(5), 15–17 (1996)
3. Linetsky, V., Mendoza, R.: The constant elasticity of variance model. In: Encyclopedia of Quantitative Finance, pp. 328–334 (2010)
4. Jordan, R., Tier, C.: The variance swap contract under the CEV process. Int. J. Theor. Appl. Finance **12**(05), 709–743 (2009)
5. Wang, H., O'Hara, J.G., Constantinou, N.: A path-independent approach to integrated variance under the CEV model. Math. Comput. Simulation **109**, 130–152 (2015)
6. Carr, P., Sun, J.: A new approach for option pricing under stochastic volatility. Rev. Deriv. Res. **10**(2), 87–150 (2007)
7. Goard, J., Mazur, M.: Stochastic volatility models and the pricing of VIX options. Math. Finance **23**(3), 439–458 (2013)
8. Feynman, R.P., Hibbs, A.R.: Quantum Mechanics and Path Integrals. McGraw-Hill, New York (1965)
9. Demeterfi, K., Derman, E., Kamal, M., Zou, J.: A guide to volatility and variance swaps. J. Deriv. **6**(4), 9–32 (1999)
10. Morette, C.: On the definition and approximation of Feynman's path integrals. Phys. Rev. **81**(5), 848 (1951)
11. Araneda, A.A., Villena, M.J.: Computing the CEV option pricing formula using the semiclassical approximation of path integral. J. Comput. Appl. Math. **388**, 113244 (2021)

12. Bennati, E., Rosa-Clot, M., Taddei, S.: A path integral approach to derivative security pricing I: formalism and analytical results. Int. J. Theor. Appl. Finance **2**(04), 381–407 (1999)
13. Grosche, C., Steiner, F.: Handbook of Feynman Path Integrals. Springer Tracts in Modern Physics, vol. 145. Springer, Cham (1998). https://doi.org/10.1007/BFb0109520

Indexing Pensions to Life Expectancy: Keeping the System Fair Across Generations

Mercedes Ayuso[1(✉)] and Jorge Miguel Bravo[2]

[1] Department of Econometrics, Statistics and Applied Economy, Riskcenter-UB, Faculty of Economics and Business, University of Barcelona, Barcelona, Spain
mayuso@ub.edu

[2] NOVA IMS - Universidade Nova de Lisboa & Université Paris-Dauphine PSL & MagIC & CEFAGE-UE, Lisbon, Portugal
jbravo@novaims.unl.pt

Abstract. Linking pensions to longevity developments and population ageing is one of the most common types of automatic adjustment mechanisms in pension schemes. Although this reform approach is primarily driven by cost-containment objectives, other dimensions of welfare restructuring are present, including pension adequacy, recalibration, introducing economic and actuarial rationality, recommodification, and blame avoidance for unpopular policies that involve retrenchments. This paper discusses how to index pensions to longevity developments and population ageing in a way that is consistent with actuarial fairness and neutrality across generations. We derive an intergenerational fairness and neutrality condition for pension reform and examine alternative policy options including modifying the contribution rate, updating the statutory retirement age, or introducing sustainability factors.

Keywords: Automatic adjustment mechanisms · Life expectancy · Pensions · Actuarial fairness · Risk-sharing · Longevity risk

1 Introduction

Pension schemes require regular adjustments to address the long-term affordability, fiscal sustainability and adequacy challenges posed by demographic (e.g., population ageing), economic (e.g., low productivity gains and economic growth, a rapidly shifting labour market) and financial (e.g., low-for-long interest rate scenario) shocks. These adjustments can be discretionary or follow some (fully or semi) automatic adjustment or stabilization mechanism (AASM), mechanically updating the scheme's parameters (e.g., retirement age) conditional on some triggering indicator (e.g., life expectancy). The introduction of automatic stabilizers replaces regular discretionary measures, contributing to enhancing the credibility of the system, social trust, and the support of the intergenerational contract by preventing otherwise unexpected public finance crises and major benefit cuts in the future [1]. About two-thirds of OECD countries employ some form of AASM in mandatory pension schemes [2].

Linking pensions to longevity developments is one of the most common types of AASM [3–6]. However, some studies pointed out several deficiencies in the way pensions have been indexed to life expectancy developments, including the use of inappropriate longevity measures, the adoption of uniform markers neglecting longevity heterogeneity and lifespan inequality, embracing compensation and obfuscation strategies such as sequencing, long-phasing in periods, and long indexation lags [7, 8]. Critical to this paper, they were not designed to keep the scheme fair across generations.

This paper extends Bravo et al. [9] and discusses how to index pensions to longevity developments and population ageing in a way that is consistent with actuarial fairness and neutrality principles. We derive the intergenerational fairness and neutrality condition for pension reform and discuss alternative automatic adjustment mechanisms including modifying the contribution rate, updating the statutory retirement age, or introducing sustainability factors, but the full policy option menu includes indexing pensions in payment, adjusting the penalties (bonus) for early (late) retirement, modifying past earnings revalorization rate. The structure of this article is as follows. Section 1 outlines the key concepts and research methods used in the paper. Section 2 presents the model setup, the intergenerational fairness and neutrality condition. Section 3 examines several alternative policy options. Section 4 concludes.

2 Intergenerational Fairness and Neutrality Condition

In this paper, we follow and extend Bravo et al. [9] and consider a stylized career average re-evaluated earnings-related non-financial defined benefit (NDB) pension scheme with entry pension actuarially computed based on the entire contribution effort. The approach is extended to account for population ageing (increase in the old-age dependency ratio) and the existence of external sources of funding in the pension scheme. The actuarial pay-as-you-go aggregate balance constraint in year t equals the revalued contribution effort and the pension wealth

$$A_t \cdot c_t \cdot V_t + EX_t = L_t \cdot \lambda_t \cdot P_{x_r(t)} \cdot a_{x_r(t)}^{\pi,y}, \tag{1}$$

where A_t is the number of active workers in the scheme; c_t is the contribution rate; $V_t \equiv V\left(x_{r(t)}, x_e, w, y_t\right)$ is the lifetime pensionable average salary w_t of all active workers, revalued using an (actuarial equilibrium, notional) rate of return y_t; x_e is the average labour market entry age; EX_t represents the external sources of funding (e.g., general or dedicated taxes); L_t is the number of pensioners; $\lambda_t \geq 1$ is the average number of pensions per pensioner (to account for the overlapping of old-age and survivor's pensions); $P_{x_r(t)}$ is the annual average pension benefit across all retirees, computed as follows:

$$P_{x_r(t)} = \theta_t\left(x_{r(t)} - x_e\right) \cdot \overline{RE}_{x_r(t)} \cdot SF_{x_r(t)} \cdot b_{x_r(t)}, \tag{2}$$

where θ_t is a linear (usually flat) accrual rate for each year of service, $\left(x_{r(t)} - x_e\right)$ is the average contribution period with $x_r(t)$ the exit (retirement); $\theta_t\left(x_{r(t)} - x_e\right)$ is the scheme's target replacement rate; $SF_{x_r(t)}$ is a life expectancy coefficient (often called sustainability factor) introduced in some countries (e.g., Finland, Portugal) to adjust

entry pensions to longevity increases; $b_{x_r(t)}$ are pension decrements $\left(b_{x_r(t)} < 1\right)$ or pension increments $\left(b_{x_r(t)} > 1\right)$ for early or delayed retirement, respectively; $\overline{RE}_{x_r(t)} \equiv \overline{RE}\left(x_{r(t)}, x_e, w_t, \upsilon_t\right)$ is the lifetime average revalued earnings of all active workers $\overline{RE}_{x_r(t)} = RE_{x_r(t)}/(x_r(t) - x_e)$ with

$$RE_{x_r(t)} = \left(w_t^{x_r(t)} + \sum_{x=x_0}^{x_r(t)-1} w_{t-x_r(t)+x}^{x_r(t)} \prod_{j=t-x_r(t)+x+1}^{t} \left(1 + \upsilon_j\right) \right), \tag{3}$$

where υ_t denotes the rate at which each year contributions are revalued; $a_{x_r(t)}^{\pi,y}$ is the life annuity factor

$$a_{x_r(t)}^{\pi,y} := \sum_{\tau=1}^{\omega-x_r} \left(\frac{1 + \pi_\tau}{1 + y_\tau} \right)^t {}_\tau p_{x_r(t)}. \tag{4}$$

where π is the uprating rate for pensions, ${}_\tau p_{x_r(t)}$ is the τ-year survival probability of a population cohort aged x_r at time t, computed using a diagonal (cohort) approach. Let D_t denote the scheme's old-age dependency ratio - the ratio between the number of pensions $L_t \lambda_t$ and the number of active workers A_t -, $D_t = L_t \lambda_t / A_t$. The balance constraint (1) can be rewritten as

$$c_t \cdot V_t + EX_t/A_t = D_t \cdot P_{x_r(t)} \cdot a_{x_r(t)}^{\pi,y}, \tag{5}$$

    If the longevity prospects of the population increase, the pension scheme parameters (e.g., the early and normal retirement ages, the contribution rate, the life expectancy coefficient, the accrual rate per year, the survivor pensions benefit formula, the indexation rate of pensions) must be updated to ensure the scheme remains actuarially fair and neutral across generations and does not require external funding. To ensure the scheme remains fair and neutral across the members of the initial (labelled 0) and the current (labelled t) generations, the following condition must hold:

$$\frac{c_t}{c_0} \cdot \frac{V_t}{V_0} + \frac{EX_t/A_t}{EX_0/A_0} = \frac{D_t}{D_0} \cdot \frac{\theta_t\left(x_{r(t)} - x_e\right)}{\theta_0\left(x_{r(0)} - x_e\right)} \cdot \frac{\overline{RE}_{x_r(t)}}{\overline{RE}_{x_r(0)}} \cdot \frac{SF_{x_r(t)}}{SF_{x_r(0)}} \cdot \frac{b_{x_r(t)}}{b_{x_r(0)}} \cdot \frac{a_{x_r(t)}^{\pi,y}}{a_{x_r(0)}^{\pi,y}}. \tag{6}$$

where we assumed the parameters that are not pension policy instruments (e.g., wages, labour market entry age) are kept constant.

    Without loss of generality, assume now that individuals of both cohorts retire at the full old-age pension age (i.e., $b_{x_r(t)}/b_{x_r(0)} = 1$), that the life expectancy coefficient is constant over time (i.e., $SF_{x_r(t)}/SF_{x_r(0)} = 1$), and that the external funding per active worker EX_t/A_t is null or remains fixed over time. The fairness condition (6) simplifies to:

$$\frac{c_t}{c_0} \cdot \frac{V_t}{V_0} = \frac{D_t}{D_0} \cdot \frac{\theta_t\left(x_{r(t)} - x_e\right)}{\theta_0\left(x_{r(0)} - x_e\right)} \cdot \frac{\overline{RE}_{x_r(t)}}{\overline{RE}_{x_r(0)}} \cdot \frac{a_{x_r(t)}^{\pi,y}}{a_{x_r(0)}^{\pi,y}}. \tag{7}$$

Equations (6) and (7) offer a complete menu of automatic adjustment mechanisms and pension policy rules to absorb the impact of economic and/or demographic shocks and

preserve actuarial fairness and neutrality across generations. Theoretically speaking, the policy interventions can take place at the three stages of pensions: accumulation (e.g., contribution rate), annuitization (e.g., retirement age, sustainability factor), and payout (pensions indexation rate), and may even combine multiple interventions in all three stages [3]. In real-world cases, it is well known that some reforms are politically and socially hard to approve and sustain over time, as recent empirical evidence shows in many OECD countries. Moreover, automatic adjustments may modify the way the cost (and the risks) of providing for pensions is shared among generations. In the next section, we summarize some of the policy options offered by the intergenerational fairness condition above.

3 Policy Options

3.1 Adjusting the Contribution Rate

In a pure NDB scheme, the natural control variable is the contribution rate. The individual benefits are defined by a set of rules and the social insurance premiums, contributions, or taxes paid to cover the benefits must adapt to accommodate to whatever is required to cover the additional costs generated by longer lives under the given set of rules including the retirement age and the benefit formula. From (7), keeping all other parameters fixed and assuming lifetime earnings are revalued at the scheme's internal rate of return (i.e., $\upsilon_t = y_t \ \forall_t$), the dynamics of the contribution rate required to cope with the population extended longevity prospects follows

$$c_t = c_0 \cdot \frac{a_{x_r(t)}^{\pi,y}}{a_{x_r(0)}^{\pi,y}} \cdot \frac{D_t}{D_0}. \tag{8}$$

From (8), we can conclude that the contribution rate updates required to cope with increasing survival rates and population ageing depend on two multiplicative factors: (i) the first is a ratio between the actuarial value of the annuity factor at time t and that of the corresponding benchmark value at time 0, $\left(a_{x_r(t)}^{\pi,y}/a_{x_r(0)}^{\pi,y}\right)$. If lower (higher) mortality is observed (and forecasted), the contribution rate must increase (decline). The second adjustment factor (D_t/D_0) captures the dynamics of the scheme's old-age dependency ratio. If the number of pensions relative to active workers augments, due to increased life expectancy and/or population ageing and/or a deterioration in the labour market conditions (reduced participation and/or higher unemployment rates), the ratio D_t/D_0 augments and the contribution must increase to keep the scheme fair and neutral across generations.

3.2 Adjusting the Retirement Age While Keeping the Replacement Rate Constant

Under this policy design, the contribution period is extended, and the retirement age increased while maintaining the macro replacement rate constant. This roughly means the additional contribution effort does not translate into higher pension entitlements.

To achieve it, the accrual rate per year must be reduced. From (7), keeping all other parameters fixed and assuming that lifetime earnings are revalued at the scheme's internal rate of return and that the uprating rate for pensions matches the discount rate (i.e., $\pi_t = y_t$ $\forall_t$), it can be shown that the dynamics of the retirement age follows

$$\frac{\dot{e}^C_{x_r(t)}}{\left(x_{r(t)} - x_e\right)} = \frac{\dot{e}^C_{x_r(0)}}{\left(x_{r(0)} - x_e\right)} \cdot \frac{D_0}{D_t}, \tag{9}$$

where $\dot{e}^C_{x_r(t)}$ is the cohort life expectancy at the retirement age. From Eq. (9), we conclude that to keep the pension scheme actuarially fair and neutral and the replacement rate constant when longevity increases, the retirement age must be updated such that the expected years in retirement relative to contribution years equal that of the benchmark (initial) generation reduced by the rate of increase in the scheme's old-age dependency ratio. In a scenario of population ageing the ratio D_0/D_1 declines ($D_0/D_1 < 1$) and future pensioners will enjoy a shorter fraction of their lives in retirement compared to previous generations.

3.3 Adjusting the Retirement Age While Improving Pension Adequacy

Under this policy design, the retirement age is increased, and the extra contribution period translates into higher pension entitlements, improved pension adequacy, and an enlarged pension scheme. This is achieved by keeping the accrual rate per year constant and the other scheme's parameters unchanged. From (7), assuming again that lifetime earnings are revalued at the scheme's internal rate of return and that the uprating rate for pensions matches the discount rate, the new equilibrium retirement age follows

$$\dot{e}^C_{x_r(t)} = \dot{e}^C_{x_r(0)} \cdot \frac{D_0}{D_t}. \tag{10}$$

Equation (10) states that to cope with increased life expectancy at retirement ages and population ageing while improving pension adequacy and keeping the scheme fair across generations, the pension age must be updated such that the expected period in retirement is reduced by a factor equal to the rate of increase in the scheme's old-age dependency ratio. Under this policy design, all extra longevity is spent working and the required pension age adjustments are higher than that obtained with (9). Stated differently, to improve pension adequacy younger cohorts must accept a reduced period in retirement.

3.4 Amending Entry Pensions Through a Sustainability Factor

For a given retirement age, sustainability factors reduce pension entitlements to compensate for the extra pension expenditures that come with increased life expectancy [1]. Sustainability factors gradually reduce the replacement rate of pensions, which is often wrongly perceived as a measure of the scheme's generosity. In some countries (e.g., Portugal), the factor introduction was originally combined with flexible retirement age approaches, including the possibility of extending working life to offset the pension cuts

introduced by the reduction factor. From (6), keeping all other parameters fixed (including the absence of external funding and a constant accrual rate per year) and considering the same assumptions as above, the dynamics of the sustainability factor follows

$$SF_{x_r(t)} = SF_{x_r(0)} \cdot \frac{a_{x_r(0)}^{\pi,y}}{a_{x_r(t)}^{\pi,y}} \cdot \frac{D_0}{D_t}, \tag{11}$$

From (11) it follows that in a scenario of increased longevity and population ageing, entry pensions must gradually be adjusted by a factor equal to the inverse of the product of the rate of change in the scheme's old-age dependency ratio and the rate of increase in the annuity factor, to keep the scheme financially balanced and fair across generations. This policy transfers, directly and indirectly, the financial burden of expanding lifetime prospects to pensioners, which are at the end of the day the main beneficiaries of longer lives.

4   Conclusion

This paper considers a simple stylized Bismarckian earnings-related NDB scheme to derive an intergenerational fairness condition on how to index pensions to longevity developments and population ageing in a way that is consistent with actuarial fairness and neutrality across generations. The results show that increases in life expectancy at retirement ages should be accompanied by either an increase in the contribution rate, by increasing the statutory retirement age while keeping the replacement rate constant or, alternatively, while expanding pension adequacy, by introducing a sustainability factor linking entry pensions to longevity gains at annuitization, or a combination of all of the above. Importantly, the results show that population ageing, as measured here by an increase in the pension scheme's old-age dependency ratio, demands an extra correction in the key parameters since this shock structurally affects the relationship between the contribution revenue and pension expenditure. Otherwise, countries will have to increasingly resort to external funding sources (or, worst, denying benefits) to restore financial balance. Further research will empirically investigate the magnitude of the adjustments prescribed by the above policy options.

References

1. Bravo, J.M., Ayuso, M.: Linking pensions to life expectancy: tackling conceptual uncertainty through Bayesian Model averaging. Mathematics **9**(24), 3307, 1–27 (2021). https://doi.org/10.3390/math9243307

2. OECD: Pensions at a Glance 2021. OECD Publishing, Paris (2021)
3. Ayuso, M., Bravo, J.M., Holzmann, R.: Getting life expectancy estimates right for pension policy: period versus cohort approach. J. Pension Econ. Finan. **20**(2), 212–231 (2021). https://doi.org/10.1017/S1474747220000050
4. Olivieri, A.: Designing annuities with flexibility opportunities in an uncertain mortality scenario. Risks **9**(11), 189 (2021). https://doi.org/10.3390/risks9110189
5. Bravo, J.M., El Mekkaoui de Freitas, N.: Valuation of longevity-linked life annuities. Insur. Math. Econ. **78**, 212–229 (2018)
6. Bravo, J.M.: Pricing participating longevity-linked life annuities: a Bayesian Model Ensemble approach. Eur. Actuar. J. 1–35 (2021). https://doi.org/10.1007/s13385-021-00279-w
7. Bravo, J.M., Ayuso, M., Holzmann, R., Palmer, E.: Addressing the life expectancy gap in pension policy. Insur. Math. Econ. **99**, 200–21 (2021a)
8. Ayuso, M., Bravo, J.M., Holzmann, R., Palmer, E.: Automatic indexation of pension age to life expectancy: when policy design matters. Risks **9**(5), 96 (2021). https://doi.org/10.3390/risks9050096
9. Bravo, J.M., Ayuso, M., Holzmann, R., Palmer, E.: Intergenerational Actuarial Fairness When Longevity Increases: Amending the Retirement Age, pp. 1–41. (CESifo Working Papers; No. 9408). Munich Society for the Promotion of Economic Research - CESifo GmbH (2021b). https://ssrn.com/abstract=3961911

Dynamic Withdrawals and Stochastic Mortality in GLWB Variable Annuities

Anna Rita Bacinello[1], Rosario Maggistro[1(✉)], and Ivan Zoccolan[2]

[1] Department of Economics, Business, Mathematics and Statistics 'Bruno de Finetti', University of Trieste, via Valerio 4/1, 34127 Trieste, Italy
bacinel@units.it, rosario.maggistro@deams.units.it
[2] Generali Italia S.p.A, Piazza Tre Torri 1, 20145 Milan, Italy

Abstract. In this paper we propose a discrete time model, based on dynamic programming, to price GLWB variable annuities under the dynamic approach within a stochastic mortality framework. Our set-up is very general and only requires the Markovian property for the mortality intensity and the asset price processes. We also show the validity of the bang-bang condition for the set of discrete withdrawal strategies of the model. This result allows to drastically reduce the computational time needed to search the optimal withdrawal in the backward recursive step of our dynamic algorithm and provides, as a by-product, an interesting contract decomposition.

Keywords: GLWB · Dynamic withdrawals · Bang-bang condition · Stochastic mortality

1 Introduction

Variable annuities (VAs) are very flexible life insurance investment products that package living and death benefits endowed with a number of possible guarantees in respect of financial or biometric risks. A rider that can be included in a VA contract in order to provide a post-retirement income is the Guaranteed Lifelong Withdrawal Benefit (GLWB), that offers a lifelong withdrawal guarantee. There has been a number of papers dealing with pricing of the VA products. Most of them are focused on pricing VA guarantees under the *static* policyholder behaviour (see e.g., [1]), meaning that the policyholder always withdraws exactly the guaranteed amount, and never surrenders the contract. Some studies include pricing under the *dynamic* approach, when the policyholder optimally decides the amount to withdraw at each withdrawal date depending on the information available at that date (see, e.g., [2]). According to whether withdrawals are assumed to occur continuously or discretely, the optimal withdrawal problem under the dynamic approach is generally solved using, respectively, stochastic control and dynamic programming [3]. In this paper we propose a discrete time model, based on dynamic programming, to price VAs with GLWB under the dynamic approach within a stochastic mortality framework. Our set-up is very

general and only requires the Markovian property for the mortality intensity and the asset price processes. Another contribution of our paper is the verification of the bang-bang condition for the set of discrete withdrawal strategies of the GLWB model. This means that the set of the optimal withdrawals consists of three choices only: zero withdrawal, withdrawal at the contractual amount, complete surrender. This result, proven in our discrete time framework, is particularly remarkable as in the insurance literature either the existence of optimal bang-bang controls is assumed or it requires suitable conditions (see e.g., [4]). The bang-bang condition, beyond drastically reducing the computational time needed to search the optimal withdrawal in the backward recursive step of our dynamic algorithm, allows to clearly separate the various contract components.

The remainder of this paper is organized as follows. In Sect. 2 we describe the structure of the VA contract. In Sect. 3 we introduce our valuation framework and define the optimal withdrawal problem. In Sect. 4 we first define the dynamic programming equations that allow to solve the problem, then we introduce the bang-bang condition and outline the proof of its validity, and after we present the contract decomposition. Finally, Sect. 5 concludes the paper.

2 The Contract Structure

In this section we describe the GLWB rider in our variable annuity contract. At time 0 (contract inception), the policyholder, aged x, pays a single premium P which is entirely invested in a well-diversified and non-dividend paying mutual fund of her own choice. We denote by S_t the market price at time t of each unit of this fund, that drives the return on the investment portfolio built up with the policyholder's payment. The value at time t of such portfolio, that is called 'personal account', is denoted by W_t. The GLWB rider gives the policyholder the right to make periodical withdrawals from her account at some specified dates for the whole life, even if the account value is reduced to zero. The cost of the guarantee is financed by periodical proportional deductions from the personal account value, while the guaranteed withdrawal amount is calculated as a fixed proportion g of the 'benefit base', denoted by A_t, which is initially set equal to the single premium. In addition, the benefit base can be adjusted upward via the 'roll-up' feature, that applies when no withdrawal is made on a specified withdrawal date. Both the complete surrender of the policy and the policyholder's death are events that cause the closure of the contract. The value that remains in the personal account when the policyholder dies is paid to the beneficiary as a death benefit. In particular, from now on we assume that: (i) withdrawals are allowed on a predetermined set of equidistant dates and we take the distance between two consecutive dates as unit of measurement of time; (ii) the death benefit is paid to the beneficiary on the next upcoming withdrawal date. Let τ denote the time of death of the policyholder, so that withdrawals are allowed only at times $i = 1, 2, \ldots$, provided that $\tau > i$. The guaranteed amount that can be withdrawn at time i is equal to gA_i, and the return on the reference fund over the interval $[i-1, i]$ is $R_i = (S_i/S_{i-1}) - 1, \ i = 1, 2, \ldots$. We denote by y_i the

actual withdrawal made by the policyholder at time i and, under our dynamic approach, we assume that the set of possible withdrawals at this time is given by the interval $[0, \max\{gA_i, W_i\}]$. If the policyholder does not withdraw anything at time i, the benefit base is proportionally increased according to the roll-up rate, that we denote by b_i (with $0 < b_i < 1$), while, if the withdrawal exceeds gA_i, it is proportionally reduced according to the so called 'pro-rata' adjustment rule. Then the benefit base evolves as follows:

$$A_{i+1} = f_{i+1}^A(W_i, A_i, y_i) = \begin{cases} A_i(1 + b_i) & \text{if } y_i = 0, \\ A_i & \text{if } 0 < y_i \leq gA_i, \\ A_i \dfrac{W_i - y_i}{W_i - gA_i} & \text{if } gA_i < y_i \leq W_i \end{cases} \quad , i = 1, 2, \ldots, \quad (1)$$

with $A_1 = P$. Moreover, in case of withdrawals exceeding the guaranteed amount, there is also a proportional penalization on the surplus according to a penalty rate, that we denote by k_i (such that $0 < k_i < 1$). Therefore, the net amount (cash-flow) received by the policyholder at time i is given by

$$B_i^{(s)} = f_i^{(s)}(y_i, A_i) = y_i - k_i \max\{y_i - gA_i, 0\}, \quad i = 1, 2, \ldots \quad (2)$$

The policy account value evolves according to the following equation:

$$W_{i+1} = f_{i+1}^W(W_i, R_{i+1}, y_i) = \max\{W_i - y_i, 0\}(1 + R_{i+1})(1 - \varphi), \quad i = 0, 1, \ldots, \quad (3)$$

where φ (such that $0 < \varphi < 1$) is the insurance fee rate, $W_0 = P$ and $y_0 = 0$. Note that 0 is an absorbent barrier for W because, once it becomes null, it remains so for ever. The contract, however, continues while $A_t > 0$ (and the insured is still alive). Finally, in case of death in the time interval $(i - 1, i]$, the death benefit, paid at time i, is

$$B_i^{(d)} = W_i, \quad i - 1 < \tau \leq i, \quad i = 1, 2, \ldots \quad (4)$$

In case of surrender at time i, i.e., when $y_i = W_i > gA_i$, the contract is automatically closed because (1) and (3) imply $A_t = W_t = 0$ for all $t > i$, hence no further withdrawals are admitted, nor a death benefit will be paid.

3   The Valuation Framework

In this section we introduce our valuation framework and define the optimal withdrawal problem. Consider a filtered probability space $(\Omega, \mathcal{F}, \mathbb{F}, Q)$ supporting all sources of financial and biometric uncertainty, where all random variables and processes are defined. The filtration $\mathbb{F} = (\mathcal{F}_t)_{t \geq 0}$ satisfies the usual conditions of right continuity and completeness, and is such that $\mathcal{F}_0$ is Q-trivial. Q is a risk-neutral probability measure selected by the insurer, for pricing purposes, among the infinitely many equivalent martingale measures existing in incomplete arbitrage-free markets. In this setting, the residual lifetime of the policyholder τ is a stochastic $\mathbb{F}$-stopping time and let $\mu_t := \mu_{x+t}(t)$ be the mortality intensity

which determines the probability of death at time t conditional on survival for the policyholder aged x at time 0. Moreover, we suppose there is independence between financial- and biometric-related variables. In this general framework we can consider for μ any (reasonable) Markovian process and denote by

$$p_i(\mu_i) = Q\left(\tau > i+1 \,|\, \tau > i, \mu_i\right) = \mathbb{E}^Q\left[\mathrm{e}^{-\int_i^{i+1} \mu_u\, \mathrm{d}u}\,|\mu_i\right], \quad i = 0, 1, \ldots, \quad (5)$$

the probability of survival up to $i+1$ for the policyholder still alive at age $x+i$ given the mortality intensity's values up to i. Consequently, $q_i(\mu_i) = 1 - p_i(\mu_i)$ is the probability of death before $i+1$ conditional on survival at time i. Concerning the financial uncertainty, we assume the instantaneous interest rate to be deterministic and constant, and denote it by r. The reference price S, instead, can be any Markovian process whose discounted value is a martingale under Q. Consider now a withdrawal strategy $y = (y_i)_{i\in\mathbb{N}^+}$, where y_i denotes the actual withdrawal made at time i (in case of survival). This is a stochastic process, adapted to the filtration $\mathbb{F}$, because at each withdrawal date the policyholder takes her withdrawal decision once she knows the values of all state variables. This strategy is *admissible* if it belongs to the set of admissible withdrawal strategies $Y = (Y_i)_{i\in\mathbb{N}^+}$, where $Y_i = [0, \max\{W_i, gA_i\}]$. Then we define the initial value of the GLWB variable annuity as the solution of the following optimization problem:

$$V_0 = \sup_{y\in Y} \mathbb{E}^Q\left[\sum_{i=1}^\infty \mathrm{e}^{-ri}\left(\mathbf{1}_{\{\tau>i\}} f_i^{(s)}(y_i, A_i) + \mathbf{1}_{\{i-1<\tau\leq i\}} W_i\right)\right], \quad (6)$$

where the account value and the benefit base satisfy (3) and (1) respectively. Hence the policyholder is assumed to maximize the present expected value, under Q, of all the future cash-flows generated by the VA contract.

4 Dynamic Programming

In this section we implement a dynamic programming algorithm for discrete stochastic control problems to solve (6). In particular, as we act in a Markovian framework, for each i we denote by $V_i(W_i, A_i, \mu_i)$ the contract value at time i (before the periodic withdrawal) and by $v_i(W_i, A_i, \mu_i)$ the contract value at the same time when, moreover, the policyholder is then alive. Clearly $V_i(W_i, A_i, \mu_i) = \mathbf{1}_{\{\tau>i\}} v_i(W_i, A_i, \mu_i)$ and $V_0 = V_0(P, P, \mu_0) = v_0(P, P, \mu_0)$.

Since the algorithm proceeds backward, we need a starting point. To this end, we assume that there is an ultimate age for the policyholder beyond which her survival probability is null. We denote by ω this age, that typically is in the range 110-120 years, and let $n = \max\{i \in \mathbb{N} : \omega - x \leq i+1\}$, hence $n < \omega - x \leq n+1$. Then Eq. (5) is valid only for $i < n$, while $p_i(\mu_i) \equiv 0$ for $i \geq n$. Therefore, the optimal problem (6) can be rewritten as

$$V_0 = \sup_{y\in Y} \mathbb{E}^Q\left[\sum_{i=1}^n \mathrm{e}^{-ri}\left(\mathbf{1}_{\{\tau>i\}} f_i^{(s)}(y_i, A_i) + \mathbf{1}_{\{i-1<\tau\leq i\}} W_i\right) + \mathrm{e}^{-r(n+1)}\mathbf{1}_{\{\tau>n\}} W_{n+1}\right]$$

$$(7)$$

We take $n+1$ as starting point of our backward dynamic algorithm, and define the following terminal condition:

$$v_{n+1}(W_i, A_i, \mu_i) \equiv 0. \tag{8}$$

Then we proceed backward and, for $i = n, n-1, \ldots, 1$, we define the Bellman recursive equation of the problem as follows:

$$v_i(W_i, A_i, \mu_i) = \sup_{y_i \in Y_i} \left(f_i^{(s)}(y_i, A_i) + q_i(\mu_i) \max\{W_i - y_i, 0\}(1 - \varphi) + \tag{9}\right.$$

$$\left. \mathbb{E}^Q \left[e^{-\int_i^{i+1} \mu_u \, du} v_{i+1} \left(f_{i+1}^W(W_i, R_{i+1}, y_i), f_{i+1}^A(W_i, A_i, y_i), \mu_{i+1} \right) e^{-r} | W_i, A_i, \mu_i \right] \right).$$

Finally, the initial contract value is given by

$$v_0(P, P, \mu_0) = q_0(\mu_0)P(1 - \varphi) + \mathbb{E}^Q \left[e^{-\int_0^1 \mu_u \, du} v_1(P(1 + R_1)(1 - \varphi), P, \mu_1) e^{-r} \right]. \tag{10}$$

4.1   Bang-Bang Analysis

At each time step $i = n, n-1, \ldots 1$, Eq. (9) requires to solve a real-valued optimization problem where the domain of y_i is the whole interval $Y_i = [0, \max\{W_i, gA_i\}]$. Moreover, this problem must be solved for every possible triplet of state variables (W_i, A_i, μ_i). Then the computational effort could be substantial. A property that drastically reduces this effort is the *bang-bang* condition, which states that the set of the optimal withdrawals consists of three choices only: zero withdrawal, withdrawal at the contractual amount, complete surrender. Such a condition is satisfied for our problem, indeed the optimal solution of (9) is $y_i = 0$, or $y_i = gA_i$, or $y_i = W_i$.

Now we outline the proof, that can be made by backward induction. First of all, through tedious computations it is easy to show that the function to maximise at step n is a continuous linear spline, defined in the closed interval Y_n, with a single knot given by $\min\{W_n, gA_n\}$, and that its maximizer belongs to the set $\{W_n, gA_n\}$. In addition, the value function at this step takes the form $v_n(W_n, A_n, \mu_n) = C_n(\mu_n)W_n + D_n(\mu_n)gA_n$, where C_n and D_n are two (constant) functions such that $0 \le C_n(\mu_n) < 1$ and $D_n(\mu_n) > 0$. Then, assuming $v_{i+1}(W_{i+1}, A_{i+1}, \mu_{i+1}) = C_{i+1}(\mu_{i+1})W_{i+1} + D_{i+1}(\mu_{i+1})gA_{i+1}$ for $i = n-1, \ldots, 1$, with $0 \le C_{i+1}(\mu_{i+1}) < 1$ and $D_{i+1}(\mu_{i+1}) > 0$ (almost surely), it is easy to show that the function to maximize at step i is a linear spline defined in the closed interval Y_i. This function is discontinuous at 0, where it takes a value strictly greater than its right limit, and continuous in the (only) knot given by $\min\{W_i, gA_i\}$. Hence the conclusion is that its maximizer belongs to the set $\{0, W_i, gA_i\}$ and also at this step the value function takes the form $v_i(W_i, A_i, \mu_i) = C_i(\mu_i)W_i + D_i(\mu_i)gA_i$, with $0 \le C_i(\mu_i) < 1$ and $D_i(\mu_i) > 0$.

4.2   Contract Decomposition

It is clear that the valuation algorithm aimed at producing the contract value under the dynamic approach can be used to obtain, as simplified cases, also the contract values under alternative policyholder behaviours, namely under the static and the mixed[1] approaches. To obtain the value under the static approach it is sufficient to fix $y_i = gA_i$ for any $i = 1, 2, \ldots, n$, without searching any maximum, while to obtain the value under the mixed approach the search of the maximum must be restricted to the subset $\{gA_i, W_i\}$. To distinguish between these three different values we denote them, respectively, by $V_0^{dynamic}$, V_0^{static} and V_0^{mixed}. Then we can see the dynamic contract as the combination of three components: the *basic GLWB contract*, i.e., the static one, the *surrender option* (with value given by $V_0^{surrender} := V_0^{mixed} - V_0^{static}$), and the *roll-up option* (whose value is $V_0^{rollup} := V_0^{dynamic} - V_0^{mixed}$):

$$V_0^{dynamic} = V_0^{static} + V_0^{surrender} + V_0^{rollup}.$$

5   Conclusion

In this paper we have proposed a discrete time model, based on dynamic programming, to price GLWB variable annuities under the dynamic approach within a stochastic mortality framework. We have verified, by backward induction, the bang-bang condition for the set of discrete withdrawal strategies of the model, and offered an interesting contract decomposition. We have considered a quite general set-up, only requiring the Markovian property for the mortality intensity and the asset price processes. However, to keep the curse of dimensionality of our valuation algorithm manageable, we have assumed constant interest rates. Our next step is the numerical implementation of the model by focussing on a square root process for the mortality intensity and an exponential Lévy process for the asset price. Moreover, the inclusion of stochastic interest rates is a challenging topic for future research.

References

1. Milevsky, M.A., Salisbury, T.S.: Financial valuation of guaranteed minimum withdrawal benefits. Insurance Math. Econom. **38**, 21–38 (2006)
2. Steinorth, P., Mitchell, O.S.: Valuing variable annuities with guaranteed minimum lifetime withdrawal benefits. Insurance Math. Econom. **64**, 246–258 (2015)
3. Bacinello, A.R., Millossovich, P., Montealegre, A.: The valuation of GMWB variable annuities under alternative fund distributions and policyholder behaviours. Scand. Actuar. J. **2016**, 446–465 (2016)
4. Azimzadeh, P., Forsyth, P.: The existence of optimal bang-bang controls for GMxB contracts. SIAM J. Finan. Math. **6**, 117–139 (2015)

---

[1] That is withdrawal of the guaranteed amount or complete surrender, see [3].

A Regression Based Approach for Valuing Longevity Measures

Anna Rita Bacinello[1], Pietro Millossovich[1,2], and Fabio Viviano[1,3(✉)]

[1] Department of Economics, Business, Mathematics and Statistics 'B. de Finetti', University of Trieste, Piazzale Europa 1, 34127 Trieste, Italy
bacinel@units.it
[2] Faculty of Actuarial Science and Insurance, Bayes Business School, City, University of London, 106 Bunhill Row, London EC1Y 8TZ, UK
pietro.millossovich.1@city.ac.uk
[3] Department of Economics and Statistics, University of Udine, Via Tomadini 30/A, 33100 Udine, Italy
viviano.fabio@spes.uniud.it

Abstract. This paper addresses the ever-prominent issue of how to evaluate and forecast future longevity dynamics. Indeed, studying the evolution of mortality and/or the cost of longevity risk is a major task for both demographers and actuaries. In contrast to the usual period-based evaluation, we consider the problem of approximating the distribution of future life expectancy with a cohort-based perspective. In particular, we suggest an application of the Least-Squares Monte Carlo approach, which allows to overcome the straightforward nested simulations method. The method is applied to the family of CBDX models, and results and comparisons between different models, males and females, and period and cohort approaches, are presented.

Keywords: LSMC · Longevity risk · Stochastic mortality

1 Introduction

The analysis of mortality, and consequently of the evolution of various longevity indices, is always under study by demographers and actuaries. Indeed, policy makers need to quantify and manage the risks deriving from unexpected changes in mortality, which would have major implications for the financial stability and solvability of insurance companies and pension providers.

In contrast to the usual period-based approach, this paper addresses the problem of approximating the distribution of future life expectancy, and provides a simulation scheme with a cohort-based perspective that depends on the future evolution of mortality obtained by relying on extrapolative methods. In this regard, one contribution can be found in [7], where the so-called SCOPE approach to forecast future life expectancy levels, i.e., by conditioning on specific future mortality scenarios, is introduced. Indeed, forecasting longevity indices

with a cohort-based method requires the computation of conditional expectations for which explicit solutions often do not exist. A simple way to solve this problem would be to rely on a nested simulations approach, which unfortunately becomes readily unmanageable and computationally intensive, especially when life expectancy estimates are needed for different cohorts and when stochastic mortality models with multiple factors are considered. To overcome this drawback, [3] proposes a Taylor-series approximation of the involved conditional expectations.

This work, instead, suggests an application of the well-known Least-Squares Monte Carlo (LSMC) approach firstly introduced in the financial field (e.g., see [6]) and then extensively adopted in the actuarial one. The main idea is to approximate conditional expectations by linear combinations of some basis functions depending on the relevant factors that affect the quantity of interest. Among the most important advantages of this method, we can mention its generality and flexibility; indeed, it can be used with any mortality model, regardless of its complexity. Essentially, the methodology proposed in this paper is based on that described in [1], where the problem of evaluating future life annuities is addressed. Even if here we focus on just life expectancy, this methodology may be adopted also for approximating other longevity measures at future dates for which cohort-based estimations are often replaced by period ones for computational simplicity.

The remainder of the paper is structured as follows: Sect. 2 states the problem and briefly explains the proposed methodology, Sect. 3 illustrates some numerical results and finally, in Sect. 4, we draw some conclusions.

2 Life Expectancy and Computational Framework

The objective of this paper is to analyse the evolution of future life expectancy levels. Indeed, even if previous studies have broadly addressed this problem, the majority of them exploited a period approach, therefore neglecting future mortality improvements. To fill this gap, we propose a methodology that allows to adopt a cohort based perspective without increasing the computational complexity.

To this end, let $\mu_{x,t}$ be the instantaneous death rate for an individual aged x at time t. Then, following [2], we assume that the force of mortality is constant over each year of age and calendar. Hence, denoting by $m_{x,t}$ the central death rate at age x in year t, and $p_{x,t}$ the 1-year survival probability of an individual aged x at time t, it follows that $m_{x,t} = \mu_{x,t}$ and $p_{x,t} = e^{-\mu_{x,t}} = e^{-m_{x,t}}$.

Now, we are interested in estimating the residual lifespan of an individual aged x at a future time $T > 0$. We define the *period life expectancy* measure as follows:

$$e_{x,T}^p = \frac{1}{2} + \sum_{i=1}^{\omega-x} {}_i p_{x,T}, \tag{1}$$

where $_i p_{x,T} = e^{-\sum_{k=0}^{i-1} m_{x+k,T}}$ represents the i-th years survival probability for an individual aged x at time T, computed by considering the age-specific mortality rates at time T, and ω is the ultimate age. It is clear from Eq. (1) that further mortality improvements after time T are ignored.

Therefore, to describe the actual life course of an individual aged x at time $T > 0$, let us introduce the concept of *cohort life expectancy* defined as

$$e_x^c(T) = \frac{1}{2} + \sum_{i=1}^{\omega-x} {_i p_x(T)}, \qquad (2)$$

where $_i p_x(T) = \mathbb{E}_T \left[e^{-\sum_{k=0}^{i-1} m_{x+k,T+k}} \right]$ represents the (conditional) i-th years survival probability for an individual aged x at time T, and $\mathbb{E}_T[\cdot]$ is the conditional expectation given the information available at the future date T. As already mentioned, cohort life expectancy is not as commonly evaluated, unlike its period counterpart, since it requires the calculation of a conditional expectation. Note that both Eqs. (1) and (2) are the discrete versions of period and cohort life expectancy measures given, for instance, in [5].

Forecasting life expectancy at future times requires projections of mortality onto the future. For this reason, we introduce the computational framework on which we build some numerical results. In particular, we make use of stochastic mortality models in order to capture the possible time evolution of mortality, and in this regard we consider the recently introduced CBDX family (see [4]). Hence, let $D_{x,t}$ denote the number of deaths at age x and calendar year t, which is assumed to be Poisson distributed with parameter $E_{x,t} m_{x,t}$, where $E_{x,t}$ denotes the central exposure. Then, according to [4], the central death rate at age x and calendar year t can be modelled as

$$\log m_{x,t} = \alpha_x + \sum_{i=1}^{N} f^{(i)}(x) \kappa_t^{(i)} + \gamma_{t-x},$$

where α_x is a static age parameter, $\boldsymbol{\kappa}_t = \left(\kappa_t^{(1)}, \ldots, \kappa_t^{(N)} \right)$ is the time index, γ_{t-x} incorporates the cohort effects, and $f^{(i)}(x)$ is a known age-modulating function. In particular, [4] considers the case of $N \in \{1, 2, 3\}$ (named CBDX1, CBDX2 and CBDX3, respectively), and proposes as modulating functions $f^{(1)}(x) = 1, f^{(2)}(x) = (x - \bar{x})$ and $f^{(3)}(x) = \left[(x - \bar{x}) - \sigma_x^2 \right]$, where $\bar{x}$ and σ_x^2 represent the mean and variance of the ages in the data. To project mortality into the future, the time indices are assumed to follow a multivariate random walk with drift, while the cohort effect is modelled as a univariate ARIMA model.

2.1   Valuation Procedure

Computing the quantity in Eq. (2) is not a trivial task since explicit expressions do not always exist. In particular, this is the case of the valuation framework previously introduced. For this reason, a straightforward solution would be a

nested simulations scheme. The latter is computationally challenging since it requires a huge number of simulations. An alternative methodology has been proposed by [3], which consists in approximating conditional expectations by Taylor-series expansions. However, also this approach would be time-demanding since multiple simulations sets are needed in order to estimate the involved coefficients. For this reason, on the basis of [1] we adopt a very flexible tool for approximating conditional expectations, i.e. the LSMC method. Indeed, to the best of our knowledge, this methodology has been extensively used in many fields but it has not yet been proposed in the demographic context. The main idea is to express conditional expectations through linear combinations of some basis functions (e.g. simple or orthogonal polynomials) depending on the relevant risk factors that affect the evolution of mortality (in our case, the time indices κ_t and γ_{t-x}), and use regression across simulations against those factors. Hence, we will evaluate Eq. (2) by regression. Moreover, we refer the readers to [1] for more details.

3   Numerical Results

In this Section we provide some numerical results based on the previously introduced framework. In particular, we analyse the evolution of life expectancy with both cohort and period life tables. The analysis considers males and females in England and Wales population. The models have been calibrated on the mortality data over the period 1965–2018 and range of ages 60–89, obtained from the Human Mortality Database, excluding the first and last 5 cohorts to avoid overfitting. We assume that year 2018 is time 0, and that life tables are closed using a log-linear procedure up to the ultimate age $\omega = 120$. Finally, all computations are based on $n = 20000$ trajectories, and the LSMC algorithm exploits as basis functions simple polynomials of order $p = 2$. Under this setting, we analyse the evolution of life expectancy of both males and females aged $x = 65$ at different future times $T = 2019, \ldots, 2053$ (35 years).

Table 1 reports a summary of the distributions of future cohort life expectancy for females at different future times T, obtained by exploiting the different stochastic mortality models. From the table, we can see that each of the proposed models suggests, as expected, an ever increasing life expectancy. In particular, the CBDX1 model provides more optimistic results, while the opposite happens for the CBDX2 model. Moreover, we can appreciate how the uncertainty increases as time passes[1]. All these features can be seen in Fig. 1 that compares the future cohort male and female life expectancy distributions. Figure 1 highlights, first, the increasing uncertainty characterizing the evolution of the longevity metric, and second, gender differences. Indeed, in line with the existing literature, our results depict future life expectancy levels for females constantly above those for males. Finally, in Table 2 we compare cohort and

---

[1] Similar results were obtained for males, not reported here for space considerations.

period[2] approaches. The table shows how, unsurprisingly, the latter approach persistently under-estimates the desired quantities. What is important to notice is, instead, the magnitude of such under-estimation, which may lead public social systems and life insurance companies to under-estimate the related risks.

Table 1. Summary of the future cohort life expectancy distributions for females aged $x = 65$ at (future) times T. LSMC based on 20000×1 simulations with monomials of degree $p = 2$.

T	Model	Mean	Std. dev.	Skew.	Kurt.	10th Perc.	Median	90th Perc.
2019	CBDX1	23.535	0.284	−0.104	3.058	23.170	23.541	23.900
	CBDX2	21.838	0.273	0.026	3.007	21.489	21.835	22.191
	CBDX3	23.215	0.283	0.085	3.081	22.859	23.211	23.580
2039	CBDX1	25.788	1.154	−0.030	3.031	24.305	25.790	27.256
	CBDX2	23.515	1.194	0.063	3.006	21.999	23.495	25.050
	CBDX3	25.292	1.221	0.026	3.028	23.741	25.286	26.863
2053	CBDX1	27.347	1.453	−0.084	2.981	25.467	27.361	29.202
	CBDX2	24.666	1.571	0.072	3.010	22.668	24.646	26.707
	CBDX3	26.700	1.557	0.031	3.012	24.717	26.690	28.714
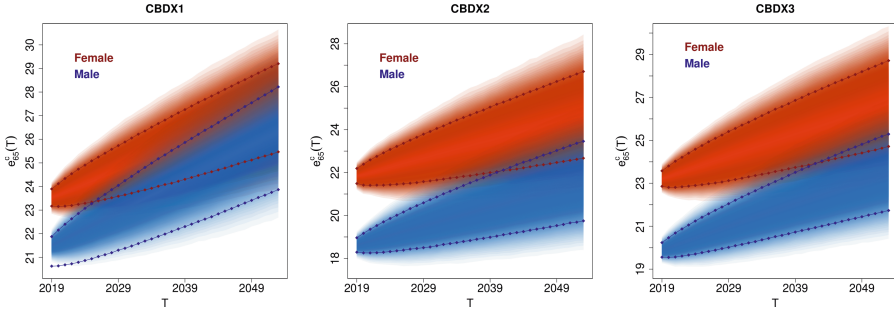
Fig. 1. Distribution of future life expectancy for a cohort of females (red) and males (blue) aged $x = 65$ at (future) times $T \in \{2019, \ldots, 2053\}$. Stochastic mortality models: CBDX1 (left), CBDX2 (centre), CBDX3 (right). LSMC based on 20000×1 trajectories with monomials of degree $p = 2$. Dotted lines represent the 90% prediction intervals.

---

[2] Period life expectancy estimates have been obtained through a simple Monte Carlo (MC) scheme based on 20000 simulations.

Table 2. Expected future cohort and period life expectancy for females aged $x = 65$ at (future) times T. LSMC (cohort) based on 20000×1 simulations and monomials of order $p = 2$. MC method (period) based on 20000 trajectories.

T	Approach	CBDX1	CBDX2	CBDX3
2019	Period	21.584	21.501	21.530
	Cohort	23.535	21.838	23.215
2039	Period	24.001	22.373	23.680
	Cohort	25.788	23.515	25.292
2053	Period	25.510	23.401	25.084
	Cohort	27.347	24.666	26.700

4   Conclusion

In this paper we addressed the ever-prominent issue of how to evaluate and forecast future longevity dynamics, and in particular we focused on life expectancy. We proposed the LSMC approach that allows to adopt a cohort based perspective, rather than a period one, without increasing the computational complexity. Our results proved to be in line with those already presented in literature. To conclude, we want to strengthen the idea that this methodology can be used to estimate any other longevity measure involving conditional arguments, where cohort measurements are often replaced by period ones for computational simplicity.

References

1. Bacinello, A.R., Millossovich, P., Viviano, F.: An efficient Monte Carlo based approach for the simulation of future annuity values. Research Paper DEAMS, n. 2/2021. EUT Edizioni Università di Trieste (2021)
2. Cairns, A.J.G., et al.: A quantitative comparison of stochastic mortality models using data from England and Wales and the United States. North Am. Actuarial J. **13**, 1–35 (2009)
3. Dowd, K., Blake, D., Cairns, A.J.: Facing up to uncertain life expectancy: the longevity fan charts. Demography **47**, 67–78 (2010). https://doi.org/10.1353/dem.0.0083
4. Dowd, K., Cairns, A.J.G., Blake, D.: CBDX: a workhorse mortality model from the Cairns-Blake-Dowd family. Ann. Actuarial Sci. **14**, 445–460 (2020)
5. Guillot, M.: Period versus cohort life expectancy. In: Rogers, R., Crimmins, E. (eds.) International Handbook of Adult Mortality. International Handbooks of Population, vol. 2, pp. 533–549. Springer, Dordrecht (2011). https://doi.org/10.1007/978-90-481-9996-9_25
6. Longstaff, F.A., Schwartz, E.S.: Valuing American options by simulation: a simple least-squares approach. Rev. Financ. Stud. **14**, 113–147 (2001)
7. Vaupel, J.W.: Forecasting life expectancy: the SCOPE approach. In: Bengtsson, T., Keilman, N. (eds.) Old and New Perspectives on Mortality Forecasting. Demographic Research Monographs (A Series of the Max Planck Institute for Demographic Research), 1st edn., pp. 73–77. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-05075-7_6

On the Assessment of the Payment Limitation for an Health Plan

Fabio Baione[1], Davide Biancalana[2(✉)], and Paolo De Angelis[3]

[1] Viale Regina Elena, 295 00161 Rome, Italy
fabio.baione@uniroma1.it
[2] Department of Statistics, Sapienza University of Rome, Rome, Italy
davide.biancalana@uniroma1.it
[3] Department of Methods and Models for Economics Territory and Finance,
Sapienza University of Rome, Rome, Italy
paolo.deangelis@uniroma1.it

Abstract. The study deals with the assessment of an optimal reimbursement strategy for an Health Plan, that covers several branches of disease (e.g. surgery, orthodontia, diagnostic tests, medical devices etc.). We start from the estimation of the expected value and variance of the health expenditure for each branch by a three-part regression model, based on Generalized Linear Models (GLM). Then, we define a reimbursement rule (e.g. deductibles, co-payments, policy limits etc.) by means of the ratio between the per payment values (reimbursement) and the per loss values (expenditure). The latter is the proportion of expenditure reimbursed and is defined as Indicated Deductible Relativity (IDR); an IDR is calculated for each branch covered by the Health Plan.

We apply the optimization problem proposed by De Finetti (1940) in the context of proportional reinsurance to calculate the IDR values, that minimize the variance of the total reimbursement of the Health Plan fixing the expected total gain.

Furthermore, an application is provided for an italian Health Plan in case of uncorrelated branches.

1 Introduction

The Italian National Health System (SSN) is based on three pillars. In particular, the second is mainly characterized by private group health plans (henceforth HP) and usually provided through labor agreements. The HP covers the expenditure for several branches of disease.

In the insurance products deductibles and/or policy limits represent a very relevant coverage modifications contract, which aim is to reduce the premium and to limit the abuse of reimbursement requests, especially in health insurance policies or plans. In particular, our goal is to use the optimization problem introduced by [2] in the context of proportional reinsurance to formalize an optimization problem for the assessment of a set of reimbursement strategies to be applied in an HP.

The first step consists of the prediction of the expected value and variance of the health care expenditure for each branch, by a three part model, following [1]. Then, considering deductibles, copayments and other limitations working on single episode, or single person or family level, we focus on the proportion of expenditure reimbursed, that is defined as Indicated Deductible Relativity (henceforth IDR).

Following [2] and [4], our final goal is the estimate of the optimal vector of IDR (whose length is equal to the number of branches) that minimizes the variance of the expenditure of the HP, fixing the total expected gain.

2   Actuarial Framework

We consider an Health Plan (henceforth HP) composed by r policyholders. In the following, we denote with:

In classical risk theory, given a specific branch j, with $1 \leq j \leq J$ the expenditure of the single policyholder i, with $1 \leq i \leq r$, is:

$$Z_j^{(i)} = \sum_{v=1}^{N_j^{(i)}} Y_{j,v}^{(i)} \tag{1}$$

where

- $Y_{j,v}^{(i)}$ is the r.v. expenditure for the i-th policyholder, j-th branch and v-th episode;
- $Z_j^{(i)}$ is the r.v. expenditure for the i-th policyholder and j-th branch;
- $N_j^{(i)}$ is the r.v. number of episodes for for the i-th policyholder and j-th branch.

It is worth noting that

$$N_j^{(i)} = \sum_{v=1}^{N^{(i)}} \mathbf{1}\left(B_v^{(i)} = j\right), \tag{2}$$

where, $\mathbf{1}(B_v^{(i)} = j)$ is the indicator r.v. for the event the episode v of the i-th policyholder belongs to the branch j, and B the r.v. branch of the requested episode. Hence, the total expenditure for single policyholder is:

$$Z^{(i)} = \sum_{j=1}^{J} Z_j^{(i)} \tag{3}$$

Under the standard actuarial assumption:

- $N_j^{(i)} \perp\!\!\!\perp Y_{j,v}^{(i)}, \forall v;$
- $Y_{j,v}^{(i)} i.i.d., \forall v;$

it is possible to state:

$$E\left[Z_j^{(i)}\right] = E\left[N_j^{(i)}\right] \cdot E\left[Y_j^{(i)}\right], \tag{4}$$

$$VAR\left[Z_j^{(i)}\right] = E\left[N_j^{(i)}\right] \cdot VAR\left[Y_j^{(i)}\right] + VAR\left[N_j^{(i)}\right] \cdot E\left[Y_j^{(i)}\right]^2 \tag{5}$$

Assuming independency between all risks insured, the expected value and variance of the expenditure for single branch are:

$$E\left[Z_j\right] = \sum_{i=1}^{r} E\left[Z_j^{(i)}\right], \tag{6}$$

$$VAR\left[Z_j\right] = \sigma_j^2 = \sum_{i=1}^{r} VAR\left[Z_j^{(i)}\right]. \tag{7}$$

Then, assuming that $C_j = E\left[Z_j\right] + m_j$ the known contribution paid for j-th branch with m_j the risk loading for branch j, the gain for the j-th branch is $G_j = C_j - Z_j$ and the expected total gain for the HP is:

$$E\left[G\right] = \sum_{j=1}^{J} E\left[G_j\right] = \sum_{j=1}^{J} E\left[m_j\right], \tag{8}$$

In order to assess the reimbursement amount,we state that the r.v. reimbursement for single episode $L_{j,v}^{(i)}$, i.e. after the application of the payment limitation is written as a function (a risk-sharing function) $h\left(L_{j,v}^{(i)}; \Theta\right)$, with Θ a set of admissible coverage modification rules. By the way of example, we introduce a standard reimbursement rule applied by Italian HPs based on the following notation:

- f_j is the deductible for single episode for j-th branch;
- M_j is the policy limit for single episode for j-th branch;

   then:

$$L_{j,v}^{(i)} = h\left(L_{j,v}^{(i)}\right) = min\left\{1\left(Y_{j,v}^{(i)} > f_j\right) \cdot Y_{j,v}^{(i)}; M_j\right\} \tag{9}$$

It is worth noting that Eq. (9) represents one of the possible reimbursement rules and other forms of reimbursement may be considered (e.g. ordinary deductible, coinsurance, copayments, out-of-pocket limit).

   In order to assess the influence of the deductibles and limits on the loss of the company, it is possible to focus on the following r.v. $R_{j,v}^{(i)} = \frac{L_{j,v}^{(i)}}{Y_{j,v}^{(i)}}$, that is the proportion of expenditure reimbursed or IDR for a single insured, branch and episode. The latter can be also considered at overall branch level:

$$R_j = \frac{\sum_{i=1}^{r} \sum_{v=1}^{N_j^{(i)}} L_{j,v}^{(i)}}{\sum_{i=1}^{r} \sum_{v=1}^{N_j^{(i)}} Y_{j,v}^{(i)}} \tag{10}$$

3   The Optimal Reimbursement Problem

Starting from Eq. (10), the HP can reduce the reimbursement by setting a reimbursement rule at episode level by introducing a "target" IDR at branch level, as $0 \leq R_j \leq 1$ limiting the benefit and reducing contribution of the same percentage.

Assuming a set of target IDR $R_j = a_j \in [0, 1], j = 1, \ldots, J$, the r.v. gain after the application of reimbursement rule for the j-th branch is $G_j^* = a_j (C_j - Z_j)$. Then, the expected gain of the HP after reimbursement rule is:

$$E(G^*) = \sum_{j=1}^{J} E[G_j^*] = \sum_{j=1}^{J} a_j \cdot m_j \tag{11}$$

and the variance is :

$$VAR(G^*) = \begin{array}{l} \sum_{j=1}^{J} a_j^2 \cdot \sigma_j^2, \qquad \text{uncorrelated branches} \\ \sum_{j=1}^{J} \sum_{w=1}^{J} a_w \cdot a_j \cdot \sigma_{wj}, \quad \text{correlated branches} \end{array} \tag{12}$$

Following [2], it is possible to define an optimal reimbursement strategy based on the following minimization problem:

$$\min_{a_1 \ldots a_J} VAR(G^*)$$
$$\sum_{j=1}^{J} a_j \cdot m_j = \hat{G} \tag{13}$$
$$0 \leq a_j \leq 1, \forall j$$

The strategy is based on finding the optimal IDR (a_j^*) for each branch, in such a way that the total variance is minimal, fixing the target gain after the application of reimbursement rule.

The optimization problem (13) is the same proposed in [2], where a solution for the optimal retention problem in reinsurance market is provided. Let A^* be the set of the optimal percentages, it corresponds to a path in the J-dimensional cube that connects the natural starting point, the vertex 1 of full retention, to the opposite vertex 0 of full reduction. In other words, gradually making choices of increasing reduction of the gain, the IDRs of the branches gradually decrease from 1 (full retention) to 0 (full reduction). This reduction takes place for each branch one by one, choosing the branch for which $F_j(a_1, \ldots, a_J) = \sum_{w=1}^{J} \frac{\sigma_{jw}}{m_j} a_w$ is maximum. The latter represents the measure of the advantage coming from a small reduction of the j-th branch. In Fig. 1, it is possible to see an example of optimal path in the case of only three branches.
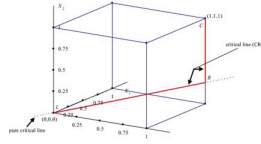
Fig. 1. Example of optimal path in case of three branches

By the way of example, we focus on the case of uncorrelated branches; in this case is $\sigma_{jw} = 0$ for $j \neq w$, then $F_j(a_1, \ldots, a_J) = \frac{\sigma_j^2}{m_j} a_j$.

In this case is simple to order the branches by the ratios $\frac{\sigma_j^2}{m_j}$. Following [2], it is possible to show that, for any $0 \leq \lambda \leq \max_j F_j(1) = \max_j \frac{\sigma_j^2}{m_j}$, an optimal retention is given by

$$a_j^* = min\left(\frac{\sigma_j^2}{m_j}; 1\right) \tag{14}$$

In case of correlated branches de Finetti suggests that the properties of the optimum path fully mimic the pattern characterizing the no-correlation case. The only difference is that we lack an a priori ordering of the risks because of a more complicated expression of $F_j(a_1, \ldots, a_J) = \sum_{w=1}^{J} \frac{\sigma_{jw}}{m_j} a_w$. An extension of de Finetti's algorithm to the general case is provided in [4], where the conditions for regularity of an optimum path are given.

4   Numerical Investigation

We set our framework on a database from an Italian HP between years 2009 and 2013. The portfolio has $r = 53{,}984$ policyholders. The number of observed episodes is 341,494 spread in $J = 5$ branches. Following [1], in order to estimate $E[Z_j]$ and $VAR[Z_j]$ we need to start by a specific probabilistic structure for the main r.v.s:

- N_i is Negative Binomial distributed;
- $B_v^{(i)}$ multinomial distributed;
- $Y_j^{(i)}$ is Gamma distributed;

In order to get an estimate of mean and variance conditioned to insurer i and branch j, a possible choice is the introduction of a dependency structure between the response variables and a set of covariates by means of a regression model. Considering the features of the r.v.s previously introduced, Generalized Linear Models (see [3]) (GLM) seem an appropriate choice.

Table 1. GLM estimate in the uncorrelated case.

Branch	C_j	$E[Z_j]$	m_j	σ_j^2	$\frac{\sigma_j^2}{m_j}$
Hospitalization class 1	1,239.90	1,127.18	112.72	10,635.00	94.35
Hospitalization class 2	357.34	324.85	32.49	2,722.00	83.79
Stomathologic diseases	22,513.81	20,467.10	2,046.71	168,940.00	82.54
Chronic diseases	850.93	773.57	77.36	6,318.00	81.67
Hearing aid	291.82	265.29	26.53	2,157.00	81.31
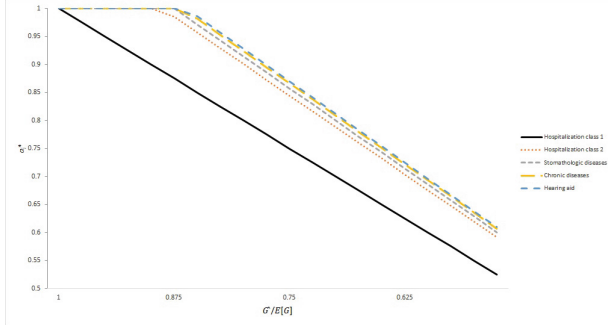
Fig. 2. Optimal a_j^* by varying the reduction of expected gain: uncorrelated case

It is important to know that GLM provides an estimate of mean and conditional variance for each r.v., then following Eq. (4), (5), (6), (7), an estimate of mean and variance for each branch is obtained. In Table 1 the mean and variance estimates provided by GLMs are reported ordering the branches by the ratio $\frac{\sigma_j^2}{m_j}$, in the uncorrelated case.

Finally, in Fig. 2 the outcome of the minimization problem (13) is reported. In the graph are reported the values of the optimal a_j^*, by varying the ratio between the target expected gain $\hat{G}$ after reimbursement rule and the expected gain $E[G]$. As one can see, by reducing the expected gain $\hat{G}$, the branches reimbursement are reduced following the order defined by the ratio $\frac{\sigma_j^2}{m_j}$.

5   Conclusions

The study deals with the assessment of an optimal reimbursement strategy for an HP, that covers several branches of disease. We estimate by a three-part model the mean and variance of the expenditure per single branch, and we consider these outputs to define an optimal strategy to fix the percentage of reimbursement, in such a way that the total variance of the HP is minimal fixing the reduction of expected gain.

We have applied the strategy for an Italian HP, with the assumption of no correlation among branches. Reducing the expected gain , the branches are

reduced one by one following the decreasing order imposed by the ratio $\frac{\sigma_j{}^2}{m_j}$; in our application the first branch to be reduced is the Hospitalization class 1 and the last is the Hearing aid.

References

1. Baione, F., Biancalana, D., De Angelis, P.: A risk based approach for the solvency capital requirement for health plans. In: Corazza, M., Gilli, M., Perna, C., Pizzi, C., Sibill, M. (eds.) Mathematical and Statistical Methods for Actuarial Sciences and Finance, pp. 63–69. (2021). https://doi.org/10.1007/978-3-030-78965-7_11
2. De Finetti, B.: Il problema dei pieni. Giornale dell'istituto italiano degli attuari (1940)
3. Mc Cullagh, P.A., Nelder, J.A.: Generalized Linear Models. Taylor & Francis, Abingdon (1989)
4. Pressacco, F., Serafini, P.: The origins of the mean-variance approach in finance: revisiting de Finetti 65 years later. Decis. Econ. Finance **30**, 19–49 (2007). https://doi.org/10.1007/s10203-007-0067-7

Reference Dependence in Behavioral Portfolio Selection

Diana Barro, Marco Corazza, and Martina Nardon[(✉)]

Department of Economics, Ca' Foscari University of Venice,
Sestiere Cannaregio 873, 30121 Venice, Italy
{d.barro,corazza,mnardon}@unive.it

Abstract. In this contribution, we address the issue of reference dependence within a behavioral portfolio model defined under Cumulative Prospect Theory. In such a framework, an investor selects the portfolio weights in order to maximize her prospect value, where portfolio returns are measured as deviations from a certain reference point. The location of this reference point affects actual investment decisions. We consider alternative hypothesis and perform an application to the European equity market.

1 Introduction

Reference dependence is an essential element in Prospect Theory (PT) [4]. A main difference between PT and Expected Utility (EU) theory relies on the fact that, under PT, investment opportunities are evaluated not in terms of final wealth, but based on potential gains and losses; outcomes are defined relative to a specific *reference point*. According to PT, individuals have different risk attitudes toward positive and negative returns: they are risk-averse for gains and risk-seeking in the domain of losses, and are more sensitive to losses than gains of comparable magnitude (they display loss aversion). The utility function is replaced by a *value function*, which is assumed continuous and strictly increasing. In correspondence with the reference point, the slope of the value function is discontinuous: above the reference point, the function is concave, as a conventional utility function; at the reference point, it displays a kink; below the reference point, it is convex and steeper.

Zero is usually taken as a reference point (the *status quo*, the individual's point of comparison to alternative scenarios), even though PT does not clarify how to locate such a value (see [5] and [7]). Its position is determined subjectively by the decision-maker, and the investment decisions will depend on the perception of the outcomes.

In this contribution, we investigate how the positioning of the reference point, r_0, may affect portfolio selection and performances within the behavioral portfolio (BP) model proposed by [1]. In particular, our goal is to analyse the optimal portfolio behavior when we remove the constant assumption about r_0 and we consider alternative criteria to set such a reference point. We present some applications to the European equity market.

The remainder of this paper is organized as follows. Section 2 synthesises the behavioral portfolio model. In Sect. 3 alternative hypotheses about the reference point definition are introduced. Section 4 presents empirical results.

2   Behavioral Portfolio Selection

We apply the BP model proposed in [1] based on Cumulative Prospect Theory (CPT) [6]. A prospect investor (PI) is risk-averse for positive outcomes and risk-seeking for negative outcomes and is loss averse. PI applies decision weights that are biased with respect to objective probabilities and are more sensitive to changes in the probability of extreme outcomes than mid outcomes; medium and high probabilities tend to be underweighted, and low probabilities of extreme outcomes are overweighted. Risk attitude and loss aversion are modeled through a value function v and probabilistic risk perception through a probability weighting (or probability distortion) function w. These functions are the two main foundations of CPT, and actual choices depend on their shape and interaction.

A PI maximizes the prospect value $V = \sum_{i=-m}^{n} \pi_i \cdot v(z_i)$, where z_i denotes negative outcomes for $-m \le i < 0$ and positive outcomes for $0 < i \le n$, with $z_i \le z_j$ for $i < j$.

In the application, we adopt the following value function:

$$v(z) = \begin{cases} v^+(z) = z^a & z \ge 0 \\ v^-(z) = -\lambda(-z)^b & z < 0, \end{cases} \tag{1}$$

with positive parameters that control risk attitude, $0 < a \le 1$ and $0 < b \le 1$, and loss aversion, $\lambda \ge 1$. Function (1) is widely used in the literature, it is continuous, strictly increasing and with the above mentioned features[1].

In CPT, the prospect value V depends also on the rank of the outcomes and the decision weights π_i are differences in transformed (through the probability distortion function) counter-cumulative probabilities of gains and cumulative probabilities of losses. The probability weighting function w is a strictly increasing function which maps the probability interval $[0, 1]$ into $[0, 1]$, with $w(0) = 0$ and $w(1) = 1$. Empirical evidence suggests a typical *inverse-S shape*: the function is initially concave (probabilistic risk seeking or optimism) for probabilities in the interval $(0, p^*)$, and then convex (probabilistic risk aversion or pessimism) in the interval $(p^*, 1)$, for a certain value of p^*. As for the reference point, PT does not indicate clearly when high probabilities should be considered extremely high probabilities (see [5]) and where elevation of the weighting function is precisely located. Various 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 which indicate that the intersection (elevation) between the weighting function and the 45° line, $w(p) = p$, is for p in the interval $(0.3, 0.4)$.

---

[1] In the numerical experiments, we use the parameters estimated by Tversky and Kahneman [6]: $\lambda = 2.25$ and $a = b = 0.88$ (referred to as TK sentiment).

A commonly applied weighting function is the following:

$$w(p) = \frac{p^\gamma}{(p^\gamma + (1-p)^\gamma)^{1/\gamma}}, \tag{2}$$

with $w(0) = 0$ and $w(1) = 1$, and $\gamma > 0$ (with some constraint in order to have an increasing function). When $\gamma < 1$, w displays the inverse-S shape[2].

In the BP model introduced in [1], a PI selects the portfolio weights in order to maximize the prospect value subject to the usual budget constraint and short selling restrictions. Let $\mathbf{x} = (x_1, \ldots, x_n)$ be the vector of portfolio weights, such that $x_j \geq 0$ $(j = 1, 2, \ldots, n)$ and $\sum_{j=1}^{n} x_j = 1$. Let us consider m possible scenarios, with r_{ij} the return of equity j in scenario i, and p_i be the probability of each i. In this work we considered equally probable scenarios. The reference point, denoted without loss of generality with r_0, will be further discussed and detailed in Sect. 3.

The portfolio returns, measured relative to r_0, are subjectively evaluated and weighted through the distorted probabilities. Formally, the BP selection model is defined as:

$$\max_{\mathbf{x}} \sum_{i=1}^{m} \pi_i \cdot v \left(\sum_{j=1}^{n} (x_j r_{ij} - r_0) \right)$$
$$\text{s.t.} \quad \sum_{j=1}^{n} x_j = 1 \tag{3}$$
$$x_j \geq 0, \quad j = 1, 2, \ldots, n.$$

The resulting optimization problem is highly non-linear and non-differentiable, hence we resort to an evolutionary metaheuristic, Particle Swarm Optimization (PSO). PSO is an iterative bio-inspired population-based approach for the solution of global unconstrained optimization problems. In order to take into account the presence of constraints, we reformulate problem (3) into an unconstrained one using a nondifferentiable penalty function method already applied in the financial context. Such an approach is known as *exact penalty method*, where the term "exact" refers to the correspondence between the optimizers of the original constrained problem and the optimizers of the unconstrained (penalized) one.

The reformulated version of BP optimization problem (3) is then

$$\max_{\mathbf{x}} \sum_{i=1}^{m} \pi_i \cdot v \left(\sum_{j=1}^{n} (x_j r_{ij} - r_0) \right) - \frac{1}{\epsilon} \left[\left| \sum_{j=1}^{n} x_j - 1 \right| + \sum_{j=1}^{n} \max(0, -x_j) \right], \tag{4}$$

where ϵ is the so-called penalty parameter. Note that a correct setting of ϵ ensures the correspondence between the solutions of the original constrained problem and problem (4).

---

[2] In the applications, we use the (TK) parameters estimated by [6]: $\gamma^+ = 0.61$ and $\gamma^- = 0.69$, for w^- and w^+, respectively.

3   The Reference Point

The reference point determines how an outcome is perceived by the investor. Nevertheless, PT does not provide specific indications on what defines the location of the reference point. A standard assumption for applications of PT is a constant reference point, usually set at zero. For monetary outcomes, the *status quo* is usually considered as the reference point, separating gains from losses. For instance, [3] adopt as reference point a benchmark wealth. In the BP model (4), outcomes are defined in terms of deviations of the portfolio (relative) returns from a constant reference point, which in [1] is set equal to $r_0 = 0\%$, and $r_0 = 2.5\%$. [2] consider outcomes as excess return on a risky asset over the risk-free interest rate.

It is worth noting that the assumption of a constant reference point identifies an absolute parameter with respect to which all possible outcomes are evaluated. However, investor attitude to gains and losses and their evaluation can reflect or be influenced by the context, as for example the current market conditions, resulting thus more in a valuation with respect to a relative reference point rather than an absolute one.

In this contribution, we aim at analyzing the impact of the introduction of a relative reference point and, specifically, we consider a reference benchmark representative of a specific market. The BP portfolio model (4) can easily handle a stochastic reference point that can reflect the risk/return profile that an investor intends to track and with respect to which deviations (gains and losses) are evaluated and weighted. To this aim, in (4) the reference point r_0 is specified through a discrete-time stochastic process characterized through a finite number of scenarios that describe possible future outcomes of the reference benchmark $\tilde{r}_i$, with $i = 1, \ldots, m$. Thus, given the scenario description of market evolution, gains and losses are evaluated and weighted with reference to each specific scenario.

A further interesting extension of this model allows for the introduction of a combination of a fixed (absolute) reference and a variable (stochastic) one in the form $r_0 = \max\{0, \tilde{r}_i\}$. With this specification, each outcome is evaluated with respect to the maximum between the fixed reference 0% and the variable market outcomes $\tilde{r}_i$.

These reference point specifications aim at accounting for the market conditions in which the investor operates and with respect to which evaluations are made. Rather than considering a discrete process with finite possible outcomes, we can also account for relative market conditions choosing as reference point significant statistics that describe the market behavior and, specifically, we can set the reference point as the mean or specific percentiles of the distribution of the reference benchmark.

4   An Application

To address the issue of reference dependence, we tested the proposed BP model (4) in an out-of-sample exercise on weekly data from the European equity market represented by the STOXX 600 Europe index.

The out-of-sample analysis is performed over the period that goes from July, 2018 to June, 2019, and is carried out as follows: a 1-year in-sample period is used to select the optimal portfolios with different reference point specifications, then the resulting optimal portfolios are evaluated in a mark-to-market procedure using realized weekly market returns for the subsequent 3-month out-of-sample period (i.e. first in-sample period, from July, 2017 to June, 2018; first out-of-sample period, from July, 2018 to September, 2018). This scheme is then rolled over for 4 times to cover the entire 1-year out-of-sample testing period.

Our purpose is to compare the performances of different investment decisions, induced by alternative definitions of the reference point, using out-of-sample mark-to-market evaluated portfolios. The relative reference points considered are: the stochastic reference point $\tilde{r}_i$, for which we use the market benchmark STOXX 600 Europe index; the enhanced stochastic reference point $\max\{0, \tilde{r}_i\}$; the mean μ_s of the market benchmark returns computed over the m scenarios generated in the considered 3-month out-of-sample period; its enhanced version $\max\{0, \mu_s\}$; and the 50%, 66.67% and 75% percentiles of the distribution of the stochastic reference.

Figure 1 shows the out-of-sample equity lines for the optimal portfolios and for the STOXX 600 Europe index (Index), using as starting portfolio value $C = 100$; Table 1 summarizes the main statistics of the portfolios' out-of-sample returns. The volatility of the market strongly affects the out-of-sample performances of the optimal portfolios, which behave similarly. The BPs for the enhanced reference cases, $\max(0, \tilde{r}_i)$ and $\max(0, \mu_s)$, perform better than the ones corresponding to the reference values $\tilde{r}_i$ and μ_s. Compare also the Sharpe ratios in columns 3–4, and 5–6 of Table 1, respectively. With regards to the different selected reference points, we can notice that the performance of the portfolios improves when we increase the considered percentiles. In particular, in the cases displayed in Fig. 1, we can observe that the behavior of the optimized portfolios improves moving from 50 % to 66.67 % and 75 % percentiles. This is also shown by the increasing values of the means and of the Sharpe ratios in columns 7 to 9 of Table 1, with substantially equal volatilities associated to the considered percentiles. Lastly, all the out-of-sample equity lines show negatively skewed distributions (see the skewnesses reported in Table 1). Likely, this is due to the difficulty of the BPs to perform well, with respect to the chosen reference points, in the pronounced up-and-down trends characterizing the considered period of the investigated market.
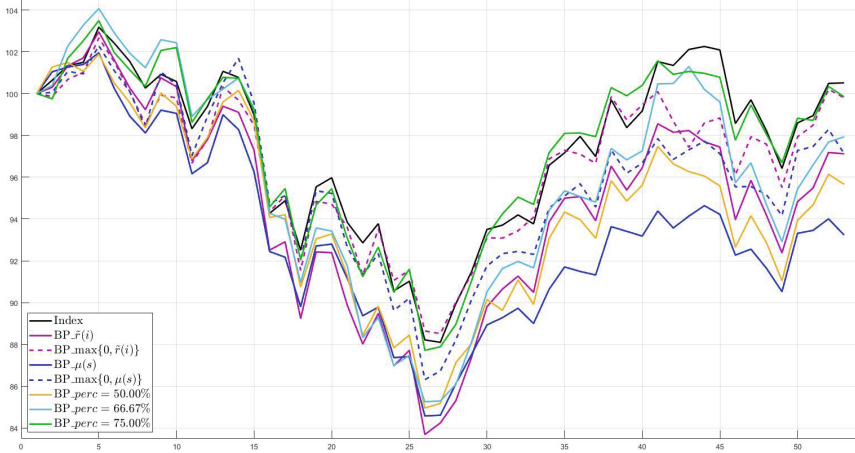
Fig. 1. Out-of-sample equity lines of the optimal BPs for different reference points and the Index

Table 1. Statistics for the out-of-sample returns of the optimal BPs for different reference points r_0 and the Index as benchmark. From left to right: Index, $\tilde{r}_i$, $\max(0, \tilde{r}_i)$, μ_s, $\max(0, \mu_s)$, $perc = 50\%$, $perc = 66.67\%$, $perc = 75\%$.

	$Index$	$\tilde{r}_i$	$\max(0, \tilde{r}_i)$	μ_s	$\max(0, \mu_s)$	50%	66.67	75%
Mean	0.0000	-0.0004	0.0001	-0.0012	-0.0004	-0.0007	-0.0002	0.0001
Stand. Dev.	0.0174	0.0203	0.0182	0.0157	0.0185	0.0188	0.0189	0.0180
Skewness	-0.4685	-0.4764	-0.2645	-0.1720	-0.4381	-0.3350	-0.5402	-0.7443
Kurtosis	3.0281	2.7912	2.6274	2.9328	3.0962	2.6706	3.0447	3.0637
Sharpe ratio	-0.0003	-0.0176	0.0068	-0.0780	-0.0212	-0.0360	-0.0119	0.0072

References

1. Barro, D., Corazza, M., Nardon, M.: Behavioral aspects in portfolio selection. In: Corazza, M., Gilli, M., Perna, C., Pizzi, C., Sibillo, M. (eds.) Mathematical and Statistical Methods for Actuarial Sciences and Finance, pp. 87–93. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-78965-7_14
2. Bernard, C., Ghossoub, M.: Static portfolio choice under cumulative prospect theory. Math. Finan. Econ. **2**, 277–306 (2010). https://doi.org/10.1007/s11579-009-0021-2
3. He, X.D., Zhou, X.Y.: Portfolio choice under cumulative prospect theory: an analytical treatment. Manage. Sci. **57**(2), 315–331 (2011)
4. Kahneman, D., Tversky, A.: Prospect theory: an analysis of decision under risk. Econometrica **47**, 263–291 (1979)
5. Shiller, R.J.: Human behavior and the efficiency of the financial system. In: Taylor, J.B., Woodford, M. (eds.) Handbook of Macroeconomics, vol. 1C, pp. 1305–1340. Elsevier (1999)

6. Tversky, A., Kahneman, D.: Advances in prospect theory: cumulative representation of the uncertainty. J. Risk Uncertain. **5**, 297–323 (1992). https://doi.org/10.1007/BF00122574
7. Werner, K.M., Zank, H.: A revealed reference point for prospect theory. J. Econ. Theory **67**, 731–773 (2019). https://doi.org/10.1007/s00199-017-1096-2

Pricing Rainfall Derivatives by Genetic Programming: A Case Study

Diana Barro, Francesca Parpinel, and Claudio Pizzi[✉]

Department of Economics, Ca' Foscari University of Venice, Venice, Italy
{d.barro,parpinel,pizzi}@unive.it

Abstract. In this contribution we consider a genetic programming approach to price rainfall derivatives and we test it on a case study based on data collected from a meteorological station in a city in the northeast region of Friuli Venezia Giulia (Italy), characterized by a fairly abundant rainfall.

Keywords: Genetic programming · Rainfall derivatives

1 Introduction

In the context of the climate change, we are witnessing increasingly meteorological uncertainty and variability, including increasing number of extreme events. Several sectors' revenues are affected by weather, not only agriculture and related sectors but, for example, also tourism, energy production, construction among others, last but not least the financial and insurance sectors themselves can be affected by losses induced by these events.

Weather derivatives, such as rainfall derivatives, represent an important tool to be placed side by side with traditional insurance dealing with meteorological risk as testified by the consistent increase in demand, [13] on the behaviour of derivative prices in response to weather outcomes.

Rainfall derivatives have been introduced at the Chicago Mercantile Exchange (CME) in 2011, relatively few contributions in the literature focused on pricing methodologies and a generally recognised pricing framework is missing. Major difficulties are associated with the modelling of the underlying, for which physical, meteorological and orographics models are involved, see, for example, [2,3]. The time series of rainfall is discontinuous and the binary event of wet or dry days is largely random. Daily data do not exhibit a trend and also seasonality is limited. The nature of the risk underlying the derivative contract, that cannot be replicated, introduces a further complexity in the pricing problem characterizing it as a problem of pricing in incomplete market. However, further testing is needed to assess applicability also to more general cases.

To tackle this problem we refer to the use of Genetic Programming (GP), a population-based search algorithm that belongs to the field of Evolutionary Automatic Programming. GP is a data-driven methodology that requires a very

limited modelling assumptions thus providing flexibility. For applications of GP in finance and economics see, among others, [1] and [5]. The first implementation of Genetic Programming for option pricing is the work by [10] that uses GP to price American put options. Other contributions specifically for derivatives and option pricing are [4,6]. More recently, GP has been used in pricing weather derivatives and specifically rainfall derivatives, see [7–9].

In this contribution we apply GP for the pricing of a strangles strategy that combines a long put and a long call on the same underlying source of risk, thus providing the holder of the contract a two-sided protection from extreme phenomena of excess or lack of rainfalls over a considered one-month period.

The structure of the paper is as follows. In Sect. 2 we briefly present GP technique and in Sect. 3 rainfall derivatives are introduced. Finally, in Sect. 4 a case study is discussed.

2   Genetic Programming

Genetic Programming (GP), introduced by [11], is a population-based search algorithm that extends genetic algorithms to a population of functions where the chromosomes may change their size and describe hierarchically the functions in tree-like structures. GP consists on a set of instructions and a fitness function to measure how well a computer program can performed a task.

We synthetically recall the structure of the GP algorithm. The first step is the generation of a random population of computer codes. The initial population is randomly generated and each individual is composed of primitive functions and terminals pertinent to the problem domain. Through the Darwinian principle of natural selection, according to which only the fittest elements in the population survive and reproduce, and the genetic recombination (crossover operation), the populations of computer programs are muted in terms of genes applying genetic operators (reproduction, crossover and mutation) probabilistically. Then they are evaluated through a fitness measure to register their performance in the specified problem environment. If a certain fitness criterion is obtained the algorithm is terminated and the computer code with the highest fitness is selected as the final result. If the criterion is not reached, the current population is replaced by a new population and the process is repeated till the point at which the desired criterion threshold is reached. Preparatory steps for GP require determining the set of terminals, the set of primitive functions, the fitness measure, the parameters for controlling the run, the termination criterion and the method of result designation for the run.

3   Rainfall Derivatives Pricing

Rainfalls derivatives are traded on rainfalls, usually measured in mm, and used to hedge risks associated with rainfall fluctuations. Modelling the underlying source of risk is particularly challenging and heavily dependent by specific local conditions. However, the pricing of rainfall derivatives presents relevant challenges

associated, not only with respect to the difficulties in modelling the weather events, but also to the lack of traded tools used to build a replicating portfolio and thus resort to standard derivative pricing techniques based on the existence of a unique martingale measure. The incompleteness of the market is a relevant issue and currently there exists no generally accepted pricing framework to value these contracts.

The contract analysed in this contribution is an Over-the-counter (OTC) strangle strategy that combines a long put and a long call on the same underlying source of risk, thus providing the holder a two-sided protection from extreme phenomena of excess or lack of rainfalls over a considered one-month period. The contract is written on an index of cumulated rainfalls in the geographic region considered over a given month. The payoff of the contract, cash settled at maturity, is expressed as a function of the value of the rainfall cumulated index measured at a given location over a given period and the strikes as follows

$$F_t(I_t, \kappa_1, \kappa_2) = \begin{cases} \kappa_1 - I_t & I_t < \kappa_1 \\ 0 & \kappa_1 \leq I_t < \kappa_2 \\ I_t - \kappa_2 & I_t \geq \kappa_1 \end{cases} \tag{1}$$

where I_t denotes the rainfall index and κ_1 and κ_2 are the strikes. The setting of two strikes is based on the ad-hoc features of agricultural and hydrological field.

Future payments of the contract depend on the evolution of the rainfall index I_t and the strikes κ_1 and κ_2. To price the contract we calculate the expected discounted value of payments

$$P_t(I_t, \kappa_1, \kappa_2, t) = \mathrm{E}^{\mathrm{Q}} \left[e^{\rho_t} F_t(I_t, \kappa_1, \kappa_2) | \mathcal{F}_t \right] \tag{2}$$

where ρ is the risk-free rate and E^{Q} denotes the expected value taken with respect to the Q measure.

In this contribution, to model the underlying we resort to GP, in which the data are the cumulative rainfalls computed over 30 days, with a sliding window accumulation method, given by: $I_t = \sum_{t_a}^{t_b} r_t$ where I_t denotes the accumulated amount of rainfall for a given period, r_t is the daily amount of rainfall with the day varying over a contract period from t_a to t_b. Finally, the Q measure, used in the application considered in this contribution, is the real-world probability measure for sake of comparison with other pricing approaches, see Sect. 4.

4   Data and Application

The dataset chosen to apply the genetic procedure is the time series of rainfalls drawn by the meteorological station (142 m above sea level) of Vivaro, a town in the Italian province of Pordenone situated in the southern part of the alluvial cone formed by the Cellina and Medusa rivers. The soil of this region is characterized by a typical karst topography which prevents crops from being subjected to water stress. The station is in an optimal position for the detection of rainy phenomena.

The time series of the daily observations starts from January, 1, 1995 until December, 31, 2021 is available on the official website of *Agenzia Regionale per la Protezione dell'Ambiente* of Friuli Venezia Giulia region (OSMER-ARPA, https://www.osmer.fvg.it/). The data presents a set of 11 missing data, whose size is irrelevant compared to the whole number of observation ($n = 9851$): it corresponds to 0.11%.

Some preliminary comparisons show that there are no differences in the main statistics considering different imputation of these missing values (seasonal imputation, a naive method of substitution for the missing values with the corresponding ones from the previous year, and substitution with data found through reports and collection of hourly data). The main sample statistics (mean, five numbers of Tukey and standard deviation) are shown in Table 1. So it is convenient to ignore the missing data in order to avoid problems with statistical assumptions.

Table 1. Data statistics (01/01/1995 to 31/12/2021)

n	mean	Std.Dev.	min	Q1	median	Q3	max	NA
9851	4.366	12.164	0	0	0	1.8	210.8	11

Looking more carefully at the data, the great part of rainfalls are in between 1.2 and 50.055 mm (respectively the quantiles of order 0.05 and 0.95 of the rainfall days). Furthermore, the recorded data greater than 100 mm are typically during the fall, and the maximum is 210.8 recorded in 2005-09-09. We recall that we can not remove the outliers that in this case coincide with the extreme values on which we base the derivative contracts. The goal of the contracts we consider in our analysis is the protection from heavy rains, that according to the WMO correspond to daily recordings greater than 50 mm. This definition is not unique, for example, according to the *Protezione Civile Italiana* the standard of heavy rains is defined as the daily recordings greater than 60 mm. This difference implies exactly 47 observations. The dataset is then divided in two parts: a training set composed by the first 25 year (from January, 1, 1995 to December, 31, 2019) and the validation set formed by following last 2 years.

Following the notes of FAO (Food and Agriculture Organization), and considering the Penman-Monteith formula in order to include also the evapotranspiration phenomenon, the inferior strike value is found using the series of rainfall data from 2015 to 2018 from Vivaro dataset and this is equal to $\kappa_1 = 74.31\,\mathrm{mm}$ in a month. As regards to the superior strike value, it can be obtained considering the definition of heavy rains, according to WMO, and estimating rain probability curves using the Gumbel random variable. It is found that this value may be considered equal to $\kappa_2 = 166\,\mathrm{mm}$ in a month.

The size of the population is fixed at 200 individuals, the crossover probability is $p_c = 0.9$, and in this first preliminary analysis we do not include the mutation probability. The stopping rule is based on computational time, we set $t_{max} = 2$

min. On an Intel I7, 3 GHz, dual core processor, this results in a number of fitness evaluations over 50 000. We consider 20 replications of the GP and 50 elite functions (i.e. the functions with the better fitness) for each replication. This produces 1000 trajectories eligible to be the best result. The fitness function is the Root Mean Square Error (RMSE) of the observed values with the fitted ones. The obtained result in terms of function is:

$$y_t = 62.06 \times \log \left\{ \log(y_{t-8}) \times \left[1.34 + \cos \left(\frac{y_{t-1}}{y_{t-4}} \right) \right] \right\}$$

As reference benchmarks we consider the actuarial valuation of the contract obtained applying the Historical Burn Analysis (HBA), a data-driven methodology where the price of the derivative is computed as the average of historical payoffs discounted at the risk-free rate [12], and the prices obtained using the Normal distribution, the Normal-inverse Gaussian distribution and the Gamma distribution. In the following table, we present the obtained derivative prices under the different assumptions and computed in absence of the market price for risk.

Table 2. Results of the pricing method

Estimated values	Normal	NIG	Gamma	HBA
17.95	20.59	14.40	19.28	15.20

The obtained prices also in this very preliminary example confirm the relevance and the difficulties associated with the statistical modelling of the rainfalls phenomena and the huge impact that this can have on the resulting price of the derivative. Furthermore, research is needed with reference to the risk-neutral pricing through the estimation of the market price of risk, using for example the Esscher transform, and the sensitivity of the obtained price to the modelling of the market price of risk.

5   Conclusion

In this contribution we aimed at testing the application of genetic programming techniques to the problem of pricing a rainfall derivative contract. The case study considered refers to an OTC long strangle contract that provides two-sided protection from risks determined by periods of insufficient rainfall and risks associated with an excess of rainfall. Compared with other traditional methodologies, GP has the advantage of being a flexible data-driven technique that allows to approach the complexity of the problem limiting the modelling assumptions.

References

1. Banzhaf, W., Nordin, P., Keller, R.E., Francone, F.D.: Genetic Programming: An Introduction on the Automatic Evolution of Computer Programs and Its Application. Morgan Kaufmann Publisher Inc., San Francisco (1998)
2. Cabrera, B.L., Odening, M., Ritter, M.: Pricing rainfall futures at the CME. J. Bank. Fina. **37**(11), 4286–4298 (2013)
3. Carmona, R., Diko, P.: Pricing precipitation based derivatives. Int. J. Theor. Appl. Financ. **08**(07), 959–988 (2005)
4. Chen, S.H., Yeh, C.H., Lee, W.C.: Option pricing with genetic programming. In: Koza, J.R., et al. (eds.) Genetic Programming 1998: Proceedings of the Third Annual Conference, pp. 32-37. Morgan Kaufmann, San Francisco (1998)
5. Chen S-H.: Genetic Algorithms and Genetic Programming in Computational Finance, Kluwer Academic Publishers, New York (2002)
6. Chidambaran, N., Triqueros, J., Lee, C.W.J.: Option pricing via genetic programming, In: Chen, S.H. (eds.) Evolutionary Computation in Economics and Finance. Studies in Fuzziness and Soft Computing, vol. 100, pp. 383–397. Physica, Heidelberg (2002)
7. Cramer, S., Kampouridis, M., Freitas, A.A., Alexandridis, A.: Pricing rainfall based futures using genetic programming. In: 20th European Conference, EvoApplications: European Conference on the Applications of Evolutionary Computation, vol. 10199, pp. 17–33. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-55849-3_2
8. Cramer, S., Kampouridis, M., Freitas, A.A.: Decomposition genetic programming: an extensive evaluation on rainfall prediction in the context of weather derivatives. Appl. Soft Comput. **70**, 208–224 (2018)
9. Cramer, S., Kampouridis, M., Freitas, A.A., Alexandridis, A.: Stochastic model genetic programming: deriving pricing equations for rainfall weather derivatives. Swarm Evol. Comput. **46**, 184–200 (2019)
10. Keber, C.: Option pricing with the genetic programming approach. J. Comput. Intell. Fin. **7**(6), 26–36 (1999)
11. Koza, J.R.: Genetic Programming. MIT Press, Cambridge (1992)
12. Odening, M., Musshaff, O., Xu, W.: Analysis of rainfall derivatives using daily precipitation models: opportunities and pitfalls. Agric. Fin. Rev. **68**, 135–156 (2007)
13. Schlenker, W., Taylor, C.A.: Market expectations of a warming climate. J. Fin. Econ. **142**(2), 627–640 (2021)

Estimation of the Gift Probability in Fund Raising Management

Luca Barzanti[1] and Martina Nardon[2(✉)]

[1] Department of Mathematics, University of Bologna,
Piazza di Porta San Donato 5, 40126 Bologna, Italy
`luca.barzanti@unibo.it`
[2] Department of Economics, Ca' Foscari University of Venice,
Sestiere Cannaregio 873, 30121 Venice, Italy
`mnardon@unive.it`

Abstract. In fund raising management, modeling the gift probability is a key point. The availability of an accurate estimate of the gift probability is relevant in order to evaluate the results of a fund raising campaign. In this contribution, we discuss a statistical methodology for modeling and estimating such quantity. To this aim, a parametric approach is suggested. In particular, we model the number of gifts as a Poisson random variable with intensity parameter which depends on individual characteristics of the Donors. The expected number of donations, and the probability of gift, can then be estimated by performing a Poisson regression.

1 Introduction

Recent innovative approaches to fund raising (FR) are characterized by a significant use of mathematical modeling, which is suitably implemented according to the considered purpose or the particular focus of the process. Along with the modeling approach, the methods differ in the use of advanced mathematical and statistical techniques (probability, linear algebra, utility functions, similarity measures, Choquet integral, non-parametric estimation) or soft computing and artificial intelligence techniques (fuzzy logic, knowledge-based approach).

An innovative approach has been performed in this field by [2], that introduces the use of mathematical modeling and Decision Support Systems (DSS) techniques, in order to help Associations both to decide the kind of campaign they have to organize and the features to implement, and the Donors of the Data Base (DB) which must be contacted, in order to maximize the expected return of the campaign, satisfying time and budget constraints. This quantitative approach has been specialized for different kind of Organizations. On one hand [3] and [7] dealt with large-sized Associations, international also, that have lists of millions of Donors and a powerful organizational system requiring a very sophisticated DSS. On the other hand, [4] consider also small-sized Organizations and developed a DSS based only on essential information with no need of an organized DB. This approach has been validated both in the operational world by

Associations that test it (as documented in [3,4] and [7]) and in the pertaining literature ([18] and [13]). The medium-sized Organizations are considered in [6] and in [5], where a DSS based on a specific mathematical model and targeted for this kind of Associations has been created and enhanced. More in general, a process of evolution, strengthening and specialization of the proposed methods and algorithms has been developed. As for the link of FR with other economic fields, [14] has shown the analogy between the FR process and some bank activities and the consequent correspondence of the employed methodologies.

In FR management modeling the gift probability, as well as the gift amount, is a prominent issue. As regards the available information, Associations are classified according to the existence of a structured DB and the presence in the DB of specific qualitative information of Donors' profile (like personal interests and attitudes, and relationship network), in addition to the usual information on the gifts (gift history), and the typical personal profile (which includes both qualitative and quantitative information). Normally this classification strictly depends on the Organization size.

The availability of accurate estimates of the gift probability is important in order to evaluate the results of a campaign. Despite its key role, this relevant topic has been approached by the business literature and practitioners without an in-depth statistical analysis.

In this contribution, we discuss statistical methodologies for modeling the gift as an individual risk, focusing on the estimation of the gift probability. To this aim, a parametric approach is suggested. In particular, in a first approach we model the number of gifts as a Poisson random variable with intensity parameter which depends on individual characteristics of the Donors (exploiting the information available in large DB of Donors). The expected number of donations, and the probability of gift, can then be estimated by performing a Poisson regression.

2   The Donor

FR strategies are crucial for the achievement of the mission of the Association and, specifically, to reaching the goal of a campaign [16]. In this context, a great importance has the role of the Donor (see e.g. [10] and [12]), and his/her efficient management. For this reason, econometric literature dealt with (potential) Donors' profiles that match some specific gift inclination (see e.g. [9]) in order to support the effectiveness of the process.

Economists agree that information on potential Donors plays a crucial role to achieve the improvement of the FR strategies [15]. Quantitative studies have shown the main factors that influence individuals in the choice of giving. For example, [1] characterizes the economic and social foundations of altruism, individuating factors such as the own community or the social network and the so called "enlightened self-interest". These variables are also modelled by [10] and [17]. [12] argue that an individual tends to assume a role-identity as Donor, that depends on his/her network of social relationship. They identify several variables

that can have impact on role-identity; all these variables influence individual preferences and attitudes, and impact on the utility people get from their decision on how and to what extent donate [8]. More generally, several factors should be considered to individuate an optimal fund raising strategy: the interests of (potential) Donors, their social network and personal profile; the operational literature and rules of thumb of the experts in the field; and the information on past campaigns. Practitioners claim that the 70–80% success rate of a fund raising campaign is determined by choosing the appropriate target (the set of Donors to whom the strategy is addressed), and only 20–30% from motivations and creativity. These factors have a strong influence on the gift probability, which is affected both by individual attitudes and economic constraints [8]: age, instruction level, place of origin, financial situation, number of children, social network and religious involvement. Therefore, the task of integrating all of this information to find an optimal fund raising strategy is very complex.

The effective use of the information on Donors and Contacts (i.e., potential Donors), which is in fact normally managed by an organized DB, is crucial for optimizing the resources for the campaign by selecting the most promising Donors/Contacts for the considered context. However, tools using a classical DB approach are not able to elaborate the knowledge available in the econometric and operational literature as an expert in the field does [9], and they are not able to suggest suitable fund raising strategies. In fact, the problems that these systems can solve are limited by the potential of such a technology. The support to the fund raiser is limited to giving general indications in relation to specific claims without adequately managing all data about people, integrating qualitative information with quantitative data.

As getting in touch with a Donor implies some costs, one main aim of FR management is to select the Donors to contact in order to maximize the expected return of the campaign and at the same time to control the return variability. In this context, the accuracy of the expected gift estimation depends on the appropriate use of the information about each Donor.

3 Modeling the Gift as an Individual Risk

As discussed in the previous sections, assessing a FR campaign expected return is crucial and, to this aim, the estimation, for each Donor, of the gift probability and the expected gift amount is required.

Here we model the 'gift' as an *individual risk* [11], in much analogy with other main domains of applications: finance, credit risk, insurance, and marketing. The gift can be viewed from four perspectives:

  i) the *occurrence* of a donation: the outcome is either 'yes' or 'no';
 ii) the *frequency* or *count* of the donations received in a period of time (e.g., a year or the duration of the campaign): the number of gifts is zero or any positive integer;
iii) the *timing* or *duration*, which is about when a donation has occurred or the interval between donations: the outcome is an interval of time, usually

measured with reference to a fixed point of origin, such as the beginning of the campaign or when the potential Donor has been contacted for the first time;

iv) the *gift amount*: how much money is given by the Donor for each donation (the outcome is measured in currency units, e.g. euros).

With respect to all these perspectives, the gift is quantifiable and can be modeled using statistical methodology, determining for any aspect listed above which kind of random variable is suitable: i) a dichotomous variable, ii) a count variable, iii) a duration variable, and iv) a continuous positive variable, respectively.

Either dichotomous or count variables can be used to model the occurrence of the gift event. Let us simply consider a dichotomous random variable[1] Y; then the probability of gift is equal to $\mathbb{E}(Y) = p$. Let X be a continuous random variable that represents the amount of money given by the Donor for a donation, or the total gift of all donations filed in the considered period. In this case the expected gift for each Donor can be computed by the product of the gift probability and expected gift amount, $\mathbb{E}(Y)\mathbb{E}(X)$. Considering the whole campaign, both the number of gifts and the gift amount are random, hence campaign's return can be modeled as a *random sum*; in order to compute its expectation, some assumptions need to be introduced (such as independence amongst Donors, and independence of gifts count and gift amounts). All these features can be modeled in alternative ways; in Sect. 4 we suggest a model for the number of gifts considering a single Donor.

We make some assumptions about the mechanism that gives rise to the gift: any gift is associated with an individual i, the Donor; a Donor can be a person, a company, or other entity that can be represented by some individual characteristics which are collected in a data set; the individual characteristics of the Donor are synthesized by a *score*; the gift history (gift events, timing and gift amounts) of the Donor is recorded.

A *score* is a statistical measure of individual risk based on individual characteristics [11]. In the context of FR, it can be used to quantify the individual propensity to donate (the higher the score, the higher the propensity to the gift), to rank Donors in a population, to distinguish between (expected) "good" and "bad" Donors. This latter procedure is called *segmentation* and in FR could be used to distinguish potential Contacts or to address *ad hoc* advertising to subclasses of Donors.

Let x_i be the vector which collects selected observable, qualitative and quantitative, individual characteristics of Donor i, in a sample of n Donors. Define z_i as the vector of transformed individual characteristics (where qualitative features are properly transformed into quantitative or dummy variables). The score can be defined as a scalar function of covariates $z_i'\theta$, where θ is a vector of parameters. The score, which summarizes the information about the Donor, can be determined by more sophisticated approaches (see [11]).

---

[1] Formally, denoting with D the gift/donation event, we have $Y = \mathbf{1}_D(\omega)$, where $\mathbf{1}_D$ is the indicator function of D, with $\mathbb{P}[Y = 1] = p$.

4   Poisson Regression in FR

Any event, such as the arrival of a new claim to an Insurer or a donation to an Association, can be viewed as the outcome of a random variable. For instance, a dichotomous variable indicates whether or not a gift is received. In this contribution, we suggest a model for the number of gifts (in a certain period) or, equivalently how frequent gifts are, in much analogy with insurance theory where count variables are used to model e.g. the number of claims on one policy in a year. The number of gifts can be modeled and estimated by the *Poisson regression model*.

Let Y represent the number of gifts in a unit of time; in a basic count variable model, we assume that Y has a Poisson distribution with intensity parameter λ. It is well known that $\mathbb{E}(Y) = \lambda$, which is equal to its variance $\mathbb{V}(Y) = \lambda$.

In the Poisson regression model, λ depends on the values of observable characteristics x_i of each individual or entity i. As the intensity varies across individuals, its specification for Donor i will be

$$\lambda_i = \exp(z_i'\theta), \tag{1}$$

where θ is the vector of unknown parameters and z_i is a vector of transformed individual characteristics; the exponential function allows for positive intensity. The score $z_i'\theta$ can be used for rating Donors with respect to their propensity to the gift; the higher the score is, the higher the expected number of gifts, which indicates "good" Donors.

Let us consider a sample of n Donors, the gift count variables $Y_1, \ldots, Y_n$ in this model are independent, conditional on the covariates, and the conditional distribution of Y_i is a Poisson distribution with parameter λ_i as in (1). It is worth noting that

$$\mathbb{E}[Y_i|x_i] = \mathbb{V}[Y_i|x_i] = \exp(z_i'\theta); \tag{2}$$

it turns out that the model is naturally heteroskedastic. Parameters θ can be estimated by maximum likelihood; the resulting log-likelihood function is

$$L(\theta) = \sum_{i=1}^{n} [y_i z_i'\theta - \exp(z_i'\theta) - \log(y_i!)]. \tag{3}$$

Once estimated, the model can be used to compute the expected number of gifts for a single Donor (or a new Contact), λ_i, and the probability of gift

$$\mathbb{P}[Y_i = y] = \exp(-\lambda_i)\frac{\lambda_i^y}{y!}, \qquad y = 0, 1, 2, \ldots \tag{4}$$

Poisson regression model is easy to interpret; a possible drawback is that it is based on some strong assumptions. Nevertheless, the model allows for various extensions. For instance, a gamma distributed *heterogeneity factor* can be introduced; as a result, one obtains a *negative-binomial* model [11]. This issue is left for future research.

References

1. Andreoni, J.: Philantropy. In: Kolm, S.C., Ythier, J. (eds.) Handbook of the Economincs of Giving, Altruism and Reciprocity, vol. 2, pp. 1201–1269. Elsevier, Amsterdam (2006)
2. Barzanti, L., Dragoni, N., Degli Esposti, N., Gaspari, M.: Decision making in fund raising management: a knowledge based approach. In: Ellis, R., Allen, T., Petridis, M. (eds.) Applications and Innovations in Intelligent Systems XV, pp. 189–201. Springer, London (2007)
3. Barzanti, L., Gaspari, M., Saletti, D.: Modelling decision making in fund raising management by a fuzzy knowledge system. Expert Syst. Appl. **36**, 9466–9478 (2009)
4. Barzanti, L., Giove, S.: A decision support system for fund raising management based on the Choquet integral methodology. Expert Syst. **29**(4), 359–373 (2012)
5. Barzanti, L., Giove, S.: A decision support system for fund raising management in medium-sized organizations. Math. Meth. Econ. Fin. **9**(10), 3–11 (2018)
6. Barzanti, L., Giove, S., Pezzi, A.: A decision support system for non profit organizations. In: Petrosino, A., Loia, V., Pedrycz, W. (eds.) Fuzzy Logic and Soft Computing Applications. LNAI, pp. 270–280. Springer, Cham (2017)
7. Barzanti, L., Mastroleo, M.: An enhanced approach for developing an expert system for fund raising management. In: Segura, J.M., Reiter, A.C. (eds.) Expert System Software: Engineering, Advantages and Applications, pp. 131–156. Nova Science Publishers, New York (2013)
8. Cappellari, L., Ghinetti, P., Turati, G.: On time and money donations. J. Socio-Econ. **40**(6), 853–867 (2011)
9. Duffy, J., Ochs, J., Vesterlund, L.: Giving little by little: dynamic voluntary contribution games. J. Public Econ. **91**(9), 1708–1730 (2007)
10. Duncan, B.: Modeling charitable contributions of time and money. J. Public Econ. **72**, 213–242 (1999)
11. Gourieroux, C., Jasiak, J.: The Econometrics of Individual Risk. Credit, Insurance, and Marketing. Princeton University Press, Princeton and Oxford (2007)
12. Lee, L., Piliavin, J.A., Call, V.R.A.: Giving time, money, and blood: similarities and differneces. Social Psycology Quart. **62**(3), 276–290 (1999)
13. Melandri, V.: Fundraising. Civil Sector Press, Toronto (2017)
14. Moro, S., Cortez, P., Rita, P.: A divide-and-conquer strategy using feature relevance and expert knowledge for enhancing a data mining approach to bank telemarketing. Expert Syst. **35**(3), e12253 (2018)
15. Nudd, S.P.: Thinking strategically about information. In: Tempel, E. (ed.) Hank Rosso's Achieving Excellence in Fund Raising, pp. 349–365. Wiley, New York (2003)
16. Sargeant, A.: Using donor lifetime value to inform fundraising strategy. Nonprofit Manag. Leadersh. **12**(1), 25–38 (2001)
17. Smith, W., Chang, C.: Shipping the good apples out: a note on contributions of time and money. Econ. Bulletin **10**(1), 1–14 (2002)
18. Verhaert, G.A., Van den Poel, D.: The role of seed money and threshold size in optimizing fundraising campaigns: past behavior matters! Expert Syst. Appl. **39**, 13075–13084 (2012)

The Estimation Risk in Credit Regulatory Capital

Roberto Baviera[✉]

Department of Mathematics, Politecnico di Milano,
32 p.zza L. da Vinci, 20133 Milan, Italy
roberto.baviera@polimi.it

Abstract. In Internal Rating Based approaches, the regulator indicates a model to determine bank's credit capital requirements. The main concern is on model's econometric usage and on the estimation of its key parameters: the probability of default and the loss given default. In this study, we point out that taking into account only parameters' expectation leads to a significant underestimation of bank's risk and its Regulatory Capital. In particular, we statistically test distributional assumptions on these two parameters and we underline the key role played by parameters' dependency.

We analyse two benchmark datasets: one with all corporations rated by Moody's and another one that includes only speculative grade firms. Results are striking: we obtain that, considering parameters' uncertainty, the Regulatory Capital should be increased by an amount in the range between 38% and 66%. A clear policy implication stems from this study: the *scaling factor* for model risk, removed by Basel III accord, should be reintroduced in the determination of credit Regulatory Capital.

Keywords: Model risk · IRB · LGD-PD dependency · Scaling factor

1 Introduction

Capital adequacy in banks is often determined via Internal Rating Based (IRB) approaches. In IRB, regulators require to measure credit Regulatory Capital as a quantile of the loss distribution (VaR) of bank's credit portfolio modeled via the Asymptotic Single Risk Factor (ASRF) model, introduced by [7]. The ASRF models with a single risk factor the credit exposures of a financial institution. The regulator, besides the relevant time horizon (1y) and the confidence level α for

the VaR (99.9%), establishes also the correlation with the single risk factor as a deterministic function of other parameters. Credit exposures are modeled by two main parameters for each obligor in bank's portfolio: the obligor's probability of default (PD) and his loss-given-default (LGD). Both parameters (PD and LGD) correspond to the one-year-horizon forecast; they are calibrated with Through-the-Cycle values, i.e. long term default and recovery rates, often provided by rating agencies. Estimation risk in credit capital requirements is related to the estimation noise of these two parameters, i.e. the risk arising from errors in model parameters when we cannot rely on the assumption that the parameters of the model are known with certainty.

Albeit the relevance of this risk is well known, there is a relative paucity of empirical studies that measure the impact risk in credit capital requirements.

The problem has been introduced by [9], even before the details of Basel II accord were introduced; then, it has been analysed by [13]. Both studies consider the impact of parameter uncertainty on measures of tail risk in a stylized credit portfolio that is homogeneous, i.e. characterized by the same exposure and the same parameters (PD and LGD) for each obligor. In this case, the capital adequacy per unit exposure at default (hereinafter regulatory capital or RC) is

$$RC = VaR_\alpha[L] - \mathbb{E}[L] , \qquad (1)$$

where L is the *portfolio loss rate*, i.e. the ratio of total losses to total portfolio exposure at default. [13] has emphasized that the estimation noise is another relevant source of uncertainty about potential losses besides credit-risk factors in capital requirements (1): thus, the *correct* regulatory capital should reflect all potential losses, whose uncertainty includes the imperfect information about risk parameters.

All studies in credit capital requirements consider LGD and PD independent. However, it is quite reasonable to observe a relationship between PD and LGD (or equivalently recovery). A dependency between PD and LGD has been first pointed out by [6] for non-financial issuers domiciled in the USA in the time interval 1982–1997, then a positive correlation between PD and LGD has been identified and measured by [1] in the speculative grade USA bond market. The economic reason of this dependency is rather simple: if an economy experiences a recession, on the one hand, the observed frequency of corporate defaults increases and, on the other hand, recoveries decrease because the assets of failed companies are sold when many other firms have defaulted and when few buyers are available at extremely discounted prices (fire sale). Here, we consider this dependency, estimate statistically it and identify the impact on capital requirement: we show that this dependency is the most relevant source of model risk in credit capital requirements.

Our contributions to the existing literature are threefold: i) we model estimation risk of PD and LGD in capital requirements via a bi-dimensional copula and we statistically test this assumption on a real dataset, ii) we compute credit capital requirements within the IRB approach in presence of estimation risk and, in particular, we analyse the impact of parameter dependency in capital requirements, and iii) we draw a relevant policy implication for credit capital requirements.

2   The Capital Requirement in the IRB Approach

In regulatory capital, econometricians play a main role. A peculiarity of the IRB approach is that the modeling framework has been established by the Basel committee, while model calibration is left to the banks.

To understand the relevance of this role in the correct determination of regulatory capital, we focus on model risk for a homogeneous portfolio [14] as in [9], [13]. The IRB approach considers the case with a large number of obligors. In this case, the expected loss conditional on the common risk factor M and given the parameters PD and LGD, is

$$\mathbb{E}[L|M, PD, LGD] = LGD \cdot \Phi\left(\frac{k - \sqrt{\rho}\, M}{\sqrt{1-\rho}}\right) , \qquad (2)$$

where $k = \Phi^{-1}(PD)$ is often referred to as the default point [13], with Φ the standard normal cumulative distribution function. Moreover, ρ is $\rho(PD)$, a deterministic function of PD established by the Basel Committee [2].

A *naïve* approximation of IRB [13] accounts for the credit risk factor M but treats PD and LGD as known and equal to $\hat{PD}$ and $\hat{LGD}$, the point estimates of the respective parameters. In this special case, the above setup reduces to a single risk factor model and the capital requirement becomes

$$RC^{naive} = \hat{LGD} \cdot \Phi\left(\frac{\Phi^{-1}(\hat{PD}) - \sqrt{\rho(\hat{PD})}\, \Phi^{-1}(1-\alpha)}{\sqrt{1 - \rho(\hat{PD})}}\right) - EL^{naive} \qquad (3)$$

where the factor that multiplies $\hat{LGD}$ is also known as the 1 year Worst Case Default Rate and

$$EL^{naive} := \hat{LGD} \cdot \hat{PD}$$

is the expected loss. The popularity of this *naïve* IRB approach among practitioners can be related to the simplicity of the analytical closed formula (3).

In general, parameters could carry a significant estimation noise that cannot be neglected. Thus, it can be proven that the *correct* capital requirement (1) is

$$RC = VaR_\alpha\left[\mathbb{E}[L|M, PD, LGD]\right] - \mathbb{E}[L] , \qquad (4)$$

where $\mathbb{E}[L|M, PD, LGD]$ is reported in (2).

Two remarks should be made. First, as already pointed out by [13], M is independent from parameters: since the uncertainty about M refers to the *ex-post* realization of the credit risk factor, while parameters are the best *ex-ante* estimation given past data, the assumed temporal independence of the risk factor M implies that it is independent from parameter uncertainty. Second, there is no closed formula for the *correct* capital requirement (4) but it can be easily obtained via a Monte Carlo simulation.

What really matters is the increase in capital requirement w.r.t. the *naïve* IRB approach. In particular we focus on the Add-on for the capital requirement, defined as the percentage increment of the *correct RC* w.r.t. its *naïve* approximation

$$add\text{-}on := \frac{(RC - RC^{naive}) + (\mathbb{E}[L] - EL^{naive})}{RC^{naive}} \ . \tag{5}$$

In this short paper, we focus on this percentage increase in RC^{naive} induced by parameter uncertainty.

3   The Dataset and Parameters' Gaussian Copula

We analyse a dataset provided by Moody's Investor Service on annual LGD rates for defaulted senior unsecured corporate bonds and on annual corporate default rates [10]. Two are the default rates considered: the first set includes all corporates rated by Moody's (hereinafter "All Ratings" or "AR") while the second is limited only to firms who have a speculative grade at the beginning of the default year (hereinafter "Speculative Grade" or "SG"). The dataset reports an annual value for the period 1983–2019 (37 years); it is used by several financial institutions either in the determination of regulatory capital or in the definition of benchmarks for measuring IRB parameters.

In this short paper, we estimate the empirical properties of one-year LGD and PD via the observed default rates in this dataset, analysing the distribution for $k = \Phi^{-1}(PD)$ and LGD. We show, not only that they can be accurately described by Gaussian marginals, but also that the joint distribution is a bi-dimensional Gaussian. We verify the normality hypothesis via a statistical test. The Shapiro-Wilk test allows to determine if the null hypothesis of univariate normality is a reasonable assumption regarding the population distribution of a random sample [12]. Moreover, [11] has extended the Shapiro-Wilk hypothesis test to the bivariate case to verify composite normality.

Table 1. Shapiro-Wilk test outcome on LGD and default point k; W is the Shapiro-Wilk test statistics. We never reject the null hypothesis of normality.

	W	$p\text{-value}$
LGD	0.983	0.840
k_{AR}	0.987	0.941
k_{SG}	0.979	0.706
Composite LGD-k_{AR}	0.973	0.509
Composite LGD-k_{SG}	0.984	0.856

The results for the Shapiro-Wilk test statistic W and the p-value are reported in Table 1. We do not reject the null hypothesis of normality with a 10% threshold. Notice that all p-values are above 50%. Thus, we can consider normal the marginal distribution of each parameter; furthermore, the couple LGD-k follows

a bivariate normal distribution in both cases. Hence, PD and LGD can be well described by a 2-d Gaussian copula.[1]

Moreover, we can easily verify whether LGD and k are correlated. We can reject the uncorrelated hypothesis with a p-value $6.12 \cdot 10^{-07}$. In the Speculative Grade case, results look similar; the uncorrelated hypothesis is rejected with a p-value $8.85 \cdot 10^{-05}$ in this case. The estimated Pearson correlation $\rho_{LGD\text{-}k}$ between LGD and k is reported in Table 2.

Table 2. Pearson correlation $LGD\text{-}k$ considering All Ratings and only Speculative Grade firms. We report the estimator $\rho_{LGD\text{-}k}$ and the 95%-confidence interval (CI).

	$\rho_{LGD\text{-}k}$	$95\% \, CI$
All Ratings (AR)	0.717	(0.511, 0.844)
Speculative Grade (SG)	0.599	(0.342, 0.773)

4   Estimation Risk in RC and Policy Implication

We conclude analysing the impact on capital requirements stemming from parameter uncertainty. First, we take into account the uncertainty due to the estimation of LGD and PD, measuring the *add-on* (5) in different cases. We analyse one parameter at a time (and impose the other parameter equal to its expected value) and both parameters at the same time. In this way we can "isolate" each contribution to the *add-on*. Table 3 shows the results obtained considering either one parameter at time or the two parameters simultaneously. Parameter dependency, that –as shown in previous section– cannot be neglected from a statistical point of view, has the most relevant impact: it determines the most relevant contribution to the *add-on* in capital adequacy. Moreover,

Table 3. Regulatory capital *add-on* due to parameter uncertainties, via a Monte Carlo with $N_{sim} = 10^7$ simulations. First, we consider the *add-on* due to LGD and k separately, keeping the other parameter constant. Then, we consider the two correlated parameters. PD-LGD dependency gives the most relevant contribution to RC-*add-on*.

		AR	SG
LGD	(only)	5.63%	9.12%
k	(only)	12.22%	28.87%
LGD, k	(correlated)	**38.48%**	**65.97%**

---

[1] Let us emphasize that the Gaussian copula among parameters has no relation with the Gaussian copula in the ASRF that models obligors' assets. In the former, other empirical analysis that confirm the Gaussian copula between PD and LGD can be found in [4]. In the latter, the Gaussian hypothesis is not empirically justified on real data and a model with heavier tails should be considered (see, e.g., [8]).

the values of *add-on* appear very large: the *correct* RC, that takes into account estimation noise, is significantly greater than the one computed with the *naïve* approach. The contribution of parameter uncertainty to capital requirements appears startling with an increase in the required capital larger than 38%, if All Ratings are considered, and almost equal to 2/3, if we consider a credit portfolio composed only by Speculate Grade corporates. This is the main result of this study.

Then, we draw a policy implication for capital requirements. For market risk, regulatory capital is calculated via a multiplication factor $m_c \geq 3$ imposed by the regulators, a multiple in line with a model risk adjustment buffer [5]. Also for credit risk, the Basel II Accord allowed to apply a multiplication factor (named *scaling factor*) –greater than 1– to the result of the credit VaR calculations, factor that corresponds to a –greater than 0– *add-on*. In Basel III, the Committee has agreed to remove this *scaling factor* [3]. The main conclusion of this study from a financial policy perspective is that, to cope with the associated estimation risk, regulators should reintroduce the *scaling factor* at least equal to 1.4, when a bank prefers to stuck with a *naïve* regulatory computation.

It is common practice by risk managers to rely on a *naïve* IRB approach for capital requirements, where parameters are estimated with the long term averages of historical rates. We have shown how to incorporate the inevitable uncertainty about the forecasted parameters in measures of portfolio credit risk: such parameter forecasting depends on statistical hypotheses that can be tested on real datasets. Ignoring estimation noise leads to a substantial understatement of the regulatory capital; in particular, we have shown that parameters' dependency plays the most relevant role in capital adequacy.

References

1. Altman, E.I., Brady, B., Resti, A., Sironi, A.: The link between default and recovery rates: theory, empirical evidence, and implications. J. Bus. **78**(6), 2203–2228 (2005)
2. Basel Committee: An explanatory note on the Basel II IRB risk weight functions. Technical report, BIS, Basel Committee on Banking Supervision (2005)
3. Basel Committee: High-level summary of Basel III reforms. Technical report, BIS, Basel Committee on Banking Supervision (2017)
4. Baviera, R.: The measure of model risk in credit capital requirements. Financ. Res. Lett. **44**, 102064 (2022)
5. Boucher, C.M., Danielsson, J., Kouontchou, P.S., Maillet, B.B.: Risk models-at-risk. J. Bank Finan. **44**, 72–92 (2014)
6. Frye, J.: Depressing recoveries. Risk Mag. **13**(November), 108–111 (2000)
7. Gordy, M.B.: A risk-factor model foundation for ratings-based bank capital rules. J. Finan. Intermed. **12**(3), 199–232 (2003)
8. Hull, J., White, A.: The risk of tranches created from mortgages. Finan. Anal. J. **66**(5), 54–67 (2010)
9. Löffler, G.: The effects of estimation error on measures of portfolio credit risk. J. Bank. Finan. **27**(8), 1427–1453 (2003)
10. Ou, S., et al.: Annual default study: defaults will edge higher in 2020. Technical report, Moody's Investor Service (2020)

11. Royston, J.: Some techniques for assessing multivarate normality based on the Shapiro-Wilk W. J. Roy. Stat. Soc. Ser. C **32**(2), 121–133 (1983)
12. Shapiro, S.S., Wilk, M.B.: An analysis of variance test for normality (complete samples). Biometrika **52**(3/4), 591–611 (1965)
13. Tarashev, N.: Measuring portfolio credit risk correctly: why parameter uncertainty matters. J. Bank. Finan. **34**(9), 2065–2076 (2010)
14. Vasicek, O.A.: The distribution of loan portfolio value. Risk **15**(12), 160–162 (2002)

Actuarial Fairness in Pension Systems: An Empirical Evaluation for Italy Using an OLG Model

Michele Belloni[1,3,4(✉)] and Magdalena Zachlod-Jelec[2]

[1] Department of Economics and Statistics "Cognetti de Martiis", University of Turin, Turin, Italy
michele.belloni@unito.it

[2] European Commission Joint Research Centre, Seville, Spain

[3] NETSPAR - Network for Studies On Pensions, Aging and Retirement, Tilburg, The Netherlands

[4] CeRP - Collegio Carlo Alberto, Turin, Italy

Abstract. The Italian pension system underwent two major reforms, in 1995 and 2011, encompassing a long transitional phase from defined benefit to notional defined contribution (NDC) rules. Automatic adjustments of benefits and eligibility age, in reaction to changes in life expectancy, were introduced to reduce political risk. In this paper, we evaluate the actuarial features of the Italian pension system using an overlapping generation (OLG) model that accounts for interdependencies between the macroeconomy, the labor market, and the pension system. This is in contrast to existing studies for Italy that use microsimulation or representative agent models. In addition to that, we evaluate two policy changes concerning the computation and adjustment of conversion coefficients used in the NDC formula. Namely, we consider the calculation of cohort-specific conversion coefficients. The first policy change exploits Eurostat projected life tables, whereas the second relies on historical mortality rates attributed to retiring cohorts following the computational mechanism existing in Sweden. The NDC system returns individuals 60% of what they paid as contributions during their working life. This value is lower than what was found by previous studies and is the result of dynamic efficiency as endogenously determined by the OLG model. Projected cohort-specific conversion coefficients lead to a slightly lower generosity, accounting for the reduction in mortality probabilities individuals experience after retirement. By implementing a narrow retirement age window, Swedish-type conversion coefficients do not suffer from obsolescence. Significantly, vis-à-vis the current legislation, they do not alter pension generosity and, as such, they look politically appealing.

Keywords: Actuarial fairness · Overlapping generation model · Notional defined contribution

1 Introduction

In the last three decades, population ageing and low economic growth have weakened the financial stability of several pay-as-you-go defined benefit (DB) pension schemes.

Some countries - including Italy, Latvia, Norway, Poland, and Sweden - have reformed their pension system by retaining their pay-as-you-go financing method but replacing their DB with a notional defined contribution (NDC) pension formula. An NDC scheme consists of an individual account system to which contributions are earmarked and interests notionally paid; at retirement, the notionally accumulated sum is converted into the pension considering life expectancy, that is, incorporating actuarial fairness.

The Italian NDC scheme was introduced in 1995 (so-called Dini reform). This reform encompasses a long transitional phase from DB to NDC rules, gradually introducing actuarial fairness in the system. Older individuals – those with more than 18 years of contributions in 1995 - were excluded from the reform. In 2011, however, a further reform (Fornero reform) extended NDC rules to these cohorts (for the part of their career after 2011). Individuals with less than 18 years of contribution in 1995 are subject to a "pro-rata" system, with a DB formula for contributions paid up to 1995 and an NDC formula for those paid later. Finally, those who started to work after 1995 are fully subject to the NDC formula.

Eurostat projects that in Italy, life expectancy at the age of 65 will increase from 22.9 (19.6) years in 2019 to 29.1 (26) years in 2100 for females (males) [1]. The implementations of actuarially fair NDC systems need to guarantee that life expectancy used to compute the annuity at retirement is as close as possible to actual, ex-post, residual life. Automatic adjustments [2] of benefits and eligibility age, in reaction to changes in life expectancy, were introduced to reduce political risk.

The critical elements of the NDC pension formula are the conversion (or transformation) coefficients (TCs), which convert the sum of contributions accumulated at retirement into the stream of future pension benefits. There is a window of possible retirement ages and one TC for each retirement age. Consistently with the principle of actuarial fairness, the higher the retirement age (thus, the lower the residual life expectancy at retirement), the higher the TC, the higher the pension. Therefore, the system in its steady-state should not provide financial incentives to retire at any particular age. TCs are computed using cross-sectional mortality rates recently published by the Italian Institute of Statistics, and they are adjusted to changes in life expectancy biannually. The 2012 reform introduced automatic biannual adjustments of pension eligibility ages to the changes in life expectancy. Currently, the old-age retirement age is 67.

The scope of this paper is to evaluate the actuarial features of the Italian pension system, analyzing both its transitional phase from DB to NDC and the steady-state NDC rules. We rely on an overlapping generation (OLG) model built upon the work by [3]. Previous Italian studies conducting similar analyses relied on either representative agent or microsimulation models. Among the first group of studies, [4] and [5] show that the former DB scheme was highly generous. In contrast, the future NDC will be almost actuarially fair. Microsimulation models have largely confirmed these findings (e.g., [6]). The main advantage of the OLG approach over microsimulation and representative agent models is that OLG can fully account for interdependencies between the macroeconomy, the labor market, and the pension system. The model is calibrated to match the main features of the Italian economy, and the output of this calibration affects the performance and the evaluation of the pension system.

We evaluate two policy changes concerning the computation and adjustment of the TCs. These proposals consider the computation of cohort-specific CTs. The first exploits projected life tables; the second relies on historical mortality rates attributed to retiring cohorts following the computational mechanism existing in Sweden and adapted to the Italian case. The next session provides more methodological details, Sect. 3 reports our main results, and Sect. 4 concludes.

2 Methods

D'Andria et al. [3] developed an OLG model (EDGE-M3) to study the effects of demographic change on taxes and pension system reforms at the EU level. We enriched it with a module on the Italian pension system, which can account for many normative details and focuses on the formulas used to compute and update TCs.

There are 80 adult generations aged 20 to 99 in EDGE-M3. The oldest one is cohort 1953, starting to work at age 20 and retiring at 67 in 2019. Over their lifetime, individuals face a mortality risk in line with Eurostat projections [1]. Individuals are further distinguished by income quintiles. All in all, there are 400 representative individuals in the model. They receive utility from consumption, leaving bequests and disutility from labor. They optimize over their lifetime, choosing their labor supply (and consumption/savings), from which they receive earnings depending on their ability type.

The Italian pension module interacts with the other modules of EDGE-M3 in a variety of ways. The most obvious way is through the modeling of taxation [3]. The calibration of the model allows coherently representing and simulating the interactions between the Italian tax system and the public pension system. E.g., an individual receives pension benefits and pays an income tax as part of the capital tax. Another interaction with the pension module is that individuals account for a marginal increase in pension benefits due to a marginal increase in hours worked.

We apply the pension rules established by the Dini and Fornero reforms for each cohort in the model. Thus, there are five "Fornero" cohorts (1953–1957) retiring in 2019–2023, eighteen pro-rata cohorts (1958–1975) retiring in 2024–2041, and NDC cohorts (1976-) retiring since 2042. We refer to [5] for a detailed description of the pension sub-schemes formulas and the TCs formula, including its computational assumptions. Here we only recall that the latter formula exploits cross-sectional mortality rates taken from recent official life tables. This exercise computes future expected values of the TCs using projected (cross-sectional) life tables from [1] (central scenario).

We assess the actuarial fairness of the Italian pension system using the present value ratio (PVR), i.e., the ratio between the expected present value of pension benefits and the lifetime present value of contributions [7]. We use cohort-specific mortality rates for Italy obtained from Eurostat projections [1] when computing present values. Namely, we get the mortality rates from the diagonal of the year times age matrix of the gender-specific life table. When PVR $= 1$, the system is actuarially fair. In NDC schemes, PVR < 1 since its rate of return is the GDP (or wage bill) growth rate (plus population growth rate) and $g < r$ in a dynamic efficient economy (quasi-actuarial fairness, [8, 9]).

We consider the following normative scenarios:

1. Baseline. The current legislation - but we keep the retirement age constant across generations and equal to its current value 67;
2. Current legislation: as the baseline, but we allow for the age of retirement to increase in line with increased longevity;
3. Policy 1: we compute TCs by cohort - using *projected* cohort-specific mortality tables;
4. Policy 2: we compute TCs by cohort – using the method applied in the Swedish NDC system (see, e.g. [10]) adapted to the Italian situation. In detail, we assume that the retirement age window is 67–71. Each cohort is then assigned a *permanent* set of coefficients by age computed when they reach age 66. Unlike Policy 1, these coefficients are thus computed using historical life tables.[1]

3 Main Results

Figure 1 reports results for the baseline scenario. While the system is almost actuarially fair for the Fornero cohorts, it becomes less than actuarially fair for the pro-rata. The PVR progressively declines as the system converges towards its NDC steady state. It is indeed well-known that the DB formula is more generous than the NDC. There is a within-cohort heterogeneity for cohorts under the Fornero and the pro-rata schemes: lower ability types seem better off than high ability types. Their relative labor supply primarily drives this result – the lower ability types working more during the lifetime than their more affluent counterparts (the latter invest more in education, for instance). Differences in the working career are irrelevant in the NDC scheme. The most remarkable finding is that the NDC system returns individuals no more than 60% of what they paid in contributions during their working life (PVR about 0.6). This value is lower than what was found by previous studies. It is the result of dynamic efficiency as endogenously determined by the OLG model: The calibrated values for $g = 1.13\%$ and $r = 2.13\%$ lead to a spread of 1%.

In the current legislation scenario, the retirement age is increased progressively from age 67 to 72 (in 2073, when the 2001 cohort retires). The main finding is that this increase in the retirement age improves adequacy substantially: for instance, for NDC cohort 1990, 3rd quintile, the PVR = 0.6 if retirement is at age 67 and PVR = 0.73 if the old-age retirement age is raised to 71.

Figure 2 presents Policy options 1 and 2 and compares them with the baseline case, focusing on the 3rd quintile.[2] Applying projected cohort-specific mortality tables in the computation of TCs leads to a minor reduction of the PVR - up to 3 p.p. lower for the cohort 1964 and around 2 p.p. lower for most analyzed cohorts - vis-à-vis the baseline.

---

[1] E.g., coefficients for cohort 1955 are attributed in year 2021, those for cohort 1954 in 2020, and so on. As a consequence, in a given calendar year, coefficients coexist for different cohorts: in 2021 there exist coefficients for cohorts from 1954 (aged 67) up to cohort 1950 (aged 71).

[2] Results for other quintiles are similar. Note that, in the current legislation scenario, a comparison of the PVR across cohorts does not isolate the effect of changes in pension rules, since the retirement age varies as well. For this reason, we compare policy options against the baseline.

There is no impact on the Fornero cohorts' PVR due to a tiny NDC component of their pension formula. Noticeably, this policy results in a smoother course of the PVR among adjacent cohorts because it eliminates the need for the periodical updating process of the TCs. Policy option 2 leads to an almost unchanged PVR compared to the baseline. Indeed, in both scenarios, historical mortality is used to compute TCs. Moreover, our policy design implies a narrow retirement age window (67–71) and thus the use of *recent* life tables.
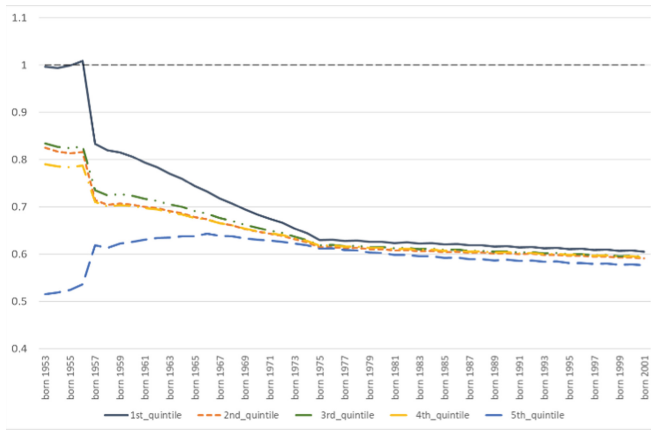
Fig. 1. PVR by cohort and income quintile – baseline scenario
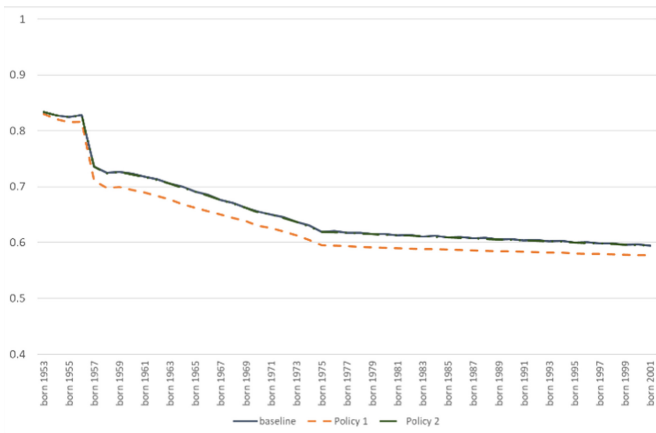
Fig. 2. PVR by cohort, 3rd income quintile – Policy options 1 and 2 versus baseline

4 Discussion and Conclusions

Our analysis points out that – if the retirement age remained constant at 67 – the future Italian NDC scheme will return to individuals 60% of what they paid in terms of contributions during their working life. This value is lower than what was found by previous studies and is the result of the OLG endogenously determined 1% spread between the GDP growth rate and the risk-free interest rate. The periodical adjustments of the TCs create discontinuities in generosity among adjacent cohorts, which, in turn, generates undesirable between-cohorts redistribution and incentives to anticipate retirement before the revision. Such incentives are stronger if one considers postponing retirement for more years, as adjustments occur more than once. The uncertainty associated with the TCs future values can also play an additional role in anticipating retirement. The retirement age increase in response to increased longevity is an effective solution to improve the adequacy of future benefits.

Cohort TCs avoid the just-mentioned adverse effects of the current updating procedure. Applying projected cohort-specific mortality tables in the computation of TCs reduces the PVR by max. 3 percentage points. This policy is preferable from the actuarial fairness viewpoint, because it incorporates declining mortality rates individuals experience during retirement. Practically, however, one should quantify the degree of uncertainty associated with projected mortality. Future developments of this study should at least check the results' robustness to different projections. By implementing a narrow retirement age window, Swedish-type conversion coefficients do not suffer from obsolescence. Significantly, vis-à-vis the current legislation, they do not alter pension generosity (i.e., the PVR) and, as such, they look politically appealing.

References

1. Eurostat homepage, EUROPOP2019. https://ec.europa.eu/eurostat/web/population-demogr aphy/population-projections/database. Accessed 18 Sept 2020
2. D'Addio, A.C., Whitehouse, E.: Towards Financial Sustainability of Pension Systems: The Role of Automatic- Adjustment Mechanisms in OECD and EU countries, OECD Social Policy Divisions, Final Report (2012)
3. D'Andria, D., DeBacker, J., Evans, R.W., Pycroft, J., Zachlod-Jelec, M.: EDGE-M3: Documentation for the European Dynamic General Equilibrium Micro Macro Model, JRC Working Papers on Taxation and Structural Reforms, 4/2020, European Commission, Joint Research Centre, Seville (2020)
4. Fornero, E., Castellino, O.: La Riforma del Sistema Previdenziale Italiano. Il Mulino, Bologna (2001)
5. Belloni, M., Maccheroni, C.: Actuarial fairness when longevity increases: an evaluation of the Italian pension system. Geneva Pap. **38**, 638–674 (2013)
6. Borella, M., Coda Moscarola, F.: Distributive properties of pension systems: a simulation of the Italian transition from defined benefit to notional defined contribution. Giornale degli Economisti e Annali di Economia **65**(1), 95–125 (2006)
7. Gruber, J., Wise, D. (eds.) Social Security and Retirement Around the World, University of Chicago Press, Chicago, pp 181–237 (1999)
8. Aaron, H.: The social insurance paradox. Can. J. Econ. Polit. Sci **32**(3), 371–374 (1966)

 9. Holzmann, R.: Toward a coordinated pension system in Europe: rationale and potential structure. In: Holzmann, R., Palmer, E. (eds.) Pension Reform: Issues and Prospects for Non-Financial Defined Contribution (NDC) Schemes. Chapter 11, pp. 225–265. The World Bank, Washington, DC (2006)
10. Palmer, E.: The Swedish Pension Reform Model: Framework and Issues. World Bank Social Protection Discussion Paper Series, No. 23086 (2000)

Forecasting VIX with Hurst Exponent

Sergio Bianchi[1], Fabrizio Di Sciorio[2], and Raffaele Mattera[3,4(✉)]

[1] MEMOTEF, Sapienza University of Rome, Rome, Italy
`sergio.bianchi@uniroma1.it`
[2] Department of Economics, University of Almeria, Almeria, Spain
`fd940@inlumine.ual.es`
[3] Department of Economics and Statistics, University of Naples "Federico II",
Naples, Italy
[4] Department of Social and Economic Sciences, Sapienza University of Rome,
Rome, Italy
`raffaele.mattera@unina.it`

Abstract. The VIX is a proxy for the implied volatility, computed considering Standard & Poor's 500 Index data. It widely regarded as a measure of turbulence in U.S. and global financial markets. Hence, forecasting the VIX is essential for both portfolio managers and policy makers. By modeling the S&P 500 Index as a multifractional Brownian motion, we exploit the relationship between its Hurst exponent and the volatility to predict the VIX by a Distributed Lag model.

Keywords: Hurst exponent · Financial markets · VIX index · Volatility · Multifractional Brownian motion

1 Introduction

Implied volatility is usually used by practitioners as a forecast of the future market volatility and is estimated from option prices. The VIX index, developed by the Chicago Board of Options Exchange (CBOE), measures the implied volatility for the Standard & Poor's 500 Index. The VIX is also widely regarded as a measure of turbulence in U.S. and global financial markets. Actually, the VIX calculation differs from the previous implied volatility-based index, called VXO. Indeed, unlike the VXO (computed by solving the Black-Scholes pricing equation), the VIX returns the following model-free value [13]:

$$VS(t,T) = \frac{2}{T-t} \sum_i \frac{\Delta K_i}{K_i^2} e^{r_i(T-t)} O_i(K_i, T) - \frac{1}{T-t} \left[\frac{F_t}{K_0} - 1 \right]^2, \quad (1)$$

where T is the common expiration date for all option contracts involved in the calculation, F_t is the forward index level from the option prices at time t, K_i is the strike price of the i-th out-of-the-money option at time t, $O_i(K_i, T)$ is the midpoint of the bid-ask spread for each out-of-the-money option with strike price K_i, K_0 is the first strike price below F_t, $\Delta K_i = \frac{K_{i+1}-K_i}{2}$ is the half-interval

between strike prices preceding and following K_i, and r_t is the risk-free rate over the period $(T - t)$.

Forecasting VIX is an important issue for many reasons [8,9]. To quote a few, investors try to get forecasts on the uncertainty levels in order to adjust their asset allocation; also, it is common nowadays to trade on derivatives built on the VIX index, such as futures or CFD contracts. Hence, forecasting the VIX can help taking both correct investment decisions for these kind of derivatives and right directions for policy makers.

Most of previous literature has tried to use implied volatility to predict realized volatility [e.g. see 6]. Moreover, [1] demonstrated a fractionally co-integrated relationship between the two. However, few papers questioned what does predict the VIX. [8] shown that the VIX index is hard to be predicted and that does not seem to be very closely connected to the volatility of the underlying index. Opposite evidences have been documented in [9] and [7]. [9] show some forecasting power with HAR (Heterogeneous Autoregressive) models when they employ macro-finance exogenous covariates such as the compounded S&P 500 returns, the first differences of S&P 500 log-volumes, the continuously compounded return on the one-month crude oil futures, or even some exchange rates. Instead, [7] found favorable results by simply implementing a non-parametric model called Singular Spectrum Analysis.

In what follows, we contribute to the VIX forecasting literature by studying the predictive power of the S&P500 Hurst exponent. Choosing the Hurst exponent as a VIX predictor can be motivated by two important evidences. First of all, we know that the VIX and S&P 500 returns are negatively correlated [e.g. see 9]. Second, it is well known that S&P 500 Index and the VIX are characterized by a long-memory processes [2,12,13]. Assuming that the S&P 500 Index follows a multifractional Brownian motion, we exploit the theoretical relationship between its Hurst exponent and the volatility [e.g. see 15].

The remainder of the paper is structured as follows. In Sect. 2 the model and the estimator are briefly described. Section 3 describes data and methodology along with the results of the analysis. Finally, Sect. 4 concludes with some final remarks and future research direction.

2 Model and Estimator

The multifractional Brownian motion (mBm) generalizes the fractional Brownian motion (fBm) by letting its exponent H to change through time [see 14]. This allows more modeling flexibility when the pointwise regularity of paths changes even abruptly, as in financial time series [3]. MBm admits the following moving average representation (to lighten notation, we write H_t in place of $H(t)$)

$$B_{H_t}(t) = \frac{C\sqrt{\Gamma(2H_t + 1)\sin(\pi H_t)}}{\Gamma\left(H_t + \frac{1}{2}\right)} \left\{ \int_{-\infty}^{0} \left[(t - u)^{H_t - 1/2} + \right. \right.$$
$$\left. \left. - (-u)^{H_t - 1/2} \, dB(u) + \int_{0}^{t} (t - u)^{H_t - 1/2} \, dB(u) \right\}$$

where C is a positive constant, $H_t \in (0,1)$ and dB is the Brownian measure. Being $E[B_{H_t}(t)] = 0$ a.s., the variance is $E[B_{H_t}(t)]^2 = C^2|t|^{2H_t}$ [5].

When $H_t > 1/2$, the increments of mBm are positively correlated. Conversely, when $H_t < 1/2$ increments are negatively correlated and there is short-memory, in the sense that the autocovariance function decreases exponentially. In the end, for $H_t = 1/2$ the mBm reduces to a standard Brownian motion.

Following [3,4], we estimate a time-varying Hurst exponent and study whether it is able to predict VIX realizations in out-of-sample. To estimate the time-varying Hurst exponent, we consider the AMBE method proposed in [3,4].

The estimator works as follows. Given the path $X_{H_t}(t)$ generated by an mBm and fixed a rolling window of size δ and a lag q, the discrete sampling $\{X_{i,n}\}_{i=1,\dots,n-1}$, for $j = i - \delta, \dots, i - q$ is considered, $i = \delta + 1, \dots, n$ and $q = 1, \dots, \delta$. Since at time t_0 the mBm behaves like an fBm with exponent $H(t_0)$, the following estimator can be built [3]:

$$\hat{H}_t = -\frac{\log\left(\sqrt{(\pi)}\mathcal{S}^k / \left(2^{k/2}\Gamma\left(\frac{k+1}{2}\right)K^k\right)\right)}{k\log\left(\frac{n+1}{q}\right)} \tag{2}$$

$$\text{with } \mathcal{S}^k = \frac{1}{\delta - q + 1}\sum_{j=i-\delta}^{i-q}|X_{j+q,n} - X_{j,n}|^k, \quad i = \delta + 1, \dots, n$$

Optimal choices to minimize the estimator's variance can be proved to be $q = 1$ and $k = 2$. The window size δ is chosen as a trade-off between the estimator's variance and the timely response to the signal. For financial application [2,4] suggest a value of $\delta = 30$.

Once the estimate of $H(t)$ has been obtained through (2), we compare the forecasts based on it using a Distributed Lag model with respect to those of a model that employs only past VIX realizations. We find that the S&P 500 Hurst exponent is a strong predictor for the VIX.

3   Empirical Analysis and Results

For the empirical analysis we consider the daily time series of the VIX and the underlying (Adjusted Price) S&P 500 between the time period 01-1990 to 12-2021 (see Fig. 1).

Before to estimate the models, the stationarity analysis is needed. In particular, the augmented Dickey – Fuller test (ADF) is carried out on both time series to verify the absence of unit roots in the time series. In unreported tables, we reject the null hypothesis of unit roots for all the involved time series.

Once the time-varying Hurst exponent H_t has been estimated, we study its predictive power by means of a Distributed Lag (DL) model. The $DL(q)$ model assumes that the time series VIX_t can be represented by a linear function of q lagged values of one or more explanatory variables. In this empirical experiment
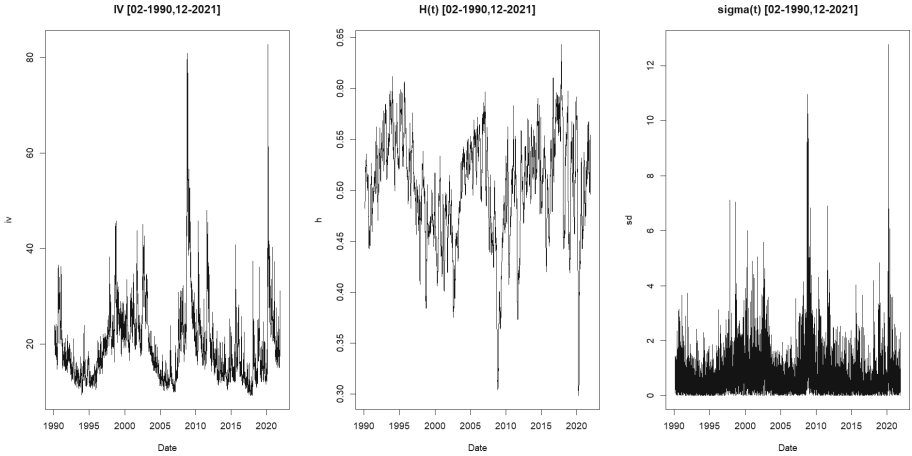
Fig. 1. Time series involved in the empirical experiment

we consider both the underlying asset Hurst exponent (i.e. the S&P 500 Hurst) H_t and a proxy of volatility, computed as the absolute value of daily log return $\sigma_t = |r_t|$ [10]. By means of step-wise regression model, we consider $q = 5$. More in detail, we compare the forecasts obtained by a simple model that employs the volatility as a predictor:

$$VIX_t = \beta_0 + \beta_1 \sigma_{t-1} + ... + \beta_5 \sigma_{t-5} + \epsilon \tag{3}$$

with an alternative model that includes the Hurst exponent:

$$VIX_t = \beta_0 + \beta_1 H_{t-1} + ... + \beta_5 H_{t-5} + \beta_6 \sigma_{t-1} + ... + \beta_{10} \sigma_{t-5} + \epsilon \tag{4}$$

In order to choose the best fitting for the model and avoid multicollinearity and overfitting, we have introduced the step-wise regression procedure [11]. This procedure starts with no predictors, then sequentially add the most contributive predictors (forward selection). After adding each new variable, remove any variables that no longer provide an improvement in the model fit (backward selection). R-squared was considered as a criterion for choosing the best model [11]. To assess the impact of the Hurst index on the predictability of VIX, two models were considered. The first adopting as regressor only the proxy daily volatility of the underlying in the step-wise regression model. The second, by adding the Hurst term to the ADL.

The model was trained on a test sample containing 90% of the time series observations, the remainder being the sample used for out-of-sample testing. The test set consists of the daily observations from 11/2019 to 12/2021. At the end of the step-wise regression procedure, the best model is as follows (see Table 1)

$$VIX_t = \beta_0 + \beta_1 H_{t-1} + \beta_2 \sigma_{t-1} + \epsilon \tag{5}$$

Table 1. Estimated models

Panel A: benchmark model				
Coefficient	Estimate	Std. Error	t	value $Pr(>|t|)$
(Intercept)	15.30556	0.10117	151.28	$<2e-16$
L(sd, 1)	5.19271	0.09192	56.49	$<2e-16$
Adjusted R-squared	0.29			
Panel B: model with Hurst exponent				
(Intercept)	83.7188	0.4674	179.1	$<2e-16$
L(h, 1)	-128.4453	0.9251	-138.8	$<2e-16$
L(σ, 1)	1.76422	0.06275	28.11	$<2e-16$
Adjusted R-squared	0.75			

Table 2 shows the results of forecast on the test set. The error metrics used (RMSE, MAE, Rsquared) show an improvement in performance with the use of the Hurst index.

Table 2. Results out of bag

Metrics	Model1	Model with Hurst Exp
MSE	65.5	29.42
RMSE	8.09	5.42
MedianAPE	18.9%	13%
MAPE	20%	15.5%
Rsquared	0.39	0.73
Corr (IV vs Fitted IV)	0.72	0.89
Test	Statistic	P-value
Diebold-Mariano Test	49.275	$2.225e-12$
Fligner-Killeen	-6.9978	$8.095e-12$

Furthermore, the non-parametric statistical tests on the variance of the error show a significant difference between the estimates of the two models used. In particular, the Hurst effect reduces the variance of the forecast error.

4   Conclusions and Further Directions

Although it is related to volatility by a closed-form formula in the case of fractional Gaussian processes, the Hurst parameter appears to be much more informative than simple volatility in VIX forecasting. This added value deserves to be

investigated along at least two directions: on the one hand, it is appropriate to compare the forecasting ability of the Hurst-based DL(q) with other performing models (e.g. HAR); on the other hand, it would be of extreme interest to analyze the forecasting ability in relation to the link that the Hurst exponent has with the market efficiency condition.

References

1. Bandi, F.M., Perron, B.: Long memory and the relation between implied and realized volatility. J. Financ. Economet. **4**(4), 636–670 (2006)
2. Bianchi, S., Frezza, M.: Fractal stock markets: international evidence of dynamical (in) efficiency. Chaos Interdiscip. J. Nonlin. Sci. **27**(7), 071102 (2017)
3. Bianchi, S., Pantanella, A., Pianese, A.: Modeling stock prices by multifractional Brownian motion: an improved estimation of the pointwise regularity. Quant. Finan. **13**(8), 1317–1330 (2013)
4. Bianchi, S., Pianese, A.: Time-varying hurst-hoelder exponents and the dynamics of (in)efficiency in stock markets. Chaos Solit. Fractals **109**(5), 64–75 (2018)
5. Cohen, S.: Fractals: theory and applications in engineering. Chapter From Self-similarity to Local Self-similarity: The Estimation Problem, pp. 3–16. Springer, New York (1999). https://doi.org/10.1007/978-1-4471-0873-3
6. Corrado, C.J., Miller, T.W., Jr.: The forecast quality of CBOE implied volatility indexes. J. Futures Mark. Futures Options Deriv. Prod. **25**(4), 339–373 (2005)
7. Degiannakis, S., Filis, G., Hassani, H.: Forecasting global stock market implied volatility indices. J. Empir. Financ. **46**, 111–129 (2018)
8. Degiannakis, S.A.: Forecasting VIX. J. Money Invest. Bank. **4**, 5–9 (2008)
9. Fernandes, M., Medeiros, M.C., Scharth, M.: Modeling and predicting the CBOE market volatility index. J. Bank. Finan. **40**, 1–10 (2014)
10. Forsberg, L., Ghysels, E.: Why do absolute returns predict volatility so well? J. Financ. Economet. **5**(1), 31–67 (2007)
11. Friedman, J., Hastie, T., Tibshirani, R., et al.: The Elements of Statistical Learning, vol. 1. Springer series in statistics New York (2001). https://doi.org/10.1007/978-0-387-84858-7
12. Granero, M.S., Segovia, J.T., Pérez, J.G.: Some comments on hurst exponent and the long memory processes on capital markets. Phys. A **387**(22), 5543–5551 (2008)
13. Ouandlous, A., Barkoulas, J.T., Alhaj-Yaseen, Y.: Persistence and discontinuity in the VIX dynamics. Chaos, Solit. Fractals **113**, 333–344 (2018)
14. Péltier, R.-F., Lévy Véhel, J.: Multifractional Brownian motion: definition and preliminary results. Technical report, RR-2645, INRIA-00074045 (1995)
15. Reed, I.S., Lee, P., Truong, T.: Spectral representation of fractional Brownian motion in n dimensions and its properties. IEEE Trans. Inf. Theory **41**(5), 1439–1451 (1995)

Modelling H-Volatility with Fractional Brownian Bridge

Sergio Bianchi[1], Massimiliano Frezza[1(✉)], Augusto Pianese[2],
and Anna Maria Palazzo[2]

[1] Department of MEMOTEF, Sapienza University of Rome, 00161 Rome, Italy
massimiliano.frezza@uniroma1.it
[2] QuantLab, University of Cassino and Southern Lazio, 03043 Cassino, Italy

Abstract. We express the realized volatility in terms of the Hurst exponent of the trajectory drawn by the market index. By analyzing distribution, stationarity, and (partial) sample autocorrelation of the estimated paths, and exploiting the empirical law of return to the central value $1/2$, we model the dynamics of $H(t)$ (and hence of the volatility) through a fractional Brownian bridge of appropriate parameter H.

Keywords: Hurst exponent · Fractional Brownian bridge · Volatility

1 Introduction

The link between the realized volatility and the dynamic Hurst exponent of a price series is now fairly well documented in literature. For an interpretation of this relationship see for example [5,14]. To make a long story short, Fig. 1 displays the realized log-volatility versus the (time-changing) Hurst parameter of the S&P500 from 1950 to 2021: the linear relationship linking the two is almost indisputable. As observed in [4], measuring volatility through the Hurst exponent has the advantage to relate the degree of variability of the log-indexes to the departures from market efficiency. In fact, while the level of volatility per se is not informative of the degree of efficiency of local market dynamics, the value of the Hurst exponent characterizes precisely how far the market is from the equilibrium represented by $H_t = \frac{1}{2}$. This is the value that is expected to be realized when (discounted) prices follow a martingale, i.e. when the market is efficient [8]. Moving from this, it seems crucial not only to estimate the Hurst exponent, but also to effectively model its dynamics. The point of this contribution is precisely this. From the estimation of the Hurst exponent we will infer some features (that seem ubiquitous with respect to time and the different markets examined) which suggest that, once properly rescaled, fractional Brownian bridges (fBb) may represent good models for the dynamics of H_t. Our results are strongly consistent with those on rough volatility [11], which model the log-volatility by fractional Brownian motion (fBm) with a very small Hurst exponent. One difference lies in the fact that, using H_t instead of the log-volatility, an additional

constraint emerges that can contribute to specify the modeling process as a fBb instead of a fBm. This follows from the empirical observation that the Hurst parameter tends to return to the value $1/2$, that is markets tend to restore the equilibrium by continuously adjusting H_t towards its central value $1/2$.
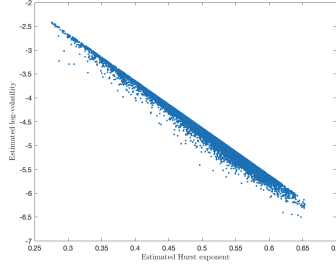
Fig. 1. S&P500 (1950–2021), 19735 observations: estimated log-volatility vs. estimated Hurst exponent

2   Fractional Brownian Bridge

While there is a substantial literature in physics and signal analysis, not many contributions have investigated the use of fBb in finance [13]. On the contrary, when one tries to model the dynamics of the Hurst parameter, this choice sounds reasonable, because generally $H_t \in [a, b] \subset (0, 1)$; to fulfill this constraint [10] uses e.g. a Fisher-like transformation. Here we will adopt the fBb approach.

To resume some results, in the following we will refer to [7].

The real-valued process X_t, with $X_0 = 0$, is called a *bridge*, denoted by X_t^B, if it is conditioned to be $X_T = a$. It can be proved that

$$\mathbb{E}\left(X_{t_1}^B\right) = a \frac{\mathbb{E}(X_{t_1} X_T)}{\mathbb{E}(X_T^2)} \tag{1}$$

and

$$\mathbb{E}\left(X_{t_1}^B X_{t_2}^B\right) = \mathbb{E}\left(X_{t_1} X_{t_2}\right) - \left[\mathbb{E}\left(X_T^2\right) - a^2\right] \frac{\mathbb{E}(X_{t_1} X_T)\mathbb{E}(X_{t_2} X_T)}{\mathbb{E}(X_T^2)^2} \tag{2}$$

The subtracted process X_t^S, defined from the original process X_t, reads as

$$X_t^S := X_t - (X_T - a) \frac{\mathbb{E}(X_t X_T)}{\mathbb{E} X_T^2}. \tag{3}$$

First, it can be noticed that $X_t^S \overset{\mathcal{L}}{=} X_t^B$, where $\overset{\mathcal{L}}{=}$ denotes the equality in law. When X_t is the Brownian motion B_t, the subtracted process - which has the same law as the Brownian bridge terminating at a - obviously reduces to

$$B_t^S := B_t - \frac{t}{T}\left(B_T - a\right). \tag{4}$$

Because of the nonzero autocorrelation functions, things are slightly trickier in the case of the fractional Brownian motion B_t^H of parameter $H \neq \frac{1}{2}$. In this case, the subtracted process is non-linear in t and reads as

$$(B_t^H)^S := B_t^H - \frac{1}{2}\left(B_T^H - a\right)\left[1 + \left(\frac{t}{T}\right)^{2H} - \left(1 - \frac{t}{T}\right)^{2H}\right] \qquad (5)$$

Using the perturbative expansion (see [15] and [6]) obtained for the standard Brownian motion and setting $H = \frac{1}{2} + \varepsilon$ (for small ε), in [7] Delorme and Wiese deduce the distribution of the maximum of the fractional Brownian bridge

$$\mathcal{P}_H^B(m) = \frac{2y^{1-8\varepsilon}}{T^H}e^{-y^2 A_\varepsilon + \varepsilon \mathcal{G}(y) + cst} + \mathcal{O}(\varepsilon^2) \qquad (6)$$

where:

- $y = \frac{m}{T^H}$;
- $A_\varepsilon = \frac{4^H}{4-4^H} = 1 + 4\ln(2)\varepsilon + \mathcal{O}(\varepsilon^2)$
- $\mathcal{G}(y) = -4(y^2-1)\mathcal{I}(y^2) + 2\sqrt{\pi}e^{y^2}y\,\mathrm{erfc}(y) + 2y^2\left[\ln(4y^2) + \gamma_E\right] - 4\gamma_E - 2$ (γ_E being the Euler-Mascheroni constant);
- $\mathcal{I}(x) = \frac{1}{2}\pi\,\mathrm{erfi}(\sqrt{x}) - x\,_2F_2\left(1,1;\frac{3}{2},2;x\right)$ (here $_2F_2$ denotes the hypergeometric function)

At small m, the distribution $\mathcal{P}_H^B(m)$ has a power law given by $m^{1-8\varepsilon+\mathcal{O}(\varepsilon^2)}$. This leads to the result $\mathcal{P}_H^B(m) \sim m^{\frac{2}{H}-3}$. At large m, $\mathcal{P}_H^B(m)$ has a Gaussian tail with dimensionless variable $y^2 = z^2/T^{2\varepsilon} = m^2/T^{2H}$.

3   Methodology and Application

First, from the daily stock indexes series of length n we estimate the Hurst exponent using the methodology in [12], where the unbiased, large-variance estimator $\hat{H}_{\nu,n}(t, A)$ is merged with the biased, low-variance estimator $\hat{H}_{\nu,n,K^*}(t)$ to obtain the unbiased, low-variance estimator

$$\widetilde{H}_{\nu,n}(t, A) = \hat{H}_{\nu,n,K^*}(t) + \frac{1}{n}\sum_{t=1}^{n}\left(\hat{H}_{\nu,n}(t, A) - \hat{H}_{\nu,n,K^*}(t)\right), \qquad (7)$$

where ν is the size of the estimation window, K^* is an arbitrary scale parameter of the process and A is a discrete differencing operator acting to make the sequence locally stationary and to weaken the dependence between the observations. Using (7), we get the estimates summarized in Fig. 2 for three main stock indexes in the period January 2, 2000–December 31, 2021. Figure 3 and Table 1 report the partial sample autocorrelation functions and the results of the normality test of $\widetilde{H}_{21,n}(t, (-1, 1))$. We refer to the partial ACF to remove the spurious autoregressive component triggered by estimator (7) and desumable by the spike in correspondence of lag ν and its multiples.

The analysis suggests that, for the three indexes, $\widetilde{H}_{21,n}(t, (-1, 1))$:

a) is trend stationary (Augmented Dickey Fuller test in column *ADF*, Table 1) and mean-reverting (Fig. 2);
b) has mean value approximately equal to $\frac{1}{2}$ (column *mean*, Table 1)
c) displays negative short-term autocorrelation (Fig. 3)
d) is roughly normally distributed. Using the Lilliefors test, normality cannot be rejected, especially net of the spikes recordered for market crashes (column *L-test*, Table 1, compared with the critical values of the last three columns).

The above claims a)-d) suggest to model H_t by means of a fBb, with terminal value $\frac{1}{2}$ and variance dictated by the estimated ones (column *std*, Table 1). To this aim, it is necessary to estimate the global Hurst exponent of the sequence $\widetilde{H}_{21,n}(t, (-1, 1))$ of each index. As noted by [10], this configures a kind of nested fractality. Denoted by $\sigma(\widetilde{H}_{21,n}(t, (-1, 1)))$ the standard deviation of the sequence $\widetilde{H}_{21,n}$, the global Hurst exponent H is estimated as the slope of the log-linear fit

$$\ln \sigma(\widetilde{H}_{21,n}(\lambda t, A)) = H \ln \lambda + \ln \sigma(\widetilde{H}_{21,n}(t, A)). \tag{8}$$

Relation (8) directly follows from the well-known scaling law $B_{\lambda t}^H \overset{d}{=} \lambda^H B_t^H$, peculiar of the fBm. The results of the log-linear fit are reproduced in Table 2.
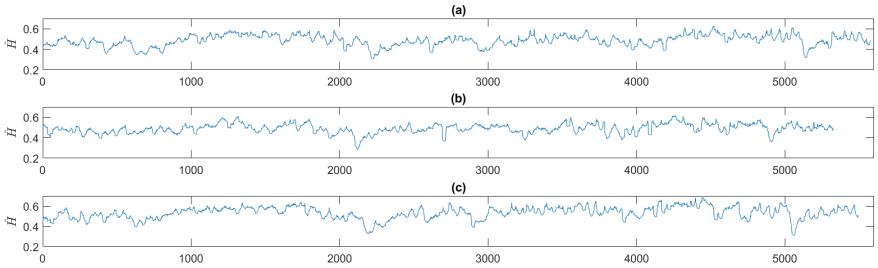
Fig. 2. $\widetilde{H}_{21,1,n}(t, (-1, 1))$: (a) ESTOXX50 (b) Nikkei 225 (c) S&P500

Table 1. Main stats for $\widetilde{H}_{21,1,n}(t, (-1, 1))$ (critical value for the ADF test: –3.4134)

	mean	std	kurt	skew	max	min	ADF	L-stat	Critical value α		
									0.01	0.05	0.1
ESTOXX50	0.486	0.0550	3.041	−0.459	0.627	0.309	−5.295	0.0617	0.0656	0.0563	0.0517
Nikkei 225	0.489	0.0468	3.965	−0.415	0.617	0.283	−5.746	0.0466	0.0671	0.0576	0.0529
S&P500	0.533	0.0600	3.539	−0.599	0.685	0.310	−4.831	0.0545	0.0662	0.0568	0.0522

Under the assumption of stationarity, let $\mathbb{A}$ in Eq. (5) be a r.v. distributed as the standardized $\widetilde{H}_{\nu,n}$, i.e. $\mathbb{A} \sim \text{PDF}\left(\frac{\widetilde{H}_{\nu,n} - \text{mean}(\widetilde{H}_{\nu,n})}{\sigma(\widetilde{H}_{\nu,n})}\right)$. Using relation (5) with H given by the slopes summarized in Table 2, we can surrogate paths of fBb of length T (representing the simulated dynamics of H_t) as follows:
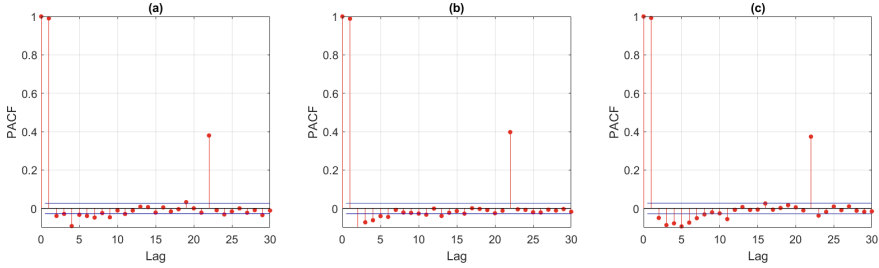
Fig. 3. Partial ACF of $\widetilde{H}_{21,1,n}(t,(-1,1))$: (a) STOXX50E (b) Nikkei 225 (c) S&P 500

Table 2. Log-linear fit (8). The slopes provide the estimated H.

		Estimated	SE	tStat	pVal	#Obs	RMSE	R-squared
ESTOXX50	Slope	0.2776	0.0033	84.69	0	532	0.0574	0.931
	Intcp	−4.450	0.0180	−247.39	0			
Nikkei 225	Slope	0.2515	0.0036	69.90	0	513	0.0614	0.905
	Intcp	−4.353	0.0196	−221.78	0			
S&P500	Slope	0.2595	0.0030	87.91	0	530	0.0515	0.936
	Intcp	−4.279	0.0162	−264.34	0			

$$H_t^{\text{sim}} = (B_t^H)^B \overset{\mathcal{L}}{=} (B_t^H)^S$$
$$= \left\{ B_t^H - \frac{1}{2} \left(B_T^H - \mathbb{A} \right) \left[1 + \left(\frac{t}{T} \right)^{2H} - \left(1 - \frac{t}{T} \right)^{2H} \right] \right\} \sigma_{\widetilde{H}_{\nu,n}} + \frac{1}{2}$$

The r.v. $\mathbb{A}$ forces the terminal values of all simulated paths of fBb to distribute as the standardized $\widetilde{H}_{\nu,n}$, while scaling the fBb by $\sigma_{\widetilde{H}_{\nu,n}}$ will ensure that the intermediate values of the surrogated fBb deviate from 0 as the observed ones. This is a crucial issue, since the condition $H \in [a,b] \subset (0,1)$ has to be fulfilled. Finally, shifting the simulated fBb by the quantity $1/2$ ensures that the surrogates H_t will fluctuate around the equilibrium value.

In the following simulations, $\mathbb{A} \sim \text{GEV}(k,\sigma,\mu)$, with the shape, scale and location parameter of the Generalized Extreme Value distribution estimated from $\widetilde{H}_{\nu,n}$ of each index. An example of the sample paths generated using the above procedure is reported in Fig. 4 for the index ESTOXX50. For each index, Fig. 5 displays the fits of the simulations at the terminal points.
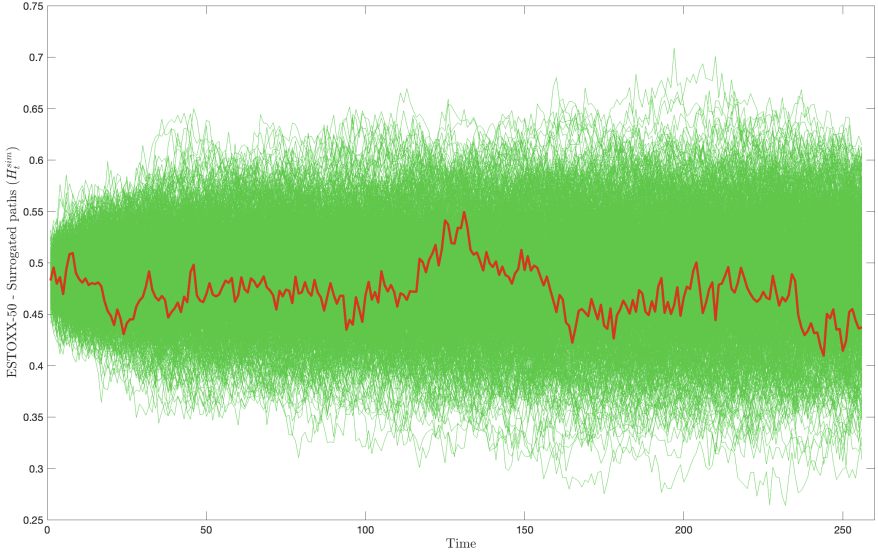
Fig. 4. H_t^{sim} (in green 1000 trajectories; in red one of the paths)
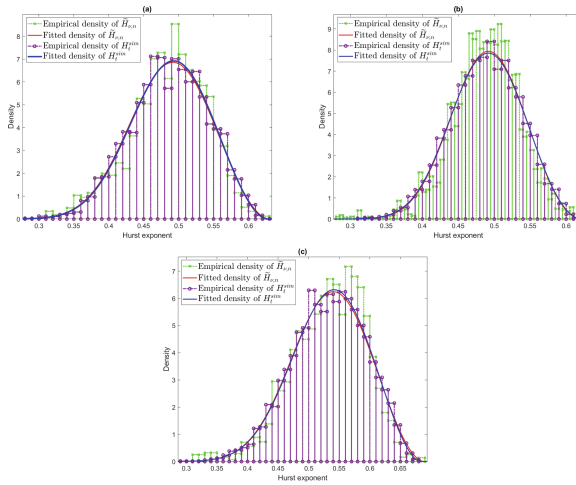
Fig. 5. (a) ESTOXX50 (b) Nikkei 225 (c) S&P500

4   Conclusion

Modeling volatility through the Hurst exponent has the added value of establishing a direct relationship with market equilibrium. The problem that this approach poses is to restrict the values of H-volatility to the range $[a, b] \subset (0, 1)$. This constraint is met through the use of an fBb properly rescaled. The preliminary results

of this contribution are consistent with the literature on rough volatility and confirm the goodness of fit of the surrogated paths.

References

1. Ayache, A., Taqqu, M.S., Multifractional processes with random exponent, Publicacions Matemátiques **49**, 459–486 (2005)
2. Bianchi, S.: Pathwise identification of the memory function of the multifractional Brownian motion with application to finance. Int. J. Theor. App. Fin. **8**(2), 255–281 (2005)
3. Bianchi, S., Pantanella, A., Pianese, A.: Modeling stock prices by multifractional Brownian motion: an improved estimation of the pointwise regularity. Quant. Finance **13**(8), 1317–1330 (2013)
4. Bianchi, S., Pianese, A., Frezza, M., Palazzo, A.M.: Stochastic dominance in the outer distributions of the α-efficiency domain. In: Corazza, M., Gilli, M., Perna, C., Pizzi, C., Sibillo, M. (eds.) Mathematical and Statistical Methods for Actuarial Sciences and Finance. Springer, Cham (2021). https://doi.org/10.1007/978-3-319-89824-7
5. Cajueiro, D.O., Tabak, B.M.: The Hurst exponent over time: testing the assertion that emerging markets are becoming more efficient. Phys. A Stat. Mech. Appl. **336**(3–4), 521–537 (2004)
6. Delorme, M., Wiese, K.J.: Maximum of a fractional Brownian motion: analytic results from perturbation theory. Phys. Rev. Lett. **115**, 210601 (2015)
7. Delorme, M., Wiese, K.J.: Extreme-value statistics of fractional Brownian motion bridges. Phys. Rev. E **94**(5), 052105 (2016)
8. Fama, E.F.: Efficient capital markets: a review of theory and empirical work. J. Finance **25**(2), 383–417 (1970)
9. Garcin, M.: Estimation of time-dependent Hurst exponents with variational smoothing and application to forecasting foreign exchange rates. Phy. A Stat. Mech. Appl. **483**, 462–479 (2017)
10. Garcin, M., Fractal analysis of the multifractality of foreign exchange rates. Math. Methods Econ. Finance **13/14**(1), 49–74 (2018/2019)
11. Gatheral, J., Jaisson, T., Rosenbaum, M.: Volatility is rough. Quant. Finance **18**(6), 933–949 (2018)
12. Pianese, A., Bianchi, S., Palazzo, A.M.: Fast and unbiased estimator of the time-dependent Hurst exponent. Chaos **28**(31102), 1–6 (2018)
13. Tapiero, C.S., Vallois, P.: Fractional Randomness and the Brownian Bridge. Phys. A Stat. Mech. Appl. **503**, 835–843 (2018)
14. Tzouras, S., Anagnostopoulos, C., McCoy, E.: Financial time series modeling using the Hurst exponent. Phys. A Stat. Mech. Appl. **425**, 50–68 (2015)
15. Wiese, K.J., Majumdar, S.N., Rosso, A.: Perturbation theory for fractional Brownian motion in presence of absorbing boundaries. Phys. Rev. E **83**, 061141 (2011)

Shapley Value in Partition Function Form Games: New Research Perspectives for Features Selection

Giovanna Bimonte[(✉)] and Luigi Senatore

Department of Economics and Statistics, University of Salerno, Fisciano, Italy
{gbimonte,lsenatore}@unisa.it

Abstract. The Shapley value assigns each game in Characteristic form a result (contribution) for each player. In games with externalities, there is a Partition Function assigned to the characteristic representation. Various generalisations or extensions of the Shapley value have been developed in the literature. The Shapley value for games in Partition Function Form can be interpreted as the ex ante value of a process of successive bilateral mergers. Game-theoretic formulations of feature importance are a way of explaining machine learning models. These methods define a cooperative game between features in a model and using Shapley's value study the influence of input features. The externality modelled in the game is read as a further measure of the contribution of features and seeks to interpret causal structures in the data. Our aim is to construct a weighted elementary marginal contribution for each feauture, in order to select attributes that have explanatory value.

Keywords: Shapley value · Games in partition function · Features contributions · Causality

1 Introduction

The theory of cooperative games associates each transferable utility game (TU) with a characteristic function that assigns real numbers (payoffs) to each coalition that can be formed. A limitation of games in the characteristic function form is that it cannot be possible to differentiate between various situations in which the payoffs that a coalition can obtain depend on the external coalitional arrangement of the players. The partition function associated with a TU game is a way of preserving information about externalities. A partition function (Thrall and Lucas (1963)) assigns a value to each pair consisting of a coalition and a coalition structure that includes that coalition. The Shapley value (Shapley 1953), one of the most important solution notions for cooperative games, has proven to be a useful solution concept for TU games as it provides a way for the division of the common profits of the grand coalition, which satisfies certain properties (axioms).

In Partition Function Form (PFF) games, several generalized and/or extended notions of the concept of Shapley value have been defined, since the final payoffs depend on the structure of the coalition that is formed (Myerson (1977), Bolger (1989) and Potter (2000)). In all these papers, authors derived an efficient extended value for TU games in PFF: in Myerson (1977) Shapley value is a natural extension of the value based on the three simple axioms; Bolger (1989) derived a value that assigns zero to dummies players and assigns non-negative values to other players; Potter (2000) modified the regular concept of a dummy player that allows the dummy player to bring a value to the game. Pham Do and Norde [2] construct a Shapley value extension under the assumption that all the players outside this coalition act as singletons, they define a collection of marginal vectors different from Bolger; instead Potter considered the sum of an 'average worth' of coalitions. The authors show that there is a unique solution, the Shapley value, satisfying efficiency, additivity, the null player property and symmetry for a TU game in PFF.

Finally, Mc Quillin [8] defined a Generalized Extended Shapley Value, excluding from consideration the payoffs to embedded coalitions involving partitions of more than two coalitions, i.e. when S forms, all the other agents form a coalition, $N \setminus S$. The extended and generalized Shapley value is the expected value of the outcome for each player. Players arrive in the coalition in a random order and receive the marginal payoff that their participation brings to the coalition formed by the players who arrived first.

2 Games in Partition Function Form

Let $N = \{1, 2, ..., n\}$ denote a finite set of players and V the set of all mappings $v : 2^N \to \mathbf{R}$ with $v(\emptyset) = 0$. Subset of N are called coalitions. We refer to $v \in V$ as a transferable utility (TU) game in characteristic function form, on N. A partition k of N is a set of non-empty coalitions, $k = \{S_1, S_2, ..., S_m\}$, such that $\bigcup_{i=1}^{m} S_i = N$ and $S_i \bigcap S_j = \emptyset$ for $i \neq j$. Let $\mathcal{P}(N)$ be the set of all partitions of N. For a partition $k \in \mathcal{P}(N)$ and $i \in N$ we denote the coalition in k to which player i belongs by $S(k, i)$. As in Pham Do and al. [2], let $k \in \mathcal{P}(N)$ and $i, j \in N$, such that $i \neq j$, we define the partition k_{ij} obtained from k interchanging the positions of i and j.[1]

A pair consisting of a coalition S and a partition $k \in \mathcal{P}(N)$ to which S belongs, $S(k, i)$, is called an embedded coalition. Let $E(N)$ be the set of al embedded coalitions:

$$E(N) = \{(S, k) \in 2^N \times \mathcal{P}(N) | S \in k\}.$$

A mapping $\omega : E(N) \to \mathbf{R}$ that assigns a real value $\omega(S, k)$ to each embedded coalition (S, k) is a partition function. Let W be the set of all mappings ω, with $S = \emptyset \to \omega(S, k) = 0$. We define a Partition Function Form Game, PFFG, an

---

[1] If i and j belong to the same coalition, i.e. $S(k, i) = S(k, j)$, then $k_{ij} = k$.

ordered pair (N, ω) and the set of partition function form games with player set N is denoted by $PFFG^N$.

2.1   The Shapley Value

The value in a PFFG represents the payoff of coalition S given that coalition structure k forms. We consider the extended generalized value defined by Mc Quillin [8]. In this model, χ is a mapping from W to W, then $\chi(\omega)$ is also an element of W. Given an embedded coalition (S, k), the real number $\omega(S, k)$ represents the utility payoff assessed for coalition S given the partition k in the game (N, ω). $\chi(\omega)(S, k)$ is the expected utility outcome associated with coalition S whenever k is the coalition structure existing in the game (N, ω). Formally, a mapping become a set of ordered pairs: $\chi \equiv \{(\omega, \chi(\omega)) : \omega \in W\}$. Note that $\{(\{i\}, [N]) : i \in N\} \subseteq E(N)$ is the set of embedded coalitions comprising singletons and the finest partition. Finally, let Π denotes the set of permutations of N, and given $\pi \in \Pi$, $S \in N$, $k \in \mathcal{P}(N)$, $\omega \in W$, we define $\pi k \in \mathcal{P}(N)$ to be the set $\{(\pi S)_{S \in k}\}$ with πS the image under π of S, and $\pi \omega \in W$ is $\pi \omega(\pi S, \pi k) = \omega(S, k)$.

We assume the following standard axioms, extended for a PFF game:

Axiom 1 (Efficiency)[2] $\sum_{i \in N} \chi(\omega)(\{i\}, [N]) = \omega(N, \{N\})$.
Axiom 2 (Symmetry) For all permutation $\pi \in \Pi$, $\chi(\pi\omega)(\{\pi(j)\}, [N]) = \chi(\omega)(\{j\}, [N])$.
Axiom 3 (Null-player) If j is a null-player in ω, then $\chi(\omega)(\{j\}, [N]) = 0$
Axiom 4 (Linearity)
 (i) $\chi(\omega + \omega')(\{j\}, [N]) = (\chi(\omega) + \chi(\omega'))(\{j\}, [N])$,
(ii) for all $\lambda \in \mathbf{R}$, $\chi(\lambda\omega)(\{j\}, [N]) = \lambda\chi(\omega)(\{j\}, [N])$.

The Shapley value can be defined as follows:

Definition 1. *The set $\{(\omega, \chi(\omega)|_{\{(\{i\}, [N]) : i \in N\}}) : \omega \in W\}$ is the Shapley value if and only if $\forall i \in N$, and for all $\omega \in W$,*

$$\chi(\omega)(\{i\}, [N]) = \sum_{S \subseteq N} \frac{(|S| - 1)!(|N| - |S|)!}{|N|!}(v(S) - v(S \setminus \{i\})) \qquad (1)$$

where v is the correspondent element in V to ω.

The term $\frac{(|S|-1)!(|N|-|S|)!}{|N|!}$ can be interpreted as the probability that in any permutation, the members of S are ahead of a distinguished player i. The term $(v(S) - v(S \setminus \{i\}))$ gives the marginal contribution of player i to the worth of the coalition S. The Shapley value gives the expected contribution of player i to the worth of any coalition.

The generalized Shapley value for games in partition function form consider the marginal contributions of the embedded coalitions, assuming that all players

---

[2] The Efficiency axiom asserts that $\omega(N, \{N\}) = \max_{k \in \mathcal{P}(N)} \sum_{S \in k} \chi(\omega)(S, k)$, and arising from superadditivity assumption (von Neumann and Morgenstern).

outside this coalition act as singletons. Pham Do and al. show that the Shapley value for PFFG, defined as the average of marginal vectors of the embedded coalitions (S, k_S), is the only solution satisfying the four axioms. McQuillin defined a Generalized Shapley Value as

$$\chi(\omega)(S, k) = \sum_{T \in k} \frac{(|T| - 1)!(|k| - |T|)!}{|k|!} (v(\bigcup_{A \in T} A) - v(\bigcup_{A \in (T \setminus \{S\})} A))$$

That is, for any partition coarser than $[N]$, only one member is the coalition representative, and all other players are null-players. To each coalition is assigned the payoff given to its representative in the game, which is thus created on the representatives and null players.

3   Shapley Values for Features Contributions

Machine learning models can produce high accuracy predictions in many applications. The accuracy in prediction from such black box models, comes at the cost of interpretability. To explain the prediction of a model using the Shapley value method, it is necessary to reformulate a cooperative game with players who correspond to the features and a payoff that fits the prediction. Given a predictive model, for each instance a vector of importance scores associated with the underlying features is produced. Thus, the importance scores can act as an explanation for the specific instance, indicating which features are fundamental for the model to make its prediction.

The Shapley Value has been widely used for the interpretation of that models; i.e. the Shapley Value is an important part of stepwise selection of the features, a modeling procedure in which features which increase the accuracy of a model are successively added to the modeling set. Following Frye and al. [4], in the machine learning models, let $f_y(x)$ represents a model's predicted probability that data point x belongs to class y. We can interpret the input features $\{x_1, x_2, ..., x_n\}$ as players that cooperate to earn a value $f_y(x)$. To define the value function, $v(S)$, to represent the model's action on a coalition $x_S \subseteq \{x_1, x_2, ..., x_n\}$ of $x's$ features. The features out of the coalition S are $x_{\overline{S}} = \{x_1, x_2, ..., x_n\} \setminus x_S$, and the value is

$$v_{f_y(x)}(S) = \cdot E_{p(x')}[f_y(x_S \sqcup x'_{\overline{S}})]$$

The average over permutations leads to Shapley value that explains the individual prediction $f_y(x)$. In the context of model explainability, the efficiency axiom implies that the model's output attribution is fully distributed over its input characteristics; The null-player axiom (nonexistence) guarantees that if a feature is completely disconnected from the model output, it receives a Shapley value of zero; the symmetry axiom (consistency) requires that the attribution is equally distributed over features that are identically informative of the model prediction; finally the additivity can be interpreted as local accuracy of the model.

Shapley explanations are largely based on the value function $v_{f_y(x)}(S)$. The unconditional marginalization is problematic, since out-of-coalition features may

not be compatible with in-coalition features. The risk is to ignore interactions between features, which are likely to be very important in applications. Constructing embedded coalitions allows to assess feature compatibility by selecting the appropriate coalition structure, and a conditional value function can be defined appropriately. A revised Shapley value of the conditional value function can respect correlations in the data and it may attribute influence to features with no interventional effect.

4   Conclusions and Further Research

The problem related to the distribution of a model's prediction score for a specific input to its underlying features, can be interpreted as the importance of the feature for the prediction. In the main literature of the application of Shapley values for feature selection emerges the necessity to quantify the importance of random input variables to a function, see Owen and Prieur [9]. Using a PFFG, the research has two objectives: first, by defining the weighted elementary marginal contribution (w-emc) for each feauture, select only those attributes that have explanatory value; second, construct an asymmetric Shapley value depending on externalities effect that takes into account the subset size.

   In non-additive model, the situational contribution of the value of one feature depends on the values of other features. This leads to considering the order in which the features are added, and the Shapley values are obtained by averaging the values of all possible orderings. The need to generalize the marginal contributions to incorporate the effect of externalities emerges. We define a new coalition structure that takes into account the dependence (externalities) effect between features. Fryer and al. [3] first attempt to incorporate causality using an asymmetric Shapley value. Asymmetric Shapley values provide a method for incorporating causal knowledge into the model's expectation explainability. Furthermore, all features are assumed to have equal weight in the model explanation. This forces Shapley values to uniformly distribute the importance of features across identically informative features. For this porpouses, we define the marginal contribution of i as a weighted average of all possible elementary marginal contributions of i. All weights must be non-negative and must not depend on the names of the agents (symmetry). Finally, for normalization, we assume that their sum is equal to one for each embedded coalition. In this way, the weights can be interpreted as the probability that a transfer of i from S to another coalition will take place. Formally, we use the notion of *weighted elementary marginal contribution* (Skibski et al. [10]) in which we take into account the new partition induced by the leaving agent.

Example 1 (Symmetric three players, superadditive and efficient game with positive externalities (Hafalir [5])). Let $N = \{1, 2, 3\}$. The values take into account externalities effect:
$v(\{i\}; \{\{1\}, \{2\}, \{3\}\}) = 4$ for all $i = 1, 2, 3$;
$v(\{jk\}; \{\{i\}, \{jk\}\}) = 9$ and $v(\{i\}; \{\{i\}, \{jk\}\}) = 1$
$v(N, \{N\}) = 14$

There are six ordered permutations of N, consider $\pi = (1, 2, 3)$. The marginal contribution depends on the process by which agents leave the grand coalition and merge into external coalitions. The *externality-free* value assumes that every leaving player will form a singletone coalition, the w-emc vector is $(5, 5, 4)$; The *McQuillin* value assumes that every leaving player will join the (only) existing coalition outside, the w-emc vector is $(5, 8, 1)$.

A second problem with using Shapley value is being able to handle the computational complexity that grows exponentially in the number of features, see Aas et al. [1]. Some authors, Jullum et al. [7] and Harris et al. [6], suggest computing Shapley values for groups of features defined with all the usual Shapley properties. In Song et al. [11] authors use a semivalue, that indicates the relative importance of a certain subset size. The research work will focus on replacing the half-value with the PFF approach: this yields an asymmetric Shapley value that depends on the effect of externalities.

References

1. Aas, K., Jullum, M., Løland, A.: Explaining individual predictions when features are dependent: more accurate approximations to Shapley values. Artif. Intell. **298**, 103502 (2021)
2. Do, K.H.P., Norde, H.: The Shapley value for partition function form games. Int. Game Theory Rev. **9**(02), 353–360 (2007)
3. Fryer, D., Strümke, I., Nguyen, H.: Shapley values for feature selection: the good, the bad, and the axioms. arXiv preprint arXiv:2102.10936 (2021)
4. Frye, C., Rowat, C., Feige, I.: Asymmetric Shapley values: incorporating causal knowledge into model-agnostic explainability. Adv. Neural Inf. Process. Syst. **33** (2020)
5. Hafalir, I.E.: Efficiency in coalition games with externalities. Games Econ. Behav. **61**(2), 242–258 (2007)
6. Harris, C., Pymar, R., Rowat, C.: Joint shapley values: a measure of joint feature importance. arXiv preprint arXiv:2107.11357 (2021)
7. Jullum, M., Redelmeier, A., Aas, K.: groupShapley: efficient prediction explanation with Shapley values for feature groups. arXiv preprint arXiv:2106.12228 (2021)
8. McQuillin, B.: The extended and generalized Shapley value: simultaneous consideration of coalitional externalities and coalitional structure. J. Econ. Theory **144**(2), 696–721 (2009)
9. Owen, A.B., Prieur, C.: On Shapley value for measuring importance of dependent inputs. SIAM/ASA J. Uncert. Quant. **5**(1), 986–1002 (2017)
10. Skibski, O., Michalak, T.P., Sakurai, Y., Wooldridge, M., Yokoo, M.: Partition decision trees: representation for efficient computation of the Shapley value extended to games with externalities. Auton. Agents Multi-Agent Syst. **34**(1), 1–39 (2020)
11. Song, E., Nelson, B.L., Staum, J.: Shapley effects for global sensitivity analysis: theory and computation. SIAM/ASA J. Uncert. Quant. **4**(1), 1060–1083 (2016)
12. Sundararajan, M., Najmi, A.: The many Shapley values for model explanation. In: International Conference on Machine Learning, pp. 9269–9278. PMLR (2020)

Nonparametric Estimation of Range Value at Risk

Suparna Biswas[(✉)] and Rituparna Sen

Indian Statistical Institute, Bangalore 560059, Karnataka, India
`suparnabsws4@gmail.com`

Abstract. Range Value at Risk (RVaR) is a two-parameter class of quantile-based risk measures. It is the conditional expectation of the loss when it lies between two values of VaR, for levels p and q, where $0 < p < q < 1$. We describe some of the nonparametric estimators of RVaR. Using Monte Carlo simulations, we compare the accuracy of these estimators under certain conditions. Our simulations provide insight into the effect of varying p and q with n on the performance of nonparametric RVaR estimators, where n is the sample size.

Keywords: Range Value at Risk · Nonparametric estimation · Monte Carlo simulations

1 Introduction

Risk management in financial institutions has been the focus of much interest in many countries. From a quantitative perspective, the use of appropriate risk measures is an important issue. The VaR and ES are the most common families of risk measures in practise; both are utilised in modern financial and insurance regulation. There has been a lot of discussion on the comparative advantages of VaR and ES during the last few years; see Embrechts et al. [6] and Emmer et al. [8] for comprehensive discussions. The one-parameter families of risk measures, VaR and ES, are unified in a more general two-parameter family of risk measures, called the Range-Value-at-Risk (RVaR). The family of RVaR was introduced in Cont et al. [4] in the context of robustness properties of risk measures. More importantly, RVaR can be seen as a bridge connecting VaR and ES. Embrechts et al. [7] addressed the problem of risk sharing among agents using the RVaR and Fissler and Ziegel [9] discussed about the elicitability of the RVaR. Much like ES, it has been shown that RVaR is not elicitable. The author shows that a triplet of RVaR with two VaR components at different levels is elicitable.

We do not find much literature regarding the estimation of RVaR. In this paper, we define some of the nonparametric estimators of RVaR and compare their finite sample performance. The advantage of the nonparametric approach is that it does not require exact specification of the data generating process and hence it is robust against mis-specification of the marginal distribution. We compare the performance of nonparametric estimators of RVaR for varying p

and q, via Monte Carlo simulations. The paper is divided into four sections. In Sect. 2, we describe five nonparametric RVaR estimators. In Sect. 3 using Monte Carlo simulations, we compare the mean squared errors (MSE) of some of the RVaR estimators for five different models. In Sect. 4 we report the findings.

1.1 Definitions

Let ψ be the set of real valued random variables. We consider a random variable X as a loss of some portfolio. Let F be the distribution function of X, then $Q_p(X) = \inf\{x : F(x) \geq p\}$, $0 < p < 1$ is the quantile function. For, $0 < p < q < 1$, the three risk measures VaR_p, ES_p and $RVaR_{p,q}$ are defined as

$$VaR_p = \inf\{x \in \mathbb{R} : F(x) \geq p\},$$

$$ES_p = \frac{1}{1-p} \int_p^1 VaR_u du, \tag{1}$$

and

$$RVaR_{p,q} = \frac{1}{q-p} \int_p^q VaR_u du. \tag{2}$$

Note that from Eqs. (1) and (2), we get

$$RVaR_{p,q} = \frac{(1-p)ES_p - (1-q)ES_q}{q-p}. \tag{3}$$

2 Nonparametric Methods for Estimating RVaR

In this section we define the estimators of RVaR using Eq. (3). So we first define the nonparametric estimators of ES and then we define the estimators of RVaR. The non-parametric estimators of RVaR are defined in the following sections.

2.1 Empirical Estimator

Let $\hat{F}$ denote the empirical distribution of the observed losses $X_1, X_2, \ldots, X_n$ i.e.

$$\hat{F}(x) = \frac{1}{n} \sum_{i=1}^n I(X_i \leq x),$$

where $I(\cdot)$ is the indicator function and X_i is i.i.d with distribution F. By standard results on empirical distribution (see Van Der Vaart [12]), the pth quantile can be estimated by:

$$\hat{F}^{-1}(p) = X_{(i)}, \ p \in \left[\frac{i-1}{n}, \frac{i}{n}\right),$$

where $X_{(1)} \leq X_{(2)} \leq \cdots \leq X_{(n)}$ are the order statistics. The empirical estimator of expected shortfall is defined as

$$Emp_p = \frac{\sum_{i=[np]+1}^n X_{(i)}}{n - [np]},$$

where $[x]$ denotes the largest integer not greater than x. The empirical estimator can be re-written as

$$Emp_p = \frac{\sum_{t=1}^n X_t I(X_t \geq \hat{q}_p)}{[n(1-p)] + 1},$$

where $\hat{q}_p = X_{([np]+1)}$. Therefore the empirical estimator of RVaR is

$$Emp_{p,q} = \frac{(1-p)Emp_p - (1-q)Emp_q}{q - p}.$$

2.2 Brazauskas et al.'s Estimator

Let us recall that expected shortfall is defined as

$$ES_p = \frac{1}{1-p} \int_p^1 Q_u du.$$

Brazauskas et al. [3] defined an empirical estimator of ES_p as follows

$$\widehat{ES}_p = \frac{1}{1-p} \int_p^1 \hat{F}^{-1}(u) du.$$

Then the estimator of RVaR following Brazauskas et al. [3] can be written as

$$\widehat{RVaR}_{p,q} = \frac{(1-p)\widehat{ES}_p - (1-q)\widehat{ES}_q}{q - p}.$$

2.3 Kernel Estimator

If we use kernel distribution function instead of empirical distribution function in the Brazauskas et al.'s estimator, then we can call it as a kernel estimator. For a given kernel function K, if $F_{n,b}(x) = \frac{1}{n} \sum_{i=1}^n K\left(\frac{x-X_i}{b}\right)$ is the kernel distribution function and $F_{n,b}^{-1}$ is its quantile function, then we can write (Biswas and Sen [2])

$$Ker_p = \frac{1}{1-p} \int_p^1 F_{n,b}^{-1}(u) du.$$

$F_{n,b}$ is estimated by using the plug-in bandwidth proposed by Altman and Leger [1] defined as

$$h_{AL} = \left(\frac{1/4\widehat{V}}{\widehat{B}}\right)^{1/3} n^{-1/3},$$

where

$$\widehat{V} = \varrho(k)\frac{1}{n(n-1)} \sum_{i=1}^n \sum_{j=1, j \neq i}^n \frac{1}{\alpha} k\left(\frac{x_i - x_j}{\alpha}\right),$$

and $\widehat{B} = 0.25\widehat{D}(F)(\mu_2(k))^2$, where $\varrho(k) = 2\int_{-\infty}^{+\infty} xk(x)K(x)dx$, $\mu_2(k) = \int_{-\infty}^{+\infty} x^2 k(x)dx$ and

$$\widehat{D}(F) = \frac{1}{n^3 \alpha_b^4} \sum_{i=1}^n \sum_{j=1}^n \sum_{l=1}^n k_b'\left(\frac{x_i - x_j}{\alpha_b}\right) k_b'\left(\frac{x_i - x_l}{\alpha_b}\right).$$

k_b' is the derivative of a kernel function k_b (not necessarily equal to k). In practice $\alpha_b = \alpha$ and $k_b = k$. If we use the kernel function as Epanechnikov kernel, [1] proved that an optimal choice is made by taking $\alpha = n^{-0.3}\hat{\sigma}(x_i)$, where $\hat{\sigma}(x_i) = \min\left\{\hat{s}, \frac{Q_3 - Q_1}{1.349}\right\}$, with $\hat{s}$ the sample standard deviation, and Q_1, Q_3 denote the first and third quartile respectively. We have estimated h_{AL} using the ALbw function in kerdiest package in R software. Then the kernel based estimator of RVaR can be written as

$$Ker_{p,q} = \frac{(1-p)Ker_p - (1-q)Ker_q}{q-p}.$$

2.4   Yamai and Yoshiba's Estimator

Yamai and Yoshiba [13] defined the following estimator of ES_p

$$ES_{p,\beta} = \frac{1}{n(\beta - p)} \sum_{i=[np]}^{n\beta} X_{(i)},$$

where β is a positive constant such that $X_{(1)} < X_{(2)} < .... < X_{([np])} < .... < X_{([n\beta])} < .... < X_{(n)}$. The empirical estimator Emp_p is similar to the above estimator for $\beta = 1$. If $1 - p \to 0$ as $n \to \infty$, we may use $\beta = 1 - r_n$ in the Yamai and Yoshiba's estimator $ES_{p,\beta}$, where r_n converges to zero at a faster rate than $1 - p$ as $n \to \infty$. We use $nr_n = max\{1, 0.25(n(1-p))^{2/3}/(ln(n(1-p)+1))^{2\iota}\}$, $\iota = 10^{-10}$. This choice is motivated by the choice of k_n in Hill's estimator [10]. The estimator of RVaR is given as

$$RVaR_{p,q,\beta} = \frac{(1-p)ES_{p,\beta} - (1-q)ES_{p,\beta}}{q-p}.$$

2.5   Filtered Historical Method

In this method a suitable time series model, such as an ARMA or a GARCH, is fitted to the asset return data. We fit a GARCH(1,1) model. Let $\hat{e}_i$, $i = 1, 2, \ldots, n$, denote the residuals of the fitted model. Then the filtered historical estimator of expected shortfall (Magadia [11]) is given by

$$FH_p = \frac{\sum_{\eta_t > q} \eta_t}{\sum_{\eta_t > q} I(\eta_t > q)},$$

where $\eta_t = \hat{e}_t - \frac{1}{n}\sum_{t=1}^{n} \hat{e}_t$ and $q = \eta_{([pn]+1)}$ is the $([pn]+1)$th order statistic of $\{\eta_1, \ldots, \eta_n\}$. The estimator of RVaR is

$$FH_{p,q} = \frac{(1-p)FH_p - (1-q)FH_p}{q-p}.$$

3   Simulation

In order to compare the behaviour of different estimators in finite samples, we compute the mean squared error (MSE) of the estimators by simulating observations from several models. We consider three models.

(i) $\{X_i\}_{i=1,2,\ldots}$ is an i.i.d. process, marginal distribution GPD with $\xi = 1/3$.

(ii) $\{X_i\}_{i=1,2,\ldots}$ is an i.i.d. process, marginal distribution Student's-t with 4 df.

(iii) $\{X_i\}_{i=1,2,\ldots}$ is an i.i.d. process, marginal distribution N(0,1).

To study the effect of dependence on the above mentioned RVaR estimators we consider the following ARMA (1,1) models in Drees [5]

$$X_i - \phi X_{i-1} = Z_i + \theta Z_{i-1},$$
$$(iv) \ \phi = 0.95, \ \theta = -0.6,$$
$$(v) \ \phi = 0.95, \ \theta = -0.9.$$

We use Monte Carlo (MC) simulation to approximate the MSE of each of these estimators. We have considered sample sizes $n = 30, 100, 250, 500, 1000$ and (p, q) to be (0.95, 0.97), (0.97, 0.99), and (0.99, 0.999). From each of the above models $(i) - (v)$ and for each combination of (n, p, q) we draw 1000 MC samples of size n. From each of these samples compute the values of the five estimators of $RVaR_{p,q}$ for various values of (p, q). From these values we compute the MC estimate of the MSE of that estimator for different choices of (n, p, q) and the underlying model. MSE1 denotes the mean squared error of $Emp_{p,q}$, MSE2 for $\widehat{RVaR}_{p,q}$, MSE3 for $Ker_{p,q}$, MSE4 for $RVaR_{p,q,\beta}$ and MSE5 for $FH_{p,q}$. In the next section we report the findings.

4   Findings

Following are the observations.

1. No estimator uniformly outperforms the other estimators. However we can identify some conditions under which some of these estimators performs well.
2. For $100 \leq n \leq 500$ and (0.99, 0.999) the estimator $\widehat{RVaR}_{p,q}$ outperform all the estimators for GPD and for $n \leq 500$ and (0.99, 0.999) the estimator $\widehat{RVaR}_{p,q}$ outperform the empirical estimator $Emp_{p,q}$ for Normal and Student's t. We also observe that for all choices of (n, p, q) the estimator $FH_{p,q}$ outperform $Emp_{p,q}$ for Normal and Student's.
3. For GPD $Emp_{p,q}$ and $\widehat{RVaR}_{p,q}$ are best except (30, 0.95, 0.97) case where $RVaR_{p,q,\beta}$ is best. For Normal $\widehat{RVaR}_{p,q}$ and $FH_{p,q}$ are best except for (30, 0.99, 0.999) case where $Ker_{p,q}$ is best. For t, $\widehat{RVaR}_{p,q}$ and $FH_{p,q}$ are best except for (1000, 0.99, 0.999) case where $Emp_{p,q}$ is best.
4. For the ARMA models the estimators $\widehat{RVaR}_{p,q}$, $FH_{p,q}$ and $Ker_{p,q}$ seems to be more accurate than $Emp_{p,q}$ for all choices of (n, p, q). $FH_{p,q}$ is almost always best except a few cases where $Ker_{p,q}$ is better, but even in those cases the difference is small.

Notes and Comments. The above observations suggest that estimators $\widehat{RVaR}_{p,q}$ and $FH_{p,q}$ are preferable choices for estimation of RVaR for $n \leq 500$ and $(0.99, 0.999)$ for all the i.i.d. models. For Normal and Student's t we can consider $FH_{p,q}$ for other choices of (p, q) and large n. If the data are generated by ARMA model then $FH_{p,q}$ and $Ker_{p,q}$ seems to perform well for all the choices of (n, p, q) considered in our study.

References

1. Altman, N., Leger, C.: Bandwidth selection for kernel distribution function estimation. J. Stat. Plan. Infer. **46**(2), 195–214 (1995)
2. Biswas, S., Sen, R.: Kernel based estimation of spectral risk measures. arXiv preprint (2019)
3. Brazauskas, V., Jones, B., Madan, L., Zitikis, R.: Estimating conditional tail expectation with acturial application in view. J. Stat. Plan. Infer. **138**, 3590–3604 (2008)
4. Cont, R., Deguest, R., Scandolo, G.: Robustness and sensitivity analysis of risk measurement procedures. Quant. Finance **10**(6), 593–606 (2010)
5. Drees, H.: Extreme quantile estimation for dependent data,with application to finance. Bernoulli **9**(1), 617–657 (2003)
6. Embrechts, P., Wang, B., Wang, R.: Aggregation-robustness and model uncertainty of regulatory risk measures. Finance Stochast. **19**(4), 763–790 (2015)
7. Embrechts, P., Liu, H., Wang, R.: Quantile-based risk sharing. Oper. Res. **66**(4), 936–949 (2018)
8. Emmer, S., Kratz, M., Tasche, D.: What is the best risk measure in practice? A comparison of standard measures. J. Risk **18**(2), 31–60 (2015)
9. Fissler, T., Ziegel, J.F.: On the elicitability of range value at risk. Stat. Risk Model. **38**(1–2), 25–46 (2021)
10. Hill, J.B.: Expected shortfall estimation and gaussian inference for infinite variance time series. Unpublished monograph (2013). http://www.unc.edu/~jbhill/expected_short_robust_JBHILL.pdf
11. Magadia, J.: Confidence interval for expected shortfall using bootstrap methods. In: 4th Annual BSP-UP Professional Chair Lectures, 21–23 February. Bangko Sentral ng Pilipinas, Malate, Manila (2011)
12. Van der Vaart, A. W.: Asymptotic Statistics. vol. 3. Cambridge University Press, Cambridge (2000)
13. Yamai, Y., Yoshiba, T.: Comparative analysis of expected shortfall and value-at-risk: their estimation error, decomposition, and optimization. Monet. Econ. Stud. **1**, 87–122 (2002)

A Fixed Career Length Versus a Fixed Retirement Age: An Analysis per Socio-Economic Groups

M. Carmen Boado-Penas[1(✉)], Pierre Devolver[2], Şule Şahin[1], and Carlos Sunyer[3]

[1] University of Liverpool, Liverpool, UK
`carmen.boado@liverpool.ac.uk`
[2] Université Catholique Louvain, Ottignies-Louvain-la-Neuve, Belgium
[3] Universidad Carlos III Madrid, Madrid, Spain

Abstract. An increase in pension age is one of the most common reforms to restore the financial sustainability of the pension scheme. However, this measure might damage individuals from lower socio-economic groups who, on average, have shorter life expectancy. This paper aims to analyse to what extent a fixed reference career duration instead of a reference retirement age acts as an automatic balancing mechanism that contributes to strength the actuarial fairness/equity in the current public pension schemes and mitigate the inequality among different socio-economic groups.

Keywords: Mortality differentials · Labour entry age · Public pension · Redistribution · Reform

1 Introduction

The decline in fertility rates, the increase in longevity and the current forecasts for the ageing of the baby boom generation, all point to a substantial increase in the age dependency ratio, and this will raise serious concerns for the sustainability of Pay-As-You-Go (PAYG) pension schemes. This is a worldwide problem, and consequently, many countries have already carried out some parametric reform, or even structural reforms, of their pension systems [1].

The most common measure to improve the financial sustainability is to strengthen the incentives to work, thereby enlarging the total contribution base of the system. This can be done through i) increases in the statutory retirement age; ii) tightening of early retirement provisions; iii) higher financial incentives to delay the retirement age; and iv) greater possibilities to combine work and pensions. In practice, on average across all OECD countries, the retirement age of those workers entering the labour market in 2020 is expected to increase from 64 to 66.1 years by the 2060s [1]. This partly explains the generalised increase in the effective retirement age observed in recent years.

However, increasing the effective retirement age, although desirable considering the system's financial sustainability, has some unintended consequences. It contributes to enlarge income disparities among pensioners with the pensions of low-income individuals relative to that of high-income retirees falling by 12% on average for OECD countries [2]. This fact is particularly acute given that since the 1980s we are in a context of growing inequality, with no prospect of this trend to be reversed soon [3].

The impact of the increase in retirement age on pensioners' inequality is mainly due to the heterogeneity in life expectancy of the socio-economic classes—normally assessed through income, education, or occupation. A close link between level of education and mortality has been well documented (see [4] and [5]). The evidence suggests that the mortality rates are negatively correlated with social-economic status (i.e., individuals with higher socio-economic status live longer than those belonging to lower socio-economic groups and the size of mortality differences across subgroups has increased over time).[1] However, the net effect of income redistribution between subpopulations considering both gender and socio-economic status is not clear for some subgroups, such as females with low socio-economic status or men from a high socio-economic class.

Far from being anecdotal evidence, the gap in life expectancy in the US between the richest 1% and poorest 1% reaches 14.6 years among men and 10.1 years among women [6]. Although narrower, it is a generally growing phenomenon that can also be extrapolated to other economies [2]. For this reason, increases in the retirement age - so popular among the reforms carried out so far - in practice generate a Matthew effect by implicitly taxing (subsidizing) people with a low (high) income level [7]. The magnitude of this implicit tax on the poorest is estimated, for instance, in the case of the US and Germany, to be as high as 21.9% and 12.3% for the lower part of the income distribution, respectively [8]. Consequently, and due to the magnitude of the issue, "limit(ing) the impact of socio-economic differences in life expectancy on pension benefits" appears as one of the key recommendations of the OECD report (2017).

2   Objective

This paper aims to analyse to what extent a fixed reference career duration instead of a reference retirement age acts as an automatic balancing mechanism that contributes to strength the actuarial fairness/equity in the current public pension schemes and mitigate the inequality among different socio-economic groups.

The reasoning behind the proposal is as follows. Low-skilled workers, since they generally spend less time on studying, are the first to enter the labour market and are also the ones who perform the most arduous jobs. This is, in turn, correlated with their life expectancy, as empirically observed. Therefore,

---

[1] Queisser and Whitehouse (2006) and Ayuso et al. (2016, 2017) provide comprehensive reviews of mortality differentials across socio-economic groups and the implications of heterogeneity for pension reform and design.

in the case of early retirement, since they have a shorter life expectancy, the pay-out period is longer than for those high-skilled individuals with higher life expectancy, on average. The opposite would be true for high-skilled workers. By investing more time in education attainment, the individuals would join the labour market later and, thus, receive their pension at a later date. This delay in retirement would be potentially compensated by a longer life expectancy.

If the pension system establishes a fixed reference career duration the desired outcome of a pension system in terms of equity would be strengthened by narrowing the relationship between contributions paid and benefits expected to be received. Moreover, the proposed scheme would be "simple, operational and transparent", as Ayuso et al. (2016) suggest, while at the same time would promote another desirable property, such as flexibility [9].

In the remainder of this paper, we first define the term actuarial fairness in pension systems. In Sect. 4, we describe the data we use while Sect. 5 focuses on policy implications.

3   Actuarial Fairness

A pension scheme is viewed as actuarially fair (or equitable) if for all individuals the (expected) present value of lifetime contributions is equal to the (expected) present value of pension benefits. Any difference between these two present values is defined as an income redistribution towards or away from the individual [10].

The measurement of the lifetime redistribution from an individual's perspective can be investigated through the ratio between the present value of benefits paid during the retirement and the contributions paid during their working career (see [11–13] and [14]). Thus, the ratio shows how much the benefit system returns to the participant for each euro paid.

A value of one indicates that the system is actuarially fair for an individual (i.e., she received pension benefits which correspond to her contributions)—by definition, there is no redistribution towards or away from any person [15]. A value greater (lower) than one indicates that an individual receives more (less) than she has been contributing-an individual faces an expected gain (loss) from the pension system.

4   Data

The continuous sample of working lives (CSWL) was first released in 2004 and has been renewed annually since then. It is composed of a random sample of 4% of the total population (approximately 1.2 million people) who had some relationship with the Spanish Social Security during the reference year, either as workers, unemployed or pension beneficiaries. For each individual, longitudinal administrative information is extracted with their social security, income tax and census records. In case an individual ceases to have a relationship with the Social Security or is deceased, he or she is replaced by another individual, drawn at random. Otherwise, he/she will continue to be part of the sample in the following

reference years. Due to these characteristics, it is an exceptionally rich database, widely used by the scientific community (e.g., [16,17] and [18]). As far as this paper is concerned, it allows us to estimate individuals' life expectancies [19], their length of working life [20] and relationships between the level of educational attainment and the type of work carried out by the individual [21]. Hence, this dataset is optimal for the objectives of this paper.

As shown in Fig. 1 the difference in the labour entry age between individuals with primary studies and those with university studies ranges between 1.5 and 4.75 years. Although the gap reaches its lowest value in 2007, since then it has been growing over time. This could be due to the use of education as a "refuge" from the labor market, which, in the case of Spain, is especially precarious for young people [21].
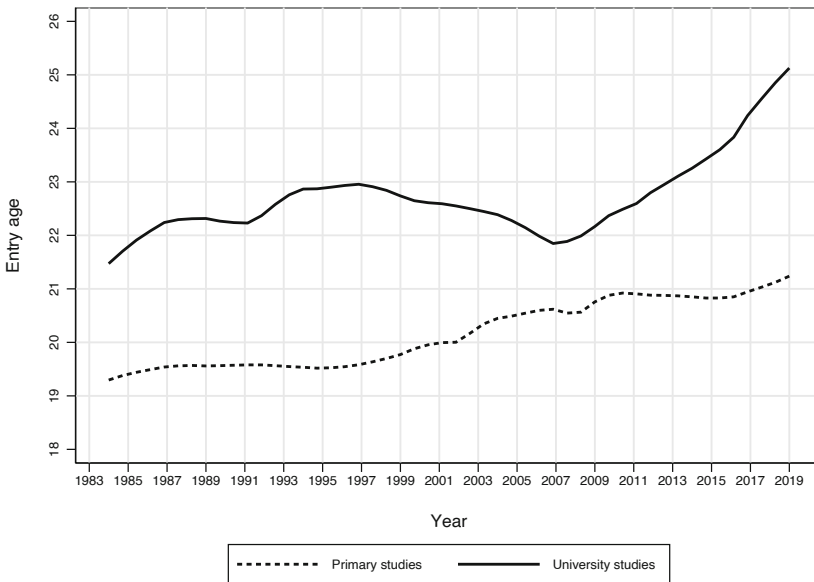
Fig. 1. Entry age (median) by educational attainment level in Spain for the period 1984–2019. Own source based on CSWL.

5   Policy Implications

Ageing populations have put existing retirement systems under significant strain. Raising the retirement age is one of the most common reforms to increase pension revenue. However, this measure might damage individuals from lower socioeconomic groups who, on average, have shorter life expectancies. Research is needed to consider both the type of work and the length of the contributory period of the individuals so that there is a closer relationship between the contributions and pensions.

Although the evaluation of the redistribution effects on the pension system has attracted wide attention, in this paper, for the very first time according to the authors' knowledge, we study the effects of fixing a career duration on actuarial fairness. This could be an alternative reform to be taken by governments, easy to understand by individuals and at the same time could improve the equity among the participants of the system. Moreover, this measure seeks not to damage individuals from lower socio-economic groups moving away from the traditional and unpopular increase of retirement age.

References

1. OECD: Pensions at a glance 2021. OECD Publishing, Paris (2021). https://www.oecd.org/publications/oecd-pensions-at-a-glance-19991363.htm. Accessed 5 Jan 2022
2. OECD: Preventing Ageing Unequally. OECD Publishing, Paris (2017). https://doi.org/10.1787/9789264279087-en
3. Piketty, T.: Capital in the Twenty-First Century. The Belknap Press of Harvard University Press (2014). ISBN-13: 978-0674430006
4. Lleras-Muney, A.: The relationship between education and adult mortality in the United States. Rev. Econ. Stud. **72**, 189–221 (2005)
5. Miech, R., Pampel, F., Kim, J., Rogers, R.G.: The enduring association between education and mortality: the role of widening and narrowing disparities. Am. Sociol. Rev. **76**, 913–934 (2011)
6. Chetty, R., et al.: The association between income and life expectancy in the United States, 2001–2014. Jama **315**(16), 1750–1766 (2016)
7. Ayuso, M., Bravo, J.M., Holzmann, R.: Adressing longevity heterogeneity in pension scheme design. J. Financ. Econ. **6**(1), 1–21 (2017)
8. Ayuso, M., Bravo, J.M., Holzmann, R.: On the heterogeneity in longevity among socioeconomic groups: scope, trends, and implications for earnings-related pension schemes (2016). https://ssrn.com/abstract=2810471
9. Schokkaert, E., Devolder, P., Hindriks, J., Vandenbroucke, F.: Towards an equitable and sustainable points system. A proposal for pension reform in Belgium. J. Pension Econ. Financ. **19**(1), 49–79 (2020)
10. Boado-Penas, M.C., Haberman, S., Naka, P.: Fairness and annuity divisors for notional defined contribution pension schemes. J. Pension Econ. Financ. **21**(2), 143–167 (2022)
11. Brown, J.R.: Differential mortality and the value of individual account retirement annuities. In: Feldstein, M., Liebman, J.B. (eds.) The Distributional Aspects of Social Security and Social Security Reform, NBER Conference Report Series, pp. 401–440. University of Chicago Press, Chicago and London (2002)
12. Brown, J.R.: Redistribution and insurance: mandatory annuitization with mortality heterogeneity. J. Risk Insur. **70**(1), 17–41 (2003)
13. Mazzaferro, C., Morciano, M., Savegnago, M.: Differential mortality and redistribution in the Italian notional defined contribution system. J. Pension Econ. Financ. **11**(4), 500–530 (2012)
14. Belloni, M., Maccheroni, C.: Actuarial fairness when longevity increases: an evaluation of the Italian pension system. Geneva Pap. Risk Insur. - Issues Pract. **38**(4), 638–674 (2013)

15. Queisser, M., Whitehouse, E.: Neutral or fair?: Actuarial concepts and pension system design. OECD Social, Employment and Migration Working Papers, No. 40 (2006)
16. Arellano, M., Bonhomme, S., De Vera, M., Hospido, L., Wei, S.: Income risk inequality: evidence from Spanish administrative records (2021)
17. La Roca, D., Puga, D.: Learning by working in big cities. Rev. Econ. Stud. **84**(1), 106–142 (2017)
18. Hospido, L., Moral-Benito, E.: The public sector wage premium in Spain: evidence from longitudinal administrative data. Labour Econ. **42**, 101–122 (2016)
19. Pérez-Salamero González, J.M., Regúlez-Castillo, M., Vidal-Meliá, C.: Differences in life expectancy between self-employed workers and paid employees when retirement pensioners: evidence from Spanish social security records. Eur. J. Popul. **37**(3), 697–725 (2021)
20. Dudel, C., López Gómez, M.A., Benavides, F.G., Myrskylä, M.: The length of working life in Spain: levels, recent trends, and the impact of the financial crisis. Eur. J. Popul. **34**(5), 769–791 (2018)
21. Bentolila, S., Felgueroso, F., Jansen, M., Jimeno, J.F.: Lost in recessions: youth employment and earnings in Spain. SERIEs (2021). https://doi.org/10.1007/s13209-021-00244-6

Nonparametric Test for Financial Time Series Comparisons

Stefano Bonnini and Michela Borghesi$^{(\boxtimes)}$

Department of Economics and Management, University of Ferrara,
Via Voltapaletto 11, 44121 Ferrara, Italy
{bnnsfn,michela.borghesi}@unife.it

Abstract. The aim of this work concerns the problem of comparing groups of time series, in particular financial time series. Some empirical studies have been published on the topic. However, there is a lack of literature about valid statistical inferential approaches regarding the comparison between groups. In particular, we focus on a two-sample testing problem with the goal of comparing two different groups of financial titles in a given time period. The dataset consists in the time series of the financial returns of the two groups of titles. The problem can be defined as a multivariate test on central tendency and the proposed solution is based on the methodology of combined permutation tests. The application presented in this study concerns the comparative evaluation of the financial performance of ESG titles.

Keywords: Nonparametric statistics · Permutation tests · Financial performance

1 Introduction

A *portfolio* is defined, in financial language, as a set of assets that can include both real assets and financial instruments. It is possible to calculate the expected return and risk of a portfolio given characteristics and weights of the single assets. If R_i represents the return of the i-th title of the portfolio and R_P the performance of the portfolio, then the expected return of the portfolio can be written as

$$E(R_P) = \sum_{i=1}^{n} w_i E(R_i) \tag{1}$$

and its standard deviation is

$$\sigma_P = \sqrt{\sum_{i=1}^{n} w_i^2 \sigma_i^2 + \sum_{i=1}^{n} \sum_{j=1 j \neq i}^{n} w_i w_j \sigma_i \sigma_j \rho_{ij}} \tag{2}$$

where σ_i represents the standard deviation of the return of i-th title and ρ_{ij} the Pearson correlation index of the returns of the i-th and the j-th title. The weight

of each individual title in the previous formulas usually refers to the proportion of the value of that asset to the total value of the portfolio. Since the evaluation of the weights' effect is not a goal of our study, from now on we will consider the case of equal weights.

Hence, in order to compare the financial performance of two different groups of financial titles, a reasonable choice seems to be a two-sample test on location using data of a given time period. The dataset consists in the time series of the financial returns of the two groups of titles. The problem can be defined as a multivariate test on central tendency where the mean returns of the two groups are compared for each of the time points in the considered time period. The motivating example in this study concerns the comparative evaluation of the financial performance of ESG titles.

The acronym ESG stands for Environmental, Social, Governance and is used in economic/financial field to indicate all those activities related to responsible investment (RI) that pursue the typical objectives of financial management taking into consideration environmental, social and governance aspects.

In the literature, many works agrees that ESG investing has become the focus of many investors as more and more research shows that it has implications for both risk and return. ESG-related investments now account for approximately 26% of all professionally managed businesses and it is now known that ESG is a shortcut to environmental, social and governance metrics.

For example [1] discuss some of the models for integrating ESG into equity portfolios and argue that integrating ESG stocks has important potential benefits for investors. In favor of sustainable investments, we find a further study [6] which examines the impact of companies' environmental, social and governance (ESG) initiatives on financial performance. Also in favor of these investments we find [4,5] who believe that the integration of ESG stocks into a portfolio has the potential to improve the valuation of the portfolio.

A recent contribution argues that it is necessary carrying out further studies on this topic, because the relationship between ESG score and brand value, while positive, has not always been constant over time (see [3]). Therefore, the need of further investigations, in order to strengthen the results obtained so far, is evident.

Section 2 focuses on the presentation of the statistical problem, followed by the description of the methodological proposal (Sect. 3). The results of the application of the proposed method to a case study are reported in Sect. 4. Section 5 includes concluding remarks.

2   Statistical Problem

The analysis of different time series may consist of methods used for predicting future values based on observed time series, or direct comparison between different time series. A descriptive method for comparing time series can be based on a similarity measure, which includes metric and non-metric methods.

However, the use of metrics is not always possible. In any case, the application of these methods directly to the original time series is computationally complex (especially within a cluster analysis) and therefore shorter representations are often created and the distance between pairs of time series approximations is estimated [12]. The existing approaches are not applicable when the two dynamic models have a different number of component variables. In this case, the only method is to establish the average behavior of each model and then compare the two univariate average time series but this approach may be useless or inappropriate for some applications [10]. Several methods have been proposed to calculate the distance between *univariate* time series. Some of the most used are Euclidean distance, Manhattan distance, Dynamic Time Warping (DTW) and Longest Common Subsequence (LCSS). [11] propose a new method: the semi-metric time series (Semi Metric Ensemble Time Series or SMETS), which is able to compare *multivariate* time series of arbitrary size. The method also takes into account the differences attributable to univariate unmatched components when one of the two time series has a higher dimensionality than the other but it is computationally complex.

We focus on the testing problem where the central tendency of the distribution of two different groups of financial titles is compared with respect to a time series with T time points. We consider observational data of the returns of two samples of titles: group 1 and group 2. The general problem can be broken down into T partial problems, one for each time point t, with $t = 1, 2, \ldots, T$. Hence, we have T partial null hypotheses $H_{01}, \ldots, H_{0T}$ and T partial alternative hypotheses $H_{11}, \ldots, H_{1T}$ and the overall problem can be defined as

$$H_0 : \bigcap_{t=1}^{T} H_{0t} \qquad H_1 : \bigcup_{t=1}^{T} H_{1t}. \qquad (3)$$

We would like to compare the performances of ESG titles with non-ESG titles for each time point of the series with T time points. Let R_{kt} be the financial performance (return) of the titles of the k-th group at time t, with $k = 1, 2$. The hypothesis of the testing problem can be written as follows:

$$H_0 : \bigcap_{t=1}^{T} \left(R_{1t} \overset{d}{=} R_{2t} \right) \qquad H_1 : \bigcup_{t=1}^{T} \left(R_{1t} \overset{d}{>} R_{2t} \right). \qquad (4)$$

In our application, $k = 1$ corresponds to ESG titles and k = 2 to non-ESG titles, hence we want to test the hypothesis that, for at least one time point t, the former have a better performance than the latter.

3   Methodological Solution

The proposed solution concerns the application of a combined permutation test [8]. This family of nonparametric methods is suitable when the problem can be broken down into q sub-problems or partial tests. Hence, we have q partial null

hypotheses $H_{01}, \ldots, H_{0q}$ and q partial alternative hypotheses $H_{11}, \ldots, H_{1q}$ and the overall problem can be defined as

$$H_0 : \bigcap_{i=1}^{q} H_{0i} \qquad H_1 : \bigcup_{i=1}^{q} H_{1i} \qquad (5)$$

Basically, it is a multiple test where each sub-problem consists in testing the null hypothesis H_{0i} versus the alternative hypothesis H_{1i}.

Permutation methods can be used, provided that mean and variance of the populations are assumed to be finite and exchangeability under the null hypothesis holds [7]. Permutation methods are preferable to parametric solutions when the underlying distribution is unknown or cannot be assumed according to asymptotic theories (hence especially for small samples). Moreover the dependence between the test statistics of the partial problems does not need be explicitly modeled, as in the likelihood approach or other parametric methods. In particular, with the combined permutation tests, the dependence structure is implicitly taken into account by permuting the rows of the dataset and the application of a suitable combining function Ψ. The sufficient statistic of permutation tests is represented by the observed dataset.

Whether variance or moments of the two distributions of returns are constant or not over time does not affect the application of the permutation multiple test. Therefore the method can also be applied in case of non-stationary behavior. Autocorrelation or other types of dependence over time are not an obstacle to the application of the test because the procedure implicitly take the dependence into account. The only important condition is exchangeability of the statistical units, the financial titles in our problem, under H_0 (satisfied in the null hypothesis under study) [2,8].

Without loss of generality, we can assume that the null (partial and overall) hypotheses are rejected for large values of the test statistics. Let $L_i(t) = P(T_i \geq t|X)$ denote the significance level function of the i-th partial test, T_i the test statistic of the i-th partial test and t_i a given value taken by T_i. The (univariate) combined test statistic is $T_\Psi = \Psi(l_1, \ldots, l_k)$ where $l_i = L_i(t_i)$. Ψ must be non-increasing function of the arguments and satisfy mild conditions such as: it tends to its supremum (possibly not finite) when one argument tends to zero and, $\forall \alpha \in (0, 1)$, the critical value of T_Ψ is assumed to be finite and strictly less than the supremum. For the problem under study we used for the test statistics the sum of values in sample 1 (permutationally equivalent to the sample mean of sample 1 and the difference of the sample means of sample 1 and sample 2). Formally

$$T_i = \sum_{u=1}^{n_1} R_{1iu} \qquad (6)$$

where R_{1iu} represents the return of the u-th title in sample 1 at time i.

When the length T of the time series tends to infinity (keeping the sample sizes fixed) the test is consistent. In fact, adding new partial tests (i.e. time

points) under H_1, determines an increase in the test power and when T diverges the power tends to one.

We will apply to our case study three main combination functions: Fisher, Liptak and Tippett (see [2]).

4   Case Study

The data used for the analysis were downloaded from the Yahoo Finance site considering a five-year period. All the securities in the S&P 500 list [9] were searched. ESG titles were differentiated from non-ESG titles and daily time series of each of them were downloaded. At the end of the data collection we obtained 453 companies divided into 139 non-ESG and 314 ESG and 1258 daily dates in which each company had a closing return value.

With the methodology explained in Sect. 3, a combined permutation test was applied to test whether ESG titles have a higher performance than non-ESG titles over time.

A significance level α equal to 0.05 was set and the p-values of the test were calculated.

In conclusion, all three mentioned combined tests were applied and no p-value was less than the significance level α (0.05). In general we obtain very high p-values, as can be seen in Table 1. This means that there is no significant

Table 1. Overall p-values

	Fisher	Liptak	Tippett
p-value	0.8241758	0.8191808	0.8441558

difference in performance between the two types of titles. The empirical findings in our study do not confirm the thesis that ESG investments are financially convenient over time, as reported by some authors in the literature of financial economics. According to the comparison of the two samples of financial time series, there is not empirical evidence in favor of this hypothesis.

5   Concluding Remarks

The goal of this study is to compare the trajectories over time of the central tendencies of two populations, in particular in the case of financial time series.

In the empirical literature on time series, there is a lack of contributions on comparing time series for inferential purposes. The few existing works adopt approaches that are not efficient, computational convenient or suitable for tests of hypotheses.

We propose the application of a procedure based on a combined permutation test. This method is very useful for solving complex problems, especially multivariate problems or problems in which a multivariate statistical test might be suitable. The main advantage, over other standard parametric methods, is that it is powerful, flexible because distribution-free, and satisfy important properties such as unbiasedness and consistency.

The application of the test to a sample concerning ESG and non-ESG titles does not bring to empirical evidence in favor of the hypothesis that ESG titles performance is greater than that of non-ESG titles.

References

1. Bender, J., Bridges, T., He, C., Lester, A., Sun, X.: A blueprint for integrating ESG into equity portfolios. J. Invest. Manag. **16**(1), 44–58 (2018)
2. Bonnini, S., Corain, L., Marozzi, M., Salmaso, L.: Non Parametric Hypothesis Testing: Rank and Permutation Methods with Application in R. Wiley (2014)
3. El Zein, S.A., Consolacion-Segura, C., Huertas-Garcia, R.: The role of sustainability in brand equity value in the financial sector. Sustain. MDPI **12**(1), 1–19 (2019). https://doi.org/10.3390/su12010254
4. Giudici, G., Bonaventura, M.: La relazione tra rating ESG e performance di mercato: uno studio sui titoli dell'indice Stoxx Europe 600. Banor (2019)
5. Khan, M., Serafeim, G., Yoon, A.: Corporate sustainability: first evidence on materiality. Harvard Business School Working Paper, No. 15-073 (2015)
6. Kiong Ting, I.W., Azizan, N.A., Bhaskaran, R.K., Sukumaran, S.K.: Corporate social performance and firm performance: comparative study among developed and emerging market firms. Sustainability **12**, 26 (2019). https://doi.org/10.3390/su12010026
7. Pesarin, F.: Multivariate Permutation Tests: With Application in Biostatistics. Wiley, Chichester, New York (2001)
8. Pesarin, F., Salmaso, L.: Permutation Tests for Complex Data. Theory, Applications and Software, Wiley, Hoboken (2010)
9. Steadman, R., Perrone, D.: The S&P 500 ESG Index: Integrating Enviromental, Social, and Governance Values into the Core. S&P Dow Jones Indices (2019)
10. Sutcliffe, C.M.S.: Stock Index Futures. Ashgate Publishing Ltd., Aldershot (2006)
11. Tapinos, A., Mendes, P.: A method for comparing multivariate time series with different dimensions. PLoS One **8**(2), e54201 (2013). https://doi.org/10.1371/journal.pone.0054201
12. Veltkam, R.C.: Shape matching: similarity measures and algorithms. In: Proceedings International Conference on Shape Modeling and Applications, pp. 188–197. IEEE (2001)

Innovative Parametric Weather Insurance on Satellite Data in Agribusiness

Maria Carannante[1(✉)], Valeria D'Amato[1], Paola Fersini[2], and Salvatore Forte[3]

[1] Università degli Studi di Salerno, Via Giovanni Paolo II, 132, 84084 Fisciano, Italy
{mcarannante,vdamato}@unisa.it
[2] LUISS Guido Carli University, Viale Romania, 32, 00197 Rome, Italy
pfersini@luiss.it
[3] Giustino Fortunato University, Via Raffaele Delcogliano, 82100 Benevento, Italy
s.forte@unifortunato.eu

Abstract. Weather Resilience against adverse meteorological events became a big issue in several sectors of the economic system. Innovative tools can complement traditional strategies and speed up recovery. In the insurance industry, the parametric coverages based on the use of the bioclimatic parameters correlated to the client's loss, represent an interesting and alternative risk solutions. Nevertheless, the risk that the index is measured with spatial distance to production location, i.e. the basis risk, can hinder the market diffusion process. The real time data collected by satellites can mitigate the basis risk and combined with main features of the Agribusiness in the personal parametric weather insurance lead to an adequate risk management.

Keywords: Weather insurance · Parametric insurance · Satellite data

1 Introduction

The climate changes represent one of the big issue of the Planet affecting several sectors of the economic system. Agricultural production depends on the weather more than any other sector, so that the extreme and not extreme adverse meteorological events cause a lot of concerns to achieve an accurate risk management. The pressing need to create tools to increase climate resilience encouraged the insurance industry to design new solutions, particularly revolved to measurable index and based on predefined triggers or pay out mechanisms without necessarily the connection to the occurrence of a physical damage.

In the weather coverage field, classical strategies are characterized by insurance product or derivative-based product. The two instruments feature different regulatory, accounting, tax, and legal issues, the risk transfer characteristics and benefits are similar. In particular, the insurance business traditionally provides Indemnity-Based Insurance, that consists in the payment of a predetermined insured amount, which can be agreed or based on insured's historical returns. Conversely, the Index-based Insurance is composed by the payment of

an indemnity on the basis of a realized average yield related to an area or a specific weather index. Nevertheless, even if the index-based schemes remove the problems of moral hazard and adverse selection, the risk that occurs if index is measured with spatial distance to production location, i.e. basis risk, can significantly affects this kind of policy [2,3]. Because no field loss assessment is made under index insurance, the payout is based entirely on the index measurement and may be either higher or lower than the actual loss [6]. As the climate-related weather risks become increasingly complex, the requests for innovative products increased. In this context, the parametric weather insurance can be designed to cover both specific catastrophic losses and frequency losses. In the paper, we propose an innovative scheme of the parametric weather insurance, we called Personal Parametric Weather Insurance (PPWI), where the bioclimatic indicator depends on the satellite data, collected by the weather stations and processed by machine learning technique. Improved data and models enable parametric cover as an increasingly efficient, affordable and viable option in the market.

The layout of the paper is the following. Section 2 introduces the complexity of the satellite data to be processed in order to obtain an accurate information about risky days in the agribusiness. Section 3 propose the model. In Sect. 4 we provides the main findings.

2   Methodology and Satellite Data

Bioclimatic risk indexes using satellite data are obtained by a data pre-processing procedure to transform data from a storage format to a matrix. In this sense, we refer to two sources of data: the NASA MERRA-2 database and the Harmonized World soil database (HWSD), both collecting the data in raster format. It is therefore necessary to transform the raster data into a row by column matrix. Before transforming the dataset it was necessary to automate the NASA data download procedure, following [4], since each dataset contained daily data and the time interval considered is from 1980/01/01 to 2020/08/30, totaling 14'854 daily datasets. Afterwards, pixels will be further refined by selecting only those containing mainland, with the support of HWSD. MERRA-2 database is structured in such a way as to have a series of daily or intraday datasets, each of which detects weather variables regarding a specific topic. For our application, the Single-Level Diagnostics (M2SDNXSLV version 5.12.4) dataset is used [1].

3   Personalised Parametric Weather Insurance

Weather risk index is used to define two index insurance strategies. Type I covers both extreme and non-extreme weather conditions and Type II covers only extreme weather conditions [5]. Let I the weather index, K the realized yield, $\overline{w}$ and $\underline{w}$ the extreme and non-extreme event thresholds, with, $\overline{w} > \underline{w}$ and μ_1 and μ_2 the expected yield associated to non-extreme and extreme events, with $\mu_1 < \mu_2$:

$$\gamma_t^I = \begin{cases} \mu_2 - K \ if & I \geq \overline{w} \\ \mu_1 - K \ if \ \underline{w} \leq I < \overline{w} \\ 0 & otherwise \end{cases} \tag{1}$$

$$\gamma_t^{II} = \begin{cases} \mu_2 - K \ if & I \geq \overline{w} \\ 0 & otherwise \end{cases} \tag{2}$$

Once defined γ_t^I and γ_t^{II}, we synthesize the financial payoff function γ as follows:

$$\gamma = \max \left\{ \mu_i - K, \ 0 \right\}, \ i = 1, 2 \tag{3}$$

Since the payoff function depends on the insurance layer, we define:

$$\eta = min \left\{ \max \left\{ \mu_i - K, \ 0 \right\}, \ \overline{w} - \underline{w} \right\}, \ i = 1, 2 \tag{4}$$

From Eq. 4 we compute the fair premium at time zero as follows:

$$\pi_0 = e^{-rT} \mathbb{E}[\eta] \tag{5}$$

where r is the risk free rate and T the time interval.

4 Numerical Application

The insurance contract we propose aims at saving the maize crop against one of the main risk affecting the agribusiness, i.e. the risk of dry conditions. The table ratings are related to the hectare, basing on the variable soil type expressed by the following modalities: $\mathcal{A}$: acidic; $\mathcal{B}$: neutral; $\mathcal{C}$: saturated. The soil type ($\mathcal{A}$, $\mathcal{B}$ or $\mathcal{C}$) is detected according to the information provided by the geographic coordinate system on the corresponding pixel. Based on the soil type, the pure price of the contract is quoted according to the fair premium due to the expected loss of the insurer at time t, L_t and the θ being the safety loading that is the double value of the cost of capital rate CoC, assuming that the insurance company follows a risk capital framework principle with a minimum solvency ratio of 200%:

$$\mathbb{E}[L_t] + \theta \tag{6}$$

where CoC represents the 6% of the difference between the 99.5% quantile $VaR_{99.5\%}(L_T)$ and the expected cash outflow. We assume a expenses loading rate of 20%, being $P_t = \frac{PurePremium}{1-20\%}$. It is noteworthy that parametric policies can offer many benefits, such as the decreasing of the premium rate due to the event that triggers the payout and the mechanism used for payouts determine what will be paid to the customers a clearer predetermined amount, instead of the traditional solutions, where a claims adjuster needs to assess and determine damages. For sake of clarity, as the soil type changes, the premium rates varies as the specific probability distribution of the indemnity. The policy is designed and adapted to customer specific needs, those are reflected on the quantiles of the following bioclimatic indicators:

1. the number of days without rainfall (IP):
   – No adverse events: IP < 175;
   – Non-extreme adverse event: $175 \leq$ IP < 200;
   – Extreme adverse event: IP ≥ 200.
2. the number of days with dry conditions (IH):
   – No adverse events: IH < 35;
   – Non-extreme adverse event: $35 \leq$ IH < 50;
   – Extreme adverse event: IH ≥ 50.

Table 1 shows the joint probabilities of the variables IP, IH and the corresponding indemnity percentages of the insured crop value.

Table 1. Analysis of the joint probability of the reimbursement, i.e. the indemnity received by the policyholder

Probability of the reimbursement			
Soil Type $\mathcal{A}$			
	IP < 175	$175 \leq$ IP < 200	IP ≥ 200
IH < 35	31.67%	13.33%	2.50%
$35 \leq$ IH < 50	5.83%	12.50%	10.83%
IH ≥ 50	0.83%	7.50%	15.00%
Soil Type $\mathcal{B}$			
	IP < 175	$175 \leq$ IP < 200	IP ≥ 200
IH < 35	0%	0%	33.33%
$35 \leq$ IH < 50	0%	22.22%	10.11%
IH ≥ 50	0%	33.33%	1.00%
Soil Type $\mathcal{C}$			
	IP < 175	$175 \leq$ IP < 200	IP ≥ 200
IH < 35	10.82%	13.16%	7.60%
$35 \leq$ IH < 50	2.34%	6.73%	8.19%
IH ≥ 50	0.29%	11.70%	39.18%

In Table 1 we can observe that for the soil type $\mathcal{A}$ there is a greater probability of absence of adverse event for precipitation deficit indicator with respect the others, while for the soil type $\mathcal{B}$ the probability of extreme or non-extreme adverse event for precipitation deficit indicator is very high, associated to non-extreme or absence of precipitation deficit. On the contrary, for the soil type $\mathcal{C}$ there is a great probability of a joint extreme event. It can also be noted that the event of no rainfall deficit has a very low probability when associated with high temperatures for all the soil type considered. In any case, despite the fact that the Italian territory has some characteristics in common, the type of soil can help to define more accurately which are the critical events to be taken into

consideration. The value of the crop and the geographic coordinate system on the corresponding pixel determine the policy quotation. The insurance coverage offered to an hypothetic agro Consortium by assuming 120 HA of the soil type $\mathcal{A}$, 9 HA of $\mathcal{B}$, 341 HA of $\mathcal{C}$, the value per hectare of the crop of Euros being 5,000 can be estimated by the Table 2:

Table 2. Valuations for the Premium of the whole Agribusiness Consortium

Premium	1,390,307
Business insured value	2,350,000
Total capital requirement	1,087,890
Fair premium	1,231,954
Pure premium	1,362,501
Expected profit	130,547
CoC	65,273
RORAC	6.00%
Cost per hectare	2,958
Rate premium	59.2%

According to the estimation, the amount of the premium paid to the Consortium will consist in Euros 1,390,307. In case of the drought, the indemnity will change according from 0 to Euros 2,350,000. The Solvency Capital Requirement for the premium risk amounts to Euros 1.087.890 with a corresponding cost of capital equal to Euros 65,273, where the expected profit is Euros 130,547, the RORAC being 6%. On the basis of the defined policy, the reimbursement is defined according to the soil type in Table 3, where we can observe that the soil type determines the percentages of reimbursement. Considering that in case of absence of adverse events there is no reimbursement, while in case of both extreme events the reimbursement is total for all types of soil, for soil $\mathcal{A}$ the reimbursement rate is higher in cases of an extreme high temperature event, associated with both the absence of a rainfall deficit and a non-extreme rainfall deficit, while for soil $\mathcal{B}$ the greater reimbursement relates the occurrence of an extreme event and a non-extreme event at the same time and for soil $\mathcal{C}$ the reimbursement rate is higher in cases of an extreme rainfall deficit, associated with both the absence of a high temperature event and a non-extreme high temperature event.

Table 3. Percentage of the reimbursement of the insured value of the crop based on the joint analysis

Percentage of reimbursement of the insured value of the crop			
Soil type $\mathcal{A}$			
	IP < 175	175 ≤ IP < 200	IP ≥ 200
IC < 35	0%	10%	40%
35 ≤ IC < 50	20%	30%	60%
IC ≥ 50	60%	70%	100%
Soil type $\mathcal{B}$			
	IP < 175	175 ≤ IP < 200	IP ≥ 200
IC < 35	0%	20%	50%
35 ≤ IC < 50	20%	40%	70%
IC ≥ 50	50%	70%	100%
Soil type $\mathcal{C}$			
	IP < 175	175 ≤ IP < 200	IP ≥ 200
IC < 35	0%	10%	90%
35 ≤ IC < 50	5%	15%	95%
IC ≥ 50	10%	20%	100%

5   Concluding Remarks

As the climate-related weather risks become increasingly complex, the requests for innovative products increased. In this context, the parametric weather insurance can be designed to cover both specific catastrophic losses and frequency losses and to complement traditional strategies. In this paper, we propose an innovative scheme of the parametric weather insurance, we called Personal Parametric Weather Insurance (PPWI), where the bioclimatic indicator depends on the satellite data, collected by the weather stations and processed by machine learning technique and combined with specific features of the agribusiness, particularly related to the soil type

References

1. Bosilovich, M.G., Lucchesi, R., Suarez, M.: MERRA-2: file specification. GMAO Office Note No. 9 (Version 1.1) (2016)
2. Carannante, M., D'Amato, V., Fersini, P., Forte, S.: Climate risk management by using index-based analytics: artificial algorithms on satellite data. Ann. Oper. Res. (2021). Under review
3. Dalhaus, T., Musshoff, O., Finger, R.: Phenology information contributes to reduce temporal basis risk in agricultural weather index insurance. Sci. Rep. **8**, 46 (2018)
4. Krehbiel, C.: DAACDataDownload.R [Source code]. https://git.earthdata.nasa.gov/projects/LPDUR/repos/daac_data_download_r/browse/DAACDataDownload.R. Accessed 27 Dec 2021

5. Salgueiro, A.M.: Weather index-based insurance as a meteorological risk management alternative in viticulture. Wine Econ. Policy **8**(2), 114–126 (2019)
6. World Bank.: Weather Index Insurance for Agriculture: Guidance for Development Practitioners. Agriculture and Rural Development Discussion Paper, 50. World Bank, Washington, DC (2011)

An Application of the Tensor-Based Approach to Mortality Modeling

Giovanni Cardillo[1(✉)], Paolo Giordani[1], Susanna Levantesi[1],
and Andrea Nigri[2]

[1] Department of Statistical Sciences, Sapienza University, Rome, Italy
{giovanni.cardillo,paolo.giordani,susanna.levantesi}@uniroma1.it
[2] Department of Social and Political Sciences, Bocconi University, Milan, Italy
andrea.nigri@unibocconi.it

Abstract. With the increasing availability of temporal data, researchers often analyze information stored in matrices, in which entries are replicated on different occasions. Such multidimensional data can be stored in 3-way arrays or tensors to be analyzed. A collection of 3-way arrays can also be available leading to 4-way arrays. In this work, we apply a tensor-based method, the Tucker4, to mortality data provided by the World Health Organization, referred to 4 dimensions (causes of death, age groups, years, and countries) and organized in a 4-way array. We carry out the analysis on the total population. Our findings reveal some peculiar aspects of the mortality phenomenon.

1 Introduction

The core of this paper is the application of multi-way models to causes of death mortality data. The area of mortality by cause includes the impact of specific causes of death on historical mortality trends, the use of "by cause" information in mortality projections, the availability and use of data suitable for underwriting, pricing, and analysis of life assurance and pensions products [8]. The twentieth century witnessed longevity improvements in many high-income countries. These improvements were determined especially by the reduction in a few specific major causes of death groups. In parallel with the growing interest in the topic of mortality, there has also been a development of statistical methodologies to analyse the mortality data, such as multi-way methods, although they were applied to mortality only more recently. In fact, the first attempts of using these models concern three dimensions [4], and the analysis of the cause-of-death mortality data [1,3,9]. A recent review of these methods is in [2]. [9] proposes a three-way extension of the Lee-Carter [5] model by considering death rates aggregated over time, age-groups and country. [1] generalizes the model used in [9] using different tensor decompositions and addresses the forecasting problem of multi-population mortality. Following this line of research, focusing on the cause of death rate, we analyze the causes of death mortality using multi-way models and considering four dimensions. The main advantage of the four-way model is that it allows using information contained in four dimensions simultaneously.

2   Methodology and Application

Four-way component analysis techniques carry out a descriptive analysis of 4-way data that can emerge in different contexts; here we refer to the causes of death mortality. The idea of this method is to efficiently summarize all the information in the four-way data stored in a four-dimensional array through a few components (P,Q, R and S) and the relation between these components (Core tensor $\underline{\mathbf{G}}$). Therefore, four-way component analysis is useful for the exploratory analysis of four-way data. We can consider a multi-way method as a generalization of Principal Component Analysis (PCA), which is a technique to explore the relationship among two-way data that can be applied when the available information can be collected in a matrix, say $\mathbf{X}$. Often the information is replicated in different (time) occasions, in this situation there is an additional mode (occasions), and the data is stored in a 3-way array or (3-way) tensor $\underline{\mathbf{X}}$. We can consider the four-way method as an extension of the three-way method, and one of the best known is the Tucker3 (T3). In our analysis, in which we consider 4-way data, there is an additional dimension (the second occasions) and the data are stored in a 4-way array or (4-way) tensor $\underline{\mathbf{X}}$ in $\Re^{I \times J \times K \times L}$ with generic element x_{ijkl} expressing the score of observation unit i for variable j at first occasion k and at the second occasion l. The Tucker4 (T4) model can be formulated as

$$x_{ijkl} = \sum_{p=1}^{P}\sum_{q=1}^{Q}\sum_{r=1}^{R}\sum_{s=1}^{S} a_{ip}b_{jq}c_{kr}d_{ls}g_{pqrs} + e_{ijkl},$$

with $i = 1, ..., I, \quad j = 1, ..., J, \quad k = 1, ..., K, \quad l = 1, ..., L$ where there are rispectively, the loadings linked to each dimension, the core elements and the generic error term. The T4 in matrix notation is:

$$\mathbf{X}_A = \mathbf{A}\mathbf{G}_A(\mathbf{D}\otimes\mathbf{C}\otimes\mathbf{B})' + \mathbf{E}_A$$

where $\mathbf{X}_A$ is a matrix of order $(I \times JKL)$ and represents the unit mode matricization of the 4-way array $\underline{\mathbf{X}}$. $\mathbf{X}_A$ is obtained by juxtaposing next to each the previously-defined matricizations of the 3-way arrays pertaining to all the second occasions. The symbol $\otimes$ denotes the Kronecker product of matrices. $\mathbf{A}$ $\mathbf{B}$, $\mathbf{C}$ and $\mathbf{D}$ are the component matrices for the 4 modes and their order is $(I \times P)$, $(J \times Q)$, $(K \times R)$ and $(L \times S)$ respectively, where P, Q, R and S denote the number of components for the units, the variables, the first occasions and the second occasions, respectively. Furthermore, $\mathbf{G}_A$ is the unit mode matricization of the core tensor $\underline{\mathbf{G}}$ of order $(P \times Q \times R \times S)$ with generic element g_{pqrs}, which expresses the quadruple interaction among component p of the unit mode, component q of the variable mode, component r of the first occasion mode and component s of the second occasion mode. A high value of g_{pqrs} in absolute sense suggests a strong relation among these components. The T4 consists in minimizing the sum squared errors $||\mathbf{E}_A||^2$ with respect to the component matrices $\mathbf{A}$ $\mathbf{B}$, $\mathbf{C}$ and $\mathbf{D}$ and the core array $\underline{\mathbf{G}}$. To choose the number of components we balance fit

and parsimony looking for a solution easy to interpret, in particular we interpret the components for all modes; next, the core array **G** summarizes the information in the original 4-way array and contains main effects and 2, 3, and 4-way interactions present in the original array [4].

We consider a dataset provided by the World Health Organization (WHO) mortality database and we refer to death rates. We choose the following 8 causes of death: Infectious diseases, Smoking-related cancer, Non-smoking-related cancer, Diabetes, Circulatory system diseases, Respiratory diseases, External causes of death, Other causes of death. Age is organized in classes from 0 to 84 years, from 5 to 5, except for the first two classes, which represent respectively the individuals aged 0 and aged 1–4. Regarding time, we focus the analysis on the years 1961–2015 to consider the same time window for each country. Only the countries for which data is available in this time frame are included in the analysis: *Australia, Austria, Canada, USA, Japan, Belgium, Denmark, Finland, France, Hungary, Ireland, Italy, The Netherlands, Norway, Spain, Sweden, Switzerland* and *UK*. The analysis is developed considering the total population. In order to apply the multi-way method, we organize the data in an array of four dimensions (cause of death × age × time × country). The dimension of the array is: $8 \times 18 \times 55 \times 18$, for a total of 142,560 entries. To choose among the multitude of possible four-way analysis solutions, we performed the T4 using a number of components from two to eight (the minimum between (I,J,K,L)) for each of the modes and computed the associated fit values, then in total, we have from 8 to 32 total components corresponding to the different solutions. We chose the solution $P = 4$, $Q = 4$, $R = 2$, $S = 2$ with the fit value of 91.3%, not only for the fit, but also for the next interesting interpretation. In Table 1, we show the scores of each component that describes the causes of death mortality. The first component mainly depends on External causes and Cardiovascular diseases (CVD), and it is also associated with other diseases and both type of Cancer diseases, so this component represents the Leading causes of death (Leading CoD). The second component is mainly associated with External diseases and with negative signs with Other Lifestyle-related (Other LR), representing that External diseases conduct to a higher incidence of mortality, while Other LR to a lower incidence of mortality. Specifically, we can stress the high positive magnitude associated with External causes, frequently related to young mortality, road accidents, and violence. On the opposite side, we see the negative component for CVD and Smoking-related cancer. The third component is more related to Other diseases (0.92) and less (with a negative sign) to CVD and External disease, so we can conclude that this component mainly portrays Other diseases. The last component is primarily associated with Infectious diseases (0.97).

Table 1. Causes of death components for total population.

Cause of death	Leading CoD	External *vs* Other LR	Other	Infectious
Infectious	0.13	0.05	0.12	**0.97**
Smoking-related cancer	**0.26**	**−0.30**	0.03	−0.17
Non-smoking-related cancer	**0.32**	−0.04	0.04	−0.01
Diabetes	0.03	−0.02	−0.00	−0.02
Cardiovascular diseases	**0.57**	**−0.65**	**−0.23**	0.03
Respiratory	0.06	−0.05	0.02	0.05
External	**0.62**	**0.70**	**−0.30**	−0.09
Other	**0.31**	0.08	**0.92**	−0.14

We can also interpret the other components for each dimension. The first age component, Adults, reflects the adult mortality (it is mainly positively associated with ages from 30 to 50), showing the typical regularity well described by a linear Gompertz law of mortality. We can relate the second component, Old *vs* Young (O *vs* Y), to the different behaviors of mortality between the old and the young, in particular it captures the excess of mortality from 60 to 80, which is opposed to a reduction of mortality from 5 to 25. The third component depicts infant mortality (Infants), which is mainly associated with 0 and 1 ages. The last one, Early adults *vs* Children (EA *vs* Child), reflects the different behaviors of mortality between early adults and children, in fact it is positively associated with age 30–35, and negatively with age 1, 5 and 10. The first year component represents the overall improvement of mortality (with a stronger association in the first few years) the second component underlines two different aspects related to "converging" and "improvements" periods, respectively. Therefore, this second component is associated with the "late" period with positive signs versus the "early" period with negative signs. Indeed, the recent historical worldwide longevity dynamics evidence the first increase after the 50s, albeit with high heterogeneity levels, that flattened in the decades around 75 s–85 s, that might be described as a global convergence [6,7]. The first-year component describes all the countries considered from the database (HMD). The second component discriminates between a group of European countries (Spain, Italy, Ireland, and Hungary) as well as Northern European countries that recently show relevant improvements in life expectancy, leading the global records as the case of Sweden [7]. Italy and Spain are usually considered similar in culture, values, and also in economic patterns. Ireland is often compared to Spain and Italy for its pattern of mortality, for example in terms of life expectancy [10]. In this group, there is also Hungary that can represent the Eastern European countries. Finally, Table 2 represents the core tensor **G**, where the higher in absolute value an element of the core, the stronger the interaction among the components involved.

Table 2. Interaction for total population.

Period: Overall improvement								
CoD components	Countries: HMD				Countries: EU *vs* North EU			
	Adults	O *vs* Y	Infants	EA *vs* Child	Adults	O *vs* Y	Infants	EA *vs* Child
Leading CoD	**13.86**	−0.29	−0.27	−0.11	0.15	−0.17	−0.11	0.02
External *vs* Other LR	−0.45	**−8.64**	−0.15	0.37	**−0.67**	0.25	−0.02	−0.13
Other	−0.02	−0.04	**3.02**	0.51	0.42	−0.26	0.03	−0.16
Infectious	−0.03	−0.00	**0.90**	−0.27	0.49	−0.29	0.42	0.03
Period: Late *vs* Early								
CoD components	Countries: HMD				Countries: EU *vs* North EU			
	Adults	O *vs* Y	Infants	EA *vs* Child	Adults	O *vs* Y	Infants	EA *vs* Child
Leading CoD	0.24	**1.22**	−0.52	**1.76**	0.25	0.00	−0.17	−0.13
External *vs* Other LR	0.16	**0.71**	**−0.73**	**1.48**	0.11	−0.44	0.00	−0.12
Other	**1.94**	0.24	**1.33**	0.53	−0.18	0.11	−0.40	−0.01
Infectious	**−0.84**	0.17	**−1.16**	−0.04	−0.33	0.21	−0.50	0.10

The highest score (13.86) refers to the interaction term between leading CoD that occurred at adult ages in a phase of overall longevity improvement for all the countries considered by HMD. It means that during this period, for the adults, the leading CoD lead to a major incidence of mortality in all the countries considered. In the same period, for the same group of countries the coefficient (–8.64) linked to External *vs* Other LR has a negative sign referred to O *vs* Y mortality; this means that during the phase of overall longevity improvement, in all the countries considered, there is a higher incidence of deaths caused by external causes for younger people compared to older people, and there is a higher incidence of CVD and Smoking-related cancer for older people compared to younger people. As for infectious diseases, they affect both periods in childhood. In the same period and in the same group of countries, infants die predominantly for other causes of death (3.02), in fact in this group there are specific causes such as perinatal infant mortality. The Infant component is also related to Infectious diseases (0.90). For the same group of countries, in the last period, there is a prevalence of deaths produced by the leading causes for older people compared to younger people, while in the early period the opposite situation occurs: there is a prevalence of deaths generated by the leading causes for younger people with respect to older people (1.22). Similarly, deaths are prevalent for early adults (compared to children) caused by leading CoD in the late period, while in the early period the opposite situation occurs (1.94). Moreover, in the late period there is a prevalence of deaths caused by external causes (compared to CVD and Smoking-related cancer) for early adults with respect to children, while in the early period there is a prevalence of deaths due external causes (compared to other LR diseases) for children with respect to early adults (1.48). In the late period, for adults there is a prevalence of deaths due to other causes, while in the early period it appears to be less evident (1.94); concerning infectious diseases, the opposite situation occurs (–0.84). Regarding infants, in the early period, there is a prevalence of deaths due to infectious diseases, much limited in the late period (–1.16); for other causes, the opposite situation occurs (1.33). Considering the second component of countries, Spain, Italy,

Ireland, and Hungary *vs* Northern EU, in the period of overall improvement, the coefficients linked to other and infectious diseases have positive signs (0.42, 0.49) referred to adult mortality; this means that there is a higher incidence of these causes for adults in the first group of countries compared to the Northern EU countries. Furthermore, in the same period and for adults, there is a prevalence of deaths caused by other LR diseases compared to external causes (–0.67).

3   Conclusions

In this paper, we have applied the T4 method to the mortality by cause of death considering four dimensions. We have shown that the model allows extracting meaningful demographic explanation by reporting a walkthrough of our results: firstly, we have summarized the interpretation of the components for all the dimensions and in the last part that of the core array. We have chosen the solution $P = 4, Q = 4, R = 2, S = 2$ with a fit value of 91.3%, which enables us to provide an interesting interpretation of the phenomenon under investigation. For example, the highest score refers to the interaction term between Leading CoD that occurred at adult ages during the phase of overall longevity improvement for all the countries considered by HMD. In this sense, the four-way component analysis is useful for 4-way data, and, in our context, the analysis has revealed some peculiar aspects of mortality by cause of death.

References

1. Dong, Y., Huang, F., Yu, H., Haberman, S.: Multi-population mortality forecasting using tensor decomposition. Scand. Actuar. J. **8**, 754–775 (2020). https://doi.org/10.1080/03461238.2020.1740314
2. Giordani, P., Kiers, H.A.L.: A review of tensor based methods and their application to hospital care data. Stat. Med. **37**, 137–156 (2018). https://doi.org/10.1002/sim.7514
3. Giordano, G., Haberman, S., Russolillo, M.: Coherent modeling of mortality patterns for age-specific subgroups. Decis. Econ. Finan. **42**, 189–204 (2019). https://doi.org/10.1007/s10203-019-00245-y
4. Kiers, H.A.L., van Mechelen, I.: Three way component analysis: principles and illustrative application. Psychol. Meth. **6**(1), 84–110 (2001). https://doi.org/10.1037//1082-989X.6.1.84
5. Lee, R.D., Carter, L.R.: Modeling and forecasting us mortality. J. Am. Stat. Assoc. **87**(419), 659–671 (1992)
6. Nigri, A., Barbi, E., Levantesi, S.: The relationship between longevity and lifespan variation. Stat. Methods Appl. (2021). https://doi.org/10.1007/s10260-021-00584-4
7. Oeppen, J., Vaupel, J.W.: Broken limits to life expectancy. Science **296**(5570), 1029–1031 (2002). https://doi.org/10.1126/science.1069675
8. Ridsdale, B., Gallop, A.: Mortality by cause of death and by socio-economic and demographic stratification, ICA2010 (2010)
9. Russolillo, M., Giordano, G., Haberman, S.: Extending the Lee-Carter model: a three-way decomposition. Scand. Actuar. J. **2011**(2), 96–117 (2011)
10. Roser, M., Ortiz-Ospina, E., Ritchie, H.: Life expectancy. Our World in Data (2013)

Cyber Risk: Estimates for Malicious and Negligent Breaches Distributions

Maria Francesca Carfora and Albina Orlando[✉]

Istituto per le Applicazioni del calcolo "Mauro Picone" Consiglio Nazionale delle Ricerche, Naples, Italy
{f.carfora,a.orlando}@na.iac.cnr.it

Abstract. Cyber risk is a fast-growing area of interest and companies have to include it in their risk management framework. Modelling frequency and severity of cyber incidents is a crucial step in actuarial valuations related to cyber insurance, a way of transferring part of the residual cyber risk to a third party. In the last years, data breaches seem to be the main cause of cyber incidents. Aim of this paper is to give further insights about frequency and severity statistical distributions, by analyzing the Chronology of Data Breaches provided by the Privacy Rights Clearinghouse.

1 Introduction

In an increasingly interconnected world, the risks arising from the cyber domain are growing more and more. Nowadays, the financial losses caused by the impact of these risks have become considerable. In general, cybercrime has an impact on the global economy of about 1 trillion dollars, more than 1 percent of global GDP. The covid pandemic has certainly worsened this situation and, to date, most companies consider cyber risks, the pandemic outbreak, business interruption and strongly linked to each other [1].

Cyber risks belong to the category of operational risks that stem from external occurrences or bad and unfruitful internal systems, people, and processes. However, they show specific characteristics that make them particularly hard to manage and mitigate. Indeed, it is well known that the environment of cyber risks constantly evolves as a consequence of new technologies and the rapid development of computer information systems. In [8] the authors state that many researchers have mentioned that cyber risks are like natural catastrophes, at least in terms of scale. In the light of these considerations, companies are increasingly aware that the best risk management practices need to consider cyber risk as a component of the overall risk management routine. After analyzing and quantifying cyber risks, each company will decide how much money to invest in order to improve security levels. Even if security investments were

significant, it would not be possible to eliminate cyber risk completely. However, there is the option of transferring part of it to a third party by signing a cyber insurance contract. Cyber insurance aims to mitigate losses from cyber incidents which include data breaches, data theft, business interruption and network damage. Insuring cyber risks is a challenging issue due to several reasons, among the others: continuous evolution of information systems, increasingly sophisticated and challenging cyber attacks, interdependence of security levels, hard impact determination, information asymmetry, lack of statistical data [7]. Regarding the lack of statistical data, it mainly arises from information sharing barriers caused by the unwillingness of companies to reveal cyber incidents. The reason lies in the concern regarding reputation damages. Reporting requirements have been introduced since 2002 in many US states. For this reason, the literature on this topic is focused on data breaches and US data. The situation is rapidly changing in Europe too with the introduction, in May 2018, of GDPR (General Data Protection Regulation). Data are essential to actuarial valuations: the premium calculation in practice requires knowledge of the full loss distribution in order to determine the right amount of safety capital. As a consequence, modelling frequency and severity of cyber incidents is a crucial step.

Aim of this paper is to give some insights regarding the statistical distributions of severity and frequency of data breaches. In particular, we refer to the Chronology of Data Breaches provided by the Privacy Rights Clearinghouse. Several contributions of the recent literature on cyber risk management analyze this data set [6,8,12]. We contribute to the existing literature offering a further analysis of this dataset. Our analysis confirms that data breaches of different types, as long as events occurred in different entities, often show a different statistical nature. The rest of the paper is structured as follows. Section 2 focuses on the main features of data breaches and relative studies in the recent literature. Section 3 discusses a case study and Sect. 4 concludes.

2   Cyber Incidents and Data Breaches

[10] gives an overview both of the different types of cyber incidents and the kind of losses that may occur. With regard to the incidents, four broad categories are included: data confidentiality breach, system malfunction/issue, data integrity/availability and malicious activity. According to [1], data breaches are the main cause of cyber incidents. Indeed, companies collect and use ever greater volumes of personal data and breaches are becoming larger and more expensive. Dealing with a mega breach (involving more than one million records) now costs \$42 mn on average. Moreover,data protection and privacy regulation, and subsequent penalties, are widening in scope and geographical reach. Many employees (intentionally or not) are often the weakest link that causes a successful cyber incident (e.g. accidental publication of confidential information, non-custody of laptop computers containing highly sensitive information) [10]. A complete review of the main data breach studies in the existing literature is given in [9]. As far as frequency and severity modelling is concerned, frequency

is generally well fitted by either a Poisson or a negative binomial distribution. Regarding severity, log-normal and skew-normal models are frequently being used in the actuarial literature. Models from extreme value theory (ETV) are popular too, because the data from operational risk show heavy tails. The peaks-over-threshold method (POT) is the most common EVT approach. Applications of these models in the cyber risk domain are proposed by [6]. It is pointed out that different kinds of cyber incidents often show a different statistical nature, requiring separate modeling [3]. Another issue concerns the dependability structure of losses, requiring a model able to deal with various, but dependent, classes of damages [2]. In order to model dependency structure, copulas are commonly used [2,5].

In the following, we analyze the data breach information obtained from the Chronology of Data Breaches compiled by the Privacy Rights Clearinghouse [11], a nonprofit organization. It is by far the most complete and reliable publicly available dataset, as already stated by other researchers [4,6,12,13].

3   Case Study

The Chronology of Data Breaches compiled by the Privacy Rights Clearinghouse contains information on data breaches occurred in the US between January 10, 2005 and December 31, 2019, always including name, location and type of the breached entity, description and type of breach, and often the number of breached records; all the reported events have been confirmed by major media sources. Organizations are categorized as businesses-financial and insurance services (BSF); businesses-retail/merchant including online retail (BSR); businesses-other (BSO); educational institutions (EDU); government and military (GOV); healthcare, medical providers and medical insurance services (MED); and nonprofit organizations (NGO). Information exposures (breaches) are reported as Debit and Credit Cards Frauds (CARD), data loss due to hacking/malware (HACK), or to insiders (INSD), physical data loss (PHYS), portable device data loss (PORT), stationary computer data loss (STAT), unintended disclosure of data (DISC), unknown cause (UNKN). The main limitation to these data is its underestimate of the phenomenon, because publicly acknowledged breaches are just a part of all the security incidents; moreover, the data do not include information on financial losses. Following [12], we decided to focus our analysis on the recent data (breaches reported after the 1st of January, 2010) assuming that they reflect the more recent cyber threat situation. Moreover, we also disregard the incomplete records with unknown/unreported/missing hacking breach sizes or unknown cause. The resulting dataset contains 4823 breach incidents in the United States between January 1st, 2010 and October, 25th, 2019. As the results of previous studies [3,4] have confirmed, the best fit for the frequency of breach incidents is provided by a negative binomial distribution, with good results on the full dataset and a quite excellent performance on any considered subset. The distribution of data breach sizes, on the other side, is just approximately described by a log-normal or a skew-normal distribution, even if the fit improves while considering smaller subsets. In general,

data breaches of different types, as long as events occurred in different entities, often show a different statistical nature so that on the entire dataset none of the considered models could give completely satisfactory results, due to the strong heterogeneity of the data. A closer inspection of data subsets have shown that medical entities (MED) suffered the largest number of Data Breaches (3591 obs., about 3/4 of the total), but most of these breaches are due to lack of vigilance (negligent breaches). Moreover, this high number of incidents corresponds to a very low number of breached records (about 2% of the total) and a severity analysis on this subset showed a very particular distribution, apparently related to the presence of different entities (big institutions and small medical practices). The empirical distributions of the data breach sizes for the MED and non-MED subgroups shown in Fig. 1 visually confirm our claim. Such anomalies led us in a previous work [3], to analyze separately the MED/negl subset, comprised of the breaches in MED entities related to negligence causes. In this study, we then exclude the MED subset and focus on breaches typologies' differences for the other organization categories, by separating the negligent breaches (i.e., DISC, PHYS, PORT, STAT) from the malicious ones (i.e., CARD, HACK, INSD).
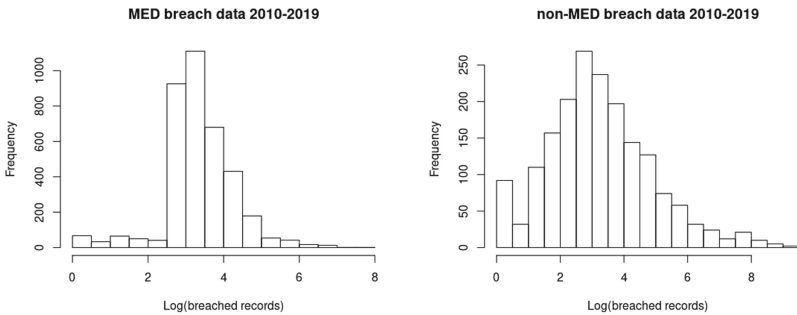
Fig. 1. Histograms of the base-10 logarithm of the data breach sizes for the MED and non-MED subgroups from the PRC data in the time range 2010–2019

We applied the data fitting analysis for the breach frequency separately to each of the two groups and find that the best fit for each of the categories (negligent, 512 obs. and malicious, 720 obs.) is given by a negative binomial distribution for daily frequency, completely confirmed by Kolmogorov-Smirnov test results (p-value = 1 for both cases).

The best fit for the severity of both malicious and negligent breaches, among all the distributions proposed in the literature, is given by the skew-normal distribution. Even not as accurate as the frequency ones, these fits are still acceptable: results of the Kolmogorov-Smirnov test give p-values of about 0.21 and 0.19, respectively for the two categories. Further analysis of the severity distributions, however, reported in Fig. 2, shows that just very few large breaches are not well fitted by the skew-normal distribution; there is not enough evidence to justify a different modeling (based on EVT) for the larger breach sizes.
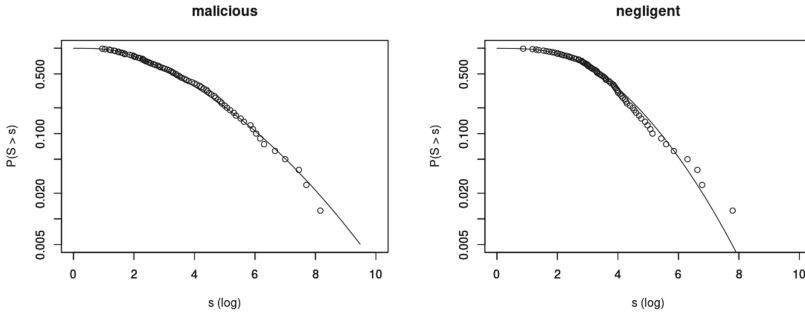
Fig. 2. The distribution of breach sizes and the fit to a skew-normal distribution: the left and right panel show the malicious and negligent subgroups, respectively.

Finally, we are able to estimate the Value at Risk and assess the goodness of fit of the model on the two data subsets. In Table 1 four confidence levels are reported (90%, 95%, 99% and 99.5%) along with the p-value for testing the difference between estimated and empirical values. These p-values confirm the evidence from Fig. 2: while the fit for malicious breaches severity gives good results up to the 95th percentile and degrades after, the one for negligent breaches poorly approximates the empirical data. Indeed, this is already evident from the 90th percentile and the worse fit concentrates in the first part of the tail, where more data are available.

Table 1. Risk measurement results: estimated (s) and empirical (m) values of the Log of breached records for the two considered categories, malicious (mal) and negligent (negl), along with the p-value of their agreement.

Type	90 s	90 m	p-value	95 s	95 m	p-value	99 s	99 m	p-value	99.5 s	99.5 m	p-value
mal	6.10	6.04	0.71	7.02	7.00	0.94	8.81	8.17	0.02	9.49	8.53	0.02
negl	5.41	5.14	0.02	6.05	6.30	0.19	7.30	7.95	0.16	7.78	8.18	0.45

4   Concluding Remarks

This study aimed at providing more detailed information both on the frequency and severity modelling of cyber incidents. Focusing on data breaches, the category of cyber incidents that causes the biggest losses nowadays, we analyzed a dataset obtained from the Chronology of Data Breaches compiled by the Privacy Rights Clearinghouse [11], a nonprofit organization. It is widely recognized as the most complete and reliable publicly available dataset. Unfortunately, the data do not include information on financial losses. Following [12], we decided to analyze the recent data (breaches reported after the 1st of January, 2010) assuming that they reflect the more recent cyber threat situation. In this study, we estimated frequency and severity for all the organization types but the Medical one, whose

empirical distribution shows several anomalies. After observing the different statistical nature (in terms of distribution parameters) of the negligent breaches and the malicious ones we proceeded by separating breaches data into these two categories. We verified that the best fit for daily frequency of both categories is given by a negative binomial distribution and this is completely confirmed by Kolmogorov-Smirnov (KS) test. As regard severity, we observe that the best fit for both malicious and negligent breaches, among all the distributions proposed in the literature, is given by the skew-normal distribution. The fits are not as accurate as frequency ones, but the KS test gives acceptable results. Indeed, just very few large breaches are not well fitted by the skew-normal distribution, where negligent breaches are slightly underestimated and malicious ones overestimated. All the conclusions are confirmed by the VaR estimates in Table 1.

Other datasets should be analysed in the future. In particular, it would be desirable to analyse data relating to European companies and the availability of such data should be facilitated by the entry into force of the GDPR.

References

1. Allianz Global Corporate and Specialty. Allianz Risk Barometer: Top Business Risks for 2021. Report (2021). https://www.agcs.allianz.com/news-and-insights/reports/allianz-risk-barometer.html. Accessed 10 Feb 2021
2. Bentley, M., Stephenson, A., Toscas, P., Zhu, Z.: A multivariate model to quantify and mitigate cybersecuity risk. Risks **8**, 61 (2020)
3. Carfora, M.F., Martinelli, F., Mercaldo, F., Orlando, A.: Cyber risk management: an actuarial point of view. J. Oper. Risk (2019). https://doi.org/10.21314/JOP.2019.231
4. Edwards, B., Hofmeyr, S., Forrest, S.: Hype and heavy tails: a closer look at data breaches. J. Cybersecur. (2016). https://doi.org/10.1093/cybsec/tyw003
5. Eling, M., Jung, K.: Copula approaches for modeling cross sectional dependence of data breach losses. Insur. Math. Econ. **82**, 167–180 (2018)
6. Eling, M., Loperfido, N.: Data breaches: goodness of fit, pricing, and risk measurement. Insur. Math. Econ. **75** , 126–136 (2017)
7. Marotta, A., Martinelli, F., Nanni, S., Orlando, A., Yautsiukhin, A.: Cyber-insurance survey. Comput. Sci. Rev. **24**, 35–61 (2017)
8. Weathley, S., Hofmann, H., Sornette, D.: Data breaches in the catastrophe framework and beyond. arXiv:1901.00699v2 (2019). arxiv.org/abs/1901.00699v2
9. Woods W.D., Böhme, R.: SoK: quantifying cyber risk. In: 2021 IEEE Symposium on Security and Privacy (SP) (2021)
10. OECD. Types of cyber incidents and losses. In: Enhancing the Role of Insurance in Cyber Risk Management. OECD Publishing, Paris (2017). https://doi.org/10.1787/9789264282148
11. Privacy Rights Clearinghouse. Chronology of data breaches (2022). https://privacyrights.org/data-breaches
12. Sun, H., Xu, M., Zhao, P.: Modeling Malicious Hacking Data Breach Risks. North Am. Actuar. J. (2021). https://doi.org/10.1080/10920277.2020.1752255
13. Xu, M., Schweitzer, K.M., Bateman, R.M., Xu, S.: Modeling and predicting cyber hacking breaches. IEEE Trans. Inf. Forensics Secur. (2018). https://doi.org/10.1109/TIFS.2018.2834227

Modeling and Forecasting Natural Gas Futures Prices Dynamics: An Integrated Approach

Oleksandr Castello[(✉)] and Marina Resta

School of Social Sciences, Department of Economics and Business Studies,
University of Genova, 16126 Genova, GE, Italy
`alexander-castello@libero.it, resta@economia.unige.it`

Abstract. We explore and test the capabilities of B-Splines and Dynamic De Rezende-Ferreira five–factor model to replicate the main dynamics and stylized facts of futures curves in the Natural Gas Futures market. Furthermore, we discuss the joint use of these models with a Nonlinear Autoregressive Neural Network for parameters fine–tuning to forecast futures curves. The simulation study highlighted the effectiveness of the proposed framework; empirical results show that the joint use of B–Splines and neural networks provides highest overall performances on the Natural Gas futures market.

Keywords: Natural Gas · Futures term structure · De Rezende–Ferreira model · B–Spline · Artificial Neural Networks (ANN) · Nonlinear Autoregressive Neural Networks (NAR–NN)

1 Introduction

Fostered by the liberalization process of commodities markets [7,8], during the last decade Energy Commodities grew in importance as an alternative financial asset. Understanding the dynamics of futures prices term structure is therefore of paramount importance for hedging, capital budgeting as well as to provide a deep insight on the investors' expectations about the behaviour of the underlying commodities. Based on these premises, our research focused on the analysis of the term structure of the Natural Gas (NG) futures market, the second largest and fastest rising source for energy demand, with the aim to identify a proper modeling approach to effectively describe stylized facts in the futures curve and to provide effective in–sample fitting and out–of–sample forecasting.

In the existing literature on the Natural Gas market there is a quite limited number of studies dealing with models of the term structure [2,13,16,17]; in fact most of the research mainly debates on the existence of possible relations with other commodities or securities [3,4,20], as well as on models for spot prices, price volatility, demand and supply [6,14,18].

The motivation for our work arised in observing that both B–Splines [19] and the Five–Factor De Rezende–Ferreira (5F–DRF) model [11] have been widely

and successfully applied in the fixed–income market [10,15] which is quite similar to that of Natural Gas, at least for what it is concerning the data structure, the maturity and the overall dynamics. We therefore introduced a framework using B–Spline and the 5F–DRF models for in–sample fitting and a Nonlinear Autoregressive Neural Network (NAR–NN) for parameters tuning in out–of–sample forecasting of NG futures curves.

Main contributions of our research are summarized as follows: we analyzed futures curves in the NG market within a new framework, providing empirical evidence of its ability to replicate patterns and dynamics of the data with high levels of in-sample matching accuracy; furthermore, we benefit of the NAR–NNs flexibility [1,12] to forecast futures curves. The analysis was carried out in the time span from August 2010 to April 2021 thus including periods of high turbulence such as the Euro Crisis of 2010, the Chinese stock market turmoil in 2015–2016 as well as the oil and pandemic crush of 2020 with the scope to validate the adequacy of the framework under very critical conditions.

The remainder of the paper is organized as follows: in Sect. 2 we present the dataset and the methodology in use; in Sect. 3 we provide and discuss the main results; Sect. 4 concludes.

2   Data and Methods

We used a dataset of daily settlement prices quoted in €/MWh of monthly futures contracts with maturities from 1 up to 12 months (Mc1 to Mc12), obtained from the Dutch Title Transfer Facility (TTF), a virtual trading hub which in recent years has become the leading gas trading platform in Europe. The dataset organized by data and maturity is plotted in Fig. 1 (see the Appendix).

The 3D–plot highlights a number of stylized facts such as upward sloping, inverted and flat curves, as well as persistence and volatility, which support the existence of strong similarities between the examined data and those in the fixed–income markets. This in turn, inspired our research study which applies on the NG futures market models of established use in the fixed–income market.

Our framework works in two phases: the first one performs in–sample fitting of the futures curve on a daily basis, while the second stage provides one–day–ahead forecasting. The modeling stage is carried out by means of B–Splines [9] and by means of the 5F–DRF factor model [11].

Parameters estimation follows the approach discussed in [15] for B–Splines and the method described in [5] for the 5F–DRF. The generated futures curve are compared in terms of Mean Square Error (MSE) and Root Mean Square Error (RMSE) statistics.

The estimated parameters enter then into a fine–tuning process driven by a NAR–NN to provide one–day ahead prices forecasts. The accuracy of the forecasting process is evaluated by the Mean Squared Forecast Error (MSFE) and Mean Square Percentage Error (MSPE) performance metrics.

3 Empirical Results

Table 1 compares the MSE and RMSE main statistics for both the B–Spline and the 5F–DRF. The B–Spline generated more accurate fits in every examined case; this, in turn, highlighted some limitations of the factor model in approximating medium and long term maturities of more twisted curves.

Table 1. Main MSE and RMSE statistics for the 5F-DRF and B–Spline models

Performance	MSE		RMSE	
Model	5F–DRF	B-Spline	5F–DRF	B-Spline
Mean	5.1600×10^{-2}	1.9187×10^{-2}	2.0118×10^{-1}	1.1958×10^{-1}
SD	5.4560×10^{-2}	2.2616×10^{-2}	1.0550×10^{-1}	6.9928×10^{-2}
Min.	3.3350×10^{-4}	1.9149×10^{-4}	1.8262×10^{-2}	1.3884×10^{-2}
Max.	3.3556×10^{-1}	2.1747×10^{-1}	5.7927×10^{-1}	4.6635×10^{-1}

To asses the overall adequacy of the models we have also performed one–day–ahead forecasts considering the first 20 working days of March 2021. We evaluated the MSPE and MSFE of both the tuned methods and we compared them to the values obtained by the NAR–NN used alone, directly on the futures price time series. The results in Table 2 suggest that not only the joint use of NAR–NNs and B–Splines works better, but also it outperforms the NAR–NN used alone on the data.

Table 2. Comparison of the forecasting performances

Model	5F–DRF	B–Spline	NAR–NN
MSPE	1.0688	0.0486	0.0883
MSFE	3.7306	0.1615	0.2963

4 Conclusion

We discussed a framework that uses B–Splines and the 5F–DRF factor models on the Natural Gas Futures market tuning their parameters with a NAR–NN for day–ahead forecasts. Both models showed high degree of flexibility and adaptability to the wide variety of situations characterizing the Natural Gas market. Furthermore, we highlighted that the joint use of B–Splines and neural networks seems being particularly effective in day–ahead forecasting, as it outperforms both 5F–DRF coupled to NAR–NN and the NAR–NN itself directly employed on the data.

A   Appendix: Figures
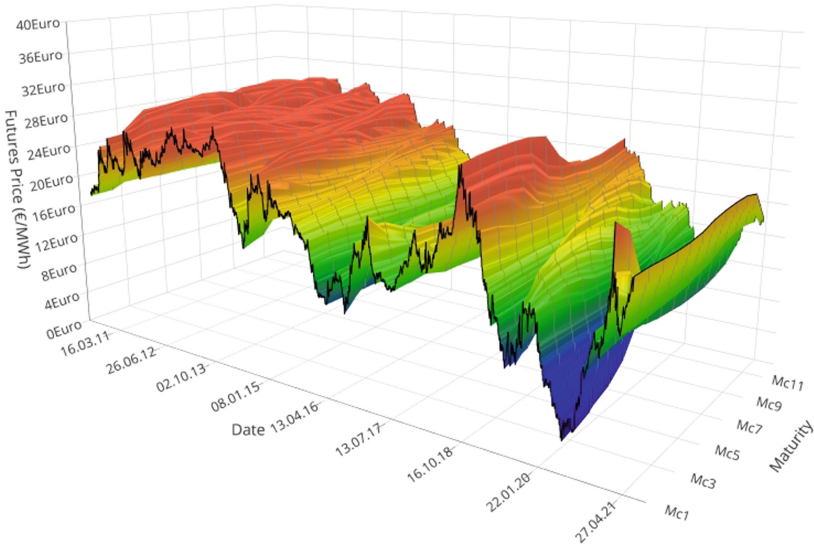
Fig. 1. Term Structure of Natural Gas Futures Prices. Prices time-series of the natural gas futures contracts with expiration date from 1 (Mc1) to 12 (Mc12) months. The data spans 2732 trading days from August 19, 2010 to April 27, 2021.

References

1. Abraham, E., dos Reis, J., Vendrametto, O., Costa Neto, P., Toloi, R., de Souza, A., Morais, M.: Time series prediction with artificial neural networks: an analysis using Brazilian soybean production. Agriculture **10**(10) (2020)
2. Alamsyah, A., Permana, M.: Artificial neural network for predicting indonesian economic growth using macroeconomics indicators. In: 2018 International Symposium on Advanced Intelligent Informatics (SAIN), pp. 15–19 (2018)
3. Behmiri, N., Manera, M., Nicolini, M.: Understanding dynamic conditional correlations between oil, natural gas and non-energy commodity futures markets. Energy J. **40**(2), 55–76 (2019)
4. Brown, S., Yücel, M.: What drives natural gas prices? Energy J. **29**(2), 45–60 (2008)
5. Castello, O., Resta, M.: Modeling the yield curve of BRICS countries: parametric vs machine learning techniques. Risks **10**(2) (2022)
6. Chen, Y., Chua, W., Koch, T.: Forecasting day-ahead high-resolution natural-gas demand and supply in Germany. Appl. Energy **228**, 1091–1110 (2018)
7. Cheng, I., Xiong, W.: The Financialization of Commodity Markets. Working Paper 19642, National Bureau of Economic Research, November 2014
8. Creti, A., Nguyen, D.: Energy markets financialization, risk spillovers, and pricing models. Energy Policy **82**, 260–263 (2015)

9. de Boor, C.: A Practical Guide to Splines, 1st edn., vol. XVIII, Springer, New York (1978)
10. De Rezende, R.: Giving flexibility to the Nelson-Siegel class of term structure models. Revista Brasileira de Financas **9**(1), 27–49 (2011)
11. De Rezende, R., Ferreira, M.: Modeling and forecasting the Brazilian term structure of interest rates by an extended Nelson-Siegel class of models: a quantile autoregression approach. resreport, Escola Brasileira de Economia e Finanças, May 2008
12. Jha, G., Sinha, K.: Time-delay neural networks for time series prediction: an application to the monthly wholesale price of oilseeds in India. Neural Comput. Appl. **24**(3), 563–571 (2014)
13. Li, B.: Pricing dynamics of natural gas futures. Energy Econ. **78**, 91–108 (2019)
14. Li, J., Wu, Q., Tian, Y., Fan, L.: Monthly Henry Hub natural gas spot prices forecasting using variational mode decomposition and deep belief network. Energy **227**, 120478 (2021)
15. Mineo, E., Alencar, A., Moura, M., Fabris, A.: Forecasting the term structure of interest rates with dynamic constrained smoothing B-splines. J. Risk Finan. Manag. **13**(4), 119–132 (2020)
16. Nuryyev, G., Hickson, C.: Term structure of natural gas futures contracts. Energy Stud. Rev. **23**(1) (2016)
17. Root, T., Lien, D.: Can modeling the natural gas futures market as a threshold cointegrated system improve hedging and forecasting performance? Int. Rev. Finan.l Anal. **12**(2), 117–133 (2003)
18. Saltik, O., Degirmen, S., Ural, M.: Volatility modelling in crude oil and natural gas prices. Procedia Econ. Finance **38**, 476–491 (2016)
19. Schoenberg, I.: Contribution to the problem of approximation of equidistant data by analytic functions: Part A - on the problem of smoothing or graduation. A first class of analytic approximation formulae. Q. Appl. Math. **4**(1), 45–99 (1946)
20. Zhang, Y., Chevallier, J., Guesmi, K.: "De-financialization" of commodities? Evidence from stock, crude oil and natural gas markets. Energy Econ. **68**, 228–239 (2017)

Modelling Life Expectancy Gender Gap in a Multi-population Framework

Leonardo Cefalo[1], Susanna Levantesi[1(✉)], and Andrea Nigri[2]

[1] Sapienza University of Rome, Rome, Italy
susanna.levantesi@uniroma1.it
[2] Bocconi University, Milan, Italy
andrea.nigri@unibocconi.it

Abstract. This paper aims at investigating whether the life expectancy gender gap follows any long-run common tendency across different countries through a model-based analysis. If these tendencies are found to exist, then a model which takes them into account should perform better than a basic and unrestricted one. Once the gap is modeled as a multivariate non-stationary stochastic process, the goal is to find any long-run equilibrium among single series via cointegration analysis, which ultimately allows estimating some stationary linear combinations of non-stationary variables referred to as the error correction terms. To achieve such a result it is preferable to work with homogeneous samples. Therefore, the first step of this analysis consists in partitioning the initial data set into five clusters. Since the input data set includes countries with different gender gap dynamics, this diversity is clearly reflected by the difference among the models employed to fit single clusters. All series result to be non-stationary. Given the model, we check the stationarity of the error correction term and apply simple backtesting to ten-years forecasting. Evidence suggests that the fifth cluster is a cointegrated series leading to postulate that an underlying long period equilibrium does exist for this cluster.

1 Introduction

Both insurance companies and pension schemes are deeply concerned with managing longevity risk. The modeling and forecasting of life expectancy play a central role (see e.g. the work of [2], which consider an indexing mechanism based on the expected residual life for adjusting the retirement age and discuss the implications of the gender gap in life expectancy). The reason is straightforward: a proper intertemporal resource allocation requires a full understanding of mortality dynamics, hence the need for a framework of stochastic models for mortality and longevity - a deterministic approach would not provide information such as risk measures. If mortality tends to converge across different countries, then forecasting should display mutual coherence. Even more so, if policy makers involve several countries in their decisions, demographic forecasting must

regard multiple populations and coherence is obviously needful. Alongside with cross-country longevity convergence, there is some evidence that also female and male life expectancies are heading towards a long-term equalisation, at least in developed countries. Modeling and forecasting the gender gap in life expectancy - henceforth GGLE - might be interesting for actuaries for the following reasons: i) male life expectancy forecasting is more reliable if they are modeled as the difference between female life expectancy and GGLE rather than directly (see, e.g. [1,8]); ii) equal life insurance premiums, imposed by the Gender Directive (Council Directive 2004/113/EC) for equal treatment between men and women, imply different risk exposures and different loss distributions; iii) the initial benefit in defined contribution pension schemes is calculated, at retirement, by converting the individual notional account into an annuity depending on the remaining life expectancy. To establish intertemporal equilibria, the knowledge of longevity gender differentials is crucial. An additional issue arises when it comes to retirement age (all increasings, especially differential ones, should be at least consistent with longevity forecasting - other than socially acceptable). However, GGLE is not a frequently discussed topic in actuarial literature. Since univariate model-based studies are not very frequent in the GGLE modeling, a multivariate time series approach may be an actual novelty (a similar approach applied to life expectancy is found in [7] and [6]). Our study collects data from a quite wide set of countries. We consider countries whose data are available from 1970 to 2018 and whose population size is larger than one million inhabitants. The data source is the Human Mortality Database [3].

To assess whether some series display a long-run equilibrium, the idea is to check for cointegration. Since present series are $I(1)$, the aim is to estimate stationary linear combinations, referred to as *error correction terms* via the *vector error correction model* (VECM). We use the maximum likelihood estimation developed by [4] to check for cointegrating relationships in the data sets with more than two variables. Since in a cointegration analysis it is preferable working with homogeneous samples, we partition the data set into five clusters using the K-means algorithm.

2   Materials and Methods

GGLE data are obtained as differences between female and male life expectancies. The data set collected in the Human Mortality Database consists of GGLE data of 25 countries in the period 1970-2018. Observations have been partitioned into a training set and a testing set; the first one includes data from 1970 to 2008, the latter from 2009 to 2018.

The K-means algorithm converges at the second iteration providing the following clusters:

- **Cluster 1**: Australia, Austria, Belgium, Canada, Switzerland, Italy, USA.
- **Cluster 2**: Belarus, Estonia, Lithuania, Latvia.
- **Cluster 3**: Czechia, Spain, Japan, Portugal.

- **Cluster 4**: Finland, France, Hungary, Poland, Slovakia.
- **Cluster 5**: Denmark, Great Britain, Netherlands, Norway, Sweden.

We focus on cluster 5, which is the most appealing for understanding the gender differences in longevity.

Cointegration Analysis

A multivariate non-stationary stochastic process is said to be cointegrated if there are some linear combinations of its single variables, which have a smaller integration order. In this paper, we work with processes integrated of order 1.

We consider a p-dimensional vector autoregression $VAR(k)$ process:

$$y_t = \Phi_1 y_{t-1} + \ldots + \Phi_k y_{t-k} + \varepsilon_t \tag{1}$$

Where y_t is the GGLE at birth, defined by $e_0^F - e_0^M$. With $E\left[\varepsilon_t\right] = 0$, $E\left[\varepsilon_t \varepsilon_t'\right] = \Sigma$ and $E\left[\varepsilon_t \varepsilon_{t-\ell}'\right] = 0$ for any integer ℓ greater than zero. The following representation is derived:

$$\Delta y_t = \Gamma_1 \Delta y_{t-1} + \ldots + \Gamma_{k-1} \Delta y_{t-k+1} + \Pi y_{t-k} + \varepsilon_t \tag{2}$$

Which is called the vector error correction model, where $\Gamma_i = -I + \Phi_1 + \ldots + \Phi_i$ and $\Pi = -I + \Phi_1 + \ldots + \Phi_k$. If $y_t \sim I(1)$, being both differences and innovations stationary, also Πy_{t-k} shall be stationary . Stationarity is possible if the Π matrix is singular. Unless its rank - henceforth, r - is zero, the process is considered to be cointegrated, and the number of cointegrating relationships is r.

For any $r \in [0; p)$, the first hypothesis of the VECM is $H_1 : \Pi = \alpha\beta'$ being α and β two $p \times r$ matrices. The $r \times 1$ vector $\beta' Z_{k,t}$ represents the error correction term. It should be noticed that α and β are not unique, so that the ones estimated with this procedure are not the only possible pair. To test H_1, a maximum likelihood estimation of model parameters is performed. In order to determine the cointegrating rank, the hypothesis to test is that there are no more than r cointegrating vectors, i.e. $H_2 : rk(\Pi) \leq r$ versus $rk(\Pi) > r$. Likelihood ratio test statistic is thus derived:

$$LR(r) = 2ln\left[\ell^*(\Pi, \Sigma, r)\right] - 2ln\left[\ell^*(\Pi, \Sigma)\right] = -T \sum_{i=r+1}^{p} ln(1 - \lambda_i) \tag{3}$$

Where $\lambda : |\lambda S_{kk} - S_{k0} S_{00}^{-1} S - 0k| = 0$ with $S_{ij} := \frac{1}{T} \sum_{t=k}^{T-k} R_{i,t} R_{j,t}'$, for $i, j = 0, 1, \ldots, k$, and $R_{0,t}$ and $R_{k.t}$ are the residuals obtained by regressing respectively Δy_t and y_{t-k} on the lagged differences $\Delta y_{t-1} \ldots \Delta y_{t-k+1}$ (see [4,5])

3   Results

In this section, we show the results concerning cluster 5 that gathers very high-longevity countries. After checking the stationarity, we find that all the GGLE series belonging to this cluster are non-stationary. We model them as linear trend

processes. The cointegration analysis is performed with a $VAR(3)$ representation. The backtesting results are reported in the following. Table 1 shows the Root Mean Square Error (RMSE) values. We can appreciate the superior accuracy of the VECM compared to the VAR model in all the selected countries. On average over the cluster, VECM produce a more than 60% reduction of RMSE with respect to the VAR model.

Table 1. RMSE values for cluster 5. Years 2009–2018.

Cluster 5	DNK	GBR	NLD	NOR	SWE	Average
VECM(3)	0.126	0.093	0.254	0.318	0.094	0.177
VAR(3)	0.358	0.175	0.727	0.661	0.435	0.471

Figure 1 illustrates the GGLE forecast with a VAR(3) derived by the VECM(3) with $r = 1$. The forecasted values (in blue) are compared with the observed value (in black). The predictions provided by the VECM(3) are particularly appreciable for Great Britain and Sweden.

Our analysis allows us to conclude that there is evidence for long period equilibria between the countries included in cluster 5. In other words, it is correct to outline a multi-population pattern of the life expectancy gender gap.
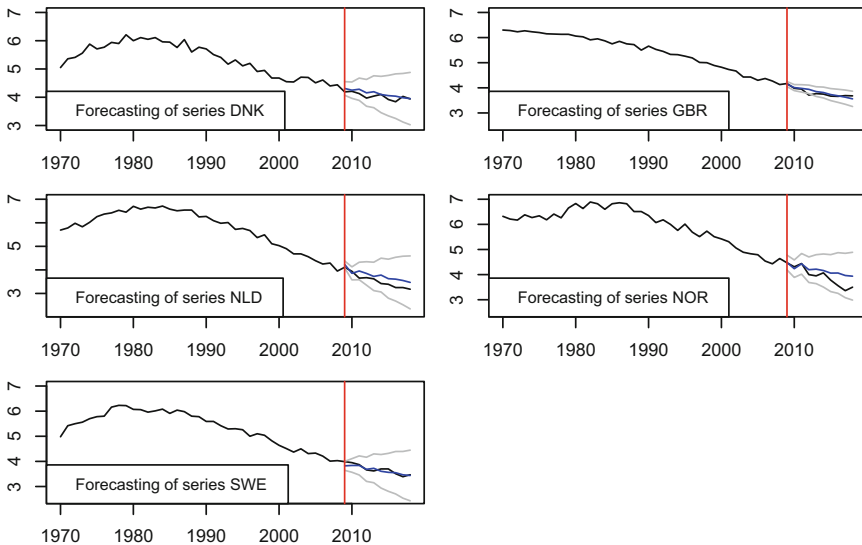
Fig. 1. GGLE forecast for Denmark (DNK), Great Britain (GBR), Netherlands (NLD), Norway (NOR) and Sweden (SWE). Years 2009–2018. Observed values (black) vs forecasted values (blue) and 5%–95% prediction intervals (grey).

4   Conclusions

The research for a model describing the data generating processes has been carried out within the vector autoregressive models framework. The results of this preliminary study show that the evolution of the gender gap in life expectancy at birth follows a homogeneous pattern. Specifically, the analysis based on cluster 5 reflects the demographic theoretical background of gender differences in life expectancy at birth. Overall our findings support the exists of a diverging trend between the male and female populations. During the early period, women experienced a higher survivor, mostly attributable to different sources of mortality. In the recent period, this gap is going to be filled in all countries due to specific policy implementation, therefore the gender gap monitoring might play a crucial role for the public and private sectors.

References

1. Bergeron-Boucher, M.P., Canudas-Romo, V., Pascariu, M., et al.: Modeling and forecasting sex differences in mortality: a sex-ratio approach. Genus **74**, 20 (2018). https://doi.org/10.1186/s41118-018-0044-8
2. Coppola, M., Russolillo, M., Simone, R.: An indexation mechanism for retirement age: analysis of the gender gap. Risks **7**(1), 21 (2019). https://doi.org/10.3390/risks7010021
3. Human Mortality Database. University of California, Berkeley (USA), and Max Planck Institute for Demographic Research (Germany). www.mortality.org
4. Johansen, S.: Statistical analysis of cointegrating vectors. J. Econ. Dyn. Control **12**, 231–254 (1988). https://doi.org/10.1016/0165-1889(88)90041-3
5. Johansen, S., Juselius, K.: Maximum likelihood estimation and inference on cointegration - with applications to the demand for money. Oxf. Bull. Econ. Stat. **52**, 169–210 (1990). http://dx.doi.org/10.1111/j.1468-0084.1990.mp52002003.x
6. Levantesi, S., Nigri, A., Piscopo, G.: Clustering-based simultaneous forecasting of life expectancy time series through long-short term memory neural networks. Int. J. Approx. Reason. **140**, 282–297 (2022). https://doi.org/10.1016/j.ijar.2021.10.008
7. Ntamjokouen, A., Haberman, S., Consigli, G.: Modeling multi-population life expectancy: a cointegration approach. Insurance Markets Companies **5**(2), 12–23 (2014)
8. Pascariu, M.D., Canudas-Romo, V., Vaupel, J.W.: The double-gap life expectancy forecasting model. Insurance Math. Econ. **78**, 339–350 (2018). https://doi.org/10.1016/j.insmatheco.2017.09.011

Decision Making in Portfolio Optimization by Using a Tri-Objective Model and Decision Parameters

Tiziana Ciano[1] and Massimiliano Ferrara[1,2(✉)]

[1] Department of Law, Economics and Human Sciences and Decisions Lab,
University Mediterranea of Reggio Calabria, 89125 Reggio Calabria, Italy
{tiziana.ciano,massimiliano.ferrara}@unirc.it
[2] ICRIOS - The Invernizzi Centre for Research in Innovation, Organization, Strategy and
Entrepreneurship - Bocconi University - Department of Management and Technology,
Via Sarfatti, 25, 20136 Milan, MI, Italy

Abstract. This study proposes a tri-objective portfolio optimization model comprising three objectives, which apart from the return, risk, modelled decision-maker preferences using a proposed composite index. In earlier studies, decision-maker preferences modelled using practical constraints; in contrast, this paper modelled these preferences as constraints along with the proposed composite index based on three decision parameters. To check the effectiveness of the proposed approach is tested on four multi-objective evolutionary algorithms i.e. NSGA-II, SPEA2, MOPSO, and MOEA/D. Finally, conclusions are drawn from the comparative study of these adapted Multi-Objective Evolutionary Algorithms (MOEAs).

Keywords: Multi-objective portfolio optimization · CVaR · Decision parameters

1 Introduction and Motivation of the Study

The transformation of the financial sector driven by technological innovation and the availability of an enormous flows of data and informations is a phenomenon that has attracted growing attention from the Institutions. In fact, huge amounts of data are generated and stored in the financial world every day. Therefore, the enormous availability of data is associated with the ability to process them and build predictive models capable of improving decision-making. Until the 1980s, the classical theory of finance that emerged culminated with the CAPM (and the APT) and the theory of efficient markets. According to the first contribution, stock returns are explained by risk factors as there are rational risk-averse agents in the market pursuing their interest, while the theory of efficient markets holds that stock prices incorporate public information or private companies present in the market making the future returns of the securities unpredictable.

The essence of portfolio optimization [1] is formulating a trade-off between computed magnitudes of returns and risk-term. Numerous internal and exterior elements

influence these two parameters. While formulating a portfolio optimization model comprised of these functions along with realistic constraints, e.g., floor and ceiling constraint, cardinality constraint, pre-assignment constraint, etc., the design can lead to a complicated assorted problem. The classical methods are inapt to generate optimal solutions considering these problems. However, multi-objective evolutionary algorithms (MOEAs) are gaining acceptance for finding solutions to a portfolio optimization problem based on mean-variance theory with realistic constraints [2–4]. The investor might be interested in setting up a trade-off between two or more objectives. Generally, one of the objective is expected return. The other objectives could be reliability [5] or the magnitude of variance. A decision parameter a_i is used for a trade-off between expected return and reliability is described in [5]. A new decision parameter a_i is used in this study for this purpose. The second decision parameter q_i is used to represent an appropriate level of the proportion of capital allocated to a category of sub-portfolios [5]. In the proposed portfolio optimization model, the third decision parameter is d_i, which signifies a level of downside risk for a sub-portfolio, and its value is calculated using CVaR. Therefore, the proposed model has the third objective represented by a single index. The proposed model presented in this paper introduces a tri-objective framework. Three significant decision parameters are used for the formulation of a composite index. An upper bound value of the composite index is also introduced in the model based on a desirable highest level of probability. The proposed tri-objective model is suitable to generate efficient and optimal solutions of the problem in the context of appropriate allocation of capital to sub-portfolios, and it considers the investors preferences regarding downside risk and a trade-off between expected return and variance. The new tri-objective portfolio optimization is described below:

$$Minimize \sigma_p = \sum_{i=1}^{N} \sum_{j=1}^{N} x_i x_j \sigma_{ij} \tag{1}$$

$$Maximize \mu_p = \sum_{i=1}^{N} x_i \mu_i \tag{2}$$

$$Minimize V_p = \sum_{i=1}^{3} dp_i \tag{3}$$

$$\sum_{i=1}^{N} x_i = 1 \tag{4}$$

Where dp_i can have one of the values from any of the three decision parameters i.e. a_i, q_i and d_i. The value of V_p is determined using aggregation of these three decision parameters namely, a_i, q_i and d_i. The user preferences about these three decision parameters allow for selecting an appropriate weight for a decision parameter in the aggregation operation.

The third objective is based on a single index value comprised of an aggregation of these three decision parameters (a_i q_i d_i). The availability of a unique index is beneficial to an investor for selecting an appropriate level for each of these essential decision

parameters. The proposed model is tested on existing MOEAs i.e. NSGA-II, SPEA2, MOEA/D, and MOPSO, and their performances are compared. The third objective is used to study the impact of decision parameters on the portfolios, validate their inclusion in the objective, and analyze the optimal solution based on these parameters to compare the outputs. The equation used for representing the third objective is given below:

$$V_p = a_i + q_i + d_i \tag{5}$$

The decision parameter (a_i) signifies the relative importance of two objectives i.e. return and variance. The decision parameter (d_i) indicates sustainable loss value, and it is related to CVaR value. The decision parameter (q_i) defines the proportion of the capital allocated to sub-portfolios.

Practically, the number of assets N could be enormous. Distributing amount amidst accessible assets needs high management and transaction cost of every asset [3]. Thus, an investor is inclined towards putting investing amount for subsets of N assets by considering a constraint on the size of assets having non-zero weights in the portfolio. This constraint that binds these non-zero weights assets using an integer value is given below:

$$k_1 \leq \sum_{i=1}^{N} z_i \leq k_2 \tag{6}$$

The z_i is a binary-valued parameter, which is used for either including i^{th} asset or excluding it in the portfolio. Therefore, $\forall_i \in \{1, 2, \ldots, N\}$, z_i is given by the following criteria:

$$z_i = \begin{cases} 0, \text{ if } i^{th} \text{ asset is not incorporated in portfolio} \\ 1, \text{ if } i^{th} \text{ asset is incorporated in portfolio} \end{cases} \tag{7}$$

These binary-valued z_is' are utilized in another way for modelling constraints on floor and ceiling, which are named as quantity constraints. These constraints restrict asset distributions inside fixed pre-decided bounds. The first one is used for limiting the cost associated with management, whereas the second one is used to avoid portfolio from exceeding focusing on a specific asset. When the cardinality constraints are used Eq. 2, the quantity constraint is defined as follows

$$z_i lower_i \leq x_i \leq z_i upper_i, \forall_i \in \{1, 2, \ldots, N\} \tag{8}$$

Where $lower_i$ and $upper_i$ are the minimum and maximum bounding value of the amount used for i^{th} asset respectively. The significance of including z_i in Eq. 4 is that when $z_i = 1$, the ith asset is included in the portfolio and amount for i^{th} asset is bounded by $lower_i$ and $upper_i$. As we have considered in this perusal $0 \leq lower_i \leq upper_i \leq 1$ which along with quantity constraints impose long-only constraints ($x_i \geq 0 \ \forall_i \in \{1, 2, \ldots, N\}$).

2   Study Framework and Experimental Results

Traditional optimization methods are incapable of maintaining the portfolio selection model in the existence of realistic constraints like cardinality, pre-assignment [4]. Consequently, the present study aims to seek the use of prevailing MOEAs (multi-objective algorithms) for the introduced model [5].

The representation of a composite index as the third objective, along with its upper bound in the model imposes multi-folded restrictions in the model. Currently, these restrictions are not represented appropriately in a model based on conventional multi-objective algorithms. The upper bound can be included in the framework with fairly less difficulty. A scheme to include the upper bound in the model is described in the current perusal. To check the usefulness of the introduced scheme and to find an optimal solution to the portfolio optimization problem, this framework is tested with four prevailing multi-objective evolutionary algorithms, i.e. NSGA-II [6], SPEA2 [7], MOPSO [8] and MOEA/D [9]. MOEAs are adapted so that the solution of the portfolio optimization problem is computed based on the fulfilment of all the restrictions in the introduced model. The crossover function that is used in the MOPSO algorithm is modified in the proposed strategy. This study evaluates the adapted MOEAs for the tri-objective portfolio optimization model using a dataset for five markets (HS33, DAX100, FTSE100, S&P100, and Nikkei225). The efficient frontiers, along with the outputs based on the hypervolume metric, are used to compare the performances of adapted algorithms. Further, the values of the third objective and the decision parameters are also used for this comparison. Based on hyper-volume statistics, NSGA2 and SPEA2 have better performance as compared to other adapted algorithms. Based on statistics of the decision parameter (a_i), NSGA2 has better performance over other adapted algorithms for market 1, market 2, and market 4. At the same time SPEA2 exceeds in performance over other adapted algorithms for market 3 and market 5. In the case of the decision parameter (q_i), MOPSO exceeds in performance over other adapted algorithms for all markets data set. Similarly, in the case of the decision parameter (d_i), MOPSO has better performance over other adapted algorithms for market 1, market 3, and market 5. At the same time, SPEA2 outperforms other algorithms for market 2, and NSGA2 outperforms other algorithms for market 4. Experimental results indicate that choosing a minimal value of the third objective is preferable for obtaining a higher expected return for a specified range of risk. Moreover, the third objective's value is highly correlated to the value of the decision parameter (a_i). A practical way of calculating the upper bound values of the third objective in the proposed tri-objective model is also illustrated.

This paper uses five market data sets for performing tests of the tri-objective portfolio optimization model taken from Chang et al. [10]. The information contained in these data sets is used for framing UCEF (Unconstrained efficient frontier), values of expected returns, and for constructing the sample matrix of covariances values of assets returns. The time duration used in the data set is comprising of 291 weeks. These data sets are publicly available from Or-library [11] (Table 1).

Table 1. Datasets Or-library [11]

Stock index	N	k_1		k_2
HS33 (Hang Seng)	31	5		10
DAX100 (DAX 100)	85	5		10
FTSE100 (FTSE 100)	89	5		10
S&P100 (S&P 100)	98	5		10
Nikkei225 (Nikkei 225)	225	5		10

The parameters used in the algorithms are adjusted to get the best optimal values before using them in the experiment for the HS33 dataset. Besides, the parameters in the algorithms, apart from the size of the population, the size of the archive, and the maximum number of generations, are calculated for their optimal values before being used in the experiment. The best values of the parameter are obtained for the algorithms (NSGA-II, SPEA2, MOEA/D, and MOPSO) are cataloged in Table 2.

Table 2. Parameters and their values that are employed in algorithms

Name of parameters	MOEA/D	MOPSO	SPEA2	NSGA-II
Size of population	100	100	100	100
Size of archive	–	100	100	–
Number of generation	1000	1000	1000	1000
BEX scale parameter	5	5	5	5
Crossover probability	0.9	0.9	0.9	0.9
Mutation probability	0.7	0.7	0.7	0.7
Total number of runs	20	20	20	20

3 Conclusions

A seminal tri-objective portfolio optimization framework has been introduced in this ongoing work based on objectives of expected return, risk, and a composite index calculated using decision parameters. The use of MOEAs is needed in this approach since traditional methods are incapable of appropriately modeling this portfolio optimization framework. Furthermore, the selection of methods amidst adapted algorithms MOEA/D, SPEA2, NSGA-II, and MOPSO for the tri-objective model is based on the hypervolume metric and diversity of the solutions. NSGA-II is preferred choice of the algorithm based on the hypervolume metric and diversity of the solutions for the portfolio optimization problem. Since the investor is more inclined to a decision that yields a higher expected return, the choice also centers on choosing a lower value of the third objective for a given or specified range of risk. Thus, choosing a minimal value of the upper bound

for the third objective value helps an investor make a decision. Further, the value of the upper bound is highly correlated with the value of the decision parameter a_i, the choice for selecting a minimal value for the a_i parameter becomes a preferable option. These aspects with others will be involved in future research on this fascinating topics and a lot of reflections into Portfolio optimization framework, with a sketch into Portfolio Management, are going to be deeply analyzed by a complex approach and multidisciplinary point of view.

References

1. Markowitz, H.: Portfolio selection. J. Finance **7**(1), 77–91 (1952)
2. Anagnostopoulos, K.P., Mamanis, G.: A portfolio optimization model with three objectives and discrete variables. Comput. Oper. Res. **37**(7), 1285–1297 (2010)
3. Meghwani, S.S., Thakur, M.: Multi-objective heuristic algorithms for practical portfolio optimization and rebalancing with transaction cost. Appl. Soft Comput. **67**, 865–894 (2018)
4. Meghwani, S.S., Thakur, M.: Multi-criteria algorithms for portfolio optimization under practical constraints. Swarm Evol. Comput. **37**, 104–125 (2017)
5. Lejeune, M.A., Shen, S.: Multi-objective probabilistically constrained programs with variable risk: Models for multi-portfolio financial optimization. Eur. J. Oper. Res. **252**(2), 522–539 (2016)
6. Deb, K., Pratap, A., Agarwal, S., Meyarivan, T.A.M.T.: A fast and elitist multiobjective genetic algorithm: NSGA-II. IEEE Trans. Evol. Comput. **6**(2), 182–197 (2002)
7. Zitzler, E., Laumanns, M., Thiele, L.: SPEA2: improving the strength Pareto evolutionary algorithm. TIK-report, 103 (2001)
8. Coello, C.A.C., Pulido, G.T., Lechuga, M.S.: Handling multiple objectives with particle swarm optimization. IEEE Trans. Evol. Comput. **8**(3), 256–279 (2004)
9. Zhang, Q., Li, H.: MOEA/D: a multiobjective evolutionary algorithm based on decomposition. IEEE Trans. Evol. Comput. **11**(6), 712–731 (2007)
10. Chang, T.J., Meade, N., Beasley, J.E., Sharaiha, Y.M.: Heuristics for cardinality constrained portfolio optimisation. Comput. Oper. Res. **27**(13), 1271–1302 (2000)
11. Beasley, J.E.: OR-Library: distributing test problems by electronic mail. J. Oper. Res. Soc. **41**(11), 1069–1072 (1990)

Bitcoin Price Prediction: Mixed Integer Quadratic Programming Versus Machine Learning Approaches

Marco Corazza[✉] and Giovanni Fasano

Ca' Foscari University of Venice, 30121 Venice, Italy
{corazza,fasano}@unive.it
https://www.unive.it/data/people/5592848,
https://www.unive.it/data/people/5592722

Abstract. Reliable Bitcoin price forecasts currently represent a challenging issue, due to the high volatility of this digital asset with respect to currencies in the Forex market. Since 2009 several models for Bitcoin price have been studied, based on neural networks, nonlinear optimization and regression approaches. More recently, Machine Learning paradigms have suggested novel ideas which provide successful guidelines. In particular, in this paper we start from considering the most recent performance of Bitcoin price, along with the history of its price, since they seem to partially invalidate well renowned regression models. This gives room to our Machine Learning and Mixed Integer Programming perspectives, since they seem to provide more reliable results. We remark that our outcomes are data–driven and do not need the fulfillment of standard assumptions required by regression–based approaches. Furthermore, considering the versatility of our approach, we allow the use of standard solvers for MIP optimization problems.

Keywords: Bitcoin · Regression problems · Support Vector Machines · Quadratic Mixed Integer Programming

1 Introduction

This paper details a novel viewpoint to study a price forecast problem, associated with *Bitcoin* [2,4,7], that represents the proposal with the largest market capitalization among the crypto assets. Bitcoin was initially created by an anonymous researcher (or possibly a team of people), under the nickname of Satoshi Nakamoto. The actual identity of such creator has never been revealed so far, and anonymity is expected to be likely maintained also in the future. No private/central bank is responsible for minting novel bitcoins[1], so that Bitcoin/USD rate is merely the result of negotiations among private stakeholders

---

[1] Observe that typically the symbol *Bitcoin* is used to identify the crypto asset, while *bitcoins* represent its coins.

through online exchanges, so that exchanges prevent from double-spending and arbitrary generation of new bitcoins. Direct peer–to–peer transactions involving bitcoins among private investors are also allowed, recurring to special secure protocols that impede abuses. In order to take record of finalized Bitcoin movements, a special distributed ledger, namely a *blockchain*, both does not allow reversibility for transactions and guarantees public information on transactions amount, though preserving anonymity.

In the last decade a number of approaches were introduced in the literature for Bitcoin price prediction (see e.g. [1] and [6], along with therein references), involving investors, researchers, practitioners, as well as private and public institutions, so that some approaches may be undoubtedly considered more methodologically sound with respect to others. Among the main difficulties which negatively affect an accurate prediction of Bitcoin price we find its high volatility, that is yielded by several causes, including a relatively small experience of investors, an unstructured market, the extreme liquidity of bitcoins and the high leverages on Bitcoin transactions.

Unlike the cited references we report here a data–driven approach, based on both Multiobjective Optimization and a Mixed Integer Quadratic Programming formulation, in order to reliably foresee Bitcoin price and exploiting its historical performance.

2   Our Problem

As from [5], in Fig. 1 we summarize some relevant information associated with the price of Bitcoin, from past transactions between 2009 and the end of September 2021. The abscissa axis reports the scaled values of the *Stock–to–Flow (SF)* associated with Bitcoin, i.e.

$$SF = \frac{Stock \ of \ Bitcoin \ at \ a \ given \ date}{Flow \ of \ bitcoins \ in \ a \ given \ time \ window},$$

where *Flow of bitcoins in a given time window* refers to an interval of 463 days (according with the suggestion from the current literature – see also [5]). On the ordinate axis of Fig. 1 we report Bitcoin price. Moreover, we compute the sets $L_{East-South}$ and $L_{West-North}$, being respectively

- $L_{East-South}$: the weak Pareto front associated with both the maximization of the stock-to-flow SF and the minimization of Bitcoin price;
- $L_{West-North}$: the weak Pareto front associated with both the minimization of the stock-to-flow SF and the maximization of Bitcoin price.

Furthermore, we also indicate in Fig. 1 some circled points, corresponding to so called *support vectors*, obtained applying a standard Support Vector Machine (SVM) approach (see also [6] and [5]) to the linear separation problem between the points in the sets $L_{East-South}$ and $L_{West-North}$. The stripe delimited by the lines through these support vectors represents an area where *no extreme* Bitcoin transactions were experienced. In other words, loosely speaking this last area

contains all the pairs of Bitcoin vs. its *SF*, obtained during its history, neither contained in $L_{East-South}$ nor in $L_{West-North}$. Thus, this (reasonably thin) area is likely expected to contain also most of the future transactions.
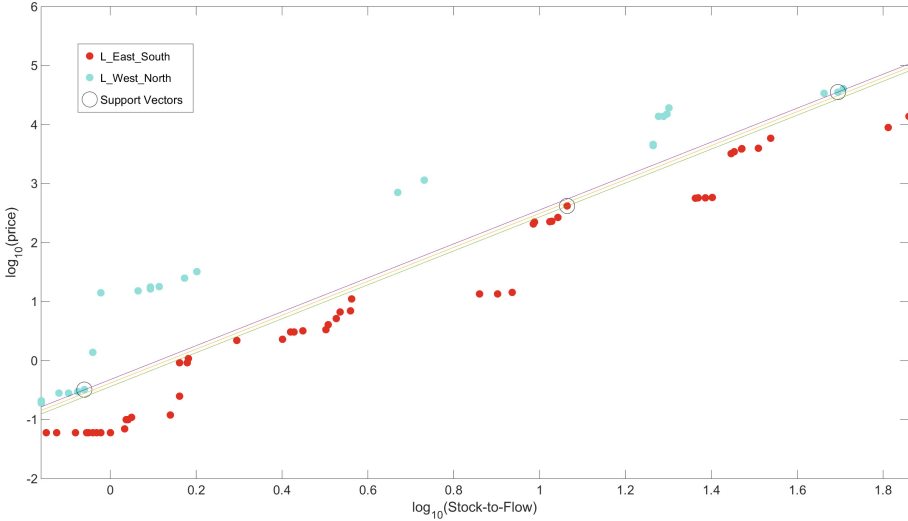
Fig. 1. The weak Pareto fronts $L_{East-South}$ and $L_{West-North}$, corresponding to *extreme* past transactions of Bitcoin price vs. its stock-to-flow ratio: cyan points correspond to relatively favorable (i.e. large priced) transactions, while red points correspond to relatively poor (i.e. poorly priced) transactions.

2.1 Our MIP Viewpoint vs. SVMs

First observe that if N represents the number of all the points in $\mathbb{R}^2$ corresponding to Bitcoin price vs. its *SF* (i.e. our dataset), we consider the next assignment for the labels $\{y_i\}$:

$$y_i = +1 \text{ if the point } P_i \text{ belongs to } L_{West-North},$$
$$y_i = -1 \text{ if the point } P_i \text{ belongs to } L_{East-South}.$$

Then, setting $N_0 = |L_{East-South} \cup L_{West-North}|$, in order to apply the SVM approach in Fig. 1, the solution of the next Convex Quadratic Programming problem is required (see also [1] and therein references for a complete justification)

$$
\begin{aligned}
\min_{\beta,\ \beta_0,\ \xi} \quad & \frac{1}{2}\|\beta\|^2 + C \sum_{i=1}^{N_0} \xi_i \\
\text{s.t.} \quad & (\beta^T x_i + \beta_0) y_i \geq 1 - \xi_i, \quad i = 1, \ldots, N_0, \\
& \xi_i \geq 0, \qquad\qquad\qquad\quad\ i = 1, \ldots, N_0,
\end{aligned}
\tag{1}
$$

In (1) the pair (β, β_0) represents the coefficients of the line (central line in Fig. 1) which best separates the sets $L_{East-South}$ and $L_{West-North}$, i.e. it corresponds to the largest possible stripe delimited by the support vectors in Fig. 1. The quantities x_i, $i = 1, \ldots, N_0$, in our application represent *scalars*, corresponding to a value (i.e. *SF*) in the abscissa axis. Nevertheless, in a more general framework they may represent n–real *vectors*, each representing a vector of values including the *SF*; this explains why in (1) we preferred to introduce the (more general) inner product between the n–real vector (of coefficients) β and the vector x_i, for any i. A similar consideration holds also for the formulations (2) and (3) below.

We highlight that some elements of the solution of this mathematical programming problem, namely β^* and β_0^*, provide the optimal estimates of the coefficients of the central line inside the area delimited by the weak Pareto fronts $L_{East-South}$ and $L_{West-North}$. Similarly, this also happens for the subsequent formulations (2) and (3). Note that these optimal estimates are determined accordingly to a data-driven SVM-based optimization approach which, loosely speaking, regresses the Bitcoin price on the *SF*. Note also that the above (optimal) central line constitutes the long term Bitcoin price forecaster; similarly, again, for the mathematical programming problems (2) and (3).

Moreover, it can be proved that the quantities $\{\xi_i\}$ will be all equal to zero if and only if the two sets $L_{East-South}$ and $L_{West-North}$ are linearly separable. We also observe that after assigning $i \geq 1$ labels $\{y_1, \ldots, y_i\} \subseteq \{+1, -1\}$ to the points $\{P_1, \ldots, P_i\}$, not contained in $L_{East-South} \cup L_{West-North}$, then in (1) we can replace N_0 by N_i, being $N_i \supset N_0$ and $N_i = N_0 + i$ (i.e. N_i points in the dataset have been labelled). Thus, it can easily be proved that adopting the procedure in [5], including one point of our dataset at a time, starting from N_0 up to N, we can iteratively refine the solution of (1) by solving

$$
\begin{aligned}
\min_{\beta, \ \beta_0, \ \xi} \quad & \frac{1}{2}\|\beta\|^2 + C \sum_{j=1}^{N_i} \xi_j \\
\text{s.t.} \quad & (\beta^T x_j + \beta_0)y_j \geq 1 - \xi_j, \quad j = 1, \ldots, N_i, \\
& \xi_j \geq 0, \qquad\qquad\qquad\quad j = 1, \ldots, N_i.
\end{aligned}
\tag{2}
$$

Thus, after solving $N - N_0 + 1$ SVMs *(i.e. $i = 1, \ldots, N - N_0 + 1$)*, each corresponding to a binary separation problem, all the N points in the dataset will be classified as either *closer* to $L_{East-South}$ or *closer* to $L_{West-North}$. We strongly highlight that, as detailed in [5], there are applications (e.g. from *semi–supervised learning*) where the labels $\{y_1, \ldots, y_N\}$ are not all known when solving the first SVM in the sequence: this justifies the above iterative procedure.

Now, considering a completely different perspective we might alternatively replace the above iterative SVM–based procedure with a unique Mixed Integer Quadratic Programming (MIQP) reformulation. In this regard, let us preliminarily set

$$
\begin{cases}
A \equiv L_{West-North}, \\[2mm]
B \equiv L_{East-South}.
\end{cases}
$$

Then, in place of the above $N - N_0 + 1$ SVMs we can propose to solve the MIQP problem (note that $M \gg 1$ is not an unknown but it represents a *Big–M*, i.e. a large enough constant value)

$$
\begin{aligned}
\min_{\beta, \ \beta_0, \ \xi, \ \gamma} \quad & \frac{1}{2}\|\beta\|^2 + C\sum_{i=1}^{N} \xi_i \\
\text{s.t.} \quad & -(\beta^T x_i + \beta_0) + 1 - \xi_i \leq (1-\gamma_i)M, \ i \ : \ x_i \notin A \cup B, \\
& (\beta^T x_i + \beta_0) + 1 - \xi_i \leq \gamma_i M, \quad\ i \ : \ x_i \notin A \cup B, \\
& -(\beta^T x_i + \beta_0) + 1 - \xi_i \leq 0, \quad\ \ \ \ i \ : \ x_i \in A, \\
& (\beta^T x_i + \beta_0) + 1 - \xi_i \leq 0, \quad\ \ \ \ \ i \ : \ x_i \in B, \\
& \xi_i \geq 0, \quad\quad\quad\quad\quad\quad\quad\quad\quad\quad i = 1, \ldots, N, \\
& \gamma_i \in \{0, 1\}, \quad\quad\quad\quad\quad\quad\quad\ \ i \ : \ x_i \notin A \cup B,
\end{aligned}
\tag{3}
$$

where:

– the objective function maintains the same structure with respect to the formulation (1) (i.e. a convex quadratic functional);
– $2(N - |A \cup B|)$ linear constraints are added with respect to (1);
– $2(N - |A \cup B|)$ binary unknowns (the unknowns $\{\gamma_i\}$) are added with respect to (1), so that if $(\beta^*, \beta_0^*, \xi^*, \gamma^*)$ is the final solution of (3) then:
  • $\gamma_i^* = 0$ means $x_i \in B$ (and in (1) we will equivalently have $y_i = -1$),
  • $\gamma_i^* = 1$ means $x_i \in A$ (and in (1) we will equivalently have $y_i = +1$).

Remark 1. We highlight that the formulation (3) is *equivalent* to solve the $N - N_0 + 1$ SVMs in (2). Indeed, under mild assumptions (3) provides the same results of the $N - N_0 + 1$ SVMs in (2). However, note that (3) unlike (2) contains integer unknowns, that typically increase the computational burden and require a more sophisticated solver.

Several additional properties can be proved for the formulation (3), with respect to considering a sequence of $N - N_0 + 1$ SVMs, including some interesting numerical results. The reader may refer to [5] and [3] for a more thorough description, along with additional suggestions and a complete analysis of the outcomes of the above methodologies on several practical applications. We also remark that C represents the unique parameter included in the formulation (3), and its assessment typically follows two guidelines: on one hand it is chosen large enough to penalize misclassification (i.e. when C is large we tend to reduce the number of nonzero unknowns $\{\xi_i\}$); on the other hand, a too large value for C may imply a relatively large time of computation. In our *Matlab* implementation we set $C = $ `inf` and no numerical odd was experienced.

References

1. Aggarwal, D., Chandrasekaran, S., Annamalai, B.: A complete empirical ensemble mode decomposition and support vector machine–based approach to predict Bitcoin prices. J. Behav. Exp. Finance **27**, 100335 (2020)

2. Blockchain information for Bitcoin (BTC). https://www.blockchain.com/charts/total--bitcoins
3. Caliciotti, A., Corazza, M., Fasano, G.: Regression models and machine learning approaches for long term bitcoin price forecast. Ann. Oper. Res. (2022, submitted)
4. Nakamoto, S.: Bitcoin: a peer-to-peer electronic cash system. http://www.bitcoin.org/bitcoin.pdf
5. Pontiggia, A., Fasano, G.: Data Analytics and Machine Learning paradigm to gauge performances combining classification, ranking and sorting for system analysis. Working Paper 05/2021, Department of Management, University Ca' Foscari of Venice (2021)
6. Sreekanth Reddy, L., Sriramya, P.: A research on bitcoin price prediction using machine learning algorithms. Int. J. Sci. Technol. Res. **9**, 1600–1604 (2020)
7. Vigna, P., Casey, M.J.: The Age of Cryptocurrency: How Bitcoin and Digital Money Are Challenging the Global Economic Order, 1st edn. St. Martin's Press, New York (2015)

Verifying the Rényi Dependence Axioms for a Non-linear Bivariate Comovement Index

Marco Corazza[1]([✉]), Elisa Scalco[2], and Claudio Pizzi[1]

[1] Ca' Foscari University of Venice, Venice, Italy
{corazza,pizzic}@unive.it
[2] London, United Kingdom

Abstract. In a paper appeared some years ago, an index for evaluating the non-linear bivariate comovement between two asset prices has been proposed. In this paper, we assess if that index satisfies the classical seven axioms formulated by Rényi that a measure of dependence should meet. In the cases in which the index does not fulfil an axiom, we propose a weakened version of that statement the index satisfies.

1 Introduction

In a paper appeared some years ago, a methodology for the non-linear evaluation of bivariate comovement between asset prices has been introduced (see [1]). That methodology proposes an index for evaluating the comovement between two random variables. In this paper, we assess if that index satisfies the classical seven axioms formulated by Rényi that a measure of dependence should meet (see [5]). In the cases in which the index does not fulfil a given axiom, we propose a weakened version of the original statement the index satisfies.

The remainder of the paper is organized as follows. In the next section we recall the definition of the index and introduce the seven axioms formulated by Rényi. In Sect. 3 we verify if and which among such axioms are satisfied by the index, for the axioms which are not satisfied we propose weakened versions of their formulations the index satisfies, and lastly we conclude with some final remarks.

2 The Comovement Index and the Rényi Dependence Axioms

The definition of the index softly draws one's inspiration from the concept of comonotonicity; limiting our interest to the bivariate case, see for details [4]. We can provide such a definition as follows. Let us start by considering two discrete-time time series, $\{X_1(t), t = t_1, \ldots, t_N\}$ and $\{X_2(t), t = t_1, \ldots, t_N\}$, and

E. Scalco—Independent.

their first order differences $\Delta_1(t) \doteq X_1(t) - X_1(t-1)$ and $\Delta_2(t) \doteq X_2(t) - X_2(t-1)$, respectively. The index for evaluating the non-linear bivariate dependence between the random variables $X_1(t)$ and $X_2(t)$ is:

$$\delta_{1,2} = \left| \frac{1}{N-1} \sum_{t=t_2}^{t_N} \Delta(t)_{1,2} \right|, \qquad \text{where} \qquad \Delta(t)_{1,2} = \begin{cases} -1 \text{ if } \Delta_1(t)\Delta_2(t) < 0 \\ 1 \text{ if } \Delta_1(t)\Delta_2(t) > 0 \end{cases}.$$

As for the Rényi dependence axioms, let us consider two random variables X_1 and X_2 and a functional $\mu(X_1, X_2) \to \mathbb{R}$. $\mu(X_1, X_2)$ is a measure of dependence between X_1 and X_2 in the Rényi sense if it satisfies the following axioms: 1) $\mu(X_1, X_2)$ is defined for any pair of random variables X_1 and X_2, neither of them being constant with probability 1; 2) $\mu(X_1, X_2) = \mu(X_2, X_1)$; 3) $0 \le \mu(X_1, X_2) \le 1$; 4) $\mu(X_1, X_2) = 0$ if and only if X_1 and X_2 are stochastically independent; 5) $\mu(X_1, X_2) = 1$ if there is a strict dependence relationship between X_1 and X_2, i.e. either $X_1 = \varphi(X_2)$ or $X_2 = \psi(X_1)$, where $\varphi(\cdot)$ and $\psi(\cdot)$ are Borel-measurable functions; 6) If Borel-measurable functions $\varphi(\cdot)$ and $\psi(\cdot)$ map the real axis in a one-to-one way onto itself, then $\mu(\varphi(X_1, X_2), \psi(X_1, X_2)) = \mu(X_1, X_2)$; 7) If the joint distribution of X_1 and X_2 is normal, then $\mu(X_1, X_2) = |\rho_{1,2}(X_1, X_2)|$, where $\rho_{1,2}(X_1, X_2)$ is the linear correlation coefficient between X_1 and X_2.

3   Is $\delta_{1,2}$ a Measure of Dependence à la Rényi?

To verify if $\delta_{1,2}$ satisfies the Rényi dependence axioms, we consider $|\delta_{1,2}|$. In fact, if we do not recur to the absolute value of $\delta_{1,2}$ it comes up that $-1 \le \delta_{1,2} \le 1$, and it is possible to prove that this property can not coexist with the property stated in Axiom 4 (see for details [2]).

Proposition 1. $|\delta_{1,2}|$ satisfies Axiom 1.

Proof. The index $|\delta_{1,2}|$ is given by the summation of the variable $\Delta(t)_{1,2}$ which assumes value -1 or 1 according to $\Delta_1(t)\Delta_2(t) < 0$ or $\Delta_1(t)\Delta_2(t) \ge 0$, respectively. For that reason, $|\delta_{1,2}|$ results to be defined for every value of $(X_1(\cdot), X_2(\cdot))$, i.e. for every bivariate time series $(X_1(t), X_2(t))$, with $t = t_1, ..., t_N$. □

Proposition 2. $|\delta_{1,2}|$ satisfies Axiom 2.

Proof. By the commutative property of multiplication, it is immediate to verify that $\Delta_1(t)\Delta_2(t) = \Delta_2(t)\Delta_1(t)$, and therefore that $|\delta_{1,2}| = |\delta_{2,1}|$. □

Proposition 3. $|\delta_{1,2}|$ satisfies Axiom 3.

Proof. First, let us consider the case in which $|\delta_{1,2}| = 1$. Such a situation might occur either if all the products of differences are negative or if all of them are non-negative, i.e. when the two random variables $X_1(\cdot)$ and $X_2(\cdot)$ are counterdependent or codependent, respectively. Then, let us consider the case in which $|\delta_{1,2}| = 0$. It represents the situation in which, if $N - 1$ is an even number,

then half of the products of differences are negative and the other half are non-negative. Finally, let us consider the case in which $0 < |\delta_{1,2}| < 1$. Such a situation might occur if at least 1 and at most $N - 2$ of the $N - 1$ products of differences is negative (non-negative) and all the others are non-negative (negative). $\square$

Proposition 4. $|\delta_{1,2}|$ does not satisfy Axiom 4.

Proof. We prove both the necessary condition and the sufficient one by counterexamples. ($\Leftarrow$) Irrespective of the hypothesis following which $X_1(t)$ and $X_2(t)$ are stochastically independent, let us consider the situation in which $N - 1$ is an odd number. In this case, the lowest value $|\delta_{1,2}|$ can reach is $1/N$, which is different from 0, when $(N - 2)/2$ products of differences are negative and other $(N - 2)/2$ are non-negative. ($\Rightarrow$) Let us consider two dependent random variables $X_1(t)$ and $X_2(t)$ such that $X_1(t) = 2 - [X_2(t)]^2$, and their discrete-time time series in Table 1.

Table 1. Discrete-time time series of $X_1(t)$ and $X_2(t)$.

Random variable	t_1	t_2	t_3	t_4	t_5
$X_1(t)$	1	1	7/4	41/25	41/25
$X_2(t)$	1	1	1/2	3/5	3/5

Thus the index $|\delta_{1,2}|$ turns out to be

$$|\delta_{1,2}| = \left| \frac{1}{4} \sum_{t=t_2}^{t_5} \Delta(t)_{1,2} \right| = \left| \frac{1}{4}(+1 - 1 - 1 + 1) \right| = 0. \square$$

Now, we propose a weakened version of the statement of Axiom 4 and prove it.

Proposition 5. Let us consider the random variables $X_1(\cdot)$ and $X_2(\cdot)$, their first order differences $\Delta_1(t)$ and $\Delta_2(t)$, with $t = t_1, \ldots, t_N$, and the probabilities $\Pr(\Delta_1(t) < 0) = p_1$, $\Pr(\Delta_1(t) > 0) = q_1$, $\Pr(\Delta_2(t) < 0) = p_2$ and $\Pr(\Delta_2(t) > 0) = q_2$, with $0 \leq p_1, q_1, p_2, q_2 \leq 0$. If the events $(\Delta_1(t) < 0)$ and $(\Delta_2(t) > 0)$ and the events $(\Delta_1(t) > 0)$ and $(\Delta_2(t) < 0)$ are stochastically independent for any t, respectively, and if $p_1 q_2 + q_1 p_2 = \frac{\lfloor (N-1)/2 \rfloor}{N-1}$, then

$$E\left(|\delta_{1,2}| \right) = 0 \text{ if } N - 1 \text{ is even and } \lim_{N-1 \to +\infty} E\left(|\delta_{1,2}| \right) = 0 \text{ if } N - 1 \text{ is odd.}$$

Proof. Let us start by determining the probability that $\Delta_{1,2}(t) = -1$, i.e. $\Pr(\Delta_1(t)\Delta_2(t) < 0)$:

$$\begin{aligned}
\Pr\left(\Delta_1(t)\Delta_2(t) < 0\right) &= \Pr\left((\Delta_1(t) < 0 \cap \Delta_2(t) > 0) \cup (\Delta_1(t) > 0 \cap \Delta_2(t) < 0)\right) \\
&= \Pr\left(\Delta_1(t) < 0 \cap \Delta_2(t) > 0\right) + \Pr\left(\Delta_1(t) > 0 \cap \Delta_2(t) < 0\right) \\
&= \Pr\left(\Delta_1(t) < 0\right)\Pr\left(\Delta_2(t) > 0\right) + \Pr\left(\Delta_1(t) > 0\right)\Pr\left(\Delta_2(t) < 0\right) \\
&= p_1 q_2 + q_1 p_2.
\end{aligned}$$

Note that: the second row of the previous relationship is due, by construction, to the fact that the events $(\Delta_1(t) < 0 \cap \Delta_2(t) > 0)$ and $(\Delta_1(t) > 0 \cap \Delta_2(t) < 0)$ are incompatible for any t; the third row of the same relationship is due, by hypothesis, to the fact that the events $(\Delta_1(t) < 0)$ and $(\Delta_2(t) > 0)$ and the events $(\Delta_1(t) > 0)$ and $(\Delta_2(t) < 0)$ are stochastically independent for any t, respectively. Of course, the probability that $\Delta_{1,2}(t) = 1$, i.e. $\Pr\left(\Delta_1(t)\Delta_2(t) \geq 0\right)$, is $1 - \Pr\left(\Delta_1(t)\Delta_2(t) < 0\right) = 1 - p_1 q_2 + q_1 p_2$. Now, let us express the probabilities above in terms of $N - 1$, i.e.:

$$\Pr\left(\Delta_1(t)\Delta_2(t) < 0\right) = p_1 q_2 + q_1 p_2 = \frac{\lfloor (N-1)/2 \rfloor}{N-1} = \begin{cases} \frac{1}{2} & \text{if } N-1 \text{ is even} \\ \frac{1}{2}\frac{N-2}{N-1} & \text{if } N-1 \text{ is odd} \end{cases}$$

and

$$\Pr\left(\Delta_1(t)\Delta_2(t) \geq 0\right) = 1 - p_1 q_2 + q_1 p_2 = 1 - \frac{\lfloor (N-1)/2 \rfloor}{N-1} = \begin{cases} \frac{1}{2} & \text{if } N-1 \text{ is even} \\ \frac{N}{N-1}\frac{1}{2} & \text{if } N-1 \text{ is odd} \end{cases}.$$

Therefore

$$E\left(\left|\delta_{1,2}\right|\right) = -1\frac{1}{2} + 1\frac{1}{2} = 0 \text{ if } N-1 \text{ is even}$$

and

$$\lim_{N-1 \to +\infty} E\left(\left|\delta_{1,2}\right|\right) = \lim_{N-1 \to +\infty}\left(-1\frac{1}{2}\frac{N-2}{N-1} + 1\frac{1}{2}\frac{N}{N-1}\right) = 0 \text{ if } N-1 \text{ is odd.} \quad \square$$

Proposition 6. $\left|\delta_{1,2}\right|$ does not satisfy Axiom 5.

Proof. We prove that the assumption of a strict dependence relationship between $X_1(\cdot)$ and $X_2(\cdot)$ is not a sufficient condition for $\left|\delta_{1,2}\right| = 1$. We prove it only for $\varphi(\cdot)$ as the proof for $\psi(\cdot)$ is absolutely similar. Let $\varphi(\cdot)$ be a Borel-measurable function such that $X_1 = \varphi(X_2)$, and let us reformulate the definition of $\left|\delta_{1,2}\right|$ taking into account the equivalence $X_1 = \varphi(X_2)$. So, we have

$$\left|\delta_{1,2}\right| \doteq \left|\delta_{1,2}(\varphi(X_2(t)), X_2(t))\right| = \left|\frac{1}{N-1}\sum_{t=t_2}^{t_N}\Delta(\varphi;t)_{1,2}\right|, \text{ where} \quad (1)$$

$$\Delta(\varphi;t)_{1,2} = \begin{cases} -1 \text{ if } [\varphi(X_2(t)) - \varphi(X_2(t-1))][X_2(t) - X_2(t-1)] < 0 \\ 1 \text{ if } [\varphi(X_2(t)) - \varphi(X_2(t-1))][X_2(t) - X_2(t-1)] \geq 0 \end{cases}.$$

Now, without lack of generality, let us assume that $\varphi(\cdot)$ is a strictly unimodal function with mode equal to m, i.e. a function which is strictly increasing for $X_2(\cdot) \leq m$ and strictly decreasing for $X_2(\cdot) \geq m$. Thus, for the points $X_2(\cdot) \leq m$ we have that $\varphi(\cdot)$ preserves the order and then $\Delta(\varphi;t)_{1,2} = 1$ is added in Eq. 1. On the contrary, for the points $X_2(\cdot) \geq m$, we have that $\varphi(\cdot)$ reverses the order and then $\Delta(\varphi;t)_{1,2} = -1$ is added in Eq. 1. As consequent result $|\delta_{1,2}| < 1$ and hence $|\delta_{1,2}|$ does not meet Axiom 5. □

Now, we propose a weakened version of the statement of axiom 5 and prove it.

Proposition 7. $|\delta_{1,2}| = 1$ if and only if there exists a dependence relationship between $X_1(\cdot)$ and $X_2(\cdot)$, i.e. $X_1 = \varphi(X_2(\cdot))$ or $X_2 = \psi(X_1(\cdot))$, where $\psi(\cdot)$ and $\varphi(\cdot)$ are either strictly decreasing functions or (not strictly) increasing functions on $\text{Range}(X_1(\cdot))$ and on $\text{Range}(X_2(\cdot))$, respectively.

Proof. First, we prove that if there exists a dependence relationship between $X_1(\cdot)$ and $X_2(\cdot)$ as defined above, then $|\delta_{1,2}| = 1$. Second, we prove the other direction of the biimplication. As done for the previous proposition, we prove it only for $\varphi(\cdot)$ as the proof for $\psi(\cdot)$ is absolutely similar. ($\Leftarrow$) Let $\varphi(\cdot)$ be a dependence relationship as specified in the statement of Proposition 7 such that $X_1 = \varphi(X_2)$, and let us reformulate the definition of $|\delta_{1,2}|$ taking into account the equivalence $X_1 = \varphi(X_2)$ as in Eq. (1). If $\varphi(\cdot)$ is strictly decreasing, then it reverses the order and $\Delta(\varphi;t)_{1,2} = -1$ is added in Eq. (1) for each product of differences. Consequently, $|\delta_{1,2}| = 1$. Similarly, if $\varphi(\cdot)$ is increasing, then it is not order-reversing and $\Delta(\varphi;t)_{1,2} = 1$ is added in Eq. (1) for each product of differences. Consequently, still $|\delta_{1,2}| = 1$. ($\Rightarrow$) By hypothesis $|\delta_{1,2}| = 1$, therefore the products of differences have to be either all negative or all non-negative. The former case implies that if $X_2(t) \lessgtr X_2(t-1)$, with $X_2(\cdot) \in \text{Range}(X_2(\cdot))$, then $\varphi(X_2(t)) \gtrless \varphi(X_2(t-1))$. Therefore, $\varphi(\cdot)$ is a strictly decreasing function. The latter case implies that if $X_2(t) \lessgtr X_2(t-1)$, with $X_2(\cdot) \in \text{Range}(X_2(\cdot))$, then $\varphi(X_2(t)) \lesseqgtr \varphi(X_2(t-1))$. So, $\varphi(\cdot)$ is a (not strictly) increasing function. □

Proposition 8. $|\delta_{1,2}|$ does not satisfy Axiom 6.

Proof. Let $\varphi(\cdot), \psi(\cdot): \mathbb{R} \to \mathbb{R}$ be one-to-one Borel-measurable functions, and let us reformulate as follows the definition of $|\delta_{1,2}|$ in terms of $\varphi(\cdot)$ and $\psi(\cdot)$:

$$|\delta_{1,2}(\varphi(X_1(\cdot)), \psi(X_2(\cdot)))| = \left| \frac{1}{N-1} \sum_{t=t_2}^{t_N} \Delta(\varphi, \psi; t)_{1,2} \right|, \text{ where} \qquad (2)$$

$$\Delta(\varphi, \psi; t)_{1,2} = \begin{cases} -1 \text{ if } [\varphi(X_1(t)) - \varphi(X_1(t-1))][\psi(X_2(t)) - \psi(X_2(t-1))] < 0 \\ 1 \text{ if } [\varphi(X_1(t)) - \varphi(X_1(t-1))][\psi(X_2(t)) - \psi(X_2(t-1))] \geq 0 \end{cases}.$$

Now, without lack of generality, let us assume that both $\varphi(\cdot)$ and $\psi(\cdot)$ are strictly unimodal functions having modes equal to m and n, respectively, with $m \neq n$ and, for instance, $m < n$. (Note that a monotone and one-to-one function is strictly monotone.) Thus, for the points $X_1(\cdot)$, $X_2(\cdot) \leq \min\{m, n\} \doteq m$ we have that both $\varphi(\cdot)$ and $\psi(\cdot)$ are order preserver and then $\Delta(\varphi, \psi; t)_{1,2} = \Delta(t)_{1,2}$. Similarly, for the points $X_1(\cdot)$, $X_2(\cdot) \geq \max\{m, n\} \doteq n$ we have that both $\varphi(\cdot)$ and $\psi(\cdot)$ are order reverser and then again $\Delta(\varphi, \psi; t)_{1,2} = \Delta(t)_{1,2}$. On the contrary, for the points $m \leq X_1(\cdot)$, $X_2(\cdot) \leq n$ we have that $\varphi(\cdot)$ reverses the order whilst $\psi(\cdot)$ preserves the order and then $\Delta(\varphi, \psi; t)_{1,2} \neq \Delta(t)_{1,2}$. So, $\left|\delta_{1,2}(\varphi(X_1(\cdot))\psi(X_2(\cdot)))\right| \neq \left|\delta_{1,2}\right|$. $\square$

Now, we propose a weakened version of the statement of Axiom 6 and prove it.

Proposition 9. If $\varphi(\cdot), \psi(\cdot) : \mathbb{R} \to \mathbb{R}$ are either strictly decreasing functions or (not strictly) increasing functions on $\text{Range}(X_1(\cdot))$ and on $\text{Range}(X_2(\cdot))$, respectively, then $\left|\delta_{1,2}(\varphi(X_1), \psi(X_2))\right| = \left|\delta_{1,2}\right|$.

Proof. Let $\varphi(\cdot)$ and $\psi(\cdot)$ be either strictly decreasing functions or (not strictly) increasing functions on $\text{Range}(X_1(\cdot))$ and on $\text{Range}(X_2(\cdot))$, respectively, and let us reformulate the definition of $\left|\delta_{1,2}\right|$ in terms of $\varphi(\cdot)$ and $\psi(\cdot)$ as in Eq. (2). Now, the following two cases are possible: First, if $\varphi(\cdot)$ and $\psi(\cdot)$ are strictly decreasing functions, then both of them are order-reversing. So, whenever $X_1(t-1) \lessgtr X_1(t)$ and $X_2(t-1) \lessgtr X_2(t)$, then $\varphi(X_1(t-1)) \gtrless \varphi(X_1(t))$ and $\psi(X_2(t-1)) \gtrless \psi(X_2(t))$, respectively. As consequence, we have $\Delta(\varphi, \psi; t)_{1,2} = \Delta(t)_{1,2}$ that implies $\left|\delta_{1,2}(\varphi(X_1), \psi(X_2))\right| = \left|\delta_{1,2}\right|$; Second, if $\varphi(\cdot)$ and $\psi(\cdot)$ are not strictly increasing functions, then both of them are not order-reversing. So, whenever $X_1(t-1) \lessgtr X_1(t)$ and $X_2(t-1) \lessgtr X_2(t)$, then $\varphi(X_1(t-1)) \lessgtr \varphi(X_1(t))$ and $\psi(X_2(t-1)) \lessgtr \psi(X_2(t))$, respectively. As consequence, we have still $\Delta(\varphi, \psi; t)_{1,2} = \Delta(t)_{1,2}$ that implies $\left|\delta_{1,2}(\varphi(X_1), \psi(X_2))\right| = \left|\delta_{1,2}\right|$. Therefore, $\left|\delta_{1,2}(\varphi(X_1), \psi(X_2))\right| = \left|\delta_{1,2}\right|$. $\square$

Proposition 10. $\left|\delta_{1,2}\right|$ does not satisfy Axiom 7.

Proof. We prove this proposition by a counterexample. Let us assume that $X_1(\cdot)$ and $X_2(\cdot)$ are uncorrelated, i.e. that $\left|\rho_{1,2}\right| = 0$, and let us consider the situation in which $N-1$ is an odd number. In this case, the lowest value $\left|\delta_{1,2}\right|$ can reach is $1/N$, which is different from the assumed value of $\left|\rho_{1,2}\right|$, i.e. 0. $\square$

For Proposition 10 it is few meaningful to propose a weakened version of its statement because it would require the relaxation of at least the normality of the bivariate joint distribution. By doing so, the weakened version of the statement would appear to much weakened.

Concluding, note that, as premised in Sect. 3, $\delta_{1,2}$ does not achieve some of the Rényi's dependence axioms. This occurrence is not surprising. Indeed, this stems from the fact that the Rényi's set of dependence axioms is probably not the most natural for a measure of dependence. Furthermore, in the literature there is a large consensus following which a scalar measure of bivariate dependence, should satisfy more suitable properties like, for example, the ones considered in [3] and [6]. Note also that we are currently investigating the possibility of formally deriving a linkage between the concept of comonotonicity and the idea of codependence expressed in terms of $\delta_{1,2}$.

References

1. Corazza, M., Malliaris, A.G., Scalco, E.: Nonlinear bivariate comovements of asset prices: methodology, tests and applications. Comput. Econ. **35**, 1–23 (2010)
2. Embrechts, P., McNeil, A., Straumann, D.: Correlation and dependency in risk management: properties and pitfalls. In: Dempster, M., Moffatt, H. (eds.) Risk Management: Value at Risk and Beyond. Cambridge University Press, Cambridge (2002)
3. Hutchinson, T.P., Lai, C.D.: Continuous Bivariate Distributions, Emphasising Applications. Rumsby Scientific Publishing (1990)
4. Jouini, E., Napp, C.: Comonotonic processes. Insur. Math. Econom. **32**, 255–265 (2003)
5. Rényi, A.: On measures of dependence. Acta Math. Acad. Sci. Hung. **10**, 441–451 (1959)
6. Szegö, G.: Measures of risk. Eur. J. Oper. Res. **163**, 5–19 (2005)

Inflation Perceptions and Expectations During the Pandemic: A Model Based Approach

Marcella Corduas[(✉)] and Domenico Piccolo

Department of Political Sciences, University of Naples Federico II, Naples, Italy
{marcella.corduas,domenico.piccolo}@unina.it
http://www.unina.it

Abstract. This article investigates how consumers' perceptions and expectations about inflation evolved during the first twelve months after the pandemic broke out in Italy. The analysis is based on data from the European business and consumer qualitative surveys and exploits an innovative dynamic model for ordinal data based on a mixture distribution with time varying parameters.

Keywords: Inflation expectations · Inflation perceptions · CUB models · Ordinal data

1 Introduction

The Covid-19 pandemic originated an unprecedent and sudden shock for the economic systems in all world countries. The crisis had unusual characteristics due to the uncertainty about the duration of the pandemic and the medium/long term effects on the economy. This uncertainty resulted in a substantial change for the worse in the economic sentiment in all EU countries [16]. The pandemic affected inflation in different ways. During the period of more drastic measures, the falling demand caused by the reduced opportunity of consumption and the disruptions of supply chains produced opposite pressures on prices. In addition, the sudden changes in the consumers' expenditure patterns introduced a bias in the measurement of inflation based on the Consumer Price Indices [2,5]. Within this framework, monitoring consumers' beliefs of current and future inflation was a challenging issue for monetary authorities and policy makers [6,10] because consumers' beliefs about inflation development influence spending and saving behavior, borrowing and wage bargaining, and affect overall macroeconomic outcomes. The pandemic changed both perceptions and expectations about the development of inflation introducing a sort of bias in the opinion formation process [1,3].

The aim of this work is to analyze how opinions of Italian consumers evolved during the pandemic period. The investigation relies on data from the qualitative EU business and economic survey and exploits an innovative dynamic model for ordinal data based on a mixture distribution with time varying parameters [7]. The article is organized as follows. Section 2 briefly illustrates the model. Section 3 discusses the results. Section 4 presents some final remarks.

2   The Model

The mixture distribution for the analysis of rating data − proposed by Piccolo [13] and developed in various directions [14] − describes the respondent's mechanism for the selection of an ordinal score. This is a combination of the respondent's deep-rooted attitude towards the item and the tendency to make a completely random choice. In a previous contribution, Corduas [7] extended the original formulation to frequency distributions of repeated ordinal responses, allowing the relationship of parameters to explanatory variables observed over time. In particular, the model is expressed as follows, for $t = 1, .., n$:

$$P(Y_t = y) = \delta_t D_{yt}^c + (1 - \delta_t) \left[\pi_t \binom{m - 1}{y - 1} (1 - \xi_t)^{y-1} \xi_t^{m-y} + (1 - \pi_t) \frac{1}{m} \right],$$
$$y = 1, 2, ..., m.$$

$$\pi_t = \frac{1}{1 + e^{-\beta_0 - \beta_1 w_{t-1}}}; \ \ \xi_t = \frac{1}{1 + e^{-\gamma_0 - \gamma_1 z_{t-1}}}; \ \ \delta_t = \frac{1}{1 + e^{-\alpha_0 - \alpha_1 v_{t-1}}}, \quad (1)$$

where w_{t-1}, z_{t-1} and v_{t-1} are lagged variables, D_{yt}^c is a degenerate distribution at a category c that represents a refuge category for respondents that are unwilling or unable to give a meditated answer. At a given time t, the parameters $(1 - \xi_t)$ and $(1 - \pi_t)$ are interpreted as the feeling and uncertainty underlying the rating process. In particular, the uncertainty is related to the heterogeneity of the distribution [4] and, in this sense, it depicts the mutual disagreement among respondents. Finally, δ_t provides the proportion of respondents that selects the refuge category in excess of that given by the combination of the shifted binomial with the uniform distribution. When $\pi \to 1$ the previous formulation collapses to the simple mixture of a shifted binomial distribution with a degenerate distribution.

The estimation can be performed by the minimum chi-square method. This seems a natural choice when data arise in form of periodically observed frequency distributions. It is worth mentioning that the chi-square estimation criterion is a special case of more general divergence measures, such as the power divergence statistic [8] and the phi-divergence [12], for which interesting properties and developments have been derived.

3   Results

The data set consisted of the frequency distributions of consumers' opinions about inflation (perceptions and expectations), collected from January 2018 to March 2021 by ISTAT within joint harmonized European programme of business and consumer surveys. Opinions were coded using a 5-points scale (ranging from 1 = decreasing price trend to 5 = rapidly increasing trend). Note that in April 2020 the consumer and business confidence survey was suspended and resumed regularly in May. Moreover, from the end of February 2020, numerous containment measures were taken by Italian authorities to cope with the spread

of the pandemic. In particular, after the initial national lockdown (from 9th March to 4th May), Italy experienced a number of stop-and-go measures characterized by different level of restrictions in accordance with the development of Covid cases. During this period the year-over-year changes in the consumer price index declined remarkably due to the weak consumer demand and persistent economic slowdown. From May 2020 the rate of inflation was negative and still declining. It began to rise only at the end of the year (Fig. 1). At the same time, consumers perceptions were altered by the changed consumption basket [9] and shopping frequency [11], by the unfairness and steep changes of prices of particular goods [15]. Moreover, expectations were influenced by the increasing uncertainty about the evolution of the health crisis.
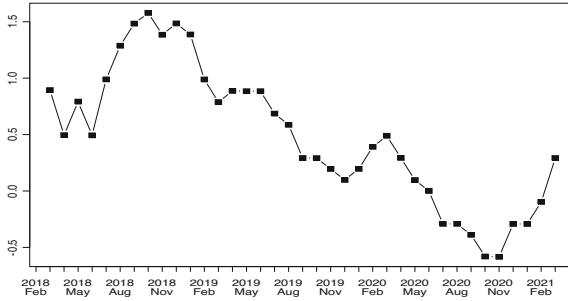
Fig. 1. Inflation rate

The frequency distributions of opinions observed over time present some common features. First, a prominent peak is located at the neutral category describing that the level of prices has remained (or is expected to remain) about the same. For this reason, both models elaborated for inflation perceptions and expectations include a refuge category $c = 2$. Secondly, a lower peak, located at the category indicating an increase (or an expected increase) in prices, is observed for most of the considered time points.

The *model of inflation expectations* is a mixture including the three components as described in (1). The parameter ξ_t depends on the mean of expectations at time $(t-1)$ (aveExp), the parameter π_t on the mean of perceptions at time $(t-1)$ (avePerc), and the parameter δ_t is function of the difference between the two mentioned means (aveExp−avePerc). Note that this difference provides a proxy of the expected change in inflation at the previous instant.

The *model of inflation perceptions*, instead, includes only two components: the shifted binomial distribution and the refuge component. In this case the time varying parameters, ξ_t and δ_t depends on avePerc and the difference (aveExp−avePerc) at time $(t-1)$, respectively.

The estimation results are illustrated in Table 1. Compared with the ignorance model (where each category is equally probable), the reduction of the chi-square discrepancy between the fitted and observed distributions [7], I_{fit},

is above 90% for both models. The scatter plots of the estimated uncertainty $(1 - \pi_t)$ versus the feeling $(1 - \xi_t)$ (left panel of Fig. 2) and versus the refuge category coefficient δ_t (right panel of Fig. 2) show the trajectories that those parameters followed during the months after the pandemic broke out.

Table 1. Estimation results (standard errors in parentheses)

Model	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\gamma}_0$	$\hat{\gamma}_1$	$\hat{\alpha}_0$	$\hat{\alpha}_1$	I_{fit}
Perceptions			1.157	-0.517	-0.847	0.214	0.93
			(0.019)	(0.006)	(0.010)	(0.008)	
Expectations	2.333	-0.370	-0.017	-0.518	-0.090	-0.897	0.98
	(0.228)	(0.077)	(0.754)	(0.255)	(0.008)	(0.051)	
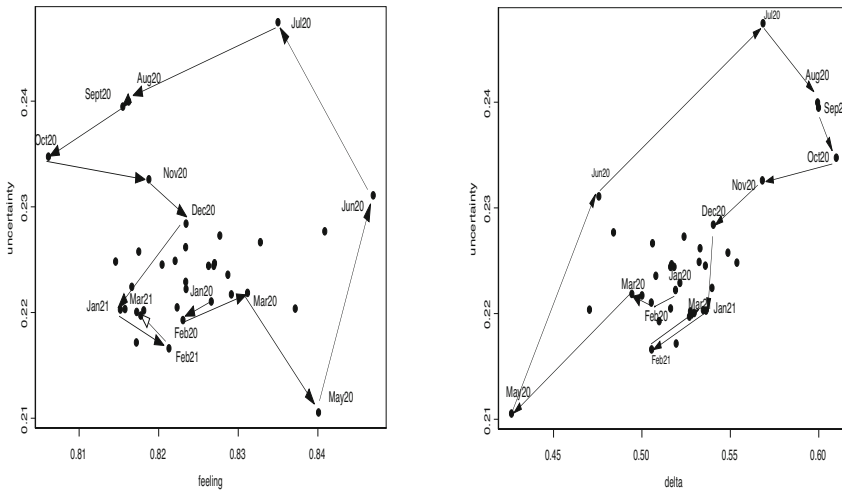
Fig. 2. Inflation Expectations: scatter plot of uncertainty $(1 - \pi_t)$ vs feeling $(1 - \xi_t)$ (left panel); uncertainty vs refuge category coefficient δ_t (right panel)

Moving from March to May 2020 (the lockdown period), consumers opinions were better defined (both the uncertainty and the weight of the neutral category decreased) and a greater number of respondents believed that inflation was going to rise (the feeling increased). As mentioned before, consumers based these opinions on a limited and biased set of information about prices. When the mobility restrictions started to be lifted, during the summer, consumer gained a wider view of the price developments. The mutual disagreement between opinions increased and the probability that respondents believed that inflation was rising began to decline (the feeling was smaller). At the end of September 2020 the contagious started to spread again, reaching the peak number of Covid cases

in November. However, as show by the trajectories in Fig. 2, the parameters followed a path that brought them closer to the bulk of data. In other words, the initial shock due to the Covid outbreak was completely absorbed. As regards to inflation perception, the scatter plot of the uncertainty and refuge components shows again a path that from March 2020 moves far from the majority of points (Fig. 3). Despite observed inflation declined until the end of 2020, the feeling increased and then the distributions of responses became more left skewed since a larger part of respondents believed that inflation was increasing. Only at the beginning of 2021 the dynamics of the time varying parameters went back to the pre-Covid pattern.
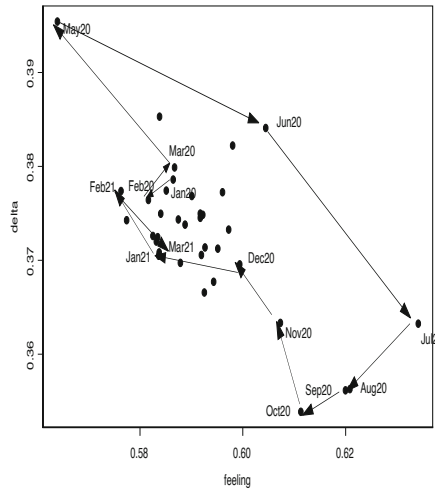
Fig. 3. Inflation perceptions: scatter plot of refuge category coefficient δ_t vs feeling $(1 - \xi_t)$

4   Conclusions

The model approach, presented in this article, represents a useful tool for the summary and interpretation of qualitative data from repeated surveys. The investigation of the evolution of the characterizing time-varying parameters helps to understand how the shape of the frequency distributions of respondents' opinions changes over time. The results concerning consumers' perceptions and expectations about inflation have confirmed that the pandemic outbreak had a significant but temporary effect on opinion formation. The end of lockdown, the progressive lift of containment measures and the start of new policies for the recovery of economic activities, helped consumers to expand the content of their consumption basket and increase the number of their price observations so that perceptions and expectation of inflation resulted less biased.

References

1. Armantier, O., et al.: How economic crises affect inflation belief: evidence from Covid-19 pandemic. J. Econ. Behav. Organ. **189**, 443–469 (2021)
2. Biancotti, C., Rosolia, A., Veronese, G., Kirchner, R., Mouriaux, F.: COVID-19 and Official Statistics: A Wakeup Call? Bank of Italy Occasional Paper, no. 605 (2021)
3. Binder, C.: Coronavirus fears and macroeconomic expectations. Rev. Econ. Stat. **102**, 721–730 (2020)
4. Capecchi, S., Iannario, M.: Gini heterogeneity index for detecting uncertainty in ordinal data surveys. METRON **74**, 223–232 (2016)
5. Cavallo, A.: Inflation with Covid Consumption Baskets. NBER Working paper 27352 (2020)
6. Coleman, W., Nautz, D.: Inflation expectations, inflation target credibility and the COVID-19 pandemic: New evidence from Germany. Center for Financial Studies Working Paper 658 (2021)
7. Corduas, M.: A dynamic model for ordinal time series: an application to consumers' perceptions of inflation. In: Balzano, S., Porzio, G.C., Salvatore, R., Vistocco, D., Vichi, M. (eds.) Statistical Learning and Modeling in Data Analysis, pp. 37–45. Springer, Cham (2019)
8. Cressie, N., Read, T.R.: Multinomial goodness-of-fit tests. J. R. Stat. Soc. B **46**, 440–464 (1984)
9. Georganas, S., Healy, P.J., Li, N.: Frequency bias in consumers perceptions of inflation: an experimental study. Eur. Econ. Rev. **67**, 144–158 (2014)
10. Gomes, S., Iskrev, N., Pires Ribeiro, P.: Euro area inflation expectations during the COVID-19 pandemic. Econ. Bull. Banco de Portugal (2021). https://www.bportugal.pt
11. Huber, O.W.: Frequency of price increases and perceived inflation: an experimental investigation. J. Econ. Psych. **32**, 651–661 (2011)
12. Pardo, L.: Statistical Inference Based on Divergence Measures. Chapmann & Hall/CRC Press, Boca Raton (2006)
13. Piccolo, D.: On the moments of a mixture of uniform and shifted binomial random variables. Quad. Stat. **5**, 85–104 (2003)
14. Piccolo, D., Simone, R.: The class of CUB models: statistical foundations, inferential issues and empirical evidence. Stat. Meth. Appl. **28**, 389–435 (2019)
15. Ranyard, R., Del Missier, F., Bonini, N., Duxbury, D., Summers, B.: Perceptions and expectations of price changes and inflation: a review and conceptual framework. J. Econ. Psychol. **29**, 378–400 (2008)
16. van der Wielen, W., Barrios, S.: Economic sentiment during the COVID pandemic: evidence from search behaviour in the EU. J. Econ. Bus. **115** (2021). https://doi.org/10.1016/j.jeconbus.2020.105970

A Proposal to Calculate the Regulatory Capital Requirements for Reverse Mortgages

Iván de la Fuente(✉) , Eliseo Navarro , and Gregorio Serna

University of Alcalá, Pl. Victoria 2, Alcalá de Henares, 28802 Madrid, Spain
ivan.fuente@uah.es

Abstract. In this paper, we propose a new method for estimating the regulatory capital requirements for a portfolio of income stream reverse mortgages owned by a financial institution, according to Basel II and III. The method considers house price risk, mortality risk and interest rate risk and regulatory capital requirements need to be computed using a Monte Carlo simulation procedure. Several scenarios for the reverse mortgage specifications are considered, including fixed or variable mortgage rates and different income stream schemes (with the lump sum as a particular case). The results for the U.K. show that the reverse mortgage provider faces higher risk in the lump-sum case, for relatively young borrowers and for the female population. Furthermore, the lender's risk grows with the percentage of the loan amount that the borrower receives on the initial date, with the lump sum (100% of the loan amount on the initial date) being the riskiest case. The lender's risk is also higher with fixed mortgage roll-up rates than with floating rates.

Keywords: Reverse mortgages · Option pricing · Mortality modeling · House price modeling · Interest rate risk · Regulatory capital requirements

JEL Codes: G21 · G22 · J14 · R3

1 Introduction

In the current context in which many people worry about the sustainability of pension systems, reverse mortgages are gaining popularity because they are a way of supplement elderly people's income. In a reverse mortgage the owner of a property receives from the lender a certain amount of money, either in form of a single payment (lump-sum solutions) or in form of several payments (income stream solutions). The loan is repaid with the sale of the borrower's property at the date of his/her death or when he/she moves to another home, with the particularity that the borrower's debt cannot be greater than the amount of

---

I. de la Fuente—PhD program in Economics and Business Management.

the sale of the house. This clause is commonly known as the non-negative-equity guarantee (NNEG), which is of course a source of risk for the lender. This NNEG can be viewed as a European put option owned by the borrower, so that he/she will exercise the option whenever on the date of his/her death the sale amount of the house is lower than the redemption amount of the mortgage.

During recent years there have been many papers addressing the problem of valuing reverse mortgages ([1,2,4,5,8,9], among others). However, less attention has been paid to the measurement and management of risk in reverse mortgages. Although in many papers ([1,3]) it is estimated the risk faced by the lender by means of computing the value at risk (VaR) or the expected shortfall (ES) of a single contract, few attention has been given to the calculation of regulatory capital requirements in practice for a financial institution owing a portfolio of reverse mortgages. [6] and [7] propose a method for computing regulatory capital requirements for a theoretical portfolio of lump-sum reverse mortgages owned by a financial institution, allowing for house price and mortality risks. However, no articles have been found addressing this problem in the case of income stream solutions.

In this paper we fill this gap by proposing a new and realistic method for calculating regulatory capital requirements for a portfolio of income stream reverse mortgages according to Basel II and III, allowing for house price risk, interest rate risk and mortality risk. Furthermore, the proposed method for income stream solutions includes the lump-sum solution as a particular case, which allows us to compare both types of reverse mortgages from the point of view of the risk faced by the lender.

2   Modeling House Price Risk, Interest Rate Risk and Mortality Rate Dynamics

As stated in the Introduction, three sources of risk will be considered when estimating the regulatory capital requirements for reverse mortgages: house price risk, interest rate risk and mortality risk.

The house price data source is the Nationwide's House Price Index from the U.K. (quarterly data from 1952 to 2020, i.e., 273 observations). The proposed model for the house price index (HPI) is the ARMA-EGARCH model, according to [8] and [7].

The interest rate data set is composed of the quarterly series of the 10-year zero-coupon government bond rate in the U.K. from 1970 to 2020, whose dynamics is modelled through the Vasicek model.

Finally, the mortality data comes from the Human Mortality Database (U.K. data from 1952 to 2018). The probabilities of survival and death, which are necessary in the regulatory capital requirements estimation process, have been estimated through the Lee-Carter model, that is the most used model in the actuarial and demographic literature.

Once the three sources of risk are modelled, the next step is to estimate the NNEG that is necessary to estimate the regulatory capital requirements for reverse mortgages. Under the assumptions above for the sources of risk, the

price at time t of a European option maturing at time T, V_t, under the equivalent risk-neutral probability measure (Q), can be computed as:

$$V_t = E_Q \left[e^{-r(T-i)} \cdot V_T \mid \Phi_t \right] \tag{1}$$

where V_T is the option's payoff at expiration and Φ_t is a set containing all market information available at time t.

Given that the date of death of the borrower is unknown, the value of the NNEG, $V_N NEG$, will be estimated as a weighted average of several European put options values, each of them maturing at each of the quarters from now till the maximum lifetime considered for a person, ω:

$$V_{NNEG} = \sum_{k=0}^{(\omega-x)\cdot 4 - 1} {}_k p_x q_{x+k} P(k + \tfrac{1}{2} + \delta, S, X, u_\kappa, r_\kappa, g) \tag{2}$$

where ${}_k p_x$ is the probability that a person aged x at inception survives to age $x + k$ and q_{x+k} is the probability that a person aged x at inception dies during the interval k to $k + 1$, with $k = 0, 1, \ldots, \omega$, and $P(k + \tfrac{1}{2} + \delta, S, X, u_\kappa, r_\kappa, g)$ is the European put option value, r_κ is the risk-free interest rate, u_κ is the floating mortgage roll-up rate (the short rate plus a certain spread) and g is the rental yield rate. In the calculation of this op-tion value, it is assumed that all deaths occur at mid-quarter and that there is a delay of six months (δ) from the home exit until the sale of the property.

Let us assume that S_0 is the current price of the house, LTV is the loan-to-value ratio, and X_κ is the value at the end of quarter κ of all quantities advanced by the lender until this moment. On the initial date, time 0, the borrower receives a percentage, α, of the loan amount: $\alpha \cdot LTV \cdot S_0$. The rest is received in annuities over j years, with payments at the end of each year: $\frac{(1-\alpha) \cdot LTV \cdot S_0}{j}$. Thus, the value at the end of quarter κ of all quantities advanced by the lender until κ, X_κ, can be computed as:

$$X_\kappa = \begin{cases} \alpha \cdot LTV \cdot S_0 \cdot e^{\frac{u_0}{4}} & \text{if } \kappa = 1 \\ X_{\kappa-1} \cdot e^{\frac{u_{\kappa-1}}{4}} & \text{if } \kappa \neq 4 \cdot j, \forall j = 1, 2, \ldots, \omega \\ \left(X_{\kappa-1} \cdot e^{\frac{u_{\kappa-1}}{4}} + \frac{(1-\alpha) \cdot LTV \cdot S_0}{\varphi} \right) & \text{if } \kappa = 4 \cdot j, \forall j = 1, 2, \ldots, \omega \end{cases} \tag{3}$$

Therefore, the put option value in expression 2, $P(k + \tfrac{1}{2} + \delta, S, X, u_\kappa, r_\kappa, g)$, can be calculated as:

$$e^{\left[-\sum_{t=1}^{(\omega-k)+\frac{1}{2}+\delta} \frac{r_t}{4} \right)} \cdot E_Q \left[\left(X_{\kappa+\frac{1}{2}} - S_{k+\frac{1}{2}+\delta} \right)^+ \right] \tag{4}$$

3   Calculation of Regulatory Capital Requirements

Let us consider a portfolio of income stream reverse mortgages owned by a certain financial institution. The portfolio is composed of $N_0 = 1{,}000$ people (all of them

men or women) aged 70, 80 or 90 years. The calculation of one-year or VaR or ES involves the simulation of (i) the house price in one year, (ii) the interest rate in one year and (iii) the number of survivors in one year. The house price and the interest rate in one year are simulated by means of the ARMA-EGARCH and the Vasicek processes respectively, taking into account the correlation among them. Finally, to simulate the number of survivors in one year it is assumed that the distribution of the number of deaths is a binomial $B(N_0, q_x)$, with q_x being the probability of death for the considered population.

Under these assumptions the value at the beginning of the year of the portfolio of income stream reverse mortgages from the lender's perspective is:

$$N_0 \cdot (Y_0 - NNEG_0) \tag{5}$$

Where Y_0 is the present actuarial value of all amounts received by the lender until his or her death and NNEG0 is the value of the no-negative-equity guarantee estimated at the beginning of the year, assuming and initial house price of $S_0 = 150,000$ pounds. Once we have simulated the number of survivors and the house price in one year, (N_1, S_1), together with the forward curve for the interest rate, we can estimate the value of the no-negative-equity guarantee in one year's time, $(NNEG_1)$. Under these specific simulated values for (N_1, S_1) and the forward curve for the interest rate, the final value of the portfolio of income stream reverse mortgages is:

$$N_1 \cdot (Y_1 - NNEG_1) + (N_0 - N_1) \cdot [Y_1' - max(Y_1' - S_1, 0)] \tag{6}$$

The second term in expression 6 accounts for the intermediate cash flows. Moreover, in expression 6 Y_1' represents the sum of all quantities received by the borrower during the year plus interest, whereas Y_1 is equal to Y_1' plus the present actuarial value at the end of the year all amounts received by the lender until his or her death. The different between Y_1 and Y_1' is due to the fact that the debt corresponding to the survivors is estimated considering all future quantities that the borrowers will receive until their dead, whereas the debt corresponding to the lenders who died during the year is limited to the amounts received during this year plus interest.

To obtain a distribution for the value of the portfolio of income stream reverser mortgages at the end of the year we repeat this process 10,000 times, so we can estimate the VaR as the 99.9th percentile of the distribution. The final step is to calculate the potential loss suffered by the lender during the year as the present value of the VaR after one year minus its initial value. In the results presented below the losses are expressed as a percentage of the initial portfolio value. Moreover, different values have been considered for the percentage (α) of the loan amount that the borrower receives on the initial date ($\alpha \cdot LTV \cdot S_0$). These values are 20%, 40%, 60% and 100%, where 100% corresponds to the lump sum case.

Tables 1 and 2 show the results for the one-year 99.9% VaR with fixed and floating roll-up rates respectively. From Tables 1 and 2 it can be seen that the risk of a floating rate of $r + 0\%$ is approximately equivalent to a fixed rate of

Table 1. One-year 99.9% VaR (%), fixed roll-up rate (u)

	Male			Female		
u/Age	70	80	90	70	80	90
1.50%	−2.85%	−1.88%	−0.13%	−3.31%	−2.33%	−0.19%
2.00%	−4.19%	−2.75%	−0.25%	−4.80%	−3.30%	−0.36%
2.50%	−6.11%	−3.65%	−0.44%	−6.86%	−4.39%	−0.60%
3.00%	−8.56%	−4.92%	−0.75%	−9.78%	−5.85%	−0.96%
3.50%	−11.79%	−6.35%	−1.07%	−13.62%	−7.30%	−1.34%
4.00%	−16.47%	−7.98%	−1.52%	−19.69%	−9.01%	−1.87%
4.50%	−22.99%	−9.58%	−2.04%	−33.57%	−10.88%	−2.45%
5.00%	−37.77%	−11.59%	−2.56%	−105.7%	−13.04%	−3.14%

Table 2. One-year 99.9% VaR (%), floating roll-up rate ($u = r + s$)

	Male			Female		
s/Age	70	80	90	70	80	90
0.00%	−9.32%	−3.67%	−1.81%	−12.53%	−4.51%	−1.82%
0.50%	−12.19%	−4.83%	−2.17%	−15.91%	−5.95%	−2.20%
1.00%	−14.86%	−6.04%	−2.57%	−18.97%	−7.28%	−2.60%
1.50%	−17.23%	−7.19%	−2.97%	−21.66%	−8.54%	−3.03%
2.00%	−19.58%	−8.32%	−3.40%	−24.30%	−9.59%	−3.52%
2.50%	−22.01%	−9.18%	−3.92%	−27.73%	−10.52%	−4.06%
3.00%	−26.00%	−10.28%	−4.39%	−34.88%	−11.44%	−4.63%
3.50%	−31.88%	−11.14%	−4.89%	−53.18%	−12.23%	−5.18%

2.5%. From there the risk grows faster in the floating case than in the fixed case. Furthermore, the risk is higher for relatively young borrowers and for the female population.

Tables 3 and 4 show the results of the comparison of the risk assumed by the lender for different values of the percentage, α, of the loan amount that the borrower receives on the initial date, with $\alpha = 100\%$ being the lump sum case. With fixed roll-up rates (Table 3), the risk is higher for higher α levels. However, with floating rates (Table 4), the evidence is mixed; for relatively young borrowers (70 years old) the risk is higher for lower values of α, however for older borrowers (80 or 90 years old) the risk is higher for higher values of α. This is because with downward interest rate dynamics the interest loss with floating rates may be higher for younger borrowers than with fixed rates.

Table 3. One-year 99.9% VaR (%), fixed roll-up rate $u = 3\%$

α/Age	Male			Female		
	70	80	90	70	80	90
100%	−8.56%	−4.92%	−0.75%	−9.78%	−5.85%	−0.96%
80%	−8.00%	−4.63%	−0.85%	−9.08%	−5.42%	−0.99%
60%	−7.44%	−4.43%	−1.14%	−8.56%	−5.19%	−1.20%
40%	−6.96%	−4.38%	−1.68%	−8.02%	−4.85%	−1.78%
20%	−6.67%	−4.24%	−2.50%	−7.40%	−4.87%	−2.33%

Table 4. One-year 99.9% VaR (%), floating roll-up rate $s = 1.5\%$

α/Age	Male			Female		
	70	80	90	70	80	90
100%	−17.23%	−7.19%	−2.97%	−21.66%	−8.54%	−3.03%
80%	−17.49%	−5.72%	−1.15%	−22.73%	−7.07%	−1.09%
60%	−17.47%	−3.71%	0.00%	−23.61%	−5.27%	0.00%
40%	−17.61%	−1.23%	0.00%	−25.40%	−2.78%	0.00%
20%	−17.87%	0.00%	0.00%	−28.31%	0.00%	0.00%

References

1. Alai, D.H., Chen, H., Cho, D., Hanewald, K., Sherris, M.: Devoloping equity release markets: risk analysis for reverse mortgages and home reversions. North Am. Actuar. J. **18**(1), 217–241 (2014). https://doi.org/10.1080/10920277.2014.882252
2. Chang, C., Wang, C., Yang, C.: The effects of macroeconomic factors on pricing mortgage insurance contracts. J. Risk Insur. **79**, 867–895 (2012). https://doi.org/10.1111/j.1539-6975.2011.01447.x
3. Cho, D., Hanewald, K., Sherris, M.: Risk analysis for reverse mortgages with different payout designs. Asia Pac. J. Risk Insur. **9**(1), 77–105 (2015). https://doi.org/10.1515/apjri-2014-0012
4. Di Lorenzo, E., Piscopo, G., Sibillo, M., Tizzano, R.: Reverse mortgages through artificial intelligence: new opportunities for the actuaries. Decis. Econ. Finance **44**, 23–35 (2021). https://doi.org/10.1007/s10203-020-00274-y
5. Di Lorenzo, E., Piscopo, G., Sibillo, M., Tizzano, R.: Reverse mortgage and risk profile awareness: proposals for securitization. Appl. Stochast. Models Bus. Ind. 1–17 (2021b). https://doi.org/10.1002/asmb.2664
6. Fuente, I., Navarro, E., Serna, G.: Reverse mortgage risks. time evolution of VaR in Lump-sum solutions. Mathematics **8**, 2043 (2020). https://doi.org/10.3390/math8112043
7. Fuente, I., Navarro, E., Serna, G.: Estimating regulatory capital requirements for reverse mortgages. an international comparison. Int. Rev. Econ. Finance **74**, 239 −252 (2021). https://doi.org/10.1016/j.iref.2021.03.001

8. Li, J.S.H., Hardy, M., Tan, K.S.: On pricing and hedging the no-negative-equity guarantee in equity release mechanisms. J. Risk Insur. **77**, 499–522 (2010). https://doi.org/10.1111/j.1539-6975.2009.01344.x
9. Shao, A.W., Hanewald, K., Sherris, M.: Reverse mortgage pricing and risk analysis allowing for idiosyncratic house price risk and longevity risk. Insur. Math. Econ. **63**, 76–90 (2015). https://doi.org/10.1016/j.insmatheco.2015.03.026

LTC of a Defined Benefit Employee Pension Scheme

J. Iñaki De La Peña[1,3,4(✉)], M. Cristina Fernández-Ramos[2], Asier Garayeta[1,3], and Iratxe D. Martín[1]

[1] Faculty of Economics and Business, University of the Basque Country, Lehendakari Agirre Avenue, 83, 48-015 Bilbao, Spain
jinaki.delapena@ehu.es
[2] Junta de Castilla y León - School of Education, Valladolid, Spain
[3] Consolidated Research Group EJ/GV: IT 897-16, Bilbao, Spain
[4] Polibienestar Research Institute, Valencia, Spain

Abstract. Private pension schemes focus on retirement savings. Perceptions of health status change over time and, as retirement age approaches, concerns about Long-Term Care grow. However, once near retirement age, there isn't enough time to plan sufficiently in advance. This paper proposes a mechanism to transform the private pension of a Defined Benefit scheme (retirement, invalidity) into an allowance. In turn, should the need arise on becoming dependant, this allowance will pay for any Long-Term Care services the beneficiary might require. Depending on the pensioner's situation, both the expected number of payments and their intensity are transformed. For this purpose, a mechanism is defined, through a multiple state Markov model, to adapt the amount of the pension to the revised life expectancy of the beneficiary. The revised life expectancy would be derived from his/her new health status. The main contribution of this work is to establish a private Defined Benefit pension scheme model capable of transforming its benefits, adding Long-Term Care support, without increasing the total pension cost to the scheme.

Keywords: Ageing · Dependence · Long-Term Care · Private pension

1 Introduction

In Defined Benefit (DB) pension schemes, an individual has no information about his/her future health status. Therefore, the individual has a contribution pattern that is independent of his/her future health status, as it depends solely on the career path of the individual. Moreover, any additional information about the true health status that arises over time, does not really affect the benefit an individual receives on retirement [1–3].

Long-Term Care (LTC) coverage is, therefore, a logical extension to a pension schemes purpose, it being: to provide an adequate financial complement to meet retirement needs, regardless of the individual's health status. Therefore, the design of the pension scheme must take into account the possible needs of LTC [4]; on the one hand,

providing an income to make up for the lack of salary and, on the other, as soon as LTC needs arise, replacing the income with a supplement to help pay for them. Hence, when the bulk of the *baby boom generation reaches* an age when LTC is needed, i.e. as the population needs and services demanded change with age, there will be resources to meet them [5].

The aim of this work is to establish a financial-actuarial model that makes it possible to transform the pension element, part of the private economic resources that individuals may have, into a supplement that helps to pay for LTC needs. This transformation would take place at the request of the beneficiary and would do so without increasing the pension's overall cost. This model provides the private pension scheme with a social vision by means of adapting its coverages to the pensioner's need for resources, whether or not, he/she is the beneficiary of a retirement or disability pension.

The second section, in order to meet this objective, considers an actuarial valuation model under different mortality rates in line with the scheme member's status. This defines an actuarial mortality correction factor to be applied to the pension's valuation depending on the beneficiary's condition: retired or invalid. The key to this lies with the difference in life expectancy according to the beneficiary's status i.e. general population, disabled or severely dependant. The third section shows the results of this correction factor applied to the Spanish mortality experience, both for generic, disabled and dependant population. The final sections present conclusions and future lines of work.

The main contribution of this work is to establish a private Defined Benefit pension scheme model capable of transforming its benefits, adding Long-Term Care support, without increasing the total pension cost to the scheme. The model allows for the adaptation of the expenditure to the reality experience by the individual. The quality of life of the dependant is improved by having part of the LTC costs covered. As indicated by [6], the dependant can live longer by improving his/her functional environment.

2 The Model

This paper develops the model initially proposed by [7–10] by incorporating it into a pension scheme and deriving the supplement that meets the needs for LTC due to higher degrees of dependency. In these cases, the dependant needs assistance to perform several basic daily activities several times a day. Due to loss of physical, mental, intellectual or sensory autonomy, the individual needs extensive support for personal autonomy. This assistance carries extra expenses. So, the aim of the transformation is to provide resources for the new expenses.

The pension is automatically increased to provide additional resources to help pay LTC costs when the pension beneficiary becomes severely dependant and requests it. Thus, there is a transfer of the value of the pension to the LTC assistance. Then, at an age $x > r$, such that the pensioner is dependant and decides to transform the pension, an equivalence is fulfilled;

$$PVFB_x = PVFLTC_x \tag{1}$$

$PVFB_x$ present value of future benefits valued at age x, such that $x > r$.

$PVFLTC_x$ present value of future benefits, including the new LTC allowance, both valued at the time of the decision, or x, such that $x > r$.

This gives the transformation factor at each age (λ_x^d). This is subject to the sum of the residual probabilities of survival in future years according to the beneficiary's status, i.e. the life expectancy according to the status (general or dependant), financially discounted to the expected return of the pension fund.

$$\lambda_x^d = \frac{\int_x^w e^{-\int_t^{t+1} \mu_t dt} \cdot e^{-\int_r^w \delta(t)dt} \cdot dt}{\int_x^w e^{-\int_t^{t+1} \mu_t^d dt} \cdot e^{-\int_r^w \delta(t)dt} \cdot dt} = \frac{\overline{a}_x^m}{{}^d\overline{a}_x^m} \tag{2}$$

$e^{-\int_t^{t+1} \mu_t^d dt}$ probability of survival of a dependant person of age t to live to age $t + 1$ as a dependant.

$e^{-\int_t^{t+1} \mu_t dt}$ probability of survival of a person of age t to live to age $t + 1$ (general mortality).

$e^{-\int_r^w \delta(t)dt}$ financial discounting factor from time t. The financial discounting function is defined by the force of interest rate $\delta(t)$ as standard in the private pension scheme.

$\overline{a}_x^m$ Continuous life annuity of a person at age x. Depending on whether or not it is indexed to an external benchmark, it can be variable or constant.

${}^d\overline{a}_x^m$ Continuous life annuity of a dependant person at age x. Depending on whether or not it is indexed to an external benchmark, it can be variable or constant.

The resulting LTC complement depends on:

- The age of decision making,
- The expected mortality of the cohort,
- The pension scheme's performance,
- The expected mortality of the dependant and
- The level of pension that the beneficiary is receiving.

Except for the expected mortality of the dependant, all other factors are standard in the design of a private pension scheme.

3   A Sample for Spain

3.1   Mortality Tables by State

In Spain, the private sector provides coverage for the highest levels of dependency almost exclusively [10], through either insurance or pension schemes [12]. To illustrate the influence of this factor, the following tables have been applied:

- For general population, the PERM/F for the year 2000, with special reference to the year 2008 [13] (different life tables for men and women).

- For invalid population, the Spanish Social Security actuarial tables for pensioners receiving an invalidity annuity [14]. This table provides information on the entire population with permanent disability, whether or not they are dependant. It is compiled for ages 16 to 108 years, both inclusive. Based on the census of the Spanish population in 2008, a separation by gender is made [15].
- For dependant mortality, those elaborated by [16], based on general mortality tables [13] and adjusted to French statistics, HID 98-01.

3.2   Results

There is a difference in mortality by gender in all age groups and for the different states. Figure 1 shows that the percentage of over-mortality in males (a) is much higher than in females (b). There is a gradual decrease in the excess mortality differential in all states. Towards the end of the estimated life expectancy, the values of the mortality ratios are almost equal showing that, excess mortality tends to decrease.
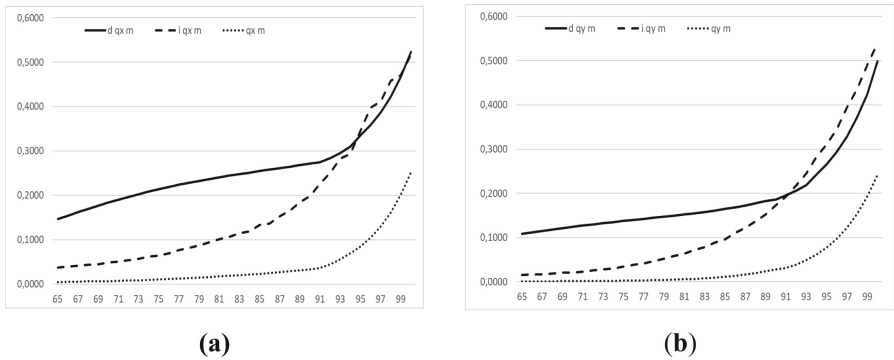
Fig. 1. Mortality differential by status and gender after retirement age (65 years). (a) Men, (b) Women. Source: Own elaboration. Age x: man. Age y: woman. $d\ qx\ m$: Dependant mortality rate at age x. $i\ qx\ m$: Invalid mortality rate at age x. $qx\ m$: General mortality rate at age x.

Analysing the last years of life (Fig. 2) the convergence of the mortality tables is greater. This shows that in the end mortality does not depend on the state at which one arrives, since towards the end of the estimated life expectancy everyone has a similar mortality.

The resulting equity/fairness factor (LTC complement percentage for a given age and gender) is much higher if it is generated in the retirement state rather than in the disability state (Fig. 3) leading to high LTC benefits. However, the longer the delay in causing severe dependency, the lower the value of the benefit. This is because of the logical confluence of the mortalities of the different states.
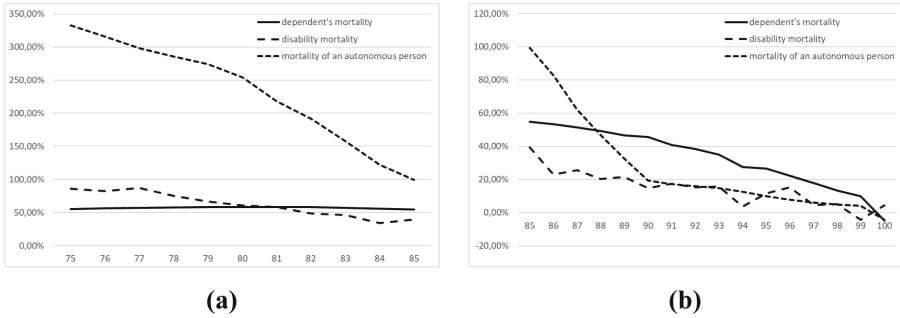
Fig. 2. Existing excess mortality in each state by gender and age (a) from 75 to 85 years; (b) from 85 to 100 years. Source: Own elaboration
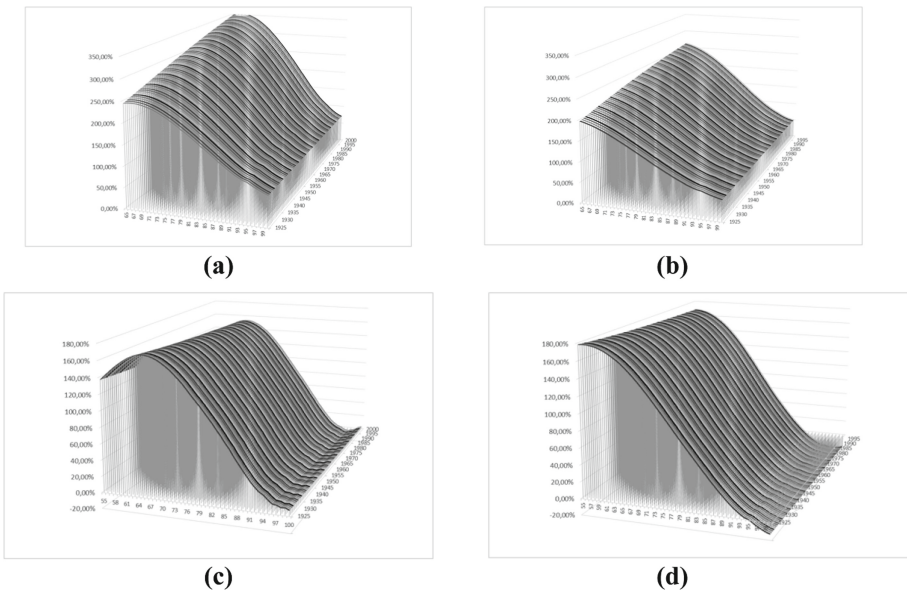
Fig. 3. Percentage of LTC supplement by age and gender. (a) Retirement men, (b) Retirement women, (c) Disability men, (d) Disability women Source: Own elaboration.

4   Conclusions

The actuarial model has been designed to be implemented without much difficulty and at no cost to the pension scheme. Enabling the pension coverage to be extended in DB and other privately funded pension schemes. At present, these privately funded pension schemes are designed assuming a general mortality rate for the insured, but not a specific one in the case of high or severe dependency.

The consideration of both the equity/fairness factor together with the use of specific mortality tables for dependants, would enable the scheme to transform its benefits, adding

LTC support, without increasing the overall pension cost to the scheme. Hence, without the need of any further contributions to be made.

What is adjusted is the expectation of payment, increasing the total pension (pension plus economic support).

Future work would entail the design of a Defined Contribution (DC) pension scheme that would include LTC coverage, although it would be in competition with other insurance products. In a DC scheme, the individual is aware of his/her health situation from year to year and, unlike in a DB scheme, can supplement the coverage with dependency insurance products.

References

1. Barr, N.: Pensions: overview of the issues. Oxf. Rev. Econ. Policy **22**(1), 1–14 (2006)
2. Finkelstein, A., Poterba, J., Rothschild, C.: Redistribution by insurance market regulation: Analysing a ban on gender-based retirement annuities. J. Financ. Econ. **91**, 38–58 (2009)
3. Rothschild, C.: Non-exclusivity, Linear Pricing, and annuity market screening. J. Risk Ins. **82**(1), 1–32 (2015)
4. Kenny, T., et al.: The future of social care funding: who pays? British Actuarial **22**(1), 10–44 (2017)
5. Hurd, M.D., Michaud, P.C., Rohwedder, S.: The lifetime risk of nursing home use. discoveries in the economics of aging. In: Chicago: National Bureau of Economic Research, pp 81–109 USA (2014)
6. Keeler, E., Guralnik, J.M., Tian, H., Wallace, R.B., Reuben, D.B.: The impact of functional status on life expectancy in older persons. J. Gerontol. Series A **65**(7), 727–733 (2015)
7. Fernández-Ramos, M.C.: Pragmatic solutions in the private sector for the coverage of dependency in Spain Thesis. In: University of the Basque Country. Spain, p. 336 (2015)
8. Fernández-Ramos, M.C., De La Peña, J.I.: Influence of dependency coverage in pension schemes. Anales de Asepuma **23**, 403 (2015)
9. De la Peña, J.I., Fernández-Ramos, M.C., Herrera, A.T., Peña-Miguel, N.: Measures of actuarial balance in Spanish social security: Back to the past. Anales del Instituto de Actuarios Españoles **23**, 129–143 (2017)
10. De La Peña, J.I., Fernández-Ramos, M.C., Garayeta, A.: Cost-free LTC model incorporated into private pension schemes. Int. J. Environ. Res. Public Health **18**, 2268 (2021)
11. Fernández-Ramos, M.C., De La Peña, J.I.: Legislative development of protection for dependence. Opportunities for the private sector: The case of the Castilla and Leon region, Spain Revista de Estudios Regionales **97**, 113–136 (2013)
12. Fernández-Ramos, M.C., De La Peña, J.I., Herrera, A.T., Iturricastillo, I., Peña-Miguel, N.: Helping long term care coverage via differential on mortality? In: Corazza, M., Durbán, M., Grané, A., Perna, C., Sibillo, M. (eds.) Mathematical and Statistical Methods for Actuarial Sciences and Finance, pp. 345–349. Springer, Cham (2018). https://doi.org/10.1007/978-3-319-89824-7_62
13. Resolution BOE-A-18295, of 3 October 2000, of the Directorate General of Insurance and Pension Funds, which complies with the provisions of number 5 of the second transitory provision of the Regulations for the Organisation and Supervision of Private Insurance, approved by Royal Decree 2486/1998, of 20 November, in relation to the mortality and survival tables to be used by insurance companies. BOE, 244: 34882–34895. Madrid. Ministry of Economy. (2000)

14. Order TAS/4054/2005, of 27 December, by which the technical criteria for the settlement of capital cost of pensions and other periodic Social Security benefits are developed. BOE, 310. Madrid: Ministry of Labour and Social Affairs
15. INE. Series of Population in Spain (2008). https://www.ine.es/jaxi/Tabla.htm?path=/t20/e245/p08/l0/&file=02003.px&L=0
16. Sánchez, E., de López, J.M., Paz, S.: The correction of dependant mortality rates: an application to the Spanish case. Anales del Instituto de Actuarios Españoles **13**, 135–151 (2008)

Socio-Economic Challenges at the Time of COVID-19: The Proactive Role of the Insurance Industry

Emilia Di Lorenzo[1], Elisabetta Scognamiglio[2], Marilena Sibillo[3(✉)], and Roberto Tizzano[1]

[1] University of Naples Federico II, Naples, Italy
{diloremi,roberto.tizzano}@unina.it
[2] Italiacamp srl, Rome, Italy
e.scognamiglio@italiacamp.com
[3] University of Salerno, Fisciano, Italy
msibillo@unisa.it

Abstract. In the paper we present a securitized bond set in the macrocosm of sustainable finance. The product is generated by a securitization scheme that involves three actors: firms, issuing insurance contracts to protect the health of workers; insurers; investors, betting on the good health management of workers. The product is a derivative linked to an underlying representing the trend of the pandemic, specifically the present COVID-19 number of infected persons compared to its expected value. Aim of the paper is to frame the product in specific categories recognized within the ESG criteria. In particular, two specific categories are identified within the Principles Responsible Investments, which integrate the bond within the ESG products. Moreover, from an insurance point of view, the product can be considered within the most innovative asset categorization for impact finance, recent for the insurance sector. This area encompasses a limited number of recently developed tools aimed to sustainability and impact.

Keywords: Impact investing · Sustainability · Securitization · Pandemic bond · Social health insurance products

1 Introduction

The worldwide spread of COVID-19 contagion has been determining long-term systemic consequences in economy, lifestyles and social cohesion, whose perspectives appear strongly related to the evolution of the health risk: innovative and resilient tools and roles are needed.

The role of health insurance entities – private and public (such as those bearing a variable part of the costs of public health, most common in European Countries) – is more than ever in protecting people and productive activities; those who are responsible for large numbers of people (cf. [3, 5]) must cope with new sustainability challenges.

Before the impact of the pandemic, the 15th edition of the Global Risk Report of the World Economic Forum addressed public health global risks through sustainability based strategic and financial resilience tools taking into account environment, climate change, energy innovation; the pandemic is a new, dramatic variable that adds to the above and requires innovative economic and financial synergies to manage health risk, to protect collective health (cf. [3]), to preserve business entity going concern, occupation and economy: differently from other health threats, the pandemic drives up the health care costs for those who are affected by the disease as well as impacting entities turnover and income as consequences of quarantines, isolations of people and lock-downs.

Considering occupational safety as a social value, we develop (cf. [3]) an insurance product that combines both the aspects of risk-sharing and incentives for virtuous behavior, according to a structure oriented to the pursuit of objectives linked to a wide perspective of common good and refer broadly to insurance contracts aiming to cover the risks of damage to workers' health, with particular regard to the risk of contagion, in which the firm is the contractor, and the insured persons are the workers. The coverage of the risk arising from the contagion is implemented through a sharing system such as securitization (cf. [3, 6, 8–10, 12]) which involves a third figure, that is the investor/underwriter of the derivative issued with the product itself.

The pursuit of the common objective of improving health is achieved through incentives granted to insured firms that develop virtuous behavior aiming at high health safety standards. The incentive may consist in the reduction of the premium at the contract renewal, to the extent of an appropriate percentage that becomes a decision variable of the scheme.

The risk arising from the contagion, particularly in the presence of pandemics, is shared through the issuance of health securities, that is health bonds whose underlying is an appropriate and specific parameter representative of the general health status of the reference community (cf. [11]). The cat bond will originate a cash flow depending on the performance of the underlying that, if it is positive, provides a percentage of bonuses to be awarded to the company insured virtuous in the renewal of the contract for the following period.

The product involves actors joined by the common goal of improving or safeguarding the general state of public health, all of whom benefit by operating into a single virtuous cycle. The first one is the insured company, which is awarded the bonus; the second is the insurer, gaining considerable reputational advantages (even political ones, in cases of public entities of agencies), increasingly recognized because of activities geared to the common good; the third, that is the investor/underwriter, who invests in the quality of life by gaining personal and community benefits in the broadest sense.

A fundamental aspect of the product is the derivative issued on the underlying connected to the state of public health. This is the use of a newly developed derivative, which is indeed in the broad category of derivatives related to insurance business. These products have diversified widely over time, starting with the securitization of risks related to unforeseen changes in the development of human survival (through mortality/longevity-linked securities that transfer these risks to the capital market, cf. [2]), up to specific categories of securities implemented in sustainable finance contexts.

2   Sustainability and Impact: A Possible Conjugation

2.1   The Guidelines of the Scheme

Let us consider insurance contracts signed by firms to cover the employees/workers from possible health damage due to COVID-19 infection. The contract is characterized by an incentive consisting into a reduction of future premiums if the general trend of the pandemic improves. The insurer issues, through a securitization process, pandemic-related bonds through a special purpose vehicle (cf. [3]). The bonds are bought by the investors who bet on the improvement of the contagion. In case of a contagion reduction, the investor will receive a cash flow being, in any case, an actor of a virtuous project aimed at collective health. Indicated by K the amount of the coupon cap, by I_t the number of infected at each maturity and by value $\bar{I}_t$ its expected value, the cash flows for the investor and the insurer respectively are the following:

$$
\text{CFInv}_t = \begin{vmatrix} K & \bar{I}_t - I_t \geq K \\ 0 & \bar{I}_t - I_t < 0 \\ \bar{I}_t - I_t & 0 < \bar{I}_t - I_t \leq K \end{vmatrix} \tag{1}
$$

$$
\text{CFIns}_t = \begin{vmatrix} 0 & \bar{I}_t - I_t \geq K \\ K & \bar{I}_t - I_t < 0 \\ K - (\bar{I}_t - I_t) & 0 < \bar{I}_t - I_t \leq K \end{vmatrix} \tag{2}
$$

The market value of the securitized product arises from the relationship between the two commercial counterparts: the insurer, which issues the security, and the vehicle company, which we will indicate with the acronym SPC: the SPC puts the securitized bond on the market and sets its price. In order to indicate the procedure for calculating the price of the securitized bond, we describe the financial relationship on which the balance between the two inflows and outflows for the SPC is based. The inflows consist of the price (that is assumed to be the maximum price) that the insurer is willing to pay to undertake the transaction referred to each bond, p_0, to which must be added the total revenue resulting from the sale of the bonds. We will indicate with P_0 the price of the single bond and with N the number of bonds being sold. By observing that the outgoing sum in the various maturities of the bond is for the SPC always equal to K, and by indicating with F the nominal value of the bond at maturity, we can write the following financial equation, in which the amount on the left-hand side must cover the outgoing at the right side (cf. [4]):

$$
Np_0 + N * P_0 = NK * E_0\left[\sum_{t=1}^{n} \exp\left(-\int_0^t \delta_s ds\right)\right]
$$
$$
+ \text{NF} * E_0\left[\exp\left(-\int_0^n \delta_s ds\right)\right] + \text{NM}_0 \tag{3}
$$

In Eq. (3) n represents the number of the maturities of the coupons, δ_s represents the instantaneous risk-free rate and M_0 is the profit margin for each bond recognized to the SPC.

In the right-hand side of Eq. (3), the future amounts have been calculated as expected values in $t = 0$ of the flows discounted at the risk-free rate.

In Eq. (3) the unknown is P_0, that is the price of the securitized bond. If we assume to operate in a risk-neutral system also in the calculation of p_0, i.e., the price paid by the insurer, it is possible to determine its value as the average, always calculated in $t = 0$ at the stochastic rate δ_s, of the random incoming flows for the insurer.

We can write:

$$P_0 = K * E_0\left[\sum_{t=1}^{n} exp\left(-\int_0^t \delta_s ds\right)\right] + F * E_0\left[exp\left(-\int_0^n \delta_s ds\right)\right] - p_0 + M_0$$

2.2 Which Category Within Socially Responsible Investments?

The proposed asset may be included in the macro area of socially responsible investments, those made by entities whose strategies, practices and ownerships incorporate environmental, social and governance (ESG) factors (cf. [13]).

The integration of these factors in investment choices, indeed, is becoming increasingly widespread. This is due to both regulatory changes and a growing market demand. These two trends have induced many investors to orient their investment strategy according to PRI (Principles Responsible Investment) and integrate the ESG factors in their choice, a clear indication of the increasing relevance of ESG criteria in the firms' decision strategies (cf. [14]). For example, is it known that BlackRock has recently announced that "almost all \$ 7 trillion assets under management would be governed by ESG principles".

Concerning the bond we propose, the ESG integration strategy can be analysed from a double point of view: the one of the investor, who considers it as a securitized instrument and the one of the insurer who manages an asset to mitigate one of the risks associated with ESG factors.

Referring to the investor perspective, it is important at first to know that currently an investor can integrate ESG criteria into his investment choices through different approaches (cf. [13] and [14]): Best in Class, Engagement and Voting, ESG Integration, Exclusions, Impact Investing, Norm Based Screenings, Sustainability Themed, Impact Investing. Among these categories, we focus on the last two, here briefly recalled (cf. [7]), that seem to fit our product well.

Sustainability Themed

> *"Investment in themes or assets linked to the development of sustainability. Thematic funds focus on specific or multiple issues related to ESG. Sustainability themed investments inherently contribute to addressing social and/or environmental challenges such as climate change, eco-efficiency and health. Funds are required to have an ESG analysis or screen of investments in order to be counted in this approach."*

Impact Investing

> *"Impact Investments are investments made into companies, organisations and funds with the intention to generate social and environmental impact alongside a financial return. Impact investments can be made in both emerging and developed markets and target a range of returns from below market-to-market rate, depending upon the circumstances. Investments are often project-specific, and distinct from philanthropy, as the investor retains ownership of the asset and expects a positive financial return. Impact investment includes microfinance, community investing, social business/entrepreneurship funds and French fonds solidaires."*

In details, we highlight why we have chosen these categories.

Sustainability Themed, because the object of the underwriting considered one of the risk categories belonging to the social area of ESG factors, as we can observe in Table 1 (cf. [14]), that shows the main risk categories related to the different ESG areas.

The product we propose deals with the fifth criteria in the Table 1. The development of strategies that reduce the risk of contagion in the company, indeed, represents a way to improve workers' condition.

Impact Investing. Considering the reward scheme associated with particularly virtuous realities, this instrument could take on a further connotation, which exceeds the logic of thematic investment; rather, it is a typical instrument of impact finance. In fact, it not only mitigates negative effects but also encourages virtuous actions. The effect generated is therefore planned ex ante (intentional), additional (because obtained thanks to the direct action of the companies) and measurable (through the evaluation of the effects related to the risk of contagion). These three elements allow us to talk about the impact generated by the underlying shares and therefore of a tool that can fit into the impact finance assets.

Even analyzing the securitized bond from an insurance point of view, you can consider it in the most innovative asset categorization for impact finance, which is recent for the insurance sector. This area, in fact, compared to investment, has only recently developed tools geared to sustainability and impact. This evolution stems from the awareness that the insurance industry has a primary role in promoting economic, social and environmental sustainability (cf. [1]).

Table 1. Risk Criteria and examples of mitigation for Social Factor

Criteria	Theme	Risk criteria	Risk mitigation examples & good practice
Social	Human rights	Child labour	Policy/statement on protecting and promoting human rights, prohibits child labour, shared with suppliers, regular audits and public findings (e.g. ILO, UNDHR)
		Human trafficking	Human rights policy that includes a statement on protecting and promoting human rights and prohibits human trafficking
		Forced labour	Human rights policy that includes a statement on protecting and promoting human rights and prohibits forced labour
		Forced resettlement (including land/water rights for native people, land grabbing)	Free, prior & informed consent (FPIC) achieved. Effective environmental & social impact assessment (ESIA) process covering consultation, resettlement, compensation aspects
		Poor worker safety record (e.g. worse than sector average record on accidents)	Effective occupational health & safety policy that defines safety responsibilities and prevention measures to minimise fatalities, injuries and health impacts
		Violation of worker rights (e.g. discrimination, collective bargaining)	Code of conduct that outlines company's commitment to respect workers' rights
		Misconduct of security personnel (e.g. physical harm to people, human rights abuses)	Whistle-blower channel to report such violations
	Controversial weapons	Controversial weapons exposure (e.g. UN conventions)	Anti-Personnel Mine Ban Convention, Convention on Cluster Munitions

Source: UN Environment Programme's Principles for Sustainable Insurance Initiative, 2020 (cf. [14]).

References

1. Allianz Research: Impact underwriting: sustainable insurance as an opportunity for society and business (2020)
2. Cairns, A., Blake, D., Dowd, K.: Pricing death: frameworks for the valuation and securitization of mortality risk. ASTIN Bull. **36**(1), 79–120 (2006)
3. Di Lorenzo, E., Sibillo, M.: Economic paradigms and corporate culture after the great COVID-19 pandemic: towards a new role of welfare organisations and insurers. Sustainability **12**(8163), 1–14 (2020)
4. Di Lorenzo, E., Piscopo, G., Sibillo, M., Tizzano, R.: Reverse mortgage and risk profile awareness: proposals for securitization. Appl. Stoch. Models Bus. Ind. (2021). https://doi.org/10.1002/asmb.2664
5. Di Lorenzo, E., Sibillo, M.: COVID-19: new role for insurers. Scholarly Community Encyclopedia (2020). https://encyclopedia.pub/3041
6. Erikson, S.: Global health futures? Reckoning with a pandemic bond. MAT Med. Antropol. Theory. **6**, 77–108 (2020)
7. Eurosif. https://www.eurosif.org/responsible-investment-strategies/
8. G20 Sustainable Finance Study Group: Towards a sustainable infrastructure securitization market: the role of collateralized loan obligations (CLO) (2018). http://unepinquiry.org/wp-content/uploads/2018/12/Towards_a_sustainable_infrastructure_securitisation_market.pdf
9. Gabor, D.: Securitization for Sustainability. Does it help achieve the Sustainability Development Goals? Heinrich Böll Stiftung Washington, D.C. (2019)
10. Gründl, H., Regele, F.: Pandemic insurance through pandemic partnership bonds: a fully funded insurance solution in a public private partnership. Frankfurt a. M.: Leibniz Institute for Financial Research SAFE (2020)
11. Hainaut, D.: An actuarial approach for modeling pandemic risk. Risks **9**, 3 (2021). https://doi.org/10.3390/risks9010003
12. Legenchuk, S., Pashkevych, M., Usatenko, O., Driha, O., Ivanenko, V.: Securitization as an innovative refinancing mechanism and an effective asset management tool in a sustainable development environment. In: The International Conference on Sustainable Futures: Environmental, Technological, Social and Economic Matters (ICSF 2020), vol. 166, pp. 1–16 (2020)
13. PRI (Principles Responsible Investment): What is responsible investment? (2020)
14. UN Environment Programme's Principles for Sustainable Insurance Initiative: PSI ESG Guide for Non-Life Insurance: Version 1.0 (2020)

Feynman-Kac Formula for BSDEs with Jumps and Time Delayed Generators Associated to Path-Dependent Nonlinear Kolmogorov Equations

Luca Di Persio[1], Matteo Garbelli[1,2(✉)], and Adrian Zalinescu[3]

[1] Department of Computer Science, University of Verona, Verona, Italy
[2] Department of Mathematics, University of Trento, Trento, Italy
`matteo.garbelli@unitn.it`
[3] Faculty of Computer Science, Alexandru Ioan Cuza University, Iasi, Romania

Abstract. We study a system of forward-backward stochastic differential equations (FBSDEs), with time delayed generator driven by a Lèvy-type noise, establishing a non-linear Feynman-Kac representation formula to associate the BSDE solution to a path dependent nonlinear Kolmogorov equation. We also provide two financial applications: a generalization of the Large Investor Problem and an insurance investment type model.

1   The Non-linear Path Dependent Kolmogorov Equation

We provide a probabilistic representation of a (mild) solution of the following path-dependent nonlinear Kolmogorov equation for a fixed time horizon $T < \infty$

$$\begin{cases} -\partial_t u(t,\phi) - \mathcal{L}u(t,\phi) - f\big(t,\phi,u(t,\phi),(\partial_x u \cdot \sigma)(t,\phi),(u(\cdot,\phi))_t,\mathcal{J}u(t,\phi)\big) = 0 \\ u(T,\phi) = h(\phi), \qquad \phi \in \Lambda \end{cases}$$

$$(1.1)$$

with $t \in [0,T)$ and $\phi \in \Lambda := \mathcal{D}([0,T];\mathbb{R}^d)$ being $\mathcal{D}$, the space of càdlàg $\mathbb{R}^d$-valued functions. The second order differential diffusion operator $\mathcal{L}$ is defined by

$$\mathcal{L}u(t,\phi) := \frac{1}{2} Tr\left[\sigma(t,\phi)\sigma^*(t,\phi)\partial^2_{xx}u(t,\phi)\right] + \langle b(t,\phi),\partial_x u(t,\phi)\rangle$$

with $b : [0,T] \times \Lambda \to \mathbb{R}^d$ and $\sigma : [0,T] \times \Lambda \to \mathbb{R}^{d\times d'}$ two non anticipative functionals. For a fixed delay $\delta > 0$, we define the delayed term by

$$(u(\cdot,\phi))_t := (u(t+\theta,\phi))_{\theta\in[-\delta,0]}$$

while $\mathcal{J}$ is the integro-differential operator associated to the jump behaviour

$$\mathcal{J}u(t,\phi) := \int_{\mathbb{R}\setminus\{0\}} \big[u\big(t,\phi+\gamma(t,\phi,z)\big) - u(t,\phi)\big]\Delta(z)\nu(dz), \qquad (1.2)$$

where $\gamma : [0,T] \times \Lambda \times \mathbb{R} \setminus \{0\} \to \mathbb{R}^d$ is a continuous, non anticipative, functional, ν is a Lèvy measure and Δ is the parameter modelling the intensity of jumps. As proved in Theorem 2, the deterministic non-anticipative functional $u : [0,T] \times \Lambda \to \mathbb{R}$ represents a mild solution to (1.1), and $u(t,\phi) := Y^{t,\phi}(t)$ holds, with $Y^{t,\phi}(t)$ the first component of the BSDE solution. We remark that the concept of mild solutions to delay equations generalizes the standard notion of viscosity solutions, see, e.g., [6] or [10].

2   The FBSDE System

Let us consider the following probability space $\big(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t\in[0,T]}, \mathbb{P}\big)$, where $\{\mathcal{F}_t\}_{t\in[0,T]}$ is jointly generated by a Brownian motion $W(s)$ and a Poisson random dom measure $N(ds,dz)$ independent from W, for all $z \in \mathbb{R}_0 := \mathbb{R} \setminus 0$ and for all $s \in [0,T]$. The compensated Poisson measure $\tilde{N}$ associated to the Lévy measure ν is defined by

$$\tilde{N}(dt,dz) := N(dt,dz) - \nu(dz)dt. \qquad (2.1)$$

We denote by $X^{t,\phi}$ the solution of the following forward SDE

$$\begin{cases} X(s) = \phi(t) + \int_t^s b(r,X)dr + \int_t^s \sigma(r,X)dW(r) \\ + \int_t^s \int_{\mathbb{R}} \gamma(X,r,z)\tilde{N}(dr,dz), \qquad s \in [t,T] \\ X(s) = \phi(s), \qquad\qquad\qquad\quad s \in [0,t] \end{cases} \qquad (2.2)$$

to emphasize the dependence on the initial time $t \in [0,T]$ and a given càdlag path $\phi \in D([0,t];\mathbb{R}^d)$. We assume that the non anticipative functions $b : [0,T] \times \Lambda \to \mathbb{R}^d$, $\sigma : [0,T] \times \Lambda \to \mathbb{R}^{d\times d'}$ and $\gamma : [0,T] \times \Lambda \times \mathbb{R} \to \mathbb{R}^d$ are continuous and they satisfy some Lipschitz-type conditions. The proof of existence and uniqueness is a classical result, proved in, e.g., [3] via Picard iterations.

We fix a delay $\delta \in \mathbb{R}^+$ and we consider a time-delayed BSDE, given by

$$\begin{cases} Y(s) = h(X^{t,\phi}) + \int_s^T f\big(r,X^{t,\phi},Y(r),Z(r),\tilde{U}(r),Y_r\big)dr \\ \quad - \int_s^T Z(r)dW(r) - \int_s^T \int_{\mathbb{R}} U(r,z)\tilde{N}(dr,dz), \qquad s \in [t,T] \end{cases} \qquad (2.3)$$

where $\tilde{U}$ is defined by

$$\tilde{U}(t) = \int_{\mathbb{R}_0} U(t,z)\Delta(z)\nu(dz)$$

being Δ the parameter defined in the jump operator (1.2) appearing in the Kolmogorov Equation (1.1). The solution of (2.3) corresponds to a triple of stochastic processes that denoted by $\left(Y^{t,\phi}, Z^{t,\phi}, U^{t,\phi}\right)$. We remark that the solution of (2.3) depends on $X^{t,\phi}$, i.e. the path of the forward dynamic (2.2) and on the delayed term $Y_r^{t,\phi}$ appearing in the generator f that is the path of $Y^{t,\phi}$ restricted to $[r-\delta, r]$, namely

$$Y_r := (Y(r+\theta))_{\theta \in [-\delta, 0]}.$$

To fully characterize an admissible solution over the whole interval $[0, T]$, we should also provide supplementary initial conditions, namely

$$Y^{t,\phi}(s) = Y^{s,\phi}(s), \qquad Z^{t,\phi}(s) = U^{t,\phi}(s, z) = 0, \qquad s \in [0, t].$$

The solution of the BSDE $(Y^{t,\phi}, Z^{t,\phi}, U^{t,\phi})$ belongs to the Banach space $\mathbb{S}_t^2(\mathbb{R}) \times \mathbb{H}_t^2(\mathbb{R}^{d'}) \times \mathbb{H}_{t,N}^2(\mathbb{R})$ where

- $\mathbb{S}_t^2(\mathbb{R})$ denotes the space of $\mathbb{F}^t$-adapted, product measurable càdlàg processes $Y : \Omega \times [0, T] \to \mathbb{R}$ satisfying $\mathbb{E}\big[\sup_{t \in [0,T]} |Y(t)|^2 \big] < \infty$;
- $\mathbb{H}_t^2(\mathbb{R}^{d'})$ denotes the space of $\mathbb{F}^t$-predictable processes $Z : \Omega \times [0, T] \to \mathbb{R}^{d'}$ satisfying $\mathbb{E}\big[\int_0^T |Z(t)|^2 dt \big] < \infty$;
- $\mathbb{H}_{t,N}^2(\mathbb{R})$ denotes the space of $\mathbb{F}^t$-predictable processes $U : \Omega \times [0, T] \times (\mathbb{R} \setminus \{0\}) \to \mathbb{R}$ satisfying $\mathbb{E}\big[\int_0^T \int_{\mathbb{R} \setminus \{0\}} |U(t,z)|^2 \nu(dz) dt \big] < \infty$.

We need the generator $F : [0, T] \times \Lambda \times \mathbb{R} \times \mathbb{R}^{d'} \times \mathbb{R} \times L^2([-\delta, 0]; \mathbb{R}) \to \mathbb{R}$ to be continuous w.r.t. ϕ, Lipschitz w.r.t. Y, Z, U and Y_r and $\mathbb{F}$-progressively measurable, for any $(Y, Z, U, Y_r) \in \mathbb{R} \times \mathbb{R}^{d'} \times \mathbb{R} \times L^2([-\delta, 0]; \mathbb{R})$. Moreover, we require that $\phi \mapsto h(t, \phi)$ is continuous and $|h(\phi)| \leq M(1 + \|\phi\|_T^p)$, for all $\phi \in \Lambda$. To prove existence, uniqueness and continuity of $Y^{t,\phi}$ w.r.t. ϕ, we need a small delay δ or small Lipschitz constant K, see or [5,7] or [9].

Theorem 1. *If suitable assumptions on the coefficients are satisfied and by requiring K or δ small enough, then there exists a unique solution $(Y^{t,\phi}, Z^{t,\phi}, U^{t,\phi})$ of the BSDE (2.3) such that $(Y^{t,\phi}, Z^{t,\phi}, U^{t,\phi}) \in \mathbb{S}_t^2(\mathbb{R}) \times \mathbb{H}_t^2(\mathbb{R}^{d'}) \times \mathbb{H}_{t,N}^2(\mathbb{R})$ for all $t \in [0, T]$ and the application $t \to (Y^{t,\phi}, Z^{t,\phi}, U^{t,\phi})$ is continuous from $[0, T]$ into $\mathbb{S}_0^2(\mathbb{R}) \times \mathbb{H}_0^2(\mathbb{R}^{d'}) \times \mathbb{H}_{0,N}^2(\mathbb{R})$.*

3   Feynman-Kac Formula

The forward process $X^{t,\phi} : [0, T] \times D([0, T]; \mathbb{R}^d) \to \mathbb{R}^d$ is not *markovian*, since its value depends on the (near) past. Nevertheless, we can enlarge the state space and consider X as a process of the path, then lifting in in an infinite-dimensional state space, to recover *markovianity*. Following, e.g., [11], we can rephrase the problem in an infinite dimensional setting, by the so-called *product-space reformulation* into the so-called *Delfour-Mitter space* $M^2 := L^2([-T, 0]; \mathbb{R}^d) \times \mathbb{R}^d$, see, e.g., [2]. More precisely, we decompose the process $X^{t,\phi}$ into two components: the solution at time s is denoted by $X(s)$, while its path, up to time s, is denoted

by $X_{[0,s]}$. Moreover, by [6] (Prop. 2.6), we have that the strong solution to Eq. (2.2) in M^2, is a *Markov process* in the sense that

$$\mathbb{P}((X^{t,\phi} \in B | \mathcal{F}_s) = \mathbb{P}(((X^{t,\phi} \in B | (X^{t,\phi}(s) = \phi(s)), \quad \mathbb{P} - a.s.$$

for all $s \in [0,t]$ and for all sets B in the Borel set $\mathcal{B}(M^2)$. Furthermore, M^2 is a separable Hilbert space, see, e.g., [2], and, since $\mathcal{D}$ is densely and continuously embedded in M^2, we can derive the following non-linear Feynman-Kac formula.

Theorem 2. *Let suitable regularity assumptions hold and K or δ be small enough. Then*

$$Y^{t,\phi}(s) = u(s, X^{t,\phi}), \qquad \text{for all } s \in [0,T] \tag{3.1}$$

$\forall (t, \phi) \in [0,T] \times \Lambda$ *being $Y^{t,\phi}$ is the solution of the BSDE (2.3) while $u(t, \phi) : [0,T] \times \Lambda \to \mathbb{R}$ is a deterministic function given by the following representation formula*

$$u(t, \phi) = Y^{t,\phi} \quad (t, \phi) \in [0,T] \times \Lambda.$$

Moreover, the solution of the decoupled system of the FBSDEs given by (2.2) and (2.3) is the quadruple (X, Y, Z, U) taking values in $\Lambda \times \mathbb{R} \times \mathbb{R}^{d'} \times \mathbb{R}$.

4   Financial Applications

Section 4.1 generalizes the Large Investor Problem with a stock price jump-diffusion dynamic and Sect. 4.2 describes an insurance problem with a payment dynamics consistent with a step process and evaluated by a dynamic risk measure.

4.1   The Large Investor Problem

Following [4] and [5], we consider an investor with strategy π and investment portfolio X^π, acting on a financial market whose strategy affects μ and r

$$\begin{cases} \dfrac{dS_0(t)}{S_0(t)} = r(t, X^\pi(t), \pi(t), X_t^\pi)dt , & S_0(0) = 1 , \\[2mm] \dfrac{dS(t)}{S(t)} = \mu(t, X^\pi(t), \pi(t), X_t^\pi)\, dt + \sigma(t, X^\pi(t), X_t^\pi)dW(t), & (4.1) \\[2mm] \qquad + \displaystyle\int_{\mathbb{R}\backslash\{0\}} \gamma(t, X^\pi(t), X_t^\pi, z)\delta(z)\nu(dz), & S(0) = s_0 > 0 , \end{cases}$$

where r, μ, σ and γ are $\mathbb{F}^{W,\tilde{N}}$-predictable processes, being $\mathbb{F}^{W,\tilde{N}}$ the natural filtration associated to the Brownian motion W and adapted to the Poisson random measure $\tilde{N}$. The total amount of the large investor portfolio reads

$$dX^\pi = \pi(t)\frac{dS(t)}{S(t)} + (X^\pi(t) - \pi(t))\frac{dS_0(t)}{S_0(t)}dt ,$$

and we aim at finding an admissible replicating strategy $\pi \in \mathcal{A}$ for a claim $h(S(T))$. By plugging (4.1), X evolves according to

$$
\begin{aligned}
dX(t) = \pi(t) \cdot \Big[& \mu\left(t, X(t), \pi\left(t\right), X_t\right) dt + \sigma(t, X(t), X_t) dW(t) \\
& + \int_{\mathbb{R}\setminus\{0\}} \gamma(t, X(t), X_t, z) \tilde{N}(dt, dz) \Big] + \left[X\left(t\right) - \pi(t)\right] \cdot r(t, X(t), \pi\left(t\right), X_t) dt
\end{aligned}
$$

with the final condition $X(T) = h\left(S\right)$. Hence, for $t \in [0, T]$, we have

$$
\begin{aligned}
X(t) = h\left(S\right) + \int_t^T & F\left(s, X\left(s\right), \pi\left(s\right), X_s, \pi_s\right) ds - \int_t^T \pi(s) \sigma(s, X(s), X_s) dW(s) \\
& - \int_t^T \int_{\mathbb{R}\setminus\{0\}} \pi(s) \gamma(s, X(s), X_s, z) \tilde{N}(ds, dz),
\end{aligned}
$$

$$(4.2)$$

The system composed by the Forward SDE (4.1) and the BSDE (4.2) replicates the structure already analysed, see (2.2) and (2.3). By imposing suitable assumptions on μ, r, σ and γ, Theorem 1 allows us to have both existence and uniqueness, while, by Theorem 2, we derive the following solution representation (X^π) of the backward component of (4.2) for every $(t, \phi) \in [0, T] \times \Lambda$, namely $u\left(t, \phi\right) = X^{t,\phi}\left(t\right)$ where $u(t, \phi)$ a mild solution of a path-dependent PDE as the same kind of Eq. (1.1).

4.2   Dynamic Risk Measure for an Insurance Payment Process

We consider a Black-Scholes setting and a filtration $\mathcal{F} = (\mathcal{F})_{0 \leq t \leq T}$. We define a $\mathcal{F}$-adapted Brownian motion W and a random measure N generated by an $\mathcal{F}$-adapted step process. The dynamics of the risk-less bond $S_0 := (S_0)_{0 \leq t \leq T}$ is described by

$$
\frac{dS_0(t)}{S_0(t)} = r(t) dt, \qquad S_0(0) = 1 \tag{4.3}
$$

where $r(t)$ denotes the risk-free rate. The stock $S := (S)_{0 \leq t \leq T}$ is described by

$$
\frac{dS(t)}{S(t)} = \mu(t) dt + \sigma(t) dW(t), \qquad S(0) = s > 0 \tag{4.4}
$$

where μ is the expected return and σ the stock volatility. According to [7], insurance claims are modelled by a step process J and a jump measure N, e.g., J can be modeled as a point process defined w.r.t. a sequence of random variables (defaults of securities, insured persons deceases or policies surrenders). The insurance payment process reads

$$
P(t) = \int_0^t H(s) ds + \int_0^t \int_{\mathbb{R}} G(s, z) N(ds, dz) + F \mathbb{1}_{\{t = T\}}, \quad 0 \leq t \leq T \tag{4.5}
$$

where the process P contains payments H continuously occurring during the contract life (annuities), claims G, randomly happening because triggered by the step process J (death benefits), and liability F settled at the end of the insurance,

(survival benefit). We consider an investor (insurer) facing liabilities' stream (4.5) and investing in the stock (4.4), assuming that the combined financial and insurance model is arbitrage-free. Her goal is to replicate the insurance by investing on assets and to quantify the risk of the investing activities.

The replicating portfolio of the investment is denoted by $X^\pi = (X^\pi(t))_{0 \le t \le T}$ under a strategy π. We assume $X^\pi := (X^\pi(t), 0 \le t \le T)$ is self-financing and its dynamic is given by the following forward SDE

$$
\begin{aligned}
dX^\pi(t) &= \pi(t) \frac{dS(t)}{S(t)} + (X^\pi(t) - \pi(t)) \frac{dS_0(t)}{S_0(t)} - dP(t) \\
&= \pi(t) \left(\mu(t)dt + \sigma(t)dW(t) \right) + (X^\pi(t) - \pi(t)) r(t)dt - dP(t) \\
X^\pi(0) &= x > 0
\end{aligned}
\tag{4.6}
$$

where π denotes the amount invested in the risky asset S and x the initial capital.

By plugging (4.5) into (4.6), we obtain

$$
\begin{aligned}
dX^\pi(t) &= \pi(t) \left(\mu(t)dt + \sigma(t)dW(t) \right) + (X^\pi(t) - \pi(t)) r(t)dt \\
&\quad -H(t)dt - \int_{\mathbb{R}} G(t,z)N(dt,dz), \qquad X^\pi(0) = x > 0
\end{aligned}
\tag{4.7}
$$

where we have to subtract the claim F from the terminal wealth $X^\pi(T)$.

To calibrate the investment, we consider a dynamic risk measure and, according to [12] or [1], we model the risk measure $Y(t)$ as a g-expectation to incorporate a memory effect by moving average, see [8]. We consider two bounded and Lipschitz functions g_1 and g_2, s.t. $g_2(0) = 0$ and a g-expectation of the form $g(x, y, z) = \beta g_1(\bar{y}) g_2(z)$, where $\bar{y}$ is the time-average in a sufficiently small time interval and a given financial weight $\beta \in \mathbb{R}$, hence dealing with a time-delayed BSDE:

$$
Y(t) = h \left(X^\pi(T) - F \right) + \frac{\beta}{\delta} \int_t^T g_1 \left(\int_{-\delta}^0 Y(s+r)dr \right) g_2(Z(s))ds - \int_t^T Z(s)dW(s)
\tag{4.8}
$$

assuming a terminal payoff $h : \Lambda \to \mathbb{R}$ depending on the final wealth of the investment discounted by F, and δ being the amount of the small delay. By associating (4.7) to the BSDE (4.8) as a risk measure, we obtain a system that is analogous to our theoretical setting given by (2.2) and (2.3).

References

1. Acciaio, B., Penner, I.: Dynamic risk measures. Risk Manag. 1–34 (2011)
2. Baños, D.R., Cordoni, F., Di Nunno, G., Di Persio, L., Røse, E.E.: Stochastic systems with memory and jumps, arXiv e-prints (2016)
3. Barles, G., Buckdahn, R., Pardoux, E.: Backward stochastic differential equations and integral-partial differential equations. Stochast. Stochast. Rep. **60**, 57–83 (1997)
4. Cvitanic, J., Ma, J.: Hedging options for a large investor and forward-backward SDEs. Ann. Appl. Probab. **6**(2), 370–398 (1996)

5. Cordoni, F., Di Persio, L., Maticiuc, L., Zălinescu, A.: A stochastic approach to path-dependent nonlinear Kolmogorov equations via BSDEs with time-delayed generators and applications to finance. Stochast. Process. Appl. (2020)

6. Cordoni, F., Di Persio, L., Oliva, I.: A nonlinear Kolmogorov equation for stochastic functional delay differential equations with jumps. Nonlinear Differ. Equ. Appl. **24**, 16 (2017). https://doi.org/10.1007/s00030-017-0440-3

7. Delong, Ł: Backward Stochastic Differential Equations with Jumps and Their Actuarial and Financial Applications: BSDEs with Jumps. EAA Series, Springer, London (2013). https://doi.org/10.1007/978-1-4471-5331-3

8. Delong, Ł: BSDEs with time-delayed generators of a moving average type with applications to non-monotone preferences. Stoch. Models **28**, 281–315 (2012)

9. Delong, Ł, Imkeller, P.: Backward stochastic differential equations with time delayed generators - results and counterexamples. Ann. Appl. Probab. **20**(4), 1512–1536 (2010)

10. Fuhrman, M., Tessitore, G.: Generalized directional gradients, backward stochastic differential equations and mild solutions of semilinear parabolic equations. Appl. Math. Optim. **51**, 279–332 (2005). https://doi.org/10.1007/s00245-004-0814-x

11. Masiero, F., Orrieri, C., Tessitore, G., Zanco, G.: Semilinear Kolmogorov equations on the space of continuous functions via BSDEs, arXiv math.PR (2019)

12. Peng, S.: Backward stochastic differential equation, nonlinear expectation and their applications. In: Proceedings of the International Congress of Mathematicians, Hyderabad, India (2010)

The Role of Stablecoins: Cryptocurrencies Sought Stability and Found Gold and Dollars

Antonio Díaz, Carlos Esparcia[✉], and Diego Huélamo

Universidad de Castilla-La Mancha, Plaza de la Universidad 1,
02071 Albacete, Spain
Carlos.Esparcia@uclm.es

Abstract. This paper empirically assesses the ability of three putative stablecoins (two dollar-backed, Tether and USD Coin; and one gold-backed, Digix Gold) to reduce the risk of a traditional cryptocurrency portfolio during the COVID-19 pandemic. A monthly rebalance experiment is conducted over an out-of-sample period, so that the effects of including stablecoins in terms of diversification can be clearly assessed. The GO-GARCH model is implemented to obtain dynamic estimates of conditional co-moment arrays up to order four. Then, assuming a CARA utility function and a risk defensive investor profile, an extension of the certainty equivalent with co-skewness and co-kurtosis is conducted for portfolio allocation purposes. Using the Cornish-Fisher expansion of the parametric VaR (i.e., the modified VaR), we evaluate how the introduction of every single stablecoin into a traditional cryptocurrency portfolio affects the downside risk of the combined strategy. The empirical evidence highlights that the two dollar-backed tokens have high diversification and hedging capabilities against traditional cryptocurrencies and can even act as safe havens, whereas Digix Gold shows a high diversification potential, but constrained by its high intrinsic volatility. In addition, our results also reveal the importance of considering higher order moments when forming cryptocurrency portfolios and measuring their risk.

Keywords: Co-skewness · Co-kurtosis · Cryptocurrency · Modified VaR · Portfolio allocation · Stablecoin

1 Introduction

Not much research has yet been done on the capabilities of stablecoins to act as diversifiers in traditional cryptocurrency portfolios, with the notable exception of [9]. Our study follows this trend by assessing stablecoins capabilities to reduce the VaR of a traditional cryptocurrency portfolio. Based on a careful and state-of-the-art methodology, we provide empirical evidence to help answer a relevant

and novel question for many academics and practitioners: do stablecoins act as diversifiers or hedgers against more volatile and traditional cryptocurrencies?

We conduct an out-of-sample experiment during the COVID-19 pandemic (which covers from March the 1st of 2020 to May the 25th of 2021, and with an in sample period ranging from October the 5th of 2018 to February the 29th of 2020), through which we obtain one-day ahead forecasted co-moment tensors by means of the calibrated Generalized Orthogonal GARCH (GO-GARCH) models and use these forecasts to reallocate the portfolios every 30 days. We consider a base portfolio formed by five traditional cryptocurrencies: Bitcoin, Ethereum, Ripple, Cardano and Litecoin; and its diversification by means of the introduction of one of the following three asset-backed cyptocurrencies: Tether, USD Coin and Digix Gold (the first two being USD backed and the latter being gold-backed).

First, we classify the considered stablecoins as diversifyiers or hedgers following the definitions proposed in [1], so that we consider a diversifying asset to be one that presents non-perfect positive correlation with another asset, while a hedging asset would be one that presents zero or negative correlation with another asset. Second, we dynamically estimate the co-moment tensors up to fourth order by making use of the GO-GARCH model proposed in [8]. Third, four portfolios are constructed over the out-of-sample period: one formed by five traditional cryptocurrencies, and three others composed by the latter base portfolio together with one of the three stablecoins. Two portfolio allocation strategies are considered: (1), the maximization of the certainty equivalent (CE) that arises from a Constant Absolute Risk Aversion (CARA) utility function of a highly risk-averse investor with preference for positive skewness and low kurtosis; (2), minimum variance (MV) portfolios are formed as a benchmark. Fourth, we dynamically estimate the downside risk of the portfolios by estimating the modified VaR (mVaR) proposed in [5], and we analyze the results in terms of the effect that the introduction of stablecoins has on portfolio risk. Thus, our main contribution consists on assessing the capabilities of stablecoins on the allocation and diversification of traditional cryptocurrency portfolios considering co-moments up to order four both in terms of optimization and risk management. This is also the first paper to apply the GO-GARCH model to portfolios composed exclusively of cryptocurrencies and also the first to apply it to stablecoins.

The results show that the two dollar-backed tokens have high diversification and hedging capabilities against traditional cryptocurrencies, whereas Digix Gold shows a significant diversification potential constrained by its high intrinsic volatility. In addition, our results also reveal the importance of considering higher order moments when forming cryptocurrency portfolios and measuring their risk.

2   Methodology

2.1   The Portfolio Allocation Method

We consider an investor who allocates her portfolio by maximizing her CE. Given a lottery L in which a particular individual can participate, the CE is the amount of wealth for which she would feel indifferent between taking it or playing the lottery. By developing a Taylor expansion on the investor's utility function we reach the function to maximize:

$$CE \approx \mu_p - \frac{1}{2}\lambda\sigma_p^2 + \frac{s_p}{6}\lambda^2\sigma_p^3 - \frac{k_p - 3}{24}\lambda^3\sigma_p^4 \tag{1}$$

which is an approximation of the investor's CE that accounts for the first four moments of the portfolio returns, being: the mean (μ_p), the variance (σ_p), the skewness (s_p) and the kurtosis (k_p). The moments are estimated dynamically by means of the GO-GARCH model proposed in [8], as it allows to estimate conditional co-moment tensors. In addition, MV portfolios are formed to serve as benchmark. To these optimization problems we add a no short-selling constraint [3,7] as this could ostensibly increase the risk and the instability of the portfolios.

2.2   Downside Risk Measures and Backtesting

As a risk measure we employ the mVaR proposed by [5], which is an extension of gVaR that accounts for skewness and excess kurtosis. The mVaR can be computed as the sum of the gVaR and a term derived from the Cornish-Fisher expansion:

$$\text{mVaR}_t(\alpha) = \text{gVaR}_t(\alpha)$$
$$- \sigma_{p,t}\left(\frac{1}{6}(z_\alpha^2 - 1)s_{p,t} + \frac{1}{24}(z_\alpha^3 - 3z_\alpha)k_{p,t} - \frac{1}{36}(2z_\alpha^3 - 5z_\alpha)s_{p,t}^2\right) \tag{2}$$

where $z_\alpha = \Phi^{-1}(\alpha)$ corresponds to the quantile α of a gaussian distribution with zero mean and unit variance, with $0 < \alpha < 1$ being the significance level (which we set at 0.04 and 0.01).

To check the effectiveness of the risk measures considered, we conduct a backtesting based on three broadly used tests: the unconditional coverage test [6], the conditional coverage test [2] and Dynamic Quantile test [4].

3   Main Results and Findings

From our preliminary analysis of the conditional correlation between stablecoins and traditional cryptocurrencies, we report on the different behaviors exhibited by the various structures of dependence, depending not only on the pairwise stablecoin but on the point in time under study. Figure 1 shows that Tether performs as a diversifier during the first months of the pandemic, during which uncertainty and turmoil in the markets was very high, and as a hedger since late 2020.

In contrast, correlations with USD Coin show a more stable and inverse behavior during most of the period considered, suggesting its role as a hedger. In the case of Digix Gold, we find a diversifying asset with a low positive correlation with traditional cryptocurrencies throughout the out-of-sample period. This low conditional correlation with traditional cryptocurrencies synergizes with the low intrinsic volatility of USD-backed tokens, which explains why they obtain average weightings close to 80% in CE portfolios and around 90% in MV portfolios. Conversely, Digix Gold's diversification potential gives it a significant position in the portfolio, with an average weighting of around 48% in the EC portfolio and 59% in the case of the MV portfolio; however, its high intrinsic volatility limits this potential and leads to a more unstable portfolio composition, drastically so in the case of the EC portfolio. From the risk impact analysis, the results of which are summarized in Fig. 2 and Table 1, we find that including USD-pegged stablecoins into the base cryptocurrency portfolio results in a significant and systematic risk mitigation of the combined strategy throughout the entire sample. In contrast, including Digix Gold conducts to a greater instability in risk reduction, even increasing it with respect to the base portfolio at certain times, which undoubtedly obeys to the high intrinsic volatility of this gold-backed currency, and which places it outside the domain of stablecoins, acting as a mere diversifier. Moreover, the empirical results highlight the importance of considering higher order moments when measuring the tail risk of cryptocurrencies, otherwise leading to an underestimation of the actual risk exposure, as mVaR exceeds gVaR by an increasing magnitude with confidence level.

Table 1. Summary statistics on the impact on the risk of CE and MV portfolios when introducing stablecoins. Note that here mVaR $(100 - \alpha\%)$ reduction and differential $(100 - \alpha\%)$ have the same interpretation as explained at the caption of Fig. 2.

With USDT	CE portfolio				MV portfolio			
	Mean	Std. dev.	Min.	Max.	Mean	Std. dev.	Min.	Max.
mVaR (99%) reduction	0.77390	0.19401	0.20444	0.95980	0.75730	0.01473	0.82673	0.94359
mVaR (96%) reduction	0.88104	0.20381	0.15097	0.94748	0.86973	0.01594	0.79228	0.92041
differential (99%)	0.01511	0.01751	−0.01243	0.06899	−0.00144	0.00629	−0.00482	0.04045
differential (96%)	0.01088	0.00577	−0.02339	0.02085	−0.00032	0.00471	−0.02203	0.01490
With USDC	Mean	Std. dev.	Min.	Max.	Mean	Std. dev.	Min.	Max.
mVaR (99%) reduction	0.77819	0.18640	0.28306	0.96878	0.92311	0.03582	0.79586	0.96927
mVaR (96%) reduction	0.76434	0.19252	0.24036	0.96731	0.88810	0.02213	0.77804	0.93133
differential (99%)	0.01198	0.01146	−0.02530	0.06325	0.01189	0.00689	−0.02261	0.02571
differential (96%)	−0.00187	0.00411	−0.02121	0.02331	−0.00140	0.00491	−0.02512	0.01521
With DGX	Mean	Std. dev.	Min.	Max.	Mean	Std. dev.	Min.	Max.
mVaR (99%) reduction	0.12419	0.56241	−3.89687	0.70818	0.22333	0.61631	−5.28361	0.67136
mVaR (96%) reduction	0.06172	0.49999	−3.27553	0.64488	0.12045	0.52273	−4.34238	0.55148
differential (99%)	0.04472	0.14250	−1.15905	0.27534	0.08215	0.18891	−1.62425	0.27227
differential (96%)	−0.01769	0.06600	−0.53773	0.09257	−0.02064	0.08459	−0.68301	0.08152
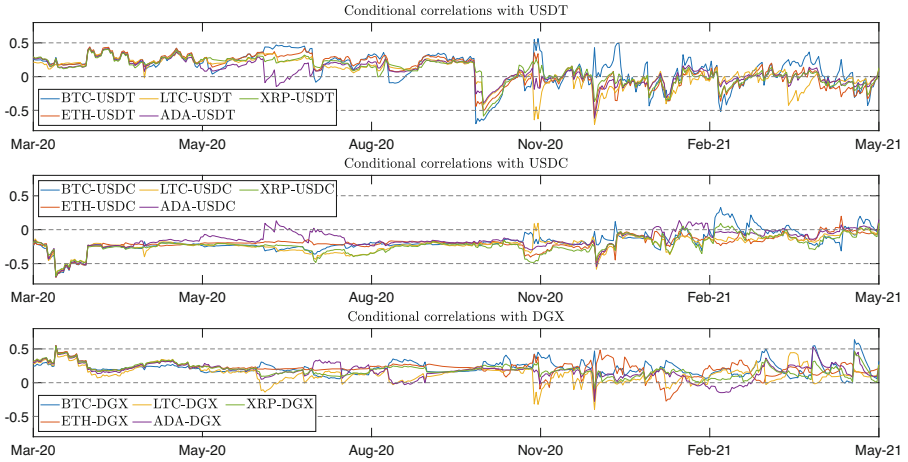
Fig. 1. Conditional correlations between traditional cryptocurrencies (Bitcoin (BTC), Ethereum (ETH), Ripple (XRP), Cardano (ADA), Litecoin (LTC)) and stablecoins (Tether (USDT), USD Coin (USDC), Digix Gold (DGX))
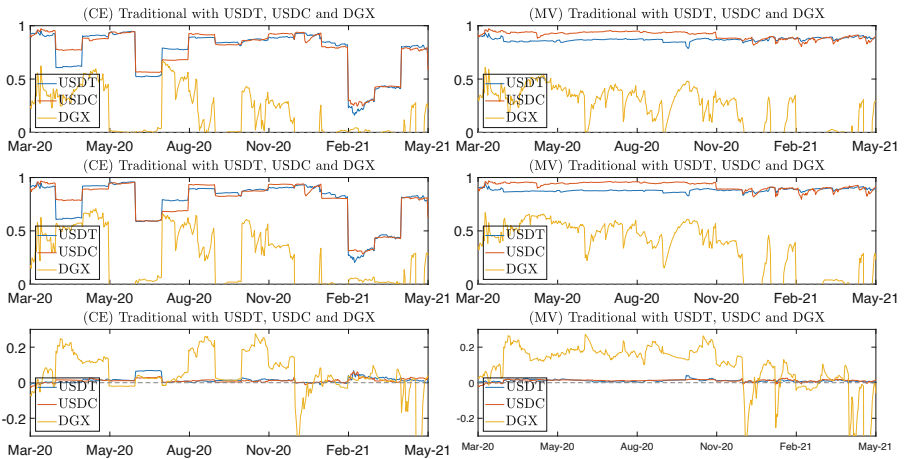
Fig. 2. Effects on VaR of introducing Tether (USDT), USD Coin (USDC) and Digix Gold (DGX). The first row shows the gVaR $(100 - \alpha\%)$ reduction, which represents the relative reduction of gVaR when comparing the portfolio that includes a stablecoin against the portfolio of traditional cryptocurrencies, i.e., $\frac{\mathrm{gVaR_{stable}}(100-\alpha\%)}{\mathrm{gVaR_{trad}}(100-\alpha\%)} - 1$, and equivalently for mVaR $(100 - \alpha\%)$ reduction shown in the second row. In the third row, the differential reduction represents the difference between the mVaR and gVaR relative reductions, i.e., $\frac{\mathrm{mVaR_{stable}}(100-\alpha\%)}{\mathrm{mVaR_{trad}}(100-\alpha\%)} - \frac{\mathrm{gVaR_{stable}}(100-\alpha\%)}{\mathrm{gVaR_{trad}}(100-\alpha\%)}$, so positive (negative) values indicate that the introduction of the stablecoin in the portfolio generates a higher (lower) relative reduction or increase in mVaR than in gVaR.

It would be interesting to extend this work in the future using Expected Shortfall (ES) and modified Expected Shortfall (mES) as complementary measures to gVaR and mVaR, or to approach risk measurement using copulas or Extreme Value Theory (EVT). On the other hand, a logical extension of this paper would be to broaden the variety of assets analyzed, especially with regard to stablecoins (Table 2).

Table 2. Backtesting gVaR and mVaR for the portfolios that maximize the CE. Traditional indicates the estimate of the model considering only the 5 traditional cryptocurrencies, while Tether (USDT), USD Coin (USDC) and Digix Gold (DGX) indicate the estimates of the models considering the 5 traditional cryptocurrencies plus the corresponding asset-backed cryptocurrency. 4 and 18 exceedances are expected at 1% and 4% significance levels, respectively. LR_{UC}, LR_{CC} and DQ denote the unconditional, conditional and Dynamic Quantile test statistics, respectively, the latter being specified with 4 lags of the endogenous variable Hit and one lag of the VaR as explanatory variables. *, ** and *** reveal significance at the 10%, 5% and 1% levels, respectively.

Portfolio	Model	Exceed.		LR_{UC}		LR_{CC}		DQ	
		1%	4%	1%	4%	1%	4%	1%	4%
Traditional	gVaR	7	11	1.1997	3.2783*	1.4214	3.8309	42.9778***	21.6749***
	mVaR	1	11	4.0192**	3.2783*	4.0237	3.8309	18.3322***	21.2561***
USDT	gVaR	8	16	2.2333	0.2401	2.5236	1.4229	23.5146***	20.3556***
	mVaR	7	15	1.1997	0.5511	1.4214	1.5882	17.4655***	22.1622***
USDC	gVaR	4	15	0.0582	0.5511	0.1302	1.5882	0.7280	2.1920
	mVaR	2	15	1.7702	0.5511	1.7881	1.5882	1.4751	2.2018
DGX	gVaR	5	11	0.0541	3.2783*	0.1667	4.5428	8.0133	11.1270*
	mVaR	2	11	1.7702	3.2783*	1.7881	4.5428	1.4545	10.7191*

References

1. Baur, D.G., McDermott, T.K.: Is gold a safe haven? International evidence. J. Bank. Finan. **34**(8), 1886–1898 (2010)
2. Christoffersen, P.F.: Evaluating interval forecasts. Int. Econ. Rev. **39**(4), 841–862 (1998)
3. Čuljak, M., Tomić, B., Žiković, S.: Benefits of sectoral cryptocurrency portfolio optimization. Res. Int. Bus. Finan. 101615 (2022)
4. Engle, R.F., Manganelli, S.: CAViaR: conditional autorregressive value at risk by regression quantiles. J. Bus. Econ. Stat. **22**(4), 367–381 (2004)
5. Favre, L., Galeano, J.-A.: Mean-modified value-at-risk optimization with hedge funds. J. Altern. Invest. **5**(2), 21–25 (2002)
6. Kupiec, P.: Techniques for verifying the accuracy of risk measurement models. J. Deriv. **3**(2) (1995)
7. Pun, C.S., Ye, Z.: Optimal dynamic mean-variance portfolio subject to proportional transaction costs and no-shorting constraint. Automatica **135**, 109986 (2022)

8. van der Weide, R.: GO-GARCH: a multivariate generalized orthogonal GARCH model. J. Appl. Economet. **17**(5), 549–564 (2002)
9. Wang, G.-J., Ma, X.-Y., Wu, H.-Y.: Are stablecoins truly diversifiers, hedges, or safe havens against traditional cryptocurrencies as their name suggests? Res. Int. Bus. Finan. **54**, 101225 (2020)

Interbank Networks and Liquidity Risk

Marina Dolfin[1,2(✉)], Leone Leonida[1,3], and Eleonora Muzzupappa[1,3]

[1] King's Business School, King's College London, London, UK
marina.dolfin@kcl.ac.uk
[2] Department of Engineering, University of Messina, Messina, Italy
[3] DES, University of Messina, Messina, Italy

Abstract. The implementation of Basel III introduces new capital requirements for liquidity risk that build on the Liquidity Coverage Ratio (LCR) and the Net Stable Funding Ratio (NSFR). We adopt a non-homogeneous Markov model framework to study liquidity dynamics on a simulated interbank network and test whether the implementation of the new regulation allows for efficient networks. The model simulates the effect of two different policies on the interbank network efficiency.

Keywords: Liquidity · Interbank network · Network efficiency

1 Introduction

Financial markets and institutions are truly interconnected to each other, both withing a certain economy and at the international level. The Global Financial Crisis (GFC) has highlighted the need for straightening the international regulatory standards, to cope with financial crashes at the global level. The unprecedented lack of banks' liquidity, which has characterized the early stage of the GFC, led to the introduction in Basel III framework (2010) of the novelty involving capital requirements for liquidity risk. The two liquidity ratios aim at a better management of the maturity mismatch between assets and liabilities, for banks to survive liquidity pressures: the Liquidity Coverage Ratio (LCR), focuses on a bank's ability to survive a 30-day period of liquidity disruptions and the Net Stable Funding Ratio (NSFR), which focuses on liquidity management over a period of one year. To this aim the European Central Bank has set up Task Force on Systemic Liquidity (TFSL) for policy responses to liquidity risk. This paper aims at studying the impact of the LCR on the efficiency of the interbank networks. Drawing from [1], we study this relationship by exploring the interrelationship with network topologies, using a different modelling framework in the context of dynamically evolving networks.

Regulatory Requirement on the LCR. The LCR requires that the amount of unencumbered High-Quality Liquid Assets (HQLA) (i.e. central bank reserves, sovereign bonds) be at least as large as the net outflow of funds under the 30-day stress scenario [5], $LCR = \frac{HQLA}{Net\,Cash\,Outflow} \geq 100\%$ where Net Cash Outflow (NCOF) stands for the total net cash outflow over the next 30 calendar days over a stress period.

2   A Model of Liquidity Dynamics on an Interbank Network

We model an economy as being composed by a large amount of homogeneous agents willing to have a buy and sell financial transaction involving the need to ask for a bank's loan. Each node of the network is a bank and a link is created between the bank that issues the loan and the bank where the account of the seller is settled. As a result of the financial transaction, whenever it occurs, liquidity flows from the bank from the bank that issues the loan to the bank where the seller's account stands and this flow results in a network link. The model builds on three concepts: a liquidity profile that, for each bank, varies over time, i.e. each node of the network; a liquidity dynamics on each node and, finally, a network dynamics. Let's start by introducing the liquidity profile. As far as the liquidity profile is concerned, for each bank assets are clusterized into different liquidity classes, ranging from the lowest to the highest liquid assets (central bank reserves or sovereign bonds for instance). The percentage of asset in the highest classes are considered as belonging to the subcluster of HQLA [5] for the purpose of computing the LCR.

$$LCR_r(t) = \frac{HQLA_r(t)}{NCOF_r}, \qquad r = 1, \ldots, M, \tag{1}$$

where $HQLA_r(t)$ is the percentage of high-quality liquid assets at time t of the r-th bank of the network and $NCOF_r$ is the 30-days Net Cash Outflow of the r-th bank and it is considered to be fixed over the time interval under observation; finally M is the number of banks.

We introduce now the liquidity dynamics on each node. As for the liquidity dynamics on each node, agents interact with each other by means of financial transactions possibily involving the issue of a loan. The agents are clusterised by means of the banks where their bank account is based. For simplicity we assume that each individual 'belongs' to just one bank. The stylised dynamics that is modelled is the following. The financial transaction is finalised if and only if the bank issues the loan and the bank issues the loan only if the LCR requirement is satisfied at that time. There are two possible outcomes:

1. The loan is issued and the payoff is stochastic. It might in fact result in a shift toward a higher liquidity class with probability α (depending on the interest rate) for the bank of the individual selling the asset and thus obtaining money and a shift towards a lower liquidity class for the bank of the individual acquiring the asset (with the same probability α for simplicity). This dynamics can be tought in terms of pairwise interaction dynamics between agents resulting in a winner and a loser, where the shift toward the higher class is the 'win' outcome (for an example in a population dynamics see [6]).
2. The loan is not issued (because the LCR requirement is not satisfied). In this case there are no payoffs for both the two nodes involved in the financial transaction.

The previous modelization assumptions result in the following system of finite-differences equations [6,7]:

$$f_{ij}(t + \Delta t) = f_{ij}(t) + \Delta t \Big(\sum_{k,l} \eta_{kl} f_{ij}(t) f_{il}(t) \psi_{kli} - f_{ij}(t) \sum_{l} \eta_{il} f_{il}(t) \Big) \qquad (2)$$

- f_{ij} probability that an entity in the 'liquidity status' x_j at time t, takes the status x_i at $t + \Delta t$;
- η_{kl} probability of the occurrence of financial transactions of an entity in the status x_k with an entity in the state x_l;
- the Markovian transition mass function ψ_{klii} is the probability that an entity in the status x_k at time t takes the status x_i at $t + \Delta t$ due to a financial transaction with an agent in the status x_l in the interval $t, t + \Delta t$.

The time-variation in the parameters of the Markov process is governed by a discrete-valued latent stochastic process with limited memory i.e. the current state is characterized only by the state from the previous period and the transition matrix [7]. Markovian transition probabilities are defined according to the previously introduced model specifications.

The initial condition is defined by each bank's liquidity profile at the initial time. A specific example is given in the numerical simulation in the next section. As a result of the dynamics the liquidity profile for each bank changes dynamically during time due to the stochastic interactions outlined by the model.

The model is completed by defining the network dynamics. The same dynamics that might determine the change in the liquidity profiles of the banks involved in the financial transactions determine the network dynamics as explained in the following. Recall that if two individuals are involved in a financial transaction the model consider that the financial transaction is finalised if and only if the bank issues the loan and the bank issues the loan only if the LCR requirement is satisfied at that time. Form the network dynamics perspective, there are then two possible outcomes.

1. The loan is issued and the link between the two involved banks is created if not already existing or its weight increases if already existing.
2. The loan is not issued (because the LCR requirement is not satisfied) and the link between the two involved banks is not created if not existing already or its weight is lowered if already existing.

3   Numerical Simulations with Diagnostic of Network Efficiency

By using the model above introduce, we conducted two different simulations in order to observe the impact of regulations requirements with different specification on the network efficiency. In this case network efficiency is quantified by using centrality measures [8,9], specifically aggregate centrality.

We compare the effect on the time evolution of aggregate centrality in two different scenarios, in order to exploit possible impact of different policy measures. Both scenarios are simulated using 100 banks, i.e. a network with 100 nodes. For both scenarios we choose an initial core-pheriphery network structure. The initial liquidity profiles on each bank have been chosen as randomly drawn from a normal distribution, using 1000 replications, with mean equal to 6 and different standard deviations, spanning from 1 to 2 (with values equally spaced). The constant NCOF is different for each bank (see (Eq. 1) and have been chosen such that the initial liquidity coverage ratios span over an interval from 80% to 125%. We have chosen 9 liquidity classes spanning from the one with the lowest liquidity quality up to the one with the highest and the 3 highest classes in terms of liquidity are classified as HQLA (Fig. 1).
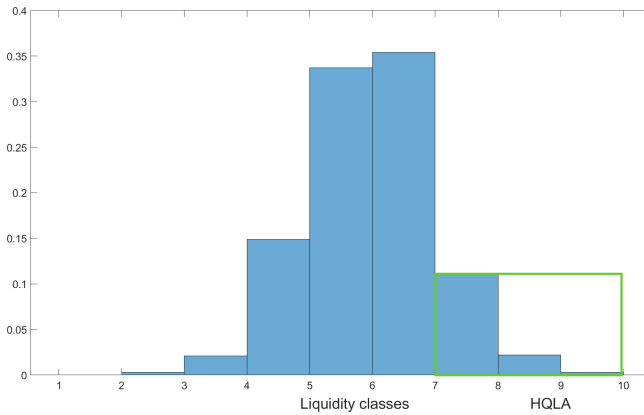
Fig. 1. A bank's liquidity profiles with 9 different liquidity classes and 3 classes classified as HQLA

For the second scenario the same conditions on the initial network topology and the initial liquidity profiles for each node as in the first scenario apply. The only difference consists in including a penalty for banks holding excesses of liquidity reserves. The penalty has been introduced by assuming that whenever $LCR > 115\%$ for a bank at a given time step, the NCOF is increased of a 5%. We report the time evolution of the aggregate centrality in the two different scenarios (Fig. 2):
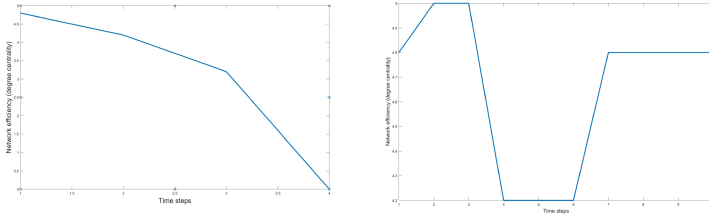
Fig. 2. Time evolution of the aggregate centrality in the first scenario (left) and in the second scenario (right).

where network efficiency is clearly falling down. All else equal, the time evolution of the aggregate centrality in the second scenario where bank are penalized for holding excesses of liquidity is clearly stabilizing the network efficiency around the initial value.

4   Conclusions and Research Perspectives

This paper presents a stylised model of liquidity dynamics on an interbank network when a specific regulatory policy is implemented in order to show how interconnectedness of institutions crucially impact network efficiency.

We explore a research perspective by considering on any node that the probability characterizing the fact that it is forced to delete a link might increase due to an increment in environmental volatility [10]. The probability has been evaluated by using a binomial distribution starting with $p = 0$ and increasing up to $p = 1$ (with steps 0.04).
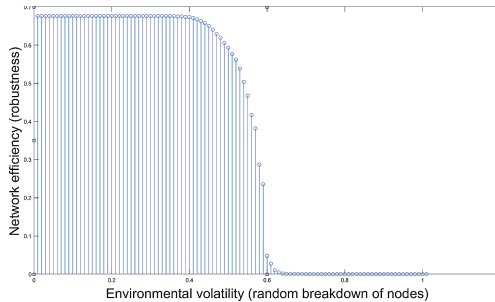
The increase in environmental volatility linked to the loss of links, determines a breakpoint in network efficiency.

References

1. Ardekani, A.M., Distinguin, I., Tarazi, A.: Do banks change their liquidity ratios based on network characteritics? Eur. J. Oper. Res. **285**, 789–803 (2020)
2. Eisenberg, L., Noe, T.H.: Systemic risk in financial systems. Manag. Sci. **47**(2), 236–249 (2001)
3. Elliot, M., Golub, B., Jackson, M.O.: Financial Networks and contagion. Am. Econ. Rev. **104**(10), 3115–3153 (2014)
4. ECB Task Force on Systemic Liquidity: Systemic liquidity concept, measurement and macroprudential instruments. Occasional Paper Series. ECB **214**, 1–68 (2018)
5. Basel Committee on Banking Supervision (Bank for International Settlemets). Basel III: The Liquidity Coverage Ratio and liquidity risk monitoring tools **214**, 1–66 (2013)
6. Jäger, E., Segel, L.A.: On the distribution of Dominane in populations of Social Organisms. SIAM J. Appl. Math. **52**, 1442–1468 (1992)
7. Kolokoltsov, V.: Nonlinear Markov Processes and Kinetic Equations. Cambridge Univ. Press, Cambridge (2010)
8. Gallegati, M., Kirman, A.: Reconstructing economic: agent-based models and complexity. Complex. Econ. **1**, 5–31 (2012)
9. Li, Y., Liu, G., Pin, P.: Network-based risk measurements for interbank systems. PLoS ONE **13**(7), 1–18 (2018)
10. Schweitzer, F., Fagiolo, G., Sornette, D., Vega-Redondo, F., Vespignani, A., White, D.R.: Economic networks: the new challenges. Science (Perspective) **325**, 422–425 (2009)

Kendall Conditional Value-at-Risk

Fabrizio Durante[1], Aurora Gatto[1], and Elisa Perrone[2(✉)]

[1] Università del Salento, Centro Ecotekne, S.P. 6, Lecce - Monteroni, Lecce, Italy
{fabrizio.durante,aurora.gatto}@unisalento.it
[2] Eindhoven University of Technology, Groene Loper 5, 5612AZ Eindhoven,
The Netherlands
e.perrone@tue.nl

Abstract. The Conditional Value-at-Risk (CoVaR) is a modified version of the Value-at-Risk (VaR) to quantify the risk of a random variable Y with respect to another random variable X. In this work, we consider a multivariate modification of CoVaR based on the Kendall distribution function. In particular, we discuss two possible hazard scenarios that generalize the standard CoVar and use the copula theory to derive the corresponding risk quantities. We consider a systemic risk exercise of the Italian banking system to demonstrate how the multivariate modification of CoVaR can be useful to analyze the resilience of a system when some parts of it are under distress.

Keywords: Copula · Systemic risk · Value-at-Risk

1 Introduction

The Conditional Value-at-Risk (CoVaR), introduced by Adrian and Brunnermeier [1], is a dependence-adjusted version of the Value-at-Risk (VaR) that represents the risk of a random variable Y with respect to another random variable X. The general idea behind CoVaR is to quantify the risk by using the conditional distribution of Y given that X is under distress. Various versions of CoVaR have been proposed in the literature depending on different definitions of the stress event: see, for instance [3,9,11].

In this work, we consider a multivariate modification of CoVaR, which has been proposed in [4] to take into account a stress event related to multiple random variables. Specifically, given a random vector $(\mathbf{X}, Y)$ that represents losses of a economic/financial system, we are interested in the VaR of Y given that $\mathbf{X}$ takes values in a stress scenario $S \subseteq \mathbb{R}^d$. Following the modified version of CoVaR presented for the first time in [9], we here consider that the probability of the event $\{\omega \in \Omega \colon X(\omega) \in S_\alpha\}$ is non-zero.

The CoVaR of Y given that $\mathbf{X} \in \mathcal{S}$ is formally defined as

$$\mathrm{CoVaR}_{\mathcal{S},\beta}(Y \mid \mathbf{X}) = \mathrm{VaR}_\beta(Y \mid \mathbf{X} \in \mathcal{S}_\alpha), \tag{1}$$

where $\beta \in (0,1)$ and $\mathbb{P}(\mathbf{X} \in \mathcal{S}) = 1-\alpha$; here, α and β belong to $\{0.90, 0.95, 0.99\}$. In other words, $\mathrm{CoVaR}_{\mathcal{S}_\alpha,\beta}(Y \mid \mathbf{X})$ can be seen as the Value-at-Risk VaR_β of

the conditional distribution of Y given that the realizations of $X_1, \ldots, X_d$ are from a hazard scenario $\mathcal{S}_\alpha$ with a specified probability size.

In the multivariate context there is no general consensus about how to define a stress scenario due to the lack of a total order in $\mathbb{R}^d$. Here, we adopt the approach based on the Kendall distribution function [8, 13, 14]. In our setting, the stress scenario is directly generated by the joint distribution function of $\mathbf{X}$ and the family of its level curves at some suitable levels. Compared with other stress scenario (like AND and OR scenarios), the advantage of the Kendall approach is that a single hazard region is associated to a given level α.

The paper is organized as follows. In Sect. 2, we revisit the notion of CoVaR based on Kendall hazard scenarios and show how the concept of copula can be helpful in deriving the risk quantities. In Sect. 3, we demonstrate the advantage of the proposed multivariate perspective to quantify the risk by considering a systemic risk exercise of the Italian banking systems.

2   The Kendall CoVaR

In this section, we present the notation and the mathematical description of the proposed multivariate approach. Given a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, let $(\mathbf{X}, Y) = (X_1, \ldots, X_d, Y)$ be a $(d+1)$–dimensional continuous random vector whose components represent losses. Consider the cumulative distribution function associated with $(\mathbf{X}, Y)$ as given by

$$F_{(\mathbf{X}, Y)}(x_1, \ldots, x_d, y) = \mathbb{P}(X_1 \le x_1, \ldots, X_d \le x_d, Y \le y),$$

which can be expressed as

$$F_{(\mathbf{X}, Y)}(x_1, \ldots, x_d, y) = C(F_{X_1}(x_1), \ldots, F_{X_d}(x_d), F_Y(y)),$$

where C is the copula of the random vector and F_{X_i}, $i = 1, \ldots, d$, and F_Y are the marginal distribution functions (see, for instance, [7]). Hereinafter, we denote by $F_{\mathbf{X}}$ the joint distribution function of $\mathbf{X}$.

Suppose that $d = 1$. Given a probability level $\alpha \in {]0, 1[}$ (e.g., $\alpha \in \{0.90, 0.95\}$), a stress scenario, also called hazard scenario [18], for X corresponds to the set $\mathcal{S} \subseteq \mathbb{R}$ of all realizations x such that the probability that X will exceed x is bounded from above by $1 - \alpha$, i.e.

$$\mathbb{P}(X \ge x) = \overline{F}_X(x) \le 1 - \alpha,$$

where $\overline{F}_X$ is the survival function associated with X. This can be equivalently expressed as

$$\mathbb{P}(X \le x) = F_X(x) \ge \alpha.$$

For continuous distribution functions F_X, we have that $\mathcal{S} = [v^*, +\infty]$, where v^* is the Value-at-Risk of X at level α, i.e. the α–quantile of X. Moreover, $\mathbb{P}(X \in \mathcal{S}) = 1 - \alpha$.

Since, for every $d > 1$, $F_{\mathbf{X}}$ and its associated survival function $\overline{F}_{\mathbf{X}}$ may not coincide, two generalizations are possible for a hazard scenario $\mathcal{S}$, namely

$$\mathcal{S}^{\check{K}} = \{\mathbf{x} \in \mathbb{R}^d \colon \overline{F}_{\mathbf{X}}(\mathbf{x}) \leq \check{t}\}, \tag{2}$$

with $\mathbb{P}(\mathbf{X} \in \mathcal{S}^{\check{K}}) = 1 - \alpha$, and

$$\mathcal{S}^{K} = \{\mathbf{x} \in \mathbb{R}^d \colon F_{\mathbf{X}}(\mathbf{x}) \geq t\}, \tag{3}$$

with $\mathbb{P}(\mathbf{X} \in \mathcal{S}^{K}) = 1 - \alpha$. These two stress scenarios are called, respectively, *survival Kendall* and *Kendall* hazard scenarios. We notice that:

– the distribution function of the random variable $\overline{F}_{\mathbf{X}}(\mathbf{X})$ is known as survival Kendall function (see, e.g., [13]).
– Analogously, the distribution function of the r.v. $F_{\mathbf{X}}(\mathbf{X})$, known as Kendall function (see, e.g., [8,17]).

An illustration of the two regions is presented in Fig. 1 by considering realizations from a bivariate Gumbel-Hougaard copula with standard exponential margins, respectively.

We now consider two different notions of multivariate CoVaR based on survival Kendall and Kendall hazard scenarios.

Kendall Scenario. First, we notice that the distribution function of Y given that $\mathbf{X} \in \mathcal{S}^{K}$ can be written as

$$\begin{aligned} F_{Y|\mathbf{X} \in \mathcal{S}^{K}}(y) &= \mathbb{P}\left(Y \leq y \mid F_{\mathbf{X}}(\mathbf{X}) \geq t\right) \\ &= \frac{F_Y(y) - \mathbb{P}\left(F_{\mathbf{X}}(\mathbf{X}) \leq t, Y \leq y\right)}{1 - \alpha}. \end{aligned}$$
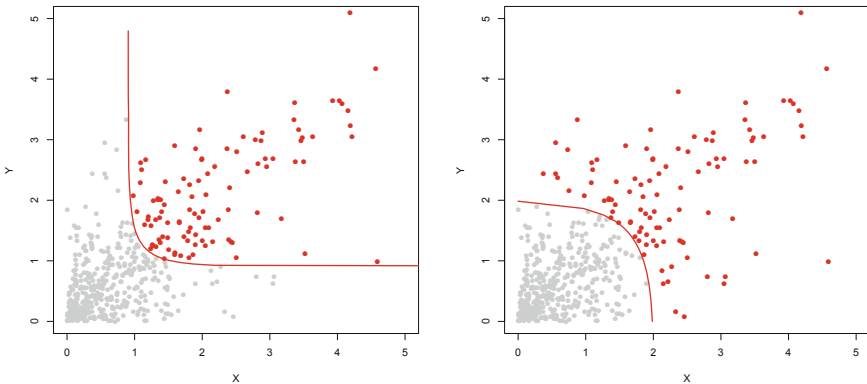
Fig. 1. Random sample of 500 points from a Gumbel-Hougaard copula model with standard exponential marginals. The red points indicate the Kendall hazard scenario (left) and the survival Kendall hazard scenario (right) with the respective level curve $\{F_{\mathbf{X}}(\mathbf{x} = t)\}$ and $\{\overline{F}_{\mathbf{X}}(\mathbf{x} = \check{t})\}$.

It follows that $\mathrm{CoVaR}_{\mathcal{S}^K,\beta}(Y \mid \mathbf{X} \in \mathcal{S}^K)$ is the value y such that

$$F_Y(y) - D_{(F_{\mathbf{X}}(\mathbf{X}),Y)}(\alpha, F_Y(y)) = (1-\alpha)\beta,$$

where D is the copula between the r.v.'s $F_{\mathbf{X}}(\mathbf{X})$ and Y.

Survival Kendall Scenario. We may write the distribution function of Y given that $\mathbf{X} \in \mathcal{S}^{\check{K}}$ as

$$\begin{aligned}
F_{Y \mid \mathbf{X} \in \mathcal{S}^{\check{K}}}(y) &= \mathbb{P}\left(Y \le y \mid \overline{F}_{\mathbf{X}}(\mathbf{X}) \le \check{t}\right) \\
&= \frac{\mathbb{P}\left(\overline{F}_{\mathbf{X}}(\mathbf{X}) \le \check{t}, Y \le y\right)}{1-\alpha} = \frac{D_{(\overline{F}_{\mathbf{X}}(\mathbf{X}),Y)}(1-\alpha, F_Y(y))}{1-\alpha},
\end{aligned}$$

where D is the copula between $\overline{F}_{\mathbf{X}}(\mathbf{X})$ and Y. Thus, $\mathrm{CoVaR}(Y \mid \mathbf{X} \in \mathcal{S}^{\check{K}})$ is the value y such that

$$D_{(\overline{F}_{\mathbf{X}}(\mathbf{X}),Y)}(1-\alpha, F_Y(y)) = (1-\alpha)\beta.$$

Summarizing, in both cases CoVaR can be calculated by using (a) the distribution function of Y, (b) the joint distribution function of $\mathbf{X}$, and (c) the copula between an aggregation of the elements of $\mathbf{X}$ and Y.

3 Illustration: Analysis of the Italian banking systems

In this section, we present a case study to illustrate the methodology introduced in Sect. 2. In this empirical analysis, we are interested in quantifying the systemic risk in the Italian banking system from a multivariate perspective. To this end, we consider log-returns of the daily prices of the ten most capitalized Italian banks listed on the FTSE MIB and FTSE Italia Mid Cap from March 15th, 2020 to November 30, 2021. The FTSE MIB is the primary benchmark Index for the Italian equity markets, and it is comprised of highly liquid, leading companies across ICB sectors in Italy. It includes the following eight banks that are analysed here: Intesa Sanpaolo (ISP.MI), Unicredit (UCG.MI), Finecobank (FBK.MI), Mediobanca (MB.MI), Banca Mediolanum (BMED.MI), Banco Generali (BGN.MI), Banco BPM (BAMI.MI) and Bper Banca (BPE.MI). Moreover, we add also Credem (CE.MI) and Monte dei Paschi di Siena (BMPS.MI) that are listed in the FTSE Italia Mid Cap.

The time series of the daily prices in EUR are obtained from the website Yahoo Finance (see https://it.finance.yahoo.com/). The database contains a total of 437 daily observations for each variable and no missing values. To focus on the right-side of the distribution, the relative log-returns are calculated with a negative sign.

To investigate the joint behaviour of the different time series we consider an ARMA-GARCH copula model (see, e.g., [15,16]) and estimate the marginals and the copula according to a two-stage procedure as illustrated in [10, section

6.2.3]. First, in order to describe separately the behavior of each time series, we estimate an appropriate ARMA(1,1)-GARCH(1,1) model with Student distribution for residuals. It allows to model the conditional mean and variance taking into account possible time-varying volatility patterns. Moreover, the obtained residuals show no strong evidence of heteroskedasticity and/or serial dependence according to Weighted Ljung-Box Test and ARCH LM Tests.

Once an appropriate ARMA-GARCH model has been estimated to each time series, the possible dependence relations among variable is investigated by means of special copula constructions built from bivariate copulas, i.e., vine copulas [2,5]. Specifically, we fit an R-Vine copula model on standardized residuals via maximum likelihood procedures as done in [6, section 4]. We use R package 'rvinecopulib', implemented in [12], to conduct estimation and model selection procedures for the vine copulas.

Once the desired joint model has been obtained for all the ten banks, we determine the predicted (one-step) CoVaR of each banking institution conditionally on a stress scenario involving all the other nine banks. The results are visualized in Fig. 2 for both the Kendall and the survival Kendall hazard scenarios. From the picture, we notice that the two approaches provide a qualitatively analogous representation of the systemic risk. In particular, it seems that the most suffering bank of the entire system is Monte Dei Paschi Di Siena (BMPS.MI), while Finecobank (FBK.MI) is the least suffering, conditioned by a stress scenario compared to all the other banks.
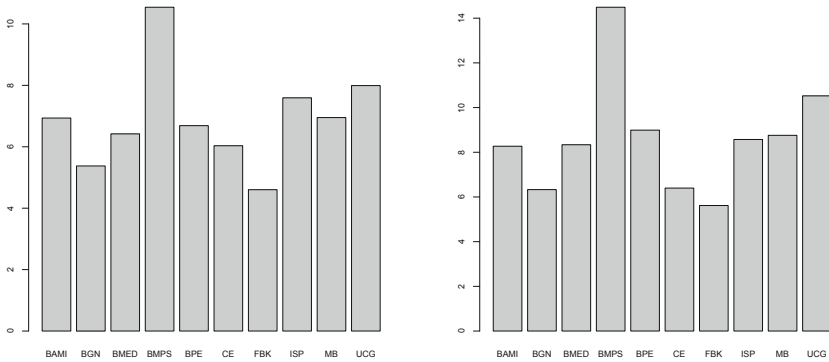
Fig. 2. CoVaR of one bank conditioned to a stress scenario of all the other banks according to Kendall hazard scenario (left) and the survival Kendall hazard scenario (right). Here, $\alpha = \beta = 0.95$.

References

1. Adrian, T., Brunnermeier, M.: CoVaR. Am. Econ. Rev. **106**(7), 1705–1741 (2016)
2. Bedford, T., Cooke, R.M.: Vines-a new graphical model for dependent random variables. Ann. Statist. **30**(4), 1031–1068 (2002)
3. Bernardi, M., Durante, F., Jaworski, P.: CoVaR of families of copulas. Statist. Probab. Lett. **120**, 8–17 (2017)
4. Bernardi, M., Durante, F., Jaworski, P., Petrella, L., Salvadori, G.: Conditional risk based on multivariate hazard scenarios. Stoch. Environ. Res. Risk Asess. **32**, 203–211 (2018). https://doi.org/10.1007/s00477-017-1425-9
5. Czado, C.: Analyzing Dependent Data with Vine Copulas. A Practical Guide with R, vol. 222. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-13785-4
6. Czado, C., Nagler, T.: Vine copula based modeling. Ann. Rev. Stat. Appl. **9**(1) (2022)
7. Durante, F., Sempi, C.: Principles of Copula Theory. CRC Press, Boca Raton (2016)
8. Genest, C., Rivest, L.P.: On the multivariate probability integral transformation. Statist. Probab. Lett. **53**(4), 391–399 (2001)
9. Girardi, G., Ergün, A.: Systemic risk measurement: multivariate GARCH estimation of CoVaR. J. Bank. Financ. **37**(8), 3169–3180 (2013)
10. Hofert, M., Kojadinovic, I., Maechler, M., Yan, J.: Elements of Copula Modeling with R. Springer Use R! Series, Springer, Heidelberg (2018). https://doi.org/10.1007/978-3-319-89635-9
11. Mainik, G., Schaanning, E.: On dependence consistency of CoVaR and some other systemic risk measures. Stat. Risk Model. **31**(1), 49–77 (2014)
12. Nagler, T., Vatter, T.: rvinecopulib: high performance algorithms for vine copula modeling (2021). R package version 0.5.5.1.1
13. Nappo, G., Spizzichino, F.: Kendall distributions and level sets in bivariate exchangeable survival models. Inform. Sci. **179**(17), 2878–2890 (2009)
14. Pappadà, R., Perrone, E., Durante, F., Salvadori, G.: Spin-off extreme value and Archimedean copulas for estimating the bivariate structural risk. Stoch. Environ. Res Risk Assess. **30**(1), 327–342 (2016). https://doi.org/10.1007/s00477-015-1103-8
15. Patton, A.J.: Modelling asymmetric exchange rate dependence. Int. Econ. Rev. **47**(2), 527–556 (2006)
16. Patton, A.J.: A review of copula models for economic time series. J. Multivar. Anal. **110**, 4–18 (2012)
17. Salvadori, G., De Michele, C., Durante, F.: On the return period and design in a multivariate framework. Hydrol. Earth Syst. Sci. **15**, 3293–3305 (2011)
18. Salvadori, G., Durante, F., De Michele, C., Bernardi, M., Petrella, L.: A multivariate copula-based framework for dealing with hazard scenarios and failure probabilities. Water Resour. Res. **52**(5), 3701–3721 (2016)

Daily Trading of the FTSE Index Using LSTM with Principal Component Analysis

David Edelman$^{(\boxtimes)}$ and David Mannion

University College Dublin, Carysfort Campus, Blackrock, County Dublin, Ireland
david.edelman@ucd.ie, david.mannion1@ucdconnect.ie

Abstract. This study comprises a preliminary investigation into the use of Long Short-Term Memory (LSTM) methodology when used in conjunction with Principal Component Analysis (PCA) for producing trading signals for daily returns of the the FTSE100 index. The model is trained on approximately 35 years of daily data and validated on six months of testing data, demonstrating a high degree of risk-adjusted trading efficacy.

Keywords: Deep learning · Recurrent networks · Time series · Ensembling

1 Introduction

The challenge of accurately predicting Time Series data, such as equity prices, bond prices and volatility levels is one of the most difficult in the finance industry. The Efficient Market Hypothesis (EMH) suggests that it is not possible, however. Over many years, there have been a vast amount of research carried out on how machine learning techniques might be applied to outperform classic time series forecasting techniques such as ARIMA and other methods.

Instead of trying to predict the returns themselves, the intention here is to find trading allocations (long or short) which optimise Optimal Growth (i.e., proceeding as if employing a logarithmic utility function). A combination of deep learning methods are used and compared in a stacked ensemble like approach in order to achieve this.

One of the main difficulties when modelling Financial Time Series data is the low signal-to-noise ratio. This makes it difficult for models to discern random fluctuations from meaningful signals that provide information on where the asset's price will go next. Many noise removal techniques have been tried over the years. For this study, a combination of truncating the returns between a certain threshold and performing a statistical transformation in the form of PCA are used in order to filter out noise while maximizing the amount of information passed to the models. As will be seen, this enables modelling of subtle relationships such as can be discovered by methods such as LSTM to be used much more effectively than many researchers have found previously in the financial forecasting context.

2   Related Work

2.1   Ensemble Methods

Meta Labelling. Meta-labelling [2] is a method whereby 'sub-learners' predict the direction of the trades, with a 1 representing a long position and a –1 representing a short position, similar to a binary classification task. These predictions, along with their corresponding probabilities are then fed into a secondary model (meta learner) as features. The Secondary model determines whether or not the transaction should take place, as indicated by a signal which can be either 0 if the meta learner determines the sub learner to be incorrect, or 1 if the sub learner deems the sub learners' prediction to be correct.

The advantage of this methodology is effectively a filter for the trading signal which has been generated: If the meta learner determines the sub learner was wrong, it does not take the opposite side of the trade, instead it takes up no position. This can be interpreted as a weak signal being present in the data at that point in time. Another advantage of this method is the output probabilities help give a confidence measure of the sub-learners predictions. When we do regression of (transformed) prices, these probabilities are not available and hence the meta learner cannot know how confident a model's prediction is.

2.2   Hybrid Methods

Hybrid methods, which combine statistical and deep learning methods have shown impressive results in time series forecasting problems. By combining these two methodologies, patterns in financial time series data can be more effectively captured due to the cancellation of errors and noise as a result of averaging. Smyl [11] used Exponential Smoothing and a LSTM, trained using the same gradient descent method to outperform deep learning and statistical methods in a Kaggle competition (M5) on time series forecasting.

2.3   Deep Learning Paradigms

The M5 competition hosted on kaggle.com has acted as an incubator for new approaches and techniques to time series forecasting. N-BEATS has shown to be an example of an ensemble based approach to time series modelling, being described as the state of the art algorithm for time series classification.

3   Model Architecture

3.1   Overview

Stacking has been shown in the past to reduce bias when used for many machine learning tasks, while Deep Learning methods such as RNNs have proven very successful in the past in modelling sequential data such as Natural Language and speech recognition problems [?]. This work aims to combine both of these

methods on Financial Time Series data, along with Principal Component Analysis (PCA) [4].

Here, a Deep Feedforward Network is applied to seven lags of transformed price return data and one lag of volume data, normalized using min-max scaling and fed into a sub learner, a Multi-Layered Perceptron. The sub-learner hyper-parameters are found using purged Cross-fold Validation. The sub-learner one-step predictions along with lag price return and volume return data are transformed using PCA and normalized using min-max scaling before being fed into the meta-learner.

The meta learner then predicts the optimal investment size to achieve log-optimal growth, based on the 'Merton Ratio' which is μ_i/σ_i^2 (equivalent to the first-order long-optimal investment amount). If σ_i is constant then the best size is proportional to μ_i, the forecast return for day i.

The Meta Learner generates one step predictions for each input sample, and the output of the meta learner is a vector of each of these one step predictions, concatenated together.

Three different Meta-Learners are considered: An MLP, Vanilla RNN and an LSTM.

3.2   Sub-Learners

Deep Multi-Layered Perceptron. A Deep Multi-Layered Perceptron with 4 hidden layers is used as the base learner. Each hidden layer gets passed through a ReLU activation function. For each of the 10 splits for generating predictions to be fed into the Meta Learner, a different model is fit. Each of these models is found by using Purged Cross fold validation with a gap of 7 for each split. For the portion of the data to be used as testing and validation data for the Meta Learner, Purged Cross fold validation is used again on the entire training dataset and the best model found is used to predict one step prediction values for these datasets.

3.3   Meta-learners

Recurrent Neural Network. Recurrent Neural Network (RNNs) are Neural Networks designed for sequential data which typically comes from areas such as speech recognition and translation, natural language processing and time series data. Classical networks such as the MLP are capable of handling sequential data, but have some limitations, such as the inability to handle sequences of variable length and to detect time invariant patterns in the data [5]. RNNs make use of recurrent connections which connect the hidden units in the network back to themselves with a time delay. RNNs' main pitfall in practice is the difficulty encountered in training them. A long sequence of data modelled by an RNN gives rise to a strong possibility of vanishing or exploding gradients [6], making it impossible to build an accurate model. Various alterations to RNNs, such as LSTM and GRU have been found useful in avoiding this issue.

Long Short Term Memory. Long short term memory (LSTM) [7] is a modified version of an RNN. It can remember both short and long term values and is capable of learning order dependence in sequence prediction problems. LSTMs have a forget gate, an input gate and an output gate. These gates are fully connected layers with weights that can be trained using well known backpropagation techniques. These gates in turn regulate the flow of information through the network.

4   Methods

4.1   Creating the Dataset

Raw Data. The raw data contains Closing Price (Fig. 1) and Volume data from the FTSE100 index, obtained from Yahoo! Finance Python API and covers a time period from 1985 until 2021.

Fig. 1. FTSE100 historical price

Feature Engineering. The first step is to identify any missing values in the dataset, and remove them, compute the logarithmic returns, and then truncate any extreme values, at approximately the 0.5% level.

Analogously, similar truncation is applied to volume data.

The dataset provided only contains volume data back as far as the year 2000, hence all data from before that period is not used in any of the model building or evaluation.

5   Experimental Setup and Evaluation

As is required with stacking, the dataset was partitioned, in this case into 10 folds. Additionally, the usual Training/validation/generalisation protocol was followed, all to ensure minimal overfitting. Data were preprocessed using

Principal Components and min-max scaling. The training criterion used was the quadtratic approximation to Optimal Growth Performance,

$$(m(x)y - 1)^2$$

where y denotes the target and $m(\cdot)$ the system's estimate of the optimal investment allocation.

An earlier draft of this article outlined the gridwise hyper-parameter optimisation which was used in the fitting of the LSTM model, but space does not allow inclusion here [interested readers are invited to contact the authors].

For simplicity, in order to assess performance in a risk-adjusted manner, we focus on the Sharpe-Ratio, namely the annualised ratio of investment return (which would have resulted from the model's investments) to its standard deviation, as measured on a daily basis. As the correlation of investments managed in this way tends to be extremely low, this relates directly to what is often referred to as 'portable alpha'.

6   Results

Table 1. Out of sample (test data) performance metrics

Description	(SL)MLP	(ML)MLP	(ML)RNN	(ML)LSTM
Sharpe ratio	0.547	0.708	**0.753**	**0.746**
Correlation	0.014	0.179	0.206	0.208
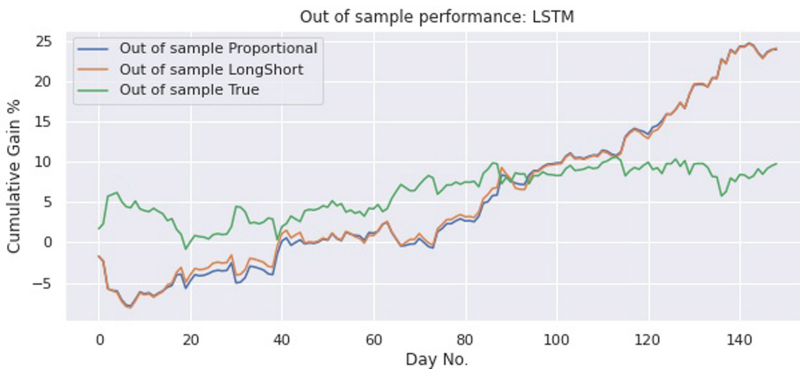
Fig. 2. Cumulative returns on the trading strategies generated by the LSTM vs actual cumulative return of the FTSE in the test period

The results show that the recurrent models (RNN and LSTM) provide the most profitable trading strategies (as well outperforming on the basis of a number of other metrics which are not included here for brevity's sake). What the results

also show is that the Meta-Learners techniques are significantly better in predicting the direction of the returns when compared to the sub learner MLP, being on average %20 better in terms of accuracy. More importantly, the meta-learners also produce trading strategies with much higher Sharpe Ratios, suggesting strong risk-adjusted performance.

7   Conclusion

Experimental results show that the methodology undertaken in this study show great potential for contributing to successful trading strategies. Further backtesting is essential before any strategy can be deployed for live trading, however. What this study appears to suggest is that LSTMs and RNNs do have the ability to accurately model financial time series data, but it is essential that the right kind of data is fed into them in order to achieve results better than random guessing. PCA helps these models filter out noise and hence helps to prevent over-fitting, which leads to better generalization ability. The custom loss function used also helps the model focus more on getting the direction of the bet right than traditional loss functions like MAE and RMSE.

References

1. Pelikan, M., Goldberg, D.E.: BOA: the Bayesian optimization algorithm. In: Proceedings of the 1st Annual Conference on Genetic and Evolutionary Computation, pp. 525–532 (1999)
2. de Prado, M.L.: Advances in Financial Machine Learning. Wiley, New York (2018)
3. Wolpert, D.H.: Stacked generalization. Neural Netw. **5**, 241–259 (1992). https://doi.org/10.1016/S0893-6080(05)80023-1
4. Wold, S., Esbensen, K., Geladi, P.: Principal component analysis. Chemometrics and intelligent laboratory systems. In: Proceedings of the Multivariate Statistical Workshop for Geologists and Geochemists, vol. 2, pp. 37–52 (1987). https://doi.org/10.1016/0169-7439(87)80084-9
5. Elman, J.L.: Finding structure in time. Cogn. Sci. **14**(2), 179–211 (1990)
6. Pascanu, R., Mikolov, T., Bengio, Y.: On the difficulty of training recurrent neural networks. In: Proceedings of the 30th International Conference on Machine Learning, vol. 28, pp. 1310–1318 (2012)
7. Hochreiter, S., Schmidhuber, J.: Long short-term memory. Neural Comput. **9**(8), 1735–1780 (1997)
8. Freund, Y., Schapire, R.E.: Experiments with a new boosting algorithm. In: International Conference on Machine Learning, pp. 148–156 (1996)
9. Jiang, M., Liu, J., Zhang, L., Liu, C.: An improved stacking framework for stock index prediction by leveraging tree-based ensemble models and deep learning algorithms. Phys. A Stat. Mech. Appl. **541**. https://doi.org/10.1016/j.physa.2019.122272
10. Sharpe, W.F.: The Sharpe Ratio. J. Portf. Manag. Fall **21**(1), 49–58 (1994). https://doi.org/10.3905/jpm.1994.409501
11. Smyl, S.: A hybrid method of exponential smoothing and recurrent neural networks for time series forecasting. Int. J. Forecast. **36**, 7585 (2020)

12. Weng, B., et al.: Macroeconomic indicators alone can predict the monthly closing price of major U.S. indices: insights from artificial intelligence, time-series analysis and hybrid models. Appl. Soft Comput. **71**, 685–697 (2018). https://doi.org/10.1016/j.asoc.2018.07.024

13. Pan, S.J., Yang, Q.: A survey on transfer learning. IEEE Trans. Knowl. Data Eng. **22**, 1345–1359 (2010). https://doi.org/10.1109/TKDE.2009.191

A Hybrid Model Based on Stochastic Volatility and Machine Learning to Forecast Log Returns of a Risky Asset

Lorella Fatone[1], Francesca Mariani[2(✉)], and Francesco Zirilli[3]

[1] Dipartimento di Matematica, Università di Camerino,
Via Madonna delle Carceri 9, 62032 Camerino, MC, Italy
lorella.fatone@unicam.it

[2] Dipartimento di Scienze Economiche e Sociali, Università Politecnica delle Marche,
Piazza Martelli 8, 60121 Ancona, AN, Italy
f.mariani@univpm.it

[3] Via F. Grossi Gondi 43, 00162 Rome, RM, Italy
apzrm1@gmail.com

Abstract. A hybrid model that combines a stochastic volatility model [2] and the K Nearest Neighbors (KNN) model [1] is proposed to obtain precision forecasts of log returns of a risky asset traded in the financial market. The precision forecasts are the sum of the forecasts obtained with the stochastic volatility model and a correction term produced by the KNN model. Numerical experiments based on real data are performed to investigate the accuracy of the precision forecasts.

Keywords: Precision forecast · Stochastic volatility model · Machine learning

1 Introduction

Recently the ubiquitous availability of big data has led to the massive use of data-driven models to forecast time series dynamics. In particular in mathematical finance and in the financial market practice non parametric Machine Learning (ML) models compete with parametric models to forecast prices or log returns of risky assets. The appeal of non parametric ML models comes from their computational efficiency and from their flexibility. However parametric models with their a priori assumptions about the dynamics of the quantities studied often are able to capture qualitative information that non parametric ML models are unable to grasp. In order to exploit the full potential of these alternative classes of models, in this paper it is developed a hybrid model that combines a non parametric ML model (i.e. KKN model [1]) and a partially specified stochastic volatility model (called optimal control model [2]) to forecast log returns of a risky asset traded in the financial market. The log return forecasts obtained with the hybrid model are called precision log return forecasts. The paper is organized as follows: Sect. 2 describes the hybrid model, Sect. 3 presents some numerical experiments based on real data.

2   The Hybrid Model

Let us introduce the optimal control model [2]. This is a partially specified stochastic volatility model that generalizes and, in some sense, enhances the Heston model [3].

The stochastic processes $S(t)$, $x(t) = \ln(S(t)/S(0))$, $v(t)$, $t \in [0, T]$, $T \in \mathbb{R}_+$, represent, respectively, the price, the log return and the stochastic variance of a risky asset traded in the financial market. The time interval $[0, T]$ is called investment period. The optimal control model assumes that the dynamics of $x(t)$, $v(t)$, $t \in [0, T]$, is defined through the following system of stochastic differential equations:

$$dx(t) = \left(f_1(t, x(t), v(t)) - \frac{1}{2}v(t) \right) dt + \sqrt{v(t)}dB(t), \qquad t \in [0, T], \qquad (1)$$

$$dv(t) = f_2(t, x(t), v(t))dt + \sigma\sqrt{v(t)}dZ(t), \qquad t \in [0, T], \qquad (2)$$

with the initial condition:

$$x(0) = 0, \quad v(0) = v_0, \tag{3}$$

where 0, v_0 are random variables concentrated with probability one in the points 0, $v_0 > 0$ and $\sigma \in \mathbb{R}_+$. Note that with abuse of notation 0, v_0 denote both the random variables of (3) and the points where these random variables are concentrated. In (1), (2) the stochastic processes $B(t)$, $Z(t)$, $t \in [0, T]$, denote Wiener processes, such that $B(0) = Z(0) = 0$, and their stochastic differentials $dB(t)$, $dZ(t)$, $t \in [0, T]$, are assumed to be such that:

$$\mathbb{E}^{\mathbb{P}}(dB(t)dZ(t)) = \rho dt, \quad t \in [0, T], \tag{4}$$

where $\rho \in (-1, 1)$ is a correlation coefficient and $\mathbb{E}^{\mathbb{P}}(\cdot)$ is the expected value of $\cdot$ with respect to the measure $\mathbb{P}$. The density of the probability measure $\mathbb{P}$ is the transition probability function of model (1), (2), (3), (4). Model (1), (2), (3), (4) is completed (when needed) with a reflecting condition to prevent the possibility that $v(t)$ becomes negative with positive probability, $t \in [0, T]$ (see [2]).

Note that when $f_1(t, x, v) = \mu$, $f_2(t, x, v) = k(\theta - v)$, $t \in \mathbb{R}_+$, $x \in \mathbb{R}$, $v \in \mathbb{R}_+$, $\mu \in \mathbb{R}$, k, $\theta \in \mathbb{R}_+$, the optimal control model (1), (2), (3), (4) coincides with the Heston model. However, in the optimal control model (1), (2), (3), (4) the analytic expression of the functions f_1, f_2 is not given. In this sense model (1), (2), (3), (4) is partially specified. The functions f_1, f_2 are determined in the calibration of model (1), (2), (3), (4).

In [2] when σ, ρ, v_0 are given (or determined from a set of data) the calibration problem of model (1), (2), (3), (4) is formulated as a stochastic optimal control problem that depends from a set of observed asset log returns and from three priors chosen by the investor determined as solution of the optimal control problem. Maximizing the objective function of this optimal control problem corresponds to pursue the following goals: i) the asset log returns implied by the model fit the observed asset log returns, ii) f_1 fits a reference drift rate,

iii) f_2 fits a reference drift, iv) at the end of the investment period the asset log return implied by the model fits a reference final asset log return. Two of these priors (i.e. the reference asset price drift rate and the reference stochastic variance drift) are chosen as suggested by the Heston model. The third prior is determined with a fitting procedure [2]. The previous goals are weighted in the objective function of the control problem. The choice of the priors made suggest that the optimal control model can be interpreted as a refinement of the Heston model. The unknown functions f_1 and f_2 are determined as solution of the optimal control problem.

In the numerical experiments the data set used in the calibration of the optimal control model is a set of log return data observed at known observation times. After being calibrated the optimal control model is used to forecast the asset log returns. The optimal control model log return forecast at a time greater than the observation time of last available calibration datum is the expected value of the log return of the calibrated optimal control model at the forecasting time conditioned to the last available log return observation. From the log return forecasts it is easy to compute the asset price forecasts. In the numerical experiments σ, ρ, v_0 are deduced from a suitable set of data.

Let us introduce the KNN correction term that must be added to the optimal control model log return forecast to obtain the precision log return forecasts. First of all let us define the asset price normalized residual time series. Let N be a positive integer, $0 \leq \tau_1 < \tau_2 < \ldots < \tau_N$ be time values such that $\tau_i - \tau_{i-1} = \Delta > 0$, $i = 2, 3, \ldots, N$, $x^{oc}(\tau_i)$, $S^{oc}(\tau_i)$, $S^{obs}(\tau_i)$, be, respectively, the log return and the asset price optimal control model forecasts and the observed asset price at time τ_i, $i = 1, 2, \ldots, N$. The asset price normalized residual time series $\{e(t), \ t = \tau_1, \tau_2, \ldots, \tau_N\}$ is the time series of the relative differences between the optimal control model forecasts and the observed asset prices, that is:

$$e(\tau_i) = \frac{S^{oc}(\tau_i) - S^{obs}(\tau_i)}{S^{obs}(\tau_i)}, \quad i = 1, 2, \ldots, N. \tag{5}$$

The KNN model is used to forecast the asset price normalized residual time series. The KNN model is based on the idea that similar time series in the past will have similar future behaviour. In particular, given the positive integers M, K, H, the KNN model collects a set of past observations (i.e. the training set of the model) that is the set of the vectors made of M consecutive outcomes of the time series under investigation (i.e. the training instances) and to each training instance associates a target vector of H values that follow in the time series the training instance considered. Given a new vector of outcomes the KNN model looks for the K training instances that are closest to the new vector in a given metric (the K nearest neighbors of the new vector in the training set) and forecasts the H future outcomes of the new vector in the time series as a linear combination of the target vectors associated to the K nearest neighbors. The vector of the H future values forecasted is the KNN model forecast associated to the new vector considered.

To tune the KNN model to the data studied it is necessary to choose the number of neighbors K used in the forecast, the dimension of the training instance

vectors M, the dimension of the target vectors H and the coefficients of the linear combination that produces the KNN forecast. The integers K, M, H and the linear combination coefficients are called hyperparameters of the KNN model and must satisfy some obvious constraints to make possible the production of forecasts.

Let $e_i = e(\tau_i), i = 1, 2, \ldots, N$, be the asset price normalized residual time series, at time τ_N we forecast the H consecutive normalized residuals that follow e_N choosing as new vector of the KNN model the last M consecutive normalized residuals observed at time $t \leq \tau_N$, that is the vector $(e_{N-M+1}, e_{N-M+2}, \ldots, e_N)$, and as training set of the KNN model the set of consecutive normalized residual vectors of dimension M that "end" before time τ_{N-M+1}. Finally we forecast the normalized residual at times $t = \tau_{N+1}, \tau_{N+2}, \ldots, \tau_{N+H}$, where $\tau_{N+i} = \tau_N + i\Delta$, $i = 1, 2, \ldots, H$, computing the arithmetic mean of the target vectors that follow the K-nearest neighbor vectors selected by the KNN model.

The precision log return forecasts produced by the hybrid model are obtained adding to the optimal control model log return forecasts a correction term easily deduced from the KNN asset price normalized residuals forecasts.

3   Numerical Experiments

Let us study the performance of the hybrid model introduced in Sect. 2 in a numerical experiment based on real data. The data considered are the daily closing values of the S&P500 (Standard & Poor 500) index of the N.Y.S.E. (New York Stock Exchange) relative to the period August, 25-th, 2008 – August, 23-th, 2021. In order to show the improvement in accuracy obtained using the hybrid model instead of the optimal control model we compare the accuracy of the precision forecasts with that of the optimal control model forecast as a function of the forecast horizon. We consider as forecast horizons: κ months in the future, that is κ months after the observation time of the last available datum, $\kappa = 1, 2, \ldots, 6$.

In what follows we assume year, month and week made, respectively, by 252, 21 and 5 (consecutive) trading days. Using a rolling window of the daily closing values of the S&P500 index relative to a time period of 30 consecutive months, we calibrate the optimal control model once a week and we forecast the log returns κ months in the future, $\kappa = 1, 2, \ldots, 6$. This is done 504 times during a period of 504 weeks, that is a period of 10 years. Note that with our definitions 504 weeks $= 2520$ (trading) days $= 10$ years. Denoting by $t = 0$ the time corresponding to August, 25-th, 2008, at calibration time $t = t_i = (630 + (i-1)5)/252$ years we calibrate the optimal control model using as data the daily closing values of the prices of S&P500 relative to the time window $[t_i - 630\frac{1}{252}, t_i]$, for $i = 1, 2, \ldots, 504$. Note that the last calibration day t_{504} corresponds to February, 23-th, 2021. The calibrated optimal control model at time t_i is used to forecast the asset log returns κ months in the future (i.e. κ months after time t_i), $\kappa = 1, 2, \ldots, 6$, $i = 1, 2, \ldots, 504$. The normalized residual with forecast horizon κ time series defined as $e_{\kappa,i} = e(t)$, where $t = t_i + \kappa\frac{21}{252}$, for $i = 1, 2, \ldots, 504$, is computed,

$\kappa = 1, 2, \ldots, 6$. Note that the time of last forecast depends on the forecasting horizon and corresponds to March, 24-th, 2021, April, 23-th, 2021, May, 24-th, 2021, June, 23-th, 2021, July, 23-th, 2021 and August, 23-th, 2021, respectively, when $\kappa = 1, 2, \ldots, 6$.

Passing from a calibration time t_i to the successive calibration time t_{i+1} we discard the first five daily closing prices belonging to the calibration time window associated to t_i and we insert the five daily closing prices following t_i, $i = 1, 2, \ldots, 504$. In this way for the forecast horizon κ a time series of 504 normalized residuals, given by $\{e_{\kappa,i} \mid i = 1, 2, \ldots, 504\}$ is obtained, $\kappa = 1, 2, \ldots, 6$. We use the KNN model to correct the last 300 outcomes of these time series. For this purpose we choose the hyperparameters of the KNN model as follows: $K = 2$, $M = 3$ and $H = 1$, that is, given a new vector of dimension $M = 3$, we consider the first $K = 2$ closest neighbors, training instances of dimension $M = 3$ and a target vector of dimension $H = 1$. The KNN forecast is obtained as the arithmetic mean of the target vectors of the two nearest neighbors selected. The hyperparameters are chosen with an elementary trial and error procedure. For the forecast horizon κ months, for $i = 203, \ldots, 504$, we use the set of the vectors made by the $M = 3$ consecutive normalized residuals belonging to the set $A_{\kappa,i} = \{e_{\kappa,j} \mid j = 1, 2, \ldots, i\}$ as training set and we forecast the next $(H = 1)$ normalized residual using the KNN model, $\kappa = 1, 2, \ldots, 6$. We denote by $\hat{e}_{\kappa,i}$ the precision forecast normalized residual at time $t = t_i + \kappa \frac{21}{252}$ relative to the forecast horizon κ months, $i = 204, 205, \ldots, 504, \kappa = 1, 2, \ldots, 6$. From the normalized residual forecast it is immediate to obtain the corresponding KNN correction term forecast that added to the optimal control model log return forecast gives precision log return forecast produced by the hybrid model.
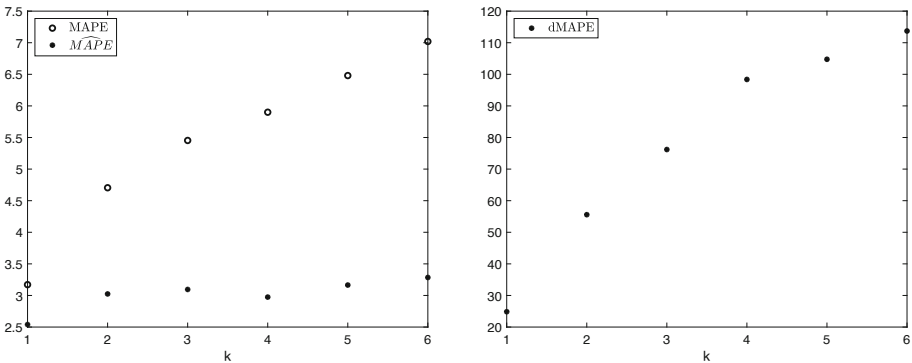
Fig. 1. $MAPE_\kappa$ and $\widehat{MAPE}_\kappa$ (left panel) and $dMAPE_\kappa$ (right panel) as a function of the forecast horizon κ months.

To compare the improvement in accuracy obtained using the precision forecasts instead of the optimal control model forecasts as a function of the forecast

horizon κ months we compute the relative difference between the mean absolute percentage error (MAPE) of the optimal control model forecast normalized residuals and the MAPE of the precision forecast normalized residuals ($\widehat{MAPE}$), $\kappa = 1, 2, \ldots, 6$. The $MAPE_\kappa$ of the optimal control model forecast normalized residuals is computed as the average of the absolute values of the optimal control model forecast normalized residuals relative to the forecast horizon κ months, $\kappa = 1, 2, \ldots, 6$, that is:

$$MAPE_\kappa = \frac{1}{300} \sum_{i=204}^{504} |e_{\kappa,i}|, \quad \kappa = 1, 2, \ldots, 6. \tag{6}$$

Analogously we compute the $\widehat{MAPE}_\kappa$ of the precision forecast normalized residuals as the average of the absolute values of the precision forecast normalized residuals relative to the forecast horizon κ months, $\kappa = 1, 2, \ldots, 6$, that is:

$$\widehat{MAPE}_\kappa = \frac{1}{300} \sum_{i=204}^{504} |\hat{e}_{\kappa,i}|, \quad \kappa = 1, 2, \ldots, 6. \tag{7}$$

In Fig. 1 in the left panel we show the values of $MAPE_\kappa$ and $\widehat{MAPE}_\kappa$ as function of the forecast horizon κ months, $\kappa = 1, 2, \ldots, 6$. In Fig. 1 in the right panel we show the relative difference between $MAPE_\kappa$ and $\widehat{MAPE}_\kappa$: $dMAPE_\kappa = (MAPE_\kappa - \widehat{MAPE}_\kappa)/\widehat{MAPE}_\kappa$ as a function of the forecast horizon κ months, $\kappa = 1, 2, \ldots, 6$. The $dMAPE_\kappa$ tells us how much better (in absolute percentage terms) or worse the precision forecasts are than the optimal control model forecasts when the forecast horizon κ months is considered, $\kappa = 1, 2, \ldots, 6$. Observing Fig. 1 we can conclude that the improvement in accuracy of the precision forecasts with respect to the optimal control model forecasts increases as the forecast horizon increases.

References

1. Altman, N.S.: An introduction to kernel and nearest-neighbor nonparametric regression. Am. Stat. **46**(3), 75–185 (1992). https://doi.org/10.1080/00031305.1992.10475879
2. Fatone, L., Mariani, F., Zirilli, F.: Calibration in the "real world" of a partially specified stochastic volatility model (2021, to be published)
3. Heston, S.: A closed-form solutions for options with stochastic volatility with applications to bond and currency options. Rev. Financ. Stud. **6**, 327–343 (1993). https://doi.org/10.1093/rfs/6.2.327

Financial Time Series Classification by Nonparametric Trend Estimation

Giuseppe Feo$^{(\boxtimes)}$, Francesco Giordano, Marcella Niglio,
and Maria Lucia Parrella

University of Salerno, 84084 Fisciano, SA, Italy
{gfeo,giordano,mniglio,mparrella}@unisa.it

Abstract. This work considers the classification of financial nonstationary time series, where the nonstationarity is due to the presence of a deterministic trend. It is evaluated in a high-dimensional context by looking at the first derivative of the trend function and without requiring a pre-specified form. This is achieved by means of a nonparametric estimator which is used in a two stage procedure: the first stage selects the time series with no trend and the second stage focuses the attention on nonlinear trends. A real data application to US Mutual Funds is conducted to demonstrate the validity and applicability of the procedure.

Keywords: High-dimensionality · Nonparametric regression · Screening procedure · Mixing processes · Financial time series

1 Introduction

Time series trend composition is a very important topic in data analysis. Checking trend composition is the first step for a further statistical analysis conducted on a time series. In the recent literature, there is an increasing interest to automate this process, see [10] and [1] among the others. In general, one can be interested in checking if the trend is absent, linear or nonlinear, but actually, the true structure of the trend is unknown, then a procedure that automatically allows this distinction is necessary before any further analysis. Suppose to observe p independent time series of the form

$$Y_{it} = m_i(t/T) + \varepsilon_{it}, \quad i = 1, \ldots, p; t = 1, \ldots, T \tag{1}$$

where p may go to infinity as a function of the time series length T, $m_i : [0, 1] \to \mathbb{R}$ are unknown trend functions and $\{\varepsilon_{it}\}_{t=1}^{T}$ are zero mean, strongly mixing error processes [11]. In particular, two classic examples of strongly mixing processes are ARMA and GARCH processes [2]. To classify the mentioned time series according to their trend composition (absent, linear or nonlinear), one can estimate, instead of the trend, its first derivative. This can be done by using a nonparametric estimator. The use of a nonparametric first derivative estimator has at least two main advantages: (i) on the mathematical point of view, the use

of the first derivative to highlight the linearity of a function is quite intuitive; (ii) one can assert if a trend is linear or not without imposing a predefined model. Given these two points, the rest of the paper is organized as follows: in Sect. 2 the proposed method for classifying the time series is presented; in Sect. 3 an application of the method is presented considering the US Mutual Funds; in Sect. 4 some conclusions and future developments are discussed.

2   The Proposed Method

Considering the representation (1) for a time series, the proposed nonparametric estimator for the trend first derivative at point $x \in [0, 1]$, has the form

$$\hat{\beta}(x) = \frac{1}{Th^2} \sum_{t=1}^{T} K_h(t/T - x)(t/T - x)Y_t, \qquad (2)$$

where $K_h(u) = \frac{1}{h} K\left(\frac{u}{h}\right)$, with $K(\cdot)$ a symmetric Lipshitz continuous kernel function with bounded support, and $h = h_T > 0$ is the bandwidth such that $h \to 0$ and $Th^4 \to \infty$ as $T \to \infty$. It can be proven that the proposed estimator, based on the guiding line of Local Polynomial estimator (see [3,5,7–9] among others) with fixed design, has the appealing characteristics that it is asymptotically normal distributed and its expected value is proportional to the true first derivative by a known quantity, as $T \to \infty$.

Under the reasonable assumption that the number of time series with nonlinear trend is finite, the proposed classification procedure consists of two stages. In the first, the proposed estimator $\hat{\beta}(x)$ is tested to be zero, by the following statistic

$$\hat{I}_\beta = \frac{T^{4/7}}{\mu_2^* c(\varepsilon)} \sum_{j=1}^{k_T} \hat{\beta}(x_j)^2, \qquad (3)$$

where $k_T = O(T)$ is the number of points considered into a subinterval of $[0, 1]$, $\mu_2^* = \int_{-1}^{1} u^2 K(u)^2 du$ and $c(\varepsilon) = \gamma_\varepsilon(0) + 2 \sum_{k=1}^{\infty} \gamma_\varepsilon(k)$, with $\gamma_\varepsilon(k) = Cov(\varepsilon_t, \varepsilon_{t-k})$. Once the sum of all autocovariancies of the process ε_t, $c(\varepsilon)$, is substituted by the nonparametric consistent estimator of the spectral density valued at frequency zero, say $\hat{c}(\varepsilon)$, $\hat{I}_\beta$ allows to distinguish the time series with no trend. Under the hypothesis that the time series has no trend, it can be proven that $\hat{I}_\beta$ follows a chi-squared distribution. This allows the p time series to be tested through the use of the Bonferroni correction. In the second stage, the difference between the estimator at different points is used in a screening approach [4] to make the further linear/non linear partition of the remaining time series from the previous stage. More precisely, defining $\hat{D}(x_1, x_2) = \hat{\beta}(x_1) - \hat{\beta}(x_2)$, where $x_1, x_2 \in (h, 1 - h)$, the statistic

$$\hat{I}_D(x) = \frac{1}{k_T} \sum_{j=1}^{k_T} D(x, x_j)^2 \qquad (4)$$

is used to rank the remaining time series. This ensures, with probability tending to 1, that one can estimate the set which contains the true set of time series with nonlinear trend under the sparsity assumption that the latter has a finite number of elements. Furthermore, the consistency results for both stages can be proven considering the rate $p = o\left(\frac{T^{1/2}}{\log T}\right)$ which also guarantees the Sure Screening Property [4].

In other words, the proposed method to classify time series according to their trend composition (absent, linear or nonlinear) consists of two stages: the first stage is used to select the time series with no trend by using a testing procedure based on the (3), while the second is a screening stage, based on the (4), which gives the set containing, with probability tending to 1, the true set of time series with nonlinear trend. The performance of the procedure has been evaluated trough a simulation study (not reported here) obtaining satisfactory results. We here evaluate the feasibility of the procedure to a large financial dataset whose details are given below.

3   Real Data Application

A Mutual Fund is a professionally managed investment fund that pools money from many investors to purchase securities. The typical classification of Mutual Funds is well known (i.e. by their principal investments or by found category) as the advantages of the latter: the opportunity for diversification, daily liquidity (quicker redemption of the net asset value) and the ability to participate in investments that may be available only to larger investors, among the others. Mutual funds have disadvantages as well. In particular, they have less predictable income and no opportunity to customize. A possible solution may be to create a portfolio of Mutual Funds. With this in mind, a preliminary classification on the long-term behaviour of the available funds is essential.

The Kaggle platform (https://www.kaggle.com) is an online community of data scientists and machine learning practitioners which allows users to find and publish data sets, explore and build models in a web-based data-science environment, work with other data scientists and machine learning engineers, and enter competitions to solve data science challenges. Among the many datasets, the one called "US Funds dataset from Yahoo Finance" was chosen. It contains 24821 daily closing prizes of US Mutual Funds scraped from the publicly available website https://finance.yahoo.com and updated to October 2021.

From this dataset, we randomly choose 200 time series and we consider the observations from 9 August 2018 to 12 October 2021, for a total of 800 observations. To make the series comparable, our classification procedure was applied on the transformed series of index numbers (i.e., the observed values of each time series are divided by the first value, observed on 9 August). After that, in order to make the assumption of independence between the time series as realistic as possible, a window of 500 time points is extracted from each time series, where the starting point of each window is uniformly selected, without replacement among the first 300 observations of the series (i.e., from 9 August 2018 to 4 June

2019). In this way, a "new" set of time series has been derived, where the 200 series start from different time points (but they essentially keep the same trend as before). Given this dataset, we apply our procedure to classify which series have no trend, linear and nonlinear trend. Finally, the above procedure is repeated 100 times, each one with different time windows, in order to derive the relative frequency of times each series is classified with "no trend", with "linear" or with "nonlinear" trend. To pursue this aim, the procedure is applied for all the 100 sets of time series using the Epanechnikov kernel $K(u) = \frac{3}{4} \max(0, 1 - u^2)$ with a FeedForward Neural Network estimator in order to obtain a plug-in estimator for the optimal bandwidths, as in [6], for each time series.

Given the results of the first stage, 100 possibly different sets of time series labelled as no trend can be obtained. It is reasonable to think that the time series with no trend will be those belonging to the intersection of all these sets. In other words, the time series with no trend will be those with relative frequency, over the 100 replicates, equal to 1. Considering now the second stage, it will produce, for each set of time series, a ranking based on (4) where the top positions include the time series with a nonlinear trend. In this way, a matrix can be created whose columns represent the 100 different ranking of the time series. To estimate the set of time series with nonlinear trend, it is reasonable to calculate the frequency with which the time series occur in the top positions of the ranking matrix. In particular, we define the following criterion. Let us consider a given threshold for the frequency, say α. For example, we set $\alpha = 0.5$. We then classify with "nonlinear trend" those series that appear in the top $\alpha\%$ positions at least $\alpha\%$ of the time. So, in our case, we are looking for those series which appear in the top 100 positions of the ranking matrix at least 50% of the time. Of course, the number of series reaching the second stage can be different for different iterations, so it may happen that this number is sometime less than 100 (so that the ranking matrix has less than 100 rows). However, this does not create problems for the aforementioned selection criterion.

Table 1. Cardinality of the classification sets obtained applying the procedure to the 200 index number time series.

♯ No trend	♯ Lin trend	♯ NoLin trend
63	123	14

The results of the overall procedure are shown in Table 1 which highlight that the majority of funds have a linear trend and only 14 funds have a non linear trend. This result is in line with the "nonlinear sparsity" assumption made in the previous section. Figure 1 shows how the procedure has classified 5 funds taken from the 200. The first two have no trend, the third has a nonlinear trend and, finally, the last two have a linear trend. Replicating the procedure on the adjusted closing prices, the number of series belonging to each class (no trend, linear, nonlinear) remains almost the same. For this reason the results are omitted but they are available from the authors.
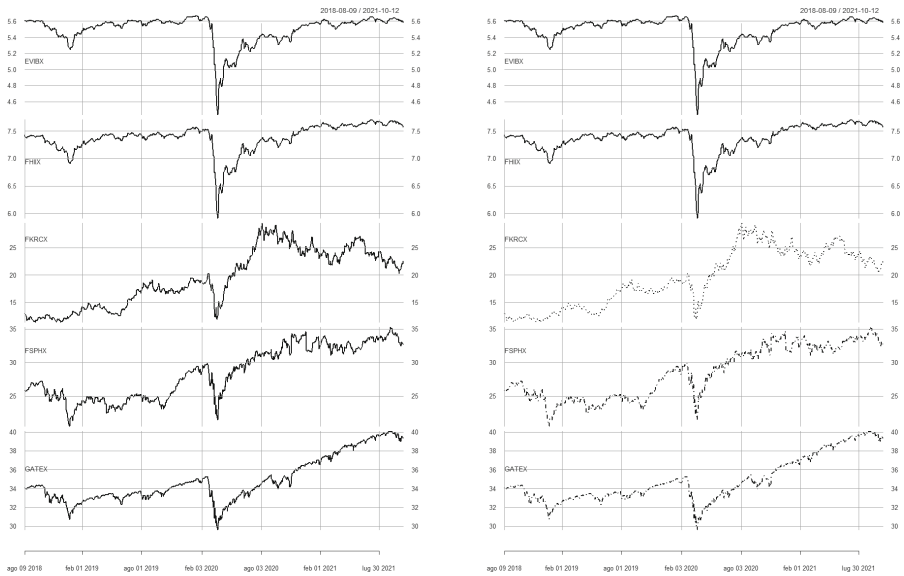
Fig. 1. On the left: 5 Mutual Funds (EVIBX, FHIIX, FKRCX, FSPHX and GATEX), from August 2018 to October 2021, before applying the procedure. On the right: the same 5 funds labelled according to the procedure. The first two are labelled to have no trend (solid line), the third to have a nonlinear trend (dotted line) and the last two to have a linear trend (dashed line).

4   Conclusions

In this work, a new procedure is presented as embryonic analysis for carrying out further analysis on time series. It deals with the classification of nonstationary time series, where the nonstationarity is given by the presence of a deterministic trend. It is based on the first derivative of the trend function in a context of high-dimensionality and without requiring a pre-specified form for the trend. This is achieved by using the proposed first derivative trend estimator $\hat{\beta}(x)$ which is based on the Local Polynomial estimator for fixed design and also presents the desirable characteristic of a simple form. Under the reasonable assumption that the number of time series with nonlinear trend is finite, the proposed procedure consists of two stages. In the first one, the proposed estimator, given in (3), is tested to be zero, which allows to distinguish the time series with no trend. In the second stage, the difference between the estimator at different points (see Eq. (4)) is used in a screening approach to make the further linear/nonlinear partition of the remaining time series from the previous stage, by means of a sort of stability selection of the number of time series with non linear trend. A real data application to 200 US Mutual Funds is conducted to demonstrate the validity and applicability of the procedure. Future developments of the procedure could include a further cluster analysis on the time series divided by type of trend. In this way an efficient procedure for determining the expected return and the

riskiness of a portfolio could be made completely automatic. Another possible development could be to consider stochastic nonstationary time series, since the empirical evidence shows almost always financial time series are stochastically nonstationary. The latter can occur both in the form of a stochastic trend and as a characteristic of the epsilon process.

References

1. Chen, L., Wu, W.B.: Testing for trends in high-dimensional time series. J. Am. Stat. Assoc. **114**(526), 869–881 (2019)
2. Doukhan, P.: Mixing: Properties and Examples. Lecture Notes in Statistics, vol. 85. Springer-Verlag, New York (1994). https://doi.org/10.1007/978-1-4612-2642-0
3. Fan, J., Gijbels, I.: Local Polynomial Modelling and Its Applications: Monographs on Statistics and Applied Probability. CRC Press, London (1996)
4. Fan, J., Lv, J.: Sure independence screening for ultrahigh dimensional feature space. J. R. Stat. Soc. Ser B (Stat. Methodol.) **70**(5), 849–911 (2008)
5. Francisco-Fernández, M., Vilar-Fernández, J.M.: Local polynomial regression estimation with correlated errors. Commun. Stat. Theory Methods **30**(7), 1271–1293 (2001)
6. Giordano, F., Parrella, M.L.: Efficient nonparametric estimation and inference for the volatility function. Statistics **53**(4), 770–791 (2019)
7. Härdle, W.: Applied Nonparametric Regression. Cambridge University Press, Cambridge (1990)
8. Masry, E., Fan, J.: Local polynomial estimation of regression functions for mixing processes. Scand. J. Stat. **24**(2), 165–179 (1997)
9. Wand, M.P., Jones, M.C.: Kernel Smoothing. CRC Press, London (1994)
10. Zhang, T.: Clustering high-dimensional time series based on parallelism. J. Am. Stat. Assoc. **108**(502), 577–588 (2013)
11. Zhengyan, L., Chuanrong, L.: Limit Theory for Mixing Dependent Random Variables. Springer Science & Business Media, Dordrecht (1997)

Differential Pursuit-Evasion Games and Space Economy: New Research Perspectives

Massimiliano Ferrara[1,2]([✉]), Gafurjan Ibragimov[3], and Bruno Antonio Pansera[1]

[1] Department of Law, Economics and Human Sciences and Decisions_Lab, University Mediterranea of Reggio Calabria, via dell'Universitá, 25, 89124 Reggio Calabria, Italy
{massimiliano.ferrara,bruno.pansera}@unirc.it
[2] I-CRIOS, Bocconi University, via Sarfatti, 25, 20136 Milan, Italy
[3] Department of Mathematics, Universiti Putra Malaysia, UPM Serdang Selangor Darul Ehsan, 43400 Seri Kembangan, Malaysia
ibragimov@upm.edu.my

Abstract. In this article, we consider a research topic in the field of game theory: the pursuit-evasion differential games. With the evolution of the world and technology, activities have transformed, and systems based on the dichotomy evader-pursuer have also evolved to answer other questions, just think for example of the management of intelligent missiles, designed to chase an object in movement. The main purpose is to provide a study proposal for the management and recovery of waste materials in space missions through an approach based on differential games.

Keywords: Differential games · Pursuit-evasion games · Space economy

1 Introduction

Studying a differential game means defining a system of time-dependent differential equations, which describes the evolution process of the phenomenon over time. The main purpose is to choose the optimal control, based on the definition of adequate escape or capture strategies. Each player defines one of the controls by choosing the optimal strategy for his purpose. The problem that arises from these observations is therefore that of the definition of the control of the actors in a process of pursuit-evasion. We are confident that the paper's idea of linking a theoretical study of this problem will be a good way to define new perspectives in space research and the increasingly sustainable management of space missions. The main purpose of this paper is to describe some results obtained in the field of pursuit-evasion games, i.e. what results have been obtained in terms of evasion and/or capture strategy. This is because we would like to start thinking about

the application of these games in the context of the recovery of waste materials in space missions. The general idea is to bring the researchers' attention to the fact that if the debris starts from certain positions and follows well-defined routes, it is possible to recover these materials. There are, in fact, meaningful results that allow us to understand under what conditions the strategies implemented by the pursuers lead to the capture of the escaped person and when they fail. The pursuit and evasion game, thus described, would present m pursuers and n evaders, where the latter are the waste materials dispersed in space and the former the tools for cleaning the space. The basic idea is still in a primordial stage, but this application would be inserted in a context of sustainability of space missions and would reduce the impact that the same conditions in cosmic space. The study of linear two person zero-sum differential games was initiated by Isaacs [14]. A fundamental contribution to the development of the theory of differential games was given in [3,4,8,9,12,19,20,23]. In [15] linear pursuit differential game was studied.

Simple motion pursuit and evasion differential games of many players is a major area of interest within the field of differential games. Traditionally, there are two constraints on controls of pursuer and evader: geometric or integral constraints. Croft in [7] made an important contribution to the area of differential games of many pursuers. In the game of many pursuers Mishchenko et al. [18] constructed a new evasion strategy. Evasion of one faster evader from many pursuers was studied by Chernous'ko [6]. The most interesting finding on simple motion differential game of many pursuers and one evader in $\mathbf{R}^n$ was obtained by Pshenichnii [21]. Another type of differential games with simple dynamics containing a group of pursuers and a group of evaders were analyzed in $\mathbf{R}^n$ by Petrov [5]. A differential game of a group of n pursuers and one evader was studied by Azamov et al. [2] on the 1-skeleton graph S of regular polyhedrons of three types. An evasion differential game of countably many pursuers and countably many evaders was studied by Idham et al. [1] in Hilbert space l_2 and a sufficient condition of evasion was obtained in terms of energies of players. If the control resource, say as energy, finance, food etc. (see, e.g. [13,17]) is bounded, then integral constraints on the control functions arise in control systems. Ibragimov and Satimov in [11] studied the following simple motion pursuit differential game of M evaders and K pursuers

$$\dot{x}_i = u_i, \quad x_i(0) = x_{i0}, \quad i = 1, \ldots, K, \tag{1}$$
$$\dot{y}_j = v_j, \quad y_j(0) = y_{j0}, \quad j = 1, \ldots, M, \tag{2}$$

with integral constraints

$$\int_0^\infty |u_i(s)|^2 ds \le \rho_i^2, \ i = 1, \ldots, K, \tag{3}$$

$$\int_0^\infty |v_j(s)|^2 ds \le \sigma_j^2, \ j = 1, \ldots, M, \tag{4}$$

in a closed convex subset of $\mathbf{R}^n$, where $x_i, y_j, u_i, v_j \in \mathbf{R}^n$, $n \geq 2$, and ρ_i, σ_j are given positive numbers, $x_{i0} \neq y_{j0}$ for all $j = 1, \ldots, M$ and $i = 1, \ldots, K$. It was established in [11] that if

$$\rho_1^2 + \ldots + \rho_K^2 > \sigma_1^2 + \ldots + \sigma_M^2, \tag{5}$$

then pursuit can be completed game (1)–(4). The evasion game described by equations (1)–(4) in $\mathbf{R}^n$ in general case, that is, when

$$\rho_1^2 + \ldots + \rho_K^2 \leq \sigma_1^2 + \ldots + \sigma_M^2 \tag{6}$$

was studied in [10].

2 Space Economy and the Detritus Management: The Role of Differential Games

In this paper we are presenting a model by which modeling the detritus dynamics and the related effects on operation of satellites, spacecrafts and related problems. The space area surrounding planet Earth is submerged by more than 300.000 particles of space-debris, characterized by a diameter of more than 1 cm. This area is divided into three levels with respect to the Earth's orbit: LEO (Low Earth Orbit), up to 2,000 Km; MEO (Medium Earth Orbit), between 2.000 Km and 36.000 Km; GEO (Geosynchronous Earth Orbit), over 36,000 Km. Space debris has various similarities to asteroids because their long-term evolution has also depended on collisions at high speeds and orbits are very chaotic. On February 10, 2009, a serious impact occurred between the Iridium 33 and Cosmos 2251 satellites which showed how the possibility of a collision is a risk not to be underestimated in any space activity. In fact, after that impact, space experts established some preventive procedures such as: validation of impact mitigation, coordination of space traffic and efficient mitigation measurements. The United States Strategic Command (USSTRATCOM) has classified the alien space probe in orbit in the TLE (Two Line Element) catalog, where about 15,000 objects are represented with their current orbital parameter. The limit dimensions of the cataloged objects are: between 5 and 10 cm under a few miles km of altitude; between 0.5 and 1 Mt for orbits up to the GEO level. In particular, 6% of the objects in the TLE catalog are operational satellites; 24% non-operational satellites; 17% missiles in the upper phase; 13 % mission debris; 40% fragments. The United States Space Surveillance Network frequently offers a wealth of optical and radar observations. Since 2008, Europe has launched the SSA program, SST segment, to increase the know-how of the space surrounding the Earth; to do this it is essential to have very accurate systems and algorithms to be able to precisely define the orbit. The field of space debris has been long and thoroughly studied by SpaceDys and its partners who, since the early 90s, have contributed to the creation of the first European models for the analysis of the evolution of all space debris in the long term, focusing precisely on the risks of possible impacts and on the related possible mitigation interventions. Compared to the

impact that occurred in 2009, for example, ten years earlier they had published an article in Nature that illustrated the possible consequences due to the fragmentation of the Constellation of Iridium. Therefore, the SpaceDys team has further deepened the theoretical and applicative research on this topic with significant results regarding the determination of the orbit and the correlation of space objects, published in numerous journals and presented in several international conferences. In addition, SpaceDys has developed another project that, thanks to radar and GPS measurements, contributes to a better determination of the orbit and to the prediction of re-entry of the Goce satellite, focusing on the definition of significant dynamics and observational uncertainties and on the correlation between position prediction and re-entry position. In this frame Space economy is a new challenge field of study which involves activities and analysis of resources that create value and benefits to human beings by exploring, managing, and utilizing space. The core of the space industry activities in space manufacturing and in satellite operations and other consumer activities that have been derived over the years from R&D especially promoted by governments. It includes all public and private plans involved in developing, providing and using space-related outputs, space derived products and services and the scientific knowledge arisen from space research. In the near future we will develop new discoveries in this framework starting from the results gathered from this work arising from space research

3   Concluding Remarks and Further Developments

Starting from October 4, 1957, over 4,900 space launches have led, as evident in the USA Space Surveillance Network graph, to a population in orbit today of over 17,000 traceable objects, with dimensions greater than 10 cm (equivalent diameter). Of these objects:

1. about 1,000 are operational spacecraft;
2. the remaining 94% is made up of space debris;
3. about 64% of the objects routinely tracked are fragments belonging to 250 space systems that have undergone fragmentation in orbit, mainly due to improvised explosions and then to collisions between satellites or launch systems (Fig. 1).

The ASI (Italian Space Agency) data describe, the population in Earth orbit of smaller objects of artificial origin, therefore not traceable, is the following:

1. objects larger than 10 cm, about 20,000;
2. objects larger than 1 cm, about 700,000 objects;
3. objects larger than 1 mm, about 170 million objects.

Due to the relative orbital speeds, which can exceed 50 thousand kilometers per hour, a debris of 1 cm can seriously damage or disable an operational spacecraft, while in the event of collisions with objects larger than 10 cm they become possible catastrophic ruptures, capable of producing clouds of dangerous debris,
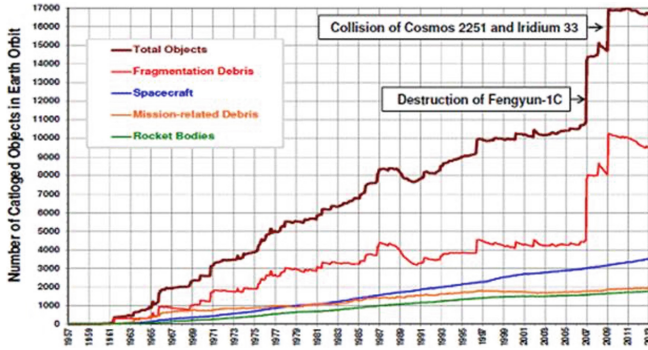
Fig. 1. Source: ASI (Agenzia Spaziale Italiana)

potentially causing further catastrophic collisions, triggering a sort of chain reaction in some orbital regions (the so-called "Kessler syndrome", named after NASA researcher Donald Kessler who highlighted the problem at the end of the 70s of the last century). Space debris mitigation measures, when properly implemented by spacecraft designers and mission operators, can reduce the growth rate of the debris population. However, the active removal of debris already in orbit, in particular large stadiums and satellites abandoned in the most crowded orbits, seems to be the only measure capable of reversing the current phase of increasing debris. The so-called space waste or the elements of rockets or satellites that remain suspended in the low orbit of the Earth when they stop functioning or during the various phases of space missions, in the next years this critical mass will grow steadily and this will involve the adoption of policies aimed at guaranteeing the sustainability of space missions. The movements and dynamics that characterize the coexistence of satellites, spacecraft and space waste represent in our opinion a field of study that in the coming years could offer interesting research perspectives and help to arrange optimization plans of space routes in order to contain risks of collision, causing costs and economic losses. Our seminal idea that inspired this article is the mathematical modeling included in the pursuit-evasion games. Our approach has as its main purpose to propose a new vision of the protagonists of a space environment such as satellites, spacecraft and debris that draw orbits, focused on a dynamics of displacements based on forecasting and crash management data. An ongoing interim study could consider debris wandering in space as pursuers and satellites and spacecraft as evaders. A second approach sees space cleaning as an application. In fact, by exchanging the role of the actors, or considering the debris as evaders and the spaceships as pursuers, the ideal conditions would be obtained to make this series of activities carried out outside the earth's orbit more and more eco-sustainable.

References

1. Alias, I.A., Ibragimov, G., Rakhmanov, A.: Evasion differential game of infinitely many evaders from infinitely many pursuers in Hilbert space. Dyn. Games App. **6**(2), 1–13 (2016)
2. Azamov, A.A., Kuchkarov, A.S., Holboyev, A.G.: The Pursuit-Evasion Game on the 1-Skeleton Graph of a Regular Polyhedron II. Autom. Remote Control **80**, 164–170 (2019)
3. Azamov, A.A.: On Pontryagin's second method in linear differential games of pursuit. Mat. Sb. (N.S.) **118**(160), 3(7), 422–430 (1983). (Math. USSR-Sb. 46(3), 429–437 (1983))
4. Berkovitz, L.D.: Differential game of generalized pursuit and evasion. SIAM J. Contr. **24**(3), 361–373 (1986)
5. Blagodatskikh, A.I., Petrov, N.N.: Conflict interaction between groups of controlled objects. Udmurt State University Press, Izhevsk (2009).(in Russian)
6. Chernous'ko, F.L.: A problem of evasion of several pursuers. Prikl. Mat. Mekh. **40**(1), 14–24 (1976)
7. Croft, H.T.: Lion and man: a postscript. J. London Math. Soc. **39**, 385–390 (1964)
8. Elliott, R.J., Kalton, N.J.: The existence of value in differential games. Mem. Am. Math. Soc. **126**, 1–67 (1972)
9. Fleming, W.H.: The convergence problem for differential games. J. Math. Anal. App. **3**, 102–116 (1961)
10. Ibragimov, G., Ferrara, M., Kuchkarov, A., Pansera, B.A.: Simple motion evasion differential game of many pursuers and evaders with integral constraints. Dyn. Games App. **8**, 352–378 (2018)
11. Ibragimov, G., Satimov, N.: A multiplayer pursuit differential game on a closed convex set with integral constraints. Abstract Appl. Anal. **2012**, 12 (2012). (Article ID 460171)
12. Hajek, O.: Pursuit Games. Mathematics in Science and Engineering, Academic Press, New York (1975)
13. Huseyin, A., Huseyin, N., Guseinov, K.G.: Approximation of the sections of the set of trajectories of the control system described by a nonlinear Volterra integral equation. Math. Model. Anal. **20**(4), 502–515 (2015)
14. Isaacs, R.: Differential Games. John Wiley & Sons, New York (1965)
15. Matychyn, I.: Pursuit strategy of motion camouflage in dynamic games. Dyn. Games App. **10**, 145–156 (2020)
16. Ivanov, R.P.: Simple pursuit-evasion on a compact convex set. Doklady Akademii Nauk SSSR **254**(6), 1318–1321 (1980)
17. Krasovskii, N.N.: Theory of Control of Motion: Linear Systems. Nauka, Moscow (1968)
18. Mishchenko, E.F., Nikol'skii, M.S., Satimov, N.Y.: Evoidance encounter problem in differential games of many persons. Trudy MIAN USSR **143**, 105–128 (1977)
19. Petrosyan, L.A.: Differential Games of Pursuit. World Scientific, Singapore, London (1993)
20. Pontryagin, L.S., Mischenko, Y.F.: The problem of evasion in linear differential games. Differencial'nye Uravnenija **7**(3), 436–445 (1971)
21. Pshenichnii, B.N.: Simple pursuit by several objects. Cybern. Syst. Anal. **12**(3), 145–146 (1976)
22. Salimi, M., Ferrara, M.: Differential game of optimal pursuit of one evader by many pursuers. Int. J. Game Theory **48**(2), 481–490 (2019)
23. Satimov, N.Y., Rikhsiev, B.B.: Methods of Solving of Evasion Problems in Mathematical Control Theory. Fan, Tashkent, Uzbekistan (2000)

Graphical Models for Commodities: A Quantile Approach

Beatrice Foroni[1(✉)], Luca Merlo[2], and Lea Petrella[1]

[1] MEMOTEF Department, Sapienza University of Rome, Rome, Italy
beatrice.foroni@uniroma.it
[2] Department of Statistical Sciences, Sapienza University of Rome, Rome, Italy

Abstract. The high level of integration of international financial markets highlights the need to accurately assess contagion and systemic risk under different market conditions. To this end, we develop a quantile graphical model to identify the tail conditional dependence structure in multivariate data across different quantiles of the marginal distributions of the variables of interest. To implement the procedure, we consider the Multivariate Asymmetric Laplace distribution and exploit its location-scale mixture representation to build a penalized EM algorithm for estimating the sparse precision matrix of the distribution by means of an L_1 penalty. The empirical application is performed on a large set of commodities representative of the energy, agricultural and metal sectors.

Keywords: EM algorithm · Energy commodities · Graphical models · Multivariate asymmetric laplace distribution

1 Introduction

The financial system is a complex, dynamic and interconnected world. Observing the extreme financial integration in the recent global financial crisis, it was soon noted the crucial importance to identify how the impact of financial stress events can spread across the whole financial global system. For this reason, network science has emerged as a useful tool for describing the propagation of systemic risk, where the interconnectedness between financial institutions is represented by a graph whose nodes stand for companies, commodities, institutions, for instance, and the edges highlight their interactions. Within this literature, Gaussian Graphical Models (GGM) have received an enormous attention because they provide a simple method to model the pair-wise conditional correlation structure of a collection of stochastic variables. As it is well known, under the assumption of normally distributed data, the underlying conditional dependence structure is completely characterized by the precision matrix, i.e. the inverse of the covariance matrix of the corresponding GGM (see [7] for a general background). In a

high-dimensional framework when a large set of random variables is considered, we are interested in identifying only a smaller subset of variables that exhibits the most relevant and strongest dependencies. Among the several methods proposed in literature, there is the popular Least Absolute Shrinkage and Selection Operator (LASSO) method by [10] which allows to improve estimation and conduct variable selection. To estimate the model parameters, the implementation of this technique often relies on the *Graphical LASSO (glasso)* algorithm of [3], which maximizes the likelihood of the model penalized by the L_1-norm of the elements of the precision matrix. However, several empirical studies show that financial returns exhibit most of the well known stylized facts like fat tails, leptokurtosis and skewness, and deviation from normality makes it harder to characterize conditional dependence structures. The literature regarding non-Gaussian graphical models is fairly limited. In this context, the work of [2] provides a tool for robust model selection using multivariate t-distributions to model the data. Moreover, these proposals are not able to recover the dependencies in the tails of the distribution. Be able to understand and focus on specific part of a distribution such as the tails can really improve the knowledge in areas like financial contagion and systemic risk, where the dynamic of extreme events is of utmost importance. In this paper we develop a quantile graphical model to estimate the conditional tail dependence structure in multivariate data at different quantile levels of interest, without relying on the restrictive assumption of normally distributed data. In order to model the conditional dependence structure of multiple random variables at quantile-specific indices, we generalize the work of [8], which consider a reparametrization of the Multivariate Asymmetric Laplace (MAL) distribution of [6] to jointly model conditional quantiles of multiple random variables in a likelihood framework. Following [2], we demonstrate that the precision matrix of the MAL distribution completely characterizes the conditional dependence structure among the random variables at each quantile level, and allows us to construct a graph whose edges correspond to relations of conditional dependency. As opposed to GGMs, the proposed methodology has several advantages. Firstly, we can assess contagion and systemic risk during different market conditions focusing on specific parts of the distributions of variables. Secondly, we can construct a collection of graphs indexed by the quantile level, which allows us to evaluate whether the financial network is more interconnected and more vulnerable to contagion during periods of financial and economic crises. To induce sparsity in the precision matrix, we exploit the Gaussian location-scale mixture of the MAL and apply the *glasso* algorithm. In particular, following [4], we build a suitable Penalized EM (PEM) algorithm based on the maximization of the likelihood of the model penalized by the L_1-norm of the off-diagonal elements of the precision matrix. The estimated networks can be analyzed with respect to centrality measures as functions of the quantile level. The relevance of our approach is shown empirically on a set of energy, metal and agriculture commodity futures, and the modeling approach we propose is able to identify the connectedness of the commodities during crisis and bullish periods, and can describe

the topological structure of the underlying graph at different quantile levels of interest.

2   Model Specification

Given p quantile indexes $\tau = [\tau_1, \ldots, \tau_p]'$, with $\tau_j \in (0, 1)$, for $j = 1, \ldots, p$, let $\mathbf{Y}_t = [Y_t^{(1)}, \ldots, Y_t^{(p)}]$ denote a continuous p-dimensional random vector for $t = 1, \ldots, T$. Generalizing the approach of [8], our objective is to develop a quantile graphical model for learning the conditional tail dependence structure among the components of $\mathbf{Y}_t$ at different quantile levels of interest τ. Specifically, we employ the MAL distribution, $\mathcal{MAL} \sim (\boldsymbol{\mu}, \boldsymbol{D}\tilde{\boldsymbol{\xi}}, \boldsymbol{D}\boldsymbol{\Sigma}\boldsymbol{D})$, (see [6]) as:

$$f_Y(\boldsymbol{y}_t) = \frac{2 \exp\left\{ (\boldsymbol{y}_t - \boldsymbol{\mu})' \boldsymbol{D}^{-1} \boldsymbol{\Sigma}^{-1} \tilde{\boldsymbol{\xi}} \right\}}{(2\pi)^{p/2} |\boldsymbol{D}\boldsymbol{\Sigma}\boldsymbol{D}|^{1/2}} \left(\frac{\tilde{m}_t}{2 + \tilde{d}} \right)^{\nu/2} K_\nu \left(\sqrt{(2 + \tilde{d})\tilde{m}_t} \right), \quad (1)$$

where $\boldsymbol{\mu}$ is the location parameter, $\boldsymbol{D}\boldsymbol{\xi} \in \mathcal{R}^p$ is the scale (or skew) parameter, with $\boldsymbol{D} = \text{diag}[\delta_1, \delta_2, \ldots, \delta_p]$, $\delta_j > 0$ and $\boldsymbol{\xi} = [\xi_1, \xi_2, \ldots, \xi_p]'$, having generic element $\xi_j = \frac{1 - 2\tau_j}{\tau_j(1 - \tau_j)}$. $\boldsymbol{\Sigma}$ is a $p \times p$ positive definite matrix such that $\boldsymbol{\Sigma} = \boldsymbol{\Lambda}\boldsymbol{\Psi}\boldsymbol{\Lambda}$, with $\boldsymbol{\Psi}$ being a correlation matrix and $\boldsymbol{\Lambda} = \text{diag}[\sigma_1, \sigma_1, \ldots, \sigma_p]$, with $\sigma_j^2 = \frac{2}{\tau_j(1 - \tau_j)}$, $j = 1, \ldots, p$. Finally, $\tilde{m}_t = (\boldsymbol{y}_t - \boldsymbol{\mu})'(\boldsymbol{D}\boldsymbol{\Sigma}\boldsymbol{D})^{-1}(\boldsymbol{y}_t - \boldsymbol{\mu})$, $\tilde{d} = \tilde{\boldsymbol{\xi}}' \boldsymbol{\Sigma}^{-1} \tilde{\boldsymbol{\xi}}$, and $K_\nu(\cdot)$ denotes the modified Bessel function of the third kind with index parameter $\nu = (2 - p)/2$. One of the key benefits of the MAL distribution is that, using (1) and following [6], the $\mathcal{MAL} \sim (\boldsymbol{\mu}, \boldsymbol{D}\tilde{\boldsymbol{\xi}}, \boldsymbol{D}\boldsymbol{\Sigma}\boldsymbol{D})$ admits the following location-scale mixture representation:

$$\mathbf{Y} = \boldsymbol{\mu} + \boldsymbol{D}\tilde{\boldsymbol{\xi}}W + \sqrt{W}\boldsymbol{D}\boldsymbol{\Sigma}^{1/2}\boldsymbol{Z} \tag{2}$$

where $\boldsymbol{Z} \sim \mathcal{N}_p(\boldsymbol{0}_p, \boldsymbol{I}_p)$ denotes a p-variate Normal distribution and $W \sim \text{Exp}(1)$ has a standard Exponential distribution, with $\boldsymbol{Z}$ being independent of W. Hence, the mixture representation in (2) represents the generating process of a MAL random vector $\boldsymbol{Y}$ from a latent Gaussian random vector $\boldsymbol{Z}$ with correlation matrix $\boldsymbol{\Psi}$ and a single latent Exponential variable with mean 1. In particular, the constraints imposed on $\tilde{\boldsymbol{\xi}}$ and $\boldsymbol{\Lambda}$ represent necessary conditions for model identifiability for any fixed quantile level $\tau_1, \ldots, \tau_p$ and guarantee that $\boldsymbol{\mu}^{(j)}$ is the τ_j-th quantile of $Y_t^{(j)}$, for $j = 1, \ldots, p$. To build the graphical model, let $G = (V, E)$ be an undirected graph where $V = \{1, \ldots, p\}$ is the set of nodes, such that each component of the random variable $\boldsymbol{Y}_t$ corresponds to a node in V, and $E \subseteq V \times V$ represents the set of undirected edges. In order to study the conditional dependence structure of $\boldsymbol{Y}_t$ through the graph G, we exploit the MAL representation in (2). For notational convenience and to illustrate the similarities with the GGM, we define the precision matrix $\boldsymbol{K} = \boldsymbol{\Psi}^{-1}$. Following [7] and the t-distribution graphical model approach in [2], we establish the following proposition.

Proposition 1. *For a fixed p-dimensional vector of quantile levels $\tau = [\tau_1, \tau_2, \ldots, \tau_p]'$ such that $\tau_j \in (0, 1)$, for $j = 1, \ldots, p$, let $\boldsymbol{Y} \sim \mathcal{MAL}(\boldsymbol{\mu}, \boldsymbol{D}\boldsymbol{\xi}, \boldsymbol{D}\boldsymbol{\Lambda}\boldsymbol{K}^{-1}\boldsymbol{\Lambda}\boldsymbol{D})$. If two nodes j and k, with $j, k \in V$ and $j \neq k$, of the graph are separated by a set of nodes $C \in V$, then $\boldsymbol{Y}^{(j)}$ and $\boldsymbol{Y}^{(k)}$ are conditionally uncorrelated given $\boldsymbol{Y}^{(C)}$.*

The proof of Proposition 1 follows directly from the mixture representation of the MAL in (2) and the closure property of the Normal distribution under conditioning of its components. Most importantly, from Proposition 1 there follow several interesting comments. Firstly, the existing edges in G indicate the allowed conditional independencies in the latent Gaussian vector $\boldsymbol{Z}$. Secondly, the zero entries in the precision matrix $\boldsymbol{K}$ imply the conditional uncorrelation between the components of $\boldsymbol{Y}_t$ at each given quantile level τ. More formally, if the (j, k)-th element of $\boldsymbol{K}$ equals 0, for $j \neq k$, this means that $\boldsymbol{Y}_t^{(j)}$ and $\boldsymbol{Y}_t^{(k)}$ are uncorrelated, given all the remaining components of $\boldsymbol{Y}_t$, and it entails a missing edge between the nodes j and k, i.e. the edge $(j, k) \notin E$. To estimate and make inference on the model parameters we develop a suitable Expectation-Maximization (EM) algorithm, which exploits the mixture representation of the MAL distribution, treating W as missing data. In order to identify only a smaller subset of variables that exhibit the most relevant and strongest dependencies, we construct a PEM algorithm by adding an L_1-norm penalty of the off-diagonal elements of $\boldsymbol{K}$ to the likelihood of the model. Specifically, for a given vector $\tau = [\tau_1, \tau_2, \ldots, \tau_p]'$, the penalized complete log-likelihood function is proportional to:

$$\ell_c(\boldsymbol{\Phi}_\tau) \propto \frac{T}{2}\log|\boldsymbol{D}^{-1}\boldsymbol{K}\boldsymbol{D}^{-1}| - \frac{T}{2}\text{tr}\{\boldsymbol{K}\boldsymbol{S}\} - \rho||\boldsymbol{K}||_1 \tag{3}$$

with

$$\boldsymbol{S} = \frac{1}{T}\sum_{t=1}^{T}\frac{1}{W_t}\boldsymbol{\Lambda}^{-1}\boldsymbol{D}^{-1}(\boldsymbol{Y}_t - \boldsymbol{\mu} - \boldsymbol{D}\boldsymbol{\xi}W_t)(\boldsymbol{Y}_t - \boldsymbol{\mu} - \boldsymbol{D}\boldsymbol{\xi}W_t)'\boldsymbol{D}^{-1}\boldsymbol{\Lambda}^{-1} \tag{4}$$

and where W_t is an Exponential random variable with mean 1.

As it can be noticed, the likelihood function in (3) is convex in $\boldsymbol{K}$. Therefore, at each iteration of the PEM, this feature allows us to adopt the *glasso* algorithm for efficient estimation of the sparse precision matrix $\boldsymbol{K}$.

3   Main Results and Conclusions

The empirical analysis is performed on the log-returns of 24 commodities representative of the energy, agriculture and metals sectors from October 3, 2005 to December 31, 2018. Table 1 reports the considered commodities. We set $\tau = \tau_j$, $j = 1, \ldots, 24$, and fit the proposed model for a sequence of 99 quantile levels $\tau = [0.01, 0.02, \ldots, 0.98, 0.99]'$. Then, for each τ, we construct the corresponding graph G_τ. In Fig. 1 we represent the estimated graph at $\tau = 0.05$ and $\tau = 0.46$ to show how the density of the network changes between the tails ($G_{\tau=0.05}$) and

Table 1. Commodities considered in the analysis. In brackets the Bloomberg tickers used in Fig. 1.

Commodity sector			
Energy	Metals	Agriculture	
Heating Oil (HO1)	Gold (GC1)	Oats (O1)	Rough Rice (RR1)
Gasoline (XB1)	Silver (SI1)	Wheat (W1)	Sugar (SB1)
Low Sulfur Gasolio (QS1)	Copper (HG1)	Soybeans (S1)	Soybean Oil (BO1)
Natural Gas (NG1)	Palladium (PA1)	Coffee (KC1)	Soybean Meal (SM1)
Ethanol (DL1)	Zinc (LX1)	Cocoa (CC1)	Orange Juice (JO1)
WTI Crude Oil (CL1)		Cotton (CT1)	Corn (C1)
Natural Gas UK (FN1)			
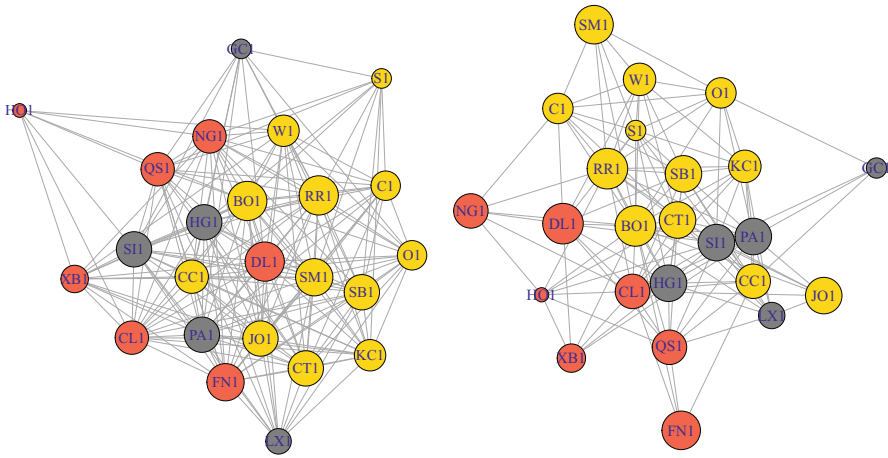
Fig. 1. Graphs for $\tau = 0.05$ (left) and $\tau = 0.46$ (right). Yellow, grey and red nodes represent respectively the agriculture, the metal and the energy sectors. The node size reflects the degree of each variable in the network.

the centre of the distribution ($G_{\tau = 0.46}$). These two quantile indices are representative for the most and the less connected network, respectively. On this matter it is also interesting to note the reduced number of connection in $G_{\tau = 0.46}$ and how in that case few nodes stand out in terms of the degree centrality.

A deeper analysis to analyze how the interconnectedness and contagion risk change as a function of the quantile index τ, is conducted by showing in Fig. 2 the following network centrality measures for the considered sequence of values of τ, namely the edge density (left), skewness of degree distribution (center) and network assortativity (right). The edge density, which is the ratio of edges in network G_τ to all possible edges $p \times (p - 1)/2$, shows a highly interconnected system, even for the smallest ratio of value 46 at $\tau = 0.46$. It is evident a strongest dependency during crisis, bearish periods, i.e. for $\tau < 0.25$, as the edge density is the largest for networks capturing the dependence in the left-
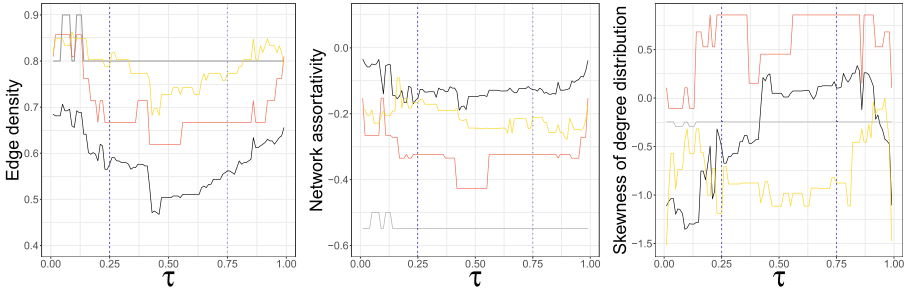
Fig. 2. Edge Density, Network Assortativity and Skewness of degree distribution as a function of τ. Blue vertical lines identify the 25th and 75th percentile. The black curve represents the centrality measure for the whole graph. Yellow, grey and red curves represent respectively the centrality measures for the agriculture, the metal and the energy sectors.

tail of the return distribution. This increased connectedness in the left tail is in line with existing studies which found that during the global financial crisis and the European sovereign debt crisis the interconnectedness of global or local commodities or financial institutions increased [5,9]. The network assortativity is a measure of dependence between the vertex degree of connected vertices. The value of the coefficient suggests whether vertices tend to be connected to each other depending on their degree. In our analysis we detect a negative value for each quantile level, which indicates that vertices of different degree are more likely to be connected, as we can expect from a highly dense system as the commodity network. Moreover, the skewness of the degree distribution expresses how the structure of connectedness changes by varying τ. Negative values at extreme quantiles evidence that no commodity dominates with respect to vertex degree, while in tranquil periods few variables have a large influence on the network, as shown in Fig. 1. In conclusion, with our approach we are able to recover valuable information at each quantile level even without the assumption of normality. The whole analysis conveys a highly connected network which becomes even more dense during bearish and bullish markets periods, and the results are in line with existing studies [1,5,9]. With this model we strengthen the existing literature in this field, implementing a technique to adjust the *glasso* algorithm to a quantile structure of dependence.

References

1. Deev, O., Lyócsa, Š.: Connectedness of financial institutions in Europe: a network approach across quantiles. Physica A **550**, 124035 (2020)
2. Finegold, M., Drton, M.: Robust graphical modeling of gene networks using classical and alternative t-distributions. Ann. Appl. Stat. **5**, 1057–1080 (2011)
3. Friedman, J., Hastie, T., Tibshirani, R.: Sparse inverse covariance estimation with the graphical Lasso. Biostatistics **9**(3), 432–441 (2008)

4. Green, P.J.: On use of the EM algorithm for penalized likelihood estimation. J. Roy. Stat. Soc.: Ser. B (Methodol.) **52**(3), 443–452 (1990)
5. Jiang, H., Su, J.J., Todorova, N., Roca, E.: Spillovers and directional predictability with a cross-quantilogram analysis: the case of US and Chinese agricultural futures. J. Futur. Mark. **36**(12), 1231–1255 (2016)
6. Kotz, S., Kozubowski, T., Podgorski, K.: The Laplace Distribution and Generalizations: A Revisit with Applications to Communications, Economics, Engineering, and Finance. Springer, Cham (2012). https://doi.org/10.1007/978-1-4612-0173-1
7. Lauritzen, S.L.: Graphical Models, vol. 17. Clarendon Press, Oxford (1996)
8. Petrella, L., Raponi, V.: Joint estimation of conditional quantiles in multivariate linear regression models with an application to financial distress. J. Multivar. Anal. **173**, 70–84 (2019)
9. Shahzad, S.J.H., Hernandez, J.A., Al-Yahyaee, K.H., Jammazi, R.: Asymmetric risk spillovers between oil and agricultural commodities. Energy Policy **118**, 182–198 (2018)
10. Tibshirani, R.: Regression shrinkage and selection via the Lasso. J. Roy. Stat. Soc. Ser. B (Methodol.) **58**(1), 267–288 (1996)

The Mardia's Kurtosis of a Multivariate GARCH Model

Cinzia Franceschini[1] and Nicola Loperfido[2(✉)]

[1] Dipartimento di Scienze Statistiche "Paolo Fortunati", Università degli Studi di Bologna, Via Belle Arti 41, Bologna, Italy
`cinzia.franceschini2@unibo.it`
[2] Dipartimento di Economia, Società e Politica, Università degli Studi di Urbino "Carlo Bo", Via Saffi 42, Urbino, Italy
`nicola.loperfido@uniurb.it`

Abstract. The Mardia's kurtosis of a random vector with nonsingular covariance matrix and finite fourth-order moments is the fourth moment of the Mahalanobis distance of the random vector from its mean. In particular, the Mardia's kurtosis of a nondegenerate random variable with finite fourth moment coincides with its fourth standardized moment. The Mardia's kurtosis is the best known measure of multivariate kurtosis and appears in normality testing, robustness studies and outlier detection. Under mild assumptions, we show that an observation generated by a multivariate GARCH model has a Mardia's kurtosis which is greater than the Mardia's kurtosis of the innovation in the same model. The result generalizes to the multivariate case a well-known feature of univariate GARCH models. The practical relevance of the result is assessed with real data.

Keywords: GARCH model · Multivariate kurtosis · Stylized fact

1 Introduction

The kurtosis of a random variable is its fourth standardized moment, if the fourth moment of the random variable is finite. Random variables whose kurtosis is smaller than three, equal to three or greater than three are platykurtic, mesokurtic and leptokurtic distributions, respectively. Leptokurtosis is a well-known stylized fact of financial returns, which partly motivated the ARCH model [7]. The same model has been generalized in several ways to describe other stylized facts of financial returns (see, for example, [4]).

Multivariate generalization of these models, as for example the multivariate GARCH model [3] and the multivariate SGARCH model [6] were introduced to investigate connections between several financial markets. In order to assess their complex tail behaviour, multivariate generalizations of kurtosis are called for. [13] generalized kurtosis to the multivariate case by defining the kurtosis

of a random vector with nonsingular covariance matrix and finite fourth-order moments as the fourth moment of the Mahalanobis distance of the random vector from its mean. More formally, let $\mathbf{x}$ be a $p-$dimensional random vector with mean $\boldsymbol{\mu}$, nonsingular covariance $\boldsymbol{\Sigma}$ and finite fourth-order moments. The Mardia's kurtosis of $\mathbf{x}$ is

$$\beta_{2,M}\left(\mathbf{x}\right) = E\left\{\left[\left(\mathbf{x} - \boldsymbol{\mu}\right)^T \boldsymbol{\Sigma}^{-1}\left(\mathbf{x} - \boldsymbol{\mu}\right)\right]^2\right\}.$$

As remarked by [9], Mardia's kurtosis is the best known measure of multivariate kurtosis. It also possesses several interesting properties. In the first place, it is invariant with respect to one-to-one affine transformations. Let $\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{b}$ be an affine, one-to-one transformation of the $p-$dimensional random vector $\mathbf{x}$. Then $\mathbf{x}$ and $\mathbf{y}$ have the same Mardia's kurtosis:

$$\beta_{2,M}\left(\mathbf{x}\right) = \beta_{2,M}\left(\mathbf{A}\mathbf{x} + \mathbf{b}\right), \mathbf{A} \in \mathbb{R}^p \times \mathbb{R}^p, \ \det\left(\mathbf{A}\right) \neq 0, \ \mathbf{b} \in \mathbb{R}^p.$$

In the second place, the tail behaviour and the Mardia's kurtosis of a random vector are related. Let $\beta_{2,M}\left(\mathbf{x}\right)$ be the Mardia's kurtosis of a random vector $\mathbf{x}$ with mean $\boldsymbol{\mu}$ and nonsingular covariance $\boldsymbol{\Sigma}$. Then the following inequalities hold true for any real ε greater than p:

$$P\left[\left(\mathbf{x} - \boldsymbol{\mu}\right)^T \boldsymbol{\Sigma}^{-1}\left(\mathbf{x} - \boldsymbol{\mu}\right) \geq \varepsilon\right] \leq \frac{\beta_{2,M}\left(\mathbf{x}\right) - p^2}{\varepsilon^2 - 2p\varepsilon + \beta_{2,M}\left(\mathbf{x}\right)} \leq \frac{\beta_{2,M}\left(\mathbf{x}\right)}{\varepsilon^2}.$$

In the third place, the Mardia's kurtosis of a random vector is not smaller than the squared number of the vector's components:

$$\beta_{2,M}\left(\mathbf{x}\right) \geq p^2.$$

The sample counterpart of Mardia's kurtosis is defined as follows. Let $\mathbf{m}$ and $\mathbf{S}$ be the mean and the variance of the $n \times p$ data matrix $\mathbf{X}$. The Mardia's kurtosis of $\mathbf{X}$, that is the sample kurtosis, is

$$b_{2,M}\left(\mathbf{X}\right) = \frac{1}{n}\sum_{i=1}^{n}\left[\left(\mathbf{x}_i - \mathbf{m}\right)^T \mathbf{S}^{-1}\left(\mathbf{x}_i - \mathbf{m}\right)\right]^2.$$

2   Main Results

Theorem 1. *Let $\mathbf{x}$ be a p-dimensional random vector with nonsingular covariance matrix and finite fourth-order moments. Also, let $\mathbf{A}$ be a nonsingular and nondegenerate $p \times p$ random matrix with finite fourth-order moments and independent of $\mathbf{x}$. Then the Mardia's kurtosis of $\mathbf{A}\mathbf{x}$ is greater than the Mardia's kurtosis of $\mathbf{x}$.*

Let's consider an application of the above theorem to multivariate financial time series. [3] defined a multivariate GARCH process as

$$\left\{\mathbf{r}_t = \boldsymbol{\Sigma}_t^{1/2}\mathbf{z}_t, \ t \in \mathbb{Z}\right\}, \ \text{where} \ \left\{\mathbf{z}_t \sim N\left(\mathbf{0}_d, \mathbf{I}_d\right), \ t \in \mathbb{Z}\right\}$$

is a strong white noise with standard normal components and $\{\mathbf{\Sigma}_t\}$ is a process satisfying

$$vec\left(\mathbf{\Sigma}_t\right) = vec\left(\mathbf{\Gamma}\right) + \sum_{i=1}^{q} \mathbf{A}_i \left(\mathbf{z}_t \otimes \mathbf{z}_t\right) + \sum_{j=1}^{p} \mathbf{B}_j vec\left(\mathbf{\Sigma}_{t-j}\right),$$

where $\mathbf{\Gamma}$ is a $d \times d$ symmetric and positive definite matrix, $\mathbf{A}_i$ and $\mathbf{B}_j$ are $d^2 \times d^2$ positive definite matrices such that $\mathbf{\Sigma}_t$ is a symmetric matrix, while $vec\left(\mathbf{M}\right)$ is the vectorization of the matrix $\mathbf{M}$, that is the vector obtained by stacking the columns of $\mathbf{M}$ on top of each other, for $i = 1, ..., q$ and $j = 1, ..., p$.

Let $r_{t,i}$ and $z_{t,i}$ be the i-th components of $\mathbf{r}_t$ and $\mathbf{z}_t$. Normality of $z_{t,i}$ and ordinary properties of GARCH models imply that $r_{t,i}$ is leptokurtic, i.e. its kurtosis is greater than three (see, e.g., [2] and [1]). The above theorem generalizes this well-known feature of univariate GARCH models to multivariate GARCH models, in that it implies that the Mardia's kurtosis of $\mathbf{r}_t$ is greater than the Mardia's kurtosis of $\mathbf{z}_t$:

$$\beta_{2,M}\left(\mathbf{r}_t\right) > \beta_{2,M}\left(\mathbf{z}_t\right) = d\left(d+2\right).$$

The last equality follows from the results in [13].

We assess the practical relevance of the above theoretical result with 1291% logarithmic daily returns recorded from 25 June 2003 to 23 June 2008 in the French, Dutch and Spanish financial markets. The same dataset already appeared in the financial literature (see, e.g., [5]). Returns are arranged in a 1291×3 data matrix where each row represents a day and each column a country.

The Mardia's kurtosis of the univariate samples associated with the French, the Dutch and the Spanish markets are 7.66, 7.15 and 11.00. The Mardia's kurtosis of the bivariate samples associated with the French and Dutch markets, the French and Spanish markets, the Dutch and Spanish markets are 14.62, 17.64, 17.80. The Mardia's kurtosis of the trivariate sample associated with the French, the Dutch and the Spanish markets is 27.07. The Mardia's kurtosis of the univariate, bivariate and trivariate samples at hand are about twice their expected values under normality, that is 3, 8 and 15. The empirical results may be explained by the above theorem, under the assumption that the data generating model is multivariate GARCH.

3   Theorem's Proof

The Mardia's kurtosis is invariant with respect to one-to-one affine transformations, so that we can assume without loss of generality that $\mathbf{x}$ is a standard random vector:

$$E\left(\mathbf{A}\mathbf{x}\right) = \mathbf{0}_p \text{ and } cov\left(\mathbf{A}\mathbf{x}\right) = \mathbf{I}_p,$$

where $\mathbf{0}_p$ and $\mathbf{I}_p$ are the p-dimensional null vector and the $p \times p$ identity matrix. The Mardia's kurtosis of a standard random vector coincides with the trace of its fourth moment [10]:

$$\beta_{2,M}(\mathbf{x}) = tr(\mathbf{M}_{4,\mathbf{x}}), \text{ where } \mathbf{M}_{4,\mathbf{x}} = E\left(\mathbf{x}\mathbf{x}^T \otimes \mathbf{x}\mathbf{x}^T\right).$$

The $p^2 \times p^2$ matrix $\mathbf{M}_{4,\mathbf{x}}$ is symmetric and positive semidefinite, with rank r at most equal to $p(p+1)/2$ [8].

Let $\boldsymbol{\Sigma}$ be the expectation of the product between the matrix $\mathbf{A}$ and its transpose:

$$\boldsymbol{\Sigma} = E\left(\mathbf{A}\mathbf{A}^T\right).$$

Independence between $\mathbf{A}$ and $\mathbf{x}$, together with the covariance structure of $\mathbf{x}$, imply that the covariance of $\mathbf{A}\mathbf{x}$ is $\boldsymbol{\Sigma}$. Let $\boldsymbol{\Sigma}^{-1/2}$ be the positive definite symmetric square root of the concentration matrix $\boldsymbol{\Sigma}^{-1}$, that is the inverse of $\boldsymbol{\Sigma}$. It follows that $\mathbf{B}\mathbf{x}$, where $\mathbf{B} = \boldsymbol{\Sigma}^{-1/2}\mathbf{A}$, is a p-dimensional standard random vector whose Mardia's kurtosis equal the Mardia's kurtosis of $\mathbf{A}\mathbf{x}$:

$$E(\mathbf{B}\mathbf{x}) = \mathbf{0}_p, \text{ cov }(\mathbf{B}\mathbf{x}) = \mathbf{I}_p \text{ and } \beta_{2,M}(\mathbf{A}\mathbf{x}) = \beta_{2,M}(\mathbf{B}\mathbf{x}) = tr(\mathbf{M}_{4,\mathbf{B}\mathbf{x}}),$$

where $\mathbf{M}_{4,\mathbf{B}\mathbf{x}}$ is the fourth moment of $\mathbf{B}\mathbf{x}$. The identity

$$\mathbf{M}_{4,\mathbf{B}\mathbf{x}} = E\left[(\mathbf{B} \otimes \mathbf{B})\mathbf{M}_{4,\mathbf{x}}\left(\mathbf{B}^T \otimes \mathbf{B}^T\right)\right]$$

follows from $\mathbf{x}$ being independent of $\mathbf{A}$ and the fourth moment of the linear transformation $\mathbf{U}\mathbf{x}$ of $\mathbf{x}$, where $\mathbf{U}$ is a $k \times d$ matrix, being

$$\mathbf{M}_{4,\mathbf{x}} = (\mathbf{U} \otimes \mathbf{U})\mathbf{M}_{4,\mathbf{x}}\left(\mathbf{U}^T \otimes \mathbf{U}^T\right)$$

[8]. Ordinary properties of a matrix's trace and of Mardia's kurtosis lead to

$$\beta_{2,M}(\mathbf{B}\mathbf{x}) = tr\left\{E\left[(\mathbf{B} \otimes \mathbf{B})\mathbf{M}_{4,\mathbf{x}}\left(\mathbf{B}^T \otimes \mathbf{B}^T\right)\right]\right\} =$$

$$E\left\{tr\left[\mathbf{M}_{4,\mathbf{x}}\left(\mathbf{B}^T \otimes \mathbf{B}^T\right)(\mathbf{B} \otimes \mathbf{B})\right]\right\}.$$

For matrices $\mathbf{X}$, $\mathbf{Y}$, $\mathbf{Z}$ and $\mathbf{W}$ of appropriate size, the identity

$$(\mathbf{X} \otimes \mathbf{Y})(\mathbf{Z} \otimes \mathbf{W}) = \mathbf{X}\mathbf{Z} \otimes \mathbf{Y}\mathbf{W}$$

holds true ([14], page 194). As a direct consequence, we have

$$\mathbf{M}_{4,\mathbf{x}}\left(\mathbf{B}^T \otimes \mathbf{B}^T\right)(\mathbf{B} \otimes \mathbf{B}) = \mathbf{M}_{4,\mathbf{x}}(\mathbf{Q} \otimes \mathbf{Q}), \text{ where } \mathbf{Q} = \mathbf{B}^T\mathbf{B}.$$

Let $\mathbf{v}_i$ be the eigenvector associated with the positive eigenvalue λ_i of $\mathbf{M}_{4,\mathbf{x}}$, so that $\mathbf{v}_i$ is the vectorization of a symmetric $p \times p$ matrix $\mathbf{V}_i$ [11]. The fourth moment of $\mathbf{x}$ might be represented as a linear combination of the Kronecker squares of the matrices $\mathbf{V}_1, ..., \mathbf{V}_r$ [12]:

$$\mathbf{M}_{4,\mathbf{x}} = \sum_{i=1}^{r} \lambda_i \mathbf{V}_i \otimes \mathbf{V}_i.$$

The above representation, together with the distributive property of the Kronecker product lead to

$$\mathbf{M}_{4,\mathbf{x}}\left(\mathbf{Q}\otimes\mathbf{Q}\right) = \sum_{i=1}^{r}\lambda_i\mathbf{V}_i\mathbf{Q}\otimes\mathbf{V}_i\mathbf{Q}.$$

The trace of the Kronecker product of two square matrices is the product of the matrices' traces, so that

$$tr\left[\mathbf{M}_{4,\mathbf{x}}\left(\mathbf{Q}\otimes\mathbf{Q}\right)\right] = \sum_{i=1}^{r}\lambda_i tr\left(\mathbf{V}_i\mathbf{Q}\otimes\mathbf{V}_i\mathbf{Q}\right) = \sum_{i=1}^{r}\lambda_i tr^2\left(\mathbf{V}_i\mathbf{Q}\right).$$

Any two matrices $\mathbf{X}$ and $\mathbf{Y}$ of the same size satisfy the identity

$$tr\left(\mathbf{X}^T\mathbf{Y}\right) = vec^T(\mathbf{Y})vec(\mathbf{X}).$$

Therefore, the following identity holds true:

$$\sum_{i=1}^{r}\lambda_i tr^2\left(\mathbf{V}_i\mathbf{Q}\right) = \sum_{i=1}^{r}\lambda_i vec^T\left(\mathbf{Q}\right)vec\left(\mathbf{V}_i\right)vec^T\left(\mathbf{V}_i\right)vec\left(\mathbf{Q}\right).$$

By definition $\mathbf{v}_i = vec\left(\mathbf{V}_i\right)$ is the eigenvector associated with the positive eigenvalue λ_i of $\mathbf{M}_{4,\mathbf{x}}$, so that

$$tr\left[\mathbf{M}_{4,\mathbf{x}}\left(\mathbf{Q}\otimes\mathbf{Q}\right)\right] = \sum_{i=1}^{r}\lambda_i vec^T\left(\mathbf{Q}\right)\mathbf{v}_i\mathbf{v}_i^T vec\left(\mathbf{Q}\right)$$

$$= vec^T\left(\mathbf{Q}\right)\left(\sum_{i=1}^{r}\lambda_i\mathbf{v}_i\mathbf{v}_i^T\right)vec\left(\mathbf{Q}\right) = vec^T\left(\mathbf{Q}\right)\mathbf{M}_{4,\mathbf{x}}vec\left(\mathbf{Q}\right).$$

By assumption, $\mathbf{A}$ is a nonsingular and nondegenerate $p \times p$ matrix, so that the covariance of $vec\left(\mathbf{Q}\right)$ is a positive definite, symmetric $p^2 \times p^2$ matrix:

$$cov\left[vec\left(\mathbf{Q}\right)\right] = E\left[vec\left(\mathbf{Q}\right)vec^T\left(\mathbf{Q}\right)\right] - E\left[vec\left(\mathbf{Q}\right)\right]E\left[vec^T\left(\mathbf{Q}\right)\right] > \mathbf{O}_{p^2\times p^2}.$$

The expected value of $\mathbf{Q}$ is the $p\times p$ identity matrix, thus leading to the inequality

$$E\left[vec\left(\mathbf{Q}\right)vec^T\left(\mathbf{Q}\right)\right] > vec\left(\mathbf{I}_p\right)vec^T\left(\mathbf{I}_p\right).$$

Let $\mathbf{S}$ be the positive semidefinite, symmetric square root of $\mathbf{M}_{4,\mathbf{x}}$ and apply linear properties of the expected value to obtain

$$E\left[\mathbf{S}vec\left(\mathbf{Q}\right)vec^T\left(\mathbf{Q}\right)\mathbf{S}\right] > \mathbf{S}vec\left(\mathbf{I}_p\right)vec^T\left(\mathbf{I}_p\right)\mathbf{S}.$$

The above inequality, ordinary properties of the trace and the definition of $\mathbf{S}$ imply the following one:

$$tr\left\{E\left[\mathbf{S}vec\left(\mathbf{Q}\right)vec^T\left(\mathbf{Q}\right)\mathbf{S}\right]\right\}$$

$$= E\left[vec^T\left(\mathbf{Q}\right)\mathbf{M}_{4,\mathbf{x}}vec\left(\mathbf{Q}\right)\right]$$

$$> vec^T\left(\mathbf{I}_p\right)\mathbf{M}_{4,\mathbf{x}}vec\left(\mathbf{I}_p\right).$$

The right-hand side of the above inequality is the Mardia's kurtosis of $\mathbf{A}\mathbf{x}$, while its left-hand side is the Mardia's kurtosis of $\mathbf{x}$ [11]. We can then write $\beta_{2,M}\left(\mathbf{A}\mathbf{x}\right) > \beta_{2,M}\left(\mathbf{x}\right)$ and complete the proof.

References

1. Bauwens, L., Laurent, S., Rombouts, J.V.K.: Multivariate GARCH models: a survey. J. Appl. Economet. **21**, 79–109 (2006)
2. Bollerslev, T.: Modelling the coherence in short-run nominal exchange rates: a multivariate generalized arch model. Rev. Econ. Stat. **72**, 498–505 (1990)
3. Bollerslev, T., Engle, R.F., Wooldridge, J.M.: A capital asset pricing model with time varying covariances. J. Polit. Econ. **96**, 116–131 (1988)
4. De Luca, G., Loperfido, N.: A skew-in-mean GARCH model for financial returns. In: Skew-Elliptical Distributions and Their Applications: A Journey Beyond Normality, pp. 205–222. CRC/Chapman & Hall (2004)
5. De Luca, G., Loperfido, N.: Modelling multivariate skewness in financial returns: a SGARCH approach. Eur. J. Finance **21**, 1113–1131 (2015)
6. De Luca, G., Genton, M., Loperfido, N.: A multivariate skew-garch model. In: Terrell, D. (ed.) Advances in Econometrics: Econometric Analysis of Economic and Financial Time Series, Part A (Special volume in honor of Robert Engle and Clive Granger, the 2003 winners of the Nobel prize in Economics), vol. 20, pp. 33–56. Elsevier, Oxford (2006)
7. Engle, R.F.: Autoregressive conditional heteroskedasticity with estimates of the variance of U.K. Inflation. Econometrica **50**, 987–1008 (1982)
8. Franceschini, C., Loperfido, N.: On some inequalities between measures of multivariate kurtosis, with application to financial returns. In: Perna, C., Sibillo, M. (eds.) Mathematical and Statistical Methods for Actuarial Sciences and Finance, pp. 211–218. Springer, Cham (2012)
9. Kollo, T.: Multivariate skewness and kurtosis measures with an application in ICA. J. Multivar. Anal. **99**, 2328–2338 (2008)
10. Kollo, T., Srivastava, M.S.: Estimation and testing of parameters in multivariate Laplace distribution. Commun. Stat.-Theory Methods **33**, 2363–2687 (2005)
11. Loperfido, N.: Spectral analysis of the fourth moment matrix. Linear Algebra Appl. **435**, 1837–1844 (2011)
12. Loperfido, N.: A new kurtosis matrix, with statistical applications. Linear Algebra Appl. **512**, 1–17 (2017)
13. Mardia, K.V.: Measures of multivariate skewness and kurtosis with applications. Biometrika **57**, 519–530 (1970)
14. Rao, C.R., Rao, M.B.: Matrix Algebra and its Applications to Statistics and Econometrics. World Scientific Co. Pte. Ltd, Singapore (1998)

Automatic Balance Mechanisms
in an NDC Pension System
with Disability Benefits

Lorenzo Fratoni[1], Susanna Levantesi[1], and Massimiliano Menzietti[2(✉)]

[1] Sapienza University of Rome, Rome, Italy
{lorenzo.fratoni,susanna.levantesi}@uniroma1.it
[2] University of Calabria, Rende, CS, Italy
massimiliano.menzietti@unical.it

Abstract. The aging population is currently impacting several countries worldwide, generating socio-economic needs such as long-term care (LTC) in old age. Developing a strategy linking these care needs to the pension scheme can help in managing the increasing cost of care of pensioners affected by disability. This paper presents a model in which the LTC benefit is integrated into a notional-defined contribution (NDC) pension system. The model's financial sustainability is investigated without and with the application of an automatic balance mechanism (ABM) founded on a Liquidity Ratio. Economic and demographic variables (including the new disability risk element generated by the LTC) are modeled in a stochastic environment.

Keywords: LTC benefits · NDC pension system · Financial sustainability · Automatic balance mechanisms

1 Introduction

An increasing number of countries - Italy included - are affected by the aging population, which has raised growing concern both for the sustainability of the social security system and for the system's ability to extend coverage to new emerging socioeconomic risks, such as the need to receive LTC in old age. Therefore, LTC expenditure is expected to grow significantly in the next years in the advanced countries. In Italy, the LTC public spending was 1.7% of GDP in 2019, and will rise to 1.9% in 2030 and 2.6% in 2050 based on the EU projections [3].

A specific literature on the combination of retirement with LTC benefits has been developed in the recent years (see e.g. [7–9]). Following this line of research, we present a model integrating the LTC benefit into an NDC pension system. The main economic and demographic variables involved in the integrated system are modeled in a stochastic environment.

The presence of LTC benefits adds new risk elements, such as the uncertainty related to the disability rates and mortality rates of disabled [5], which makes the introduction of automatic balance mechanisms (ABMs) to guarantee the system's financial sustainability even more necessary. Specifically, we apply an ABMs based on the liquidity ratio, acting on the indexation of pensions and notional rate as in [1].

2   The Model

Our model considers two benefits: old-age pension and LTC in the form of a life care annuity (LCA). LCA is a coverage providing a LTC benefit which is a percentage increase, η, of the basic pension, b, (with $b^{(h)} = b$): $b^{(d)} = b(1 + \eta)$. The uplift is financed during the accumulation period by an extra contribution. Other benefits are disregarded. We consider a multiple state model with four states: contributor (1), pensioner (2), disabled (3), and dead (4). The transition probabilities between states (contributor $\rightarrow$ pensioner, contributor $\rightarrow$ dead, pensioner $\rightarrow$ disabled, pensioner $\rightarrow$ dead, disabled $\rightarrow$ dead) depend on age and time and not from the years of service. We define the transition probability of an individual aged x in state i at time t to arrive in state j at time $t + h$ as $_h p^{ij}(x, t)$, and the probability for the same individual to remain in state i for time h as $_h p^{ii}(x, t)$.

Denoting $N^k(x, t)$ as the number of individuals in state k at age x at time t and $Z^k(x, t)$ the new entrants, the population dynamics at each time t is given by:

$$N^k(x, t) = N^k(x - 1, t - 1)p^{kk}(x - 1, t - 1) + Z^k(x, t) \quad k = 1, 2, 3, 4 \qquad (1)$$

The total population in the state k at time t is given by $N^k(t) = \sum_x N^k(x, t)$.

Let us consider an overlapping generation model starting at time $t = 0$ with $N^k(0) = 0 \ \forall k$. x_a is the entry age into the contributor state and x_r is the retirement age, both are assumed to be constant over time. $Z^1(x_a, t)$ is assumed constant $\forall t$. Therefore, assuming that there are no deaths among contributors, the active population stabilizes in $x_r - x_a$ years.

Denoting $s(x, t, i)$ as the wage for the i-th active aged x at time t and assuming the same wage for all the individuals belonging to the same generation, $s(x, t, i) = s(x, t)$ for all i. The individual wage evolves according to a given growth rate, as follows:

$$s(x, t) = s(x, t - 1) [1 + \xi(t)] \quad \text{for } x < x_r \qquad (2)$$

Where $\xi(t)$ is the growth rate of individual wage from $t - 1$ to t.

The total wage earned by the active population aged x at time t is given by $S(x, t) = N^1(x, t)s(x, t)$ (for $x < x_r$). While, denoting $S(t) = \sum_x S(x, t)$ as the total wage at time t, the average wage is equal to: $s(t) = \frac{S(t)}{N^1(t)}$.

The individual contribution paid by an active aged x at time t is $c(x,t) = c(t)s(x,t)$ (for $x < x_r$). Where $c(t)$ is the contribution rate of the pension system at time t. The total contribution for the active population aged x at time t is given by: $C(x,t) = c(t)S(x,t)$ (for $x < x_r$). We denote with $C(t)$ the amount of total contribution earned by the system, $C(t) = \sum_x C(x,t)$.

In an NDC system, the contribution rate is fixed and assumed constant over time, $c(t) = c$ for all t. Each participant's contributions are booked on an individual notional account and remunerated each year t at the same rate of return, $g(t)$. The individual notional account for an active i aged x at the end of year t is calculated as:

$$m(x,t,i) = [m(x-1,t-1,i) + c(t)s(x,t)][1+g(t)] \quad \text{for } x < x_r \quad (3)$$

At retirement, the initial benefit is determined by converting the individual notional account into an annuity consistently with the remaining cohort life expectancy, the expected indexation rate, λ^*, and the expected rate of return, g^*. Considering a new healthy pensioner aged x_r at time t and assuming benefits paid in advance, the annuity rate $\ddot{a}(x_r,t)$ is calculated as:

$$\ddot{a}^{22}(x_r,t) = \sum_{h=0}^{t+h} {}_hp^{22*}(x_r,t) \cdot \prod_{k=t}^{t+h} [1+j^*(k)]^{-1} \quad (4)$$

where $j^*(k) = \frac{1+g^*(k)}{1+\lambda^*(k)} - 1$ is the expected rate measuring the amount by which the notional rate deviates from pension indexation. Similarly, the annuity rate for a new disabled pensioner aged $x \geq x_r$ at time t is determined by: $\ddot{a}^{23}(x,t) = \sum_{h=0} {}_hp^{23*}(x,t) \cdot \prod_{k=t}^{t+h} [1+j^*(k)]^{-1}$.

Considering LCA benefits, the basic pension amount b paid to a new (healthy) pensioner is the same as what would be paid in an NDC system without LTC:

$$b(x_r,t,i) = \frac{m'(x_r,t,i)}{\ddot{a}^{22}(x_r,t) + (1+\eta)\ddot{a}^{23}(x_r,t)} = \frac{m(x_r,t,i)}{\ddot{a}^{22}(x_r,t) + \ddot{a}^{23}(x_r,t)} \quad (5)$$

where $m'(x_r,t,i)$ is the value of the notional amount to be accumulated to finance both old-age pension and LCA benefits ($m'(x_r,t,i) > m(x_r,t,i)$). Starting from Eq. 5 we can determine the new contribution rate $c'(x,t)$:

$$\frac{c'(x,t)}{c(x,t)} = \frac{m'(x_r,t,i)}{m(x_r,t,i)} = \frac{\ddot{a}^{22}(x_r,t) + (1+\eta)\ddot{a}^{23}(x_r,t)}{\ddot{a}^{22}(x_r,t) + \ddot{a}^{23}(x_r,t)} \quad (6)$$

When the LCA is introduced, we define $C'(x,t)$ and $C'(t)$ consistently to $C(x,t)$ and $C(t)$.

We denote $B_z^2(x_r,t)$ and $B_z^3(x,t)$ the total benefits paid to the new healthy and disabled pensioners in the year t, respectively. The total pensions paid to all retirees aged x in the year t evolve as follows:

$$B^2(x,t) = B^2(x-1,t-1)p^{22}(x-1,t-1)[1+\lambda(t-1)] \quad \text{for } x > x_r \quad (7)$$

$$B^3(x,t) = B^3(x-1,t-1)p^{33}(x-1,t-1)\left[1 + \lambda(t-1)\right] + B_z^3(x,t) \quad \text{for} \quad x > x_r \tag{8}$$

where λ, could be different from its estimated value, λ^*.

Finally, let $B(t) = \sum_{x,i} B^i(x,t)$, $i = 2,3$, be the amount of total pensions paid to all retirees in the year t. The corresponding average pension is given by: $b(t) = \frac{B(t)}{N^2(t)+N^3(t)}$.

3   Financial Sustainability and ABMs

In a balanced PAYG scheme we have $C'(t) = B(t)$. Considering the previous equations, such condition can be expressed as: $N^1(t) \cdot c'(t) \cdot s(t) = (N^2(t) + N^3(t)) \cdot b(t)$. The contribution rate satisfying the equilibrium equation is defined as:

$$\hat{c}'(t) = \frac{N^2(t) + N^3(t)}{N^1(t)} \cdot \frac{b(t)}{s(t)} = D(t) \cdot r(t) \tag{9}$$

where $D(t) = \frac{N^2(t)+N^3}{N^1(t)}$ denotes the dependency ratio, and $r(t) = \frac{b(t)}{s(t)}$ denotes the average replacement rate of the system in the year t.

A PAYG system could experience periods with cash-flow deficit (surplus), $C'(t) < B(t)$ ($C'(t) > B(t)$). We measure the liquidity of the system in the year t by the unfunded liabilities (UL), defined as: $UL(t) = B(t) - C'(t)$.

This implies the accumulation of a reserve fund $F(t)$ (if $UL < 0$), or a pension liability (if $UL > 0$). Assuming the rate of return of $F(t)$ equal to the notional rate, the evolution of $F(t)$ is given by: $F(t) = F(t-1)\left[1 + g(t-1)\right] + C'(t) - B(t)$.

Demographic evolution and/or economic dynamics could undermine the PAYG equilibrium. In an NDC system, as the contribution rate is constant, the equilibrium can be reached by changing the replacement rate, which means changing the notional rate and the expected probabilities used in Eq. 4. But, as observed by [2], there are situations where an NDC system is not able to immediately restore the equilibrium, thus remaining vulnerable to demographic and economic shocks. We consider an ABM based on liquidity ratio defined as follows, and designed to attain a liquidity ratio of 1:

$$LR(T) = \frac{C'(t) + F^-(t)}{B(t)} \tag{10}$$

Both the notional rate $g(t)$ and the pension indexation $\lambda(t)$ are adjusted:

$$(1 + g^{LR}(t)) = (1 + g(t))I^{LR}(t) \tag{11}$$

$$(1 + \lambda^{LR}(t)) = (1 + \lambda(t))I^{LR}(t) \tag{12}$$

where $I^{LR}(t) = \frac{C(t)+F^{LR-}(t)}{B^{LR}(t)}$.

4   Numerical Application and Conclusions

We model both transition probabilities and macroeconomic variables in a stochastic environment to take into account demographic and economic risks, which are assumed to be independent. The economic data are taken from the Italian National Institute of Statistics (ISTAT) for the period 1983-2019. We model the inflation rate and wage growth rate with a VAR(1) with an exogenous long-run trend of 2% and 3.5%, respectively (see [4] for further details). We assume that the transition probabilities $p^{12}(x,t)$ are deterministic, while we model $p^{23}(x,t)$, $p^{24}(x,t)$ and $p^{34}(x,t)$ with CBD models including an age-dependent cohort effect (see [6] for a detailed description of the model and parameter estimation based on the data of the Italian National Institute of Social Security (INPS) on people qualified to a universal disability benefit over the period 2002–2012 for ages 40–89). We disregard contributors mortality: $p^{14}(x,t) = 0$ $\forall x, t$. Finally, $p^{i1}(x,t)$ with $i = 2, 3$ are assumed to be 0 $\forall x, t$. x_a and x_r are respectively set at 25 and 65.

We consider two scenarios: one refers to a standard NDC pension system without LTC benefits ("base scenario"), the other refers to an NDC system with an LCA benefit ("LCA Scenario"). The time horizon is set to $T = 150$ years in order to achieve the full development of the overlapping generation model.

Based on the features of the Italian NDC system, the notional rate is equal to the 5-years moving average of the GDP growth rate, and the pension indexation rate is equal to the inflation rate. We assume that the GDP growth rate is equal to the sum of the growth rate of the active population and the growth rate of the individual wage. In the actuarial valuation, we consider a constant discount rate equivalent to the long run trend of notional GDP growth rate.

Expected values and confidence intervals of the main indicators of the pension system at the end of forecasting period are reported in the following table (Table 1).

Table 1. Expected values (and confidence interval at 90%) of $D(T)$, $r(T)$ and $\hat{c}(T)$, without and with ABM.

Indicator	Base scenario		LCA scenario	
	No ABM	With ABM	No ABM	With ABM
$D(T)$	0.658			
	(0.547; 0.744)			
$r(T)$	0.461	0.459	0.479	0.475
	(0.426; 0.499)	(0.403; 0.545)	(0.437; 0.526)	(0.416; 0.564)
$\hat{c}'(T)$	0.303	0.299	0.314	0.309
	(0.255; 0.346)	(0.296; 0.302)	(0.273; 0.353)	(0.306; 0.312)

The results show that the introduction of the LCA benefit involves an increase in the average replacement rate of the system $(+1.8\%)$ together with an increase in the equilibrium contribution rate $(+1.1\%)$. The introduction of the ABM based on the liquidity ratio slightly reduces the replacement rates, producing an alignment of the contribution rates to 0.30 and 0.31, respectively. Obviously, with the introduction of the ABM, the uncertainty of $\hat{c}'(t)$ is significantly reduced (with and without LCA benefit), but, at the same time, the uncertainty of the replacement rate increases $(+94.5\%$ in the base scenario, $+66.3\%$ in the LCA scenario). The results show that the introduction of an ABM could help to improve the stability of the system, in both base and LCA scenarios. On the other hand, the introduction of an ABM based on the liquidity ratio transfers risk to retirees, increasing the uncertainty of the replacement rate. Future researches may analyze the effect of alternative ABMs to reduce the replacement rate uncertainty. Further development could consist of studying alternative LTC benefits, for example, the enhanced pension.

References

1. Alonso-García, J., Boado-Penas, M.D.C., Devolder, P.: Automatic balancing mechanisms for notional defined contribution accounts in the presence of uncertainty. Scand. Actuar. J. **2**, 85–108 (2018)
2. Devolder, P., Levantesi, S., Menzietti, M.: Automatic balance mechanisms for notional defined contribution pension systems guaranteeing social adequacy and financial sustainability: an application to the italian pension system. Ann. Oper. Res. **299**, 765–795 (2020). https://doi.org/10.1007/s10479-020-03819-x
3. European Commission: 2021 Long term care report. Trends, challenges and opportunities in an ageing society, vol. 2. Joint report prepared by the Social Protection Committee and the European Commission. European Union (2021). https://doi.org/10.2767/183997
4. Fratoni, L., Levantesi, S., Menzietti, M.: Measuring financial sustainability and social adequacy of the Italian NDC pension system under the COVID-19 pandemic. Rapporto Tecnico del Dipartimento di Scienze Statistiche, 1/2021, Roma (2021). ISSN 2279-798X
5. Levantesi, S., Menzietti, M.: Managing longevity and disability risks in life annuities with long term care. Insurance Math. Econ. **50**, 391–401 (2012). ISSN 0167-6687
6. Levantesi, S., Menzietti, M.: Natural hedging in long term care insurance. Astin Bull. **48**(1), 233–274 (2018). https://doi.org/10.1017/asb.2017.29. ISSN 0515-0361
7. Pla-Porcel, J., Ventura-Marco, M., Vidal-Meliá, C.: Life care annuities (LCA) embedded in a notional defined contribution (NDC) framework. Astin Bull. **46**(2), 331–363 (2016). https://doi.org/10.1017/asb.2015.27
8. Ventura-Marco, M., Vidal-Meliá, C.: Integrating retirement and permanent disability in NDC pension schemes. Appl. Econ. **48**(12), 1081–1102 (2016). https://doi.org/10.1080/00036846.2015.1093084
9. Vidal-Meliá, C., Ventura-Marco, M., Pla-Porcel, J.: An NDC approach to helping pensioners cope with the cost of long-term care. J. Pension Econ. Finance **19**(1), 80–108 (2020). https://doi.org/10.1017/S1474747218000070

Deep Neural Network Algorithms for Parabolic PIDEs and Applications in Insurance Mathematics

Rüdiger Frey and Verena Köck[✉]

Institute for Statistics and Mathematics, Vienna University of Economics and Business, Welthandelsplatz 1, 1020 Vienna, Austria
{rfrey,verena.koeck}@wu.ac.at

Abstract. In recent years a large literature on deep learning based methods for the numerical solution partial differential equations has emerged; results for integro-differential equations on the other hand are scarce. In this short paper we study deep neural network algorithms for solving linear parabolic partial integro-differential equations with boundary conditions in high dimension. To show the viability of our approach we discuss a test case study from insurance.

Keywords: Deep neural networks · Parabolic partial integro-differential equations · Machine learning · Insurance

1 Introduction

Many problems in insurance and finance lead to terminal or boundary value problems involving parabolic partial integro-differential equations (PIDEs). Examples include option pricing in models with jumps, the valuation of insurance contracts, ruin probabilities in non-life insurance, optimal reinsurance problems and many applications in credit risk. These PIDEs can be linear (such as PIDEs arising in risk-neutral pricing) or semilinear (such as the dynamic programming equation in many stochastic control problems). Practical applications often involve several underlying assets or economic factors, so that one has to deal with PIDEs in a high-dimensional space. These PIDEs do typically not admit an analytic solution, making the design of suitable numerical methods an ongoing challenge. Existing numerical methods include deterministic schemes such as finite difference and finite element methods and random schemes based on Monte-Carlo methods. However, finite difference and finite element methods cannot be used in the case of high-dimensional PIDEs as they suffer from the curse of dimensionality. Monte-Carlo methods on the other hand are suitable for problems in higher dimensions. However, these methods only provide a solution for a single fixed time-space point (t, x). This is problematic in risk management applications, where one needs to find the solution of a pricing problem for a large set D of future scenarios. The naive solution via nested Monte Carlo is in most

cases computationally infeasible. For these reasons many recent contributions study machine learning techniques for the numerical solution of PDEs. A large strand of this literature is based on the representation of semilinear parabolic PDEs via backward stochastic differential equations (BSDEs), see e.g. the seminal papers [6] and [4]. Applications of deep learning methods to partial *integro* differential equations on the other hand are scarce.

In this short paper we consider a deep neural network (DNN) algorithm for linear parabolic PIDEs that generalizes the regression approach of [2]. In the extended version [5] we additionally consider an algorithm for the semilinear case generalizing the deep splitting method of [1]. In the semilinear case we first linearize the equation locally in time using a time grid t_n, $n = 0, \ldots, N$. Then we perform a backward induction over the grid points, using in each step the DNN algorithm for the linear case. Moreover, we propose a alternative linearization procedure to [1] and we apply our DNN algorithms also to boundary value problems, while [2] and [1] consider only pure Cauchy problems.

Our focus is on applications to insurance and finance. To assess performance of our methodology we carry out extensive tests for PIDEs arising in actuarial mathematics. A formal error analysis is left to future research. In this short paper we consider pricing of a stop-loss type reinsurance contract in a model where claims arrive with stochastic intensity (see, e.g. [3]) as a first test case. In the extended version [5] we present additional insurance and finance related examples both for the linear and the semilinear case. For other applications of neural networks to insurance see e.g. [7]. Our experiments show that the performance is satisfying in terms of accuracy and speed. The relative L^1-approximation error is close to zero and the computation time is significantly smaller compared to Monte-Carlo methods in cases when the solution is computed on a set D.

The paper is organized as follows. Section 2 introduces the general setting; Sect. 3 deals with linear PIDEs.

2   Modeling Framework

We fix a probability space $(\Omega, \mathcal{F}, \mathbf{P})$, a time horizon T and a right continuous filtration $\mathbb{F}$. Consider measurable functions $\mu \colon [0, T] \times \mathbb{R}^d \to \mathbb{R}^d$, $\sigma \colon [0, T] \times \mathbb{R}^d \to \mathbb{R}^{d \times d}$ and $\gamma^X \colon [0, T] \times \mathbb{R}^d \times E \to \mathbb{R}^d$, where $(E, \mathcal{E})$ is a separable measurable space. We assume that $(\Omega, \mathcal{F}, \mathbf{P})$ supports a d-dimensional Brownian motion W and a Poisson random measure J on $[0, T] \times E$. The compensator of J is given by $\nu(\mathrm{d}z)\mathrm{d}t$ for a finite measure ν on E. We consider a d-dimensional process X that is the unique strong solution to the SDE, $X_0 = x \in \mathbb{R}^d$,

$$\mathrm{d}X_t = \mu(t, X_t)\mathrm{d}t + \sigma(t, X_t)\mathrm{d}W_t + \int_E \gamma^X(t, X_{t-}, z)J(\mathrm{d}t, \mathrm{d}z), \qquad (1)$$

We restrict ourselves to finite activity processes to simplify the exposition as it is sufficient for most applications in insurance. It is well known that the SDE (1) has a unique strong solution under mild conditions on the coefficients.

Define the matrix $\Sigma(t,x) = (b_{i,j}(t,x), i,j = 1, \ldots, d)$, by $\Sigma(t,x) = \sigma\sigma^\top(t,x)$ and consider for $u \in \mathcal{C}^{1,2}([0,T] \times \mathbb{R}^d)$ the integro-differential operator $\mathcal{L}$

$$\mathcal{L}u(t,x) := \sum_{i=1}^{d} \mu_i(t,x)u_{x_i}(t,x) + \frac{1}{2}\sum_{i,j=1}^{d} b_{i,j}(t,x)u_{x_i x_j}(t,x) \tag{2}$$

$$+ \int_{\mathbb{R}^d} [u(t, x + \gamma^X(t,x,z)) - u(t,x)]\nu(dz), \quad x \in \mathbb{R}^d, \ t \in [0,T].$$

The operator $\mathcal{L}$ is the generator of X solving the martingale problem for $\mathcal{L}$.

Consider functions $c\colon [0,T] \times \mathbb{R}^d \to \mathbb{R}$, $r\colon [0,T] \times \mathbb{R}^d \to \mathbb{R}$ and $g\colon [0,T] \times \mathbb{R}^d \to \mathbb{R}$, and let D be an open subset of $\mathbb{R}^d$. In Sect. 3 we are interested in the following boundary value problem

$$u_t(t,x) + \mathcal{L}u(t,x) - r(t,x)u(t,x) + c(t,x) = 0, \quad (t,x) \in [0,T) \times D, \tag{3}$$

and $u(t,x) = g(t,x)$ for $(t,x) \in \big([0,T) \times (\mathbb{R}^d \backslash D)\big) \cup \big(\{T\} \times \mathbb{R}^d\big)$. The special case $D = \mathbb{R}^d$ corresponds to a pure Cauchy problem without boundary conditions; in that case we use the simpler notation $g(T,x) =: \varphi(x)$ to denote the terminal condition. It follows from the Feynman-Kac formula that under some integrability conditions a classical solution u of (3) has the probabilistic representation

$$u(t,x) = \mathbb{E}_{t,x}\Big[\int_t^{\bar\tau} e^{-\int_t^s r(u,X_u)du} c(s,X_s)ds + e^{-\int_t^{\bar\tau} r(u,X_u)du} g\big(\bar\tau, X_{\bar\tau}\big)\Big], \tag{4}$$

where $\bar\tau := T \wedge \tau$ and $\tau := \inf\{s \geq t\colon X_s \notin D\}$.

In Sect. 3 we propose a deep neural network (DNN) algorithm to approximate the function u defined in (4). In the extended version [5] we are interested in semilinear problems of the form $u_t(t,x) + \mathcal{L}u(t,x) + f(t,x,u(t,x),\nabla u(t,x)) = 0$, $(t,x) \in [0,T) \times D$, where $f\colon [0,T] \times \mathbb{R}^d \times \mathbb{R} \times \mathbb{R}^d \to \mathbb{R}$ is a nonlinear function such as the Hamiltonian in a typical Hamilton Jacobi Bellman equation.

3   Deep Neural Network Approximation for Linear PIDEs

3.1   Representation as Solution of a Minimization Problem

Fix some time point $t \in [0,T)$ and a closed and bounded set $A \subset \overline{D}$. Define the function $u\colon [0,T] \times \mathbb{R}^d \to \mathbb{R}$ by the Feynman-Kac representation (4). We want to compute an approximation to the function $u(t,\cdot)$ on the set A. The key idea is to write this function as solution of a minimization problem.

Consider some random variable ξ whose distribution is absolutely continuous with respect to the Lebesgue measure such that the corresponding density has support A (in applications the distribution of ξ is often the uniform distribution on A) and denote by X^ξ the solution of the SDE (1) with initial value $X_t = \xi$. Define the random variable

$$Y^\xi := \int_t^{T \wedge \tau} e^{-\int_t^s r(u,X_u^\xi)du} c(s,X_s^\xi)ds + e^{-\int_t^{T \wedge \tau} r(u,X_u^\xi)du} g\Big(T \wedge \tau, X_{T \wedge \tau}^\xi\Big). \tag{5}$$

Assume that $\mathbb{E}[|Y^\xi|^2] < \infty$ and that the function $u(t, \cdot)$ belongs to $\mathcal{C}^0(\overline{A})$. Since X^ξ is a Markov process it holds that $u(t, \xi) = \mathbb{E}[Y^\xi \mid \sigma(\xi)]$, where $\sigma(\xi)$ is the sigma-field generated by ξ. Since Y^ξ is square integrable we thus get from the $\mathcal{L}^2$-minimality of conditional expectations that

$$\mathbb{E}\left[\left|Y^\xi - u(t, \xi)\right|^2\right] = \inf\left\{\mathbb{E}\left[\left|Y^\xi - Z\right|^2\right] : Z \in L^2(\Omega, \sigma(\xi), \mathbf{P})\right\}. \qquad (6)$$

Since $u(t, \cdot) \in \mathcal{C}^0(A)$ and since the density of ξ is strictly positive on A we conclude that $u(t, \cdot)$ is the unique solution of the minimization problem

$$\min \mathbb{E}\left[\left|Y^\xi - v(\xi)\right|^2\right], \quad v \in \mathcal{C}^0(A). \qquad (7)$$

which can be solved with deep learning methods, as we explain next.

3.2 The Algorithm

The first step in solving (7) with machine learning techniques is to simulate trajectories of X^ξ up to the stopping time τ. The simplest method is the Euler-Maruyama scheme. Here we choose a time discretization $t = t_0 < t_1 < \cdots < t_M = T$, $\Delta t_m = t_m - t_{m-1}$, generate K simulations $\xi^{(1)}, \ldots, \xi^{(K)}$ of the random variable ξ and simulate K paths $X^{(1)}, \ldots, X^{(K)}$ of X^ξ up to the stopping time τ by the following recursive algorithm. We let $X_t^{(k)} = \xi^{(k)}$, and for $m \geq 1$,

$$X_{t_m \wedge \tau}^{(k)} := X_{t_{m-1} \wedge \tau}^{(k)} + \mathbf{1}_{(0,\tau)}(t_{m-1})\left(\mu(t_{m-1}, X_{t_{m-1}}^{(k)})\Delta t_m + \sigma(t_{m-1}, X_{t_{m-1}}^{(k)})\Delta W_{t_m}^{(k)} \right.$$
$$\left. + \int_{t_{m-1}}^{t_m} \int_{\mathbb{R}^d} \gamma(t_{m-1}, X_{t_{m-1}}^{(k)}, z)\, J^{(k)}(\mathrm{d}z, \mathrm{d}s)\right). \qquad (8)$$

Note that the integrand in the integral with respect to $J^{(k)}$ is evaluated at t_{m-1} so that this integral corresponds to the increment of a standard compound Poisson process. Using these simulations we compute for each path

$$Y^{(k)} := \int_t^{\bar{\tau}} e^{-\int_t^s r(u, X_u^{(k)})\mathrm{d}u} c(s, X_s^{(k)})\mathrm{d}s + e^{-\int_t^{\bar{\tau}} r(u, X_u^{(k)})\mathrm{d}u} g(\bar{\tau}, X_{\bar{\tau}}^{(k)}), \qquad (9)$$

with $\bar{\tau} := T \wedge \tau$. The integrals can be approximated by Riemann sums.

In the next step we approximate $u(t, \cdot)$ by a deep neural network $\mathcal{U}_t(\cdot) = \mathcal{U}_t(\cdot; \theta) \colon A \to \mathbb{R}^d$. We determine the network parameters θ (training of the network) by minimizing the loss function $\theta \mapsto \frac{1}{K} \sum_{k=1}^K \left(Y^{(k)} - \mathcal{U}_t(\xi^{(k)}; \theta)\right)^2$. For this we rely on stochastic gradient-descent methods; algorithmic details are given in [5]. This approach can be considered as a *regression-based* scheme since one attempts to minimize the squared error between the DNN approximation $\mathcal{U}_t(\cdot; \theta)$ and the given terminal and boundary values of the PIDE.

To test the proposed DNN algorithm we price a reinsurance contract in the model of [3], where the claims process follows a doubly stochastic risk process.

3.3 Example

Valuation of an Insurance Contract with Doubly Stochastic Poisson Arrivals.
We consider an insurance company and a reinsurer who enter into a reinsur-
ance contract with a given maturity $T = 1$. The risk-free interest rate is r. To
model the losses in the insurance portfolio underlying this contract we consider
a sequence $\{T_n\}_{n\in\mathbb{N}}$ of claim arrival times with nonnegative intensity process
$\lambda^L = (\lambda_t^L)_{t\geq 0}$ and a sequence $\{Z_n\}_{n\in\mathbb{N}}$ of claim sizes that are iid strictly posi-
tive random variables independent of the counting process $N = (N_t)_{t\geq 0}$ defined
by $N_t = \sum_{n=1}^\infty \mathbf{1}_{\{T_n \leq t\}}$. The loss process $L = (L_t)_{t\geq 0}$ is given by $L_t = \sum_{n=1}^{N_t} Z_n$.
We assume that the Z_n are Gamma (α, β) distributed with density $f_{\alpha,\beta}(z)$. This
is a common choice in insurance. Moreover, the Gamma distribution is closed
under convolution so that the sum of independent Gamma distributed random
variables can be generated with a single simulation, which speeds up the sam-
pling of trajectories from L. The claim-arrival intensity process λ^L satisfies the
SDE

$$d\lambda_t^L = b(\lambda_t^L)dt + \sigma(\lambda_t^L)dW_t, \quad \lambda_0^L = \lambda_0 \in \mathbb{R}_+, \tag{10}$$

where W is a standard Brownian motion. In this example it is convenient to
write the process X in the form $X_t = (L_t, \lambda_t^L)$. We assume that the reinsurance
contract is a *stop-loss* contract, i.e. the indemnity payment is of the form $\varphi(L_T)$
with $\varphi(l) = [l - K]^+$, with $[z]^+ = \max\{z, 0\}$. The market value u at time $t \in
[0, T]$ of the reinsurance contract is for $(l, \lambda) \in \mathbb{R}_+^0 \times \mathbb{R}_+$ defined by $u(t, l, \lambda) :=
\mathbb{E}_t^{l,\lambda}[e^{-r(T-t)}\varphi(L_T)]$, where $\mathbb{E}_t^{l,\lambda}[\,\cdot\,] := \mathbb{E}[\,\cdot\,|L_t = l, \lambda_t^L = \lambda]$. The authors of [3]
show that u is the unique solution of the PIDE $u_t(t, l, \lambda) + \mathcal{L}u(t, l, \lambda) = 0$ with
terminal condition $u(T, l, \lambda) = \varphi(l)$ and generator

$$\mathcal{L}u(t, l, \lambda) = u_\lambda(t, l, \lambda)b(\lambda) + \frac{1}{2}u_{\lambda\lambda}(t, l, \lambda)\sigma(\lambda)^2 \tag{11}$$

$$+ \lambda \int_\mathbb{R} [u(t, l + z, \lambda) - u(t, l, \lambda)]f_{\alpha,\beta}(z)dz,$$

for $(l, \lambda) \in \mathbb{R}_+^0 \times \mathbb{R}_+$, $t \in [0, T]$. There is no explicit solution for this PIDE, and
we approximate $u(0, l, \lambda)$ on the set $A := \{(l, \lambda) : l \in [0, 30], \lambda \in [90, 130]\}$ with a
deep neural network $\mathcal{U}_0(l, \lambda)$. Parameters used for the valuation of the stop-loss
contract are $b(\lambda) = 0.5(100 - \lambda)$, $\sigma(\lambda) = 0.2\lambda$, $\alpha = 1$, $\beta = 1$, $r = 0$ and $K = 90$.
Paths of the processes L and λ^L are simulated with the Euler-Maruyama scheme
and $\xi \sim \text{Unif}(A)$.

Figure 1 shows the approximate solution $\mathcal{U}_0$ obtained by the DNN algo-
rithm. As a reference we compute for fixed (l, λ) approximate values $U^{MC}(l, \lambda) \approx
u(0, l, \lambda)$ with Monte-Carlo (MC) using 10^6 simulated paths for each point (l, λ)
(paths are simulated with the Euler-Maruyama scheme).

The *relative L^1-error* between the DNN approximation $\mathcal{U}$ and the MC-
solution U^{MC} is defined as $\epsilon := \mathbb{E}\left[\left|\frac{\mathcal{U}(\xi^l, \xi^\lambda) - U^{MC}(\xi^l, \xi^\lambda)}{U^{MC}(\xi^l, \xi^\lambda)}\right|\right]$. Using 1000 simula-
tions of $\xi^l \sim \text{Unif}([0, 30))$, $\xi^\lambda \sim \text{Unif}([90, 130))$ we obtained a relative error of
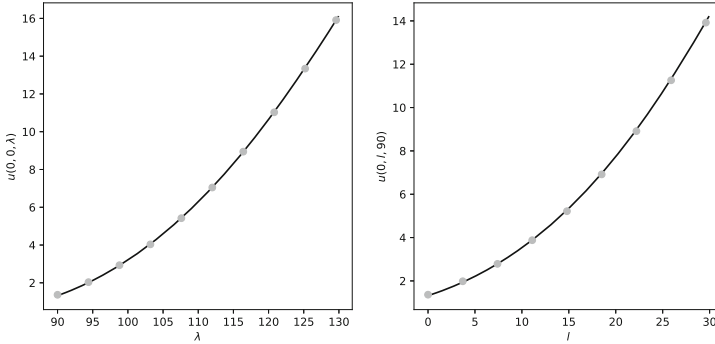
Fig. 1. Solution $u(0, l, \lambda)$ for $l = 0$, $\lambda \in [90, 130]$ (*left*) and $l \in [0, 30]$, $\lambda = 90$ (*right*) computed with the DNN-algorithm (*black*) and reference points with MC (*grey*).

$\epsilon = 0.0018$. On a Lenovo Thinkpad notebook with an Intel Core i5 processor (1.7 GHz) and 16 GB of memory the computation of $\mathcal{U}$ via the training of a DNN took around 322 s, whereas the computation of $U^{MC}(l, \lambda)$ with Monte-Carlo for a fixed point (l, λ) took around 4.3 s. This shows that the DNN approach is faster than the MC approach if one wants to compute $u(t, l_i, \lambda_i)$ for a grid (l_i, λ_i) with more than 100 grid points.

References

1. Beck, C., Becker, S., Cheridito, P., Jentzen, A., Neufeld, A.: Deep splitting method for parabolic PDEs. SIAM J. Sci. Comput. **43**(5), 3135–3154 (2021)
2. Beck, C., Becker, S., Grohs, P., Jaafari, N., Jentzen, A.: Solving the Kolmogorov PDE by means of deep learning. J. Sci. Comput. **88**(3), 1–28 (2021). https://doi.org/10.1007/s10915-021-01590-0
3. Ceci, C., Colaneri, K., Frey, R., Köck, V.: Value adjustments and dynamic hedging of reinsurance counterparty risk. SIAM J. Financ. Math. **11**(3), 788–814 (2020)
4. Weinan, E., Han, J., Jentzen, A.: Deep learning-based numerical methods for high-dimensional parabolic partial differential equations and backward stochastic differential equations. Commun. Math. Stat. **5**, 349–380 (2017). https://doi.org/10.1007/s40304-017-0117-6
5. Frey, R., Köck, V.: Deep neural network algorithms for parabolic PIDEs and applications in insurance mathematics. arXiv preprint arXiv:2109.11403 (2021)
6. Han, J., Jentzen, A., Weinan, E.: Solving high-dimensional partial differential equations using deep learning. Proc. Nat. Acad. Sci. USA **115**, 8505–8510 (2018)
7. Wüthrich, M., Buser, C.: Data analytics for non-life insurance pricing. Swiss Finance Institute Research Paper, pp. 16–68 (2020)

Ergodic Behavior of Returns in a Buy Low and Sell High Type Trading Strategy

Hedvig Gál[1(✉)] and Attila Lovas[2,3]

[1] Corvinus University of Budapest, Fővám Square 8, Budapest 1093, Hungary
hedvig.gal@stud.uni-corvinus.hu
[2] Alfréd Rényi Institute of Mathematics, Reáltanoda Street 13-15,
Budapest 1053, Hungary
lovas.attila@renyi.hu
[3] Budapest University of Technology and Economics, Budapest, Hungary

Abstract. In algorithmic trading strategies aiming at "Buying Low and Selling High" a given asset is a recurrent topic for many practitioners and still pose challenges for researchers. We may ask, for example, what happens in the long run if we set price levels $\underline{\theta} < \overline{\theta}$ and buy the asset if its price goes below $\underline{\theta}$ and sell it if the price exceeds $\overline{\theta}$? In their recent paper, Lovas and Rásonyi proved that under suitable conditions, the distribution of the log-rate of returns realized in each cycle converges to a unique limit distribution in total variation at geometric speed. Furthermore, the law of large numbers holds for bounded and measurable functionals of the returns. We tested these findings by executing the strategy on real stock exchange data consists of in about 2.3 million records, providing empirical evidence for the law of large numbers.

Keywords: Algorithmic trading · Threshold-type strategies · Optimal investment · Stochastic stability · Markov chain · Minorization condition

1 Introduction

The good old but naive advice for beginner traders is to "(B)uy (L)ow and (S)ell (H)igh" (BLSH) a given financial asset. Although, the majority of expert traders refuse to take this obvious approach serious or at least they operate with a more refined versions of such strategies, BLSH has been studied by several authors in the past twenty years from various aspects (see for example [1] and [8]). We focus on two main question arising in this context: (1) When it is advisable to buy/sell a given asset? (2) What happens in the long run if we fix thresholds $\underline{\theta} < \overline{\theta}$ and buy the asset if its price goes below $\underline{\theta}$ and sell it if the price exceeds $\overline{\theta}$ exploiting the so-called up-crossings $[\underline{\theta}, \overline{\theta}]$ to profit from price oscillations?

For instance, the first question was addressed by Dai *et al.* who formulated stochastic optimal stopping problems and elaborated a trading strategy to determine the proper time for a single buy and sell under the assumption that the underlying asset price follows geometric Brownian motion (GBM) [3]. Authors in [9], studied optimal sequential investment strategies. They showed that GBM price process does not allow for optimal buying and selling strategies that have a sequential nature. They also proved that mean-reverting constant elasticity of variance processes may be optimal to sequentially buy and sell. Earlier Zhang and Zhang [10] considered the problem of buying and selling a single asset sequentially to maximize a discounted utility in a model, where the evolution of the underlying asset price is described by a mean-reverting process

$$dY_t = a(b - Y_t)dt + \sigma dB_t, \ Y_0 = y_0, \tag{1}$$

where $b, y_0 \in \mathbb{R}$, $a, \sigma > 0$, and $(B_t)_{t \geq 0}$ denotes the standard Brownian motion. Using the dynamic programming approach, they showed that the optimal trading rule can be determined by two threshold levels $\underline{\theta}, \overline{\theta}$, as in the second question, moreover $\underline{\theta}$ and $\overline{\theta}$ can be expressed in terms of a, b and σ.

Concerning the second question, Lovas and Rásonyi studied the long term behavior of the above-mentioned fixed-threshold version of the BLSH [5]. Inspired by the recent results of Mijatovic and Vysotsky on the stability of the overshoots of random walks with i.i.d. increments [7], they proved for a broad class of price processes that the distribution of the gain and the time between buying and selling converges in the long run, moreover for bounded utility functions the law of large numbers holds in the strong sense. Although these results cover many stochastic volatility models including certain price processes even with non-Markovian increments (see Section 6 in [5]), the assumptions are too restrictive, and the applicability of the theorem to real-life situations is questionable.

It is the purpose of this numerical study to examine whether the law of large number remains valid if we omit some of these restrictive technical conditions. We simulated trading with fixed thresholds on artificial data, and also on S&P 500 tick data consists of more than 2.3 million records. Our investigations provide empirical evidence for that the ergodic averages of logarithmic rate of returns converge to a limit, and thus the law of large numbers may hold.

2   Trading with Fixed Thresholds

In this section, we formulate the elements of trading mechanism, and briefly outline our recent advancements in [5] regarding the stability and ergodic properties of returns if the investor follows a vanilla BLSH-type trading strategy.

We consider an asset which logarithmic price at $t \in \mathbb{N}$ is denoted by A_t, and assume that for some $\mu \in \mathbb{R}$,

$$A_t = \mu t + S_t \tag{2}$$

holds, where μt represents the drift and $S_t = S_0 + \sum_{j=1}^{t} X_j$ is a random walk that describes the fluctuations around the linear trend, moreover S_0 is an almost surely finite random variable independent of $\sigma(X_k : k \in \mathbb{N})$.

To set up a "buying low, selling high"-type strategy, we fix thresholds $\underline{\theta}, \overline{\theta} \in \mathbb{R}$, satisfying $\underline{\theta} < 0 < \overline{\theta}$, and define the sequences of buying and selling times $0 =: L_0 < T_1 < L_2 < T_2 < \ldots$ such that $T_{n+1} := \min\{k > L_n : S_k < \underline{\theta}\}$ and $L_{n+1} := \min\{k > T_{n+1} : S_k > \overline{\theta}\}$, $n \in \mathbb{N}$. By buying the asset at T_n and selling it at L_n, the logarithmic return over the n-th investment period will be

$$R_n = A_{L_n} - A_{T_n} = \mu(L_n - T_n) + S_{L_n} - S_{T_n}, \ n \geq 1. \tag{3}$$

We restrict ourselves to the case when for some $M > 0$, $(X_t)_{t \in \mathbb{N}}$ is a time-homogeneous Markov chain on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$ with state space $(-\infty, M]$, satisfying the uniform minorization condition i.e. there exists $\alpha, h > 0$ such that, for all $x, y \in (-\infty, M]$,

$$\mathbb{P}\left(X_{t+1} \in (-\infty, y] \mid X_t = x\right) \geq \alpha \frac{1}{2h} \mathrm{Leb}([-h, h] \cap (-\infty, y]), \tag{4}$$

where $\mathrm{Leb}(.)$ stands for the standard Lebesgue measure on $\mathbb{R}$. Roughly speaking, the minorization condition ensures that the whole state space itself is a small set of the chain $(X_t)_{t \in \mathbb{N}}$, on the other hand, the random movements of S have a small diffuse component. By Theorem 16.2.2 of [6], there is a unique probability law π_* on $(-\infty, M]$ such that the distribution of X_t converges to π_* in total variation as $t \to \infty$, at a geometric speed. Let us assume that π_* has zero mean, that is

$$\int_{(-\infty, M]} x \, \pi_*(\mathrm{d}x) = 0. \tag{5}$$

According to Lemma 2.4 of [5], under the minorization condition for $(X_t)_{t \in \mathbb{N}}$ and the assumption that π_* has zero mean, the random variables $(T_n, L_n)_{n \in \mathbb{N}}$ are well-defined and almost surely finite.

The following theorem plays an important role because it opens door to the statistical analysis of investment returns.

Theorem 1. *Under our standing assumptions, the law of $U_n := (L_n - T_n, R_n)$ converges to a unique limiting law Π_* on $\mathbb{N} \times \mathbb{R}$ in total variation, as $n \to \infty$. Furthermore, for any bounded and measurable function $\phi : \mathbb{N} \times \mathbb{R} \to \mathbb{R}$,*

$$\frac{\sum_{j=1}^{n} \phi(U_j)}{n} \to \int_{\mathbb{N} \times \mathbb{R}} \phi(u) \, \Pi_*(\mathrm{d}u), \ n \to \infty, \tag{6}$$

almost surely.

Proof. This follows from Theorems 2.5 of [5]. We give a brief sketch of the main ideas nonetheless. Let us consider the process $W_n := (X_{T_n}, S_{T_n}, X_{L_n}, S_{L_n}, L_n - T_n)$ instead of $(U_n)_{n \in \mathbb{N}}$. It is straightforward to verify that $(W_n)_{n > 1}$ is a time-homogeneous Markov chain on the state space $(-\infty, 0) \times (-\infty, \underline{\theta}) \times (0, M] \times$

$(\overline{\theta}, \overline{\theta} + M) \times (\mathbb{N} \backslash \{0\})$. Furthermore, by taking into account only those trajectories of S that consist of just one decreasing and one increasing segment, it can be shown that the chain $(W_n)_{n>1}$ also satisfies the uniform minorization condition. By Theorem 16.2.2 of [6] again, we have that the law of W_n converges in total variation to a unique limit distribution, as $n \to \infty$, moreover for bounded functionals of $(W_n)_{n>1}$, the strong law of large numbers holds. Since $(U_n)_{n>1}$ is the projection of $(W_n)_{n>1}$, the same is true for $(U_n)_{n>1}$ as well which completes the proof.   □

Corollary 1. *Due to the time value of money, especially in presence of a potentially risk-free asset with positive interest rate, the logarithmic rate of return is more informative, and thus has greater importance for investors, than the logarithmic return over one investment period. For the logarithmic rate of return, we have*

$$r_n = \frac{R_n}{L_n - T_n} = \mu + \frac{S_{L_n} - S_{T_n}}{L_n - T_n}. \tag{7}$$

Notice that for $n \in \mathbb{N}$, r_n is a bounded and measurable function of U_n, where U_n is as in Theorem 1, provided that the sequence $(X_t)_{t \in \mathbb{N}}$ is bounded hence by Theorem 1, the sequence $(r_n)_{n \in \mathbb{N}}$ admits an ergodic behavior.

Remark 1. Section 6 of [5] extends Theorem 1 under suitable assumptions to certain price models of the form (2) with non-Markovian summands in their martingale part i.e. $(S_t)_{t \in \mathbb{N}}$. The motivation for such generalization is that discrete time analogues of continuous-time stochastic volatility models can be treated in this way, where the latter are of the form

$$\mathrm{d}A_t = \mu \mathrm{d}t + \nu_0(\Sigma_t)\mathrm{d}B_t, \tag{8}$$

where $(B_t)_{t \geq 0}$ is a Brownian motion, ν_0 is a suitable function, moreover in the volatility term, it is imaginable that $(\Sigma_t)_{t \geq 0}$ is non-Markovian process, driven by e.g. fractional Brownian motion, see [2,4].

3   Numerical Simulations

First, we considered situations when $(X_t)_{t \in \mathbb{N}}$ is AR(1) process with model parameters $p = 0$ (uncorrelated case), $p = 0.1$, and $p = 0.9$ that is $X_{t+1} = pX_t + \sigma\varepsilon_{t+1}$, where $(\varepsilon_t)_{t \in \mathbb{N}}$ is a sequence of i.i.d. Gaussian random variables, $X_0 = 0$, $\sigma = 0.01$. In the price model (2), we set $\mu = 0.001$ and $S_0 = 0$. Notice that $(X_t)_{t \in \mathbb{N}}$ is still Markovian but it is not bounded from above, moreover it fails to satisfy the minorization condition (4). We evaluated the ergodic average of the logarithmic rate of returns over 200 investment periods (see Fig. 1), and experienced convergence.

Next, we downloaded free S&P 500 Value Index tick data from http://www.kibot.com/free_historical_data.aspx, took a period between 28-03-2017 and 23-10-2019 that consists of 2338793 records of close prices. After separating the drift and "martingale part" S, we set $\underline{\theta} = -8.36 \times 10^{-4}$ and $\overline{\theta} = 9.39 \times 10^{-4}$ that
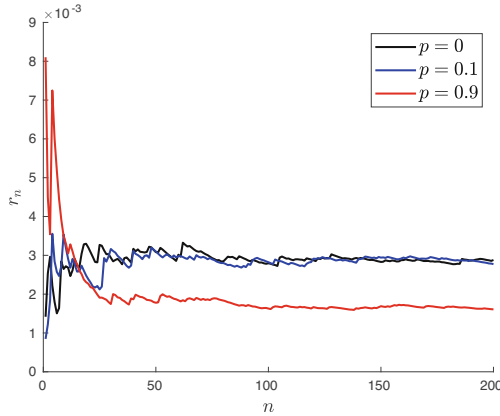
Fig. 1. Simulated trading on AR(1) data, where p denotes the model parameter. Ergodic average of the logarithmic rate of returns (r_n) over 200 investment periods.

are equal to the 45th and 55th percentiles of the empirical S-values which means roughly speaking, that the process S spends only 10% of its time between $\underline{\theta}$ and $\overline{\theta}$. Finally, we run the trading algorithm as described in Sect. 2, and obtained sequences of buying and selling times that determine 56 consecutive investment periods: $T_1 < L_1 < T_2 < \ldots < T_{56} < L_{56}$ (see Fig. 2).
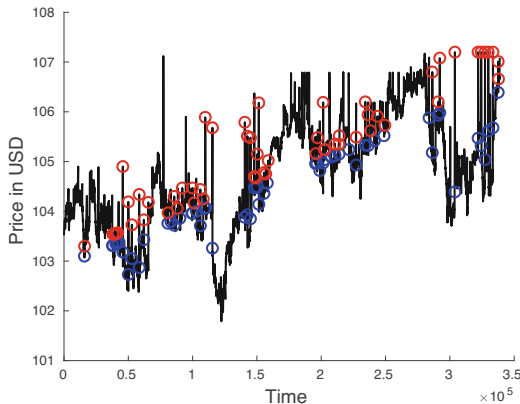
Fig. 2. S&P 500 Value Index over the considered trading period. Blue circles indicate times when it is advisable to buy i.e. $(T_n)_{n\in\mathbb{N}}$, and red circlies signalize times when it is recommended to sell i.e. $(L_n)_{n\in\mathbb{N}}$.

The logarithmic rate of return has been computed for each investment period, and also we calculated the sequence of ergodic averages $\hat{r}_n := \frac{1}{n}\sum_{k=1}^{n} r_k$, $n = 1,\ldots,56$. From Fig. 3, we can observe that the sequence $(\hat{r}_n)_{n\in\mathbb{N}}$ apparently

tends to a limit demonstrating that Theorem 1 may be applicable to "real-life" price data.
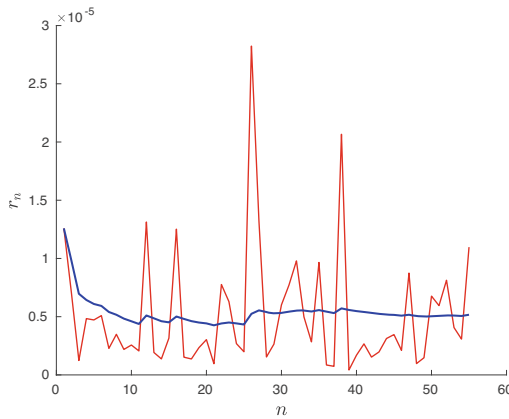
Fig. 3. Simulated trading on S&P 500 data. Logarithmic rate of returns (r_n) over the 56 investment periods – red line, and their ergodic average – blue line.

References

1. Cartea, Á., Jaimungal, S., Ricci, J.: Buy low sell high: a high frequency trading perspective. SIAM J. Financ. Math. **5**(1), 415–444 (2014)
2. Comte, F., Renault, É.: Long memory in continuous-time stochastic volatility models. Math. Finance **8**, 291–323 (1998)
3. Dai, M., Jin, H., Zhong, Y., Zhou, X.Y.: Buy low and sell high. In: Chiarella, C., Novikov, A. (eds.) Contemporary Quantitative Finance: Essays in Honour of Eckhard Platen, pp. 317–333. Springer, Heidelberg (2010). https://doi.org/10.1007/978-3-642-03479-4_16
4. Gatheral, J., Jaisson, T., Rosenbaum, M.: Volatility is rough. Quant. Finance **18**, 933–949 (2018)
5. Lovas, A., Rásonyi, M.: Ergodic aspects of trading with threshold strategies. arXiv preprint arXiv:2111.14708 (2021)
6. Meyn, S.P., Tweedie, R.L.: Markov Chains and Stochastic Stability. Springer, London (1993). https://doi.org/10.1007/978-1-4471-3267-7
7. Mijatović, A., Vysotsky, V.: Stability of overshoots of zero mean random walks. Electron. J. Probab. **25**, 1–22 (2020). https://doi.org/10.1214/20-EJP463
8. Wouter, K., Vladimir, V.: Buy low, sell high. Theor. Comput. Sci. **558**, 144–158 (2014)
9. Zervos, M., Johnson, T.C., Alazemi, F.: Buy-low and sell-high investment strategies. Math. Finance **23**(3), 560–578 (2013)
10. Zhang, H., Zhang, Q.: Trading a mean-reverting asset: buy low and sell high. Automatica **44**(6), 1511–1518 (2008)

Improving Decision Making Information: "Table 29" to an Actuarial Balance Sheet

Anne M. Garvey[1]([✉]) [iD], Juan Manuel Pérez-Salamero González[2] [iD],
Manuel Ventura-Marco[2] [iD], and Carlos Vidal-Meliá[2,3] [iD]

[1] Department of Economics and Management Sciences, University of Alcalá, Madrid, Spain
anne.garvey@uah.es
[2] Department of Financial Economics and Actuarial Science, University of Valencia, Valencia, Spain
[3] Instituto Complutense de Análisis Económico (ICAE), Complutense University of Madrid (Research Affiliate), Madrid, Spain

Abstract. This paper considers how to improve pension information, by making it more relevant for decision making purposes. Since 2017, it has been mandatory for the Member States of the European Union to report pension entitlement information, in a supplementary report referred to as Table 29. The usefulness of this unaccompanied liability information is analyzed, and a proposal is made to complement it with an actuarial balance sheet (ABS) and its corresponding income statement. More specifically, we propose a methodology to easily convert Table 29 into an ABS which estimates the assets that accompany the liabilities in the Table and which undoubtedly provides more useful information for taking decisions. To a large extent the considerations follow the requirements of the International Public Sector Accounting Standards Board (IPSASB), which explain in its conceptual framework (CF) that the aim of financial reporting by public sector entities is to provide information about any particular entity for the purposes of accountability and decision-making. It is the information needs of their users, therefore, that condition the content of financial reports.

Keywords: Accountability · Actuarial balance sheet · Income statement · Pension liabilities · Table 29 · Useful information

1 Introduction

European Union (EU) regulations require the disclosure of the accrued-to-date pension liability (ADL) by all Member States since 2017. The information is prepared using a standard actuarial cost method and some common assumptions and the information provided relates to both Social Security (SS) schemes and unfunded defined benefit (DB) schemes covering civil servants. The pension liabilities[1] are disclosed in a supplementary

---

[1] Liabilities are expressed in terms of "actuarial present value" (APV). APV is the sum of money needed now which, invested over the duration of the scheme's pension commitments, is expected to be sufficient to pay out all the pensions promised.

table, known as "Table 29". The objective of this supplementary table is to provide information on the pension obligations to households covered by any social insurance pension scheme.

Table 29 is a step forward, but it has its limitation for reporting and decision-making purposes because it does not provide information on the sustainability of SS schemes, which are typically financed on a pay-as-you-go/partially funded basis. The ADL method captures only one side of a pension scheme's balance sheet and according to previous literature an ABS (actuarial balance sheet) is the most valuable instrument for reporting the system's financial status from a SS administration's point of view (Vidal-Meliá et al. 2018; Garvey et al. 2021; TSPS 2020).

In this paper we propose the introduction of an actuarial balance sheet (ABS) and its associated income statement (IS) to improve pension information disclosure. An ABS can be defined as a financial statement that details a pension system's obligations to contributors and pensioners at a particular date, together with the amounts of the assets (financial and in particular those from contributions) that underwrite those commitments.

The proposal comes from converting the information included in "Table 29" into an ABS through a methodology designed for that purpose. This will allow the "Table 29" liabilities to be accompanied by their corresponding assets and offer information for decision making on SS pension schemes. In addition, it offers useful information as to its solvency. Barr and Diamond (2009) mention that ignoring explicit or implicit assets is misleading.

The inclusion of an Income statement provides a breakdown of the variations in assets and liabilities during the reporting period which makes pension information more understandable, transparent, relevant and useful for the purposes of taking decisions. It shows the actuarial gain or loss for each accounting period formulated which can be examined in detail, showing the breakdown of the financial result and giving important information on the origin of the variations.

We consider the disclosure of "Table 29" to be an opportunity to develop a methodology to transform the information into a more useful report for all EU Member States and offer useful and relevant information to policymakers, statisticians, public accountants, SS actuaries and public finance economists, among other users.

2   The European Union Requirement for "Table 29"

The information in the supplementary table was submitted by Member State countries of the EU for the first time in 2017 using 2015 as the reference year, and again in 2020 with data from the year 2018. Data must be transmitted compulsorily every three years for data relating to year t-2 and the liability shows the perspective of the debtor and also of the creditor.

"Table 29" pension reporting is based on the ADL method, which is consistent with the closed-group approach (CG) excluding future accruals. A plan's liabilities are equal to the present value of all expected future benefits to pensioners and all accrued rights of current affiliates derived from social contributions already paid by current workers and the remaining pension entitlements payable to existing pensioners.

The supplementary table (Table 29) records all positive and negative movements relating to the different elements making up pension obligations for all social insurance pension schemes according to the debtor and the creditor.

Table 1. Rows of the supplementary table showing accrued-to-date pension entitlements in social insurance.

Row no.	Item
1	**Pension entitlements at 31 December t (Opening Balance Sheet)**
2 (2.1 + 2.2 + 2.3 + 2.4–2.5)	**Increase in pension entitlements due to social contributions**
2.1	Employer actual social contributions
2.2	Employer imputed social contributions
2.3	Household actual social contributions
2.4	Household social contribution supplements
2.5	Less: pension scheme service charges
3	**Other (actuarial) changes in pension entitlements in social security pension schemes**
4	**Reduction in pension entitlements due to payment of pension benefits**
5 (2 + 3–4)	**Changes in pension entitlements due to social contributions and pension benefits**
6	**Transfers of pension entitlements between schemes**
7	**Change in entitlements due to negotiated changes in scheme structure**
8 + 9	**Changes in entitlements due to revaluations and other changes in volume**
8	Changes in entitlements due to revaluations
9	Changes in entitlements due to other changes in volume
10 (1 + 5 + 6 + 7 + 8 + 9)	**Pension entitlements at 31 December t + 1 (Closing Balance Sheet)**

Source: Eurostat (2013).

The information is included in Table 29 according to whether it is recorded or not in the core national accounts. This refers to whether they are funded or unfunded schemes. On a second level, the information is classified according to whether the funds are managed by general government organizations or by non-general government organizations.

At the next level, workplace pension schemes are included in **Columns A to G** and SS schemes are in **Column H**. Workplace schemes are further divided according to defined benefit (DB), defined contribution (DC) or hybrid schemes. **Column I** includes the total for all the schemes and **Columns J and K** show the breakdown of the pension liability between resident and non-resident households (a mandatory requirement from 2015).

As previously mentioned, Table 1 above shows the rows (**1 to 10**) where the variations in entitlements during the year are recorded according to different categories. **Row 1** offers an opening balance of the pension entitlements on the first day of the reporting year and **Row 10** shows the closing balance of pension entitlements on the last day of the year being reported, offering a reconciliation of the information from the start to the close of the reporting period.

Rows 2 to 9 are each related to a different item, and they include the variations in that category during the year. For example, **Row 2** includes the increase in pension entitlements due to social contributions. This row is further subclassified to include the amounts that correspond to employer actual or imputed contributions and household actual or supplementary contributions. Pension scheme service charges are then deducted before reaching the total for **Row 2.**

As another example, **Row 3** represents the difference between the current service cost and the sum of the employees' and employer's actual social contributions paid during that year. It also includes any "experience effects" observed for SS pension schemes which could include for example wage growth rate, inflation rate and discount rate where the outcome differs from the assumptions initially made.

3 Improving Pension Information for Decision Making

The supplementary table is a step forward in reporting pension information, but the usefulness of this liability information presented unaccompanied is in doubt and we question whether it complies with useful and relevant information for decision making. We contemplate the possibility of creating a methodology to transform easily, this detailed table into information that can provide useful information for decision making. The proposal is to transform Table 29 into an actuarial balance sheet (ABS) and construct its corresponding income statement. This information would certainly be a step forward to complying with the requirements of the International Public Sector Accounting Standards Board (IPSASB 2020).

According to the conceptual framework (CF) of the IPSASB (2020), General Purpose Financial Reports (GPFRs) should be prepared primarily considering the needs of primary users. Citizens are considered to be primary users and are both beneficiaries and major contributors to the pension system. The CF (IPSASB 2020) gives importance to accountability and decision-making requirements of public sector information and clearly explains that governments and other public sector entities are accountable to those who provide them with resources and to those who depend on them to use those resources to deliver services during the reporting period and longer periods of time. Information about the entity's management of the resources entrusted to it form part of accountability. The CF specifically mentions giving information on liquidity, solvency and the sustainability of the entity in question.

The CF (IPSASB 2020) outlines six qualitative characteristics of all financial and non-financial information reported in GPFRs relating to historical, prospective and explanatory information: relevance, faithful representation, understandability, timeliness, comparability and verifiability. Materiality and cost-benefit are included as two pervasive constraints that are used to achieve a suitable balance between the qualitative
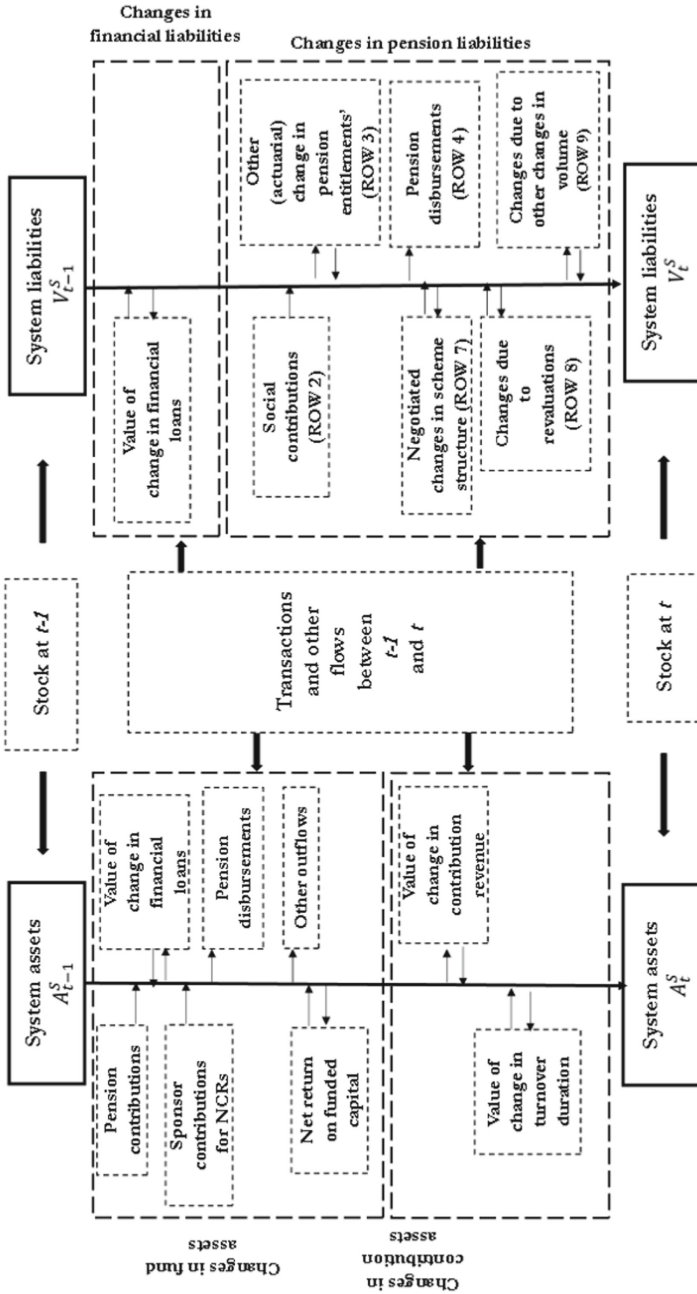
Fig. 1. The system's assets and liabilities on two consecutive valuation dates and their changes

characteristics. Relevance and faithful representation are considered fundamental, while the others are seen as enhancing qualitative characteristics. Both types of characteristics are important and when used together contribute to the usefulness of the information.

According to the CF (IPSASB 2020), information is relevant if it can make a difference in achieving the objectives of financial reporting. For information to be capable of making a difference, it needs to have confirmatory and predictive value. Confirmatory value is said to be achieved if the information confirms or changes past or present expectations. This includes confirmation of expectations that management will discharge their responsibilities for the efficient and effective use of resources and that delivery of specified service objectives and compliance with budgetary, legislative and other requirements will be achieved.

From the above we can see that there is increasing demands on public sector entities to provide useful information for decision making purposes and on the accountability of the resources entrusted to these entities.

Very briefly, our proposal consists of taking the liabilities for pensions included in Table 29 and developing a methodology of introducing mathematical formula and using available data to calculate the assets that back up the reported liabilities. This would allow us to calculate the net worth corresponding to the pension system on a specified date and by comparing that information a year later for example, the calculation of actuarial profits or losses for the period (Fig. 1). At the same time, an income statement could complement the net worth by showing a breakdown of the actuarial profit or loss for the period.

4   Conclusions

Pension information would be more relevant for decision making purposes if the pension obligations included in Table 29 were accompanied by their corresponding assets. As Member States of the EU have the mandatory obligation to prepare detailed information on pension entitlements, this information could be made more relevant by developing a methodology to transform Table 29 into more relevant and useful information for decision making purposes, including an ABS and an Income Statement. To a large extent, the above-mentioned proposal would conform with the requirements of the International Public Sector Accounting Standards Board.

References

Barr, N., Diamond, P.: Pension Reform: A Short Guide. Oxford University Press, New York (2009)

Eurostat: European system of accounts ESA 2010. Luxembourg (2013)

Garvey, A.M., Ventura-Marco, M., Vidal-Meliá, C.: Does the pension system's income statement really matter? A proposal for an NDC scheme with disability and minimum pension benefits. Econ. Res.-Ekonomska Istraživanja **34**(1), 292–310 (2021). https://doi.org/10.1080/1331677X.2020.1782246

International Public Sector Accounting Standards Board (IPSASB): IPSAS: The Conceptual Framework for Generally Purpose Financial by Public Sector Entities. In International Federation of Accountants – International Public Sector Accounting Standards Board (ed), Handbook of international public sector accounting pronouncements 2020, vol. 1. New York,

NY (2020). https://www.ifac.org/system/files/publications/files/IPSASB-HandBook-2020_Volume-1_W_0.pdf. Accessed 4 Oct 2021

TSPS. Orange Annual Report 2019. Settergren O (ed), Stockholm: Swedish Pensions Agency (Pensionsmyndigheten). Stockholm, Sweden (2020). https://www.pensionsmyndigheten.se/other-languages/english-engelska/english-engelska/publications0. Accessed 21 Oct 2021

Vidal-Melía, C., Ventura-Marco, M., Pérez-Salamero González, J.M.: Social insurance accounting for a notional defined contribution scheme combining retirement and long-term care benefits. Sustainability **10**, 2832 (2018). https://doi.org/10.3390/su10082832

Revisiting Risk Premia in Electricity Markets

Angelica Gianfreda[1]([⊠]) and Giacomo Scandolo[2]

[1] University of Modena and Reggio Emilia, Modena, Italy
angelica.gianfreda@unimore.it
[2] University of Firenze, Florence, Italy
giacomo.scandolo@unifi.it

Abstract. Motivated by the controversial empirical evidence, we deeply investigate the model proposed by Bessembinder and Lemmon (2002) and their hypotheses. We inspect the accuracy of the forward premium approximation when expressed in terms of wholesale prices, and try to unveil its dependency on price kurtosis. We derive the analytical formula of the forward risk premium in terms of the first two moments of demand and relax the assumption of normality. Finally, we study the dependence of the premium in terms of the demand skewness and kurtosis.

Keywords: Electricity · Demand · Prices · Simulations

1 Background on the Bessembinder and Lemmon's Model

The electricity system is organized in a day-ahead (or spot) and in a forward market, where P_i with $i = 1, 2, \ldots, N_P$ *Producers* produce power and sell it (either spot or forward) to retailers. *Retailers*, R_j with $j = 1, 2, \ldots, N_R$, buy power from producers and resell it to final consumers. *Consumers* cannot access the spot/forward market and can only buy electricity from Retailers at a fixed price. In this setting, P_W is the *spot* or *wholesale* or *day-ahead* price, at which one unit of power is traded on the spot market at time T. P_F is the *forward* price at which the contract is agreed upon at time 0 for delivery of (a unit of) power at the maturity T. Finally, P_R is the *retail* price, at which (unit) power is sold to consumers at time T; and it is reasonably higher than the spot price to account for taxes and levies. At time 0, each *producer sells* the quantity Q_{Pi}^F at a forward price P_F in the forward market with delivery on T. Whereas, at the maturity T, the producer generates the quantity $Q_{Pi} \geqslant 0$, which can be seen as a combination of two parts: the first one being Q_{Pi}^F, delivered to honour the forward contracts, granting a revenue given by $P_F \, Q_{Pi}^F$; and, the second one being the residual $Q_{Pi}^W = Q_{Pi} - Q_{Pi}^F$ which is instead sold on the spot market, with a revenue given by $P_W \, Q_{Pi}^W$. Note that Q_{Pi}^F or Q_{Pi}^W may be negative, which means that the producer *buys* $|Q_{Pi}^F|$ or $|Q_{Pi}^W|$ on the forward/spot market. However, the sum $Q_{Pi} = Q_{Pi}^F + Q_{Pi}^W$ cannot be negative. At time T, producer i incurs into the *total production costs* $TC(Q_{Pi})$

to generate the quantity Q_{Pi}. The cost function is assumed to be the same for all producers, and defined as $TC(Q) = C_0 + aQ^c/c$, with $a > 0$ and $c \geqslant 2$ and C_0 being fixed costs. Marginal costs are increasing and convex in production Q. The (ex-post) profit for producer i at time T is the same (concave) function of (Q_{Pi}^F, Q_{Pi}^W) for all producers. On the other hand, at time 0, each *retailer buys* the quantity Y_{Rj}^F at the forward price P_F in the forward market for delivery on T, facing at time T the *aggregated demand* D_j requested from all consumers, which is covered by the quantity bought at time 0 on forward markets, Y_{Rj}^F (with an outflow at time T equal to $P_F Y_{Rj}^F$) and the remaining part, $Y_{Rj}^W = D_j - Y_{Rj}^F$, is bought on the spot market (on T, at a cost of $P_W Y_{Rj}^W$). Note that Y_{Rj}^F or Y_{Rj}^W may be negative, but the sum (D_j) cannot be negative. In addition, at time T, the retailer sells the total purchased quantity to consumers at the retail price P_R, receiving a revenue of $P_R D_j$. Hence, at time T, the (ex-post) profit is a linear function of (Y_{Rj}^F, Y_{Rj}^W), and again is equal for all retailers. To summarize, at the initial time 0, the forward and retail prices (P_F and P_R) are known together with the quantities traded forward by producers and retailers, whereas, $P_W, Q_{Pi}^W, Q_{Pi}, \pi_{Pi}, Y_{Rj}^W, D_j,$ π_{Rj} will become known only at the maturity time T. The optimal choice is determined at the maturity time T, when spot prices and demand are known. Then, to determine the optimal choice at the initial time 0, P_W and D are unknown and considered as random variables. At time T, each Producer/Retailer has the same Mean-Variance utility function for random profits to be maximized, where $A > 0$ is the common risk aversion coefficient. The equilibrium forward price at time 0 is then determined by matching the total quantity sold by producers with the total quantity purchased by retailers. The solution is obtained and the forward premium $\mathrm{FP} = P_F - \mathbb{E}\, P_W$ is derived. Writing, here and in what follows, W instead of P_W we have[1]

$$\mathrm{FP} = \omega \cdot \mathrm{cov}(W,\ W^{b+1} - c\, P_R\, W^b) \quad \text{with} \quad \omega = \frac{AN_P}{(N_P + N_R)\, c\, a^b} > 0, \quad (1)$$

where $b = 1/(c - 1)$ (notice that $0 < b \leqslant 1$). This equation shows that FP is a function of the distribution of the wholesale prices W; and, in particular, it depends on its moments. Using a quadratic Taylor expansion around μ_W, it is possible to write the premium in terms of price moments, that is

$$\mathrm{FP} \approx \gamma_2\, \sigma_W^2 + \gamma_3\, \sigma_W^3 \xi_W \qquad (2)$$

where μ_W, σ_W^2 and ξ_W denote the mean, variance and skewness of W, respectively, with

$$\gamma_2 = \omega\, (b+1)\mu_W^{b-1}\, (\mu_W - P_R) \quad \text{and} \quad \gamma_3 = \omega\, \frac{b+1}{2}\mu_W^{b-2}\, (b\mu_W - (b-1)P_R).$$

Recalling that $\omega > 0$ and $0 < b \leqslant 1$, we see that $\gamma_3 > 0$, while $\gamma_2 < 0$ if $P_R > \mu_W$ (which is reasonable). All this leads Bessembinder and Lemmon (2002, henceforth BL) to state their first two hypotheses: (H1) FP is decreasing in σ_W; and, (H2) FP is increasing in ξ_W.

---

[1] The covariance is finite provided W has finite moment of order $b+2$, which is between 2 and 3.

FP as a Function of Demand. BL derive the equilibrium relation between demand and wholesale price as $W = a(D/N_P)^{c-1}$. Plugging this expression into (1) we obtain the following expression (not provided by BL)[2]

$$\text{FP} = \omega' \cdot \text{cov}(D^{c-1}, \ D^c - \omega'' D) \tag{3}$$

where

$$\omega' = \omega \cdot \frac{a^{b+2}}{N_P^{2c-1}} > 0 \quad \text{and} \quad \omega'' = \frac{c\, P_R\, N_P^{c-1}}{a} > 0$$

This equation allows us to investigate the dependence of FP on the distribution of D. Concerning this dependence, BL propose these further hypotheses: (H3) FP is convex in σ_D (first decreasing, then increasing); and, (H4) FP is increasing in μ_D. However, they did not formulate any hypothesis about the dependence of FP on the skewness and kurtosis of demand (incidentally, neither [5] did; henceforth SVK).

2   Simulations and Results

Testing the Accuracy of the BL Approximation (2) **for FP.** Hypotheses H1 and H2 in BL are based on the approximate expression (2) for FP. However, they did not investigate the accuracy of such an approximation. We do this investigation next, using the following parameter choices, introduced in BL and used in subsequent studies[3]: $A = 0.8/c^2$, $N_P = N_R = 20$, $a = 30(N_P/100)^{c-1}$ and $P_R = \lambda \cdot \mu_W$, where $\lambda > 1$ is a constant; BL propose $\lambda = 1.2$, but higher retail markups may be expected, so we also considered $\lambda = 1.5$. In line with previous research, we assume that W is normally distributed, with given mean and variance. Through simulation[4] (discarding negative values of W), we can compute the exact FP using Eq. 1 and compare it with the *BL approximation* in (2). Table 1 shows the results for selected combinations of μ_W, σ_W and c, with $\lambda = 1.2$. We observe that the approximation works reasonably well when σ is small compared to μ, and when c is low. However, accuracy worsens when $c \geqslant 4$ or σ increases. We obtain similar conclusions in the case $\lambda = 1.5$ (numerical results are omitted).

Testing the BL Hypotheses H1 and H2, and Inspecting for Dependence on Price Kurtosis. In order to check the validity of hypothesis H1

---

[2] The covariance is finite provided D has finite moment of order $2c - 1 \geqslant 3$.

[3] Note that we focus on forward prices, whereas recently [1] investigate risk premia from options contingent on electricity futures.

[4] The *exact* forward premium in terms of W comes in the form of a covariance between W and a transformation of W, precisely $h(W) = W^{b+1} - cP_R W^b$ (see formula (1)). Then, we have generated $M = 10^7$ iid replicates for W (according to a selected distribution) and computed the sample covariance between the two series, $(W_i)_{i=1,...,M}$ and $(h(W_i))_{i=1,...,M}$. We proceeded similarly for the forward premium in terms of the demand D (using formula (3)). We use Matlab R2021 and its built-in functions for normal and student-t distributions and implemented the algorithm presented in [2] for the skew-t distribution.

Table 1. Simulations results for the exact and approximated forward premium for a given set of parameters. μ_W and σ_W are selected according to [3] and [4].

μ_W	σ_W	c	Exact FP	BL approx FP
100	10	3	-4.17	-4.05
100	30	3	-47.35	-36.06
100	10	4	-1.29	-1.25
100	30	4	-15.94	-11.04
100	10	5	-0.56	-0.54
100	30	5	-7.32	-4.83

in BL, we consider a normal distribution for W with fixed $\mu_W = 100$ and σ_W varying in the range $(0, 30)$. Simulating W and discarding negative price values, we can compute the premium as a sample covariance, using (1) with $\lambda = 1.5$ and $c = 2, 3, 4, 5$. In all cases, the premium is decreasing in σ_W, thus confirming hypothesis H1. A similar behaviour is found for $\lambda = 1.2$ and/or other fixed values of μ_W and ranges for σ_W. For testing the hypothesis about dependence of premia on price skewness, we consider the family of (location-scale) skew-t distributions for W, as introduced in [2]. This family is specified in terms of 4 parameters that allow to adjust for mean, variance, skewness and kurtosis. In particular, by carefully selecting the parameters we have considered a subgroup of such price distributions, with fixed mean (100), standard deviation (20) and kurtosis (10) and skewness ranging in $(0.8, 1.25)$. For each such distribution, via simulation we have computed FP. Results for the case $c = 3$ and $\lambda = 1.5$ are reported in Fig. 1 (on the left). A similar behaviour can be observed for other values of c, λ and the variance. We can state that, at least in the skew-t case, BL hypothesis H2 is confirmed (i.e. FP increasing in ξ_W). Finally, and as additional contribution, for investigating the dependence of premia on price kurtosis, we consider a (location-scale) Student-t distribution for W with $\mu_W = 100$, $\sigma_W = 30$ and varying ν (degrees-of-freedom), in such a way that the kurtosis[5] ranges in $(4, 15)$. Results are presented in Fig. 1 (on the right). We can see that, at least in this case, FP seems to be *increasing* in the kurtosis.

Testing the BL Hypotheses H3 and H4, Inspecting Also for Demand Skewness and Kurtosis. We investigate the behaviour of the forward premium with respect to the demand mean, variance, skewness and kurtosis, using the exact formula that we derived in Eq. 3. Note that both BL and SVK studied the dependence of FP on μ_D and σ_D assuming a normal distribution for demand. Here, we inspect the dependence on the mean; in detail, we consider a normal distribution for D with fixed $\sigma_D = 5$ (then at 15, 25, 35) and μ_D ranging in $(50, 150)$. Via simulation, we compute the exact value of FP (with $\lambda = 1.2$); results coincide with those in Fig. 2 (bottom row) in [5] and therefore are omitted.

---

[5] Recall that the kurtosis of such a distribution is given by $\kappa = (3\nu - 6)/(\nu - 4)$, while the skewness is null, and μ and σ^2 are the actual mean and variance.

All dynamics seem to indicate that hypothesis H4 holds, at least in the normal case, even though there are some occasional departures. Results for $\lambda = 1.5$ are similar. Relaxing the normality assumption, we assume that D is distributed as a Student-t, adopting 3 different values[6] for ν; see Fig. 2 (on the left). We see that BL hypothesis H4 (FP increasing in μ_D) receives further confirmation. Moving to the dependence on variance, we consider a normal distribution for D with fixed $\mu_D = 50$ (then at 75, 100, 125, 150) and σ_D ranging in $(0, 40)$. The results corresponding to $\lambda = 1.2$ agree with those in Fig. 2 (top row) in [5], hence are omitted; results with $\lambda = 1.5$ are similar. As a robustness check, we also provide the results for D distributed as a Student-t with 3 different values for ν (Fig. 2, on the right). We can observe that the Hypothesis H3 is generally supported by these results. Moving forward to demand skewness, we then consider a (location-scale) Skew-t distribution for D and, as before, we carefully select the parameters in order to keep μ_D, σ_D and κ_D constant, respectively at level 100, 20 and 10, while varying the skewness in $(0.8, 1.25)$. Results for the case $c = 3$ and $\lambda = 1.5$
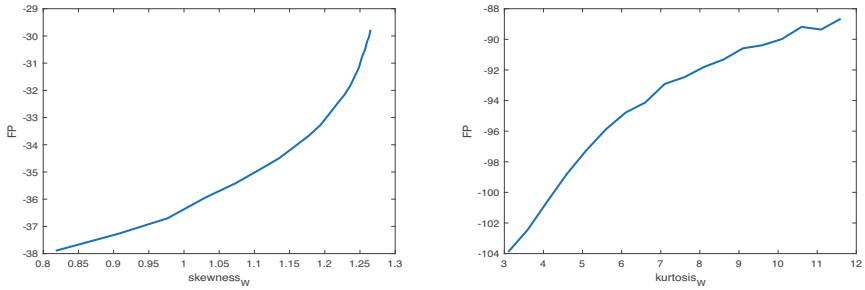
Fig. 1. On the left: FP vs ξ_W for W Skew-t with $\mu = 100$, $\sigma = 20$, $\kappa = 10$ ($c = 3$, $\lambda = 1.5$). On the right: FP vs κ_W for W Student-t with $\mu = 100$, $\sigma = 30$ ($c = 3$, $\lambda = 1.5$).
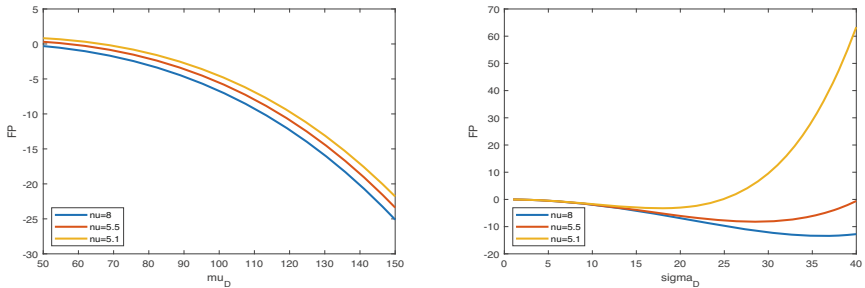
Fig. 2. On the left: FP vs μ_D for D Student-t with $\sigma = 20$ and three values for ν ($c = 3$, $\lambda = 1.5$). On the right: FP vs σ_D for D Student-t with $\sigma = 20$ and three values for ν ($c = 3$, $\lambda = 1.5$)

---

[6] Note that $c = 3$ imposes finiteness of the moment of order 5, hence $\nu > 5$.
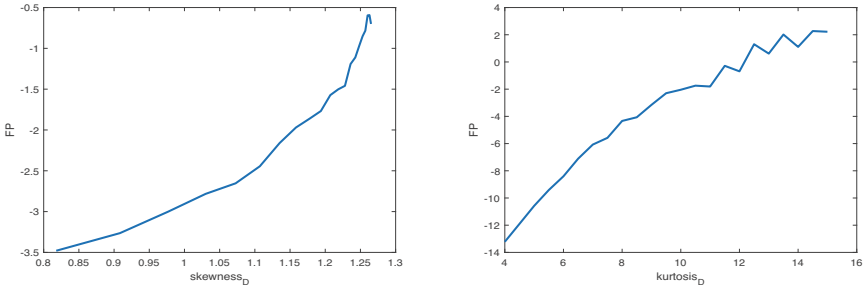
Fig. 3. On the left: FP vs ξ_D for D Skew-t with $\mu = 100$, $\sigma = 20$, $\kappa = 10$ ($c = 3$, $\lambda = 1.5$). On the right: FP vs κ_D for D Student-t with $\mu = 100$, $\sigma = 30$ ($c = 3$, $\lambda = 1.5$)

are presented in Fig. 3 (on the left). We obtain a similar behaviour with other values of σ_D, λ and/or c. We can observe that the FP is increasing w.r.t. the skewness of D, at least when using Skew-t distributions. Finally, for investigating the dependence on kurtosis, we consider a (location-scale) Student-t distribution for D with $\mu_D = 100$, $\sigma_D = 30$ and varying ν_D, in such a way that the kurtosis ranges in $(3, 12)$. Results are reported in Fig. 3 (on the right). We obtain a similar behaviour with other values of σ, λ and/or c. We can conclude that the FP is increasing w.r.t. κ_D, at least using Student-t distributions.

3   Conclusions

Given previous controversial results, we replicate the BL model and contribute firstly deriving the analytical formula of the forward risk premia with respect to the first two moments of demand. Then, we relax the assumption of normal distributed demand and simulate its dependence in terms of (demand) skewness and kurtosis. Additionally, we investigate the accuracy of the forward premium approximation when expressed in terms of wholesale prices. Finally, we try to discover its dependency on price kurtosis.

References

1. Algieri, B., Leccadito, A., Tunaru, D.: Risk premia in electricity derivatives markets. Energy Econ. **100**, 105300 (2021)
2. Azzalini, A.: The Skew-Normal and Related Families (Institute of Mathematical Statistics Monographs). Cambridge University Press, Cambridge (2013)
3. Bessembinder, H., Lemmon, M.: Equilibrium pricing and optimal hedging in electricity forward markets. J. Finance **3**, 1347–1382 (2002). LVII
4. Bunn, D.W., Gianfreda, A.: Integration and shock transmissions across European electricity forward markets. Energy Econ. **32**, 278–291 (2010)
5. Van Koten, S.: Forward premia in electricity markets: a replication study. Energy Econo. **89**, 104812 (2020)

A Semi-Markov Model for Stock Returns with Momentum and Mean-Reversion

Javier Giner[1] and Valeriy Zakamulin[2]([✉])

[1] Department of Economics, Accounting and Finance, University of La Laguna, Camino La Hornera s/n, 38071 Santa Cruz de Tenerife, Spain
`jginer@ull.edu.es`
[2] School of Business and Law, University of Agder, Service Box 422, 4604 Kristiansand, Norway
`valeriz@uia.no`

Abstract. A vast body of empirical literature documents the existence of short-term momentum and medium-term mean reversion in various financial markets. By contrast, there is still a great shortage of theoretical models that explain the presence of these two common phenomena. We develop a semi-Markov model where the return process randomly switches between bull and bear states. In our model, the state duration times are governed by a negative binomial distribution that exhibits a positive duration dependence. We demonstrate that this model induces return momentum at short lags and reversal at subsequent lags. We calibrate our model to empirical data and show that the model-implied autocorrelation function fits reasonably well to the empirically estimated autocorrelation function.

Keywords: Time-series momentum · Mean reversion · Duration dependence · Bull and bear markets · Semi-Markov model

1 Introduction

There is a vast body of empirical literature on momentum and mean reversion in various financial markets. By contrast, there is still a great shortage of theoretical models that explain the presence of short-term momentum and subsequent mean reversion. Almost exclusively, these models are equilibrium models that assume the existence of several types of traders in a financial market: rational traders, noisy traders, momentum traders, and contrarian traders. These models are elaborate and complicated theoretical models that are difficult to solve analytically. This paper is the first to entertain a fundamentally different approach to the theoretical modeling of momentum and mean reversion in financial markets. In our model, the return process randomly switches between two possible states commonly referred to as bull and bear markets.

Conventional Markov-switching models (MSM) of [3] have been traditionally used to model stock returns in bull and bear market states. In an MSM, the

return process exhibits a positive autocorrelation that decreases as the lag length increases (see [1] and [2]). Thus, an MSM can explain the short-term momentum, but is not able to explain the medium-term mean reversion.

A severe limitation of an MSM is that the state duration times are governed by a geometric distribution that is memoryless. As a result, the state termination probability does not depend on the time already spent in that state. By contrast, many empirical studies document that the longer a bull (bear) market lasts, the higher its probability of ending. Consequently, an MSM does not correctly represent the bull and bear market duration times.

We incorporate the duration dependence in a regime-switching model using an expanded-state MSM (ESMSM), where several Markovian states represent one semi-Markovian state. We use a specific topology where the state duration times are governed by a negative binomial distribution that exhibits a positive duration dependence. Our ESMSM provides some degree of analytical tractability and is easy to solve numerically.

We present the theoretical construction of our ESMSM, where the return process randomly switches between bull and bear states. For the simplest case, where two Markovian states represent each semi-Markovian state, we offer the analytical solutions to the return autocorrelation function. We demonstrate that the return autocorrelation function exhibits short-term momentum and medium-term mean reversion. Qualitatively, this autocorrelation function remains the same in the general case where many Markovian states represent each semi-Markovian state.

2   Return Autocorrelation in a Regime-Switching Model

We assume that the period-t log return X_t is a discrete-time stochastic process that randomly switches between two states (regimes): A and B. The return distribution depends on the state S_t in the following manner:

$$X_t = \begin{cases} \mu_A + \sigma_A z_t & \text{if } S_t = A, \\ \mu_B + \sigma_B z_t & \text{if } S_t = B, \end{cases} \tag{1}$$

where μ_A and σ_A are the mean and standard deviation of returns in state A, μ_B and σ_B are the mean and standard deviation of returns in state B, and z_t is an i.i.d. random variable with zero mean and unit variance. We assume that state A is a bull state of the market, while state B is a bear state of the market.

The conditional probabilities $Prob(S_{t+n} = J | S_t = I) = p_{IJ}(n)$ are called the multi-period transition probabilities. The n-period transition probability distribution of the process can be represented by a 2×2 transition probability matrix $\mathbf{P}(n)$:

$$\mathbf{P}(n) = \begin{pmatrix} p_{AA}(n) & p_{AB}(n) \\ p_{BA}(n) & p_{BB}(n) \end{pmatrix}. \tag{2}$$

Denote by $\boldsymbol{\pi} = [\pi_A, \pi_B]$ the vector of the steady-state (stationary or ergodic) probabilities. Specifically,

$$\pi_A = Prob(S_t = A), \quad \pi_B = Prob(S_t = B). \tag{3}$$

The expression for the lag-n autocorrelation can be re-written in the following form (see [1] and [2, Chapter 10]):

$$\rho_n = \frac{\pi_A \pi_B (\mu_A - \mu_B)^2 - (\mu_A - \mu_B)(\pi_A \, p_{AB}(n) \, \mu_A - \pi_B \, p_{BA}(n) \, \mu_B)}{\sigma^2}, \quad (4)$$

where

$$\sigma^2 = \pi_A \sigma_A^2 + \pi_B \sigma_B^2 + \pi_A \pi_B (\mu_A - \mu_B)^2. \quad (5)$$

It is important to note that the return autocorrelation function depends on n only through transition probabilities $p_{AB}(n)$ and $p_{BA}(n)$.

3   Return Autocorrelation in a Markov Model

The conventional MSM is defined by the following one-period transition probability matrix:

$$\mathbf{P} = \begin{pmatrix} p_{AA} & p_{AB} \\ p_{BA} & p_{BB} \end{pmatrix} = \begin{pmatrix} 1 - \alpha & \alpha \\ \beta & 1 - \beta \end{pmatrix}, \quad (6)$$

where α (β) denotes the one-period transition probability from state A to state B (state B to state A). In an MSM, the return autocorrelation function is given by (see [1])

$$\rho_n = \frac{\pi_A \pi_B (\mu_A - \mu_B)^2}{\sigma^2} (1 - \alpha - \beta)^n. \quad (7)$$

The autocorrelation exponentially decreases towards zero as n increases. Consequently, in an MSM, the return process exhibits a short-term momentum only.

4   Topology of ESMSM

In the simplest case, each macro-state A and B is represented by two sub-states. Specifically, macro-state A consists of sub-states 1 and 2, while macro-state B consists of sub-states 3 and 4. The one-period transition probability matrix is given by:

$$\mathbf{P} = \begin{bmatrix} p_{11} & p_{12} & p_{13} & p_{14} \\ p_{21} & p_{22} & p_{23} & p_{24} \\ p_{31} & p_{32} & p_{33} & p_{34} \\ p_{41} & p_{42} & p_{43} & p_{44} \end{bmatrix} = \begin{bmatrix} 1 - 2\alpha & 2\alpha & 0 & 0 \\ 0 & 1 - 2\alpha & 2\alpha & 0 \\ 0 & 0 & 1 - 2\beta & 2\beta \\ 2\beta & 0 & 0 & 1 - 2\beta \end{bmatrix}. \quad (8)$$

Each element p_{ij} of the transition probability matrix is defined in the usual manner: $p_{ij} = Prob(S_{t+1} = j | S_t = i)$. Note that the self-transition probabilities of sub-states 1 and 2 (3 and 4) are the same $p_{11} = p_{22}$ ($p_{33} = p_{44}$). As a result, the transition probabilities from one sub-state of macro-state A (B) to either another sub-state or another macro-state are the same $p_{12} = p_{23}$ ($p_{34} = p_{41}$).

In the ESMSM specified by the transition probability matrix in (8), the self-transition probability of macro-state A is computed as follows. If we know that the process is in macro-state A, then the process is equally likely to be either in

sub-state 1 or 2. If the process is in sub-state 1, then the probability of remaining in macro-state A is $p_{11} + p_{12}$. If the process is in sub-state 2, then the probability of remaining in macro-state A is $p_{21} + p_{22}$. Consequently, the probability p_{AA} is computed as $(p_{11} + p_{12})/2 + (p_{21} + p_{22})/2$. All other transition probabilities are computed in the same manner:

$$p_{AA} = (p_{11} + p_{12} + p_{21} + p_{22})/2, \quad p_{BA} = (p_{31} + p_{32} + p_{41} + p_{42})/2,$$
$$p_{AB} = (p_{13} + p_{14} + p_{23} + p_{24})/2, \quad p_{BB} = (p_{33} + p_{34} + p_{43} + p_{44})/2. \tag{9}$$

It is easy to check that both the ESMSM and corresponding MSM have the same one-period transition probabilities for states A and B. For example, $p_{AA} = 1 - \alpha$ in both the ESMSM and MSM. However, the multi-period transition probabilities are different.

The n-period transition probability matrix in the ESMSM is given by

$$\mathbf{P}(n) = \mathbf{P}^n = \begin{bmatrix} p_{11}(n) & p_{12}(n) & p_{13}(n) & p_{14}(n) \\ p_{21}(n) & p_{22}(n) & p_{23}(n) & p_{24}(n) \\ p_{31}(n) & p_{32}(n) & p_{33}(n) & p_{34}(n) \\ p_{41}(n) & p_{42}(n) & p_{43}(n) & p_{44}(n) \end{bmatrix}. \tag{10}$$

The n-period transition probabilities of macro-states A and B are computed similarly to (9). For example, the n-period self-transition probability of state A is computed as $p_{AA}(n) = (p_{11}(n) + p_{12}(n) + p_{21}(n) + p_{22}(n))/2$.

5   Analytical Solutions

This section presents a number of propositions. All proofs are available from the authors upon request.

Proposition 1. *The solutions to the n-period state transition probabilities of macro-states A and B, with two sub-states for each macro-state, are given by*

$$p_{AB}(n) = \pi_B - \frac{1}{4\beta}\psi(n), \quad p_{BA}(n) = \pi_A - \frac{1}{4\alpha}\psi(n), \tag{11}$$

where function $\psi(n)$ is given by

$$\psi(n) = \frac{(\delta + C)^2}{4C}\lambda_3^n - \frac{(\delta - C)^2}{4C}\lambda_4^n - \frac{(\alpha - \beta)^2}{\delta}(1 - 2\delta)^n, \tag{12}$$

π_A and π_B are the stationary probabilities, $\delta = \alpha + \beta$, $\lambda_3 = 1 - \delta - C$, $\lambda_4 = 1 - \delta + C$, and $C = \sqrt{\alpha^2 + \beta^2 - 6\alpha\beta}$ assuming that $C \neq 0$.

Therefore, the solution for the lag-n autocorrelation yields

$$\rho_n = \frac{(\mu_A - \mu_B)^2}{4\sigma^2(\alpha + \beta)}\psi(n). \tag{13}$$

Function $\psi(n)$ determines the functional form of the lag-n autocorrelation in the ESMSM. This function represents the sum of three exponential functions, where the first two are functions of C. Note that C can be either a real non-zero number, zero, or a complex number depending on the sign and value of $\alpha^2 + \beta^2 - 6\alpha\beta$. In particular:

$$
C \text{ is } \begin{cases} \text{a complex number} & \text{if } (3 - \sqrt{8})\beta < \alpha < (3 + \sqrt{8})\beta, \\ \text{a real number} & \text{if } \alpha < (3 - \sqrt{8})\beta \text{ or } \alpha > (3 + \sqrt{8})\beta. \end{cases} \tag{14}
$$

One can easily deduce that C is a real number when the mean duration of one state is approximately more than six times greater than the mean duration of the other state. We do not observe such a notable difference between the mean durations of bull and bear markets. Consequently, in the context of the stock market cycles, we expect that C is a complex number. In this case, λ_3 and λ_4 is a complex conjugate pair, and the analytical solution to function $\psi(n)$ is provided by the following proposition.

Proposition 2. *If C is a complex number, then function $\psi(n)$ given by Eq. (12) can be rewritten in the following form:*

$$
\psi(n) = R\lambda^n \cos(n\varphi + \theta) - \frac{(\alpha - \beta)^2}{\delta}(1 - 2\delta)^n, \tag{15}
$$

where

$$
\lambda = \sqrt{1 - 2\delta + 8\alpha\beta}, \quad \varphi = \arctan\left(\frac{\sqrt{6\alpha\beta - \alpha^2 - \beta^2}}{1 - \delta}\right), \tag{16}
$$

$$
R = \sqrt{\delta^2 + \frac{(\alpha - \beta)^4}{6\alpha\beta - \alpha^2 - \beta^2}}, \quad \theta = \arctan\left(\frac{(\alpha - \beta)^2}{\delta\sqrt{6\alpha\beta - \alpha^2 - \beta^2}}\right). \tag{17}
$$

Consequently, if C is a complex number, the expression for the n-period state transition probabilities represents the difference between two components. The first component is a damped cosine wave with a phase shift, while the second is exponential decay. Therefore, ρ_n approaches zero in an oscillating manner as n increases. To gain further insight into the behavior of the lag-n autocorrelation, let us assume that $\alpha = \beta$. In this case, the expression for ρ_n can be simplified to[1]

$$
\rho_n = \frac{(\mu_A - \mu_B)^2}{4\sigma^2}\lambda^n \cos(n\varphi). \tag{18}
$$

Under this simplified assumption, it is clear-cut that a damped cosine function without a phase shift represents the shape of the lag-n autocorrelation. In particular, ρ_n periodically changes sign beginning from a positive one.[2] Typically,

---

[1] If $\alpha = \beta$, then $R = \alpha + \beta$ and $\theta = 0$. Besides, the second term on the right-hand side of Eq. (15) disappears. Finally, $\pi_A = \pi_B = 0.5$.

[2] This function crosses zero each time when $n\varphi = k\pi$ radians, where k is a positive integer value.

because the cosine wave decays rather fast, the full oscillating behavior is hard to notice. However, one can clearly see a positive autocorrelation over the short run and a subsequent negative autocorrelation over the medium run. That is, the return process exhibits both short-term momentum and medium-term mean reversion.

6 Empirical Application

We calibrate our model to monthly returns on the Dow Jones and Standard and Poor's Composite indices. Figure 1 plots the results and demonstrates that the fit is reasonably good. In particular, our model correctly captures the duration of short-term momentum that lasts about 10–12 months and subsequently reverses.
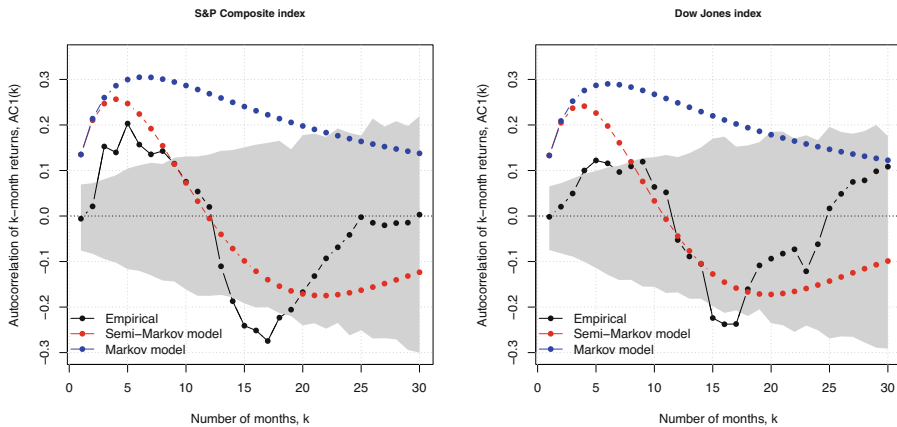
Fig. 1. The results of estimations and calibrations. The black lines with points show the empirically estimated autocorrelations. The shaded areas indicate the 90% confidence interval for the estimated autocorrelation under the null hypothesis of i.i.d. returns. The blue lines with points depict the autocorrelations implied by the fitted conventional Markov model. The red lines with points depict the autocorrelations implied by the fitted semi-Markov model, where four Markovian states represent one semi-Markovian state.

References

1. Timmermann, A.: Moments of Markov switching models. J. Econom. **96**, 75–111 (2000). https://doi.org/10.1016/S0304-4076(99)00051-2
2. Frühwirth-Schnatter, S.: Finite Mixture and Markov Switching Models. Springer, New York (2006). https://doi.org/10.1007/978-0-387-35768-3
3. Hamilton, J.D.: A new approach to the economic analysis of nonstationary time series and the business cycle. Econometrica **57**(2), 357–384 (1989)

A Variable Selection Method
for High-Dimensional Survival Data

Francesco Giordano, Sara Milito[(✉)], and Marialuisa Restaino

University of Salerno, Via Giovanni Paolo II, 132, 84084 Fisciano, Salerno, Italy
{giordano,smilito,mlrestaino}@unisa.it

Abstract. Survival data with high-dimensional predictors are regularly collected in many studies. Models with a very large number of covariates are both infeasible to fit and likely to incur low predictability due to overfitting. The selection of significant variables plays a crucial role in estimating models. Even if several approaches that identify variables in presence of censored data are available in literature, there is not unanimous consensus on which method outperforms the others. Nonetheless, it is possible to exploit the advantages of methods to get the final set of covariates as good as possible. Therefore, we propose a method that combines different variable selection procedures by using the subsampling technique, for identifying as relevant those covariates that are selected most frequently by the different variable selectors on subsampled data. By a simulation study, we evaluate the performance of the proposed procedure and compare it with other techniques.

Keywords: Variable selection · High-dimension · Survival data

1 Introduction

In recent years the classical problem of variable selection has enjoyed increased attention thanks to a massive growth of high-dimensional data available in many scientific disciplines. In modern statistical applications, the number of variables often exceeds the number of observations. In such contexts, the true model is often assumed to be sparse, meaning that only a small fraction of the variables are actually related to the response. Therefore, the selection of the relevant variables is of fundamental importance in the analysis of high-dimensional data.

Survival analysis deals with the expected time until one or more events occur. It is frequently used in the field of economics, where the event of interest is the failure of companies (mainly due to bankruptcy) or the reasons for which customers choose to stop their relationship with company. In regression analysis of survival data, the Cox Proportional Hazard model, proposed by Cox in 1982 [2], is the most used to explore the relationship between subjects' survival and some explanatory variables.

Like linear regression models, traditional variable selection methods such as subset selection, forward selection, backward elimination, and a combination

of both are among the most common applied for choosing the set of relevant variables under survival framework. However, these methods have computational difficulties in presence of high-dimensional data. Therefore, other methods have been proposed to overcome this problem. Lasso, firstly proposed for linear regression models [5], is then extended to the Cox model [6]. Subsequently, some authors have developed some penalized shrinkage techniques such as SCAD introduced by [3] specifically for Cox models. On one hand, the above methods of variable selection have been shown to be successful in theoretical properties and numerous experiments. On the other hand, their performance is highly dependent on the correct choice of the tuning parameter and these approaches can be unstable, especially in the high-dimensional data setting.

Among the problems encountered in identifying relevant variables, the choice of the best selector from those available is the most relevant. Unfortunately, the set of covariates selected by one method may be different from that selected by another. Even if it might be seen as a disadvantage, analysing the differences and similarities among the various methods can provide useful information. For example, a covariate chosen from all methods can be considered as actually relevant, while ones selected only by one method cannot be related to the response. In order to take into account this insight, following the idea of [7] for linear model, we propose a method called Combined Variable Selector with Subsample (CVSS) that combines different variable selection procedures by using the subsampling technique. We record the percentage of times a covariate is selected among the procedures and we get the final set by identifying as relevant those covariates that are selected most frequently. The main difference between our procedure and [7] consists in the choice of the tuning parameter in the various methods used. In fact, while in [7] for each method the authors take into consideration some vectors of covariates selected by different penalty coefficients, we consider only one vector of betas referring to the best tuning parameter. Thus, we extract only one set of variables for each approach with the advantage that the procedure becomes very fast.

The paper is organized as follows. In Sect. 2, we introduce our proposed approach. In Sect. 3, we show the simulation results. We conclude this work with a discussion in Sect. 4.

2   The Proposed Procedure

Suppose there are n observations $\{(y_i, \mathbf{x}_i, \delta_i)\}_{i=1}^n$ of survival data. For an individual i, y_i denotes its survival time and $\mathbf{x}_i = (x_{i1}, x_{i2}, \ldots, x_{ip})^T$ represents the observed data for the p covariates. At the same time, $\delta_i \in \{0, 1\}$ is a variable indicator of censorship, where $\delta_i = 0$ means that y_i is right-censored. We assume also that the censoring mechanism is non-informative and independent of the event process. Let $h(t)$ be the hazard rate at a time t; the generic form of the Cox proportional hazards model can be expressed as

$$h(t \mid \mathbf{x}) = h_0(t) \exp(\mathbf{x}^T \boldsymbol{\beta})$$

where $\boldsymbol{\beta} = (\beta_1, \beta_2, \ldots, \beta_p)^T$ denotes a p-dimensional vector of unknown regression coefficients and $h_0(t)$ is the baseline hazard function, that is the hazard function at time t when all the covariates take value zero. In general, $\boldsymbol{\beta}$ can be estimated by maximizing the partial likelihood function [2].

In order to identify the set of true relevant variables, it is possible to use a penalized variable selection method among those proposed in the last years. For example, the Lasso is able to select the non-zero components in setting with large p, it is computationally efficient and it uses an L_1 type penalty, while the SCAD is a regularized regression methods with non-convex penalties and it is designed to reduce estimation bias. Although in the literature there are several approaches for selecting variables in presence of censored data, there is not unanimous consensus on which method outperforms the others. Then, how to select a method remains an open question. Since choosing a method rather than another influences the selection of relevant variables, it is very important to identify the best variable selection method for the data under analysis.

In order to solve this open question, we propose to implement different variable selection methods on the sampled data and to check similarities between different variable selectors. Combining the models with subsampling is used to improve the variable selection performance of a single variable selection method. For example, RBVS proposed by [1] uses subsampling to identify the set of highly-ranked covariates, while Stability Selection proposed by [4] repeatedly samples observations and fits the sampling data using a variable selection method (e.g. the Lasso). It therefore keeps covariates with a selection frequency above a certain threshold.

Similarly to the methods above, our proposal fits variable selection methods to the subsampled data and it identifies as non-zero components those covariates appearing most frequently. Unlike these other approaches, however, our procedure uses various variable selection methods. In fact, we observe that no method outperforms all other methods in all settings, since different variable selection methods optimize different objective functions. In the case of regularized regression, the difference among methods is usually in terms of the penalty. If a covariate is selected by the majority of methods, it means that the covariate is chosen to minimize many various objective functions. We expect that a true covariate should frequently be chosen regardless of the objective function used. We repeat the fitting on subsampled data to incorporate the variability in selection due to the variability in the data.

The variable selection procedure proposed can be summarized as follows. First, we consider mutually exclusive subsets $I_{b1}, \ldots, I_{br}$ of size m, drawn uniformly from $\{1, \ldots, n\}$ without replacement, where $r = \lfloor n/m \rfloor$, $b = 1, \ldots, B$ and $B \in \mathbb{N}$ is the number of replicates. Assume that the sets of subsamples are independently drawn for each b. Second, we fit different variable selection methods on the sets $I_{b1}, \ldots, I_{br}$ and we collect the estimated model in $\mathcal{M}$, where $|\mathcal{M}| = r \times B \times k$ and k is the number of variable selector used. For each subset and for each procedure, we obtain a vector of $\hat{\beta}$. Third, we measure the relative frequency of times the jth covariate is selected given by

$$\hat{\tau}_j = \frac{1}{|\mathcal{M}|} \left(\sum_{M_i \in \mathcal{M}} I_{(\hat{\beta}_j^{M_i} \neq 0)} \right)$$

where $\hat{\beta}_j^{M_i}$ is the estimated coefficient of the jth covariate on the fitted model $M_i \in \mathcal{M}$, and I_x is the indicator function. Fourth, we identify as relevant those variable such that

$$\hat{S} = \{j : \hat{\tau}_j \geq q\}$$

where q is a fixed threshold. For the practical use, the number of replicates B should be large enough to stabilize the value of j and at the same time, it should be small enough to not increase the computational time. Following [1], we set $r = 2$ and $B = 50$, so we obtain 100 sets each with $n/2$ number of observations. In this paper, we set $q = 1/2$, which means that covariates with $\hat{\tau}_j \geq 1/2$ are selected.

The choice of the different methods to be used within our procedure is based on the following considerations. Each method must have good variable selection performance and it is required some variability among methods. In this article, we choose Lasso, MCP, SCAD, Elastic Net and Ridge since they optimize different objective functions, as they use various penalty terms. Furthermore, such methods are also computationally feasible in high-dimensional setting.

3   Simulation Study

We compare the variable selection performance among different methods by the number of false positive (FP), the number of false negative (FN), the total number of variable selection error (FN+FP) and the size of selected set. For comparison, we also consider other variable selector methods applied on the whole dataset: the Lasso, the Elastic net, the Ridge regression, the SCAD and the MCP.

In our simulation study we generate survival times $t_i, i = 1, 2, \ldots, n$, as exponential distributions with subject-specific parameters $h_i = h_0(t_i) \exp(\beta^T X_i)$, baseline $h_0(t_i) = 1$ and $\beta = (2_5, 0_{p-5})$. Thus the true size of model is $s = 5$. The variables $X_1, \ldots, X_p$ are sampled from a multivariate normal density $N(0, \Sigma)$ where the entries of Σ are fixed to $corr(X_j, X_k) = \rho^{|j-k|}$ with $\rho \in \{0, 0.3, 0.6\}$. The percentage of censorship c is setting to 20% or 40%. We set $n = 150$ and $p = \{100, 200\}$. The results are shown in Table 1.

In all scenarios our procedure has the best performance in terms of both total error FP+FN and FP. When $p = 100$ the highest value of FP for CVSS is 1.46, this means that at most 1.46 of the variables identified as relevant are not related to the response. MCP procedure is the only selector for which in Setting 3 the FN is not equal to zero: the final set contains in this case variables that are not relevant in the model. Looking at the size, our procedure selects a number of covariates that is very close to the real size 5. As we expected, the

Table 1. Simulation results for different combination of ρ, c and p. A dark grey cell represents the best results, while a grey one represents the worst. Standard errors are shown in the parentheses

Parameters	Methods	p = 100				p = 200			
		FP+FN	FP	FN	Size	FP+FN	FP	FN	Size
Setting 1	LASSO	12.91 (13.52)	12.91	0	16.92 (13.56)	13.26 (18.68)	13.26	0	17.27 (18.73)
$\rho = 0$	Elastic Net	10.22 (10.59)	10.22	0	14.32 (10.62)	11.57 (16.64)	11.57	0	15.50 (16.70)
$c = 20\%$	Ridge	10.13 (12.46)	10.13	0	14.14 (12.50)	13.28 (17.91)	13.28	0	17.29 (17.96)
	SCAD	2.54 (2.74)	2.54	0	6.55 (5.86)	3.01 (3.51)	3.01	0	7.02 (3.56)
	MCP	1.85 (1.67)	1.85	0	5.86 (1.71)	1.63 (1.40)	1.63	0	5.64 (1.45)
	CVSS	1.14 (0.38)	1.14	0	5.14 (0.38)	1.09 (0.29)	1.09	0	5.09 (0.29)
Setting 2	LASSO	11.32 (15.39)	11.32	0	15.33 (15.43)	12.89 (15.92)	12.89	0	16.90 (15.96)
$\rho = 0.3$	Elastic Net	9.9 (12.35)	9.90	0	13.91 (12.40)	9.38 (11.25)	9.38	0	13.39 (11.29)
$c = 20\%$	Ridge	12.54 (13.84)	12.54	0	16.55 (13.87)	11.26 (12.88)	11.26	0	15.27 (12.91)
	SCAD	2.16 (2.49)	2.16	0	6.17 (2.54)	2.53 (3.23)	2.53	0	6.54 (3.28)
	MCP	1.58 (1.40)	1.58	0	5.60 (1.50)	1.36 (0.80)	1.36	0	5.38 (0.87)
	CVSS	1.04 (0.20)	1.04	0	5.04 (0.20)	1.02 (0.14)	1.02	0	5.02 (0.14)
Setting 3	LASSO	9.84 (14.17)	9.84	0	13.85 (14.22)	11.61 (14.54)	11.61	0	15.62 (14.57)
$\rho = 0.6$	Elastic Net	8.17 (10.08)	8.17	0	12.18 (10.12)	12.11 (13.71)	12.11	0	16.12 (13.75)
$c = 20\%$	Ridge	12.97 (17.27)	12.97	0	16.98 (17.32)	13.49 (15.97)	13.49	0	17.50 (16.00)
	SCAD	2.57 (2.52)	2.57	0	6.58 (2.56)	2.50 (2.67)	2.50	0	6.51 (2.71)
	MCP	1.66 (1.27)	1.64	0.02	5.63 (1.32)	1.51 (1.01)	1.49	0.02	5.48 (1.07)
	CVSS	1.03 (0.17)	1.17	0	5.03 (0.17)	1.01 (0.10)	1.01	0	5.01 (0.10)
Setting 4	LASSO	13.05 (12.13)	13.05	0	17.06 (12.18)	18.22 (18.82)	18.22	0	22.23 (18.86)
$\rho = 0$	Elastic Net	13.66 (12.34)	13.66	0	17.67 (12.38)	16.19 (17.63)	16.19	0	20.20 (17.68)
$c = 40\%$	Ridge	12.67 (12.52)	12.67	0	16.68 (12.56)	14.81 (17.37)	14.81	0	18.83 (17.45)
	SCAD	2.61 (2.82)	2.61	0	6.62 (2.87)	3.34 (3.25)	3.34	0	7.35 (3.29)
	MCP	1.51 (1.27)	1.51	0	5.52 (1.32)	1.64 (1.48)	1.64	0	5.66 (1,56)
	CVSS	1.46 (0.70)	1.46	0	5.46 (0.70)	1.25 (0.48)	1.25	0	5.25 (0.48)
Setting 5	LASSO	10.76 (11.52)	10.76	0	14.77 (11.55)	13.46 (17.42)	13.46	0	17.47 (17.46))
$\rho = 0.3$	Elastic Net	11.59 (12.95)	11.59	0	15.60 (12.98)	13.80 (20.75)	13.80	0	17.81 (20.80)
$c = 40\%$	Ridge	9.36 (9.39)	9.36	0	13.37 (9.43)	13.19 (19.45)	13.19	0	17.20 (19.49)
	SCAD	2.17 (1.99)	2.17	0	6.18 (2.04)	2.38 (2.64)	2.38	0	6.40 (2.72)
	MCP	1.45 (0.99)	1.45	0	5.46 (1.03)	1.45 (0.99)	1.45	0	5.47 (1.08)
	CVSS	1.17 (0.45)	1.17	0	5.17 (0.45)	1.12 (0.38)	1.12	0	5.12 (0.38)
Setting 6	LASSO	14.02 (16.99)	14.02	0	18.03 (16.93)	15.98 (19.02)	15.98	0	19.99 (19.05
$\rho = 0.6$	Elastic Net	12.19 (13.33)	12.19	0	16.20 (13.32)	18.03 (21.05)	18.03	0	22.04 (21.09)
$c = 40\%$	Ridge	11.32 (13.43)	11.32	0	15.33 (13.51)	14.65 (16.69)	14.65	0	18.66 (16.73)
	SCAD	2.46 (2.03)	2.46	0	6.48 (2.10)	3.33 (2.86)	3.33	0	7.35 (2.92)
	MCP	1.54 (1.09)	1.54	0	5.55 (1.31)	1.68 (1.41)	1.68	0	5.69 (1.45)
	CVSS	1.07 (0.26)	1.07	0	5.07 (0.26)	1.02 (0.14)	1.02	0	5.02 (0.14)

procedures with highest FP (the Lasso, the Elastic Net and the Ridge) are also the procedures that select a higher number of covariates compared to s. In fact, as the total error increases, also the size increases. While the other approaches suffer when the correlation increases, CVSS, Lasso and Elastic Net give better results in terms of selection performances. On the other hand, the increase of censoring percentage worsens the selection for all the methods.

When $p = 200$, our procedure is still the best one. If we compare the total error for two values of p, it is possible to notice that FP+FN is lower when $p =$

200. This characteristic is not shared with the competitors. Other approaches, such as Lasso and Ridge, suffer the increase of the number of variables in the dataset. The size of CVSS is the closest to the true size $s = 5$ in all scenarios and the best performance is related at high correlation value.

4   Conclusion

In this work we proposed a new method to choose the relevant covariates with high-dimensional survival data. Although survival analysis was initially used to study death as a specific event in medical studies, these statistical techniques have increasingly been used in economics and social sciences. Given the relevance of the topic, it is important to be able to find a method that selects the relevant variables related to the response variable as good as possible. In particular, we proposed to combine several variable selectors available in literature with the subsample technique. Simulation study has shown that our approach works better than its competitors. For future work we will evaluate this approach from a theoretical point of view and apply it to real data.

References

1. Baranowski, R., Chen, Y., Fryzlewicz, P.: Ranking-based variable selection for high-dimensional data. Stat. Sin. **30**(3), 1485–1516 (2020)
2. Cox, D.R.: Regression models and life-tables. J. Roy. Stat. Soc.: Ser. B (Methodol.) **34**(2), 187–202 (1972)
3. Fan, J., Li, R.: Variable selection for Cox's proportional hazards model and frailty model. Ann. Stat. **30**(1), 74–99 (2002)
4. Meinshausen, N., Bühlmann, P.: Stability selection. J. Roy. Stat. Soc. Ser. B (Stat. Methodol.) **72**(4), 417–473 (2010)
5. Tibshirani, R.: Regression shrinkage and selection via the lasso. J. Roy. Stat. Soc.: Ser. B (Methodol.) **58**(1), 267–288 (1996)
6. Tibshirani, R.: The lasso method for variable selection in the Cox model. Stat. Med. **16**(4), 385–395 (1997)
7. Yuen, C., Fryzlewicz, P.: Exploiting disagreement between high-dimensional variable selectors for uncertainty visualization. J. Comput. Graph. Stat. 1–9 (2021). https://doi.org/10.1080/10618600.2021.2000421

Ranking-Based Variable Selection for the Default Risk of Bank Loan Holders

Francesco Giordano, Marcella Niglio$^{(\boxtimes)}$, and Marialuisa Restaino

Dipartimento di Scienze Economiche e Statistiche, Università degli Studi di Salerno,
84048 Fisciano, SA, Italy
{giordano,mniglio,mlrestaino}@unisa.it

Abstract. In this paper we extend the Ranking Based Variable Selection technique (Baranowsky *et al.*, Statistica Sinica 30, 1485–1516 (2020)) to the framework of general linear regression models. After the presentation of the main steps of the algorithm, it is applied to select the variables affecting the repayment ability of bank loan holders. We give evidence that, unlike some largely applied selection methods, the algorithm is robust to the presence of high correlated variables and the number of features selected does not change even when the dataset is contaminated with irrelevant artificial covariates.

Keywords: Screening · Variable selection · High dimension · Ranking

1 Introduction

In the last years, in many fields of sciences such as medicine, finance, social sciences, the availability of datasets of large dimensions is increased and is often characterized by a huge number of variables (p) that does not correspond to an adequate number of observations (n).

In practice, it occurs that the number of variables significantly exceeds the number of observations and then the identification of the relevant variables is a critical and essential challenge because among thousand of variables, only a subset of them could affect the phenomena under analysis.

The growing availability of data and the high computational power of processors have increased the attention on the variable selection problem that has been differently faced.

An approach of variable selection, largely examined in the literature, is based on penalized estimation procedures (among the others see LASSO [8], SCAD [4], MCP [10]) that, unfortunately, in presence of high dimensional datasets are computationally very heavy. It has lead to face the problem of variable selection, considering two steps: the first is devoted to the data reduction (*screening step*) and, more precisely, to the reduction of the covariates such that they become less than n; the second step (*selection step*) performs the true variable selection

on the reduced dataset where, for example, penalized estimation procedures can be applied without computational difficulties.

The first step, devoted to the variable reduction, has been largely faced in [5] where the Sure Independence Screening (SIS) is introduced. It ranks the p covariates using their marginal utility and allows to identify (using proper metrics) the top variables that are then considered in the next selection step. The original idea behind the SIS has been largely extended, but even further approaches have been proposed to screen variables. Among them, [6] extend the SIS to generalized linear models, [9] identifies the relevant covariates of linear regression models using the forward regression or, more recently, [2] screen the variables by maximum likelihood estimators.

Within the literature of high-dimensional regression problem we focus on the variable selection for generalized linear models. In particular, we here extend the Ranking-Based Variable Selection (RBVS) procedure of [1], originally proposed for linear regression models, to the generalized linear models (glm). In Sect. 2 we briefly introduce the setup of the regression model under analysis, in Sect. 3 we clarify how the RBVS procedure can be extended to the model setting under consideration. In Sect. 4 we apply our proposal to a large dataset where the selected variables can be used to evaluate the default risk of bank loan holders.

2   The Model

The analysis of financial datasets are often characterized by binary response variables related to the status of the units under analysis (banks, borrowers, credit card holders, credit card transactions, etc.). In most cases this dichotomy is obtained by the creditworthiness of the units (solvent $= 1$, no solvent $= 0$) or the type of transaction (fraudulent $= 1$, no fraudulent $= 0$) which can be related to a large set of covariates that have to be selected from huge datasets.

As largely known, given the response variable $Y \sim Ber(\pi)$, with $\pi \in (0, 1)$, and the matrix $(n \times p)$ of covariates $\mathbf{X}$, the *logit* model is:

$$g(\pi_i) = \mathbf{x}_i'\boldsymbol{\beta}, \qquad i = 1, \ldots n \tag{1}$$

where $\mathbf{x}_i$ is the i-*th* row of $\mathbf{X}$, $g(\pi_i) = \log\left(\frac{\pi_i}{1-\pi_i}\right)$ is the link function and $\boldsymbol{\beta}$ is the $(p \times 1)$ vector of coefficients.

One feature that often characterizes the response variable Y in financial datasets, is the small number of ones that makes inappropriate the use of the symmetric *logit* link function (see among the others [3]). In this context with unbalanced number of zeros and ones, we model Y by the Gumbel distribution such that the link function (called *complementary log-log function*) becomes $g(\pi_i) = \log[-\log(1 - \pi_i)]$, $i = 1, \ldots, n$.

After the selection of the model, to estimate the probabilities π_i, $i = 1, \ldots, n$ we need to select the variables that are evaluated as "important". The availability of high dimensional datasets, that are very common in financial domain, requires the use of proper variable selection procedures. We here extend, to the

generalized linear models domain, the RBVS procedure of [1], whose details are shortly given in the following.

3   RBVS Algorithm

A variable selection procedure that includes the two steps described in Sect. 1 is introduced in [1]. The procedure aims to rank the variables of the dataset to define the top-ranked variables (*screening step*) that are then evaluated to select the relevant covariates for Y (*variable selection step*).

The algorithm of [1] is based on the following idea: given the set of p covariates included in $\mathbf{X}$, the variables with higher influence on Y are those that even in presence of randomly selected subsamples, extracted from $\mathbf{X}$, exhibit consistent relationship with Y.

We here shortly describe the RBVS algorithm, whereas all technical details are in [1].

Let $Z_i = \{Y_i, X_{i1}, X_{i2}, \ldots, X_{ip}\}$, for $i = 1, 2, \ldots, n$ and with p that grows with n, be the observed dataset used to select the subset of covariates $\{X_1, \ldots, X_p\}$ relevant for Y. Further, let $\mathcal{A} \subset (1, \ldots, p)$ be the indices that identify a subset of covariates and let $|\mathcal{A}| = k$ be the cardinality of $\mathcal{A}$, for $k = 0, 1, \ldots, p$. Let $R_{nj}(Z_1, \ldots, Z_n)$ be the ranking of the jth covariate, based on a measure $\hat{\omega}_j = \hat{\omega}_j(Z_1, \ldots, Z_n)$ assessing the importance of each covariate where, following [6], $\omega_i = |\hat{\beta}_j|$, with $\hat{\beta}_j$ the estimated coefficient of the marginal logit model (1), $g(\pi) = \beta_0 + \beta_j$ for $j = 1, \ldots, p$. Then consider the probability:

$$\pi_n(\mathcal{A}) = \mathbf{P}(\{R_{n1}(Z_1, \ldots, Z_n), \ldots, R_{n|\mathcal{A}|}(Z_1, \ldots, Z_n)\} = \mathcal{A})$$

with $\pi_n(\mathcal{A}) = 1$, if $\mathcal{A} = \emptyset$.

Correspondingly define:

$$\pi_{n,m}(\mathcal{A}) = \mathbf{P}(\{R_{n1}(Z_1, \ldots, Z_m), \ldots, R_{n|\mathcal{A}|}(Z_1, \ldots, Z_m)\} = \mathcal{A}),$$

the probability of $\mathcal{A}$ obtained from a subset of m observations, with $1 \leq m \leq n$.

It follows that if $\mathcal{A}$ corresponds to the indices of the top-ranked covariates, $\pi_{n,m}(\mathcal{A})$ is its probability computed on a randomly selected subset of m observations.

To estimate $\pi_{n,m}(\mathcal{A})$, [1] use a bootstrap approach that for each $b = 1, \ldots, B$ (with B the number of bootstrap replicates) and given $r = \lfloor n/m \rfloor$, extract from Z_i, for $i = 1, \ldots, n$, r independent subsets without replacement $(I_{b_1}, \ldots, I_{b_r})$ and for each bootstrap replicate compute the empirical relative frequency of $\mathcal{A}$, given by $r^{-1} \sum_{j=1}^{r} \mathbf{1}(\mathcal{A}|I_{b_j})$, with $\mathbf{1}(\cdot)$ an indicator function. Then $\pi_{n,m}(\mathcal{A})$ is estimated by:

$$\hat{\pi}_{n,m}(\mathcal{A}) = B^{-1} \sum_{b=1}^{B} r^{-1} \sum_{j=1}^{r} \mathbf{1}(\mathcal{A}|I_{b_j}), \tag{2}$$

where $\mathcal{A}|I_{b_j} = \{R_{n1}(Z_i)_{i \in I_{b_j}}, \ldots, R_{n|\mathcal{A}|}(Z_i)_{i \in I_{b_j}}\}$.

The probability (2) allows to define the top-ranked variables (first step: *screening step*) given by:

$$\hat{\mathcal{A}}_{k,m} = \underset{\mathcal{A} \in \Omega_k}{\arg\max}\, \hat{\pi}_{m,n}(\mathcal{A}),$$

with Ω_k the set of all permutations of $\{1, \ldots, k\}$.

Starting from $\hat{\mathcal{A}}_{k,m}$, we need to detect the "important" variables for Y (second step: *selection step*). It is common at this stage of the variable selection algorithm, to introduce a threshold ζ that allows to select the subset:

$$\hat{s} = \min\left\{ k : \hat{\pi}_{n,m}(\hat{\mathcal{A}}_{k+1,m}) \leq \zeta \right\}, \qquad \zeta > 0.$$

Given the difficulty to choose the value of ζ, the alternative introduced in [1] is the estimation of the ratio $\hat{\pi}_{n,m}^{\tau}(\hat{\mathcal{A}}_{k+1,m})/\hat{\pi}_{n,m}(\hat{\mathcal{A}}_{k,m})$ with $\tau \in (0,1]$ such that the relevant covariates are the s top-ranked variables where:

$$\hat{s} = \underset{k=0,\ldots,k_{\max}-1}{\arg\min}\, \frac{\hat{\pi}_{n,m}^{\tau}(\hat{\mathcal{A}}_{k+1,m})}{\hat{\pi}_{n,m}(\hat{\mathcal{A}}_{k,m})}. \tag{3}$$

In practice, given the estimated probabilities of $\hat{\pi}_{n,m}(\hat{\mathcal{A}}_{k,m})$, for $k = 0, \ldots, k_{\max} - 1$, with $k_{\max}$ a fixed large integer, the number of relevant variables is related to the evaluation of the magnitude of the estimated probability and $\hat{s}$ corresponds to the case where the ratio in (3) has the greatest decrease, whereas $\hat{\mathcal{S}} = \hat{\mathcal{A}}_{\hat{s},m}$ is the subset of $\{0, 1, \ldots, p\}$ that contains the indices of the relevant variables.

4   Default Risk of Bank Loan Holders

The RBVS algorithm has been applied to predict the repayment ability of bank loans holders. It is a very critical element for all banks and then providing a procedure able to select the most important variables that affect the repayment ability of their clients, could be a substantial concern.

We have considered a huge dataset downloaded from Kaggle (named "Bank loan status dataset") characterized by millions of operations (that involve bank accounts) made by a sample of bank loan holders. Then the first main step has been the merging of the dataset containing the general information of the bank clients with other nine datasets where the operations made through their bank account, credit card etc. are stored. The merging key has been the ID of the loan, that has lead to ascribe, to each bank loan holder, all collected data.

After the merging of the ten datasets, we have extracted the numeric variables that are of interest in our case and that make the data matrix consistent with the RBVS algorithm, so removing all qualitative variables, all variables with missing values and other variables without any contribution for the present analysis. The dataset considered is finally characterized by 29 variables classified in four main groups, as presented in Table 1.

Table 1. List of the main variables in the dataset.

Class	Description
Loan	Amount loan; loan annuity; price goods for which the loan is given
Personal information	Income; seniority of employment; number population region where client lives; rating of the region; rating of the city; client's age; client's age at the time of application; family members number; days (before application) the client change the identity document; days (before application) client change registration
Credit position	Number enquiries to Credit Bureau (CB) in one year; number of days past due on CB for related loans; times the CB credit prolonged; days the client applies for Credit Bureau credit; number of times the CB credit prolonged; current CB credit; Current amount overdue on CB credit; days from the last information about the CB credit; days past due, during the month of previous credit
Previous credits (*)	AMT: Balance; credit card limit; amount drawing; how much did the client pay; amount receivable; number of drawings
	Other credits: Days past due, Days past due with tolerance (debts with low loan amounts are ignored)

(*) These variables are only related to previous credits.

Further, the binary response variable Y assumes value one if the client has payment difficulties and these difficulties are related to the fact that he/she has late payments.

It is interesting to note that some covariates under analysis are strongly correlated (as it is often the case in financial domain) and 9 of them have correlation, in absolute value, greater than 0.90. This collinearity could affect the variable selection stage but, as it will be shown in the following, the RBVS algorithm is robust to the presence of linear relation among variables.

To perform the RBVS algorithm (described in Sect. 3) to select variables in the glm domain, we have extracted from the dataset a sample of $n = 640$ units, ensuring that the proportion of ones of the response variable in the subset corresponds to that of the whole dataset. This proportion is equal to 0.0875 in the dataset and leads to select the Gumbel distribution for the response variable and consequently a complementary log-log as link function.

After the standardization of the data, we have run the RBVS algorithm fixing $m = n/2$, $r = 100$, $\tau = 0.5$ in (3) (as suggested in [1]) whereas $\hat{\omega}_j = |\hat{\beta}_j|$, for $j = 1, \ldots, p$, is chosen as measure to assess the importance of each covariate (as described in the algorithm in Sect. 3).

In this context, where the "true important" variables are unknown, the evaluation of the procedure has been performed introducing new variables in the dataset. In particular we have added, to the 29 observed variables, 100 variables artificially generated from a Uniform distribution with mean zero and unit variance. The RBVS algorithm has been run again on this augmented dataset to evaluate the stability of the selected variables.

The performance of the RBVS procedure has been finally compared with the outcomes of the glm-LASSO [7], both in the presence of the original and the augmented datasets.

Table 2. RBVS selection on the original and augmented datasets. In brackets the artificial variables selected in the augmented dataset.

Procedure	Original dataset	Augmented dataset (Artificial selected variables)
RBVS	4	4 (0)
glm-LASSO	14	13 (10)

The results of the two variable selection procedures are summarized in Table 2 where it can be noted that in the RBVS case the number of selected variables does not change when the dataset is artificially augmented and in both cases the variables are given by the rating of the region, the rating of the city (that are strictly related to the economic position of the loan holder), the days past due (during the month) on the previous credit and the days the client applied for Credit Bureau credit. In contrast, in the glm-LASSO case, the variable selection is definitely less parsimonious and even less stable because, with the augmented dataset, it selects 10 artificial variables whereas, among the remaining selected variables, two are common with those selected on the original dataset.

These results give evidence of the main findings of the proposed procedure: it is robust to the collinearity in the data, it combines the screening and variable selection steps and it is not based on any form of penalization. Even if this first example of application of the RBVS algorithm in financial domain considers a dataset with $p < n$, the algorithm has been evaluated in high-dimension in a Monte Carlo study (not included in the paper) where its performance has been assessed and compared to competing approaches.

References

1. Baranowski, R., Chen, Y., Fryzlewicz, P.: Ranking-based variable selection for high-dimensional data. Stat. Sin. **30**, 1485–1516 (2020)
2. Chen, X., Chen, J.: The sparse MLE for ultrahigh-dimensional feature screening. J. Am. Stat. Assoc. **109**, 1257–1269 (2021)
3. Dobson, A.J., Bannet, A.G.: An Introduction to Generalized Linear Models, 3rd edn. CRC Press, Boca Raton (2008)
4. Fan, J., Li, R.: Variable selection via nonconcave penalized likelihood and its oracle properties. J. Am. Stat. Assoc. **96**, 1348–1360 (2001)
5. Fan, J., Lv, J.: Sure independence screening for ultrahigh dimensional feature space (with discussion). J. Roy. Stat. Soc. B **70**, 849–911 (2008)
6. Fan, J., Song, R.: Sure independence screening in generalized linear models with NP-dimensionality. Ann. Stat. **38**, 3567–3604 (2010)
7. Park, M.Y., Hastie, T.: L1-regularization path algorithm for generalized linear models. J. Roy. Stat. Soc. B **69**, 659–677 (2007)
8. Tibshirani, R.: Regression shrinkage and selection via the lasso. J. Roy. Stat. Soc. B **58**, 267–288 (1996)
9. Wang, H.: Forward regression for ultra-high dimensional variable screening. J. Am. Stat. Assoc. **104**, 1512–1524 (2009)
10. Zhang, C.H.: Nearly unbiased variable selection under minimax concave penalty. Ann. Stat. **38**, 894–942 (2010)

Exploring Non Linear Structures in Range-Based Volatility Time Series

Michele La Rocca and Cira Perna[✉]

University of Salerno, Fisciano, Italy
{larocca,perna}@unisa.it

Abstract. In this paper we focus on the use of Extreme Learning Machines (ELMs) to appropriately capture the nonlinear dynamics of the range based estimators. The results on all the assets in the S&P500 index show that ELMs produce residuals without neglected nonlinearities

Keywords: Extreme learning machines · Nonlinear dynamics · Range based estimators

1 Introduction

Volatility plays a key role in many areas of financial econometrics including derivative pricing, asset allocation, investment decisions and risk analysis. The most popular parametric approach to model the time varying conditional volatility is the family of GARCH-type models. This class of models are generally return-based models and use only the data available at closing prices. Therefore, they are not able to catch the intraday price movements and related information. As a consequence, volatility measures based on the range are becoming popular alternative measures, since they use functions of the high and low prices of an asset within a given time interval [1].

The aim of this paper is to explore the use of Extreme Learning Machines (ELMs) as a tool to appropriately capture the nonlinear dynamics of the range based estimators. The paper is organized as follows. In Sect. 2 the price range estimators are briefly reviewed and discussed. In Sect. 3, the use of Extreme learning machines are introduced, highlighting their advantage with respect the traditional models generally employed in this context. In Sect. 4 an application to real data is presented and discussed along with some remarks.

2 The Range-Based Volatility Measures

For an asset, define the following variables: O_t = the opening price of the tth trading day; C_t = the closing price of the tth trading day; H_t = the highest price of the tth trading day; L_t = the lowest price of the tth trading day. The classical range estimator was introduced by [6]; it is defined as:

$$\widehat{\sigma}_P^2 = \left(\frac{1}{4\log 2}\right)(\log(H_t) - \log(L_t))^2. \tag{1}$$

Another estimator, which uses extra information such as the opening and the closing prices, has been proposed by [2]; it is defined as:

$$\widehat{\sigma}_{GK}^2 = 0.51\left[\log(H_t/L_t)\right]^2 - 0.019\left\{\log(C_t/O_t)\left[\log(H_t) + \log(L_t) - 2\log(O_t)\right]\right.$$
$$\left. + -2\left[\log(H_T/O_t)\log(L_T/O_t)\right]\right\} - 0.383\left[\log(C_t/O_t)\right]^2 \tag{2}$$

This estimator can be expressed as:

$$\widehat{\sigma}_{GK}^2 = 0.5\left[\log(H_t/L_t)\right]^2\left[2\log(2) - 1\right]\left[\log(C_t/O_t)\right]^2 \tag{3}$$

Both the estimators implicitly assume that the asset log-price follows a geometric Brownian motion without a drift. In term of efficiency with respect to the close-close estimator, the theoretical relative efficiency gain of the Parkinson estimator ranges from 2.5 to 5, which means that the estimation variance is 2.5–5 times lower. The Garman and Klass estimator has an efficiency of 7.4.

3   ELM for High-Low Range Volatility Models

Let R_t, $T = 1, \ldots, T$ be one of the definitions of the range-based volatility time series as reviewed in the previous section. It can be modeled as an autoregressive process:

$$R_t = m\left(R_{t-1}, \ldots, R_{t-p},\right) + \varepsilon_t \tag{4}$$

where $m(\cdot)$ is an unknown (possibly nonlinear) function, ε_t are i.i.d. innovations with mean zero and finite variance.

The function m can be approximated by using neural networks with a single output and additive nodes (NNs) in the class:

$$\mathcal{F} = \left\{f(\mathbf{z}, \boldsymbol{\eta}) = \sum_{k=1}^r \beta_k \psi\left(\mathbf{a}_k'\mathbf{z} + b_k\right) : \mathbf{z} \in \mathbb{R}^{p+d}, \boldsymbol{\eta} \in \mathbb{R}^{r(p+d+2)}\right\} \tag{5}$$

where r is the hidden layer size, $\psi(\cdot)$ is a sigmoidal activation function. The vector $\boldsymbol{\eta}$ is defined as $\boldsymbol{\eta} = (\beta_1, \ldots, \beta_r, \mathbf{a}_1', \mathbf{a}_2', \ldots, \mathbf{a}_r', b_1, \ldots, b_r)'$ where $\{\mathbf{a}_k\}$ are the $(p + d)$ dimensional vectors of weights for the connections between input layer and hidden layer; $\{\beta_k\}$ are the weights of the link between the hidden layer and the output; $\{b_k\}$ are the bias terms of the hidden neurons. However, despite their proven theoretical capabilities of non-parametric data driven universal approximation, NNs face challenging issues, concerning the specification of the network topology in accordance with the underlying structure of the series and the estimation of the parameters, which results in an heavy computational task. This latter issue makes difficult the use of neural networks on large financial temporal databases. Recently, ELM has attracted the attention in the literature (see [3] for a review) as a fast and effective alternative method to specify and estimate

neural network models for time series data. The essence of ELMs is that, unlike the other traditional learning algorithms, such as back propagation, the hidden nodes weights are randomly generated and they need not to be tuned, so that the algorithm analytically determines the output weights of NNs.

Basically, ELM trains a NN in two main stages. In the first stage, once fixed the activation function, the hidden node parameters $(\mathbf{a}, b)$ are randomly generated according to any continuous probability distribution so that the matrix:

$$\mathbf{H} = \begin{bmatrix} \psi(\mathbf{a}_1'\mathbf{z}_1 + b_1) & \cdots & \psi(\mathbf{a}_r'\mathbf{z}_1 + b_r) \\ \vdots & \ddots & \vdots \\ \psi(\mathbf{a}_1'\mathbf{z}_T + b_1) & \cdots & \psi(\mathbf{a}_r'\mathbf{z}_T + b_r) \end{bmatrix} \tag{6}$$

is completely known. In the second stage, the output weights $\boldsymbol{\beta}$ are estimated by solving the following minimization problem:

$$\hat{\boldsymbol{\beta}} = \mathrm{argmin}_{\boldsymbol{\beta}} \|\mathbf{H}\boldsymbol{\beta} - \mathbf{y}\| \tag{7}$$

where $\mathbf{y}$ is the training data target vector and $\|\cdot\|$ denotes the L_2-norm.

If $\mathbf{H}^\dagger$ denotes the Moore-Penrose generalized inverse of matrix $\mathbf{H}$, the optimal solution to the previous optimization problem is:

$$\hat{\boldsymbol{\beta}} = \mathbf{H}^\dagger \mathbf{y} \tag{8}$$

The matrix $\mathbf{H}^\dagger$ can be calculated by using one of the numerous methods proposed in the literature which include orthogonal projection, orthogonalization method, iterative method and the single value decomposition, the last one being the most general. The estimation of the parameter vector $\boldsymbol{\beta}$ can also be obtained via regularized ELM [3].

The ELM approach have several advantages (see [5] for a discussion). Firstly, it has good generalization performance in the sense that it reaches the small training error and, contemporaneously, the smallest norm of output weights. Secondly, learning can be done without iteratively tuning the hidden nodes which can be independent of training data. This has a big advantage in terms of computational burden for ELMs that can be thousand of time faster with respect to traditional NNs trained by using back-propagation or NLS. Moreover, ELMs, preserve the property of being universal approximators for a large class of non-linear functions. The results in [4] state the universal approximation capability of ELMs without imposing any restrictive assumption on the activation function as in the case of NN paradigm in which, on the contrary, a continuous and differentiable activation function is needed. In practice, being the hidden layer randomly generated, ELMs usually require more hidden neurons than NNs to obtain a given performance. However, this does not seem to be a serious problem due to the computational efficiency of ELMs. Moreover, ELMs are well suited for large data processing and, even if a model selection process is implemented for an optimal structure searching, the running time of ELMs is always lower than other competing strategies. In any case, parallel and cloud computing techniques can also be used for even faster implementation of ELMs.

4   Application to Real Data and Concluding Remarks

In this application we will consider all the assets in the S&P500 index. The time series have been downloaded from Yahoo finance and they cover the period from 30/11/2017 to 30/11/2019. The analysis has been conducted firstly considering the S&P500 index and then all its constituencies. In Fig. 1 the time plot of the Parkinson and Garman Klass definition of the High-Low range for the S&P500 index, while in Table 1 are reported the results of White's linearity and Teräsvirta's linearity test for the original series, the residuals from a linear model and the residuals from a nonparametric NAR model estimated via ELM. All tests p-values have been adjusted for multiple testing by controlling the FDR using the Benjamini-Hochberg procedure. Clearly, the linear specification does not seem able to model the nonlinear dynamic futures of the series.

In Fig. 2, 3 and 4 are reported the distributions of the White's linearity test p-values for, respectively, the original time series, the residuals from linear autoregressive models and the residuals from nonlinear autoregressive models estimated via ELMs. Again, the linear model specification is not able to model the nonlinear dynamics of the High-Low range time series correctly, while the ELMs produce residuals without neglected nonlinearities.
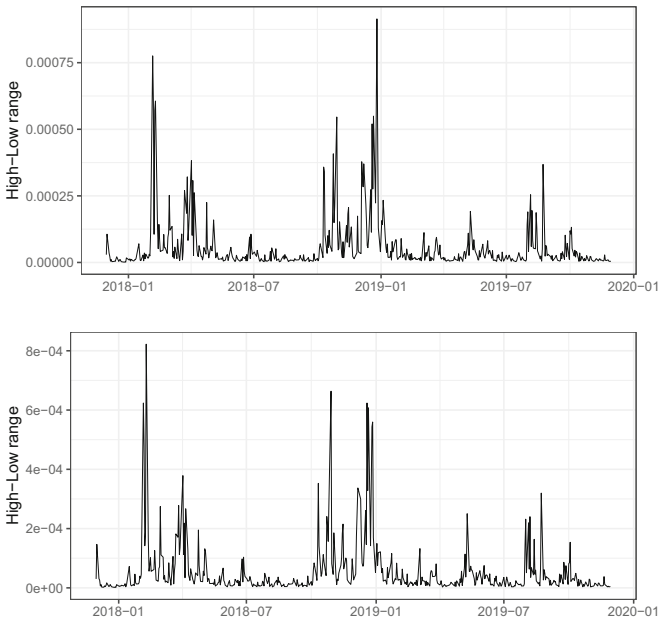
Fig. 1. Time plot of the S&P 500 high-low range with Parkinson (upper plot) and Garman and Klass (lower plot) definition
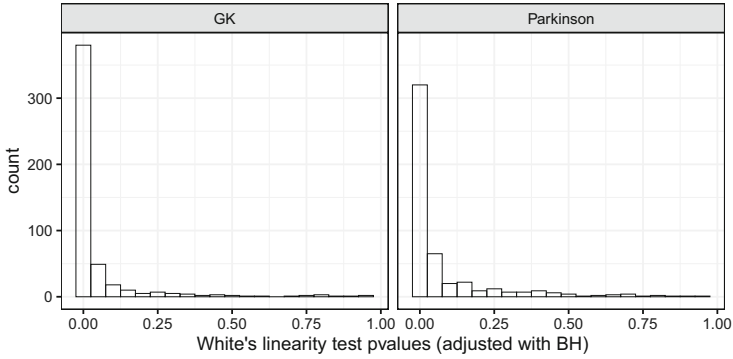
Fig. 2. Distribution of White's linearity test p-values for all the constituencies of the S&P500, adjusted for multiple testing by controlling the FDR using the Benjamini-Hochberg procedure
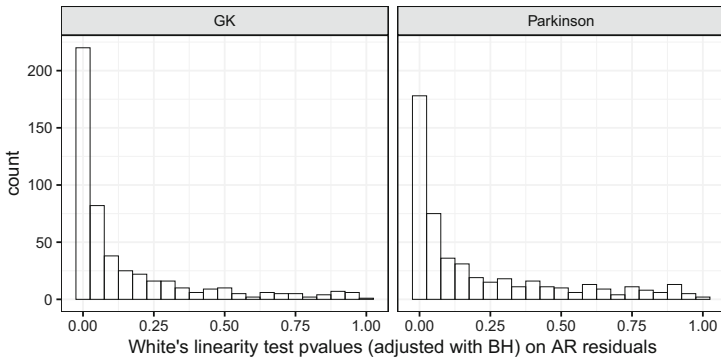
Fig. 3. Distribution of White's linearity test p-values for the residuals from AR models for all the constituencies of the S&P500, adjusted for multiple testing by controlling the FDR using the Benjamini-Hochberg procedure

Table 1. Test statistics and p-values for the Teräsvirta and White linearity test for the S&P500 index

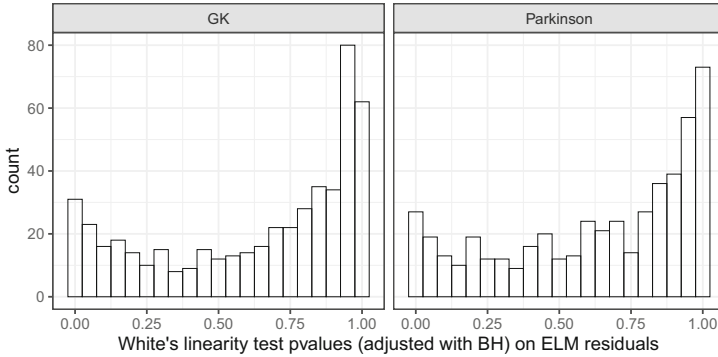|		Parkinson		Garman and Klass	
		Teräsvirta	White	Teräsvirta	White
Original series	Statistics	28.7966	30.0607	46.5113	42.3297
	pvalue	0.0000	0.0000	0.0000	0.0000
AR residuals	Statistics	10.6461	1.9668	7.4286	11.5053
	pvalue	0.0049	0.3740	0.0244	0.0032
ELM residuals	Statistics	3.1749	0.7008	3.7553	1.8316
	pvalue	0.2045	0.7044	0.1530	0.4002

Fig. 4. Distribution of White's linearity test p-values for the residuals from ELM models for all the constituencies of the S&P500, adjusted for multiple testing by controlling the FDR using the Benjamini-Hochberg procedure

References

1. Chou, R.Y., Chou, H., Liu, N.: Range volatility models and their applications in finance. In: Lee, C.F., Lee, A.C., Lee, J. (eds.) Handbook of Quantitative Finance and Risk Management, pp. 1273–1281. Springer, Boston (2010). https://doi.org/10.1007/978-0-387-77117-5_83
2. Garman, M., Klass, M.: On the estimation of security price volatilities from historical data. J. Bus. **53**, 67–78 (1980)
3. Huang, G., Huang, G.-B., Song, S., You, K.: Trends in extreme learning machines: a review. Neural Comput. **61**, 32–48 (2015)
4. Huang, G.-B., Chen, L., Siew, C.-K.: Universal approximation using incremental constructive feedforward networks with random hidden nodes. IEEE Trans. Neural Netw. **17**, 879–892 (2006)
5. La Rocca, M., Perna, C.: Nonlinear autoregressive sieve bootstrap based on extreme learning machines. Math. Biosci. Eng. **17**, 636–653 (2020)
6. Parkinson, M.: The extreme value method for estimating the variance of the rate of return. J. Bus. **53**, 61–65 (1980)

Mortality Risk. Incorporating the New Seasonal-Ageing Index (SAI) into a Pricing Strategy

Josep Lledó[(✉)] and Jose M. Pavía

Universitat de València, Valencia, Spain
{Josep.Lledo,Jose.M.Pavia}@uv.es

Abstract. Insurance companies use annual life tables to manage mortality risks despite intra-annual mortality risks showing sub-annual fluctuations. The difficulty of (accurately) measuring and computing these fluctuations is likely behind this decision. The research carried out by [5], however, offers new opportunities by developing a methodology that, for the first time in the literature, allows actuaries and statisticians to derive quarterly life tables from annual tables simply by using the so-called SAI (Seasonal-Ageing Indexes) coefficients. SAIs capture mortality risks taking into account both source of fluctuations, age and calendar quarters. This paper aims to discern the implications of using an approach based on SAIs and study how the new methodology could be employed in a competitive market where different insurance companies operate. On the one hand, we exemplify the new procedure on a real insurance portfolio by calculating premiums for different quarters. This gives, for each age and calendar quarter, the distance (measured in euros) between using annual and SAI-quarterly life tables. On the other hand, we study the market opportunities that the new methodology offers by simulating a two-company market, where one of the companies incorporates SAI-quarterly tables in its pricing processes. We analyse the impact of the use of these different market strategies on profit and loss accounts of both companies. The results of this research highlight some of the practical advantages of using the SAI-based approach.

1 Introduction

The introduction of a new pricing methodology has an impact on any business environment, and the insurance sector is no stranger to this fact. The price offered by an insurance company is determined by the premium of the insurance product. In life-insurance product pricing processes, it is common to use a life table to calculate the pure premiums [4]. Life tables are composed of multiple biometric functions, such as the probability that a person of age x does not reach age $x+1$, q_x, or life expectancy at birth, e_0^0. All the variables have a characteristic in common; they are expressed in annual frequency. However, several studies show that the seasonal behaviour of mortality (see [2] has a negative impact on the insurance sector [3,5].

The intra-annual behaviour patterns of the number of deaths of mortality risks point to the need to adopt a frequency other than the annual one. The study published by [5] enables quarterly mortality tables to be produced that collect intra-annual mortality patterns without the need to perform additional calculations, simply by applying the new SAI coefficients (Seasonal-Ageing Index) to the annual mortality tables. [5] develop this new methodology from a theoretical perspective and study its effects in practice on just one classic life insurance product. Thus, the impact of this methodology on the competitive market of the insurance sector as a whole, where multiple companies operate with the classic methodology of annual life tables, is still to be estimated. Among the numerous research questions, as yet unresolved, are: **(Q1)** What pricing strategy should a company that uses the SAI methodology adopt? and **(Q2)** How different would the returns per policy of two companies be if one adopts a classic methodology and the other uses a methodology based on SAIs?

To answer the research questions **Q1** and **Q2**, this study follows a two-pronged approach. On the one hand, we analyse the strategy to be adopted by an insurance company that uses the SAI approach in a insurance market in which competing companies operate using an approach based on annual life tables. On the other hand, we calculate the (hypothetical) income statement for each of the two approaches.
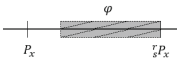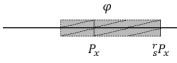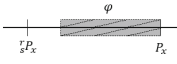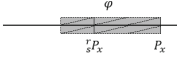
2   Benefits and Losses. Pricing Strategies

In the insurance sector the companies (insurers) establish their pricing policies and the consumer (insured) chooses between the different products. Logically, the behaviour of consumers is rational and, for the same product, they tend to choose the lowest price. At the moment, the dominant pricing methodology for life insurance products is based on the use of annual life tables. For a risk-life insurance product, whose main cover is death, the insurance price, also denoted as a pure premium or risk premium, P, is calculated as the sum insured (capital at risk) in the event of death, C, by the probability that a person of age x does not reach age $x+1$, q_x, $P = C \cdot q_x$. The main hypothesis of this classic approach is to assume that the risk of mortality between age x and $x+1$ is constant and independent of calendar.

This hypothesis, however, is debatable, if not clearly incorrect. [5] show the existence of intra-annual patterns in mortality risks, which present certain characteristics of quarterly periodicity (age and calendar). In this context, [5] make a double methodological contribution, openning up new possibilities by allowing pure premiums to be obtained for age x for each age quarter r and calendar quarter s, ${}_s^r P_x$. In this scheme, the corresponding annual premium is calculated as the sum of the four quarterly premiums, one for each quarter that the person remains in the insurance portfolio until the next renewal. To calculate the annual premium both the moment of the hiring and the exact age of the insured, (r, s), must be taken into account to proper compute the probability of dying in each quarter. For example, for a person aged 55 and 2 quarters (2Q) who takes out insurance in the third calendar quarter (summer), the premium for a capital of C is calculated as: ${}_3^2 P_{55} = C \cdot {}_3^2 q_{55} + C \cdot (1 - {}_3^2 q_{55}) \cdot {}_4^3 q_{55} + C \cdot (1 - {}_3^2 q_{55}) \cdot (1 - {}_4^3 q_{55})$ ${}_1^4 q_{55} + C \cdot (1 - {}_3^2 q_{55}) \cdot (1 - {}_4^3 q_{55}) \cdot (1 - {}_1^4 q_{55}) \cdot {}_2^1 q_{56}$. The use of quarterly life tables makes it possible to calculate 16 annual premiums for each age, one for each calendar s and

age r quarter, depending on the exact age and time when the insured takes out the policy. These annual premiums, obtained by applying the SAI procedure, can be compared with the annual premium calculated using the classic methodology.

In this context, let us consider two companies A and B. A uses the classic approach while B uses the SAI procedure. In a competitive market, given a fixed portfolio of potential clients, there are four possible scenarios, taking into account the relative prices of the two companies for the same product (see Fig. 1). When the two companies (A and B) compete in a competitive market with the same product, the insured, who behaves rationally, chooses the policy with the lowest premium (price). However, there may be certain situations in which the customer is willing to pay a higher price, within a threshold, φ, for the same product. This could be due to factors such as the influence of advertisements, the administrative cost of cancelling a pre-existing insurance policy or loyalty to an insurance company.

Scenario	Scheme	Market share for A	Market share for B
I	φ P_x .... $_s^r P_x$	1	0
II	φ P_x .... $_s^r P_x$	$1 - \theta$	θ
III	φ $_s^r P_x$ .... P_x	0	1
IV	φ $_s^r P_x$ .... P_x	θ	$1 - \theta$

Fig. 1. Possible scenarios depending on the price offered by Company A, P_x, and by Company B, $_s^r P_x$.

In scenarios I and III (see Fig. 1) (almost) all policies will be sold by the company with the cheapest price (A and B, respectively). In scenario I (esc. III) Company A (Co. B) takes over the market with its products at a price P_x ($_s^r P_x$). In scenarios II and IV, with smaller price differences, companies A and B will share a percentage of the market. In scenario II (esc. IV), Company B (Co. A) will enjoy a percentage θ of the market despite having a higher price. It is, there is a maximum price difference, φ, below which companies A and B will share a percentage of the portfolio's premiums. This occurs when the price difference between $_s^r P_x$ and P_x does not exceed a percentage β of the upper price between P_x and $_s^r P_x$. This parameter, β, measures the sensitivity that consumers have to prices between companies. A direct relationship between the parameters β and φ can be obtained. For example, $\varphi = \beta \cdot _s^r P_x$ in scenario I. If $\beta = 0$ ($\varphi = 0$), consumers will (strictly) prefer the lowest price. This sensitivity to prices, or equivalently the market share, can be proportional to the distance between both prices (through a linear function, which will depend on φ or β) or depend on a non-linear function between both prices, such as logarithmic or exponential functions. In any case, it seems reasonable to assume that when $_s^r P_x = P_x$ both companies will share the market equally: $\theta = 0.50$. Likewise, for $|_s^r P_x - P_x| \geq \varphi$ it is verified that $\theta = 0$.

To give an example, let us consider a 40-year-old man in his age quarter 1, Q1, $(r = 1)$, who takes out insurance in calendar quarter 1, winter, $(s = 1)$. Suppose that for this profile, the annual premium of a product that uses quarterly mortality tables is $^1_1P_{40} = 44€$ and that the premium of the same product that uses annual mortality tables is $P_{40} = 44.82€$. If we assume that $\beta = 0.05$, then $\varphi = P_{40} \cdot \beta = 2.24€$. Company B, which sells its products at a lower price, would gain most of the total premiums, but not all, since $P_x - ^r_s P_x < \varphi \rightarrow 44.82 - 44 = 0.82 < 2.24$ (scenario IV of Fig. 1). Company A will take a share of the policies despite selling its product at a higher price. Specifically, if we assume linearity, the percentage of policies taken out with Company B would be $1 - \theta = 0.5 + \frac{(P_x - ^r_s P_x) \cdot 0.5}{P_x - P_x \cdot (1 - \beta)} = 68.3\%$ while Company A would take the rest of the market, 31.7%.

Knowing the premiums that each company (A or B) gains from the total number of insured enables the total income of their income statements to be calculated (Technique Income Statements). To calculate this income, we start with an insurance portfolio made up of the total number of insured persons for age x in the age quarter r and calendar quarter s, $^r_s T_x$. The technical benefit is then calculated by deducting the benefits paid out as claims (insured sum of the people who die) from the total paid in as premiums. The total loss ratio, the total benefits paid as claims, will coincide with the probable payment flow of Company B (if it covers the entire market) since the quarterly tables capture the incidence of mortality for each quarter and this is calculated taking into account all demographic events [5]. Thus, for a life insurance product between the ages π_1 and π_2, the technical profit (loss), BT, of Company A can be calculated based on the policies that Company A acquires in scenarios I, II and IV, using $BT_A = \sum_{i=\pi_1}^{\pi_2} T_i (P_x - C \cdot ^r_s q_i) \vee (P_i < ^r_s P_i - \varphi) \vee (P_i - ^r_s P_i > \varphi) \vee (^r_s P_i - P_i > \varphi)$ and, for Company B, the policies it acquires in scenarios II, III and IV are considered: $BT_B = \sum_{i=\pi_1}^{\pi_2} \sum_{s'=1}^{4} \sum_{r'=1}^{4} {}^{r'}_{s'} T_i ({}^{r'}_{s'} P_i - C \cdot {}^{r'}_{s'} q_i) \vee ({}^r_s P_i - P_i > -\varphi) \vee ({}^r_s P_i < P_i - \varphi) \vee (P_i - ^r_s P_i > \varphi)$

3   Results

The previous section shows the different scenarios that can occur in a competitive market with the entry of a company that implements the use of quarterly life tables through the SAIs. To verify the results of these scenarios, real data from an insurance portfolio of a Spanish insurance company is used. Its activity involves the commercialisation of products whose main cover is death (life insurance) through a renewable annual insurance premium. The annual mortality tables used are those known as ($PASEM2019secondorder$) according to the Spanish insurance regulations [1]. The quarterly life tables are calculated applying the SAI methodology. The age range is defined in the underwriting policy and covers the interval $x \in [18, 75]$. The insurance portfolio corresponds to the year 2020 and its total size, $\sum_s^r T_x$, is 72,957 insured, of which 45,327 are men and 27,630 are women.

The calculations are carried out for all the policies of the insurance portfolio assuming $r = 1$ and using different values for $\beta = 0\%, 1\%, 2\%, 5\%, 10\%$. The hypothesis $r = 1$ is less strong than it seems since in annual insurance when $r > 2$ the probabilities that apply for an insured person of whole age x are those that correspond to age $x + 1$.

More realistic results would have been obtained assuming that for half of the portfolio $r = 1$ and for the other half $r = 2$.

Man

β	Concept	Winter	Spring	Summer	Autumn	Total
0%	Income	0 €	2,002,050 €	1,992,590 €	1,531,501 €	5,526,142 €
	Expenses	0 €	2,007,677 €	2,017,875 €	1,536,082 €	5,561,635 €
	Profit (Loss)	0 €	-5,627 €	-25,285 €	-4,581 €	-35,493 €
	Profit (Loss) (%)	0.00%	-0.28%	-1.27%	-0.30%	-0.64%
1%	Income	0 €	1,281,548 €	1,959,778 €	1,154,837 €	4,396,163 €
	Expenses	0 €	1,285,174 €	1,984,883 €	1,158,014 €	4,428,071 €
	Profit (Loss)	0 €	-3,626 €	-25,106 €	-3,177 €	-31,908 €
	Profit (Loss) (%)	0.00%	-0.28%	-1.28%	-0.28%	-0.73%
2%	Income	114,368 €	1,141,287 €	1,620,112 €	1,055,090 €	3,930,857 €
	Expenses	112,349 €	1,144,506 €	1,641,067 €	1,057,729 €	3,955,651 €
	Profit (Loss)	2,019 €	-3,220 €	-20,954 €	-2,639 €	-24,794 €
	Profit (Loss) (%)	1.77%	-0.28%	-1.29%	-0.25%	-0.63%
5%	Income	641,277 €	1,057,130 €	1,245,822 €	993,457 €	3,937,687 €
	Expenses	629,932 €	1,060,106 €	1,261,789 €	995,752 €	3,947,579 €
	Profit (Loss)	11,346 €	-2,976 €	-15,967 €	-2,295 €	-9,892 €
	Profit (Loss) (%)	1.77%	-0.28%	-1.28%	-0.23%	-0.25%
10%	Income	816,914 €	1,029,078 €	1,121,058 €	972,913 €	3,939,963 €
	Expenses	802,459 €	1,031,972 €	1,135,363 €	975,093 €	3,944,888 €
	Profit (Loss)	14,455 €	-2,895 €	-14,305 €	-2,180 €	-4,925 €
	Profit (Loss) (%)	1.77%	-0.28%	-1.28%	-0.22%	-0.12%

Woman

β	Concept	Winter	Spring	Summer	Autumn	Total
0%	Income	228,625 €	98,652 €	257,140 €	434,648 €	1,019,065 €
	Expenses	98,774 €	258,283 €	435,524 €	235,118 €	1,027,699 €
	Profit (Loss)	129,851 €	-159,631 €	-178,384 €	199,530 €	-8,634 €
	Profit (Loss) (%)	-0.60%	-0.12%	-0.44%	-0.20%	-0.34%
1%	Income	229,654 €	192,076 €	245,031 €	277,020 €	943,781 €
	Expenses	191,819 €	245,800 €	277,571 €	234,434 €	949,624 €
	Profit (Loss)	37,835 €	-53,724 €	-32,540 €	42,586 €	-5,843 €
	Profit (Loss) (%)	-0.44%	0.13%	-0.31%	-0.20%	-0.22%
2%	Income	233,023 €	218,780 €	239,985 €	255,579 €	947,368 €
	Expenses	218,383 €	240,444 €	256,069 €	231,446 €	946,342 €
	Profit (Loss)	14,641 €	-21,664 €	-16,083 €	24,132 €	1,026 €
	Profit (Loss) (%)	-0.27%	0.18%	-0.19%	-0.19%	-0.13%
5%	Income	233,001 €	235,130 €	239,426 €	242,713 €	950,271 €
	Expenses	234,644 €	239,622 €	243,168 €	231,826 €	949,260 €
	Profit (Loss)	-1,644 €	-4,492 €	-3,741 €	10,887 €	1,011 €
	Profit (Loss) (%)	-0.12%	0.21%	-0.08%	-0.19%	-0.05%
10%	Income	232,762 €	240,580 €	239,276 €	238,425 €	951,044 €
	Expenses	240,065 €	239,384 €	238,867 €	232,190 €	950,506 €
	Profit (Loss)	-7,303 €	1,197 €	409 €	6,235 €	538 €
	Profit (Loss) (%)	-0.07%	0.21%	-0.04%	-0.19%	-0.02%

Fig. 2. Technique Income Statement for Company A. Left table for men, right table for women.

As can be seen in Fig. 2, when $\beta = 0\%$ (only scenarios I and III apply), Company A has expected technical losses in the three quarters that it manages to sell its products since it sells them at a price lower than expenses (claims), both for men (0.64%) and for women (0.34%). These (total) differences are not uniform for the three quarters. Continuing with this β level, in summer the differences in premiums ${}_3^r P_x$ and P_x are greater than between P_x and the other quarterly premiums, ${}_2^r P_x$ and ${}_4^r P_x$. This leads to a loss of 1.27% (0.44%) in men (women).

However, when $\beta > 0\%$, their expected technical losses are reduced, both for men and women, because Company B shares clients with Company A (scenarios II and IV). For all β, the expected technical benefit of Company B is 0 since its expected premium income coincides with the expected cost for claims. Undoubtedly, the pricing strategy to be followed by a company that uses quarterly tables (Company B) is to aim to attract customers born in winter and to get customers born in summer to contract the product with the competition, Company A.

4   Conclusions and Future Research

The research carried out by [5] provides quarterly life tables without the need for additional calculations, simply by using SAIs. This study analyses the pricing strategy that a company adopting this new methodological procedure needs to use in a competitive market where consumers (insured) behave rationally in relation to price. Specifically the company that uses quarterly life tables needs to capture customers born in winter and encourage customers born in summer to choose the company that uses annual life tables. The main results suggest that not taking into account the exact times (quarters) of birth and when the policy is taken out represents a loss compared to a company that does take these variables into account. The differences found in a real portfolio, for a

classic annual renewable insurance (covering death), would be 0.64% and 0.34% for men and women, respectively.

Of course, the conclusions reached in this study still leave many research opportunities open. On the one hand, all the scenarios carried out in this research are based on pure premiums and no other factors are taken into account. The impact of the benefit surcharge that insurance companies incorporate in their premiums to deal with unexpected claims remains to be analysed. On the other hand, it would be helpful to have an alternative methodological procedure that allows quarterly life tables to be obtained without the need to perform the SAI procedure (the calculations carried out by [5] required the processing of more than 180 million microdata).

References

1. BOE: Resolución de 17 de diciembre de 2020, de la Dirección General de Seguros y Fondos de Pensiones, relativa a las tablas de mortalidad y supervivencia a utilizar por las entidades aseguradoras y reaseguradoras. Boletín Oficial del Estado **338**, 121566–121602 (2020)
2. Díaz, J., García, R., López, C., Linares, C., Tobías, A., Prieto, L.: Mortality impact of extreme winter temperatures. Int. J. Biometeorol. **49**, 179–183 (2005)
3. Fernández-Durán, J.J., Gregorio-Dominguez, M.M.: Seasonal mortality for fractional ages in short term life insurance. Scand. Actuarial J. **2015**, 266–277 (2015)
4. Lledó, J., Pavía, J.M., Morillas, F.G.: Incorporating big microdata in life table construction: a hypothesis-free estimator. Insur.: Math. Econ. **88**, 138–150 (2019)
5. Pavía, J.M., Lledó, J.: Estimation of the combined effects of ageing and seasonality on mortality risk. An application to Spain. J. R. Stat. Soc. Series A Stat. Soc. (2021). https://doi.org/10.1111/rssa.12769

Credit Spreads, Leverage and Volatility: A Cointegration Approach

Federico Maglione[(✉)]

Scuola Normale Superiore, Piazza dei Cavalieri 7, 56126 Pisa, Italy
federico.maglione@sns.it
https://www.sns.it/en/persona/federico-maglione

Abstract. This work documents the existence of a cointegration relationship between credit spreads, leverage and volatility for a large set of US companies. It is shown that accounting for the long-run equilibrium dynamic between these variables is essential to correctly explain credit spread changes. Using a novel structural model in which equity is modelled as a compound option on the firm's assets, a new methodology for estimating the unobservable market asset value and volatility is developed. The proposed model allows to reduce pricing errors in predicting credit spreads. In terms of correlation analysis, it is shown that not accounting for the long-run equilibrium equation embedded in an Error Correction Mechanism (ECM) results into a misspecification problem when regressing a set of explanatory variables onto the spread changes. Once credit spreads, leverage and volatility are correctly modelled, the fit of the regressions sensibly increases if compared to the results of previous research.

Keywords: Credit spreads · Financial leverage · Asset volatility · Cointegration · Compound options

1 Credit Risk Model and Calibration

The model generalises the one in Geske (1977): there equity is modelled as a compound option written on the firm assets, whilst here equity is seen as a n–fold compound call option, thus allowing a more realistic description of the firm's capital structure. The firm has indeed issued n bonds, maturing at t_i, with $i \in I = \{1, \dots, n\}$.

Assuming that default occurs the first time at which the firm is unable to issue new equity to repay its debt, and the firm's asset value process V_t follows a geometric Brownian motion with volatility σ_V and drift $(r - \varpi)$, where r is the risk-free rate and ϖ the payout rate, the firm's equity is

$$S\left(V_t, \sigma_V\right) = e^{-\varpi(t_n - t)} V_t \Phi_n\left(\mathbf{d}^{\mathbb{M}}; \boldsymbol{\Gamma}_n\right) - \sum_{i=1}^{n} e^{-r(t_i - t)} F_i \Phi_i\left(\mathbf{d}_i^{\mathbb{Q}}; \boldsymbol{\Gamma}_i\right) \qquad (1)$$

where $\mathbf{d}^{\mathbb{M}} := \left(d_i^{\mathbb{M}}\right)_{1 \leq i \leq n}$ and $\mathbf{d}_i^{\mathbb{Q}} = \left(d_j^{\mathbb{M}} - \sigma_V \sqrt{t_j - t}\right)_{1 \leq j \leq i}$ with

$$
d_i^{\mathbb{M}} = \frac{\ln\left(V/\bar{V}_i\right) + \left(r - \varpi + \sigma_V^2/2\right)(t_i - t)}{\sigma_V \sqrt{t_i - t}}, \quad
\boldsymbol{\Gamma}_i = \begin{pmatrix}
1 & \sqrt{\frac{t_1-t}{t_2-t}} & \cdots & \sqrt{\frac{t_1-t}{t_i-t}} \\
 & 1 & \cdots & \sqrt{\frac{t_2-t}{t_i-t}} \\
\cdots & \cdots & \cdots & \cdots \\
 & & 1 & \sqrt{\frac{t_{i-1}-t}{t_i-t}} \\
 & & & 1
\end{pmatrix}
$$

and $\Phi_i(\mathbf{z}; \boldsymbol{\Gamma})$ the cumulative distribution function of a i-dimensional normal random vector with zero mean and covariance matrix $\boldsymbol{\Gamma}$ calculated over the set $\bigtimes_{j=1}^{i}(-\infty, z_j)$. Also, $(\bar{V}_i)_{1 \leq i \leq n}$ is the latent sequence of default thresholds embedded in the firm's capital structure.

The unobservable parameters of the model are the value of the firm assets, V, and the asset volatility, σ_V. As the sequence of risk-neutral probabilities $\mathbb{Q}(\tau \geq t_i)$ can be estimated from the CDS spreads in a model-free fashion (Brigo 2005), the following system of non-linear equations can be employed to estimate both variables,

$$
\begin{cases}
S(V_t, \sigma_V) = S_t \\
\Phi_i^{\mathbb{Q}}(V_t, \sigma_V) = \widehat{\Phi}_i^{\mathbb{Q}} \quad \forall i \in I.
\end{cases}
\tag{2}
$$

Here, the functional form of $S(V_t, \sigma_V)$ and $\Phi_i^{\mathbb{Q}}(V_t, \sigma_V) = \Phi_i\left(\mathbf{d}_i^{\mathbb{Q}}(V_t, \sigma_V); \boldsymbol{\Gamma}_i\right)$ are obtained from (1). S is the observed stock price, whist $\widehat{\Phi}_i^{\mathbb{Q}}$ are the model-free risk neutral probability of survival (for maturity t_i) estimated from the CDS spread.

Once the estimates of V_t and σ_V are obtained, the volatility of the equity and the firm's leverage are calculated accordingly, that is

$$
\text{VOL}_t = \sigma_V \frac{V_t}{S_t} \frac{\partial S}{\partial V}, \qquad
\text{LEV}_t = \frac{V_t e^{-\varpi(t_n - t)} - S_t}{S_t}.
\tag{3}
$$

This novel estimation technique is applied to a set of 64 US companies, constituents of the S&P100 during the period January 2013–December 2017.

2   The Error Correction Mechanism

The structural approach to credit risk prescribes that the probability of default of a firm, and hence its credit spread, is driven by the firm's leverage and the volatility of its cash-flows. Models such as (1) clearly link these variable; moreover, changes in any of the inputs would affect the others (e.g. a change in the firm riskiness would impact the firm's equity and hence its leverage). Given the evidence in Collin-Dufresne and Goldstein (2001) in which average firms are shown to target stationary leverage ratios, a long-run equilibrium between the drivers of credit spread would then appear.

Despite the functional form in (1) being nonlinear, an Error Correction Mechanism could be used as first-order approximation for describing such link, once

such variables display stochastic trends. Empirical test confirms both the presence of unit roots and the cointegration.

Based on the structural approach of default, the spread is likely to follow upon changes on the firm's financial leverage (LEV) and riskiness (VOL) and not vice versa. Therefore, the model is implemented à la Engle-Granger instead of using a VECM (that is, only one cointegrating vector is estimated).

Assume the long-run equilibrium equation to be

$$\text{CDS}_{i,t} = \theta_{i,0} + \theta_{i,L}\text{LEV}_{i,t} + \theta_{i,V}\text{VOL}_{i,t} + \varepsilon_{i,t}, \tag{4}$$

in which $(\text{CDS, LEV, VOL})_{i,t}$ are, respectively, the CDS spread (for a given maturity), model-implied market leverage and equity volatility of firm i at time t. CDS is observed, whilst LEV and VOL are estimated as in (3).

The autoregressive distributive lag, $\text{ARDL}(1,1,1)$, dynamic panel specification of (4) (with exogenous variables, $\Delta\mathbf{X}$ below) is defined as

$$\begin{aligned}
\text{CDS}_{i,t} = \alpha_i + \phi_i\text{CDS}_{i,t-1} + \beta_{i,0}\text{LEV}_{i,t} + \beta_{i,1}\text{LEV}_{i,t-1} + \gamma_{i,0}\text{VOL}_{i,t} \\
+ \gamma_{i,1}\text{VOL}_{i,t-1} + \boldsymbol{\xi}^\top\Delta\mathbf{X}_t + \eta_{i,t},
\end{aligned} \tag{5}$$

and the error correction reparameterization of (5) is

$$\begin{aligned}
\Delta\text{CDS}_{i,t} &= \lambda_i\left(\text{CDS}_{i,t-1} - \theta_{i,0} - \theta_{i,L}\text{LEV}_{i,t-1} - \theta_{i,V}\text{VOL}_{i,t-1}\right) \\
&\quad + \beta_{i,0}\Delta\text{LEV}_{i,t} + \gamma_{i,0}\Delta\text{VOL}_{i,t} + \boldsymbol{\xi}^\top\Delta\mathbf{X}_t + \eta_{i,t} \\
&= \lambda_i\varepsilon_{i,t-1} + \beta_{i,0}\Delta\text{LEV}_{i,t} + \gamma_{i,0}\Delta\text{VOL}_{i,t} + \boldsymbol{\xi}^\top\Delta\mathbf{X}_t + \eta_{i,t}
\end{aligned} \tag{6}$$

where $\lambda_i = -(1 - \phi_i)$, $\theta_{i,0} = \frac{\alpha_i}{1-\phi_i}$, $\theta_{i,L} = \frac{\beta_{i,0}+\beta_{i,1}}{1-\phi_i}$, and $\theta_{i,V} = \frac{\gamma_{i,0}+\gamma_{i,1}}{1-\phi_i}$. $\lambda_i = 0$ is expected to be significantly negative under the prior assumption that the variables show a return to a long-run equilibrium. Of particular importance is the vector $\boldsymbol{\theta} = (\theta_L, \theta_V)$, which contains the long-run relationships between the variables driving the spreads.

3   The Main Result

Most of the results are qualitatively similar when 1-, 5- and 10-year spreads are used: all the variables, both endogenous and exogenous, display the sign that the theory prescribes and previous empirical research has found. Table 1 reports the results for the 1-year spread (5- and 10-year are) and Table 2 reports the adjusted R^2 of the models for different terms. Based on the latter, it is clear that the cointegration mechanism is more successful at explaining shorter maturities than long-term ones, thus suggesting a more responsiveness of short-term spreads to changes in the firm's leverage and risk.

This novel estimation technique is applied to a set of 64 US companies, constituents of the S&P100 during the period January 2013 – December 2017. Companies with either preferred equity or subject to merges or acquisitions are excluded. Also, only companies for which CDS spreads are available are included.

Table 1. ECM for 1−year CDS spreads. All the variables which structural models predict to influence the change in spreads are statistically significant and have the predicted signs. The loading on the cointegrating equation (ε) is negative and statistically significant, thus confirming the existence of a long-term equilibrium which spreads, volatility and leverage converge to. Number of observations: 16,640; number of groups: 64; observations per group: 260. Significance levels: 10% (*), 5% (**), 1% (***).

1−year CDS spread				
Long-run equilibrium	Coefficient	t–stat	p–value	
VOL	0.0028	9.88	0.000	***
LEV	0.0024	14.45	0.000	***
Short-term adjustment	Coefficient	t–stat	p–value	
ε	−0.1005	−11.52	0.000	***
ΔVOL	0.0074	6.29	0.000	***
ΔLEV	0.0036	5.23	0.000	***
ΔLevel	−0.0566	−4.55	0.000	***
ΔSlope	0.0213	4.56	0.000	***
ΔCurvature	1.2409	3.77	0.000	***
$\Delta \ln$(S&P500)	−0.0013	−5.13	0.000	***
ΔSkew	5E−07	2.37	0.018	**
Constant	−0.0001	−9.66	0.000	***

Table 2. Adjusted R^2s of the firm-specific time-series regressions in (6) (short-term adjustments). As shown by both the mean and median adjusted R^2, the explanatory power of the variables which should affect credit spread changes as predicted by structural models diminishes with the maturity of the spread.

	1−year	5−year	10−year
	adj–R^2		
Mean	0.69	0.45	0.30
Median	0.71	0.41	0.24
Min	0.22	0.07	0.01
Max	0.94	0.91	0.90

For what concerns the long-run equilibrium, both volatility (VOL) and leverage (LEV) display a positive and statistically significant loading: an increase in either VOL or LEV induce a larger level of the spread in the long-run. Focusing on the short-term adjustment, changes in both the firm's equity volatility and its financial leverage increase the change in the spread. In terms of economic significance, an increase of 1% in the firm's volatility increases the CDS spread of 0.7 bps. Similarly, an identical change in the firm's financial leverage induces the spread to increase of 0.4 bps.

Following Collin-Dufresne et al. (2001), exogenous variables, in changes ($\Delta\mathbf{X}$), are also added. These are the change in level, slope and curvature of the term structure of interest rates, the log-return on the S&P500, and the change in the CBOE Skew Index. The change in the Skew Index is used to proxy the effect of systematic jumps on credit risk.

First, the changes in the level of interest rates have a negative impact on the credit spread as in Longstaff and Schwartz (1995) and Duffee (1998). Likewise, the positive coefficients of the changes on the slope and curvature of the term structure are consistent with the findings of previous studies. Also, positive returns in the S&P500 – which accounts for growing economy and therefore an increasing expected recovery rate – have the effect to reduce the spread as suggested by economic intuition.

The variable ΔSkew is the only one that changes sign and significance between short-term spread (1-year) and medium- and long-term spreads (5- and 10-year). As shown in Zhou (2001), Zhang et al. (2009) and Du et al. (2019), jumps are necessary to explain the level of short-term spreads: structural models which account only for diffusive shocks in the asset value process imply zero instantaneous probability of default and therefore cannot meet the observed level of 6-month and 1-year spreads. Hence, the coefficient of ΔSkew is positive for 1-year spread changes as expected.

Finally, the estimated coefficient of the long-run equation (ε) is negative, within the unit circle and statistically significant. The closer the estimate is to zero, the slower is the adjustment. As expected, the size of the coefficient is larger, in absolute value, for shorter maturities: short-term spreads adjust faster to shocks in the firm's volatility and leverage. The associated t–statistic is also larger for 1–year spread changes. Conversely, the degree of cointegration becomes stronger at longer horizons: the t–statistics of the long-run equilibrium equation increase with the maturity of the CDS.

The cointegrating mechanism is also able to enhance the fit of the regressions on the spreads as compared with Collin-Dufresne et al. (2001) and other studies. Average adjusted–R^2s of 69%, 45% and 30% are obtained for 1-, 5- and 10-years spread changes and are reported in Table 2.

As the goodness-to-fit of the ECM model is evidently superior to the ones of a simple regression on changes, this provides extra evidence of the importance of a long-run equilibrium dynamic which must be taken into account to correctly identify how credit spreads change.

4   Conclusions

This paper develops a new estimation technique for the unobservable firm's asset value and volatility which relies only on the observable equity value, risk-neutral probability of default and the face value of the firm's debt.

The estimated parameters are used to investigate the existence of cointegration between credit spreads and those variables which structural models of default predict driving their level. Estimations confirm the presence of an error-correction mechanism which leads to a long-equilibrium between the level of the

spreads, financial leverage and the volatility of the firm's equity. Once the cointegration equation is accounted for, the goodness-to-fit of the regressions on the changes improves substantially compared to previous studies.

In conclusion, a structural model where equity is modelled as a compound option provides substantial improvement in predicting spreads out-of-sample, thus suggesting its superior ability in capturing firms' default dynamics. Most importantly, this work is the first to document the cointegration between CDS spreads, financial leverage and the firm's risk in a large panel of US firms. Once the cointegration equation is added to the regressions on credit spread changes, the selected variables do explain quite well their variation. Consistently with previous findings and the economic intuition, it is shown that short-term spreads react more quickly to shocks to the long-run equilibrium and that jumps affect short- and long-term spreads differently. Also, most of the variation in the cross-section appears to be driven by firm-specific characteristics rather than systematic factors.

References

Brigo, D.: Market models for CDS options and callable floaters. Risk Mag. 89–94 (2005)

Collin-Dufresne, P., Goldstein, R.S.: Do credit spreads reflect stationary leverage ratios? J. Finan. **56**(5), 1929–1957 (2001)

Collin-Dufresne, P., Goldstein, R.S., Martin, J.S.: The determinants of credit spread changes. J. Finan. **56**(6), 2177–2207 (2001)

Du, D., Elkamhi, R., Ericsson, J.: Time-varying asset volatility and the credit spread puzzle. J. Finan. **74**(4), 1841–1885 (2019)

Duffee, G.R.: The relation between treasury yields and corporate bond yield spreads. J. Finan. **53**(6), 2225–2241 (1998)

Geske, R.: The valuation of corporate liabilities as compound options. J. Finan. Quant. Anal. **12**(4), 541–552 (1977)

Longstaff, F.A., Schwartz, E.: A simple approach to valuing risky fixed and floating rate debt. J. Finan. **50**(3), 789–821 (1995)

Zhang, B.Y., Zhou, H., Zhu, H.: Explaining credit default swap spreads with the equity volatility and jump risks of individual firms. Rev. Finan. Stud. **22**(12), 5099–5131 (2009)

Zhou, C.: The term structure of credit spreads with jump risk. J. Banking Finan. **25**(11), 2015–2040 (2001)

Business Intelligence Modelling for Studying Science Parks Externalities

Valentina Mallamaci[1] and Massimiliano Ferrara[1,2(✉)]

[1] Department of Law, Economics and Human Sciences and Decisions Lab, Mediterranea University, 89124 Reggio Calabria, Italy
{valentina.mallamaci,massimiliano.ferrara}@unirc.it

[2] ICRIOS, The Invernizzi Centre for Reasearch in Innovation, Organization, Strategy and Entrepreneurship - Bocconi University, Department of Management and Technology, 20136 Milan, Italy

Abstract. The key to business success for many companies is the correct use of data to make better decisions. Companies need to use robust and efficient tools such as Business Intelligence (BI) as positive catalysts to achieve this goal, which can assist them in mechanizing the tasks of analysis, decision making, strategy formulation and forecasting.

Therefore, the main objective of the work is to answer the question whether operationalization of Business Intelligence, Organizational Learning (OL) and Innovation can provide financial performance enhancement for companies. It is an applied research as it examines the theoretical structures in a real context of start-ups located in the Shanghai Zizhu Science-based Industrial Park to demonstrate what kind of externality it generates on participating companies.

Research findings demonstrate that Business Intelligence and innovation have a critical influence on the companies conduct. But there was no meaningful relationship between Organizational Learning and financial performance of the same companies.

Keywords: Science Park · Business Intelligence · Financial Performance

1 Introduction

The main corporate objective is to generate profit and increase the well-being of both the shareholders and the various stakeholders who will interact with the company.

Science parks are indeed complex institutions designed to be applied as policy tools to support the development of innovative startups and regional clusters.

In fact, they facilitate the exchange of knowledge between universities, research and development bodies and companies. For the pursuit of an efficient interaction between the parties involved, companies and Science Parks, it is essential to examine the factors that influence the financial performance of the company, considering the limited availability of resources and information.

Therefore, in addition to paying attention to factors that are necessary for sustainability in other companies, managers should pay attention to factors such as business intelligence (BI), organizational learning (OL), innovation and etc. in order to increase the efficiency, effectiveness and durability of innovative companies.

According to many economic experts, entrepreneurship as an economic driver, plays a variety of roles in society and is the basis of all human developments and progress[1].

So, the objective of this paper is twofold: as regards companies, to analyze the impact of BI, OL and innovation on the financial performance of startups located within a science park.

In the literature, International Association of Science Parks[1] provided the main definition of Science Park as follows: a Science Park is a business support and technology transfer initiative that.

1. encourages and supports the startup and incubation of innovation led, high growth knowledge based businesses;
2. provides an environment where larger and international businesses can develop specific and close interactions with a particular center of knowledge creation for their mutual benefit;
3. has formal and operational links with poles of knowledge creation such as universities, higher education institutes and research organizations [2, 8].

In recent years, an increasing number of empirical studies has examined the relationship between innovation and firms' performance considering different types of models, estimation methods, measures of growth and innovation activity.

In particular, corporate performance, and therefore investments in innovation, are strictly connected to the financial performances that influence the entire management. These are usually measured in terms of profitability through three accounting variables: Return On Assets, Return On Equity and Return On Sales which provide a system of measurements complete enough for potential investors to optimally assess the company's financial performance [2].

In the study on the relationship between corporate governance and financial performance firms by Kyere and Ausloos (2021), ROA and Tobin's Q were used as dependent variables. ROA gives an indication of how best the assets of a company is utilized to generate profit and it is not affected by leverage or other extraordinary and discretionary items; it is calculated by dividing annual earnings of the company by its total assets. The Tobin's Q is usually used to capture existing assets, future growth potentials of the company and investors' expectations to future events, including evaluation of current business strategies; it is a ratio of market value of company outstanding stock and debt divided by replacement cost of the company's assets ("book value"). Instead, the independent variables considered in the same study are related to corporate governance mechanisms: insider shareholding, board size, independent board, CEO duality, audit committee meetings [3].

---

[1] IASP International Board, 6th February, 2002.

As anticipated, technological innovation is also correlated to the business intelligence mechanisms introduced in the company to be able to make all available data productive. In fact, BI methodologies, through an assortment of coordinated operational just as choice help applications and databases, convert raw data sets into information to support decision-making processes to optimize knowledge and make processes faster and more efficient. Therefore, the development of BI is one of the fundamental elements of innovation within a company and determines the improvement of its overall performance [4].

The last aspect considered is the organizational learning which is determined by the set of organizational features, practices and issues that enable the learning processes. Thus, the OL translates into the ability to convert new and past ideas into actions that can improve business performance ahead of competitors in order to gain a sustainable competitive advantage. This occurs through the generation, acquisition, dissemination and integration of knowledge and the modification of its behavior. In fact, it is argued that firms better at learning get a better chance of increasing market share thanks to a responsive structure that can quickly react to new challenges [5].

Finally, Ferrara and Mavilia (2014), despite constituting a pioneering analysis, has showed a lot of interesting features of TP/SP's effects on both regional economic growth and associated (or incubated) firms' growth in terms of revenues. In particular, they found both that TP/SP impacts are different if analyzed by geographic location and that their effects are still evident even in the aggregate model. The number of TP/SP per region seems to display a positive role in sustaining the economic growth of corresponding regions. In addition, the patenting activity and the creation of research centers foster the growth of affiliated firms, which in turns affects regional economy's parameters. To the contrary, the distance between the TP/SP and affiliated firms reduces the growing potential of the latter. In addition, firms within a TP/SP turn out outperforms (largely) the regional average and more recent structures tend to be more prone to both patenting activity and high-level growth. Younger structures are also characterized by higher dispersion rates [7].

2 Problem Statement and Methodology

2.1 Research Method and Data Collection

Statistical population of the research is considered start-up companies located in the Science Park (Shanghai Zizhu Science-based Industrial Park, located in the southeast of Minhang District, Shanghai). The sampling of this research is simple random. Morgan table was employed to demonstrate the sample size as shown in Table 1. It was found that the number of executives of start-up companies located in the Science Park is 400 people; therefore, the number of samples requested is 196 people and 280 questionnaires were distributed to compensate for any statistical bias due to the choice of sample.

Table 1 Krejcie and Morgan table (1970) for determining sample size of a known population (N is the population size; S is the sample size.)

N	S	N	S	N	S	N	S	N	S
10	10	100	80	280	162	800	260	2800	338
15	14	110	86	290	165	850	265	3000	341
20	19	120	92	300	169	900	269	3500	346
25	24	130	97	320	175	950	274	4000	351
30	28	140	103	340	181	1000	278	4500	354
35	32	150	108	360	186	1100	285	5000	357
40	36	160	113	380	191	1200	291	6000	361
45	40	170	118	**400**	**196**	1300	297	7000	364
50	44	180	123	420	201	1400	302	8000	367
55	48	190	127	440	205	1500	306	9000	368
60	52	200	132	460	210	1600	310	10000	370
65	56	210	136	480	214	1700	313	15000	375
70	59	220	140	500	217	1800	317	20000	377
75	63	230	144	550	226	1900	320	30000	379
80	66	240	148	600	234	2000	322	40000	380
85	70	250	152	650	242	2200	327	50000	381
90	73	260	155	700	248	2400	331	75000	382
95	76	270	159	750	254	2600	335	1000000	384

The online questionnaire (according to the Corona pandemic) includes demographic information and 47 questions of research variables (financial performance – 12 questions; business intelligence – 12 questions; innovation – 12 questions; organizational learning – 11 questions) and its validity and reliability have been appropriately verified. In particular, Cronbach's alpha approach has been applied to compute the internal consistency of the measuring instrument; since the value attributed to all variables is greater than 0.7 (0.913 for the whole questionnaire), the case has agreeable dependability.

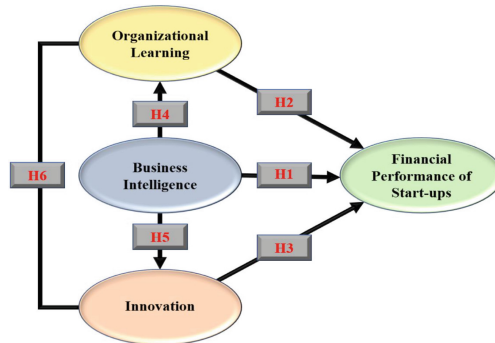Figure 1 shows the research' conceptual model [1].

Fig. 1. Conceptual model of research.

2.2 Research Data Analysis Method

In the study, the demographic data of the research were described employing descriptive statistics, frequency tables and pie charts and the data were analyzed employing SPSS software.

2.3 Structural Equation Modeling

There are two important issues in measuring variables in the behavioral and cognitive sciences. A. Measurements, and B. Cause and effect relationships between variables. Structural equation models contain two parts, the measurement model (external model) and the structural function model (internal model). Structural Equation Models by integrating the two models of confirmatory factor analysis and structural function analysis, many problems and difficulties of measuring latent variables and inferring causal relationships between these latent variables are solved. One of the strongest and most appropriate methods of analysis is multivariate analysis, which means the analysis of different variables that in a theory-based structure, demonstrates the simultaneous influences of variables on each other. The default partial least squares method does not require the type of distribution of the measurement variables. Therefore, it is suitable and practical for data with abnormal distribution or unknown distribution [1].

3 Main Results

Employing the internal model, hypotheses could be inspected. By correlating the value of "t" computed for the coefficient of every path, we could confirm or reject the research hypothesis. Therefore, if certain rate of the "t" statistic is larger than 1.96, it is meaningful at the 95% trust level. In analysis, "t" statistics value for all paths except the path of OL to financial performance, is larger than 1.96 and as a result are meaningful at the 95% confidence level.

The coefficient of perception for the financial performance variable is estimated to be 0.62 and shows that the variables of OL, BI and Innovation, together, could explain 62% of the changes in financial performance. According to the value of standard coefficient

and "t" statistics, variables of Innovation (0.568) and BI (0.233) had the topmost impact on the financial performance variable, respectively, and the OL variable did not have a significant effect. Similarly, the variables of OL and BI explain a total of 0.66 of the changes in Innovation. According to the value of standard coefficient and "t" statistics, OL variable has a greater impact on the innovation variable than the BI variable. According to the number of coefficients of determination, it can be said that the BI variable explains 27% of the changes in the OL variable [1].

4   Conclusions

The illustrated research aims to answer the question whether operationalization of BI, OL and Innovation can provide financial performance enhancement for these organizations. Based on the analysis, out of 6 hypotheses, 5 hypotheses were confirmed and 1 hypothesis was rejected in the study population. In particular, the main results are:

- Business intelligence has a positive and critical influence on organizational learning. Particularly, improving BI can improve OL.
- Organizational learning has a positive and critical influence on innovation. Particularly, OL can increase innovation.
- Business intelligence has a positive and critical influence on innovation. Particularly, BI can increase innovation.
- Following the existence of a relationship between organizational learning on the financial performance of innovative companies, the results showed a lack of relationship between these two variables in the statistical population of the study and this hypothesis was rejected. One of the reasons is the discrepancy between organizational maturity formed in start-up companies versus large companies.
- Business intelligence has a positive and critical influence on financial efficacy. Particularly, BI can improve financial performance. This path coefficient is significant at the error level of 0.05.
- Innovation has a positive and critical influence on financial efficacy. Particularly, innovation can improve financial performance [1].

References

1. Yang, M., Sulieman, R., Yin, Y., Mallamaci, V., Alrabaiah, H.: The effect of business intelligence, organizational learning and innovation on the financial performance of innovative companies located in Science Park. Inf. Process. Manage. **59**(2), 102–852 (2021)
2. Waddock, S.A., Graves, S.B.: The corporate social performance– financial performance link. Strateg. Manag. J. **18**(4), 303–319 (1997)
3. Kyere, M., Ausloos, M.: Corporate governance and firms financial performance in the United Kingdom. Int. J. Financ. Econ. **26**(2), 1871–1885 (2021)
4. Nithya, N., Kiruthika, R.: Impact of business intelligence adoption on performance of banks: A conceptual framework. J. Ambient. Intell. Humaniz. Comput. **12**(2), 3139–3150 (2021)

5. Hailekiros, G.S., Renyong, H.: The effect of organizational learning capability on firm performance: mediated by technological innovation capability. Eur. J. Bus. Manag. **8**(30), 87–95 (2016)
6. Lamperti, F., Mavilia, R., Castellini, S.: The role of Science Parks: a puzzle of growth, innovation and R&D investments. J. Technol. Transf. **42**(1), 158–183 (2017)
7. Ferrara, M., Mavilia, R.: The effects of technopoles and science parks on regional economies in Italy. WSEAS Trans. Bus. Econ. **11**(1), 537–549 (2014)
8. Ferrara, M., Mavilia, R.: Dai distretti industriali ai poli di innovazione. L'Italia nel Mediterraneo. 2nd edn. Egea, Milano (2013)

Surrender and Path-Dependent Guarantees in Variable Annuities: Integral Equation Solutions and Benchmark Methods

Antonio L. Martire[1], Emilio Russo[2(✉)], and Alessandro Staino[2]

[1] Department of Methods and Models for Economics, Territory and Finance, University of Rome 'La Sapienza', Via del Castro Laurenziano, 9, 00161 Rome, Italy
antonioluciano.martire@uniroma1.it

[2] Department of Economics, Statistics, and Finance, University of Calabria, Ponte Bucci cubo 1C, 87036 Rende, CS, Italy
{emilio.russo,alessandro.staino}@unical.it

Abstract. We investigate the evaluation problem of variable annuities by considering guaranteed minimum maturity benefits with constant or path-dependent guarantees of barrier and lookback type. We propose to solve the non-standard Volterra integral equations associated with the policy valuations through a randomized trapezoidal quadrature rule combined with an interpolation technique. Such a rule improves the convergence rate with respect to the classical trapezoidal quadrature, while the interpolation technique allows us to obtain an efficient algorithm that produces an accurate approximation of the early exercise boundary. The method accuracy is assessed by constructing two benchmarks based on lattice approaches and the least-squares Monte Carlo simulations. In the first case, a novel algorithm for the lookback path-dependent guarantee is obtained thanks to the lattice convergence properties.

Keywords: Variable annuity · Guaranteed minimum maturity/accumulation benefit · Volterra integral equation · Randomized trapezoidal quadrature · Lattice model · Monte Carlo simulation

1 The Framework

We consider a variable annuity (VA) policy issued at time 0 and maturing at time T for which the policyholder pays a single premium, then invested by the insurer in a fund presenting the following dynamics under the risk-neutral probability measure $\mathbb{Q}$, $dS_t = rS_t dt + \sigma S_t dB_t$, where r is the constant instantaneous risk-free rate of return, σ is the return volatility, and $(B_t)_{0 \leq t \leq T}$ is a standard Brownian motion in the probability space $(\Omega, \mathcal{F}, \mathbb{Q})$, endowed with the filtration $(\mathcal{F}_t)_{0 \leq t \leq T}$.

In absence of the possibility of surrendering the contract early, the policy pays at maturity T the maximum between the policyholder account value at time T, W_T, and the value of the guaranteed minimum payoff at the same time, G_T. Part of the premium is paid for the guarantee and it is supposed to be collected as a fee continuously withdrawn from the policyholder account at a constant rate α, i.e., at any time $0 \leq t \leq T$, $W_t = e^{-\alpha t} S_t$, with $dW_t = (r - \alpha) W_t dt + \sigma W_t dB_t$.

Whenever the VA policy presents a surrender option for the embedded guarantee, the policyholder may decide to exercise the option at any time before maturity. Clearly, being a rational individual, the policyholder optimally exercises the surrender option according to a strategy that is strictly dependent upon the type of the embedded guarantee, which we suppose to be of three different types as detailed hereafter in **Model 1**, **Model 2**, and **Model 3**.

In **Model 1**, we consider a VA contract with a guaranteed minimum maturity benefit (GMMB) characterized by constant guarantees. Decomposing the contract payoff according to Brennan and Schwartz [1], the policyholder will receive the account value W_T at maturity T and, whenever the surrender option on the guarantee is exercised at time t before maturity, the surrender benefit modelled as $(G - e^{-\kappa(T-t)} W_t)^+$, where G represents the constant minimum guaranteed amount and a penalty for early surrender, having the form $e^{-\kappa(T-t)}$, is applied to the account value W_t. To sum up, the value of the VA contract at a generic time t may be written as

$$VA_1(t, W_t) = e^{-r(T-t)} \mathbb{E}^{\mathbb{Q}}[W_T | \mathcal{F}_t] + \sup_{\tau^* \in \mathcal{S}(t,T)} \mathbb{E}^{\mathbb{Q}}\left[e^{-r(\tau^*-t)} (G - e^{-\kappa(T-\tau^*)} W_{\tau^*})^+ | \mathcal{F}_t \right],$$

$$(1)$$

where $\mathcal{S}(t,T)$ indicates the set of all the stopping times of the filtration $(\mathcal{F}_t)_{0 \leq t \leq T}$.

Model 2 extends **Model 1** by embedding in the VA contract with GMMB a path-dependent guarantee having the form of an up-and-out barrier option. In this case, if the policyholder account value hits the up-and-out barrier H from below, the embedded guarantee disappears and the policyholder receives the account value at maturity. At a generic time t where the surrender option has not yet been exercised, this VA contract has value

$$VA_2(t, W_t) = e^{-r(T-t)} \mathbb{E}^{\mathbb{Q}}[W_T | \mathcal{F}_t] +$$
$$+ \sup_{\tau^* \in \mathcal{S}(t,T)} \mathbb{E}^{\mathbb{Q}}\left[e^{-r(\tau^*-t)} (G - e^{-\kappa(T-\tau^*)} W_{\tau^*})^+ \mathbb{1}_{\{\tau_H > T\}} | \mathcal{F}_t \right],$$

where $\mathbb{1}_{\{.\}}$ is the indicator function and $\tau_H = \inf\{s > 0 | W_s = H\}$ represents the first barrier hitting time.

Model 3 complicates further **Model 2** by adding to **Model 1** a path-dependent guarantee with a lookback feature that is, clearly, more complex to be managed. Indeed, the guarantee is now described as a constant proportion φ of the historical maximum registered by the policyholder account value during the policy lifetime, defined as $M_t = \max_{0 \leq \pi \leq t} W_\pi$. The result is that, in this

case, the guarantee is time-dependent being its value defined as $G_t = \varphi M_t$, with $\varphi \in [0,1]$, and the VA contract value is given by

$$V A_3(t, W_t, M_t) = e^{-r(T-t)} \mathbb{E}^Q \left[W_T | \mathcal{F}_t \right] +$$

$$+ \sup_{\tau^* \in \mathcal{S}(t,T)} \mathbb{E}^Q \left[e^{-r(\tau^*-t)} \left(\varphi M_{\tau^*} - e^{-\kappa(T-\tau^*)} W_{\tau^*} \right)^+ | \mathcal{F}_t \right].$$

2   Volterra Integral Equation Approach

When considering the VA with GMMB in **Model 1**, the embedded guarantee in (1), indicated as

$$V_1(t, f) = \sup_{\tau^* \in \mathcal{S}(t,T)} \mathbb{E}^Q \left[e^{-r(\tau^*-t)} \max \left(G - e^{-\kappa(T-\tau^*)} W_{\tau^*}, 0 \right) | W_t = f \right],$$

satisfies the following variational inequality

$$\begin{cases} \mathcal{L} V_1(t, f) \le 0 & \text{if} \quad V_1(t, f) = \max \left(G - e^{-\kappa(T-t)} f, 0 \right), \\ \mathcal{L} V_1(t, f) = 0 & \text{if} \quad V_1(t, f) > \max \left(G - e^{-\kappa(T-t)} f, 0 \right), \\ V_1(T, f) = \max(G - f, 0), \end{cases} \quad (2)$$

on the domain $\mathcal{D}_1 = \{(t, f) | \ 0 \le t \le T, 0 < f < \infty\}$. In problem (2), the continuation region is $\mathcal{C}_1 = \left\{ (t, f) \in \mathcal{D}_1 | V_1(t, f) > \max \left(G - e^{-\kappa(T-t)} f, 0 \right) \right\}$ and the surrender region is $\mathcal{S}_1 = \left\{ (t, f) \in \mathcal{D}_1 | V_1(t, f) = G - e^{-\kappa(T-t)} f \right\}$. Defining the optimal surrender boundary as $B_1(t) = \inf \left\{ e^{-\kappa(T-t)} f \in \mathbb{R}_+ | (t, f) \in \mathcal{C}_1 \right\}$, the regions $\mathcal{C}_1$ and $\mathcal{S}_1$ may be redefined as $\mathcal{C}_1 = \left\{ (t, f) \in \mathcal{D}_1 | e^{-\kappa(T-t)} f > B_1(t) \right\}$ and $\mathcal{S}_1 = \left\{ (t, f) \in \mathcal{D}_1 | e^{-\kappa(T-t)} f \le B_1(t) \right\}$. In addition, at the free boundary $B_1(t) = e^{-\kappa(T-t)} f$, the following conditions hold:

$$V_1 \left(t, e^{\kappa(T-t)} B_1(t) \right) = G - B_1(t), \quad \frac{\partial V_1}{\partial f} \left(t, e^{\kappa(T-t)} B_1(t) \right) = -e^{-\kappa(T-t)}. \quad (3)$$

The value of the embedded guarantee $V_1(t, f)$ satisfies the problem

$$\begin{cases} \mathcal{L} V_1(t, f) = - \left(rG - (\alpha - \kappa) e^{-\kappa(T-t)} f \right) \mathbb{1}_{\{f < e^{\kappa(T-t)} B_1(t)\}}, \\ V_1(T, f) = \max(G - f, 0), \end{cases}$$

which presents the solution in the Volterra integral equation form given by

$$V_1(t, f) = Ge^{-r\tau} \Phi \left(-d^- \left(\tau, \frac{f}{G} \right) \right) - fe^{-\alpha\tau} \Phi \left(-d^+ \left(\tau, \frac{f}{G} \right) \right) +$$

$$+ rG \int_0^\tau e^{-r(\tau-\xi)} \Phi \left(-d^- \left(\tau - \xi, \frac{f}{e^{\kappa\xi} \widetilde{B}_1(\xi)} \right) \right) d\xi +$$

$$- (\alpha - \kappa) f \int_0^\tau e^{-\kappa\xi} e^{-\alpha(\tau-\xi)} \Phi \left(-d^+ \left(\tau - \xi, \frac{f}{e^{\kappa\xi} \widetilde{B}_1(\xi)} \right) \right) d\xi, \quad (4)$$

where $\tau = T - t$, $\widetilde{B}_1(\tau)$ is the time reversed free-boundary, $\varPhi(\cdot)$ is the cumulative normal distribution function, and $d^{\pm}(t,f) = \frac{\log f + \left(r - \alpha \pm \frac{1}{2}\sigma^2\right)t}{\sigma\sqrt{t}}$. Given equation (4) and taking into account the defined optimal surrender boundary and the conditions in (3), the time-reversed free boundary is the solution of $G - \widetilde{B}_1(\tau) = V_1\left(\tau, e^{\kappa\tau}\widetilde{B}_1(\tau)\right)$, and the optimal surrender boundary is obtained by reverting the free-boundary, i.e., $B_1(T - \tau) = \widetilde{B}_1(\tau)$. Furthermore, it results that[1]

$$B_1(T^-) = \widetilde{B}_1(0^+) = \lim_{\tau \to 0^+} \widetilde{B}_1(\tau) = \min\left(1, \frac{r}{\max(\alpha - \kappa, 0)}\right) G. \tag{5}$$

We propose to solve the non-standard integral equations presented above through an approximation of the integrals that is based on the randomized trapezoidal quadrature method presented by Wu [6], coupled with linear interpolation. To detail the algorithm, let N be a positive integer and partition the interval $[0, T]$ into N intervals of equal length $\Delta = \frac{T}{N}$, $0 = t_0 < t_1 < t_2 < \cdots < t_N = T$, where $t_i = i\Delta, i = 0, \ldots, N$. In addition, as argued in Wu [6], let $\{v_i\}_{i=0}^{N-1}$ be a sequence of independent standard uniform random variables and $\tilde{v}_i = 1 - v_i$. The Volterra integral formulations involve two or more integrals of the form

$$I(t, s) = \int_0^t e^{-\lambda(t-s)} \varPhi\left(\pm d^{\pm}\left(t - s, \frac{B(t)}{B(s)}\right)\right) ds,$$

where $t \in [0, T]$ and λ is a constant. By considering a generic time $t_i > 0$, when $i = 1, 2, \ldots N$, following Wu [6] we can approximate such integral as

$$\int_0^{t_i} e^{-\lambda(t_i-s)} \varPhi\left(\pm d^{\pm}\left(t_i - s, \frac{B(t_i)}{B(s)}\right)\right) ds \approx$$
$$\frac{\Delta}{2} \sum_{w=0}^{i-1} \left[e^{-\lambda(t_i - t_w - v_w\Delta)} \varPhi\left(\pm d^{\pm}\left(t_i - t_w - v_w\Delta, \frac{B(t_i)}{B(t_w + v_w\Delta)}\right)\right) + \right.$$
$$\left. + e^{-\lambda(t_i - t_w - \tilde{v}_w\Delta)} \varPhi\left(\pm d^{\pm}\left(t_i - t_w - \tilde{v}_w\Delta, \frac{B(t_i)}{B(t_w + \tilde{v}_w\Delta)}\right)\right) \right].$$

Now, suppose to analyze **Model 1**.[2] Preliminarily, we define the quantity

$$V_1^E\left(\tau, e^{\kappa\tau}\widetilde{B}_1(\tau)\right) = Ge^{-r\tau}\varPhi\left(-d^-\left(\tau, \frac{e^{\kappa\tau}\widetilde{B}_1(\tau)}{G}\right)\right) + $$
$$-\widetilde{B}_1(\tau)e^{-(\alpha-\kappa)\tau}\varPhi\left(-d^+\left(\tau, \frac{e^{\kappa\tau}\widetilde{B}_1(\tau)}{G}\right)\right), \tag{6}$$

---

[1] For **Model 2** and **Model 3**, the integral equation is obtained similarly.

[2] What we present here may be easily adapted to the integral formulations of **Model 2** and **Model 3**.

and, for $w = 0, 1, \ldots, i - 1$, we introduce the quantities $z_w = t_w + \upsilon_w \Delta$ and $\tilde{z}_w = t_w + \tilde{\upsilon}_w \Delta$, so that $z_w, \tilde{z}_w \in [t_w, t_{w+1}]$. By virtue of equation (5), $\widetilde{B}_1(t_0) = \min\left(1, \frac{r}{\max(\alpha - \kappa, 0)}\right) G$, and for each $t_i, i = 1, \ldots, N$, we can write

$$
\begin{aligned}
\tilde{B}_1(t_i) = {} & G - V_1^E(t_i, e^{\kappa t_i} \tilde{B}_1(t_i)) + \\
& - rG \frac{\Delta}{2} \sum_{w=0}^{i-1} \left[e^{-r(t_i - z_w)} \varPhi\left(-d^-\left(t_i - z_w, \frac{e^{\kappa t_i} \widetilde{B}_1(t_i)}{e^{\kappa z_w} \widetilde{B}_1(z_w)} \right) \right) + \right. \\
& \left. + e^{-r(t_i - \tilde{z}_w)} \varPhi\left(-d^-\left(t_i - \tilde{z}_w, \frac{e^{\kappa t_i} \widetilde{B}_1(t_i)}{e^{\kappa \tilde{z}_w} \widetilde{B}_1(\tilde{z}_w)} \right) \right) \right] + \\
& + (\alpha - \kappa) e^{\kappa t_i} \widetilde{B}_1(t_i) \frac{\Delta}{2} \sum_{w=0}^{i-1} \left[e^{-\kappa z_w} e^{-\alpha(t_i - z_w)} \varPhi\left(-d^+\left(t_i - z_w, \frac{e^{\kappa t_i} \widetilde{B}_1(t_i)}{e^{\kappa z_w} \widetilde{B}_1(z_w)} \right) \right) + \right. \\
& \left. + e^{-\kappa \tilde{z}_w} e^{-\alpha(t_i - \tilde{z}_w)} \varPhi\left(-d^+\left(t_i - \tilde{z}_w, \frac{e^{\kappa t_i} \widetilde{B}_1(t_i)}{e^{\kappa \tilde{z}_w} \widetilde{B}_1(\tilde{z}_w)} \right) \right) \right].
\end{aligned}
\tag{7}
$$

Considering simultaneously, for all i, the non-linear equation (7), we obtain a non-linear system of N equations in $3N$ unknowns. In order to avoid this problem, we approximate each $\widetilde{B}_1(z_w)$ and $\widetilde{B}_1(\tilde{z}_w)$ with the interpolated values

$$
\widetilde{B}_1(z_w) \approx F_A\left(t_w, t_{w+1}, \widetilde{B}_1(t_w), \widetilde{B}_1(t_{w+1}), z_w \right),
$$

$$
\widetilde{B}_1(\tilde{z}_w) \approx F_B\left(t_w, t_{w+1}, \widetilde{B}_1(t_w), \widetilde{B}_1(t_{w+1}), \tilde{z}_w \right),
$$

where F_A and F_B are two functions indicating a linear interpolation, thus reducing the $3N$ unknowns to the N unknowns represented by the quantities $\widetilde{B}_1(t_i)$, with $i = 1, \ldots, N$. Solving the arising system of N equation in N unknowns and considering the exact value of $\widetilde{B}_1(t_0)$, we obtain the approximate solution

$$
\left\{ \tilde{B}_1(t_0), \tilde{B}_1(t_1), \tilde{B}_1(t_2), \ldots, \tilde{B}_1(t_N) \right\}.
$$

To present how to extend the previous algorithm for computing the provision α, suppose to fix the initial investment value at level W and consider the first case analyzed in **Model 1**. The provision α making the contract fair may be evaluated applying the proposed algorithm and considering the additional equation $W = VA_1(0, W)$. A system of $N + 1$ equations in $N + 1$ unknowns is solved in order to find both the early boundary exercise and the provision.[3]

3   Hints to the Benchmark Methods and Numerical Results

Some hints of the benchmark methods will be focused on **Model 3**, because the cases reported in **Model 1** and **Model 2** are quite standard.

---

[3] The same considerations may be applied to **Model 2** and **Model 3**.

Concerning the lattice procedure, while a straightforward application of the Cox, Ross, and Rubinstein (CRR) [2] model allows to evaluate the VA contract appearing in **Model 1** and a little effort is needed to take into account the up-and-out barrier in **Model 2**, a novel approach is established for the VAs described in **Model 3**, where the policy embeds a lookback path-dependent guarantee. In detail, we propose an algorithm that detects all the possible maximum values registered on the trajectories reaching a given node of the lattice used to discretize the policyholder account dynamics. The convergence properties of the CRR model coupled with the consideration of all the possible maximum account values for each lattice node make the proposed approach a natural benchmark when lookback features are embedded in the policy guarantees. Clearly, the proposed algorithm is of immediate application for valuing European or American lookback options, thus providing a useful tool not only for actuarial products like VAs but also for pure financial instruments.

Concerning the least squares Monte Carlo method (cf. Longstaff and Schwartz [4]), the main problem concerns the treatment of the lookback feature that we propose to manage through the consideration of basis functions as functions of two variables, i.e., the maximum of the policyholder account values registered before the evaluation time t and the same account value at time t. In detail, we propose to choose for such basis functions monomials of order up to $M = 3$.

The idea of developing two different benchmark methods rises when, running the proposed numerical algorithm to solve the non-standard Volterra integral equations, in some cases we find significant discrepancies with respect to the results reported in the papers of Shen et al. [5] when considering **Model 1**, and of Jeon and Kwak [3] when considering **Model 3**. In addition, we remark that the proposed procedure provides VA evaluations that are completely coherent with the policy values obtained by applying both the lattice and the Monte Carlo benchmark methods. These empirical evidences highlight that the contributions of Shen et al. [5] limitedly to **Model 1**, and Jeon and Kwak [3] limitedly to **Model 3**, present some problems affecting the cases in which their algorithms do not perform properly. On the contrary, the proposed numerical algorithm provides results that are coherent with the developed benchmarks in all the analyzed cases.

References

1. Brennan, M., Schwartz, E.: The pricing of equity-linked life insurance policies with an asset value guarantee. J. Financ. Econ. **3**, 195–213 (1976)
2. Cox, J., Ross, S., Rubinstein, M.: Option pricing: a simplified approach. J. Financ. Econ. **3**, 229–263 (1979)
3. Jeon, J., Kwak, M.: Optimal surrender strategies and valuations of path-dependent guarantees in variable annuities. Insurance Math. Econom. **83**, 93–109 (2018)
4. Longstaff, F., Schwartz, E.: Valuing American options by simulation: a simple least-squares approach. Rev. Financ. Stud. **14**(1), 113–147 (2001)

5. Shen, Y., Sherris, M., Ziveyi, J.: Valuation of guaranteed minimum maturity benefits in variable annuities with surrender options. Insurance Math. Econom. **69**, 127–137 (2016)
6. Wu, Y.: A randomized trapezoidal quadrature. Int. J. Comput. Math. (2021). https://doi.org/10.1080/00207160.2021.1929194

Weather Index-Based Insurance in Agricultural Risk Management

Massimiliano Menzietti and Marco Pirra[✉]

University of Calabria, Rende, CS, Italy
{massimiliano.menzietti,marco.pirra}@unical.it

Abstract. The increase in the frequency and severity of extreme weather events associated with climate changes has relevant impacts for farmers. Weather parametric insurance schemes are a possible option in managing agricultural risks because they refer to objective and immediate data to assess the payouts (as e.g. weather station data) accelerating time of reimbursement and reducing disputes with respect to conventional crop insurance coverages. These insurance can be considered an attractive opportunity, due to their advantages: low costs, no information asymmetry, abundant data, wide spectrum of activities covered, flexibility. The goal of the study is to investigate the potential benefits that the improvement in the design of insurance solutions (and the predictive analytics techniques) could offer in this area. Specifically, with reference to grape production in two Italian regions, we study which meteorological indices are most suitable as predictors of agricultural production and the predictive efficacy of different models: GLM, Neural Network, Random Forest.

Keywords: Agribusiness insurance · Weather events · Parametric insurance · Risk management · Machine learning

1 Introduction

Climate change causes shifts in average weather conditions and an increase in the weather variability due to changes in the frequency and occurrence of extreme events. Weather-related insurance losses have increased in recent years, much faster than non-weather related events. Insurers have historically provided insurance solutions for weather-related losses and have started to pay attention to the implications of climate change for their business. One of the most obvious applications of weather risk management products is in agriculture and farming. Weather impacts many aspects of the agricultural supply and demand chain. Identifying weather risk for an agricultural grower or producer involves three steps: identifying the regions at risk to weather and the weather stations that reflected that risk; identifying the time period during which risk is prevalent; and identifying the weather index that is the best proxy for the weather exposure.

Weather parametric insurance schemes are a possible option in managing weather and climate: they refer to weather station data or grid data (rainfall, temperature) in

order to assess the payouts and these are therefore immediate and with fewer disputes than conventional crop insurance coverages. An index insurance contract pays out according to the value of an 'index', not on losses measured in the field. The index should be a variable that is highly correlated with losses and that cannot be influenced by the insured (e.g.: rainfall, temperature, regional yield, river levels etc.). On one hand, index insurance contracts have many advantages: they overcome most of the supply side problems of traditional Multi-Peril Crop Insurance, are objective and transparent, provide timely payout, reduce administrative costs, facilitate international reinsurance. On the other hand, these contracts also have disadvantages: there is a potential mismatch between losses and payouts (basis risk), they provide single-risk protection, require high inputs during development phase (weather data and networks), require local adaptation (slow the scaling up). The construction of a weather index-based (WII) insurance requires the selection of one or more indices correlated as much as possible with yield. The index selection is therefore of fundamental importance for these products (see [2,5,10,11] and literature therein). Meteorological data can be derived from different types of data sets and, as shown by Parkes et al. [9], these differences create a wide uncertainty in estimated crop yield responses and exposure to variability in growing season weather". With reference to grape production in two Italian regions, the first objective of this work is to identify which of the six most frequently used indicators in the literature is effective as a regressor in yield estimation. We consider the following indexes:

Winkler

$$WI = \sum_{apr}^{oct} T_{a/d} - 10\,^\circ C \tag{1}$$

where $T_{a/d}$ is the daily mean temperature in Celsius grade. The index take into account all days from April 1st (apr) till October 31st (oct).

Huglin

$$HI = K \cdot \sum_{apr}^{oct} \frac{T_{a/d} + T_{m/d}}{2} - 10\,^\circ C \tag{2}$$

where $T_{m/d}$ is the daily maximum temperature in Celsius grade and K is a parameter dependent on the latitude of the location.

Winkler Normalized

$$WIN = H \cdot \sum_{apr}^{oct} 10\,^\circ C < T_{h/d} \leq 35\,^\circ C \tag{3}$$

where H is a normalization factor that takes into account the hours included in the April-October period.

Branas, Bernon and Levadoux

$$BBL = \sum_{m=1}^{M} T_m * R_m \tag{4}$$

where T_m and R_m are the mean monthly temperature and the cumulative monthly rainfall on month m, respectively.

Ribéreau-Gayonand Peynaud

$$RGP = \sum_{d=1}^{DM} max(T_d - 10^\circ C; 0) - R_d \qquad (5)$$

where T_d and R_d are the mean daily temperature and the cumulative daily rainfall on day d, respectively.

Growing Season Precipitation

$$GSP = \sum_{d=1}^{DM} R_d \qquad (6)$$

The first three indicators include only one factor and account for the accumulated heat over the growing period, while the forth and the fifth take into account the accumulated heat over the growing period and rainfall. The last one considers only rainfall. Several studies have analyzed the ability of different weather-yield indices to represent the weather-yield relationship, but this is the first analytical study on the production of wine grapes in Italy that analyzes such a wide range of indices.

A second decisive element for an effective WII product is the accurate representation of the relationship between agricultural yield and meteorological indicator. The literature [10] shows how the assumption of the existence of linear correlations (as in linear regression models) is not always valid and it may be necessary to use alternative models capable of representing more complex forms of dependence. Some authors model the dependence between yield and weather using copulas [3] and [10]. Also machine learning algorithms can be used in predicting the relationship between weather conditions and yield, e.g. [1] uses recursive partitioning to build classification and regression trees (CART) to evaluate the predictive power of the weather variable. Our contribution in this field consists in the adoption of alternative algorithms such as neural networks and random forests (compared to a traditional GLM model) as well as in the application to a crop and a set of indicators not yet tested in the literature.

2   Methodology

Machine learning (ML) approaches are used for yield prediction adopting several mathematical and statistical methods, namely artificial neural networks, fuzzy information networks, decision tree, regression analysis, clustering, principal component analysis, Bayesian belief network, time series analysis, and Markov chain model. The application of these machine learning techniques in agricultural risk management shows more tremendous advantages due to the availability of many data from several resources to obtain hidden knowledge. In this paper we use Neural Networks and Random Forests to predict and check yield to support the construction of an optimal index and the design of an index-based insurance. Deep neural networks (DNN): Deep neural networks techniques belong to the class of representation learning methods with multiple levels of

abstraction, which can learn the underlying representation of data without the need for handcrafted features (see [7] for details). Representation learning is a set of methods that allows a model to be fed with raw data and to automatically discover the relationships between inputs and outputs. Deep neural networks apply a nonlinear function to the output of each hidden layer which makes them highly nonlinear. DNN models are trained with gradient-based optimization methods to minimize the desired error function for the task for which they are used. DNN models have recently been used for crop yield prediction which has shown great success by outperforming other traditional machine learning methods (as discussed in [6] and related works). Random forests (RF): RF include multiple individual trees and uses Breiman's "bagging" idea to ensemble many decision trees into a single but strong model (as described in [4]). It uses the self-help method (i.e., the bootstrap resampling technology) to generate new training sample sets from the original training samples of N by repeatedly selecting random k sets of samples. During the overall selecting process, some samples may be collected more than once. The training sample is used to generate k buffeting the decision or regression trees (CART) for the development of random forests, and then classify the test sample by majority vote decision or use the average as return values. Given the fact that randomness can effectively reduce model variance, random forests, in general, can achieve good generalization ability and low variance resistance without additional pruning. As far as the model accuracy regards, different standard statistical performance evaluations evaluate various conventional predictor model performances. The most widely used statistical measures are coefficient of determination (R^2), root mean square error (RMSE), mean absolute error (MAE), mean squared error (MSE), mean absolute percentage error (MAPE), coefficient of variance (CV) and normalized mean squared error (NMSE).

3   Numerical Application

In this section the models introduced in Sect. 2 are applied to historical datasets of two of the most relevant regions of wine in Italy, Veneto and Piemonte. Data on daily weather conditions for the period 2006–2020, which have been used for the construction of the bioclimatic indicators described in Sect. 1 were downloaded from the Modern-Era Retrospective analysis for Research and Applications, Version 2 database managed by the NASA Goddard Earth Sciences (GES) Data and Information Services Center (DISC) [https://gmao.gsfc.nasa.gov/reanalysis/MERRA-2/]: for the two regions considered in the analysis, Piemonte and Veneto, the list of locations, identified by longitude and latitude, corresponds to the list of provinces. Grape yield data (in q/ha) for the period 2006–2020 are the ones provided by the Italian National Institute of Statistics, a public research organisation, the main producer of official statistics, including Agricultural ones [https://dati.istat.it/].

Table 1 shows the improvements using Neural Networks and Random Forests in terms of reduction of the MSE. The architecture of the Neural Network model has been fitted calibrating the number of neurons and layers that minimizes the error. As for the Random Forest the selection of the seed and the number of trees is the combination that produces the minimum value of the squared distances. Comparing the performance of

Table 1. Performance analysis

Model	MSE	Difference vs MSE.GLM
GLM	1007.4710	
Neural Network	821.8806	−18.42%
Random Forest	544.8351	−45.92%

different learning methods for yield prediction based on all the bioclimatic indicators introduced in Sect. 1 it can be observed that in the GLM model the significance codes indicate that only Winkler, BBL and GSP have an impact on the dependent variable while in the other models all the indicators contribute to the increase of the percentage of variance explained. As expected, contrary to linear regression models, neural networks and tree partitioning allow to capture non-linear and high order interactions. The optimal selection of the forecasting model is used to design the payoff of a weather index-based product with the benefit of reducing basis risk which refers to a mismatch between the index that determines the indemnities and the actual loss experienced and has historically been one of the limitations to the development of this type of solutions for agriculture.

4   Conclusions

In this work we have investigated the benefits that some machine learning techniques can bring in agricultural risk management with respect to the development of alternative solutions. The results indicate that nonlinear models outperform the linear one and are more effective in finding the functional relationship between the yield and the input data. The result is relevant to increase the accuracy in forming the best predictors and constructing optimal indexes. The comparative analysis of the models might support the construction of an optimal index and the design of an WII. Indeed, the choice of the index has a decisive impact on the payoff of the insurance contract. There are several possible directions to improve future research. First, it would be interesting to extend the set of explanatory variables to include other indicators that might be relevant to crop yield, such as variables representing global warming, as well as remote sensing data, as suggested in [8]. It could be worth exploring other techniques for variable selection and index identification, testing the potential improvements in forecasting and reducing basis risk. Future research will focus on building an insurance product based on the models presented here and measuring the effectiveness of insurance coverage in stabilizing farmers' income.

References

1. Biffis, E., Chavez, E.: Satellite data and machine learning for weather risk management and food security. Risk Anal. **37**, 1508–1521 (2017). https://doi.org/10.1111/risa.12847
2. Biffis, E., Chavez, E., Louaas, A., Picard, P.: Parametric insurance and technology adoption in developing countries. Geneva Risk Insur. Rev. **47**, 7–44 (2022). https://doi.org/10.1057/s10713-020-00061-0
3. Bukusheva, R.: Using copulas for rating weather index insurance contracts. J. Appl. Stat. **45**(13), 2328–2356 (2018). https://doi.org/10.1080/02664763.2017.1420146
4. Breiman, L.: Random forests. Mach. Learn. **45**(1), 5–32 (2001). https://doi.org/10.1023/A:1010933404324
5. Bucheli, J., Dalhaus, T., Finger, R.: The optimal drought index for designing weather index insurance. Eur. Rev. Agric. Econ. **48**(3), 573–597 (2021). https://doi.org/10.1093/erae/jbaa014
6. Khaki, S., Wang, L.: Crop yield prediction using deep neural networks. Front. Plant Sci. **10**, 621 (2019). https://doi.org/10.3389/fpls.2019.00621
7. LeCun, Y., Bengio, Y., Hinton, G.: Deep learning. Nature **521**, 436–444 (2015). https://doi.org/10.1038/nature14539
8. Li, H., Porth, L., Tan, K.S., Zhu, W.: Improved index insurance design and yield estimation using a dynamic factor forecasting approach. Insur. Math. Econ. **96**, 208–221 (2020). https://doi.org/10.1016/j.insmatheco.2020.11.003
9. Parkes, B., Higginbottim, T.P., Hufkens, K., Ceballos, F., Kramer, B., Foster, T.: Weather dataset choice introduces uncertainty to estimates of crop yield response to climate variability and change. Environ. Res. Lett. **14**(12), 1–11 (2019). https://doi.org/10.1088/1748-9326/ab5ebb
10. Salgueiro, A.M.: Weather index based insurance as a meteorological risk management alternative in viticulture. Wine Econ. Policy **8**(2), 114–126 (2019). https://doi.org/10.1016/j.wep.2019.07.002
11. Zara, C.: Weather derivatives in the wine industry. Int. J. Wine Bus. Res. **22**(3), 222–237 (2010). https://doi.org/10.1108/17511061011075365

Lattice Cryptalization and Cybersecurity: New Findings in Analyzing Cryptovalues Dynamics

Domenica Stefania Merenda[1] and Massimiliano Ferrara[1,2(✉)]

[1] Department of Law, Economics and Human Sciences and Decisions_Lab, University "Mediterranea" of Reggio Calabria, via dell'Universitá, 25, 89124 Reggio Calabria, Italy
`{stefania.merenda,massimiliano.ferrara}@unirc.it`
[2] Department of Management and Technology, ICRIOS-The Invernizzi Centre for Research in Innovation, Organization, Strategy and Entrepreneurship Bocconi University, via Sarfatti, 25, 20136 Milan, Italy
`massimiliano.ferrara@unibocconi.it`

Abstract. RSA-based scheme are the most popular asymmetric cryptosystems to date, being deployed in billions of security systems and applications. This RSA cryptosystem comprises of two important features that are needed for encryption process known as the public parameter and the modulus. Such constructions are extensively used in financial applications such as Bitcoins, where more than one key is required in order to authorize Bitcoin transactions. Many of the current state-of-the-art schemes are based on the RSA assumptions, however this may be-come insecure in the future, for example due to the possibility of quantum attacks. In this paper we propose a scheme based on recent theoretical advances in lattice-based cryptography.

Over the last years lattice-based cryptography has received much attention as it is a very versatile tool in different fields.

In this work we explore the implementations of cryptographic systems presenting geometric considerations for an application approach to the financial sector, in particular emphasizing the feasibility of the special algebraic structure of ideal lattice on the spread of "cryptocurrencies" (or "virtual currencies"), the best known of which is bitcoin.

Keywords: Encryption · RSA cryptanalysis · Lattice Based Cryptanalysis

1 Introduction

The term Cryptography comes from Kryptós (hidden) and graphía (writing), in fact it is a system designed to make a message unreadable to those who do not have the solution to decode it.

Cryptography consists in the study of methods to encrypt a message in such a way that only those interested can decrypt and read the content of this.

$$\text{Cryptography} := \text{Encryption} \oplus \text{Decryption}$$

Obviously, in parallel with this discipline has developed Cryptoanalysis, ie the study of various attacks on encryption methods.

The purpose of modern cryptography is to find mathematical methods to ensure the main properties of communication: confidentiality, privacy, integrity, authentification of the sender.

In this paper a brief introduction to the currently most used cryptoanalysis model, RSA showed, and then a family of cryptosystems that could potentially resist possible attacks illustrated based on the lattice.

This article has been divided into the following sections. Section 2 describes encryption with a focus on asymmetric encryption. Section 3 explains the RSA algorithm while Sect. 4 presents a new approach using the lattice method. Finally, Sect. 6 provides the conclusions of this study.

2   Encryption: A Brief Introduction

This section briefly present basics yet important materials on Encryption.

2.1   Asymmetric Encryption Algorithms

Cryptography is divided into symmetrical and asymmetrical. The initial idea of encryption started with a symmetrical idea that implied that users used the same key to encrypt and decrypt data. However, the problem of how to distribute keys efficiently arose as an increase in the number of users. Two cryptographers, Diffie and Hellman [1] have introduced Public Key Cryptography (PKC) or also known as asymmetric cryptography that led to a success in the mass use of cryptography.

Asymmetric algorithms use two interdependent keys, one to encrypt data, and the other to decrypt it. One private and one public. If one key used for the encryption operation, the other must be used for the decryption and vice versa. As can be understood from the names themselves, the private key known only by the owner, it must be kept secret and must not be shared with anyone else. The public key shared by all correspondents. The fact of being aware of the public key does not allow to trace the private key in any way (Table 1).

Table 1. Public Key Cryptography scheme.

Public Key Cryptography (PKC)	
Private key (for decryption)	*kept secret and never shared*
Public key (for encryption)	*advertised publicly as part of a "digital certificate" that includes also the name of the owner, the details of the PKC scheme and some other information*

An important feature of PKC is that it uses a one-way function along with its hatch information.

The function is an easy function to calculate but computationally impossible to invert unless you have the information of the trap door that allows the inverse calculation in polynomial time [2]. In 1978, Rivest, Shamir and Adleman invented a cryptographic system called RSA [3] and it implemented globally to provide communications security and information protection.

The algorithm RSA, often used in e-commerce protocols as SSL, is also one of the most used algorithms for the encryption of digital signatures: RSA considered safe due to sufficiently long keys and the use of up-to-date implementations.

The main elements of RSA are the modulo where is a product of two distinct large and balanced prime numbers called p and q, a parameter and which is set as the public key and relatively prime to the Euler function, and a private exponent connected via the relation

$$N\phi(N)d\ ed\ \equiv\ 1\ mod\ (N) \tag{1}$$

3   RSA Cryptosystem

The RSA public key cryptography system [3] can be defined as follows:

$$RSA = (M, C, K, e, d, N, E, D) \tag{2}$$

where:

1. M is the space for unencrypted messages.
2. C is the space for encrypted messages.
3. K is the space of the keys.
4. $N = pq$ is the module with p, q primes, usually the latter have at least 100 digits.
5. $\{e; N\}, \{d; N\} \in K\ with\ e \neq d$ are the encryption and decryption keys, which satisfy respectively

$$ed\ \equiv\ 1\ (mod\ \phi(N)) \tag{3}$$

   where $\phi(N) = (p - 1)(q - 1)$ is the Euler function, defined by $\phi(N) = \#(Z_N^*)$, which is therefore the number of invertible elements in the multiplicative group Z_N^*

6. E is the encryption function:

$$E_{e,N} : M \rightarrow C \tag{4}$$

   which operates as follows:

$$c\ \equiv\ m^e(mod\ N) \tag{5}$$

7. D is the decryption function:

$$D_{d,N} : C \rightarrow M \tag{6}$$

   which operates as follows:

$$m \equiv c^d \equiv \left(m^e\right)^d (mod\ N) \tag{7}$$

4   Lattice Based Cryptanalysis

Lattice-based cryptographic [4] constructions are prime candidates for post-quantum public key cryptography.

Lattice-based cryptography uses two-dimensional algebraic constructs known as "lattices", resistant to quantum computational schemes.

A lattice is an infinite grid of points; the computational problem on which lattice-based technology based is the "Shortest Vector Problem", which requires identifying the point in the grid that is closest to a fixed central point in space, called the origin. This is an easy problem to solve in a two-dimensional grid, but as the number of dimensions increases, even a quantum computer can no longer solve the problem efficiently.
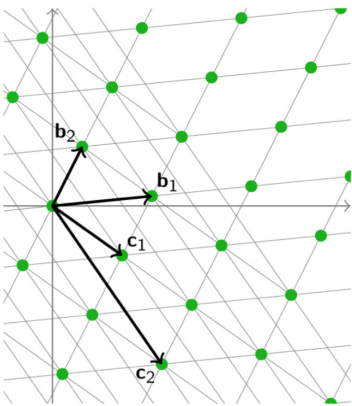
Fig. 1. Integer lattice

In SVP we try to find, given a basis of a geometric lattice, a non-zero vector of minimum norm belonging to the lattice; an approximation of it, very useful in cryptography is SVP that is to find the shortest non-zero vector within an approximation. The two problems are both considered computationally very difficult.

A lattice is an additive subgroup of R^n, in particular any subgroup of $\mathbb{Z}^n$ called an integer lattice.

An integer lattice is the set of all integer linear combinations of a set of linearly independent vectors (b_i)

$$L = \left\{ \sum_{i=1}^{n} n_i b_i | n_i \in \mathbb{Z} \right\} \tag{8}$$

The set of these linearly independent vectors called the lattice base (Fig. 1).

All the bases have the same number of elements, which identifies the size of the latter, indicated with $dim(L)$.

Minimum distance:

$$\lambda_1 = min_{x,y \in L, x \neq y} \|x - y\| \tag{9}$$
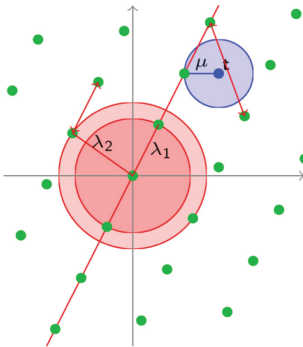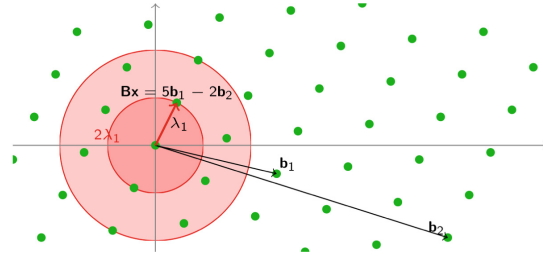
$$min_{x \in L, x \neq 0} \|x\| \tag{10}$$

Fig. 2. Minimum distance and distance function

Distance function (Fig. 2):

$$(t; L) = min_{x \in L,} \|t - x\| \tag{11}$$

Definition Shortest Vector Problem (SVP). Given a basis $B \in Z^{nxn}$ *of L*, find the non-zero vector of the lattice Bx ($x \in Z^n = \{0\}$) of length at most $\|Bx\| \leq \lambda_1$ (Fig. 3).

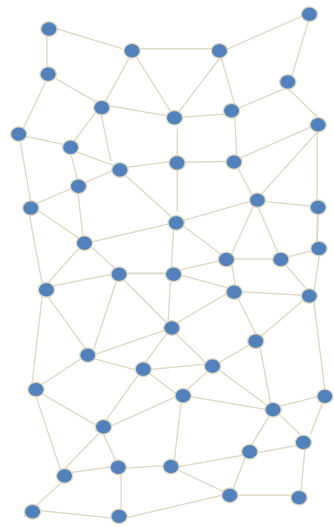**Fig. 3.** Shortest Vector Problem

5 Application to Cryptovalues Dynamics

New technologies, favored by advances in cryptography - that is the application of methods that serve to make a message understandable/intelligible only to people authorized to read it - and by the evolution of the internet, are causing a radical change in the global economy, with particular reference to the financial sector, in terms of the methods of exchange of goods, services and any financial activity.

One of the most significant applications of digital technology to the financial sec-tor is the birth and spread of "cryptocurrencies" (or "virtual currencies"), the best known of which is bitcoin. A distributed ledger or blockchain is an open and distributed ledger that can safely store transactions between two parties, verifiable and permanent. The participants in the system are defined as 'nodes' and are connected to each other in a distributed manner. In essence, it is an ever-growing list of records, called blocks, which are linked together and secured through the use of cryptography. The data

Fig. 4. Distributed network

in a block are by their nature immutable (they cannot be retroactively altered without all subsequent blocks being modified; to do this, given the nature of the protocol and the validation scheme, the consent of the majority of the network would be required).

The distributed nature and the cooperative model make the validation process particularly safe and stable, even though it has to resort to non-negligible times and costs, largely referable to the price of the electricity necessary to validate the blocks (this in the case of the Blockchain of bitcoin) and the computational capacity necessary to solve complex algorithmic calculations (an activity commonly referred to as 'mining'). In our idea the new aspects introduced in this ongoing research could pro-mote these geometrical structure to better studying this dynamics in terms of distributed chain or graph (Fig. 4).

6 Conclusion

In this paper, the theme of Cryptography has been addressed, considering first of all the RSA Method that has been found computationally complex but currently the most used one, a valid alternative can be the Lattice Based Cryptanalysis.

Because cryptography is a computationally difficult problem, the best post-quantum algorithms to solve it have complexities of exponential order compared to the size of the lattice.

In fact through the new reticular approach the calculations are performed easily, with low computational cost. In this way it would be possible to carry out encryption operations also using devices with a reduced computing power and therefore cheaper.

References

1. Diffie, W., Hellman, M.: New directions in cryptography. IEEE Trans. Inf. Theory **22**(6), 644–654 (1976)
2. Hoffstein, J., Pipher, J., Silverman, J.H., Silverman, J.H.: An Introduction to Mathematical Cryptography, vol. 1. Springer, New York (2008). https://doi.org/10.1007/978-1-4939-1711-2
3. Rivest, R., Shamir, A., Adleman, L.: A Method for obtaining digital signatures and public-key cryptosystems. Commun. ACM **21**, 120–126 (1978)
4. Micciancio, D., Regev, O.: Lattice-based cryptography. In: Bernstein, D.J., Buchmann, J., Dahmen, E. (eds.) Post-quantum Cryptography, pp. 147–191. Springer, Heidelberg (2009). https://doi.org/10.1007/978-3-540-88702-7_5

The Impact of Newspaper-Based Uncertainty Indices on Tail Risk Forecasting

Antonio Naimoli[(✉)] and Giuseppe Storti

Dipartimento di Scienze Economiche e Statistiche (DISES), Università degli Studi di Salerno, Via Giovanni Paolo II, 132, 84084 Fisciano, Salerno, Italy
{anaimoli,storti}@unisa.it

Abstract. This paper investigates the impact of economic policy uncertainty on tail risk forecasting. We refer to the Realized Exponential GARCH model as it can directly incorporate information from realized volatility measures and newspaper-based uncertainty indices. An application to the prediction of daily Value-at-Risk and Expected Shortfall for the S&P 500 provides evidence that combining realized volatility and uncertainty measures can lead to significant accuracy gains in forecasting tail risk.

Keywords: Economic policy uncertainty · Uncertainty indices · Realized GARCH · Tail risk forecasting

1 Introduction

The growing importance of risk management in banking and industrial sectors has led to Value at Risk (VaR) becoming a widely used benchmark for measuring risk. However, VaR fails to meet the requirements of a coherent risk metric [1]. The Basel Committee on Banking Supervision has moved toward using the Expected Shortfall (ES), as a complement to VaR, to calculate minimum capital requirements for market risk.

The volatility forecasting literature provides extensive evidence that the use of realized volatility (RV) measures can be beneficial in improving the accuracy of volatility forecasts, with several papers also supporting the usefulness of including realized measures in risk forecasting models [2,3].

The question we attempt to answer is whether or not economic policy uncertainty matters in forecasting tail risk. Recently, [4] developed an economic policy uncertainty (EPU) index that relies on newspaper coverage to measure policy-related economic uncertainty. Along the same lines, [5] introduced the Equity Market Volatility (EMV) Index that moves with VIX and with the RV of returns on the S&P 500. Currently, social media provide one of the main sources of information. Given the popularity of Twitter, Twitter-based Economic Uncertainty indices have been proposed by [6].

The existing studies have mainly focused on analyzing the link between economic policy uncertainty and financial markets within a GARCH-MIDAS [7] framework, using monthly-scale uncertainty indices [8–10]. Our paper presents two main contributions compared to previous research. First, we examine the impact that recently developed newspaper-based uncertainty indices have on tail risk forecasting, by using the EPU, EMV and four Twitter-based economic uncertainty indicators observed on a daily basis. Second, we employ the Realized Exponential GARCH (REGARCH) [11] to account for the contribution of a set of newspaper-based indices on volatility dynamics.

An empirical application on the Standard & Poor's 500 index reveals that incorporating newspaper-based uncertainty indices as additional predictor variables in REGARCH specifications, based only on the use of the 5-min RV, leads to significant accuracy gains in forecasting tail risk, for both VaR and ES.

The remainder of this paper is organized as follows. Section 2 briefly describes the REGARCH model. Section 3 focuses on the estimation of the models under analysis. Section 4 discusses the data, while the empirical results are shown in Sect. 5. Finally, Sect. 6 concludes.

2   The Realized Exponential GARCH

The Realized Exponential GARCH (REGARCH) model by [11] represents a flexible framework for jointly modelling returns and realized volatility measures. The REGARCH differs from standard GARCH in several respects: (i) in the volatility dynamics, the squared returns are replaced by a more efficient proxy, i.e., a realized volatility measure; (ii) it relates the realized measure to the latent volatility through a measurement equation, by also including an asymmetric response to shocks; (iii) the model is very flexible and dynamically complete, allowing the generation of multi-step forecasts.

Let $\{r_t\}$ be a time series of daily log-returns and $\boldsymbol{x}_t = (x_{1,t}, \cdots, x_{K,t})'$ be a vector of K realized and uncertainty measures; the REGARCH is defined as

$$r_t = \sqrt{h_t}\, z_t \tag{1}$$

$$log(h_t) = \omega + \beta \, log(h_{t-1}) + \delta(z_{t-1}) + \boldsymbol{\gamma}' \, \boldsymbol{u}_{t-1} \tag{2}$$

$$log(x_{j,t}) = \xi_j + \varphi_j \, log(h_t) + \tau_j(z_t) + u_{j,t} \qquad j = 1, ..., K \tag{3}$$

where $\boldsymbol{u}_t = (u_{1,t}, ..., u_{K,t})'$ and $\boldsymbol{\gamma} = (\gamma_1, ..., \gamma_K)'$. Also, $h_t = var(r_t|\mathcal{F}_{t-1})$ is the conditional variance and $\mathcal{F}_{t-1}$ is the information set at time $t - 1$. Finally, the functions $\tau(z_t) = \tau_1 z_t + \tau_2(z_t^2 - 1)$ and $\delta(z_t) = \delta_1 z_t + \delta_2(z_t^2 - 1)$ are used to model leverage-type effects, with z_t and $\mathbf{u}_t$ mutually independent.

3   Estimation

The models are estimated by the maximum likelihood (ML) approach, assuming a standardized Student-t distribution for innovations $z_t \overset{iid}{\sim} t(0, 1, \nu)$, since, as

confirmed by empirical evidence, the distribution of daily log-returns is typically peaked and characterized by fat tails. Also, following [11], we assume that $\mathbf{u}_t \overset{iid}{\sim} \mathcal{N}_K(\mathbf{0}, \Sigma)$, where $\mathcal{N}_K(\mathbf{0}, \Sigma)$ denotes a K-variate Normal distribution with mean $\mathbf{0}$ and variance-covariance matrix Σ. Therefore, the log-likelihood function for REGARCH models is given by

$$\mathcal{L}(r, u; \boldsymbol{\theta}) = -\frac{1}{2} \sum_{t=1}^{T} -\mathcal{A}(\nu) + log(h_t) + (1+\nu) log\left(1 + \frac{r_t^2}{h_t(\nu-2)}\right) \quad (4)$$

$$-\frac{1}{2} \sum_{t=1}^{T} K log(2\pi) + log\left(|\Sigma|\right) + \mathbf{u}_t' \, \Sigma^{-1} \, \mathbf{u}_t \,, \quad (5)$$

where $\mathcal{A}(\nu) = log\left[\Gamma\left(\frac{\nu+1}{2}\right)\right] - log\left[\Gamma\left(\frac{\nu}{2}\right)\right] - \frac{1}{2} log[\pi(\nu-2)]$.

Therefore, the overall log-likelihood incorporates both the contribution of the realized measures from (5) and the returns from (4).

4   The Data

The data set used in this paper consists of two categories: (i) daily observations of open-to-close log-returns and 5-min RV of the Standard & Poor's 500 index, publicly available at https://realized.oxford-man.ox.ac.uk; (ii) daily indicators based on text-counts of newspaper articles that include several keywords related to the U.S. economy or stock market volatility. In particular, to investigate the role of policy uncertainty, we refer to the daily Economic Policy Uncertainty (EPU) index. The EPU is calculated by counting the occurrence of words related to uncertainty in leading U.S. newspapers that contain a trio of terms related to the economy (E), policy (P), and uncertainty (U) [4].

On the other hand, to account for the public attention on future market volatility we consider the newspaper-based Equity Market Volatility (EMV) Index [5]. This Index moves with the CBOE Volatility Index (VIX) and with the realized volatility of returns on the S&P 500. The EMV tracker is calculated through an analysis of newspaper articles containing terms related to stock market uncertainty, i.e., by counting the following keywords occurring in eleven major U.S. newspapers: E {economic, economy, financial}; M {"stock market", equity, equities, "Standard and Poors" (and variants)} and V {volatility, volatile, uncertain, uncertainty, risk, risky}.

We also consider Twitter-derived measures of economic uncertainty. Twitter Economic Uncertainty (TEU) indicators are based on counts of tweets about the "economy" and "uncertainty" [6]. We examine four variants of daily TEU indices in our analysis. The first, TEU-ENG, consists of the total number of daily English-language tweets that contain both uncertainty and economic terms. The second, TEU-USA, is constructed by isolating the number of these tweets that come from users in the U.S. using a geotagged-based classifier. The TEU-WGT is a variant of TEU-USA as it weights each tweet by (1+log(1+# of retweets)).

Finally, in order to track changes in Twitter usage intensity over time, the TEU-SCA index scales the number of tweets each day by the number of tweets on that day containing the word "have". The data and and further details on these indices can be found at https://www.policyuncertainty.com/index.html.

Based on the availability of daily Twitter uncertainty measures, the sampling period is from 01 June 2011 to 01 June 2021, for a total of 2507 observations. The uncertainty indices were adjusted by excluding non-trading days (Saturday and Sunday), in order to have the same observations as the financial variables. Also, the uncertainty indices were rescaled to have the same range as the 5-min RV. To provide more insight on the characteristics of the variables, Table 1 reports the main descriptive statistics of log-returns, 5-min RV, and all daily indicators based on key words related to economic and policy uncertainty.

Table 1. Summary statistics for the full sample period 01/06/2011–01/06/2021

	Mean	Sd	Median	Min	Max	Skew	Kurt
r_t	0.00	0.01	0.00	−0.07	0.05	−0.55	6.68
RV5†	0.00	0.00	0.00	0.00	0.00	10.95	156.78
EPU‡	0.12	0.09	0.09	0.00	0.81	2.47	8.41
EMV‡	0.06	0.08	0.03	0.00	0.94	4.04	24.08
TEU.ENG‡	0.10	0.09	0.08	0.01	1.48	3.41	26.75
TEU.USA‡	0.10	0.11	0.07	0.00	1.56	3.86	25.11
TEU.WGT‡	0.10	0.12	0.07	0.00	1.66	4.13	27.12
TEU.SCA‡	0.10	0.10	0.07	0.00	2.04	6.03	77.95

Key to table. r_t: daily open-to-close log-returns; RV5: daily 5-min RV; EPU: Economic Policy Uncertainty Index; EMV: Equity Market Volatility Index; TEU.∗: Twitter-based Economic Uncertainty Index.∗, with $* \in$ (ENG,USA,WGT,SCA); $† = \times 100$; $‡ = /1000$.

5 Empirical Findings

We compare several models obtained by considering different configurations of the REGARCH. In particular, the standard REGARCH based on the 5-min RV, REG(RV5), is compared with specifications combining the 5-min RV with uncertainty measures: REG(RV5, EPU); REG(RV5, EMV); REG(RV5, TEU.ENG); REG(RV5, TEU.USA); REG(RV5, TEU.WGT) and REG(RV5, TEU.SCA).

The analyzed specifications are used to generate one-step-ahead forecasts of VaR and ES, where model parameters are recursively estimated daily via ML with a rolling window of 1500 observations. The out-of-sample forecasting performance of the models is evaluated by considering different loss functions and three distinct α-risk levels: 1%, 2.5% and 5%. The significance of the differences in the forecasting performance of the models is tested through the Model Confidence Set (MCS) [12], considering the confidence levels of 75% and 90%.

Specifically, to rank models based on their ability to accurately forecast VaR, we rely on the Quantile Loss (QL) function [13]

$$QL_t(\alpha) = (\alpha - l_t)(r_t - VaR_t(\alpha)), \tag{6}$$

where $l_t = \mathcal{I}_{(r_t < VaR_t(\alpha))}$. This loss function is well known to be a strictly consistent scoring rule for VaR prediction.

Furthermore, although several strictly consistent scoring rules have been proposed for the pair (VaR, ES), to evaluate the ability of the proposed models to jointly forecast VaR and ES, we consider the zero degree homogeneous loss function [14]:

$$FZ_t^0 = \frac{1}{\alpha ES_t(\alpha)} l_t (r_t - VaR_t(\alpha)) + \frac{VaR_t(\alpha)}{ES_t(\alpha)} + log(-ES_t(\alpha)) - 1, \tag{7}$$

where VaR and ES are assumed to be strictly negative and $ES_t(\alpha) \leq VaR_t(\alpha) < 0$. As usual, models characterized by lower average values of QL and FZ_t^0 are preferred.

Table 2 reports the ratios of QL and FZ^0 loss functions of all models to those of the REG(RV5); values smaller than 100 denote improvements over the benchmark. Overall, it emerges that: (i) the REGARCH specification combining the 5-min RV and EPU minimizes the considered losses in five out of six cases, with REG(RV5, EMV) completing the picture; (ii) the REGARCHs using EPU and EMV information are the only models always entering the 75% MCS; (iii) the Twitter uncertainty measures appear to be less influential in forecasting VaR and ES than EPU and EMV; (iv) the FZ^0 is more discriminating than the QL; (v) according to the MCS, the main benefits in combining economic and financial uncertainty indices with the 5-min RV occur at the 1% risk level.

Table 2. VaR and ES loss functions comparison: QL and FZ^0.

| | $\alpha = 0.01$ | | $\alpha = 0.025$ | | $\alpha = 0.05$ | |
	QL	FZ^0	QL	FZ^0	QL	FZ^0
REG(RV5)	100.00	100.00	100.00	100.00	100.00	100.00
REG(RV5, EPU)	**96.40**	**97.98**	99.23	**99.27**	99.51	**99.56**
REG(RV5, EMV)	96.89	98.23	**99.05**	99.32	99.68	99.64
REG(RV5, TEU.ENG)	101.98	100.81	100.52	100.34	100.57	100.29
REG(RV5, TEU.USA)	97.47	98.79	99.67	99.59	99.77	99.76
REG(RV5, TEU.WGT)	98.12	99.14	100.03	99.74	99.96	99.84
REG(RV5, TEU.SCA)	99.64	99.93	100.15	99.99	100.03	100.00

The table shows the ratios of the loss functions QL and FZ^0 at the risk levels of $\alpha = \{0.01; 0.025; 0.05\}$ of all models to those of the REGARCH (RV5) (benchmark model). Values smaller than 100 denote improvements over the benchmark. The best model is given in bold, while models $\in 75\%$ and $\in 90\%$ MCS are shaded in gray and light-gray, respectively.

6   Conclusion

In this paper, we examined the forecasting power of newspaper-based uncertainty indices on daily returns of the S&P 500 index. Using the REGARCH model, we find evidence that combining the information contained in the RV and the EPU and EMV uncertainty indices yields significant accuracy gains in predicting VaR and ES at different risk levels. On the other hand, the benefits of incorporating Twitter-based uncertainty indices are less evident, if any. A natural extension of our work would be to combine multiple daily uncertainty indicators based on text-counts and to use alternative proxies to measure economic policy uncertainty and market uncertainty.

References

1. Artzner, P., Delbaen, F., Eber, J.M., Heath, D.: Coherent measures of risk. Math. Finance **9**(3), 203–228 (1999)
2. Gerlach, R., Naimoli, A., Storti, G.: Time-varying parameters realized GARCH models for tracking attenuation bias in volatility dynamics. Quant. Finance **20**(11), 1849–1878 (2020)
3. Naimoli, A., Gerlach, R., Storti, G.: Improving the accuracy of tail risk forecasting models by combining several realized volatility estimators. Econ. Model. **107**, 105701 (2022)
4. Baker, S.R., Bloom, N., Davis, S.J.: Measuring economic policy uncertainty. Q. J. Econ. **131**(4), 1593–1636 (2016)
5. Baker, S.R., Bloom, N., Davis, S.J., Kost, K.J.: Policy news and stock market volatility. Technical report, National Bureau of Economic Research (2019)
6. Baker, S.R., Bloom, N., Davis, S., Renault, T.: Twitter-derived measures of economic uncertainty (2021)
7. Engle, R.F., Ghysels, E., Sohn, B.: Stock market volatility and macroeconomic fundamentals. Rev. Econ. Stat. **95**(3), 776–797 (2013)
8. Zhu, S., Liu, Q., Wang, Y., Wei, Y., Wei, G.: Which fear index matters for predicting US stock market volatilities: text-counts or option based measurement? Phys. A **536**, 122567 (2019)
9. Yu, X., Huang, Y.: The impact of economic policy uncertainty on stock volatility: evidence from GARCH-MIDAS approach. Phys. A **570**, 125794 (2021)
10. Lang, Q., Lu, X., Ma, F., Huang, D.: Oil futures volatility predictability: evidence based on Twitter-based uncertainty. Finance Res. Lett. 102536 (2021)
11. Hansen, P.R., Huang, Z.: Exponential GARCH modeling with realized measures of volatility. J. Bus. Econ. Stat. **34**(2), 269–287 (2016)
12. Hansen, P.R., Lunde, A., Nason, J.M.: The model confidence set. Econometrica **79**(2), 453–497 (2011)
13. González-Rivera, G., Lee, T.H., Mishra, S.: Forecasting volatility: a reality check based on option pricing, utility function, value-at-risk, and predictive likelihood. Int. J. Forecast. **20**(4), 629–645 (2004)
14. Patton, A.J., Ziegel, J.F., Chen, R.: Dynamic semiparametric models for expected shortfall (and value-at-risk). J. Econom. **211**(2), 388–413 (2019)

The Impact of Collateralization on Longevity Swap Transactions

Selin Özen[1(✉)] and Şule Şahin[2]

[1] Ankara University, Ankara, Turkey
`ozens@ankara.edu.tr`
[2] University of Liverpool, Liverpool L69 3BX, UK
`Sule.Sahin@liverpool.ac.uk`

Abstract. Index-based longevity swaps provide many advantages over the other hedging instruments to life insurance companies and pension plans. Insurers and pension plan providers can transfer their longevity exposures to the capital markets at lower costs by using these securities. Hence, significant growth has been seen in longevity swap transactions in the longevity-linked securities and derivatives markets since 2008. However, since longevity-linked instruments are traded OTC, each involved party is exposed to the counterparty default risk. Therefore, regulators have emphasised the role of credit risk mitigation tools such as collateralization for the improvement of swap contracts' credit quality. In this paper, our aim is to construct a hedging strategy for longevity risk by using collateral. As the first step, the Lee-Carter with renewal process and exponential jumps model proposed by Özen and Şahin [7] and the Lee-Carter model without jumps are used to project the future mortality rates and to price the index-based longevity swaps. Additionally, re-hypothecation is allowed for the parties of the swap to increase the benefits of the collateralization. As a result, for both mortality models, insurers and pension plan providers obtain more effective risk reduction levels with the inclusion of the collateral. However, the Lee-Carter model with renewal process and exponential jump model provides more risk reduction.

Keywords: Longevity risk · Collateral · Hedging strategy · Longevity swaps

1 Introduction

Longevity risk has become a major issue along with the development in medicine, lifestyle, and health care for insurance companies. It is a crucial financial concern for both pension plans and life insurers since they may have to make more payments than expected.

Generally, three approaches are used for financial institutions to manage and mitigate their longevity risk. Index-based hedging solutions, which involve longevity-linked derivatives or securities, provide many advantages over other hedging solutions, such as faster execution, liquidity potential and lower costs

[6]. Due to offering significant capital savings and effective risk management, index-based longevity instruments attract increased interest from within and outside of the worlds of insurance and pensions [2].

The first step of the longevity risk assessment and thus the valuation of index-based financial products is mortality modelling. The choice of the appropriate model is crucial to quantify the risk and provide a foundation for pricing and reserving. In this paper, a different approach proposed by Özen and Şahin [7] is used for mortality modelling. Their approach is a specification of the Lee-Carter model with jump effects. In the model, the history of catastrophic events is included in the jump frequency modelling process by using the renewal process. It is important to incorporate the mortality jumps to estimate the uncertainty surrounding a central mortality projection. Incorporating the jumps into the modelling process allows us to estimate the probability of catastrophic mortality deterioration when pricing securities for hedging extreme mortality risk [8].

While trading mortality, the counterparty default risk is another risk factor that should be considered. Since longevity-linked instruments are traded in OTC, each involved party will be exposed to the counterparty default risk which can be defined as the risk that the counterparties might not meet their obligations regarding swap payments [2]. Historical experiences show that counterparty default risk often leads to significant losses, and this risk has become particularly apparent following the global financial crisis. Therefore, regulators have emphasized the role of the credit risk mitigation tools such as clearing and collateralization to improve the swap contracts' credit quality [1]. The International Swap and Derivatives Association [3] indicates that bilateral collateral posting is the most credit-enhancing way for counterparty risk and hence it provides an effective risk reduction.

Collateralization is a hedging strategy that includes exchanging the assets between two parties to reduce the counterparty default risk. The main idea is quite simple: securities, financial instruments, or cash are passed to the counterparty to provide hedging for the default exposure. Collateralization has many benefits, both privately and socially. For example, collateralization decreases losses conditional on default. Whichever party received the collateral will keep the collateral, meaning that the maximal loss will be total exposure with the subtraction of any collateral that was posted [4].

The aim of this paper is to build an effective hedging framework for longevity risk under collateralization. To illustrate the idea, we construct a hedging strategy for a hypothetical pension plan by using collateral. Two different mortality models are used to quantify jump effects on the hedge. The Lee-Carter model with exponential jumps and renewal process model, which is proposed by Özen and Şahin [7], and the original Lee-Carter model without jumps are chosen to model mortality. Moreover, different amounts of collateral are considered here to show the effects of collateralization on the hedge. It is found that posting collateral decreases the longevity risk and default risk exposure.

The remainder of this paper is as follows. Section 2 presents the mortality models. In Sect. 3, a hedging framework under collateralization is described and

hedging results of the hypothetical pension plan are given. Section 4 concludes the paper.

2   Mortality Models

2.1   The Lee-Carter Model

This model describes the logarithm of central death rates in the following way:

$$\ln(m_{x,t}) = a_x + b_x k_t + e_{x,t}. \tag{1}$$

Here, a_x is an age-specific component; the time-varying k_t parameter summarizes the general mortality level; and the other age-specific parameter, b_x, explains how slowly or how rapidly mortality varies for each age as the mortality index changes. $e_{x,t}$ is an error term reflecting age-specific influences that are not captured by the model.

The Lee-Carter model is an overparameterized model. For obtaining a unique solution, a_x is taken to be the arithmetic mean of $\ln(m_{x,t})$ over time and the sums of b_x and k_t are respectively normalized to unity and zero. A two-stage estimation procedure is used for the parameters. We can apply the singular value decomposition (SVD) method to the matrix of $ln(m_{x,t}) - a_x$ for obtaining the estimates of b_x and k_t as the first step. As the second step, the time-varying terms are iteratively reestimated given the values of a_x and b_x, which are obtained in the first step. The autoregressive integrated moving average (ARIMA) model is applied for the modeling of the dynamics of k_t to project the future mortality rates in the original model [5].

2.2   The Lee-Carter Model with Exponential Transitory Jumps and Renewal Process

The model used here is a Lee-Carter model with exponential transitory jumps and renewal process. Our proposed model is given as follows:

$$log(m_{x,t}) = a_x + b_x k_t, \tag{2}$$

$$k_t = k_0 + (\mu - \frac{1}{2}\sigma^2)t + \sigma W(t) + \sum_{i=1}^{N(t)} Y_i. \tag{3}$$

Here, $m_{x,t}, a_x, b_x$ and k_t are the same as in the original Lee-Carter model. $W(t)$ is standard Brownian motion, and Y_i denotes a sequence of iid exponential random variables representing the size of the jumps with the η parameter. Finally $N(t)$ denotes the renewal process that has a lognormal distribution with α and β parameters.

We estimate the model's parameters using the MLE method. Afterwards, Eq. (3) is used to calibrate the time-varying mortality index. We need to find the

density function of the independent one-period increments, $\Delta k_i = r_i = k_i - k_{i-1}$, to estimate the parameters of the calibrated model.

Let $D = \{k_0, k_1, ..., k_T\}$ represents the mortality time series at times of $t = 1, 2, ..., T$, which have equal spacing. The one-period increments are iid. Unconditional density for the one-period increment $f(r)$ is given as follows:

$$f(r_i) = P(0)f(r_i|0) + \sum_{n=1}^{N(t)} P(n)f(r_i|n). \tag{4}$$

$P(0) = 1 - F(t)$, $P(n) = \int_0^t P_{n-1}(t-s)f(s)ds$ are the probability of no jump and n jumps occur in the renewal process, where $F(t)$ and $f(t)$ are the distribution and density functions of inter-arrival times between two jumps. $f(r_i|0)$, $f(r_i|n)$ are conditional densities for a one-period increment; more specifically, they are conditional on the given numbers of jumps and expressed as:

$$f(r_i|0) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(r-\mu+0.5\sigma^2)^2}{2\sigma^2}}$$

$$f(r_i|n) = \frac{\eta^n}{(n-1)!\sqrt{2\pi}\sigma} \int_0^\infty X^{n-1} e^{-\eta X - \frac{1}{2\sigma^2}(r-X-\mu+0.5\sigma^2)^2} dx$$

Then, we can write the log-likelihood of the model as follows:

$$L(D; \mu, \sigma, \eta, \alpha, \beta) = \sum_{i=1}^T \ln(f(r_i)).$$

3   Hedging with Collateral

In this section, we consider a hypothetical pension plan to construct a hedging framework for longevity risk with collateral by using a 10-year index-based longevity swap. The pension plan members are assumed to have underlying mortality rates that are the same as England & Wales male population for a sample period from 1961 to 2016. Suppose that all members of the pension plan are aged 65 and pay £1 per year on survival from age 66 to 90. The risk-free interest rate r is used during the whole period as in Biffis et al. [2]. The current date is taken as the start of the calendar year 2016.

We assume that S_t denotes the longevity swap value at time t. As a floating-leg receiver, the present value of the longevity swap's future cash inflows without counterparty default risk, S_t, can be written as:

$$S_t = \sum_{t=1}^T \left(_t p_{65} -_t p_{65}^{forward}\right)(1+r)^{-t} \tag{5}$$

where $S_0 = 0$, $_t p_{65}$ is random future survivor index, and $_t p_{65}^{forward}$ is the forward survivor index which is specified at the beginning of the swap. After specifying the longevity swap payment structure, the counterparty default risk is

considered, and hence collateralization for hedging this risk. Using the notation $X^+ := max(X, 0)$ and $X^- := max(-X, 0)$, all possible liabilities under collateralization and rehypothecation will be obtained here. It should be noted that the cost of posting collateral is ignored in this study. Furthermore, we assume that $(C_t)_{t \geq 0}$ is the collateral process and it indicates the quantity of cash, C_t, that will be posted at each time t before default as in the work of Biffis et al. [2].

Under the collateralization rules the liabilities of insurer can be expressed as follows:

1. On the insurer's default event, $t = \tau = \tau_i$, the counterparty posts collateral and the longevity swap value is paid to the insurer if $S_{t^-} \geq 0$. The exposure of counterparty is reduced by collateral and the swap value is paid to the insurer. The posted collateral rehypothecated by the insurer and the remaining amount is paid to the counterparty. Then the insurer has:

$$1_{\{\tau = \tau_i\}} 1_{\{S_{t^-} \geq 0\}} \left((S_t - C_t)^+ + REC'_i (S_t - C_t)^- \right)$$

2. On the insurer's default event, $t = \tau = \tau_i$, if $S_{t^-} < 0$, then the insurer posts collateral. The counterparty uses the collateral to reduce exposure and only a fraction of the swap is exchanged. The remaining collateral returned to the investor (if any). If the remaining collateral is not enough, then insurer would have a loss for remaining value. Then the insurer has:

$$1_{\{\tau = \tau_i\}} 1_{\{S_{t^-} < 0\}} \left(- REC_i (S_t - C_t)^- - (S_t - C_t)^+ \right)$$

3. On the counterparty's default event, $t = \tau = \tau_c$, the insurer pays the swap value to the counterparty if $S_{t^-} < 0$ and the collateral is posted by the insurer. Then the insurer uses the collateral to reduce exposure. The collateral rehypothecated by the counterparty and remaining collateral is returned to the investor (if any). Then the insurer has:

$$1_{\{\tau = \tau_c\}} 1_{\{S_{t^-} < 0\}} \left(- (S_t - C_t)^- + REC'_c (S_t - C_t)^+ \right)$$

4. On the counterparty's default event, $t = \tau = \tau_c$, if $S_{t^-} \geq 0$, the counterparty posts some collateral which is used by insurer to reduce its exposure. Only a fraction of the swap is exchanged and the remaining collateral is returned to the counterparty (if any). Then the insurer has:

$$1_{\{\tau = \tau_c\}} 1_{\{S_{t^-} \geq 0\}} \left(REC_c (S_t - C_t)^+ + (S_t - C_t)^- \right)$$

where REC_i (REC_c) denotes the recovery fraction of the swap value that will be received by the counterparty (insurer) upon the defaulting of the insurer (counterparty), REC'_i (REC'_c) is the recovery fraction of collateral rehypothecated by the defaulted insurer (counterparty), and S_{t^-} denotes the swap value before the default [2].

In this paper, we calculate recovery fraction ratio for different collateral amounts in the hedged portfolio. We assume that both parties are subject to the same default probabilities, and we consider the basic case in which

$REC = REC_i = REC_c = REC_i' = REC_c'$ to examine the impact of collateral on the hedge for longevity risk. The aggregated cash flows are rearranged to determine liabilities arising from the swap payments for insurer. Using the fact that the market value of the swap is zero at initiation, the swap is discounted for ten years and the recovery rates are obtained for different collateral amounts. The results show that the recovery rates increase while the posted amount of collateral increases for both mortality models. An increasing recovery rate means that the insurer gets more protection in the case of default and hence this transaction will reduce hedger's default as well as longevity exposure. The insurer would recover their exposure more with the increased recovery rate.

4   Conclusions

In this paper, the effect of collateral on the longevity swap transaction is examined. Different mortality models and different amount of collaterals are used to obtain recovery fractions of the swap. We use the Lee-Carter with exponential jumps and renewal process model and the original Lee-Carter model to calculate the legs of the swap. The analysis show that, the recovery rates increase as the collateral amount increases for both mortality models. However, the Lee-Carter with exponential jumps and renewal process model provides more hedge effectiveness compared to the other model and it has the highest risk reduction levels. Recovery fractions reflect the amount of insurer gets in the case of default. Obtaining higher recovery values indicates that this transaction will reduce longevity and default risk exposure of hedger.

References

1. Bauer, D., Biffis, E., Sotomayor, L.R.: Optimal collateralization with bilateral default risk (2015). https://doi.org/10.2139/ssrn.2320108
2. Biffis, E., Blake, D., Pitotti, L., Sun, A.: The cost of counterparty risk and collateralization in longevity swaps. J. Risk Insur. **83** (2012). https://doi.org/10.2139/ssrn.1801826
3. ISDA: International Swaps and Derivatives Association. Collateral Review (1999)
4. Johannes, M., Sundaresan, S.: The impact of collateralization on swap rates. J. Finance **62**, 383–410 (2007). https://doi.org/10.1111/j.1540-6261.2007.01210.x
5. Li, S.H., Chan, W.S.: The Lee-Carter model for forecasting mortality. N. Am. Actuar. J. **11**, 68–89 (2007). https://doi.org/10.1007/s10680-020-09559-9
6. Li, J., Li, J.S.H., Tan, C.I., Tickle, L.: Assessing basis risk in index-based longevity swap transactions. Ann. Actuar. Sci. **13**, 1–32 (2018). https://doi.org/10.1017/S1748499518000179
7. Özen, S., Şahin, Ş: Transitory mortality jump modeling with renewal process and its impact on pricing of catastrophic bonds. J. Comput. Appl. Math. **376**, 1–15 (2020). https://doi.org/10.1016/j.cam.2020.112829
8. Zhou, R., Li, J.S.H., Tan, K.S.: Pricing standardized mortality securitizations: a two-population mortality model with transitory jump effects. J. Risk Insur. **80**, 733–774 (2013). https://doi.org/10.1111/j.1539-6975.2013.12015.x

Time-Varying Assets Clustering via Identity-Link Latent-Space Infinite Mixture: An Application on DAX Components

Antonio Peruzzi[(✉)] and Roberto Casarin

Ca' Foscari University of Venice, Venice, Italy
{antonio.peruzzi,r.casarin}@unive.it

Abstract. Finance literature suggests that cross-correlations among assets increase during periods of financial distress, and that cross-correlation's very own clustering structure varies over time. This work proposes an Identity-Link Latent-Space Infinite-Mixture model to analyze the clustering structure of cross-correlation over time. The model allows for the representation of stocks on a d-dimensional Euclidean space and the clustering of assets into groups. Model estimation is carried out within a Bayesian framework, which allows including prior extra-sample information in the inference and accounting for parameter uncertainty. We apply the model to time-varying correlations among the DAX components. We find evidence of clustering effects and positive dependence between the number of clusters and both annualized volatility and average cross-correlation.

Keywords: Latent space models · Bayesian inference · Non-parametric methods

1 Introduction

It is a well-established fact in the finance literature that stocks cross-correlation increases during periods of financial turmoil [2,9,11,14]. There is also empirical evidence of clustering effects in assets cross-correlation and of time variations in the clustering structure [1,8,10]. On a preliminary visual inspection, these two stylized facts seem to hold for the components of the DAX 40 index. Panel A in Fig. 1 reports the box plots of the yearly cross-correlation of DAX components through time. The average cross-correlation increases during the 2008–2009 financial crisis, the European debt crisis (year: 2011), the stock-market sell-off (2015), and the covid-19 breakout (2020). Panel B in Fig. 1 reports the Epanechnikov kernel density estimates of cross-correlation of DAX components over the years. The presence of bi-modality signals a change in the clustering structure of stock returns over time.

The analysis of stocks cross-correlation dynamics and clustering is of paramount importance, especially for practitioners. Stock market portfolios originally devised to achieve diversification may experience higher components cross-correlation in periods of financial distress. In such a context, having some insight into which components do not experience an increase in cross-correlation may be relevant information to consider.

In this work, we propose a Latent-Space (LS) Infinite-Mixture model. We build on the LS modeling strategy of D'Angelo [3] and propose a new LS model with an identity-link function well suited for symmetric and positive-definite matrices. The model allows for clustering in the correlation among asset log-returns with an unknown number of clusters. We follow a Bayesian approach to inference. This framework allows for the inclusion of extra-sample information and makes it easier to handle latent variables and identification issues. The model inference estimates the asset coordinates on a latent d-dimensional Euclidean space, the number of clusters, and the cluster composition.

We apply the model to the cross-correlations of a set of DAX components in a time-lapse raging from 2003 to 2021 and investigate the relationship between the estimated number of clusters over time and two relevant market features, which are average cross-correlation and annual volatility. We find a $+65.6\%$ correlation with DAX volatility and a $+27\%$ correlation with the average cross-correlation of DAX components. Our finding is coherent with what is found in previous works (see [1]).

The structure of the work is the following. Section 2 provides a review of the literature on LS models. Section 3 introduces the proposed LS model. Section 4 provides an application to correlations among DAX components and presents the main results. Eventually, Sect. 5 wraps up this work.

2   Literature Review

LS models have been introduced by Hoff et al. [6] and allow to filter a lower d-dimensional latent space from network-structured data. Several modifications to the early logistic-link model have been implemented. Friel et al. [4] propose a dynamic latent space model for the analysis of the bipartite network of interlocking directors. A similar dynamic model is proposed by Sewell et al. [12] accommodating for the use of count and non-negative real-valued weighted edges. In an early follow-up, Handcock et al. [5] provide an extension of the model allowing for clustering by drawing latent positions from a finite Gaussian mixture. Yet in the discussion section of that work, Trevor Sweeting suggests the adoption of an infinite mixture of Gaussian distributions by means of a Dirichlet Process. To the best of our knowledge, a first static implementation of such a model is present in the doctoral work of D'Angelo [3]. Our model represents an adaptation of the aforementioned model in two main directions. On the one hand, the logistic link is replaced by an identity link to gain flexibility in modeling assets cross-correlations. On the other hand, our model takes into account also the time domain, as we are interested in the dynamic evolution of both the latent positions and the number of clusters.
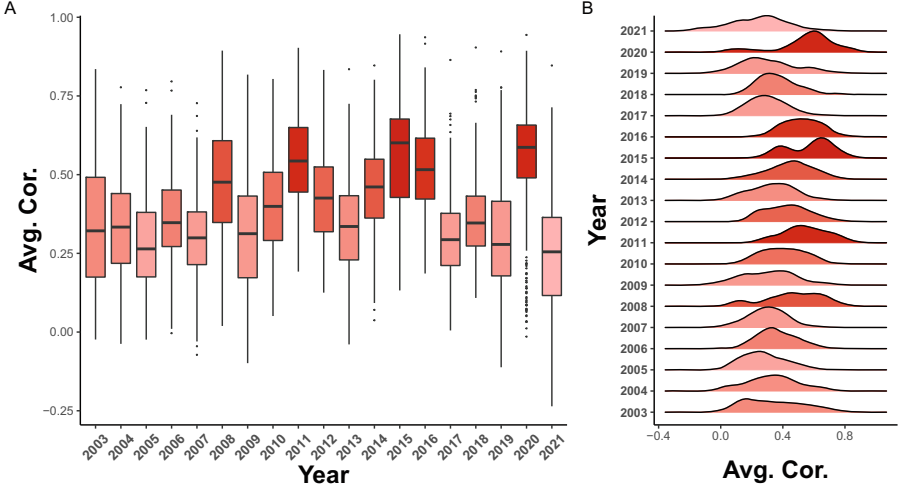
Fig. 1. DAX Components Cross-correlation Panel A: Boxplots of DAX's components cross-correlations over time. Average cross-correlation increases in periods of financial turmoil. Panel B: Epanechnikov Kernel Density Estimates of DAX's components cross-correlation distribution through time with smoothing parameter $h = 0.04$. Bi-modality emerges during years of financial distress such as 2008, 2015, and 2020.

A related application of LS models to financial data can be found in Aheleg-bey et al. [1]. The authors combine LS models and covariance-structure learning algorithms to extract latent asset positions from a Graphical VAR model. Our model differs from their model in many aspects. First, we assume the number of clusters is random and is not given a priori, allowing for time variations in the number of clusters. Secondly, our Dirichlet Process Prior assumption (DPP) combined with a Bayesian inference procedure allows us to estimate the number of clusters over time and account naturally for cluster uncertainty in the predictive distribution. The empirical results in the two works provide evidence of an increasing number of clusters during periods of financial distress.

3   LS Infinite-Mixture Model with Identity Link

Let $\hat{\mathbf{C}}_\tau \ \tau = 1, 2, \ldots, T$ be a sequence of $N \times N$ correlation matrices. Each matrix is estimated in one of T time intervals on a set of N log-return time series $\mathbf{Y} = \{\mathbf{Y}_1, \mathbf{Y}_2, ..., \mathbf{Y}_N\}$. Let $\mathbf{w}_\tau$ be the $M \times 1$ vector of unique cross-correlations of $\hat{\mathbf{C}}_\tau$ at the interval τ with characteristic element $w_{ij\tau}$. We assume $w_{ij\tau} \sim \mathcal{N}(\mu_{ij\tau}, \gamma_\tau^2)$ _i.i.d._ where $\mathcal{N}(\mu, \gamma^2)$ denotes a Normal distribution described by a location and scale parameters μ and γ, respectively. We further assume that the location parameter is inversely proportional to the squared distance in the latent space for any pair of assets. We model such a relationship as follows:

$$\mu_{ij\tau} = \alpha_\tau - ||\mathbf{x}_{i\tau} - \mathbf{x}_{j\tau}||^2 \tag{1}$$

where $||\cdot||$ denotes the Euclidean distance and $\mathbf{x}_{i\tau} \in \mathbb{R}^d$ is a d-dimensional vector of latent coordinates. Latent variable models pose inference difficulties which naturally call for Bayesian inference. In this work, we assume a non-parametric hierarchical prior on the latent vector $\mathbf{x}_{i\tau}$. At the first stage of the prior, we assume $\mathbf{x}_{i\tau}|\boldsymbol{\theta} \sim \mathcal{N}(\boldsymbol{\mu}, \sigma^2 I)$ where $\boldsymbol{\theta} = \{\boldsymbol{\mu}, \sigma^2\}$ denotes the hyper-parameter vector. At the second stage, we assume a DPP with concentration parameter ψ and base measure S_0, that is: $\boldsymbol{\theta}|S \sim S$ with $S \sim DP(\psi, S_0)$. We choose the base measure S_0 as given by the product of $\mathcal{N}(\mathbf{0}, \omega^2 I)$ and $Inv\chi^2(\zeta)$, respectively a multivariate normal distribution with mean $\mathbf{0}$ and variance-covariance matrix $\omega^2 I$ and an inverse chi-squared distribution with degrees of freedom parameter ζ. Following the DPP assumption the latent positions model admits the following infinite mixture representation:

$$\mathbf{x}_{i\tau}|S \sim \sum_{k_{i\tau}=1}^{\infty} \lambda_{k_{i\tau}} \mathcal{N}(\boldsymbol{\mu}_{k_{i\tau}}, \sigma^2_{k_{i\tau}} I) \tag{2}$$

Our LS model identifies latent factors (i.e., the latent coordinates) driving the asset correlations, the correlation clustering structure, and allows for the estimation of the number of clusters in the latent positioning. The inference problem becomes more tractable within a Bayesian framework since data augmentation can be applied and simulation methods can be easily included in the inference procedure. We adopt a Gibbs Sampling procedure adapting to our model the slice-sampling technique proposed in [7, 13].

4   Empirical Application

By means of the identity-link LS infinite-mixture model, we investigate the changes in the clustering structure of the stocks included in the DAX 40 index as of December 10th, 2021. Daily closing prices for the DAX components are obtained from the Bloomberg platform for the period ranging from January 1st, 2003, to December 10th, 2021. We drop from the dataset all the components with missing observations and obtain 29 price series.

For each year τ from 2003 to 2021, the correlation matrix $\hat{\mathbf{C}}_\tau$ among log-return series is estimated and an LS model, LS_τ, is fitted on each $\hat{\mathbf{C}}_\tau$. As an illustrative example, Panel A in Fig. 2 reports the latent coordinates estimated by the LS_{2020} model. The model detects the presence of two main clusters in the latent positioning of the index components. The great majority of stocks belong to a single cluster while QIA and SRT3 belong to a second separated cluster. This result is coherent in light of the covid-19 breakout. In fact, the stock price of QIA and SRT3 – two bio-medical companies – behaved counter-cyclically w.r.t. the market, probably discounting their expertise in molecular testing.

Panel B in Fig. 2 reports the bar plots for the most frequent number of clusters estimated by each LS_τ model. The red bars in the picture are proportional to the number of times a given number of clusters has been drawn by the MCMC algorithm. Inspection suggests that the number of clusters increases during years
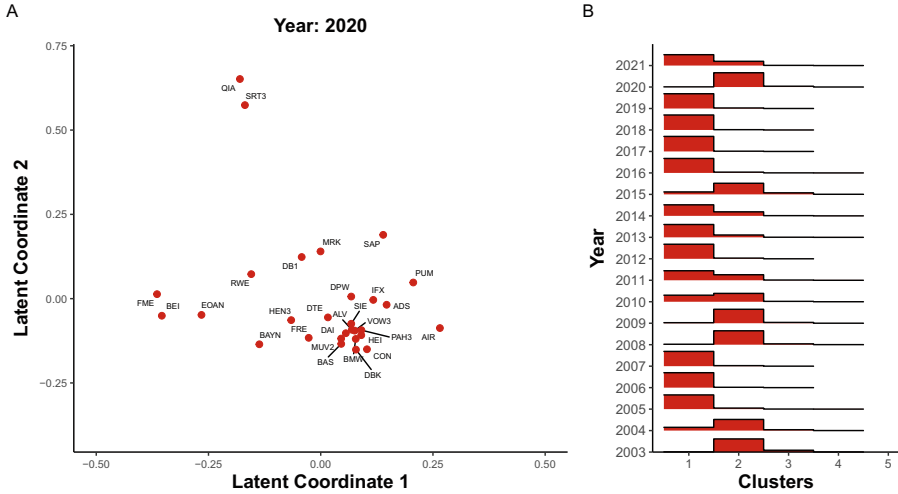
Fig. 2. Empirical Application Panel A: Latent Positioning of Dax components in the year 2020 (covid-19 breakout). The algorithm detects the presence of two clusters. The majority of the stocks belong to a single cluster, while QIA and SRT3 – two pharmaceutical companies – belong to a second distinct cluster. Panel B: bar plot representing the number of sampled clusters year by year

of financial turbulence. This seems to be true for the years of the financial crisis (2008–2009), stock market selloff (2015), and the covid-19 crisis (2020). Overall, the yearly time series of the median number of clusters shows a +65.6% correlation with the DAX-index yearly volatility and a +27% correlation with DAX-components average cross-correlation.

5 Conclusion

In this work, we proposed an identity-link LS infinite-mixture model to extract the latent positions of a set of assets from returns cross-correlation. The model not only provides a graphical representation of such positions but offers information about assets' clustering structure without the need of specifying the number of clusters a-priori. We applied our model to the components of the DAX index estimating latent positions on a yearly basis. We found that the number of clusters positively correlates with market volatility and, more loosely, with average cross-correlation. Such results are coherent with existing evidence found in other works.

References

1. Ahelegbey, D.F., Carvalho, L., Kolaczyk, E.: A Bayesian covariance graph and latent position model for multivariate financial time series. SSRN 3090236 (2020)
2. Balogh, E., Simonsen, I., Nagy, B.Z.: Persistent collective trend in stock markets. Phys. Rev. E **82**, 066–113 (2010)
3. D'Angelo, S.: Latent space models for multidimensional network data, pp. 61–83, PhD Thesis (2018)
4. Friel, N., Rastelli, R., Wyse, J., Raftery, A.E.: Interlocking directorates in Irish companies using a latent space model for bipartite networks. Proc. Nat. Acad. Sci. **113**(24), 6629–6634 (2016)
5. Handcock, M.S., Raftery, A.E., Tantrum, J.M.: Model-based clustering for social networks. J. Roy. Stat. Soci. Ser. A (Stat. Soc.) **170**(2), 301–354 (2007)
6. Hoff, P.D., Raftery, A.E., Handcock, M.S.: Latent space approaches to social network analysis. J. Am. Stat. Assoc. **97**(460), 1090–1098 (2002)
7. Kalli, M., Griffin, J.E., Walker, S.G.: Slice sampling mixture models. Stat. Comput. **21**(1), 93–105 (2011). https://doi.org/10.1007/s11222-009-9150-y
8. Kocheturov, A., Batsyn, M., Pardalos, P.M.: Dynamics of cluster structures in a financial market network. Physica Stat. Mech. Appl. **413**, 523–533 (2014)
9. Münnix, M.C., et al.: Identifying states of a financial market. Sci. Rep. **2**(1), 1–6 (2012)
10. Nie, C.X.: Dynamics of cluster structure in financial correlation matrix. Chaos, Solitons Fractals **104**, 835–840 (2017)
11. Preis, T., Kenett, D.Y., Stanley, H.E., Helbing, D., Ben-Jacob, E.: Quantifying the behavior of stock correlations under market stress. Sci. Rep. **2**(1), 1–5 (2012)
12. Sewell, D.K., Chen, Y.: Latent space models for dynamic networks with weighted edges. Soc. Netw. **44**, 105–116 (2016)
13. Walker, S.G.: Sampling the Dirichlet mixture model with slices. Commun. Stat. Simul. Comput. **36**(1), 45–54 (2007)
14. Zheng, Z., Podobnik, B., Feng, L., Li, B.: Changes in cross-correlations as an indicator for systemic risk. Sci. Rep. **2**(1), 1–8 (2012)

Demographic Risks Associated with a Tontine Investment

Peter Pflaumer[(✉)]

Department of Statistics, Technical University of Dortmund, Dortmund, Germany
`peter.pflaumer@tu-dortmund.de`

Abstract. This article analyzes the stochastic aspects of a tontine using a Gompertz distribution. In particular, the probabilistic and demographic risks of a tontine investment are examined. The expected value and variance of tontine payouts are calculated. Both parameters increase with age. The stochastic present value of a tontine payout is compared with the present value of a fixed annuity. It is shown that only at very high ages the tontine is more profitable than an annuity. Finally, the demographic risks associated with a tontine are discussed. Elasticities are used to calculate the impact of changes in modal age on the tontine payout. It is shown that the tontine payout is very sensitive to changes in modal age.

Keyword: Gompertz distribution · Life table · Annuity · Demography · Mortality

1 Introduction

A tontine is a rising life annuity, in which the annual constant interest yield or dividend of the total investment sum is paid out to the surviving subscribers of the tontine. It is commonly believed that the Italian banker Lorenzo Tonti invented tontines in the 17th century. In the 18th and 19th centuries, tontines were a popular means of public financing (see, e.g., Hellwege [1]). The tontine ends with the death of the last subscriber. The capital falls to the initiator of the tontine. Recently, the tontine is considered as an attractive alternative to life annuities (see, e.g., Milevsky [3, 4] or Hellwege [1]).

2 Gompertz Distribution

Benjamin Gompertz proposed in 1825 a life table function, which is one of the oldest and most famous models of demography. It states that the mortality intensity exponentially increases with age in adulthood. It has been much applied in the life table analysis and insurance mathematics with various modifications (e.g., Gompertz-Makeham law). Due to declining children and youth mortality, it has again become essential in order to describe "modern" life tables with low mortality. The Gompertz model allows to fully describe the present and future life tables in industrialized countries by using only two

parameters, in principle, which are both easy to estimate from data. It provides a good approximation of life tables in these populations (see, e.g., Pollard [7] or Pflaumer [5, 6]).

The survival function is given by

$l(x) = \exp(e^{-k \cdot m} - e^{k \cdot (x-m)})$, where m is the modal age and k the growth rate of the force of mortality function $\mu(x) = k \cdot e^{k \cdot (x-m)}$. The life expectancy at birth is calculated by $e_0 = m - \gamma/k$ with $\gamma = 0,577221566...$(Euler-Mascheroni-Constant).

The rectangularization of the survival curve is defined as a trend toward a more rectangular shape of the survival curve, due to increased survival and concentration of deaths around the mean age at death. The variability in the age at death declines and deaths are being compressed into the upper years of life. An increase in the parameter k causes rising rectangularization.

3   Basic Formulas of the Tontine Payout

It is assumed that all n investors buy at age u. The tontine share per investor is equally high. The mortality is subject to a Gompertz distribution with the parameters m and k.

l(x): Probability of surviving from birth to age x
l(u): Probability of surviving from birth to age u
N(x,u): Number of investors at age x alive who buy at age u
$N(0, u) = \frac{n}{l(u)}$ Number at age x = 0 of the life table
Z: Total yearly payout
n = N(u,u): Number of subscribers of the tontine at age u
i: Interest rate
Z/n: Initial payout
t(x,u): (Expected) tontine payout per surviving investor at age x $x \geq u$

$$t(x, u) = \frac{Z}{N(x, u)} = \frac{Z}{n \cdot \frac{l(x)}{l(u)}} = \frac{Z}{\frac{n}{l(u)} \cdot \exp(e^{-k \cdot m} - e^{k \cdot (x-m)})}$$

$$= \frac{Z}{n \cdot \exp(e^{k \cdot (u-m)} - e^{k \cdot (x-m)})}; \quad x \geq u$$

The growth rate of the tontine is $\frac{\frac{\delta t}{\delta x}}{t} = \mu(x) = k \cdot e^{k \cdot (x-m)}$.

t(x,u) is a random variable, since it is assumed that N(x,u) has a binomial distribution (see Li and Tuljapurkar [2]). The variance is:

$$Var(N(x, u)) = n \cdot \exp\left(e^{k \cdot (u-m)} - e^{k \cdot (x-m)}\right) \cdot \left(1 - \exp\left(e^{k \cdot (u-m)} - e^{k \cdot (x-m)}\right)\right)$$

$$= n \cdot \left(\exp\left(e^{k \cdot (u-m)} - e^{k \cdot (x-m)}\right) - \exp\left(2 \cdot e^{k \cdot (u-m)} - 2 \cdot e^{k \cdot (x-m)}\right)\right)$$

The 0.05- and 0.95-quantiles of N(x,u) are N*(x,u)$_{0,05}$ and N*(x,u)$_{0,95}$, which are obtained from the corresponding binomial distribution.

The lower and upper limits of a 90%-confidence interval of the tontine payouts are given by the quantiles

$$L_{x,0.05} = \frac{Z}{N^*(x,u)_{0.95}} \text{ and } U_{x,0.05} = \frac{Z}{N^*(x,u)_{0.95}}.$$

Approximative $(1 - \alpha)$-confidence interval for the survivors:

$$N(x, u) \pm u_{1-\frac{\alpha}{2}} \sqrt{Var(N(x, u))}$$

Approximative $(1 - \alpha)$-confidence interval for the tontine payouts.

$$\text{Lower limit: } \frac{Z}{N(x, u) + u_{1-\frac{\alpha}{2}} \sqrt{Var(N(x, u))}} ;$$

$$\text{Upper limit: } \frac{Z}{N(x, u) - u_{1-\frac{\alpha}{2}} \sqrt{Var(N(x, u))}}$$

u_α is the α-quantile of the standard normal distribution.

Present value of the tontine payout up to age a (with a first payout at age u + 1):

$$\int_u^a e^{-i(u-x)} \cdot \frac{Z}{n \cdot \exp\left(e^{k\cdot(u-m)} - e^{k\cdot(x-m)}\right)} dx$$

$$= \frac{Z}{n} \cdot \exp\left(i \cdot u - e^{k\cdot(u-m)}\right) \cdot \int_u^a \exp\left(e^{k\cdot(x-m)} - i \cdot x\right) dx$$

Present value of the tontine payout up to age a (with a first payout at age u):

$$\int_u^a e^{-i(u-x)} \cdot \frac{Z}{n \cdot \exp\left(e^{k\cdot(u-m)} - e^{k\cdot(x-m)}\right)} dx + \frac{Z}{n}$$

$$= \frac{Z}{n} \cdot \left(1 + \exp\left(i \cdot u - e^{k\cdot(u-m)}\right) \cdot \int_u^a \exp\left(e^{k\cdot(x-m)} - i \cdot x\right) dx\right)$$

4 Demographic Risks

In recent decades, life expectancy has risen sharply worldwide. While the increase was initially due to a decline in infant mortality, it is now the result of lower mortality in old age. For example, the life expectancy of a man at age 60 today (2016/18 life table of Germany) is 22 years. Around 140 years earlier, it was only 12 years, and in 1970/1972 it was 15 years. The demographic risk will be explained by a rise in life expectancy or in modal age.

Elasticity of t(x,u) with respect to m:

$$\varepsilon_{t,m}(x) = \frac{\frac{\delta t}{\delta m}}{t} \cdot m = \left(-k \cdot e^{k\cdot(x-m)} + k \cdot e^{k\cdot(u-m)}\right) \cdot m = \left(-r(x) + k \cdot e^{k\cdot(u-m)}\right) \cdot m$$

$$m \gg u : \quad \rightarrow \hat{\varepsilon}_{t,m}(x) = -r(x) \cdot m; \; \varepsilon_{t,m}(m) \approx -k \cdot m$$

Elasticity of t(x,u) with respect to k:

$$\varepsilon_{t,k}(x) = \frac{\frac{\delta t}{\delta k}}{t} \cdot k = k \cdot e^{k \cdot (x-m)} \cdot (x - m) + k \cdot e^{k \cdot (u-m)} \cdot (m - u)$$

$$= r(x) \cdot (x - m) + k \cdot e^{k \cdot (u-m)} \cdot (m - u)$$

Approximation: $k \cdot e^{k \cdot (x-m)} \cdot (x - m) = r(x) \cdot (x - m)$

$\varepsilon_{t,k}(m) = k \cdot e^{k \cdot (u-m)} \cdot (m - u) > 0$; $\varepsilon_{t,k}(x)$ has a minimum at $x = m - \frac{1}{k}$.

Approximate solution: $\varepsilon_{ts,k}(x)$ has a zero point at $m - e^{k \cdot (u-m)} \cdot (m - u)$ (Taylor series of order 1 around $x = m$). From the total derivative $(x > m)$ $dt = f_m \cdot dm + f_k \cdot dk = 0$ follows $\frac{dk}{dm} = -\frac{f_m}{f_k} = \frac{k \cdot (e^{k \cdot x} - e^{k \cdot u})}{e^{k \cdot x} \cdot (x-m) + e^{k \cdot u} \cdot (m-u)}$ or $\Delta k \approx \frac{k \cdot (e^{k \cdot x} - e^{k \cdot u})}{e^{k \cdot x} \cdot (x-m) + e^{k \cdot u} \cdot (m-u)} \cdot \Delta m$.

Example:

The mortality follows a Gompertz distribution with m = 88.7 and k = 0.1152 (data for estimation: German life table for females (2013/2015). It is assumed that all n = 400 investors buy at age u = 65 with l(65) = 0.93691. The number at age x = 0 of the life table is therefore 427. The tontine share per investor is 100. Thus, the total yearly payout is 1,600 and the initial payout per investor is 4. The comparable annuity can be calculated as 7.3 ($ä_{y,65}$ = 13.734 at i = 4%). At age 86 the tontine payout exceeds the annuity of 7.3; at age x = 96, e.g., the payout varies between 30.8 and 50.0; the expected value is 38.1 (see Fig. 1). The present value of tontine payouts exceeds at age 92 the initial investment of 100 (see Fig. 2). The present value increases rapidly after that age with increasing confidence intervals. At the modal age of m = 88.7, the elasticity is −10 (see Fig. 3); an increase of the modal age reduces the tontine payout substantially (e.g., an increase of m by 1% reduces the tontine payout by about 10% at age 88.7). The reduction can be partly compensated by an increase of k, which means a rectangularization of the life table. For example, if at age x = 95 the modal age increases by 1 year, k would have to increase by nearly 0.02 for the payout to remain the same. This increase in k is unrealistically high. Therefore, an increase in m will always lead to a reduction in the payout.
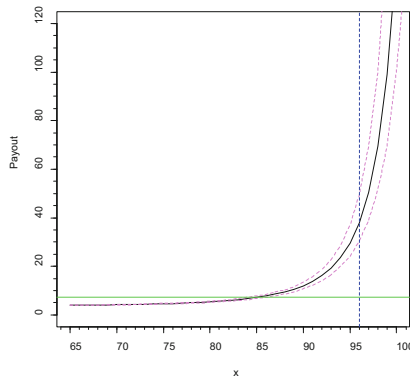
Fig. 1. Annuity (green) and tontine payout with 90% - confidence intervals
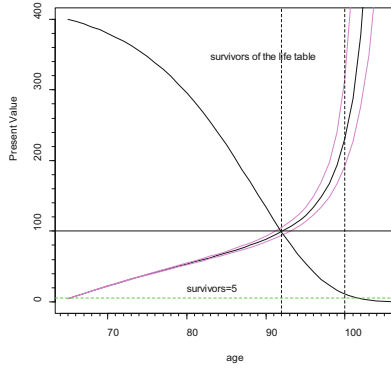
Fig. 2. Present value of the tontine payout with 90% - confidence intervals
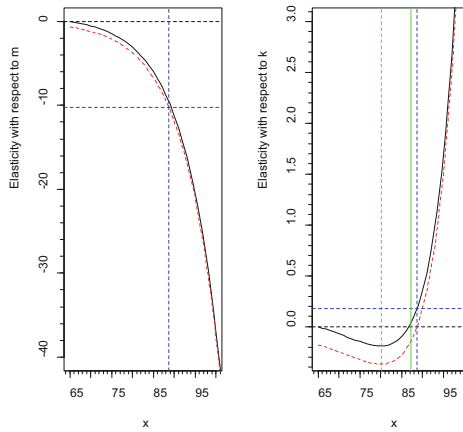
Fig. 3. Demographic influences (black: exact, red: approximative formulas)

5 Conclusion

A tontine is a worthwhile investment for people who reach a very old age. But it is subject to three types of risk: financial (interest rate, default, inflation), probabilistic and demographic. The probabilistic risk can be reduced by increasing the number of investors. The demographic risk should not be underestimated. Decreasing mortality and increasing life expectancy will lead to a significant reduction in tontine payout.

References

1. Hellwege, P. (ed): The Past, Present, and Future of Tontines, Duncker & Humblot, Berlin (2018)
2. Li, N., Tuljapurkar, S.: The probabilistic life table and its applications. Paper presented at the 2013 Annual Meeting of the PAA, New Orleans (2013)

3. Milevsky, M.: Portfolio choice and longevity risk in the late seventeenth century: a re-examination of the first English tontine. Financ. Hist. Rev. **21**(3), 225–258 (2014)
4. Milevsky, M.: King William's Tontine: Why the Retirement Annuity of the Future Should Resemble Its Past, Cambridge University Press, Cambridge (2015)
5. Pflaumer, P.: Demographic Applications of the Gompertz Distribution: Quantitative Analyses with Special Reference to Ulpian's Table and Epitaph Inscriptions, Kempten (2018)
6. Pflaumer, P.: Methods for estimating selected life table parameters using the Gompertz distribution. In: JSM Proceedings, Social Statistics Section, pp. 733–747. American Statistical Association, Alexandria (2011)
7. Pollard, J.: An old tool – modern applications, actuarial studies and demography, Research Paper No. 001/98, August 1998

A Geographical Analysis of the Systemic Risk by a Compositional Data (CoDa) Approach

Francesco Porro[(✉)]

Università degli Studi di Genova, Genoa, Italy
`francesco.porro@unige.it`

Abstract. After the financial crises in the last decades, the systemic risk is now recognized as an unavoidable issue to be constantly monitored. In this paper the contributions to the global systemic risk of different geographic areas are investigated. The analysis is performed by considering the aggregate systemic risk of the firms located in a specific geographic area as a part of the distribution of the global systemic risk. The techniques used for a such investigation are based on Compositional Data (CoDa) methodology, a quite recent approach very useful when the relevant information convoyed by the data is in the proportions among the parts and not in their absolute values or in their sum.

Keywords: Systemic risk · SRISK · Compositional data · Aitchison geometry

1 Introduction

The relevance of the systemic risk has been globally emphasized by the global financial crisis of 2007-2009. Initially, there was no commonly accepted definition of the systemic risk, but now it can be identified as "the propensity of a financial institution to be undercapitalized when the financial system as a whole is undercapitalized" [6]. Many papers deal with the issue of the assessment of the systemic risk: for an overview on this topic, see for example [4,6], and the references therein. An important measure for evaluating the systemic-risk degree associated with a single firm is the so-called SRISK, which has been firstly introduced in [1] and later extended in many other papers. More details about this topic can be found in the interesting review in [7].

The value of SRISK related to a single specific firm can be intended as the amount of money needed by the firm to rise to function normally in the case of a financial crisis (such as a very significant loss of the financial markets). Formally, at a given time t, the value of SRISK for an individual firm i is calculated by:

$$SRISK_{it} = k\left[D_{it} + (1 - \phi_{it})W_{it}\right] - (1 - \phi_{it})W_{it}, \qquad \text{where:}$$

- k is the prudential capital ratio;
- D_{it} the book value of total liabilities;
- W_{it} the market capitalization (or market value of equity);
- ϕ_{it} the Long-Run Marginal Expected Shortfall (LRMES), which corresponds to the expected drop in equity value conditional on the market falling by more than 40% within the next six months.

The SRISK can be decomposed in the difference between two quantities: the first one, which represents the *Required Capital*, is equal to $k[D_{it}+(1-\phi_{it})W_{it}]$, while the second one, given by $(1 - \phi_{it})W_{it}$ corresponds to the *Available Capital*. As stated in [5], "SRISK is a function of the size of the firm, its degree of leverage, and its expected equity devaluation conditional on a market decline. SRISK is higher for firms that are larger, more leveraged, and with higher sensitivity to market declines." A negative (or null) value of SRISK for a specific firm, means that such firm can overcome the market shock with no capital injections by the government. For this reason, a measure of the financial distress for a system (with N firms) is the aggregate SRISK, that, at a given moment t, is defined by:

$$SRISK_t = \sum_{i=1}^{N} (SRISK_{it})_+ \tag{1}$$

where $(x)_+$ denotes $\max(0, x)$. This measure represents the total amount of bailing out a financial system (with N firms), conditional on a systemic event. In the computation of the aggregate SRISK, the negative amounts of SRISK do not effectively contribute in the sum in formula (1), because in a crisis it is unlikely that surplus capital will be easily mobilized through mergers or loans among firms (cfr. [4]). The fact that, financially speaking, the world is a global village and the regulators must consider "a large picture" to make decisions, is now accepted by the most part of the researchers. For this reason, an analysis about the influence of geographical factors to the systemic risk can help to monitor and to better understand the dynamics of such kind of risk.

   Keeping this in mind, in this paper a geographical analysis of the composition of the systemic risk is performed. The analysis is conducted with a dataset provided by the Volatility Laboratory (V-Lab) (more information about the data can be found in Sect. 3). The approach used is based on the Compositional Data (CoDa) methodology, a set of quite recent techniques, which is getting more and more attention in the literature (cfr. [2] and [3]).

2   The Compositional Approach

The Compositional Data (CoDa) are multivariate observations where relative rather than absolute information is relevant. This means that they represent a quantitative description of the parts of some whole. The basic pillar in compositional methods are the compositions, defined as follows.

Definition 1. *A composition vector is a real-valued vector with all (strictly) positive components. A D-part composition is a class of equivalence which contains all the* compositionally equivalent *vectors in* $\mathbb{R}^D$*, where two compositions* $\mathbf{x} = (x_1, x_2, \ldots, x_D)$ *and* $\mathbf{y} = (y_1, y_2, \ldots, y_D)$ *are compositionally equivalent if there exists a positive constant* $\lambda \in \mathbb{R}^+$ *such that* $\mathbf{x} = \lambda \cdot \mathbf{y}$.

A suitable sample space for the equivalence classes is the D-part simplex $\mathbb{S}^D$, defined as:

$$\mathbb{S}^D = \{(x_1, x_2, \ldots, x_D) \in \mathbb{R}^D \ : \ x_i > 0 \ \forall i; \ \sum_{i=1}^{D} x_i = c\}, \tag{2}$$

where c is a positive arbitrary constant. For further details, see [8,9] and the references therein. Usually in compositional analysis, the vectors of proportions (which sum to 1) are used as representatives of an equivalence class: this corresponds to select $c = 1$ in the previous definition, and in the next one.

Definition 2. *The closure (to c) of the D-part composition* $\mathbf{x} = (x_1, x_2, \ldots, x_D)$ *is given by:*

$$\mathscr{C}(\mathbf{x}) = \left(\frac{c \cdot x_1}{\sum_{i=1}^{D} x_i}, \frac{c \cdot x_2}{\sum_{i=1}^{D} x_i}, \ldots, \frac{c \cdot x_D}{\sum_{i=1}^{D} x_i} \right).$$

Starting from these initial two definitions, it is possible to create a coherent geometry, called Aitchison geometry on the simplex, which allows a deep analysis of compositional data (see [9] for further details).

A typical compositional dataset $\mathbf{X}$ is a sample of n observations of D-part compositions $\mathbf{X} = (\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n)'$, with $\mathbf{x}_i = (x_{i1}, x_{i2}, \ldots, x_{iD})$, $i = 1, 2, \ldots, n$. Since in such a dataset, the standard statistical descriptive measures, based on the real Euclidean structure, applied to compositional data may lead to erroneous conclusions (see for example [9]), an alternative set of descriptive measures based on the Aitchison geometry can more properly be used. In the following, just the most common two are reported.

Definition 3. *An indicator of central tendency for the compositional dataset* $\mathbf{X}$ *is the closed geometric mean. This vector is called* center*, and it is defined as*

$$cen(\mathbf{X}) = \mathscr{C}(g_1, g_2, \ldots, g_D),$$

where g_j *denotes the geometric mean of the n observations related to the j-th component of the vectors in* $\mathbf{X}$*:* $g_j = \left(\prod_{i=1}^{n} x_{ij} \right)^{1/n}$*,* $j = 1, 2, \ldots, D$.

The dispersion in a compositional dataset $\mathbf{X}$, can be described by the variation matrix, defined by:

$$T = \begin{pmatrix} 0 & t_{12} & \cdots & t_{1D} \\ t_{21} & 0 & \cdots & t_{2D} \\ \vdots & \vdots & \ddots & \vdots \\ t_{D1} & t_{D2} & \cdots & 0 \end{pmatrix}, \quad \text{where } t_{ij} = var\left(\ln \frac{x_i}{x_j} \right).$$

For measuring the variability of a dataset by a single value, the following definition of *Total Variance* has been introduced.

Definition 4. *The Total Variance of the compositional sample* **X** *is a measure of its global dispersion. It is based on the entries of the variation matrix T:*

$$TotVar(\mathbf{X}) = \frac{1}{2D} \sum_{i=1}^{D} \sum_{j=1}^{D} var\left(\ln \frac{x_i}{x_j}\right) = \frac{1}{2D} \sum_{i=1}^{D} \sum_{j=1}^{D} t_{ij}.$$

As the information convoyed by the compositions is relative, an usual practice is to apply transformations, mapping the compositions into real vectors in order to exploit the usual Euclidean structure. In the literature there are several transformations based on the logratios: the additive logratio (alr), the centered logratio (clr), and the isometric logratio (ilr). Unfortunately, even a simple overview of these transformations is out of scope of this paper: the interested reader can see [9] and [8], among the others, for further details. For the provided analysis in the following, it can just be reported that the clr-transformation is basically characterized by two important properties: the first one is that it does not change the number of parts, since a D-part composition is mapped in a vector in $\mathbb{R}^D$. The second one is that it preserves the distances and the angles: this implies that the Aitchison distance in the simplex of two compositions is equal to the distance of the corresponding transformed vectors in $\mathbb{R}^D$ (see [9] for details). This feature is fundamental in exploratory analyses based on metrics, like clr-biplots and ternary or De Finetti diagrams. The definition of the centered logratio transformation, is the following one.

Definition 5. *The centered logratio transformation (clr) of a composition* $\mathbf{x} = (x_1, x_2, \ldots, x_D)$ *is given by*

$$clr(\mathbf{x}) = \ln\left(\frac{x_1}{g_m(\mathbf{x})}, \frac{x_2}{g_m(\mathbf{x})}, \ldots, \frac{x_D}{g_m(\mathbf{x})}\right),$$

where $g_m(\mathbf{x})$ denotes the geometric mean of the D parts: $g_m(\mathbf{x}) = \left(\prod_{i=1}^{D} x_i\right)^{1/D}$.

In every compositional data analysis, graphics are usually used to visualize and interpret the data. The most common one is the PCA biplot. Generally speaking, the biplots permit the representation of a rank-2 approximation of the data, and they are based on the Single Value Decomposition (SVD) of the centered (or standardized) data matrix. In Compositional Data analysis, there are two basic kinds of biplots: the *form biplot*, which favours the display of the units, and the *covariance biplot*, which favours the display of the variables. More details on them can be found in [9], and in [8].

3   Application

The data used in the application are provided by the Volatility Laboratory (V-Lab) of the New York University Stern School of Business, and they are available

at the website: https://vlab.stern.nyu.edu. The dataset consists of the value of
the aggregate SRISK for four macro-areas: Africa, Americas, Europe, and Asia.
Such values are related to each quarter from 2014 to 2021: the value for each
quarter is the value of SRISK in the last month of the period (March, June,
September, and December). The only exception regards the year 2021: since at
the moment, the SRISK detection of December is not available, the last quarter
of such year is not considered. The aggregate SRISK of each macro-area is based
on the SRISK of all the firms in such geographical area, monitored by the V-Lab
team, therefore enough big to have a systemic relevance. In practice, the dataset
can be seen as a compositional sample of 31 different 4-compositions, grouped in
8 different years (2014-2021). The left panel of Table 1 reports the sample center
and the right panel, the variation matrix of the dataset. The center highlights
an important predominance of Asia, showing a geometric mean that exceeds
more than the 50% SRISK share over the sample period. The right panel shows
that the maximum variability is associated with Africa, and the Total Variance
is equal to 0.2890. In order to identify patterns in the data, Fig. 1 shows the
covariance biplot. The first two Principal Components are representative of the
dataset variability, since the first one explains the 66.43% of the total variance,
and the second one adds an other 24.83%, bearing the proportion of the cumula-
tive explained variance at more than 90% (91.26%). By using some interpretation
rules of the compositional biplot, some remarks can be achieved from Fig. 1. The
first one is related to the interpretation of the Principal Components: the first
component discriminates Africa from the other macro-areas, likely because it
captures a sort of "evolution degree" of the financial markets in the macro-area;
the second one sharply distinguishes between Asia and Americas. By an obser-
vation of the length of the four rays (the segments joining each vertex to the
center), it can be stated that the smallest one is related to Europe, suggesting
that the logratio of such macro-area has the smallest contribution to the total
variability. The links (segments joining two vertices) corresponding to the pairs
Asia-Americas and Europe-Africa are nearly orthogonal: this suggests that the
corresponding logratios should be checked for zero correlation. The observation
that the four vertices are very spread out indicates that all the variances of the
logratios are very far from being null, highlighting the lack of proportionality of

Table 1. The center of the 4-part composition (left panel) and the Logratio variances
of the 4-part composition for the macro-areas (right panel).

Macro-Areas	Center		Africa	Americas	Asia	Europe	Clr variances
Africa	0.0031	Africa	0	0.2751	0.2525	0.3219	0.1061
Americas	0.1299	Americas	0.2751	0	0.1438	0.0840	0.0606
Asia	0.5262	Asia	0.2525	0.1438	0	0.0786	0.0629
Europe	0.3408	Europe	0.3219	0.0840	0.0786	0	0.0594
					Total Variance		0.2890

the parts. Since the projections of the points of the quarters in 2014 on the link Asia-Americas are (quite) close to the projection of the center on the same link, it can be state that such quarters have the value of the logratio corresponding to Asia and Americas (quite) equal to its average on the whole dataset. The projections of the points of years 2014, 2015, 2018 and 2020 are (quite) far from the projection of the center on the link Europe-Africa: this shows that the values of the corresponding logratios are (quite) very different from the average on the whole dataset. All these findings can be considered coherent with the results of other analysis, but their added value is that they came from the application of the quite recent CoDa methodology.
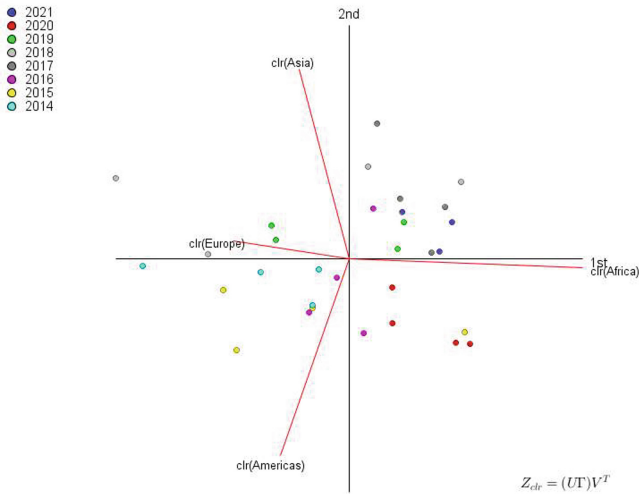
Fig. 1. The covariance biplot of the dataset of the four macro-areas.

References

1. Acharya, V., Richardson, M.: Restoring Financial Stability: How to Repair a Failed System. Wiley, Hoboken (2009)
2. Aitchison, J.: The statistical analysis of compositional data. J. R. Stat. Soc. Series B Stat. Methodol. **44**(2), 139–160 (1982)
3. Aitchison, J.: The Statistical Analysis of Compositional Data. Monographs on Statistics and Applied Probability, Chapman & Hall Ltd., London (1986)
4. Benoit, S., Colliard, J.E., Hurlin, C., Pérignon, C.: Where the risks lie: a survey on systemic risk. Rev. Financ. **21**(1), 109–152 (2017)
5. Brownlees, C., Engle, R.F.: SRISK: a conditional capital shortfall measure of systemic risk. Rev. Financ. Stud. **30**(1), 48–79 (2017)
6. Engle, R., Jondeau, E., Rockinger, M.: Systemic risk in Europe. Rev. Financ. **19**(1), 145–190 (2015)

7. Engle, R.: Systemic risk 10 years later. Annu. Rev. Financ. Econ. **10**, 125–152 (2018)
8. Filzmoser, P., Hron, K., Templ, M.: Applied Compositional Data Analysis. Springer, Switzerland (2018)
9. Pawlowsky-Glahn, V., Egozcue, J., Tolosana-Delgado, R.: Modeling and Analysis of Compositional Data. Wiley, Hoboken (2015)

Jump-Telegraph Market Model: Barrier Binary Options

Nikita Ratanov[✉]

Chelyabinsk State University, Br. Kashirinykh, 129, Chelyabinsk, Russia
rtnnkt@gmail.com

Abstract. The paper continues the study of a market model based on jump-telegraph processes. It is assumed that the price of a risky asset follows the stochastic exponential of a piecewise linear process, equipped with jumps that occur at the moments of a pattern change. In this case, the standard option pricing formula was derived earlier and it is very similar to the classic Black-Scholes formula, see [7]. Meanwhile, exotic options for this model have not been studied yet.

Within this framework, we are developing procedures for pricing binary barrier options. This article concerns the "cash-(at hit)-or-nothing" binary barrier option. The cases "down-and-in" and "up-and-in" are studied separately. The main tools of this analysis are methods developed for first pass probabilities. Some known results related to the ruin probabilities follow directly from these settings. On the same basis, some advanced versions of binary options can also be developed and studied.

Keywords: Jump-telegraph process · Martingales · Binary option · Barrier option

1 Market Model and Measure Transform

Let $\varepsilon = \varepsilon(t) \in \{0, 1\}$, $t \geq 0$, be a random process with two states, switching at random times $\{\tau_n\}_{n \geq 1}$, $\tau_0 = 0$, which form a Markov flow controlled by two alternating rates λ_0 and λ_1. Let $N(t) = \max\{n : \tau_n \leq t\}$ be a counting process.

Let $X(t) = \int_0^t c_{\varepsilon(u)} \mathrm{d}u$, $t \geq 0$, be an asymmetric telegraph process with parameters (c_0, λ_0), (c_1, λ_1), $c_0 \geq c_1$, and $J = J(t) = \sum_{n=1}^{N(t)} Y_n$ is a jump process with independent random jump amplitudes Y_n, $Y_n > -1$. Consider a market model consisting of one risky asset. Let the stock price $S(t)$, $t \in [0, T]$, follows the stochastic equation

$$\mathrm{d}S(t) = S(t-)\mathrm{d}\left(X(t) + J(t)\right), \qquad 0 < t < T. \tag{1}$$

Assume that the distributions h_0 and h_1 of the jump amplitudes Y_n interchange alternately as the market state $\varepsilon = \varepsilon(t)$ switches. For simplicity, let the interest rate be constant r, $r \geq 0$, so that the bond price is $B_t = e^{rt}$, $0 \leq t \leq T$.

Integrating (1), we have $S(t) = S_0 \mathscr{E}_t (X + J)$, $t \in [0, T]$, where $S_0 = S(0)$. Here $\mathscr{E}_t(\cdot)$ is the stochastic exponential, so that $S(t) = S_0 e^{X(t)} \kappa_{N(t)}$, where $\kappa_{N(t)} = \prod_{k=1}^{N(t)} (1 + Y_n)$. See, for example, [1].

This idea is now well understood. On the basis of such jump-telegraph and similar processes, various market models have been constructed, see the overview in [2]. Note that the standard call option pricing formula has the same structure as the classic Black-Scholes formula. Meanwhile, exotic options have been studied much less.

Recall that the jump-telegraph process $X + J$ and its stochastic exponential $S(t)$, $t \geq 0$, are martingales if and only if $c_0 + \lambda_0 y_0 = 0$, $c_1 + \lambda_1 y_1 = 0$, where y_0 and y_1 are the average amplitudes of the jumps associated with the state 0 and 1, respectively, see [4]. Since jump amplitudes are random, this model usually has infinitely many risk-neutral measures. Following in the footsteps of R.C.Merton, [5], consider the measure transformation based on the Radon-Nikodym derivative

$$\frac{d\mathbb{Q}}{d\mathbb{P}} = \mathscr{E}_t(X^* + J^*) = \exp(X^*(t))\kappa^*_{N(t)},$$

where $X^*(t) = \int_0^t c^*_{\varepsilon(u)} du$, $J^*(t) = \sum_{n=1}^{N(t)} y^*_{\varepsilon(\tau_n-)}$, $\kappa^*_{N(t)} = \prod_{n=1}^{N(t)} (1 + y^*_n)$. The deterministic constants c^*_0, c^*_1 and y^*_0, y^*_1 satisfy the martingale condition, that is,

$$c^*_0 + \lambda_0 y^*_0 = 0, \qquad c^*_1 + \lambda_1 y^*_1 = 0. \tag{2}$$

Given the measure $\mathbb{Q}$ in this way, the distribution of jump amplitudes does not change, but the market regimes switch with changed intensities λ^*_0, λ^*_1,

$$\lambda^*_0 = \lambda_0(1 + y^*_0) = \lambda_0 - c^*_0, \qquad \lambda^*_1 = \lambda_1(1 + y^*_1) = \lambda_1 - c^*_1.$$

By virtue of (2), one can see that the martingale measure for this market model is given by $c^*_i = \lambda_i + c_i/y_i$ and $y^*_i = -c^*_i/\lambda_i = -1 - c_i/(\lambda_i y_i)$, $i \in \{0, 1\}$.

2   Cash-(At Hit)-or-Nothing Barrier Binary Option

This paper deals with binary barrier options. In particular, we are interested in the price of the option in the simple case when the payoff is received at the moment the barrier x is breached. To be specific, let, first, $S_0 < x$.

Consider an option with a payoff function $H(x) = \mathbb{1}\left\{\max_{t \in [0,T]} S(t) > x\right\}$, where $\mathbb{1}\{A\}$ is the indicator of the event A, and the payoff is received at the time

$$T(x) = \inf\{t \in (0, T] \mid S(t) > x\}$$

when the stock price $S(t)$, $t \in (0, T]$, passes through x for the first time.

The option price is determined by a pair of bond-discounted payoff expectations

$$\mathfrak{c}_0 = \mathbb{E}_{\mathbb{Q}}\left[\mathrm{e}^{-rT(x)} \mid \varepsilon(0) = 0\right], \qquad \mathfrak{c}_1 = \mathbb{E}_{\mathbb{Q}}\left[\mathrm{e}^{-rT(x)} \mid \varepsilon(0) = 1\right],$$

with respect to the martingale measure $\mathbb{Q}$, and depending on the state of the market at underwriting.

Note that the first passage time $T(x)$ for the stock price coincides with the first passage time for the process $X(t) + \log \kappa(t)$ through the threshold $\log[x/S_0]$.

In what follows, we study the first passage time problem for the process $X(t) + J(t)$, $0 \leq t \leq T$. Therefore, to describe the market model (1), it suffices to modify the results obtained below by replacing $x \to \log(x/S_0)$, $Y_n \to \log(1+Y_n)$ with appropriate changes in the jump distributions.

Let the measure be already martingale.

2.1 "Bull Market" and Positive Threshold

We consider first the model with alternating positive trends $c_0 > c_1 > 0$, and negative jump amplitudes. The latter means that the alternating distributions h_0 and h_1 of Y_n-*corrections* to rising bullish market (log-)prices are supported by $(-\infty, 0]$. Let $x > 0$.

Let us denote the first passage time through the threshold x by the log-price process $X(t) + J(t)$ as $\tau_x := \inf\{t > 0 : X(t) + J(t) > x\}$. Let $\phi = (\phi_0(x), \phi_1(x))'$ be the Laplace transform of τ_x,

$$\begin{aligned}
\phi_0(x) &= \mathbb{E}\left[\mathrm{e}^{-r\tau_x}\mathbb{1}_{\{\tau_x < T\}} \mid \varepsilon(0) = 0\right], \\
\phi_1(x) &= \mathbb{E}\left[\mathrm{e}^{-r\tau_x}\mathbb{1}_{\{\tau_x < T\}} \mid \varepsilon(0) = 1\right].
\end{aligned} \tag{3}$$

By definition $0 \leq \phi_i(x) \leq 1$, $\forall x$. Because $\tau_x = 0$ for non-positive x, we set $\phi_0(x) \equiv 1$, $\phi_1(x) \equiv 1$, if $x \leq 0$. Further, $\phi_0(x)$, $\phi_1(x)|_{x > c_0 T} = 0$, since the threshold x, $x > c_0 T$, is never reached till time T.

By conditioning on the first Markov switching, we obtain the pair of coupled integral equations. First, in the case $0 < x < c_1 T$, we have

$$\phi_0(x) = \mathrm{e}^{-(\lambda_0 + r)x/c_0} + \int_0^{x/c_0} \lambda_0 \mathrm{e}^{-(r+\lambda_0)t}\, \mathscr{H}_0^{x-c_0 t}[\phi_1]\mathrm{d}t, \tag{4}$$

$$\phi_1(x) = \mathrm{e}^{-(\lambda_1 + r)x/c_1} + \int_0^{x/c_1} \lambda_1 \mathrm{e}^{-(r+\lambda_1)t}\, \mathscr{H}_1^{x-c_1 t}[\phi_0]\mathrm{d}t. \tag{5}$$

Here, the first terms correspond to the price movement without speed switching, and the operators $\mathscr{H}_i^z$, catching the first jump, are defined by the convolutions $\mathscr{H}_i^z[\phi] := \int_{-\infty}^{\infty} \phi(z - y)h_i(\mathrm{d}y)$, $i \in \{0, 1\}$, for any test-function $\phi = \phi(z)$.

In the case $c_1 T < x < c_0 T$, the process $X(t)$ does not reach the threshold x without switching from the low velocity c_1 on the time interval $[0, T]$. Therefore, in this case, Eq. (5) should be written as

$$\phi_1(x) = \int_0^T \lambda_1 \mathrm{e}^{-(r+\lambda_1)t}\, \mathscr{H}_1^{x-c_1 t}[\phi_0]\mathrm{d}t. \tag{6}$$

Both systems of integral equations, (4)–(5) and (4), (6), have a unique boundary solution, [3]. To get it, note that $\psi_\alpha = \exp(-\alpha x)$, $\text{Re}(\alpha) > 0$, is an eigenfunction of the operator $\mathscr{H}^x$,

$$\mathscr{H}^x [\psi_\alpha] = \int_{-\infty}^0 \exp\left(-\alpha(x-y)\right) h(\mathrm{d}y) = \widehat{h}(-\alpha)\psi_\alpha(x), \qquad x > 0,$$

with the eigenvalue $\widehat{h}(-\alpha) = \int_{-\infty}^0 \mathrm{e}^{\alpha y} h(\mathrm{d}y)$. For a real positive α, the eigenvalue $\widehat{h}(-\alpha)$ is real, $0 \le \widehat{h}(-\alpha) \le 1$, and the function $\alpha \to \widehat{h}(-\alpha)$, $\alpha \in (0, \infty)$, decreases.

Let

$$\phi(x) = \sum_{k=1}^N \exp(-\alpha_k x)\mathbf{A}_k, \qquad x > 0, \qquad \text{Re}\alpha_k > 0, \tag{7}$$

with indefinite coefficients $\mathbf{A}_k = (A_{k0}, A_{k1})'$ and exponents α_k, $k = 1, \dots, N$. Substituting this function into (4)–(5), after a simple algebra, we find that $N = 2$, and α_0, α_1 are the roots of the equation

$$(\alpha c_0 - \lambda_0 - r)(\alpha c_1 - \lambda_1 - r) = \lambda_0 \lambda_1 \widehat{h}_0(-\alpha)\widehat{h}_1(-\alpha). \tag{8}$$

For system (4), (6), α_0 and α_1 are roots of

$$(\alpha c_0 - \lambda_0 - r)(\alpha c_1 - \lambda_1 - r) = \lambda_0 \lambda_1 \widehat{h}_0(-\alpha)\widehat{h}_1(-\alpha) \cdot \left(1 - \mathrm{e}^{-(r+\lambda_1 - \alpha c_1)T}\right). \tag{9}$$

Each of Eqs. (8) and (9) has exactly two real and positive roots α_1 and α_2, such that

$$0 < \alpha_1 < \min\left(\frac{\lambda_0 + q}{c_0}, \frac{\lambda_1 + q}{c_1}\right) \le \max\left(\frac{\lambda_0 + q}{c_0}, \frac{\lambda_1 + q}{c_1}\right) < \alpha_2,$$

since the right parts of these equations are decreasing and positive. The same simple algebra leads to explicit formulae for the coefficients $\mathbf{A}_1$ and $\mathbf{A}_2$, see [8] for details.

In a particular case of the Cramér-Lundberg model, our previous achievements help to find the ruin probability explicitly. Precisely, suppose that our market model the trends are positive and equal, $c_0 = c_1 = c > 0$, $\lambda_0 = \lambda_1 = \lambda$, and the negative jumps have exponential distribution with the mean value $1/b$, $b > 0$. That is, $\widehat{h}_0(-\alpha) = \widehat{h}_1(-\alpha) = b/(b+\alpha)$. This is a compound Poisson process with i.i.d. negative exponential jumps and a positive trend.

We are interested in the distribution of the ruin time τ_x, $x > 0$.

In this case, Eq. (8) can be simplified to the pair of quadratic equations,

$$(c\alpha - \lambda - r)(\alpha + b) = \pm\lambda b,$$

which give two positive roots α; the first one is obtained from $(c\alpha - \lambda - r)(\alpha + b) = -\lambda b$,

$$\alpha_1 = \frac{\lambda + r - bc + \sqrt{(\lambda + r - bc)^2 + 4bcr}}{2c} = -b + \frac{1}{2c}\left(q + \sqrt{q^2 - 4bc\lambda}\right). \tag{10}$$

where $q = q(r) = r + \lambda + bc$, and the second is found from $(c\alpha - \lambda - r)(\alpha + b) = +\lambda b$,

$$\alpha_2 = \frac{\lambda + r - bc + \sqrt{(\lambda + r - bc)^2 + 4bc(r + 2\lambda)}}{2c}.$$

Due to (7) and the explicit formulae for $\mathbf{A}_1$ and $\mathbf{A}_2$, the Laplace transform of τ_x is given by

$$\phi_0(x) = \phi_1(x) = A_1 e^{-\alpha_1 x} + A_2 e^{-\alpha_2 x},$$

where $A_1 = 1$ and $A_2 = 0$.

Since the limit of the root $\alpha_1 = \alpha_1(r)$ at $r \downarrow 0$ depends only on $\lambda/c - b$, $\lim_{r\downarrow 0} \alpha_1(r) = \dfrac{\lambda - bc}{c} \mathbb{1}_{\{\lambda > bc\}}$, the ruin probability is given by

$$\mathbb{P}\{\tau_x < \infty\} = \lim_{r\downarrow 0} \phi(x) = \begin{cases} \exp\left(-(\lambda/c - b)x\right), & \text{if } \lambda/c > b, \\ 1, & \text{otherwise,} \end{cases}$$

which coincides with known results.

Further, by [6, formula (2.2.5-18)] the inverse Laplace transform of the function

$$q \to \exp(aq - a\sqrt{q^2 - z^2}) - 1, \qquad q > z,$$

is given by

$$\frac{az}{\sqrt{t^2 + 2at}} I_1\left(z\sqrt{t^2 + 2at}\right),$$

where $I_1(\cdot)$ denotes the modified Bessel function of the first order. Due to (10), the probability density function $p(t; x) = \mathscr{L}_{q\to t}^{-1}(\exp(-\alpha_1(q)x))$,

$$p(t; x) = e^{-\lambda t}\left[\delta\left(t - x/c\right) + \frac{x\sqrt{\lambda b}}{\sqrt{t(ct - x)}} I_1\left(2\sqrt{\lambda bt(ct - x)}\right)\right], \qquad t > -x/c,$$

where $\delta(\cdot)$ is the Dirac delta-function, corresponding to movement without switching.

2.2 "Bull Market" and Negative Threshold

A process with positive trends $c_0 > c_1 > 0$ breach the threshold $x < 0$ only by jumping. It cannot pass this level without switching, and the problem becomes much more difficult. Similarly to Sect. 2.1, we obtain the equations

$$\phi_0(x) = \int_0^T \lambda_0 e^{-(r+\lambda_0)t} H_0(x - c_0 t) dt$$
$$+ \int_0^T \lambda_0 e^{-(r+\lambda_0)t} dt \int_{x-c_0 t}^0 \phi_1(x - c_0 t - y) h_0(dy), \qquad (11)$$

$$\phi_1(x) = \int_0^T \lambda_1 e^{-(r+\lambda_1)t} H_1(x - c_1 t) dt$$
$$+ \int_0^T \lambda_1 e^{-(r+\lambda_1)t} dt \int_{x-c_1 t}^0 \phi_0(x - c_1 t - y) h_1(dy), \qquad (12)$$

where $H_i(x)$, $i \in \{0, 1\}$, is the accumulated probability distribution function of the jump occurring in state i. The first terms of these equations appear under the condition of breakdown of the threshold after the first jump, the second terms correspond a small jump, not enough to immediately exercise the option.

In the case of infinite expiration time (perpetual option with no fixed maturity) and exponentially distributed jump sizes, Eqs. (11)–(12) become a bit simpler. Indeed, let (negative) jumps have alternating exponential distributions $H_i(y) = \mathbb{P}_i\{Y < y\} = e^{b_i y} \wedge 1$, $b_i > 0$, $i \in \{0, 1\}$. Equations (11)–(12) take the form

$$\phi_0(x) = \frac{\lambda_0}{r + \lambda_0 + b_0 c_0} \left[e^{b_0 x} + b_0 \int_x^0 e^{b_0(x-z)} \phi_1(z) \mathrm{d}z + b_0 \int_{-\infty}^x e^{-(r+\lambda_0)(x-z)/c_0} \phi_1(z) \mathrm{d}z \right],$$

$$\phi_1(x) = \frac{\lambda_1}{r + \lambda_1 + b_1 c_1} \left[e^{b_1 x} + b_1 \int_x^0 e^{b_1(x-z)} \phi_0(z) \mathrm{d}z + b_1 \int_{-\infty}^x e^{-(r+\lambda_1)(x-z)/c_1} \phi_0(z) \mathrm{d}z \right].$$

$$(13)$$

With the additional symmetry assumptions $\lambda_0 = \lambda_1 = \lambda$, $c_0 = c_1 = c$ and $b_0 = b_1 = b$, these equations give the ruin probability. In this symmetric case, $\phi_0 = \phi_1 = \phi$, and, under the net profit condition $bc > \lambda$, the ruin probability $\mathbb{P}\{\tau_x < \infty\}$ is given explicitly:

$$\mathbb{P}\{\tau_x < \infty\} = \phi(x)|_{r=0} = \frac{\lambda}{bc} \exp\left(x(b - \lambda/c)\right),$$

which can be proved by plugging into Eqs. (13). This coincides with the known results, see, e.g. [9, (5.3.8)].

2.3   "Bear Market" and the Cramér-Lundberg Ruin Model

Let both trends be negative, $0 > c_0 > c_1$, $x < 0$, and jumps positive. This case is symmetric to bullish market with negative corrections described in Sect. 2.1. In this case, the Laplace transform ϕ of τ_x is given by formula (7) with the *negative* exponential rates α_1, α_2 given by the Eqs. (8) and (9).

3   Conclusion

A new approach to barrier options pricing is proposed. The model is based on simple piecewise-linear processes with random jumps occurring at velocity switching. Due to space constraints, we've only mentioned the cash-(at hit)-or-nothing type here. Of course, a complete overview of binary options within this framework can also be given and will be published elsewhere later.

References

1. Jeanblanc, M., Yor, M., Chesney, M.: Mathematical Methods for Financial Markets. Springer, Heidelberg (2009)
2. Kolesnik, A.D., Ratanov, N.: Telegraph Processes and Option Pricing, 2nd edn. Springer, Heidelberg (2013)
3. Linz, P.: Analytical and Numerical Methods for Volterra Equations. SIAM, Philadelphia (1985)
4. López, O., Ratanov, N.: Option pricing driven by a telegraph process with random jumps. J. Appl. Probab. **49**, 838–849 (2012). https://doi.org/10.1239/jap/1346955337
5. Merton, R.C.: Option pricing when underlying stock returns are discontinuous. J. Financ. Econ. **3**, 125–144 (1976). https://doi.org/10.1016/0304-405X(76)90022-2
6. Prudnikov, A.P., Brychkov, Y.A., Marichev, O.I.: Integrals and Series, vol. 5. Inverse Laplace Transforms. Gordon and Breach Science Publication (1992)
7. Ratanov, N.: A jump telegraph model for option pricing. Quant. Financ. **7**, 575–583 (2007). https://doi.org/10.1080/14697680600991226
8. Ratanov, N.: First crossing times of telegraph processes with jumps. Methodol. Comput. Appl. Probab. **22**, 349–370 (2020). https://doi.org/10.1007/s11009-019-09709-5
9. Rolski, T., Schmidli, H., Schmidt, V., Teugels, J.: Stochastic Processes for Insurance and Finance. Wiley, Hoboken (1999). https://doi.org/10.1002/9780470317044

Estimating Recovery Curve for NPLs

Roberto Rocci[1], Alessandra Carleo[2(✉)], and Maria Sole Staffa[3]

[1] Sapienza University, Rome, Italy
[2] Roma Tre University, Rome, Italy
`alessandra.carleo@uniroma3.it`
[3] European University, Rome, Italy

Abstract. The objective of the present paper is to propose a new method to measure the recovery performance of a portfolio of non-performing loans (NPLs) in terms of recovery rate and time to liquidate. The fundamental idea is to draw a curve representing the recovery rates during time, here assumed discretized, for example, in years. In this way, the user can get simultaneously information about recovery rate and time to liquidate of the portfolio. In particular, it is discussed how to estimate such a curve in presence of right censored data, i.e. when the NPLs composing the portfolio have been observed in different periods. Uncertainty about the estimates is depicted trough confidence bands obtained by using the non-parametric Bootstrap. The effectiveness of the proposal is shown by applying the method to a real financial data set about some portfolios of Italian unsecured NPLs taken in charge by a specialized operator.

Keywords: Recovery rate · Time to liquidate · NPLs · Censored data

1 Introduction

Non-Performing Loans (NPLs) are exposures in state of insolvency, that means loans whose collection by banks is uncertain. As Resti and Sironi [9] point out, an effective recovery depends on the characteristics of the exposure, of the counterparty, on macroeconomic and on internal (to the bank) factors. There is a NPL market that offers banks the opportunity to get rid of non-performing loans by selling them to specialized operators who deal with recovery. The main method for determining the value of Non-Performing Loans is that of discounted financial flows, according to which the value of the loans is equal to the sum of the expected income flows, discounted at a rate consistent with the expected unlevered return of the investor and net of the related recovery costs.

In the case of a performing loan, the borrower is expected to pay principal and interest at the agreed deadlines with a high level of probability (one minus the probability of default, generally low). In this case, the uncertainty in the valuation is limited to the determination of the discount rate to consider the general market conditions and the specific risk of the debtor. In the case of Non-Performing Loans, the uncertainty concerns not only the discount rate but also the amount that will be returned and the time of return. In fact, the probability of default is now equal to one, or is in any case

very high, if the credit is in other categories of impaired loans (unlikely to pay). The valuation methodologies currently used on the market are therefore based primarily on forecast models of the amount of net repayments expected from receivables and related collection times.

The estimation methodology for recovery rate, which we are interested in for *NPLs*, is faced in the more general context of Basel II. It is well known that under internal ratings-based (IRB) approach to determine capital requirements for credit risk, banks are required to estimate the following risk components: probability of default (PD), loss given default (LGD), exposure at default (EAD) and maturity (M). While the estimation of PD is well established, LGD is not so well investigated and still subject to research. Given the borrower has already defaulted, LGD is defined as the proportion of money financial institutions fail to gather during the collection period and, conversely, Recovery Rate (RR) is defined as the proportion of money financial institutions successfully collected. That means $LGD = 1 - RR$.

Recovery rate (or LGD) can be estimated using both parametric and non-parametric methods. Mainly, recovery rate is estimated using parametric methods and considering a one-year time horizon. Methods used in literature, among others, are: classical linear regression, regularized regression like Lasso, Ridge, Elastic-net, etc. [7], Beta regression, inflated Beta regression, two-stage model combining a Beta mixture model with a logistic regression [10].

In the case of NPLs, in our opinion, in investigating the recovery process of defaulted exposures the focus must be not only on the recovered amounts, but also on the duration of the recovery process, the so-called time to liquidate (TTL).

Cheng and Cirillo [4] propose a model that can learn, using a Bayesian update in a machine learning context, how to predict the possible recovery curve of a counterpart. They introduce a special type of combinatory stochastic process, based on a complex system of assumptions, referring to a discretization of recovery rates in *m* levels.

Our purpose is to introduce a particular non parametric method to measure the performance of a NPLs portfolio in terms of recovery rate (RR) and time to liquidate (TTL) jointly, without assuming any particular model and/or discretization of the RR. The idea is to represent the recovery process as a curve showing how the RR is distributed during the time without assuming a particular parametric model. We will also propose a method to estimate such a curve when some data are censored. The plan of the paper is the following. In Sect. 2 we show how the recovery curve is defined, while in Sect. 3 the method of estimation in case of censored data is introduced. In Sect. 4, the effectiveness of the proposal is shown through an application on real data, while some conclusions and final remarks are discussed in Sect. 5.

2 Recovery Rate and Time to Liquidate of a Portfolio

The definition of recovery rate (RR) and time to liquidate (TTL) of a NPLs portfolio is not trivial because the two quantities are strictly connected. Since it is crucial to decide when to measure the RR and TTL – that is when each NPL in the portfolio has been entirely liquidated or after a given period to be defined - the measurement of the RR cannot disregard the measurement of the TTL and vice versa.

First, we note that to measure the TTL when the last NPL has been liquidated could lead to measures highly affected, and biased, by anomalous NPLs with long TTLs and small EAD. It follows that the TTL should be measured when the RR becomes significant. It remains to understand what is "significant". Second, in many cases the user needs a more complete information rather than only two numbers: RR and TTL. It would be better to know how the RR increases during the time. This would also help in choosing at what RR point to measure the TTL according to whatever optimality criterion the operator decides to adopt. For the aforementioned reasons, we decide to measure the behavior of the RR during the time through what we called the "recovery curve". Such a curve is built in the following way.

Let us consider a portfolio of K NPLs. For each of the K NPLs the debt exposure at default is EAD_k (exposure at default of the k-th NPL) and the total portfolio exposure . Assume I time intervals (of the delay of payment) from the default (time t_0) to the valuation date (time t_I). Let $p_{k,i}$ be the recovery of the k-th , in the i-th interval (of delay), i.e. $(t_{i-1}, t_i]$, with $k \in \{1, 2, …, K\}$ and $i \in \{1, 2, …, I\}$. The portfolio recovery in time interval i equals , that is the total recovery, for all the K debt positions, in the i-th time interval of delay. Consequently, after i time intervals of delay, i.e. by the end of the interval $(t_0, t_i]$ we define as the total portfolio "recovery value until time t_i", i.e. the total recovery, for all the K debt positions, in the first i periods from the default date. We could also define the total recovery, being $V(p_i)$ the value of P_i^* capitalized at an appropriate interest rate. In this initial study, we (like many other, i.e. [10]) do not consider the interest because we consider time and recovery rate together and also because the recovery curve, even if lower, would have the same trend. We define also $R_i = P_i/EAD$ as the portfolio "recovery rate until time i", while $r_i = p_i/EAD$ equals the portfolio recovery rate in the i-th time interval. Since , $r_i = R_i - R_{i-1}$ and $r_1 = R_1$ we can refer in an equivalent way to R_i or to r_i for $i \in \{2, …, I\}$.

In Table 1 there is an example of portfolio with $K = 4$ debt positions.

Table 1. Portfolio with $K = 4$ debt positions.

k	EAD_k	$p_{k,1}$	$p_{k,2}$	$p_{k,3}$	$p_{k,4}$
1	100	10	0	0	0
2	200	20	15	0	0
3	300	20	25	10	15
4	400	30	35	10	#N/D

We are interested in measuring the portfolio performance in 3 years after default, i.e. $I = 3$ periods of delay. It can be measured in terms of recovery rates until year i as

Table 2. Portfolio ($EAD = 1000$) performance in 3 years ($I = 3$).

i	1	2	3
p_i	80	75	20
P_i	80	155	175
r_i	8.00%	7.50%	2.00%
R_i	8.00%	15.50%	17.50%

We see that, for example, in the first 2 years the portfolio recovers the 15.5% of the total initial exposure: 8% in the first year and 7.5% in the second.

Sometimes the available data are incomplete, in particular, censored, i.e. the $p_{k,i}$ are not available from a certain date on for some k. In our example, this happens in the fourth period for the NPL ($k = 4$). In this case, it is not possible to compute the recovery curve for the fourth interval without further hypotheses. In the next section, we will discuss some of them and how to estimate the recovery curve from the incomplete data.

3 Estimating the Recovery Rate Curve from Censored Data

The estimation of the recovery curve in the presence of censored data is carried out in a way similar to the estimation of a survival curve (e.g. [8]). First, we note that sometimes it is interesting to consider the "conditional recovery rate" c_i in each delay period i. Let E_i be the effective portfolio exposure at the beginning of period i

$$E_i = \begin{cases} EAD & i = 1 \\ \sum_{k=1}^{K} \left(EAD_k - \sum_{i'=1}^{i-1} p_{k,I'} \right) & i > 1 \end{cases} \quad (1)$$

that means $E_i = EAD - P_{i-1}$ with $P_0 = 0$ by convention. The conditional recovery rate is defined as $c_i = p_i/E_i$. In words, it is the recovery rate with respect to the effective portfolio exposure at the beginning of the period (E_i) rather than to the initial one (EAD). We observe that it is possible to obtain r_i from c_i and R_{i-1}:

$$r_i = \frac{p_i}{EAD} \cdot \frac{E_i}{E_i} = \frac{p_i}{E_i} \cdot \frac{EAD - P_{i-1}}{EAD} = c_i \left(1 - \frac{P_{i-1}}{EAD} \right) = c_i(1 - R_{i-1}) \quad (2)$$

It means that the recovery rate is the conditional recovery of the percentage of how much still has to be recovered. This way of computing r_i is convenient when there are censored data in the database, i.e. for some NPLs the recovery $p_{k,i}$s are observed only until a particular time. In this case, since $r_i = p_i/EAD$ cannot be used, the idea is to apply formula (2) by computing the conditional recovery rate c_i using only the available data. In details, let us suppose that $K_i = \{k = 1, ..., K \mid \exists \, p_{k,i}\}$ is the subset of indexes k corresponding to the NPLs for which at delay time i the value $p_{k,i}$ is not censored. In this case the effective portfolio exposure, for $i > 1$, is a generalization of (1):

$$E_i = \sum_{k \in K_i} \left(EAD_k - \sum_{i'=1}^{i-1} p_{k,i'} \right) \quad (3)$$

and the conditional recovery rate is

$$c_i = \left(\sum_{k \in K_i} p_{k,i}\right)/E_i \qquad (4)$$

Let's consider the previous example. If we want to consider more than 3 intervals of delay, assuming we are interested in measuring the performances in 4 years, i.e. $I = 4$ periods of delay, then we obtain the same results of Table 2 for the first 3 years, and for $i = 4$ we get: $p_4 = 15$, $P_4 = 190$, $E_4 = 500$, $r_4 = 2.48\%$, $c_4 = 3.00\%$, $R_4 = 19.98\%$. This method of measuring performances allows not only to measure jointly the recovery rate and the time to liquidate, but also to deal with censored data.

Obviously, it is wrong to imagine the censored data equal to 0, meaning no inflows instead than no information about that inflow. With the same example, substituting $p_{4,4} = 0$, we would obtain the same results of Table 2 for the first 3 years, but for $i = 4$ we would get: $p_4 = 15$, $P_4 = 190$, $E_4 = 825$, $r_4 = 1.50\%$, $c_4 = 1.82\%$, $R_4 = 19.00\%$. That is, probably, an underestimate of the true curve.

The results would have been different if we simply did not consider in the portfolio the NPLs for which the data are censored. In the example, considering $I = 4$ periods of delay excluding NPL_4 would lead to different results for all the durations, as it is shown in the table below. Such estimates are of lower quality than the proposed ones because obtained by using less data, i.e. information (Table 3).

Table 3. Portfolio ($EAD = 600$) performance for $K = 3$ loans.

i	1	2	3	4
p_i	50	40	10	15
r_i	8.33%	6.67%	1.67%	2.50%
R_i	8.33%	15.00%	16.67%	19.17%

4   Application

We analyse a data set of Italian NPLs supplied by a specialized operator. We examine two portfolios of unsecured loans with different initial debt size. The portfolios have the same year of acceptance by the operator: year 2005. In particular:

Portfolio 1: $5000 <EAD_k < 25000$, $K = 4732$, Average $EAD_k = 14709$;
Portfolio 2: $100000 <EAD_k < 250000$, $K = 876$, Average $EAD_k = 151117$.

We consider as time t_0 the year of acceptance (2005), rather than the exact time of default, because this is the moment in which the operator starts the recovery procedure. We follow the recovery history for 9 years. We observe that both portfolios have approximately 5% censored data in the last year and about 2.5‰ censored data in the penultimate year. The results in terms of r_i and R_i are reported in the plots below,
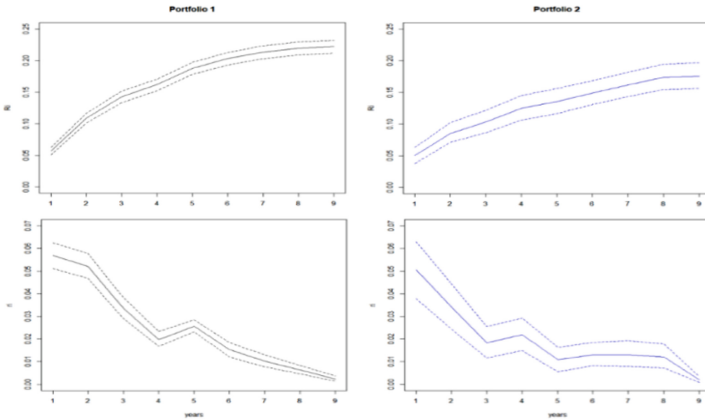
Fig. 1. Recovery rate until time i (R_i) recovery rate (r_i) of Portfolio 1 and Portfolio 2.

where the dotted lines are the boundaries of the confidence intervals computed pointwise by using a non-parametric bootstrap [6].

Obviously, the highest values of the recovery rate are at the beginning of the period ($i = 1$) and as time passes the recovery rate tends to decrease, even if not monotonically. To compare the results we discuss R_i, that in our opinion is the most explicative ratio. Even considering the width of the confidence intervals, it appears that the recovery is greater for the portfolio with smaller credits. Probably, this is because taking charge by specialized operators has greater effect on those who must return lower amounts.

In the extended version of the paper other comparisons will be presented.

5 Conclusions and Final Remarks

According to the objective of this paper, we propose a kind of measurement that takes in consideration both the recovery rate, the time to liquidate and how they interact. This is obtained by estimating a "recovery curve" displaying the behaviour of the recovery rate during the time.

In doing that, we faced the problem of censored data and we suggest to use a method of measuring performances that allows not only to measure jointly the recovery rate and the time to liquidate, but also to deal with censored data. This method is based on an algorithm that is usually used in the construction of survival curves.

Our next goal is to use our method to compare performance of portfolios with different characteristics by using non-parametric boostrap tests for clustered observations. taking into account. Another idea is to extend and test the validity of the method to cases where the database has missing data not only at the end of the observation period, but also at the beginning of it.

References

1. BIS: The Internal Ratings-based Approach. Bank for International Settlements, Basel (2001). https://www.bis.org/publ/bcbsca05.pdf

2. Bruche, M., Gonzales-Aguado, C.: Recovery rates, default probabilities, and the credit cycle. J. Bank. Finance **34**(4), 754–764 (2010)
3. Campbell, J.Y., Lo, A.W., MacKinlay, C.A.: The Econometrics of Financial Markets. Princeton University Press, Princeton (1997)
4. Cheng, D., Cirillo, P.: A reinforced urn process modeling of recovery rates and recovery times. J. Bank. Finance **96**, 1–17 (2018)
5. Devjak, S.: Modeling of cash flows from nonperforming loans in a commercial bank. Naše Gospodarstvo/Our Economy **64**(4), 3–9 (2018)
6. Efron, B.: Censored Data and the Bootstrap. J. Am. Stat. Assoc. **76**(374), 312–319 (1981)
7. Hastie, T., Tibshirani, R., Friedman, J.: The Elements of Statistical Learning. Springer, New York (2009). https://doi.org/10.1007/978-0-387-84858-7
8. Kalbfleisch, J.D., Prentice, R.L.: The Statistical Analysis Failure Time Data. Wiley, New York (2002)
9. Resti, A., Sironi, A.: Risk Management and Shareholders' Value in Banking. Wiley, New York (2007)
10. Ye, H., Bellotti, A.: Modelling recovery rates for non-performing loans. Risks **7**(1), 19 (2019)

An Application of the Pair-Copula Construction to a Non-life Dataset

Mariagrazia Rositano[(✉)] and Fabio Baione

Sapienza University, Rome, Italy
{mariagrazia.rositano,fabio.baione}@uniroma1.it

Abstract. The modern *pair-copula construction* (PCC) approach, which defines complex multivariate structures through the use of bivariate copulas, it proves to be an extremely effective tool for respond to problems in various fields of application including the actuarial one. The aim of this paper is to analyze the PCC methodology through an application to a non-life insurance portfolio in presence of categorical and continuous data. The aim is to define a multivariate distribution, highlighting the technical and operational limits in applications in the insurance field. This methodology allows to overcome both the limits of the "traditional" dependence structures and of the more "modern" copula functions. However, since each varied n-distribution has a considerable number of decompositions, the multivariate distribution was determined using Dißmann's algorithm.

Keywords: Pair-copula construction · Model selection · Insurance

1 Introduction

In the actuarial context the use of the correlation coefficient for the analysis of the dependence between pairs of random variables has led, due to the nature of the phenomena represented, such as the asymmetry of the distribution of claims and the increasing dependence between risks on the queues, to distorted or even misleading results. These limitations have driven the use of copula functions in insurance to a massive extent, and in fact, to date, there are numerous families of copulas existing in the literature and multiple approaches to define them [13]. Moreover, although there is a wide range of bivariate copulas, the choice narrows significantly for larger dimensions, since it is not always possible to use a bivariate copula extension to the n-dimensional case, assuming the latter exists.

The need for methodologies that could easily define multivariate distributions with a large number of marginals stimulated the development of *pair-copula construction* (hereafter PCC). The theory of PCC was introduced by the pioneering work of Joe [8], later taken up in the papers of Bedford and Cooke [2] and [3], who first used the term *trees regular vines*. This theory was put into inferential context by Aas et al. [1] and, subsequently, more recent studies such as Kurowicka

and Cooke [9] have defined the sequences of *vines* in the form currently used. In Czado [4] and in later work by the same author, the problem of *pair-copula selection* is dealt with extensively. Stöber and Czado al. [15] introduce R-vine PCC for discrete and continuous marginal distributions, which generalizes those of Panagiotelis et al. [14] for discrete data and Aas et al. [1] for continuous data. One of the main advantage in the presence of discrete marginal distributions of the approach they presented is the significant computational advantages since it requires only the evaluation of bivariate instead of higher dimensional copula functions. Furthermore, the number of required evaluations of copula functions to calculate the probability mass function (pmf) grows only quadratically with the number of discrete variables. As observed in [5] the use of this approach in the insurance field is complex, therefore we explore this methodology by means of an application to a non-life *database* widely used in the actuarial field [7]. This application is performed by using a very recent package called `rvinecopulib` [12] that extents some features of the most popular `VineCopula` package with the main advantages of the introduction of nonparametric and multi-parameter families and the ability to model discrete variables. In this short paper we limit ourselves to an analysis of the *fitting* and of the operational limitations that can be encountered using an non-life database.erdh, with the aim of a more in-depth study in subsequent research work, to explore how the use of the PCC can be declined in the perimeter of pricing, also using multivariate regression tools.

2   An Hint on *Pair-Copula Construction*

The development of PCC methodologies fully meets the need for methodologies that could easily define multivariate distributions with a high number of marginals.

The model is based on the decomposition of the multivariate density into a series of pair-copula, i.e., pairs of distributions joined via copula function, which can be applied to the original variables or their conditional or unconditional distribution functions. In other words, the idea behind PCC is to rewrite the density f as a product of possibly conditional *pair-copule* densities. For example, a *pair copula decomposition* of an arbitrary three-dimensional density is given by the following expression:

$$
\begin{aligned}
f(x_1, x_2, x_3) = f_1(x_1) \cdot f_2(x_2) \cdot f_3(x_3) \\
\cdot c_{12}\big(F_1(x_1), F_2(x_2)\big) \cdot c_{23}\big(F_2(x_2), F_3(x_3)\big) \\
\cdot c_{13;2}\big(F_{1|2}(x_1|x_2), F_{3|2}(x_3|x_2); x_2\big)
\end{aligned}
\tag{1}
$$

In order to represent in a simple way the many possible PCC, Cooke in 1997 expanded the concept of *tree* addressed by Cayley in 1889 by generalizing it into the modern concept of *vine*. Bedford and Cooke [2] and [3] conjugated the theory of PCC to that of graphs and introduced the so-called regular vines or R-vines. As the size increases, the number of decomposition of f using R-vine increases very rapidly. In fact, Morales-Nápoles [10] showed that on n dimensions the

number of regular R-vine structures is equal to $\frac{n!}{2} \cdot 2^{\binom{n-2}{2}}$. Therefore, given the size of this class, many recent studies have focused on two specific subclasses of R-vines: *Canonical vine* (*C-vine*) whose trees have a star-like structure; *Drawable vine* (*D-vine*) whose trees have a linear path structure. These subclasses were introduced by Aas et al. [1]. The example given above for the dimension $n = 3$ is an example of *D-vine copula*. The introduction of these classes allows the number of possible combinations of C-vine to be reduced, which is identical to the more reasonable D-vine and equal to $\frac{n!}{2}$.

In the modeling of real phenomena, it is useful to simplify the model in Eq. (1) by requiring that the copula density $c_{13;2}$ depends on x_2 only through the partition functions $F_{\cdot|2}(x.|x_2)$ (for further details see [5]). Defining the multivariate distribution through the use of *semplified* vine copulas greatly reduces the risks of *overfitting* from which insurance portfolios generally have been shown to be affected.

3   Numerical Application

The following study considered data[1] regarding a portfolio of 67,856 Motor risk contracts entered into on a one-year basis between 2004 and 2005. In the portfolio, we observe that 6,8% reported at least one claim. The variables contained in the dataset are both quantitative and categorical.

Very recently the new 'R' package `rvinecopulib` handles these types of variables essentially by transforming the categorical variables into a set of dummy variables. The resulting model will therefore have many more variables than initially supplied to the modelling function.

To find the *vine copula* that best fits the data, the *vine* function of `rvinecopulib` implements Dißmann's algorithm [6] and allows to find a "locally optimal" *R-vine copula* in reasonable time without analyzing all the $\frac{n!}{2} \times 2^{\binom{n-2}{2}}$ *R-vine copule* of size n. It is observed that Dißman's algorithm called *the greedy algorithm* does not necessarily find C- or *D-vine copula* but also a generic *R-vine*. Furthermore, it is important to note that when the number of variables is modest, it may be convenient to search for the optimal C- or *D-vine copula* through an arbitrary comparison, i.e., without making use of this algorithm. In our case, however, even considering all the variables as quantitative, an exhaustive analysis of the dataset via C- or D-vine copulas is not feasible (i.e. $\frac{10!}{2} = 1,814,400$).

In order to show a brief example, we analyzed two models using four variables: The `numclaims` measures the number of claims whereas the `claimcst0` is the loss amount; the `gender` variable (2 levels - F: female and M: male) and the `area` (6 levels identified by the letters from A to F) are pure categorical. The algorithm used transforms the gender into a binary variable assigning value 1

---

[1]   www.businessandeconomics.mq.edu.au/our_departments/Applied_Finance_and/ Actuarial_Studies_research/books/GLMsforInsuranceData/data_sets   source: *Macquarie University Department of Applied Finance and Actuarial Studies*, Sydney. Data may also be found in the `insuranceData` package of `R`.

to the M level, and the `area` is splitted into 5 dummy variables. Therefore, the generated models will have 8 variables.

The difference between the two models is that in the former we left the choice of copulas to the greedy algorithm while in the latter we restricted the choice to only elliptical copulas (hereafter model M1 and M2 respectively). Traditionally, BIC and AIC indicators are used to compare the goodness of fit of statistical models. In the case of PCC, an ad-hoc method was proposed by Nagler et al. [11], called *Modified vine copula Bayesian information criterion*, and denoted *mBICv*. It is defined as follows:

$$\text{mBICv} = -2\log \text{lik} + v \log n - 2 \sum_{t=1}^{n-1} (q_t \log \phi_0^t - (n - t - q_t) \log 1 - \phi_0^t) \quad (2)$$

Where $\log \text{lik}$ is the log-likelihood, v is the (effective) number of model parameters, t is the tree level, ϕ_0 is the a-priori probability of having a non-independent copula, and q_t is the number of independence copulas in the tree t and n is the dimension of the model. As expected, the model M1 (i.e. the one that could select all copula families implemented in the package), succeeds in modeling the dataset better. In fact, the $mBICv(M1) = -573,920.9$ while $mBICv(M2) = -220,452.6$ is significantly higher. On the other hand, the massive use of non-parametric copulas makes it more difficult to interpret, as can be seen from the contour plot in Fig. 1). In Table 1 we just report the first and the last trees outcomes although the output provides all trees involved in the PCC. The first tree shows that the two models coupled the same edges. However, none of the distributions selected by the algorithm adopted in M1 belongs to the elliptic family as imposed in M2. Moreover, for the categorical variable `area` the dummy transformation gives rise to a relevant differences in terms of Normalized bivariate copula contours plot as observable in Fig. 1. Indeed, the M2 (see Fig. 2) does not

Table 1. Table of the R-vine copula: First Tree (1) and Last Tree (7).

| Tree	Edge	Conditioned	Conditioning	Variable types	Copula type	Rotation	Copula parameters	df	τ	Log lik
R-vine copula associated with model M1										
1	1	2, 1		c, c	bb7	0	1.405, 14.165	2	0.855	106963.490
1	2	1, 6		c, d	joe	270	1.031	1	−0.017	4.829
1	3	6, 5		d, d	tll	0	[30 × 30 grid]	0	−0.068	4074.412
1	4	4, 5		d, d	bb8	90	7.997, 0.990	2	−0.779	6121.241
1	5	7, 5		d, d	bb8	90	7.996, 0.990	2	−0.779	2871.696
1	6	3, 8		d, d	joe	180	1.168	1	0.087	19.253
1	7	5, 8		d, d	tll	0	[30 × 30 grid]	0	−0.032	2936.725
7	1	2, 4	3, 7, 8, 5, 6, 1	c, d	t	0	−0.021, 2	2	−0.013	119.291
R-vine copula associated with model M2										
1	1	2, 1		c, c	t	0	0.982, 2.081	2	0.880	87891.573
1	2	1, 6		c, d	Gaussian	0	−0.028	1	−0.018	4.671
1	3	6, 5		d, d	t	0	−0.999, 5.0108	2	−0.999	3183.100
1	4	4, 5		d, d	t	0	−0.999, 5.001	2	−0.999	5488.691
1	5	7, 5		d, d	t	0	−0.999, 5.0199	2	−0.999	2251.425
1	6	3, 8		d, d	Gaussian	0	0.062	1	0.039	19.253
1	7	5, 8		d, d	t	0	−0.999, 5.026	2	−0.990	1332.590
7	1	2, 3	4, 7, 8, 5, 6, 1	c, d	Gaussian	0	0.009	1	0.005	1.736
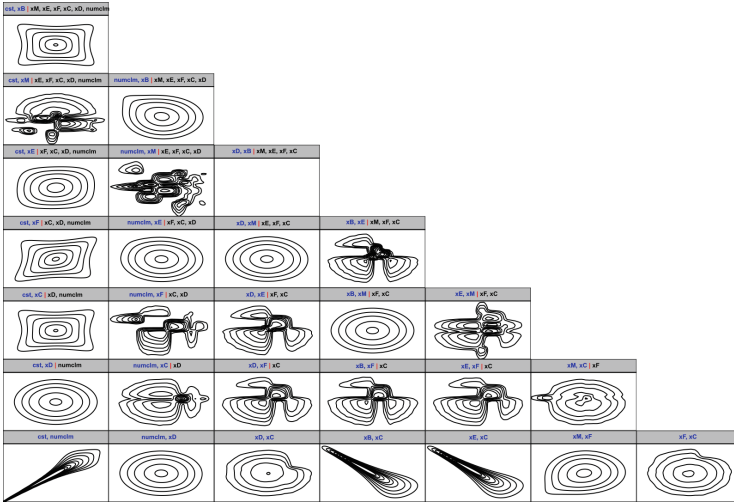
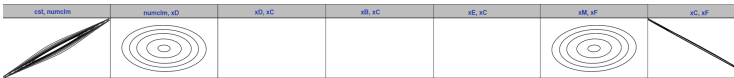**Fig. 1.** Normalized bivariate copula contours of Model 1

Fig. 2. Normalized bivariate copula contours of the first tree of Model 1

produce a plot when Kendall's tau is closed to -0.999 given the link with the basic parameter of the exponential family distribution, generally indicates as ρ. Finally, non parametric transformation kernel bivariate copulas, (indicated in Table 1 with the acronym "tll") often occurs in M1. As observable contour plots are very irregular (e.g. see plot in the third row and second column in Fig. 1) making it difficult to interpret the results obtained by the model in terms of dependence structure.

4   Conclusion and Further Research

The empirical analysis shows that qualitative variables are not always handled by the PCC algorithm in a "rational" manner. For example, in our study, the variable `area` is transformed into 5 dummy variables that are sometimes coupled together showing dependency shapes that seem to be not appropriate and, not secondarily, the approach is time consuming. Therefore, in our opinion, in case of categorical variable, the algorithm can be improved first by assuming a priori the dependency between dummy variables, generated by the same categorical variable, second by constructing a vine tree sequence from the regular vine matrix. In addition, a mixed tailored model could not only improve the fit of the model to the phenomenon but would also speed up the estimation process by skipping some steps of the algorithm to search for the best copula in the PCC.

References

1. Aas, K., Czado, C., Frigessi, A., Bakken, H.: Pair-copula constructions of multiple dependence. Insur. Math. Econ. **44**, 182–198 (2009)
2. Bedford, T., Cooke, R.M.: Monte Carlo simulation of vine dependent random variables for applications in uncertainty analysis. In: Proceedings of ESREL 2001, Turin, Italy (2001)
3. Bedford, T., Cooke, R.M.: Vines: a new graphical model for dependent random variables. Ann. Stat. **30**(4), 1031–1068 (2002)
4. Czado, C.: Pair-copula constructions of multivariate copulas. In: Jaworski, P., Durante, F., Härdle, W., Rychlik, T. (eds.) Copula Theory and Its Applications. Lecture Notes in Statistics, vol. 198, pp. 93–109. Springer, Heidelberg (2010). https://doi.org/10.1007/978-3-642-12465-5_4
5. Czado, C.: Analyzing Dependent Data with Vine Copulas: A Practical Guide With R (2019). https://doi.org/10.1007/978-3-030-13785-4
6. Dißmann, J., Brechmann, E.C., Czado, C., Kurowicka, D.: Selecting and estimating regular vine copulae and application to financial returns. Comput. Stat. Data Anal. **59**, 52–69 (2013)
7. Frees, E.: Regression modeling with actuarial and financial applications (2010)
8. Joe, H.: Families of m-variate distributions with given margins and $m(m-1)/2$ bivariate dependence parameters. In: Rüschendorf, L., Schweizer, B., Taylor, M.D. (eds.) Distributions with Fixed Marginals and Related Topics (1996)
9. Kurowicka, D., Cooke, R.: Uncertainty Analysis with High Dimensional Dependence Modelling. Wiley, Chichester (2006)
10. Morales-Nápoles, O.: Counting vines. In: Kurowicka, D., Joe, H. (eds.) Dependence Modeling: Vine Copula Handbook. World Scientific Publishing Co., Singapore (2011)
11. Nagler, T., Bumann, C., Czado, C.: Model selection in sparse high-dimensional vine copula models with application to portfolio risk. J. Multivar. Anal. **172**, 180–192 (2019)
12. Nagler, T., Vatter, T.: rvinecopulib: High Performance Algorithms for Vine Copula Modeling. https://cloud.r-project.org/web/packages/rvinecopulib/index.html
13. Nelsen, R.: An Introduction to Copulas. Springer, New York (2006)
14. Panagiotelis, A., Czado, C., Joe, H.: Pair copula constructions for multivariate discrete data. J. Am. Stat. Assoc. **107**(499), 1063–1072 (2012). https://doi.org/10.1080/01621459.2012.682850
15. Stöber, J., Czado, C.: Sampling pair copula constructions with applications to mathematical finance. In: Simulating Copulas, Stochastic Models, Sampling Algorithms, and Applications (2012)

New Insights on Loss Given Default for Shipping Finance: Parametric and Non-parametric Estimations

Aida Salko[1] and Rita D'Ecclesia[2(✉)]

[1] Department of Economics and Social Sciences, Sapienza University of Rome, Piazzale Aldo Moro 5, 00185 Rome, Italy
aida.salko@uniroma1.it
[2] Department of Statistics, Sapienza University of Rome, Piazzale Aldo Moro 5, 00185 Rome, Italy
rita.decclesia@uniroma1.it

Abstract. This study analyzes different parametric and non-parametric modeling methods for estimating the Loss Given Default (LGD) of bank loans for shipping companies. The shipping industry is subject to several risks which create the need to accurately measure the possible losses in order to estimate the LGD for the banking industry. We use a unique database of defaulted loans in European banks involved in shipping finance. The aim of this study is twofold: to compare the performance of alternative LGD modeling methodologies in shipping finance and to provide some insights into what drives LGD in the shipping industry. We find that non-parametric methods, especially random forest, lead to a remarkable increase in the prediction accuracy and outperform the traditional statistical models in terms of both in-sample and out-of-sample results. To investigate the risk drivers in the shipping business, we use a variable importance measure built on the idea of the permutation importance and find the energy index to be of paramount importance the most important risk factor. We find that crude oil prices play a big role and may affect the financial health of shipping firms and then the LGDs of shipping loans.

Keywords: Loss Given Default · Shipping finance · Global Credit Data

1 Introduction

The shipping industry is the leading mode of transportation worldwide and is considered the backbone of the global economy. The need for high levels of capital investment is one of the main indicators affecting the development of this sector. The debt capital provided by banks is considered historically the largest source and the most common way of financing vessels in the shipping industry [1,10,11]. Accurate estimates of potential losses are essential for financial institutions. The

Loss Given Default (LGD)concerning shipping finance is of great relevance for banks taking into account several risks associated with the sector, mainly driven by the very high volatility in oil prices and freight rates. However, it still remains an unexplored topic in the academic literature and the lack of data availability remains the main reason. We use a unique database of shipping finance loss data and investigate different parametric and non-parametric modeling methods to estimate and forecast the LGD for shipping finance. The aim of this study is to explore different approaches in estimating the LGD in shipping finance and to identify the main risk drivers of LGD in the shipping industry. To the best of our knowledge, this is the first study that investigates modeling methods for LGD of bank loans in shipping finance and provides new insights into what drives LGD in the shipping industry. The remainder of the paper is organized as follows. Section 2 briefly describes our dataset. Section 3 introduces the parametric and non-parametric methods that are used for the estimation as well as the metrics used for performance assessment. Section 4 and Sect. 5 report the empirical results and the concluding remarks.

2 Data

We use a unique loss database provided by Global Credit Data (GCD).[1] The data is of defaulted shipping borrowers and it provides us with information on: a) the defaulted borrower (residence, industry, financial health etc.); b) the characteristics of the ships serving as collateral (vessel type, size, valuation etc.), and (3) loan-related factors (exposure at default, seniority, facility type etc.). A full list of variables is available upon author's request. In our paper, we analyze 363 defaulted loans with a shipping collateral whose country of jurisdiction is located in European countries. LGD[2] is given as one minus the recovery rate where the last is the difference between the discounted incoming cash flows and discounted direct and indirect costs, divided by the exposure at default (EAD).
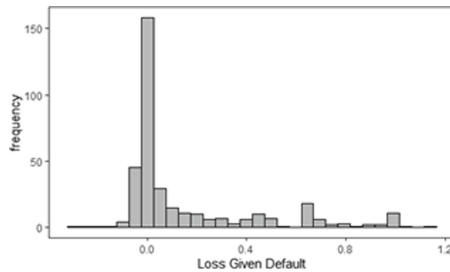
Fig. 1. LGD distribution

---

[1] The GCD association consists of 55 member banks from all over the world and the data collected comes across the span of 20 years https://globalcreditdata.org/.

[2] In this study we use the loan-level LGD. The economic LGD calculation is used where principal advance and financial claim are parts of the recovered amount.

We refer to the default definition [3] and restrict our data sample from the year 2000 to ensure a consistent default definition. In addition, we do not account for defaults after 2019 since the workouts of recent defaults may not be necessarily completed.[3]. Finally, we include the GDP growth, the uncertainty index and the energy commodity index as macroeconomic factors so the models are sensitive to macroeconomic characteristics [7]. As Fig. 1 indicates, there is the presence of highly left-skewed LGDs. In our sample, the lowest LGD is -28.61% while the mean is given by 15.54%.

3   Methodology and Performance Assessment

We consider four parametric models in our study, a simple OLS regression, ridge regression [12], least absolute shrinkage selector operator (LASSO) regression [14], and elastic net (EN) regression [15]. Ridge, LASSO and EN are a modification of simple OLS and the main idea behind these regression models stands in shrinking the regression coefficients by imposing a penalty on their size to mitigate the overfitting. When this penalty factor is equal to zero, we get the standard OLS regression. In addition, we include five non-parametric methods including bagging (BG) [4], random forest (RF) [5], boosting (BS) [9], artificial neural networks (ANN) [2], and multivariate adaptive regression splines (MARS) [8]. Evaluating the predictive accuracy of the models is an essential part of the study so we use the root mean squared error (RMSE) and the mean absolute error (MAE) as the most commonly used measures of model performance. Since we are interested to assess the RMSE and MAE on a sample that is independent of that used in building the models, we will split our sample into two sets using a standard 70% (training set) - 30% (test set) random split. In addition, models hyperparameters were tuned by using ten-fold cross-validation on the training set. All the models are trained using the latest version of the Caret library in RStudio [13].

4   Results

The performance matrices of in-sample and out-of-sample results are presented in Table 1. We also add R^2 as the most intuitive measure of explanatory power. Regarding errors between the realized and forecasted LGDs, we find that non-parametric methods produce the best forecasting results, and outperform parametric methods in terms of both in-sample and out-of-sample results. Non-parametric methods exhibit a proportion of explained variation in terms of R 2 measure, ranging from 44.20% to 81.15 % for in-sample results and from 14.31% to 55.76% for out-of-sample results. Additionally, we observe that the random forest algorithm is superior to all the other methods presenting not only the highest proportion of explained variation of 81.15% (in-sample) and 55.76% (out-of-sample) but also the lowest errors. On the other hand, we observe that parametric

---

[3] The resolution bias is addressed according to GCD methodology: https://www.globalcreditdata.org/syst.

models are associated with a weaker performance in terms of predictive accuracy and explained variation.

Table 1. Performance matrix

Method	In-sample performance			Out-of-sample performance		
	RMSE	MAE	R^2	RMSE	MAE	R^2
OLS	0.2149 (5)	0.1447 (5)	0.4319 (6)	0.2922 (9)	0.2002 (8)	0.1064 (9)
Ridge	0.2512 (9)	0.1887 (9)	0.3106 (9)	0.2752 (8)	0.2059 (9)	0.1104 (8)
LASSO	0.2364 (7)	0.1676 (7)	0.3453 (7)	0.2671 (6)	0.1907 (6)	0.1313 (6)
EN	0.2404 (8)	0.1761 (8)	0.3270 (8)	0.2696 (7)	0.1959 (7)	0.1148 (7)
BG	0.1943 (3)	0.1273 (3)	0.5712 (3)	0.2598 (5)	0.1584 (3)	0.2461 (3)
RF	0.1265 (1)	0.0776 (1)	0.8115 (1)	0.1777 (1)	0.1213 (1)	0.5576 (1)
BS	0.1728 (2)	0.1168 (2)	0.6908 (2)	0.2457 (3)	0.1473 (2)	0.3321 (2)
ANN	0.2174 (6)	0.1646 (6)	0.4552 (4)	0.2499 (4)	0.1809 (5)	0.1431 (5)
MARS	0.2106 (4)	0.1431 (4)	0.4420 (5)	0.2286 (2)	0.1658 (4)	0.1901 (4)

Note: The numbers in brackets state the ranks of the models in terms of performance measures. The ranks range from 1 (best) to 9 (worst).

4.1   Variable Importance

We investigate the importance of LGD risk drivers in our models mainly to identify the factors which drive the LGD dynamics. We generate a visual comparison of all the input variables for every model by constructing a measure built on the idea of the permutation importance [5]. Figure 2 presents variable importance rankings for all the non-parametric methods that best performed in our study. The variables are ranked by their importance in a decreasing order and the results reveal some important insights into what drives LGD of shipping finance. The most important variables considered by all the non-parametric methods are a mixture of macroeconomic indicators and loan file information. All these algorithms generally agree in their decisions and rank the energy index and the jurisdiction country as the main input variables in forecasting LGDs of shipping loans. Considering only the rankings of the random forest model, that best performed across all models, we observe that the energy index is the first most important input variable, followed by the country of jurisdiction, and the uncertainty index. This result highlights the significant role of the macroeconomic environment in the shipping industry.
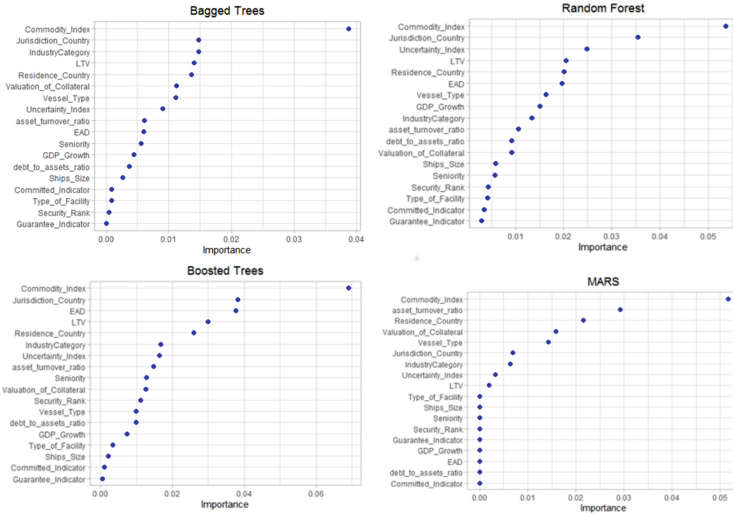
Fig. 2. Variable importance in non-parametric methods

4.2 Effect of Energy Index in Model Prediction

The most important risk driver of LGD in shipping finance. To further investigate the relationship between model forecasts and this risk driver, we use the Partial Dependence plots (PDP) [9], a popular tool used in the field of explainable machine learning. They are calculated after the model has been fit and attempt to visually explain what the model predicts on average when the value of the feature changes. Figure 3 (left) reports the dynamic of the energy commodity index which shows its peak in correspondence of the Global Financial Crisis (GFC) and then shows a very volatile dynamic from 2010 onward. Figure 3 (right) reports the PDP of this index. As it can be seen, the highest peak of the index (around 170) is achieved during GFC, the period which was also associated with the highest losses for the banks. Given that crude oil presents an 84.6% share of the energy index, the results are obviously driven by the price of crude oil which plays an essential role in the global economy and particularly in the shipping industry. The model suggests that as the index is increasing, up to the value of 120, there is a negative downtrend, meaning that it is associated with lower losses for the bank. This may be explained by a positive relationship between an increase in oil prices due to increase demand from oil, and shipping stock returns [6]. However, when it goes above 120, which is the value of the energy index during GFC, a sharp increase of LGD occurs. This may be explained by the fact that when the oil prices increase and therefore the energy index, shipping companies experience losses due to higher operating costs and this may lead to higher number of defaults and higher LGDs. We find that large volatility of crude oil prices affects LGDs in two directions: at first when prices increase due to an initial higher demand of crude oil, shipping companies benefit from a first

increase of business, but when oil prices keep on increasing, the positive effect on shipping companies revenues is offset by the increase in operating costs.
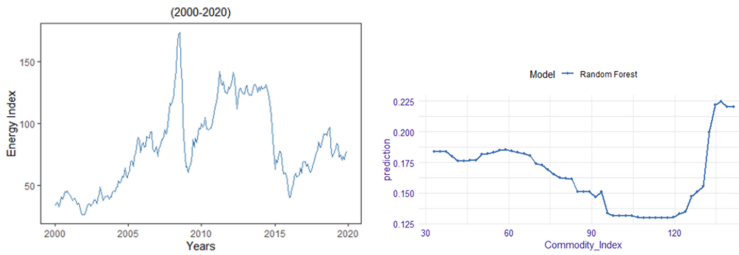
Fig. 3. Energy index in years (left) and PDP of energy index (right)

5   Conclusion

This study was based on the shipping finance loss data and we compared different parametric and non-parametric models that attempt to predict the LGD for shipping finance. We find that non-parametric methods lead to a remarkable increase in the prediction accuracy and outperform the traditional statistical models in term of both insample and out-of-sample results. The random forest model stood out as having the best forecasting performance among all the models. Furthermore, we use a variable importance measure built on the idea of the permutation importance, to analyze the risk drivers with the greatest effects on the LGD for shipping finance prediction accuracy for each method. We observe that all the methods consider the energy index, which is mainly driven by the crude oil prices, as the main input variable in forecasting LGDs of shipping loans. Furthermore, using PDP, we found that the model captures a positive signal in terms of higher expectation of losses, as the volatility of the crude oil market increases sharply, as it happened during the last financial crisis. The result highlights the dominant role played by crude oil prices which can deteriorate the financial health of shipping firms and therefore affect the LGDs of shipping loans. Other inputs such as the freight rates can be considered in LGD modeling of shipping finance for further research.

References

1. Albertijn, S., Bessler, W., Drobetz, W.: Financing shipping companies and shipping operations: a risk-management perspective. J. Appl. Corp. Financ. **23**(4), 70–82 (2011)
2. Bandy, H.: Developing a neural network system. Technical report, AAII Working Paper (1997)
3. Basel, I.: International convergence of capital measurement and capital standards: a revised framework. Bank for international settlements (2004)

4. Breiman, L.: Bagging predictors. Mach. Learn. **24**(2), 123–140 (1996)
5. Breiman, L.: Random forests. Mach. Learn. **45**(1), 5–32 (2001)
6. El-Masry, A.A., Olugbode, M., Pointon, J.: The exposure of shipping firms' stock returns to financial risks and oil prices: a global perspective. Marit. Policy Manag. **37**(5), 453–473 (2010)
7. European Banking Authority: Consultation paper: on guidelines for the estimation of LGD appropriate for an economic downturn ('downturn LGD estimation') (2018)
8. Friedman, J.H.: Multivariate adaptive regression splines. Ann. Stat. **19**(1), 1–67 (1991)
9. Friedman, J.H.: Greedy function approximation: a gradient boosting machine. Ann. Stat. 1189–1232 (2001)
10. Grammenos, C.T.: Revisiting credit risk, analysis and policy in bank shipping finance. In: The Handbook of Maritime Economics and Business, pp. 807–840. Informa Law from Routledge (2013)
11. Harwood, S.: Shipping Finance. Euromoney Books (2006)
12. Horel, A.: Applications of ridge analysis toregression problems. Chem. Eng. Progress. **58**, 54–59 (1962)
13. Kuhn, M.: Caret: classification and regression training. R package version 6.0-80 (2018)
14. Tibshirani, R.: Regression shrinkage and selection via the lasso. J. Roy. Stat. Soc.: Ser. B (Methodol.) **58**(1), 267–288 (1996)
15. Zou, H., Hastie, T.: Regularization and variable selection via the elastic net. J. R. Stat. Soc.: Ser. B (Stat. Methodol.) **67**(2), 301–320 (2005)

Real R&D Options Under Sentimental Information Analysis

Domenico Santoro[(✉)] and Giovanni Villani

Department of Economics and Finance, Largo Abbazia Santa Scolastica,
University of Bari, 70124 Bari, BA, Italy
{domenico.santoro,giovanni.villani}@uniba.it

Abstract. The problem of understanding how to modify the probability of success for a stage in an R&D project is still open. Primarily in cases where it is impossible to compare a project with other competitors, the probability of passing a certain phase of the experimentation is determined by taking into account only information from within the company and not from external information.

In this paper, we propose to use Natural Language Processing techniques to obtain a sentiment score for the news from the outside world. In this way, we can transform sentences expressed in natural language into a numerical value which, in addition to the internal information, allows us to better "direct" the probabilities of success in a stage.

Keywords: Real options · Sentiment analysis · Information revelation

1 Introduction

In the various stages of a project's development, information from external sources plays an important role. In particular, information expressed in natural language is fundamental (such as news in newspapers or information leaks). The branch of Artificial Intelligence that deals with the study of natural language is *Natural Language Processing* (NLP). Among the different tasks that can be pursued through the study of words, there is also the sentiment analysis, which allows to determine the polarity of a sentence (positive/negative) and indicate a sentiment score. In the academic literature, several papers have emphasized the importance of uncertainty. For example, Dixit [7] assumes a constant hazard rate, Weeds [14] defines how a player should make an irreversible decision, Cassimon et al. [3] analyze multi-stage R&D in the pharmaceutical sector, Kellogg and Charnes [9] consider the technical uncertainty of drug projects, or D'Amico and Villani [4] consider the probabilities of success as generated by a Markov chain. The literature has in common that these probabilities of success in each phase of the R&D process are determined relying on internal analyzes of the company and reshaped in various ways by taking into account two types of risks. At the

same time, what we want to introduce in this paper is the possibility of updating these probabilities based on news coming from sources outside the company, especially in the case of projects that are not comparable with others.

Dias [6] defined a mathematical framework for considering uncertainty in R&D project evaluation, creating a metric that simplifies investment decisions under technical uncertainty. With this in mind, the objective of this paper is to use the previous framework by introducing the sentiment score determined through NLP models as a learning measure, with some adaptations.

2 AlBERTino for Sentiment Analysis

Over time, NLP models have evolved. One of the newer models, Bidirectional Encoder Representations from Transformers (BERT [5]), was born from the combination of the best elements of its predecessor models, such as Embeddings from Language Models (ELMo) and Generative Pre-trained Transformer (GPT). In particular, this new model based on an encoder-decoder network that uses the transformers architecture (developed by Vaswani et al. [13]) solves the problems of its predecessors, such as encoding context bidirectionally and requesting minimal changes for various NLP tasks. BERT consists of a set of transformer encoders which perform two fundamental tasks: *Masked Language Modeling* (Masked LM), in which it defines the language modeling by randomly masking 15% of the tokens being in the corpora and the *Next Sentence Prediction*, that is the ability to predict if two sentences follow each other.

In our hypothesis, we are interested in evaluating projects for which there are not (or are not very representative) probabilities of success in the different phases. In this sense, the information obtained from outside the company has a significant impact. For example, we can consider news from the economic/financial world that are interdisciplinary to different sectors in most cases. Aract [1] was the first to develop a BERT model for sentiment classification and regression about that kind of information. In particular, BERT's training was carried out by using three datasets: the pre-training was carried out through *TRC2-financial*, filtering the *TRC2* corpus based on financial keywords (obtaining a dataset consisting of over 29M words and 400k phrases); *FinancialPhrase-Bank*, the main sentiment analysis dataset [10] and *FiQA Sentiment*, a dataset created for the WWW 18 conference challenge. This model, called by the authors FinBERT, despite failing in some cases (as demonstrated in cases where it has difficulty in distinguishing phrases generally used in the business environment from positive ones), turns out to be the best compared to other state-of-the-art models and allows to give great decision support for investors.

However, our goal is to use a BERT model that considers the facets of the Italian language to use the language very often used by journalists rich in technical/financial terminology. For this reason, we used **AlBERTo**, a BERT model developed by Polignano et al. [12] that has been trained on the Italian language. This model, which was trained on the TWITA [2] dataset, contains a large corpus of Tweets (191 GB) recorded from 2012 to 2018. As the authors define,

AlBERTo focuses mainly on the Masked LM task leaving out the Next Sentence Prediction. The model was tested on the EVALITA 2016 dataset (in particular on the SENTIPOLC - SENTIment POLarity Classification task), demonstrating accuracy in the prediction of positivity and negativity, measured by F1 (an indicator for accuracy on the test set based on precision and recall), equal to 0.72. Given the remarkable capacity of this model in the different tasks, such as polarity and irony detection, we have carried out a fine-tuning of this model based on the *FinancialPhrasebank* dataset collected by Malo et al. [10] appropriately translated into Italian. This dataset contains about 4,000 sentences relating to financial news with the annotation of positivity or negativity for each of them. In this way, it was possible to build an Italian model focused on news from the financial world, **AlBERTino**, by adding a layer on AlBERTo trained on this translated dataset.

3   Probability of Success in R&D Stages

As defined by Cassimon et al. [3] (in the pharmaceutical case, but extendable to other types of projects), we can divide the realization of a project into different stages, each of which is characterized by its own probability of success p_i. We can also assume that the various probabilities p_i are independent of each other in the various i-th stages. For example, in Fig. 1 there is a representation of an R&D project divided into two stages. Let us consider the division into only two stages in this test phase to study the functioning of sentiment analysis in research projects, as defined by Perlitz et al. [11] and Jensen and Warren [8].

Since our aim is to incorporate the sentiment score defined by *AlBERTino* in the various success probabilities, we can start from the framework defined by Dias [6]. In particular, in each state i we can study the dependence relationship between two Bernoulli random variables: X, a variable with technical uncertainty that is important for investment decisions; and S, also called *binary signal*, representing events that make it possible to determine the success or failure of a certain action. For example, let us consider the development project of a new biomaterial as a real-world application. Then, we can focus on the pre-production stage of studying the constituent element relationships. For example, we could say that success is described by a Bernoulli variable $X \sim Be(p)$ and that a signal $S \sim Be(q)$ could be the recently discovered news of the possible scarcity of one constituent selected element. We can thus introduce the element of *information revealed*, that is, the impact of the news (or the signal S) on the probability of success p in that i-th phase. Dias [6] defines a learning measure $\eta^2(X|S)$ which links the two Bernoulli distributions, so that the probability p evolves according to the signal S in two ways:

$$p^+ = P[X = 1|S = 1] = p + \sqrt{\frac{1-q}{q}}\sqrt{p(1-p)}\sqrt{\eta^2(X|S)} \qquad (1)$$

$$p^- = P[X = 1|S = 0] = p - \sqrt{\frac{q}{1-q}}\sqrt{p(1-p)}\sqrt{\eta^2(X|S)} \qquad (2)$$
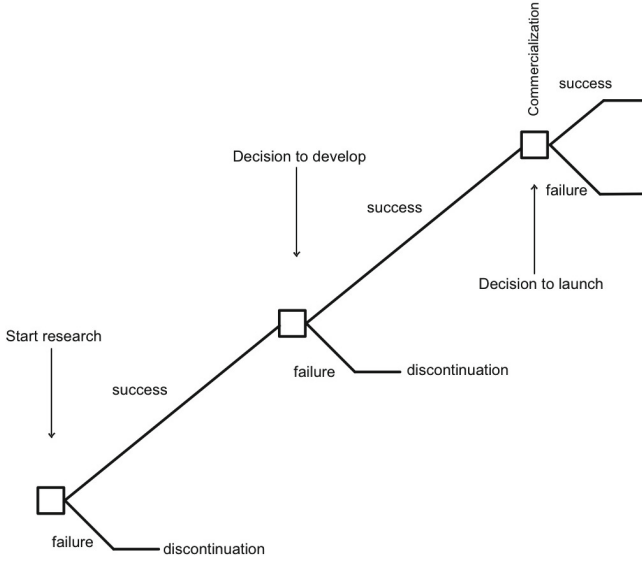
Fig. 1. Structure of $n = 2$ compound option as decsribed by Cassimon et al. [3] and D'Amico and Villani [4].

in case of positive dependence; in case of negative dependence by inverting the sign after p. It is precisely in this case that the determined sentiment score intervenes: the possibility of transforming sentences in natural language into a polarity value allows to obtain a score $\gamma \in [-1, 1]$ for *positive* sentences $\gamma \in [0, 1]$, while for *negative* sentences $\gamma \in [-1, 0]$. Through manipulations, we can make sure that even in the case of negative polarity $\gamma \in [0, 1]$, so that γ responds to the axioms defined by Dias [6] (in particular the Fréchet-Hoeffding bound and the equality of marginal distribution) and can be used as a learning measure $\eta^2 = \gamma$. Thanks to these assumptions, we can determine the new probabilities revealed only in the case of positive dependence, such as:

$$p^+ = p + (1 - p)\sqrt{\gamma} \tag{3}$$
$$p^- = p - p\sqrt{\gamma}. \tag{4}$$

If the polarity is positive, the new probability of success revealed will be p^+, and if the polarity is negative, it will be p^-. In this way, returning to developing a new biomaterial, the news (learned, for example, from a newspaper) of the possible scarcity would lead to a low probability of success in that stage. Thus, having a numerical value related to success, the company can decide in advance whether to continue the experimentation or modify the constituent elements of the material. All this allows us to keep the assumptions of Dias [6] valid in determining the revelation process in the presence of multiple S signals, but it benefits us because, thanks to the polarity, we can follow a single path of the tree.

4 Conclusions

In this paper, we propose the use of an NLP model, *AlBERTino*, to determine the sentiment score of some news that can modify the probability of success in the different stages of an R&D project. In particular, we have improved the AlBERTo model on the sentiment analysis of Italian sentences to focus on the economic world, which generally allows information to be extracted from heterogeneous sectors. Furthermore, this sentiment score, based on certain assumptions and axioms, was used as a revealed information tool for a certain stage of the project and therefore made it possible to improve the probability of overcoming that phase, taking into account information coming from the outside. The next objective is to verify how we can improve a probability based on different signals occurring in one stage, determine a revelation process, and numerically test this model with real information.

References

1. Aract, D.: FinBERT: financial sentiment analysis with pre-trained language models. arXiv:1908.10063v1 (2019)
2. Basile, V., Lai, M., Sanguinetti, M.: Long-term social media data collection at the university of Turin. In: Fifth Italian Conference on Computational Linguistics (CLiC-it 2018), pp. 1–6 (2018)
3. Cassimon, D., De Backer, M., Engelen, P., Van Wouwe, M., Yordanov, V.: Incorporating technical risk in compound real option models to value a pharmaceutical r&d licensing opportunity. Res. Policy **40**(9), 1200–1216 (2011)
4. D'Amico, G., Villani, G.: Valuation of R&D compound option using Markova chain approach. Ann. Finance **17**, 379–404 (2021)
5. Devlin, J., Chang, M.W., Lee, K., Toutanova, K.: BERT: pre-training of deep bidirectional transformers for language understanding. In: Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, vol. 1 (Long and Short Papers), pp. 4171–4186. Association for Computational Linguistics (2019)
6. Dias, M.A.G.: Real options, learning measures, and Bernoulli revelation processes. In: Proceedings of the 9th Annual International Conference on Real Options, pp. 1–41. Real Options Group, Paris, June 2005
7. Dixit, A.: A general model of R&D competition and policy. RAND J. Econ. **19**(3), 317–326 (1988)
8. Jensen, K., Warren, P.: The use of options theory to value research in the service sector. R&D Manag. **31**(2), 173–180 (2008)
9. Kellogg, D., Charnes, J.M.: Real-options valuation for a biotechnology company. Finan. Anal. J. **56**(3), 76–84 (2000)
10. Malo, P., Sinha, A., Korhonen, P., Wallenius, J., Takala, P.: Good debt or bad debt: detecting semantic orientations in economics texts. J. Assoc. Inf. Sci. Technol. **65**(4), 782–796 (2014)
11. Perlitz, M., Peske, T., Schrank, R.: Real options valuation: the new frontier in R&D project evaluation? R&D Manag. **29**(3), 255–270 (2002)

12. Polignano, M., Basile, P., de Gemmis, M., Semeraro, G., Basile, V.: AlBERTo: Italian BERT language understanding model for NLP challenging tasks based on tweets. In: Proceedings of the Sixth Italian Conference on Computational Linguistics (CLiC-it 2019), vol. 2481. CEUR (2019)
13. Vaswani, A., et al.: Attention is all you need. In: Advances in Neural Information Processing Systems, pp. 5998–6008, Morgan Kaufmann Publishers Inc., San Francisco (2017)
14. Weeds, H.: Strategic delay in a real options model of R&D competition. Rev. Econ. Stud. **69**(3), 729–747 (2002)

A Multi-population Locally-Coherent Mortality Model

Salvatore Scognamiglio[(✉)]

Department of Business and Quantitative Studies, University of Naples Parthenope,
Naples, Italy
`salvatore.scognamiglio@uniparthenope.it`

Abstract. This paper proposes a simple and fully-interpretable neural network model for multi-population mortality modelling and forecasting. The architecture is designed to be interpretable in the Lee-Carter framework and induces a massive reduction of the parameters to optimise. The model structure leads the creation of clusters of countries with similar mortality patterns during the fitting procedure highlighting differences and commonalities among the clusters. Numerical experiments performed on the Human Mortality Database Data show that the proposed model produces reliable estimates and very accurate forecasts.

1 Introduction

Accurate modelling and forecasting are important for social welfare policies and resource budgeting among the government and insurance industry sectors. With this aim, several contributions were proposed in the literature. The Lee-Carter model proposed in [6] represents a milestone. It decomposes logarithmic death rates into an age-specific base level and a time-varying component (period effect) multiplied by an age-modulating parameter (age effect). Forecasts are obtained by projecting the time-dependent index with classical time-series models. Several extensions of the LC model have been proposed. [1] improved the Ordinary Least Squared estimation approach of the classical LC method by modelling the number of deaths directly by a Poisson distribution and employing maximum likelihood for parameter estimation. [11] explored multi-factor extensions of the LC model while [12] extended the model incorporating cohort effects. Even though these approaches often describe the mortality of a single population well, there are cases in which it could be necessary to simultaneously model the mortality of multiple populations [2]. To this end, the researchers started exploring multi-population models. One of the most famous models is the Li and Lee (LL) model proposed in [8], which provides a double log-bilinear mortality model augmenting common age and period effects with subpopulation-specific age and period effects. An attractive property of this model is to ensure long-term not-divergent forecasts among the populations, which could be not biologically reasonable. An attractive property of this model called is to ensure long-term

not-divergent forecasts among the populations, which could be not biologically reasonable. Despite the coherent mortality forecasting has been investigated in other papers such as [4,5], some authors highlights that the coherent assumption could be suitable for specific populations and over limited time windows [7]. Recently, the application of neural networks to mortality modelling and forecasting is becoming popular. [3] used a neural analyser for detecting and forecasting non-linearities of the log-mortality in a single-population setting. [9] introduced the Long Short Term Memory networks to model the dynamics of the LC time-index into the classical two-stage approach. [13] developed a neural network model based on fully connected and embedding layers for large-scale mortality forecasting, and [10] further extended their model by introducing Recurrent and Convolutional Neural Networks. This paper introduces a simple and fully-interpretable neural network model for large-scale mortality modelling and forecasting. The architecture is designed to be easily interpretable induces a substantial reduction in the number of parameters to estimate. Some homogeneous clusters are formed within the model fitting, highlighting differences and commonalities among populations. In addition, numerical experiments show that the model produces very accurate forecasts.

2   Multi-population Mortality Modelling

Let $X = \{x_0, x_1, \ldots, x_\omega\}$ be the set of the age categories, $T = \{t_0, t_1, \ldots, t_n\}$ be the set of calendar years and $I = R \times G = R \times \{\texttt{male}, \texttt{female}\}$ be the set of the populations considered (where R is the set of the countries). The LC model defines the logarithm of the central death rate $\log m_{x,t}^{(i)} \in \mathbb{R}$ at age $x \in X$ in the calendar year $t \in T$ in the population $i = (r, g) \in I$ as

$$\log m_{x,t}^{(i)} = a_x^{(i)} + b_x^{(i)} k_t^{(i)} + e_{x,t}^{(i)}, \tag{1}$$

where $a_x^{(i)} \in \mathbb{R}$ is the average age- and population-specific pattern of mortality, $b_x^{(i)} \in \mathbb{R}$ represents the age- and population-specific patterns of mortality change and indicates the sensitivity of the logarithm of the force of mortality at age x to variations in the time index $k_t^{(i)}$, $k_t^{(i)} \in \mathbb{R}$ is the population-specific time index describing mortality trend and $e_{x,t}^{(i)} \in \mathbb{R}$ is the error term. Since the model in is over-parameterised, the following constraints are imposed

$$\sum_{x \in X} b_x^{(i)} = 1 \qquad \sum_{t \in T} \frac{k_t^{(i)}}{|T|} = 0. \tag{2}$$

To obtain forecasts, the LC model assumes that the parameters $(a_x^{(i)})_x$ and $(b_x^{(i)})_x$ are constant over time while the time indices $k_t^{(i)}$ are modelled as random walk with drift (RWD)

$$k_t^{(i)} = k_{t-1}^{(i)} + \theta^{(i)} + e_t^{(i)} \qquad with\ i.i.d,\ e_t^{(i)} \sim N(0, (\sigma_\epsilon^{(i)})^2) \tag{3}$$

where $\theta^{(i)} \in \mathbb{R}$ is the population-specific drift term. The application of individual Lee-Carter models for each population can produce divergent long-term predictions while it might be reasonable to assume that the differences in mortality between these populations should not increase over time indefinitely if they share similar socioeconomic conditions. To avoid long-run divergence, Li and Lee proposed a model where all the populations share the same age and period ($b_x^{(i)} = B_x \in \mathbb{R}$ and $k_t^{(i)} = K_t \in \mathbb{R}, \forall i \in I$) with additional population-specific age and period effects to improve fitting and forecasting. The full LL model reads:

$$\log m_{x,t}^{(i)} = a_x^{(i)} + B_x K_t + b_x^{(i)} k_t^{(i)} + e_{x,t}^{(i)}, \tag{4}$$

where K_t is a RWD and $k_t^{(i)}$ evolves as a first-order autoregressive model (AR(1)).

3   A Multi-population Locally-Coherent Mortality Model

The coherent modelling prevents diverging long-term forecasts, which do not seem biologically reasonable. However, the coherence assumption may be perceived as too strong and is not always supported by empirical observations. This paper proposes a mortality model that relaxes the assumption of coherent modelling by requiring it to work only within sub-groups of similar populations. This property can be defined as local-coherence, and we define our model as Locally-Coherent Neural Network (LCNN) mortality model. The model assumes that a small set of latent factors $\kappa_t = (\kappa_t^{(1)}, \kappa_t^{(2)}, \ldots, \kappa_t^{(P)})' \in \mathbb{R}^P$ driving the mortality of groups of countries exists, where $P < |I|$. Each latent factor $\kappa_t^{(j)}$ describes the mortality of the countries belonging to the same j-th cluster. In the case that the assumption $b_x^{(i)} = B_x$ still holds, the model can be written as:

$$\log m_{x,t}^{(i)} = a_x^{(i)} + B_x \kappa_t^{c(r)} + e_{x,t}^{(i)}, \tag{5}$$

where $c : R \mapsto \{1, 2, \ldots, P\} \subset \mathbb{N}$ is a surjective function which assigns to every population $r \in R$ a cluster number $c(r) \in \{1, \ldots, P\}$ and $C^{(j)} := c^{-1}(\{j\}), j = 1, \ldots, P$ denote the set of countries belonging of the j-th clusters. The model in (5) can be also written as

$$\log \mathbf{m}_t^{(i)} = \mathbf{a}^{(i)} + \mathbf{B} \cdot \left\langle \gamma^{c(r)}, \kappa_t \right\rangle + e_{x,t}^{(i)} \qquad \forall i \in I \tag{6}$$

where $\log \mathbf{m}_t^{(i)} = \{\log m_{x,t}^{(i)}\}_{x \in X}$, $\mathbf{a}^{(i)} = \{a_x^{(i)}\}_{x \in X}$, $\mathbf{B} = \{B_x\}_{x \in X}$, $\langle \cdot, \cdot \rangle$ is the scalar product and $\gamma^{c(r)} = (\gamma_1^{c(r)}, \gamma_2^{c(r)}, \ldots, \gamma_P^{c(r)})' \in \{0, 1\}^P$ with

$$\gamma_l^{c(r)} = \begin{cases} 1 & l = c(r) \\ 0 & l \neq c(r) \end{cases} \quad l = 1, \ldots, P. \tag{7}$$

We perform the model fitting in a neural network setting. A network architecture replicating the model structure in (6) is provided and each model term is approximated through a dedicated subnet. More in detail:

- $\mathbf{a}^{(i)}$ can be considered a function of i and is approximated by a neural network composed of two embedding layers and a fully connected layer.
- $\gamma^{c(r)}$ determines the cluster to which the country r belongs and is function of r. The subnet that approximate this term consists of embedding layer and a custom layer that forces the output vector to have a single non-zero element equal to 1.
- The vector κ_t summarises the mortality of all the clusters and it is assumed as a function of the set of log-mortality curves $\log \mathbf{m}_t^{(i)}, i \in I$. It is extracted using a subnet with two fully connected layers. The first 1-dimensional layer is applied to the log-mortality curves individually and compresses each curve to a single value. The second P-dimensional layer processes these values simultaneously to extract latent factors.
- A 1-dimensional fully layer further processes the outcome $\left\langle \gamma^{c(r)}, \kappa_t \right\rangle \in \mathbb{R}$. The weight vector of this layer represents the age-specific coefficients B_x.

The sum of $\mathbf{a}^{(i)}$ and $B_x \cdot \left\langle \gamma^{c(r)}, \kappa_t \right\rangle$ gives an estimate of the mortality curve $\widehat{\log \mathbf{m}_t^{(i)}} \in \mathbb{R}^{|X|}$. The coefficients of these layers are calibrated simultaneously in order to minimise a given loss function.

4   Results

In this section, we perform some numerical experiments using the data of the Human Mortality Database [14]. We set $X = \{x \in \mathbb{N} : 0 \leq x < 100\}, T = \{t \in \mathbb{N} : 1960 \leq t < 2021\}$ and I contains male and female populations of the following countries: USA, RUS, JPN, DEUTWN, FRATNP, ITA, GBRTENW, UKR, ESP, POL, CAN, AUS, NLD, DEUTE, HUN, PRT, BLR, CZE, BEL, SWE, AUT, BGR, CHE, SVK, DNK, FIN, GBR_SCO, NOR, IRL, LTU, NZL_NM, LVA, GBR_NIR, EST, LUX and ISL (we observe $|I| = |R| \cdot |G| = 36 \cdot 2 = 72$). We considered all populations for which data from 1960 onwards were available in the HMD. The set T is splitted in $T_1 = \{t \in T : t < 2000\}$ and $T_2 = \{t \in T : t \geq 2000\}$ such that $T_1 \cup T_2 = T$. The data relating to the first set of years are used for the model calibration while the data of the second part are employed to measure the forecasting accuracy. Let $P = 4$, the LCNN network model is trained for 2000 epochs with the aim of minimising the Mean Squared Error (MSE). Figures 1 and 2 show the estimates of the model parameters. Figure 2 presents the estimations of the $(a_x^{(i)})_x$ for all populations considered $i \in I$. We observe that the curves are pretty smooth, suggesting that the estimates produced by the network are few sensible to the fluctuations often found in mortality data. Figure 2 (left) illustrates the estimates of $(B_x)_{x \in X}$ which is the same for all populations which appears pretty smooth as well while, Figure 2 (right) also shows the extracted $\kappa_t^{(j)}$ factors relating to the different clusters of populations. We observe that the factors present a decreasing trend highlighting that mortality is generally decreasing. However, the four factors have different rates of decline, showing that the mortality of some countries is falling faster than others. Interestingly, we observe that the factor relating to Cluster 4 declines rapidly.

This cluster encompasses countries such as ITA, JPN, ESP and SWE. The factor of Cluster 3 shows a less pronounced reduction than Cluster 4. In addition, it appears similar to Cluster 2. These two groups include countries that have observed reductions in mortality on average compared to the group of countries considered. Cluster 3 includes DEUTE, DEUTW, GBR_NIR GBR_SCO, while Cluster 2 includes AUT, FIN, FRATNP, USA. The factor of Cluster 1 shows a fluctuating dynamics, highlighting both increases and decreases in mortality over time. This cluster contains countries where mortality evolution is less regular than the other countries in the set, such as BLR, EST, LTU, LVA, RUS, and UKR. Finally, we compare the proposed model against some benchmarks in terms of forecasting accuracy and number of parameters to optimise. Table 1 lists the MSE forecasting and the number of parameters of our LCNN model and the LC_SVD [6], the LC_Poisson [1], the LL [8] models. We observe that all models perform pretty well, but the LCNN model outperforms the others. In addition, we also note that the LCNN model is the most parsimonious. In future works we intend to analyse in more detail similarities and differences between the clusters obtained and to investigate different neural network architectures for the extraction of the factors κ_t.
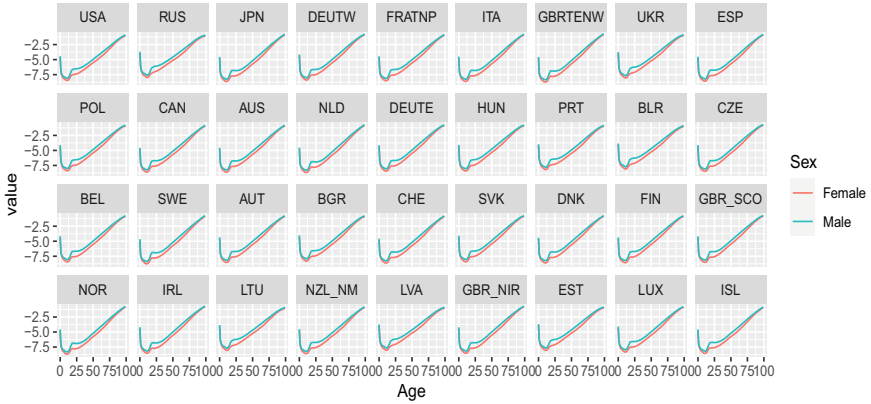
Fig. 1. $(a_x^{(i)})_{x \in X}$ estimates for all the populations considered; fitting period 1960–1999; countries are sorted by population size in 2000.
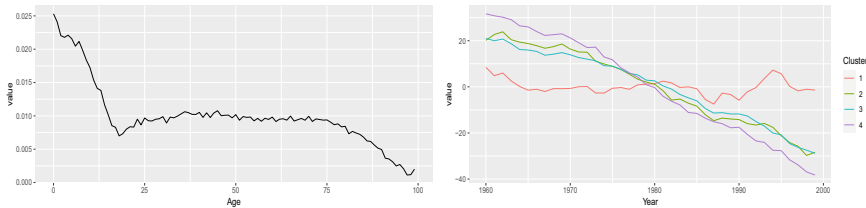
Fig. 2. $(B_x)_{x \in X}$ estimates (left); $(k_t^{(j)})_{t \in T}$ estimates for the four clusters (right); fitting period 1960–1999.

Table 1. Comparisons of the LC_SVD, LC_Poisson, LL and LCNN model: forecasting MSE and number of parameters.

Model	LC_SVD	LC_Poisson	LL	LCNN
MSE	.4794 * 10^{-3}	.4581 * 10^{-3}	.3500 * 10^{-3}	.3179 * 10^{-3}
# Parameters	17.280	17.280	17.420	1.642

References

1. Brouhns, N., Denuit, M., Vermunt, J.K.: A Poisson log-bilinear regression approach to the construction of projected lifetables. Insur. Math. Econ. **31**(3), 373–393 (2002)
2. Enchev, V., Kleinow, T., Cairns, A.J.: Multi-population mortality models: fitting, forecasting and comparisons. Scand. Actuar. J. **2017**(4), 319–342 (2017)
3. Hainaut, D.: A neural-network analyzer for mortality forecast. ASTIN Bull. J. IAA **48**(2), 481–508 (2018)
4. Hatzopoulos, P., Haberman, S.: Common mortality modeling and coherent forecasts. an empirical analysis of worldwide mortality data. Insur. Math. Econ. **52**(2), 320–337 (2013)
5. Hyndman, R.J., Booth, H., Yasmeen, F.: Coherent mortality forecasting: the product-ratio method with functional time series models. Demography **50**(1), 261–283 (2013)
6. Lee, R.D., Carter, L.R.: Modeling and forecasting us mortality. J. Am. Stat. Assoc. **87**(419), 659–671 (1992)
7. Li, J.S.H., Chan, W.S., Zhou, R.: Semicoherent multipopulation mortality modeling: the impact on longevity risk securitization. J. Risk Insur. **84**(3), 1025–1065 (2017)
8. Li, N., Lee, R.: Coherent mortality forecasts for a group of populations: an extension of the Lee-Carter method. Demography **42**(3), 575–594 (2005)
9. Nigri, A., Levantesi, S., Marino, M., Scognamiglio, S., Perla, F.: A deep learning integrated Lee-Carter model. Risks **7**(1), 33 (2019)
10. Perla, F., Richman, R., Scognamiglio, S., Wüthrich, M.V.: Time-series forecasting of mortality rates using deep learning. Scand. Actuar. J. **2021**, 1–27 (2021)
11. Renshaw, A.E., Haberman, S.: Lee–carter mortality forecasting with age-specific enhancement. Insur. Math. Econ. **33**(2), 255–272 (2003)
12. Renshaw, A.E., Haberman, S.: A cohort-based extension to the Lee-Carter model for mortality reduction factors. Insur. Math. Econ. **38**(3), 556–570 (2006)
13. Richman, R., Wüthrich, M.V.: A neural network extension of the Lee-Carter model to multiple populations. Ann. Actuar. Sci. **15**(2), 346–366 (2021)
14. Wilmoth, J.R., Shkolnikov, V.: Human mortality database. University of California, Berkeley (US), and Max Planck Institute for Demographic Research (Germany) (2021)

RVaR Hedging and Market Completions

Ilia Vasilev$^{(\boxtimes)}$ and Alexander Melnikov

University of Alberta, Edmonton, AB T6G 2R3, Canada
{ivasilev,melnikov}@ualberta.ca

Abstract. The problem of partial hedging is very important in the modern risk management industry. It is known that the key element of this problem is a risk measure chosen for assessment of risks. Two of the most widely used risk measures in the industry nowadays are Value-at-Risk (VaR) and Expected Shortfall (CVaR). However, it has been demonstrated recently that both of these measures could be incorporated into one two-parametric risk measure called Range Value-at-Risk (RVaR). In this paper we focus on demonstration that partial hedging problem with respect to RVaR in incomplete market could be approached with the help of method of market completions through the Utility Maximization task embedded into RVaR optimization problem.

Keywords: Range Value at Risk · Incomplete market · Method of market completions · Partial hedging

1 RVaR Hedging Problem

1.1 Multidimensional Diffusion Market Model

We will work with well-known *Standard Multidimensional Market Model* that consists of one risk-free asset and n risky assets (stocks) and can be defined as $(B, S) = (B_t, S_t^1, ..., S_t^n)_{t \leq T}$ where:

$$dB_t = B_t r_t dt, \quad B_0 = 1 \tag{1}$$

$$dS_t^i = S_t^i \left(\mu_t^i dt + \sum_{j=1}^k \sigma_t^{ij} dW_t^j \right)$$

and a k-dimensional vector $W = (W^1, ..., W^k)$ consists of independent standard Brownian motions on a stochastic basis $(\Omega, \mathcal{F}, (\mathcal{F}), P)$. We will call $(\mathcal{F})_{t \leq T}$-adapted process $\pi = (\beta_t, \pi_t^1, ..., \pi_t^n)_{t \leq T}$ *portfolio (strategy)* with its value process $V_t^\pi = \beta_t B_t + \sum_{i=1}^n \pi_t^i S_t^i$.

Let us also denote class of *admissible* portfolios with initial capital x as

$$\mathcal{A}(x) = \{\pi : V_0^\pi = x, \exists K(\pi) \geq 0 \quad \text{s.t.} \quad V_t^\pi \geq -K \quad \forall t \leq T\}.$$

Admissible strategy π is called *self-financing* if the following conditions hold

$$\int_0^T \sum_{i=1}^n \left(|\pi_t^i \mu_t^i| + (\pi_t^i)^2 \sum_{j=1}^k (\sigma_t^{ij})^2 \right) dt < \infty, \tag{2}$$

$$V_t^\pi = V_0^\pi + \sum_{i=1}^n \int_0^t \pi_s^i dS_s^i. \tag{3}$$

The set of strategies, satisfying (2)–(3) is denoted by SFa. We assume that the market (1) is arbitrage-free, which (see details in [2]) holds true if there exists a $(\mathcal{F}_t)_{t \leq T}-$progressively measurable process $\theta = (\theta_t^1, ..., \theta_t^k)_{t \leq T}$ that satisfies

$$\sum_{j=1}^k \sigma_t^{ij} \theta_t^j = \mu_t^i - r, \quad i = 1, ..., n, \quad P - a.s. \tag{4}$$

and

$$\mathbb{E} \left[\exp \left(\frac{1}{2} \int_0^T \sum_{j=1}^k (\theta_t^j)^2 dt \right) \right] < \infty. \tag{5}$$

Definition 1 (Market Completeness). *The market is called complete if for each $\mathcal{F}_T-$measurable payment function $H = H_T(\omega) \geq 0$, such that $\mathbb{E}[H] < \infty$ there exists a strategy $\pi \in SF_a$ such that $\mathbb{P} - a.s.$*

$$V_T^\pi = H$$

Completeness of the market (1) is closely connected with the shape of "volatility" matrix $\Sigma = \{\sigma^{ij}\}_{i=1..n, j=1..k}$. In case of complete market one observes $n = k$, matrix has full rank and we expect it to be non-degenerative and square one $k \times k$.

1.2   Optimization Problem

Construction of an optimal hedging portfolio is a very important practical problem in financial markets and actuary industry. For complete and incomplete markets, there is a chance that perfect- and super-hedging strategies respectively, are not feasible as their initial price is out of investors budget. In this case, market agent can focus on partial-hedging strategies that minimize risk exposure measured by some chosen risk-measure. Assume investor holds a position which, potentially, can result in loss X, which is a random variable defined on space $(\Omega, \mathcal{F}, P)$.

Definition 2 (VaR). *Value-at-Risk (VaR) measure of a loss X at the level α can be defined as*

$$VaR_\alpha(X) = \inf\{a : P(X > a) \leq \alpha\} \tag{6}$$

This measure is not coherent, that is why alternative one was offered

Definition 3 (CVaR). *Conditional Value-at-Risk (CVaR) measure of a loss X at the level α can be defined as*

$$CVaR_\alpha(X) = \frac{1}{\alpha} \int_0^\alpha VaR_\alpha(x)dx, \tag{7}$$

These two measures were suggested by famous Basel accord and are widely used in the Risk Management industry. It was recently demonstrated in [7] that these two measures can be considered as a limit cases of more general two-parametric risk measure called *Range Value at Risk*, or *RVaR*.

Definition 4 (RVaR). *Range Value-at-Risk at levels (α, β) of a loss X is defined as*

$$RVaR_{\alpha,\beta}(X) = \frac{1}{\beta} \int_\alpha^{\alpha+\beta} VaR_s(X)ds, \; \beta > 0$$

$$RVaR_{\alpha,\beta}(X) = \quad VaR_\alpha(X), \; \beta = 0$$

RVaR hedging problem consists in finding a strategy, that satisfies budget constraint (V_0) and minimizes RVaR measure of possible future loss. This problem is solved in two steps: first - find an optimal function for determining hedged-loss proportion of risk, second - use portfolio that replicates this proportion of risk as optimal strategy for RVaR hedging problem.

As shown in [6], the solution to this problem depends on the size of initial capital that investor possess. If investor has enough money to pay for hedging of $f^*(x) = x \cdot I_{x \leq VaR_{\alpha+\beta}(x)}$ then an optimal hedged loss proportion will be $f^*(x)$, otherwise, optimal hedged loss function will be $f^*(x) = f(x, d^*, u^*) = ((x - d^*)^+ + (x - u^*)^+) \cdot I_{x \leq VaR_{\alpha+\beta}(B)}$ which is formulated in the following theorem.

Theorem 1. *If $\Pi(B \cdot I_{B \leq VaR_{\alpha+\beta}(B)}) \leq V_0$, then the optimal hedged loss function is:*

$$f^*(x) = x \cdot I_{x \leq VaR_{\alpha+\beta}(x)} \tag{8}$$

If $\Pi(B \cdot I_{B \leq VaR_{\alpha+\beta}(B)}) > V_0$, then the optimal hedged loss function is:

$$f^*(x) = f(x, d^*, u^*) = ((x - d^*)^+ + (x - u^*)^+) \cdot I_{x \leq VaR_{\alpha+\beta}(B)}, \tag{9}$$

where (d^*, u^*) is the solution to

$$\begin{cases} \min_{\{0 \leq d \leq VaR_\alpha(B), d \leq u \leq VaR_{\alpha+\beta}(B)\}} \left\{ d + \frac{1}{\beta} \int_\alpha^{\alpha+\beta} (VaR_s(B) - u)^+ ds \right\} \\ s.t. \quad \Pi(f(x, d, u)) \leq V_0 \end{cases} \tag{10}$$

In other words, to choose proper optimal hedged loss function, it is crucial to assess if investor has enough capital, which is done with the help of some pricing functional. As it was mentioned in [1], such pricing functional should preserve stop-loss ordering and allow no-arbitrage on the market. One of the famous types of pricing functionals that fits this goal perfectly comes from Utility Based Indifference Pricing (UBIP) approach. However, for this approach, market incompleteness is a serious complication, that can easily be overcome with the help of Method of Market completions.

1.3   Method of Market Completions

In order to complete market, we propose to add necessary amount of "auxiliary" assets into consideration. These are special mathematical objects that represent assets of the same structure as existing ones and intended to "extend" observed Σ to square matrix with full rank.

Assume canonical market model (1) with n risky assets for which $n < k$. As always, asset price dynamics is defined on measure space $(\Omega, \mathcal{F}, P)$ equipped with filtration $\mathbb{F}$ generated by k-dimensional Brownian motion. We will call assets that form this incomplete model *primary assets*, or *existing assets*.

Denote S^c a $(k - n)-$dimensional $(\mathcal{F}_t)_{t \leq T}-$adapted process $S^c = (S_t^{n+1}, ..., S_t^k)_{t \leq T}$ with the same structure as primary assets:

$$dS_t^i = S_t^i \left(\mu_t^i dt + \sum_{j=1}^k \sigma_t^{ij} dW_t^j \right), \quad i = n+1, ..., k.$$

With the help of newly introduced assets, we can "fix" initially rectangular volatility matrix for a set of existing risky assets σ by adding $k - n$ auxiliary assets with coefficients $\sigma_t^{j,1}, \ldots, \sigma_t^{j,d}, j = n+1, \ldots, k$. Then we arrive to properly shaped $(k \times k)$ volatility matrix $\tilde{\Sigma}$.

Definition 5. *The $(k - n)-dimensional$ $(\mathcal{F}_t)_{t \leq T}-adapted$ process $S^c = (S_t^{n+1}, ..., S_t^k)_{t \leq T}$ is called a completion for the (B,S) market if the resulting volatility matrix $\tilde{\sigma}_t$ has full rank for all $t \leq T$.*

1.4   Pricing Functional for RVaR Problem

Useful approach for pricing was offered by Davis [5] and consists in embedding pricing problem into Utility Maximization task. Assume that investors risk appetite is described with concave, non-decreasing utility function U with $U \in C^2$ on $\mathbb{R}^+$ with $U' > 0$, $\lim_{x \to 0} U'(x) = \infty$ and $\lim_{x \to \infty} U'(x) = 0$. Having an initial capital x investor forms a portfolio π with terminal value $V_T^\pi(x)$. His objective is to maximize expected utility of terminal wealth:

$$v(x) = \sup_{\pi \in \mathcal{A}(x)} \mathbb{E} \left[U(V_T^\pi(x)) \right] \tag{11}$$

It was demonstrated in [5] that for a differentiable v with $v'(x) > 0$, $x \in \mathbb{R}^+$ the fair price $\hat{p}(x)$ is given by

$$\hat{p}(x) = \frac{\mathbb{E} \left[U' \left(V_T^{\pi^*}(x) \right) B \right]}{v'(x)} \tag{12}$$

Applying Method of Market Completions *in case of incomplete markets*, one can introduce $k - n$ auxiliary assets in addition to n existing ones on incomplete market, driven by the same k-dimensional Brownian motion as n existing

tradeable assets. Then, the problem of utility maximization can be solved in completed market with fictitious assets. Further, relative risk process can then be represented as $\tilde{\theta}_t \equiv \theta_t + \nu_t$ with $\theta_t^T \nu_t = 0$, and completions are parametrized by ν which is square integrable, adapted and R^d valued process. Define exponential local martingale:

$$Z_t^\nu \equiv \exp\left\{-\int_0^t \tilde{\theta}_s^T dW_s - \frac{1}{2}\int_0^t (\|\theta_s\|^2 + \|\nu_s\|^2)ds\right\}, \tag{13}$$

the function

$$\forall y > 0, \quad \mathcal{X}_\nu(y) \equiv E[\beta_T Z_T^\nu I(y\beta_T Z_T^\nu)], \tag{14}$$

and $\forall \nu \in K_1(\Sigma) \; \{\nu \in K(\Sigma), \mathcal{X}_\nu(y) < \infty, \forall y > 0\}$:

$$\xi_\nu^x \equiv I(\mathcal{Y}_\nu(x)\beta_T Z_T^\nu) \tag{15}$$

where $\mathcal{Y}_\nu$ is the inverse function of $\mathcal{X}_\nu$

Attainable solution will give us value less or equal than that. If we find a strategy π with initial capital x, which doesn't require purchase of the artificial stocks and completion $\lambda_t \in K_1(\Sigma)$ such that

$$\sup_{\pi \in \mathcal{A}(x)} E[U(V_T^\pi)] = E[U(\xi_\lambda^x)], \tag{16}$$

then, $(\hat{\pi}, \lambda)$ would be optimal. In [4] it was proven that if we call

1. Optimality of $\hat{\pi}$: $EU(V_T^\pi) \leq EU(V_T^{\hat{\pi}}) \quad \forall \pi \in \mathcal{A}(x)$
2. Financiability of ξ_λ^x: $\exists \hat{\pi} \in \mathcal{A}(x)$ such that $V_T^{\hat{\pi}} = \xi_\lambda^x$
3. Least Favorability of λ: $EU(\xi_\lambda^x) \leq EU(\xi_\nu^x) \quad \forall \nu \in K_1(\Sigma)$
4. Parsimony of λ: $E[\beta_T Z_T^\nu \xi_\lambda^x] \leq x, \quad \forall \nu \in K_1(\Sigma)$

then 2. $\Longleftrightarrow$ 4.$\Rightarrow$3. Furthermore, if 2. holds, then $\hat{\pi}$ in 2. satisfies 1.

This theorem provides powerful instrument in verifying if one can build optimal strategy without artificial assets in use. In other words, when $\lambda = 0$ will satisfy necessary criteria above, and (12) will be solved on incomplete market.

1.5 Application

Consider utility function $U(x) = \ln(x)$ for which:

$$\mathcal{X}_\nu(y) = \frac{1}{y}, \quad \mathcal{Y}_\nu(x) = \frac{1}{x} \tag{17}$$

and optimal terminal capital can be calculated as

$$\xi_\nu^x = \frac{x}{\beta_T Z_T^\nu}. \tag{18}$$

One could check that completion with parameter $\lambda = 0$ satisfies 4.

$$E[\beta_T Z_T^\nu \xi_0^x] = x \cdot E\left[\exp\left\{-\int_0^T \nu_s^T dW_s - \frac{1}{2}\int_0^T \|\nu_s\|^2 ds\right\}\right] \leq x \quad \forall \nu \in K(\Sigma) \tag{19}$$

as the process under expectation is a supermartingale. It means that investor would not use auxiliary stocks ($\lambda = 0$) to form an optimal portfolio even for hedging purposes. Consequently, RVaR problem (10) here becomes

$$
\begin{cases}
\min_{\{0 \leq d \leq VaR_\alpha(B), d \leq u \leq VaR_{\alpha+\beta}(B)\}} & \left\{ d + \frac{1}{\beta} \int_\alpha^{\alpha+\beta} (VaR_s(B) - u)^+ ds \right\} \\
s.t. \quad E[\beta_T Z_T^\lambda B] \leq V_0
\end{cases}
\tag{20}
$$

It is also worth mentioning that proposed approach is not limited by logarithmic utility function and can also be applied to other popular utility functions such as $U(x) = \frac{x^p}{p}, p < 1, p \neq 0$. Method of market completions was developed in [3] for discrete markets as well as for jump-diffusion model in [8]. In addition, introducing necessary technical complications, demonstrated approach and results obtained for diffusion market model can be extended to other models.

References

1. Cong, J., Tan, K.S., Weng, C.: Conditional value-at-risk-based optimal partial hedging. J. Risk **16**(3), 49–83 (2014)
2. Melnikov, A., Vasilev, I.: On market completions approach to option pricing. Rev. Bus. Econ. Stud. **9**(3), 77–93 (2021)
3. Melnikov, A.V., Feoktistov, K.M.: Non-arbitrage and completeness of discrete markets and contingent claims pricing. Obozreniye Prikladnoy i Promyshlennoy Matematiki **8**(1), 28–40 (2001)
4. Karatzas, I., Lehoczky, J.P., Shreve, S.E., Xu, G.: Martingale and duality methods for utility maximization in a incomplete market. SIAM J. Control Optim. **29**(3), 702–730 (1991). https://doi.org/10.1137/0329039
5. Davis, M.: Option pricing in incomplete markets. Math. Deriv. Secur. **15**, 216–226 (1997)
6. Melnikov, A., Wan, H.: On RVaR based optimal partial hedging. In: Annals of Actuarial Sciences. Cambridge University Press (2022)
7. Embrechts, P., Liu, H., Wang, R.: Quantile-based risk sharing. Oper. Res. **66**(4), 936–949 (2018)
8. MacKay, A., Melnikov, A.: Price Bounds in Jump-Diffusion Markets Revisited via Market Completions. In: Springer Proceedings in Mathematics and Statistics, vol. 259, pp. 553–563

External Spillover Index and Its Relation with GDP per Capita on European Countries

Xenxo Vidal-Llana[1]([✉]) [ID], Jorge M. Uribe[2] [ID], and Montserrat Guillen[1] [ID]

[1] Universitat de Barcelona, Barcelona, Spain
`juanjose.vidal@ub.edu`
[2] Universitat Oberta de Catalunya, Barcelona, Spain

Abstract. We propose a spillover index of external connectedness that measures the outer influences among countries from estimated return volatilities of 645 European firms. We find that Gross Domestic Product per capita is directly related to this index, as countries with lower Gross Domestic Product per capita are influenced in a greater way than they influence others, while higher Gross Domestic Product per capita countries influence others more than they receive. From a risk management perspective, firms should quantify the influence from markets located in other countries, in order to predict possible future movements.

Keywords: Financial contagion · Return spillovers · Gross Domestic Product

1 Introduction

Analyses on how economy moves altogether intrigue most researchers. The interconnection of the markets present relevance also for investors and banks [6], as capital management forms a crucial part of risk assessment. Staying with one market information has been proven biased, as international effects present anomalous relationships [11].

Another common point in the literature about financial contagion is the index usage for calculation of international movements (e.g. [1–4,12]), but while the results are valid, aggregation always carries error. On the counterpart, firm level studies usually reduce their analyses to one country, which, as stated before, will present induce estimation bias. Nevertheless, zooming down to the firm level is necessary to answer questions about market co-movements. A summarized index for explaining price foundation of domestic markets biases the estimation of the models, as we omit information only present at low level granularity, especially in regions with a high level of economic integration.

Less studied, is literature about relation of stock price variation with Gross Domestic Product (GDP onwards). [8] study GDP across countries simulating shocks and calculating spillovers. They conclude that under floating exchange regimes, spillovers decrease for small systems. More closely to our contribution is [13], which correlates oil prices with GDP across oil producing countries like the United States, Russia, Saudi Arabia or China, and prove relation of oil prices between Russia and Saudi Arabia. Our contribution goes one step forward and generalizes the spillovers of a country, with a firm-level starting point of view, proposing the External Spillover Index, an indicator to measure the direction of external spillovers of a country.

We estimate risk spillovers in the European stock market. We create a network using data from October 2, 2015 to March 26, 2021. We use firm level stock return conditional volatilities, estimated daily for 645 firms from 35 European countries.

2   Methodology

We estimate latent autoregressive processes to model the logarithm of the squared volatilities proposed by [14], as the logarithm smooths the variation of the volatility series, and make them more suitable for modeling using a vector autorregression. We take use of [10] approximation with Markov Chain Monte Carlo sampling, implemented in the R-package 'stochvol' [9]. The main input to connectedness analysis is the variance-covariance matrix, for which calculation we use generalized linear regression solved via elastic net regularization [15].

Following a series of papers [6,7] and [5], we estimate the system's connectedness. The moving average representation of a N-variable VAR of order p, is given by $x_t = \sum_{l=1}^{t} A_l \varepsilon_l$, where A_l are a succession of $N \times N$ recursive matrices and A_0 is the identity matrix. We denote as σ_{ij} the elements of the variance-covariance matrix Σ calculated before. A selector vector e_i is included, which consists of ones in the ith coordinates and zeros elsewhere. Thus, similar to [5], we define the spillover from firm j to firm i in the system, H steps ahead, as:

$$\theta_{ij}^g(H) = \frac{\sigma_{jj}^{-1} \sum_{h=0}^{H-1} (e_i' A_h \Sigma e_j)^2}{\sum_{h=0}^{H-1} (e_i' A_h \Sigma A_h' e_j)}, \quad H = 1, 2, 3, \dots \quad (1)$$

In order to compare results, we normalize each term defining the directional connectedness from firm j to firm i as:

$$\tilde{\theta}_{ij}^g(H) := C_{i \leftarrow j}^H = \frac{\theta_{ij}^g(H)}{\sum_{j=1}^{N} \theta_{ij}^g(H)} \ . \quad (2)$$

Suppose a division of the N firms in K subgroups. Subgroups, in our particular case, will be the country that each firm belongs to. In particular, for a subgroup $k \in \{1, \dots, K\}$, we have a list of indexes $I_k \subseteq \{1, \dots, N\}$ that define which firms belong to which subgroup. Therefore, we can calculate the spillover from a subgroup k to all other firms, and the spillover from all the firms in the system that do not belong to a subgroup k, to the same subgroup:

$$C_{\bullet \leftarrow k}^{H} = \frac{\sum_{i \notin I_k, j \in I_k} \tilde{\theta}_{ij}^{g}(H)}{|I_k|}, \tag{3}$$

$$C_{k \leftarrow \bullet}^{H} = \frac{\sum_{i \in I_k, j \notin I_k} \tilde{\theta}_{ij}^{g}(H)}{|I_k|}, \tag{4}$$

Where $|I_k|$ is the number of firms included in subgroup k. Both these equations represent the part of the weighted connectedness of the system that influences and receives a subgroup. Thus, the construction of the External Spillover Index (EI onwards) for a subgroup k in a horizon H appears directly from the division of both equations (3) and (4):

$$EI_k^H = \frac{C_{\bullet \leftarrow k}^{H}}{C_{k \leftarrow \bullet}^{H}} . \tag{5}$$

A value of the EI greater than 1 represents an economy that has more influence over other countries that other countries bring back itself. On the counterpart, an EI under 1 indicates a subgroup that receives spillovers in a greater way from outside than the opposite.

3   Data

Stock prices data have been retrieved from Bloomberg. They consist on daily closing prices of 645 firms listed on one of the 35 national stock markets in Europe from October 2, 2015 to March 26, 2021. The maximum difference in the closing time of the markets within our sample is 3 h, which minimizes the issue of asynchronous trading and different market information enclosed by the prices. This would be an important concern for databases including very distant markets, such as U.S. and China. The GDP per capita for each country has been retrieved from the webpage World Bank Open Data. We estimate the log differences of the original series of prices and the stochastic volatility of the returns for each firm in our sample. Thus, we calculate the spillovers of the system and aggregate following methodology stated in Sect. 2.

4   Results

Calculating connectedness spillovers is useful in terms of risk management. Moreover, it is possible to understand firm behaviors during, for example, turmoil periods. The perspective that we propose is the aggregation of the connectedness of several firms that form a subgroup, in our case their home country, and study their directed spillovers relationship. The horizon used in this study is 10 d.

First, we address the directional spillovers separately, which we present in Fig. 1. In this figure we show results from both Eqs. (3) and (4) in map form. On the left, we find the weighted influence that a country has on others. On the

right, we show the weighted spillovers that a country receives. We see how the influence on others, left map, has more determinant color variety, as the Northern and Western Europe present darker colors than the countries in Eastern Europe. Regarding the influence received, right map, no particular pattern or organization is clear, as colors dilute across the map. In general terms, the countries that influence others is a relevant characteristic for country segmentation, while the influence that a country receives is indeterministic.
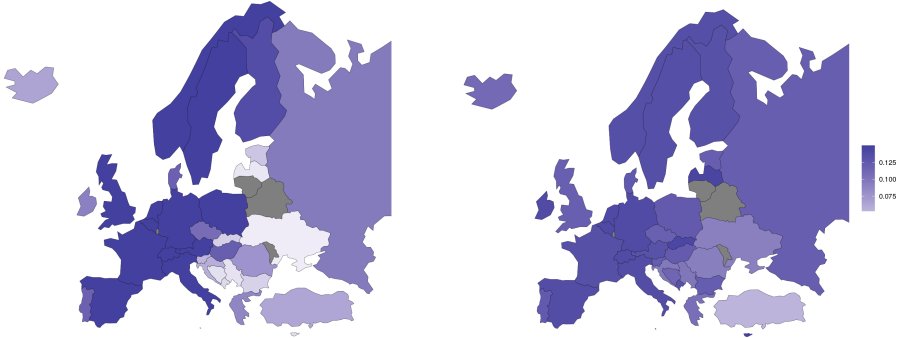
Fig. 1. Weighted spillover maps. Left weighted spillovers on others, and right weighted received spillovers.

		$|I_k|$	EI_k^{10}
Austria	AUT	15	1.2
Belgium	BEL	13	1.4
Bosnia and Herz.	BIH	13	0.2
Bulgaria	BGR	10	0.3
Croatia	HRV	12	0.6
Cyprus	CYP	26	0.2
Czech Rep.	CZE	8	0.9
Denmark	DNK	13	0.9
Estonia	EST	11	0.3
Finland	FIN	16	1.0
France	FRA	32	1.6
Germany	DEU	20	1.5
Greece	GRC	37	0.8
Hungary	HUN	10	1.0
Iceland	ISL	12	0.6
Ireland	IRL	22	0.7
Italy	ITA	18	1.4
Latvia	LVA	12	0.1
Malta	MLT	16	0.1
Montenegro	MNE	7	0.2
Netherlands	NLD	17	1.4
Norway	NOR	16	1.2
Poland	POL	19	1.5
Portugal	PRT	11	0.9
Romania	ROU	11	0.8
Russia	RUS	27	0.8
Serbia	SRB	8	0.2
Slovakia	SVK	4	0.2
Slovenia	SVN	7	0.4
Spain	ESP	22	1.4
Sweden	SWE	26	1.3
Switzerland	CHE	17	1.5
Turkey	TUR	61	1.2
Ukraine	UKR	5	0.1
United Kingdom	GBR	71	1.5

Fig. 2. External Spillover Index of each country in relation with its GDP per capita at 2020. Red line represents a linear regression in logarithmic scale. Dotted horizontal line marks an EI is equal to 1. The table displays the number of firms in each country together with the EI. (Color figure online)

We calculate the External Spillover Index for each country presented in (5), dividing the weighted addition of spillovers from its firms to other countries by the weighted addition of all spillovers from foreign firms to firms that belong to that country.

In Fig. 2 we show the relation of EI and GDP per capita for all 35 studied European countries. In addition, we present a table with the number of firms and the EI for each country in our analysis. In the figure we find how the majority of firms accumulate around the red line, which represents a linear regression with logarithmic scale, meaning possible relation, and proved true. We find how there is a hub on the bottom left of the plot. This hub represents countries with lower GDP per capita and lower EI, meaning countries that receive more influence from other countries than the spillover that they apport to the system. Examples of countries here are Bosnia and Herzegovina (HRV), Ukraine (UKR) and Latvia (LVA). On the center-top of the figure we find another hub, which represents medium-high GDP per capita countries together with an EI greater than one, meaning countries that influence others more than they receive influence in terms of return volatility spillovers. We find here Germany (DEU), United Kingdom (GBR) or Austria (AUT). There are some examples of countries that do not follow the general regression. Bottom-right we find Iceland (ISL) and Ireland (IRL), with a large GDP per capita but low EI. On the counterpart, top-left we find Turkey (TUR) and Poland (POL), whose influence on other countries is greater than received (EI greater than one), but who have a low GDP per capita across Europe.

5   Conclusions

We calculate stochastic volatilities together with the spillover network using individual stock returns of 645 European firms. Our objective was to show the difference of spillover relations of countries between them regarding their GDP per capita. We provided the External Spillover Index, an indicator that shows the direction of external spillovers of a subgroup of firms, in our case the country. We show that countries with higher GDP per capita tend to influence more other countries than they receive influence, and we proved an exponential relation between GDP and EI. However, there are some heterogeneities, as countries like Ireland and Iceland present a greater received influence from other countries, while their GDP per capita is higher than most European countries. On the counterpart, countries like Turkey or Poland have lower GDP per capita but show a greater influence on other countries relative to their received spillovers. From a risk management perspective, we believe that the External Spillover Index is a useful tool for asset management when predicting future volatility returns.

References

1. Baruník, J., Křehlík, T.: Measuring the frequency dynamics of financial connectedness and systemic risk. J. Finan. Economet. **16**(2), 271–296 (2018)

2. Baumöhl, E., Shahzad, S.J.H.: Quantile coherency networks of international stock markets. Finan. Res. Lett. **31**, 119–129 (2019)
3. BenSaïda, A.: Good and bad volatility spillovers: an asymmetric connectedness. J. Finan. Mark. **43**, 78–95 (2019)
4. Chuliá, H., Fernández, J., Uribe, J.M.: Currency downside risk, liquidity, and financial stability. J. Int. Money Finan. **89**, 83–102 (2018)
5. Demirer, M., Diebold, F.X., Liu, L., Yilmaz, K.: Estimating global bank network connectedness. J. Appl. Economet. **33**(1), 1–15 (2018)
6. Diebold, F.X., Yilmaz, K.: Measuring financial asset return and volatility spillovers, with application to global equity markets. Econ. J. **119**(534), 158–171 (2009)
7. Diebold, F.X., Yilmaz, K.: Better to give than to receive: Predictive directional measurement of volatility spillovers. Int. J. Forecast. **28**(1), 57–66 (2012)
8. Douven, R., Peeters, M.: GDP-spillovers in multi-country models. Econ. Model. **15**(2), 163–195 (1998)
9. Kastner, G.: Dealing with stochastic volatility in time series using the r package stochvol. arXiv:1906.12134 (2019)
10. Kastner, G., Frühwirth-Schnatter, S.: Ancillarity-sufficiency interweaving strategy (ASIS) for boosting MCMC estimation of stochastic volatility models. Comput. Statist. Data Anal. **76**, 408–423 (2014)
11. Lewis, K.K.: Global asset pricing. Annu. Rev. Financ. Econ. **3**(1), 435–466 (2011)
12. Lyócsa, Š, Výrost, T., Baumöhl, E.: Return spillovers around the globe: a network approach. Econ. Model. **77**, 133–146 (2019)
13. Nyangarika, A.M.: Correlation of oil prices and gross domestic product in oil producing countries (2018)
14. Taylor, S.J.: Financial returns modelled by the product of two stochastic processes- a study of the daily sugar prices 1961–75. Time Ser. Anal. Theory Pract. **1**, 203–226 (1982)
15. Zou, H., Zhang, H.H.: On the adaptive elastic-net with a diverging number of parameters. Ann. Stat. **37**(4), 1733 (2009)

Author Index