Marco Corazza
Frédéric Gannon
Florence Legros
Claudio Pizzi
Vincent Touzé
Editors

Mathematical and Statistical Methods for Actuarial Sciences and Finance

MAF 2024

Springer

Mathematical and Statistical Methods for Actuarial Sciences and Finance

Marco Corazza · Frédéric Gannon ·
Florence Legros · Claudio Pizzi · Vincent Touzé
Editors

Mathematical and Statistical Methods for Actuarial Sciences and Finance

MAF2024

Springer

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

Florence Legros
ICN Business School Paris La Défense
Nancy, France

Vincent Touzé
OFCE
Sciences Po
Paris, France

Frédéric Gannon
Department of Economics and Management
University of Le Havre
Le Havre, France

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

ISBN 978-3-031-64272-2 ISBN 978-3-031-64273-9 (eBook)
https://doi.org/10.1007/978-3-031-64273-9

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

Preface

This year marks the twentieth anniversary of the first edition of the *International Conference Mathematical and Statistical Methods for Actuarial Sciences and Finance*. The conference is now in its eleventh edition: *International Conference MAF2024 – Mathematical and Statistical Methods for Actuarial Sciences and Finance.*

This sequence of biennial conferences was initiated by the *Department of Economics and Statistics* of the University of Salerno, Italy. The idea underlying this series of scientific meeting is that collaboration and cross-pollination among mathematicians and statisticians engaged in actuarial sciences, insurance, and finance could enhance research in these fields. The effectiveness of this idea has been demonstrated by the widespread participation in all editions, held at various locations including the University of Salerno, Italy (2004, 2006, 2010, 2014, and 2022); Ca' Foscari University of Venice, Italy (2008, 2012, and 2020); University Paris-Dauphine in Paris, France (2016); and University Carlos III of Madrid, Madrid (2018). This effectiveness has also been demonstrated by the attention that both the scientific community and the community of professionals have consistently shown towards the volumes of peer-reviewed papers that have accompanied all past editions of the MAF.

The current international conference *MAF2024* took place in Le Havre, France, in April 4–6, 2024. It was organized by the *University of Le Havre Normandie*, the *ICN Business School* of Paris, France, and the *Department of Economics* of the University Ca' Foscari of Venice, Italy, with the collaboration of the *Department of Economics and Statistics* of the University of Salerno, Italy.

This volume presents a compilation of peer-reviewed papers selected from the submissions intended for presentation at *MAF2024*.

It covers a wide variety of subjects: actuarial models; artificial intelligence and machine learning for finance; clustering of financial data; credit risk methods and models; derivatives; ESG finance; financial econometrics; FinTech and InsurTech; forecasting of actuarial and financial phenomena; fundraising; life insurance; longevity; methods for time series analysis; natural language processing for finance; optimization methods for insurance and finance; pensions; probability in actuarial sciences and finance; real-world case analyses; risk assessment and management; solvency analysis; sustainability; static and dynamic portfolio management; and trading systems.

Of course, both *MAF2024* and this volume would not have been possible without the valuable collaboration of the members of the Scientific and Organizing Committees, without the support of the sponsors, namely: the Italian *Association for Mathematics Applied to Social and Economic Sciences – AMASES*, Italy; the *Department of Economics* of the University Ca' Foscari of Venice, Italy; the *Department of Economics and Statistics* of the University of Salerno; and *Egonon SA – Risk management and advisor*, Appenzell, Switzerland. Additionally, we extend our gratitude to some important partners, including the *Italian Statistical Society*, and the *Italian National Council of Actuaries.*

MAF2024 was further enriched by the contributions of three distinguished plenary session speakers: Stéphane Loisel of the *National Conservatory of Arts and Crafts* of Paris, France; Marie Brière of *AMUNDI* in Paris, France; and Antonio Mele of the *University of Italian Switzerland* of Lugano, Switzerland.

To all of them, our sincere thanks.

Finally, we are pleased to inform you that the Steering Committee is already working on the next edition of MAF in 2026.

We look forward to seeing you!

February 2024

Marco Corazza
Florence Legros
Frédéric Gannon
Claudio Pizzi
Vincent Touzé

Contents

The Cost of Retirement Income Provision: Some Quantitative Insights in Life Insurance

Giovanna Apicella[1], Emilia Di Lorenzo[2], Giulia Magni[3(✉)], and Marilena Sibillo[4]

[1] Department of Economics and Statistics, University of Udine, via Tomadini 30/A, Udine, Italy
giovanna.apicella@uniud.it
[2] Department of Economic and Statistical Sciences, Complesso Monte S. Angelo, University of Naples Federico II, via Cintia, Naples, Italy
diloremi@unina.it
[3] Department of Statistical Sciences, Sapienza University of Rome, Piazzale Aldo Moro 5, Roma, Italy
giulia.magni@uniroma1.it
[4] Department of Economics and Statistics, University of Salerno, via Giovanni Paolo II 132, Fisciano, Italy
msibillo@unisa.it

Abstract. Humanity has observed remarkable improvements in life expectancy at birth. These improvements imply a greater longevity risk in the life insurance field, for pension systems and for the individuals involved in retirement planning. Our work aims to give sense of how longevity evolution reverberates into increasing costs for pension provision. We use the Lee-Carter model and the Human Mortality Database (HMD) data of six EU countries. We assess the dynamics of the price of a temporary life annuity, issued in different calendar years. With respect to the past, in all the countries under study, underwriting a life annuity is becoming progressively more expensive, with similarities between countries especially in relation to the cost of longevity risk protection for females.

Keywords: Life expectancy at birth · longevity risk · life annuities

1 Introduction

Population ageing is both a social achievement and a core problem because of its important implications that relate, for instance, to pension provision and health care supply (e.g., [3]). Indeed, it is well-established in the literature that higher survival prospects imply challenges for public pension systems and private pension plans, with welfare effects determined by the higher longevity prospects for female retirees (e.g., [1,10]) Not all countries have experienced the same pace in longevity improvements. For instance, as stressed by [4], countries of Eastern

M. Corazza et al. (Eds.): MAF 2024, *Mathematical and Statistical Methods for Actuarial Sciences and Finance*, pp. 1–6, 2024.
https://doi.org/10.1007/978-3-031-64273-9_1

Europe have witnessed a rapid increase in the proportion of older people, but later than other European countries. Nevertheless, the issue of ageing population has acquired cruciality in both low-mortality and high-mortality countries [8]. Our study aims to assess how the phenomenon of ageing population, in its different features for a range of European countries, has impacted on the cost of annuitization, in a very simple insurance context. Indeed, we assume that the single premium paid by the annuitant, namely the single amount to be paid at the time of the annuity issue, reflects only the so-called financial-demographic equivalence principle, without any additional charging or adjustment. By way of an example, we focus of an immediate annuity, whose duration is 15 years, issued on an annuitant aged 60. We determine the premium that the annuitant would pay in different calendar years in the past, namely 1960, 1980, 2000, 2020, and in the two future years 2030 and 2040. This procedure involves computing the actuarial present value (APV) of the considered annuities and thus implies making assumptions about how mortality is expected to evolve during the policy duration, since annuity payouts are paid by the insurer upon the annuitant's survival at each policy anniversary. We select 60 as the age for our annuitant based on the average retirement age of males and females from 1958 to 2010 in the countries under study. Our evidence reveals, for 15-year durations, differences in the APV across calendar years, that give sense of the economic implications of the evolving pace of longevity improvement. Such evidence persists even for longer policy durations. We consistently use, throughout the analysis, the Lee-Carter (LC) model [6], because of its simplicity and its wide use at statistical offices to obtain mortality forecasts [2]. The increasing trend in longevity brings with it the increase in the APV of the considered type of annuity. Within our study, by observing the dynamics of such APVs over time, we gain insights into the pace of longevity improvements at different stages of the considered time horizon, for six different European countries, thus uncovering both commonalities and distinct patterns. Our work thus gives sense of how longevity evolution reverberates into increasing costs for pension provision, for countries characterized by different levels of mortality. The paper consists of the following sections: Sect. 2 addresses the used data, the fitting and forecasting procedure of the Lee-Carter model and its application to our case study addressing the computation of the actuarial present values of temporary life annuities. Section 3 presents the results. Section 4 concludes.

2 Data and Methodology

We use the available mortality data, relative to deaths and exposures to risk, from the Human Mortality Database (HMD) [5] for six European countries: Bulgaria, France, Italy, Slovakia, Spain, and Sweden. We choose some countries of Western Europe and some of Eastern Europe to examine variations in survival prospects, the rate of longevity improvement, and the evolution of APV values over time between these two groups. We compute the actuarial present value (APV) of temporary life annuity contracts, issued in different calendar years:

1960, 1980, 2000, 2020, 2030 and 2040. We use the Lee-Carter model projections of mortality rates to describe the expected mortality profile of annuitants aged 60 at the considered times of policy issue. In this respect, we calibrate the Lee-Carter model to the 30 years of data preceding the issue time and project it for the next 15 years. For the issue times 2030 and 2040, we only make use of projected mortality trajectories. Using the same pricing procedure throughout the consistent time horizon allows us to consistently catch the dynamics of the underlying data and to coherently assess how much the cost of a temporary annuity is changing over time, based on the same assumption about the law governing mortality evolution. We use the StMoMo package [9] to implement the fitting and the forecasting of the Lee-Carter model, whose predictor structure is as follows [6]:

$$\ln m_{(x,t)} = \alpha_x + \beta_x k_t \ , \tag{1}$$

where $m_{(x,t)}$ is the central death rate for a person aged x in the calendar year t and k_t is a period parameter describing the general level of mortality over time, specifically it controls the rate at which mortality changes over time. We use the methods from actuarial mathematics (see, e.g., [7]) to assess the actuarial present value (APV) of a temporary immediate life annuity, which expires in 15 years, for individuals (males and females) aged 60 in 1960, 1980, 2000, 2020, 2030 or 2040. The actuarial present value (APV) of the annuity is obtained as in Eq. 2:

$$\pi = b \sum_{k=1}^{15} v^k \ {}_kp_{60,t} \ , \tag{2}$$

where π is the single premium paid by the annuitant to be determined, b refers to the annual amount received by the annuitant (assumed equal to 1000), v^k is the discount factor, ${}_kp_{60,t}$ is the probability that the annuitant aged 60 in calendar year t survives age $60 + k$ We assume that the interest rate is constantly equal to 2% since we are mainly interested in assessing the impact of the demographic dynamics on the single premium π. Let t be the calendar year of the annuity issue. The probability that a person aged 60 in t survives age $60 + k$ can be expressed as follows:

$${}_kp_{60,t} = \prod_{h=0}^{k-1} p_{60+h,t+h} = \prod_{h=0}^{k-1} e^{-m_{60+h,t+h}} \ , \tag{3}$$

where $m_{60+h,t+h}$ is a predicted value by the Lee Carter model, while $p_{60+h,t+h}$ is the annual survival probability obtained from it.

3 Results

In Fig. 1, we display the actuarial present value of the temporary life annuity described in Sect. 2, issued on a woman aged 60 in t=1960, 1980, 2000, 2020, 2030, 2040, under projected LC survival probabilities. For each considered country (Bulgaria, France, Italy, Slovakia, Spain, Sweden), we report, from

the left to the right, the bars showing the amount of the APV for the considered calendar years displayed in ascending order. In Table 1, we report the relative difference between APVs computed every 20 years starting from 1960, when female annuitants are considered. In Fig. 2, we display the actuarial present value of the temporary life annuity described in Sect. 2, issued on a man aged 60 in t = 1960, 1980, 2000, 2020, 2030, 2040, under projected LC survival probabilities. In Table 2, we report the corresponding relative differences, as for the female case. Even though gender is not a pricing factor in the insurance markets, considering both the female and the male case allows us to assess the different cost implied by differing survival prospects and thus to gain intuition about the potential longevity risk management and welfare effects.

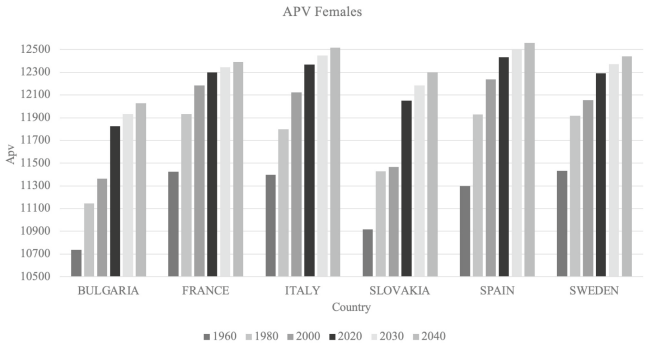


Fig. 1. Actuarial Present Value of Temporary Life Annuity issued on a female aged 60 in t = 1960, 1980, 2000, 2020, 2030, 2040 with projected LC survival probabilities.

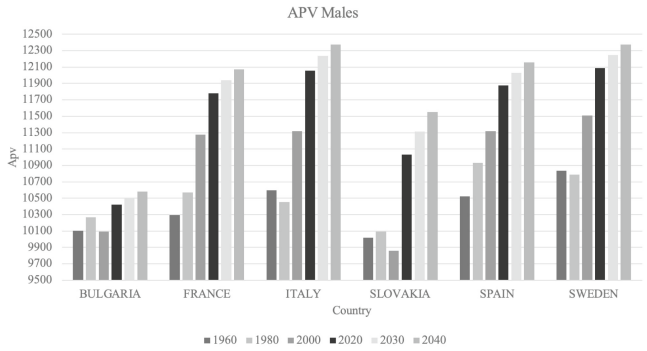


Fig. 2. Actuarial Present Value of Temporary Life Annuity issued on a male aged 60 in t = 1960, 1980, 2000, 2020, 2030, 2040 with projected LC survival probabilities.

The considered type of life annuity is becoming progressively more expensive, with high similarities between countries especially in relation to the cost

of longevity risk protection for females. Since Bulgaria and Slovakia are characterized by a higher mortality risk for both genders than the other countries under study, a life annuity contract is historically less expensive. When looking at the relative differences in the APV, shown in Tables 1 and 2, we obtain evidence of the economic implications of the different pace of longevity improvements under a gender and a cross-country perspective. Indeed, for females, the highest increase in the survival prospects, and the corresponding increase in the APV of a life annuity contract, has affected the 20-years span between 1960 and 1980 for France, Italy, Spain and Sweden. For males and for the same countries, except for Spain, the largest increase has arisen later, between 1980 and 2000. It is also interesting to notice that Bulgaria and Slovakia have experienced a distinct pattern, also in terms of the APV change rates. For both counties, the largest change has occurred between 2000 and 2020 for both genders, with a remarkable increase for males in Slovakia. With respect to past, the change of the APV between 2020 and 2040 is expected to occur at a lower rate than in the past, with the changes for male annuitants having larger magnitudes compared to female annuitants, and with Slovakia marking the highest increase for both genders.

Table 1. Relative percent variation of the actuarial present value of a temporary life annuity for females, between two consecutive (20-years apart) issue years.

ΔAPV	Bulgaria	France	Italy	Slovakia	Spain	Sweden
$(APV_{1980}\text{-}APV_{1960})/APV_{1960}$	3,79%	4,45%	3,51%	4,68%	5,56%	4,25%
$(APV_{2000}\text{-}APV_{1980})/APV_{1980}$	1,96%	2,10%	2,76%	0,36%	2,60%	1,17%
$(APV_{2020}\text{-}APV_{2000})/APV_{2000}$	4,05%	0,93%	2,02%	5,08%	1,59%	1,96%
$(APV_{2040}\text{-}APV_{2020})/APV_{2020}$	1,73%	0,76%	1,20%	2,05%	1,00%	1,21%

Table 2. Relative percent variation of the actuarial present value of a temporary life annuity for males, between two consecutive (20-years apart) issue years.

ΔAPV	Bulgaria	France	Italy	Slovakia	Spain	Sweden
$(APV_{1980}–APV_{1960})/APV_{1960}$	1,59%	2,67%	–1,38%	0,73%	3,90%	–0,44%
$(APV_{2000}\text{-}APV_{1980})/APV_{1980}$	–1,69%	6,70%	8,26%	–2,32%	3,51%	6,70%
$(APV_{2020}\text{-}APV_{2000})/APV_{2000}$	3,28%	4,46%	6,52%	11,88%	4,92%	4,99%
$(APV_{2040}\text{-}APV_{2020})/APV_{2020}$	1,54%	2,50%	2,66%	4,71%	2,39%	2,37%

4 Conclusions

Humanity has observed remarkable improvements in life expectancy at birth. These improvements imply a greater longevity risk in the life insurance field, for pension systems and for the individuals involved in retirement planning. Our

work aims to give sense of how longevity evolution reverberates into increasing costs for pension provision. We assess the dynamics of the price of a temporary life annuity, issued in different calendar years, of six EU countries: Bulgaria, France, Italy, Spain, Slovakia and Sweden. We use the Lee-Carter model to obtain the survival projections needed to price the contract, over the different time horizons of the policy duration under study. When looking at the relative differences in the APV, we obtain evidence of the economic implications of the different pace of longevity improvements under a gender and a cross-country perspective. The considered type of life annuity is becoming progressively more expensive, but at a different pace over the considered time horizon, for the two genders and for the countries under study.

References

1. Apicella, G., De Giorgi, E., Di Lorenzo, E., Sibillo, M.: Gender-inclusive financial and demographic literacy: monetizing the gender mortality gap. Appl. Stochast. Models Bus. Ind. 1–23 (2024). https://doi.org/10.1002/asmb.2876
2. Basellini, U., Camarda, C.G., Booth, H.: Thirty years on: a review of the Lee-Carter method for forecasting mortality. Int. J. Forecast. (2022). https://doi.org/10.1016/j.ijforecast.2022.11.002
3. Brouhns, N., Denuit, M., Vermunt, J.K.: A Poisson log-bilinear regression approach to the construction of projected lifetables. Insur. Math. Econ. **31**, 373–393 (2002)
4. Gavrilova, N. S., Gavrilov, L. A.: Rapidly aging populations: Russia/Eastern Europe. In: Uhlenberg, P. (ed.) International Handbook of Population Aging. International Handbooks of Population, vol. 1. (2009). Springer, Dordrecht. https://doi.org/10.1007/978-1-4020-8356-3_6
5. HMD 2023. Human Mortality Database. Max Planck Institute for Demographic Research (Germany), University of California, Berkeley (USA), and French Institute for Demographic Studies (France). https://www.mortality.org/. Accessed 23 Dec 2023
6. Lee, R.D., Carter, L.R.: Modeling and forecasting U.S. mortality. J. Am. Stat. Assoc. **87**(419), 659-671 (1992). https://doi.org/10.1080/01621459.1992.10475265
7. Pitacco, E., Denuit, M., Haberman, S., Olivieri, A.: Modelling Longevity Dynamics for Pensions and Annuity Business. Oxford University Press, Oxford (2009). ISBN:9780199547272
8. Rabbi, A. M. F., Mazzucco, S.: Mortality and life expectancy forecast for (comparatively) high mortality countries. Genus **74**(18) (2018)
9. Villegas, A.M., Kaishev, V.K., Millossovich, P.: StMoMo: an R package for stochastic mortality modeling. J. Stat. Softw. **84**(3), 1–38 (2018). https://doi.org/10.18637/jss.v084.i03
10. Wingenbach, R., Kim, J.-M., Jung, H.: Living longer in high longevity risk. J. Demogr. Econ. **86**(1), 47–86 (2020). https://doi.org/10.1017/dem.2019.20

Time Preference over the Life-Cycle: Expanding Saver's Rationality

Luc Arrondel[1] and André Masson[2(✉)]

[1] CNRS & Paris School of Economics, Paris, France
luc.arrondel@psemail.eu
[2] CNRS, Paris School of Economics & EHESS, Paris, France
amasson@pse.ens.fr

Abstract. The existence of a pure and rational (time consistent) preference over the life cycle is problematic. We propose an 'existential' approach to time preference where the subject only cares about his 'future selves' insofar as his reasons for living today that involve those selves: his time preference is inversely proportional to the robustness and scope of his current life projects. The empirical measure of such a preference involves building ordinal scores derived from a series of questions, mostly concrete or relating to everyday life. We obtain overall stability of time preference over the medium term, from 2007 to 2020 (until the first Covid lockdown). The variations in this preference for the same individual panel member appear poorly explained by usual covariates, apart from a negative effect of age and of getting married.

Keywords: Time preference · Life-cycle model · Savings · Subjective rationality

1 Introduction: The Problem of a Pure and Rational Time Preference

Time preference plays a role in many areas of the economic literature, including savings and investment, economic growth, interest rate determination, labour supply, health, addiction behaviour and the value placed on human life.

Yet, as emphasized by John Elster (1986, p. 138), the concept of a pure and rational (time consistent) preference for the present "is always a problem" when analysing an individual's choices over his life-cycle, that is on a finite horizon. Many eminent economists and philosophers (Jevons, Ramsey, Harrod, Tobin, Rawls) consider such a preference to be a flaw in individual rationality that can and should be eliminated. For the psycho-economic approach (Strotz 1956; Akerlof 1991; Laibson 1997), on the other hand, short-term impatience is an intrinsic component of human action, but it reflects a limit to rationality (behavioural biases) against which appropriate policies can protect people for their own good.

M. Corazza et al. (Eds.): MAF 2024, *Mathematical and Statistical Methods for Actuarial Sciences and Finance*, pp. 7–12, 2024.
https://doi.org/10.1007/978-3-031-64273-9_2

This problem of the existence of a pure and rational time preference over the life cycle has been largely ignored in the recent economic literature which has focused on the issue of the stability of this preference over time – particularly in the face of major shocks such as the Great Recession or the Covid-19 pandemic – without any empirical consensus finally emerging.

This paper draws heavily on Arrondel and Masson (2024) who focus on two theoretical questions: What meaning should be attributed to a pure and rational time preference over the life-cycle, and what dimensions of choice might it represent? In what way would this preference be operational, and what original and relevant predictions could it lead to?

To be more specific, consider the saver who follows a life-cycle model, with *C(t)* consumption at age t in continuous time and end of life T. Assumed to be time-additive, his or her utility function is written in a situation of certainty at time s:

$$U_s\left[C\left(s\right)...C\left(T\right)\right] = \int_{t=s}^{T} \alpha\left(t\right) u\left[t, C\left(t\right)\right] dt, \alpha\left(t\right) \geq 0, \tag{1}$$

where the discount factor $\alpha(t)$ reflects the decreasing weight (from $\alpha(s) = 1$) given to the instantaneous utility flow $u(t,.)$ due to the existence of a preference for the present. The rate of depreciation of the future δ(t) is (minus) its logarithmic derivative:

$$\delta\left(t\right) = -\frac{\alpha^{'(t)}}{\alpha\left(t\right)} \geq 0 \ ; \ \alpha\left(t\right) = e^{-\int_0^t \delta(t)dt} = e^{-\delta t} \text{ if } \delta\left(t\right) = \delta \tag{2}$$

A *rational* preference for the present corresponds to choices that are time consistent, i.e. a system of preferences that is stable over time. The rate δ can depend on age t, but not on the distance to the present $(t - s)$, that will create a conflict between the desires of the present self and the future self.

Short-term impatience, denoted here by β (corresponding to Laibson's $(1-\beta)$) is inherent in human behaviour, thus justifying a behavioural or psycho-economic approach. It reflects the specific discounting of the immediate future in relation to the present:

$$U_t\left(C_t, C_{t+1}, \ldots, C_T\right) = u_t\left(C_t\right) + (1-\beta)\sum_{k=1}^{T-t}(1+\delta)^{-k} u_{t+k}(C_{t+k}) \tag{3}$$

with $0 \leq \beta \leq 1$. $\beta > 0$ generates a time inconsistency of choices. It may reflect deficit in imagination or foresight as well as lack of willpower and self-control. It explains that a large proportion of long-term savings are in a *contractual* form, with the money tied up in investment products not benefiting from any premium in terms of risk or return but allowing the saver to self-discipline.

2 An Existential Approach to Time Preference

What about the meaning and usefulness of δ? In Arrondel and Masson (2024), we follow an 'existential' approach to that time preference, assumed to expresses

the very nature of human subjectivity in self-to-self relationships over the life cycle. The subject only cares about his 'future selves' insofar as his reasons for living today involve those selves: his time preference is inversely proportional to the robustness and scope of his current life projects (marriage, job, housing, children, preparing for retirement, etc.). The continuity of life, represented by these coefficients $\alpha(t)$, is then no longer a given but a *work-in-progress*, which depends on life choices made on the spectrum between two poles:

1. $\alpha(t)$ close to 0, which may reflect a disjointed existence with "no rhyme or reason", or also correspond to a "carefree" life (advocated by Parfit 1984);
2. $\alpha(t)$ close to 1, which represents a unified life, striving towards a goal, like the edifying lives of the saints, devoted to a noble quest.

This preference α may reveal discontinuities at the nodes of existence – whether desired or experienced, whether a twist of fate or the fulfilment of a current project – that divide the life cycle into successive phases. This idiosyncratic preference extends the person's intertemporal rationality. It thus generates *rational myopia*, as in Becker and Murphy (1988) model of rational addiction, leading to self-destructive behaviour. It also leads to *temporary myopias* at the nodes of existence, producing 'little deaths' followed by 'little rebirths' (Parfit 1984).

3 Empirical Analysis: An Ordinal Time Preference Score

The empirical approach to this 'existential' concept of time preference must be specific. It involves the measurement of *composite individual scores* based on a wide range of questions, often concrete or relating to everyday life, rather than the simple intertemporal trade-off questions usually used (or even Likert scales). For our study, these ordinal scores were established on the basis of the five waves of the PATER survey on household savings and investment (2007, 2009, 2011, 2014 and 2020), with large panel subsamples.

These composite ordinal scores are calculated from some thirty questions, the same in each wave, covering a wide range of areas of life, such as consumption, leisure, investments, work, family, health, retirement, etc. The questions were of a different kind, often concrete or relating to everyday life: "Is retirement something you are preparing for a long time in advance?"; "Are you concerned about keeping fit?"; "Should you instil in your children the taste for saving?"; "Are you prepared to deprive yourself of some of life's pleasures in order to live longer?"; "Do you approve of children who prioritise leisure activities over their studies?"; "Are you someone who generally makes plans?". Others concerned reactions to fictional scenarios as well as lottery choices. The aim was to construct a coherent relative indicator or 'score' of each respondent's time preference. The score is therefore intended to be an aggregate, qualitative and ordinal measure, representative of the answers given by the respondent to a varied set of questions. We code the responses into five categories: far-sighted: -2 or -1; neutral: 0; short-sighted: $+1$ or $+2$. The individual's score is finally the sum of the answers given reduced to those items which, ex post, were found to form a statistically

coherent whole (according to Cronbach's alpha criterion). The underlying idea is that no question is sufficient in itself, but that the score reveals a dimension common to all the questions, ruling out polluting factors.

Carried out on the five waves of the PATER survey, this analysis leads to highly heterogeneous measures of time preference between individuals and to concordant cross-sectional results from one wave to the next: income, age, being a woman, being married, having received an inheritance, and level of education all have a negative effect of their own on time preference – conclusions that largely match Fisher's (1930) intuitions. Time preference δ also has a significant negative effect (comparable from one wave to the next) on the amount of wealth held. The probability of owning shares decreases with preference for the present and risk aversion, while the probability of being home-owner decreases with preference for the present but increases with risk aversion. Finally, foresight strongly increases the holding of life insurance (annuities). All these effects are compatible with the theoretical predictions.

Let us come to the issue of preference stability. We have shown (Arrondel and Masson (2017) that, using this scoring method, attitudes to time remained stable during the Great Recession, between 2007 and 2014: individuals' psyches did not change during the crisis. But here we are first interested in the potential impact of the Covid crisis on time preference.

Figure 1 plots the distribution of the panel sample in 2014 and 2020 according to time preference measured by the value of the score, with a higher score indicating less foresight (stronger preference for the present).

The histograms for the two waves (2014 and 2020) are very similar for the panelled sample. Kolmogorov-Smirnov tests verify that the distributions are not significantly different. In addition, the correlation between the time preference score measured in 2014 and that measured in 2020 is 0.60 for individuals surveyed on both dates.

An econometric study allows us to test the temporal stability of the time preference score at the individual level over the period 2007–2020. Panellised individuals of a given age seem to have become less far-sighted after the Great Recession (2009 wave), but this effect soon diminishes (from the 2011 wave) and the score remains then stable (no Covid effect in 2020).

Consider finally individual variations of time preference for panellised individuals. Over short panel data, of two or three years (between 2007 and 2009, 2009 and 2011, 2011 and 2014, 2014 and 2020), those variations in the time preference score δ appear to be poorly explained by the observed variables or life cycle events: they appear to be akin to white noise or measurement errors, as suggested by Meier and Sprenger (2015) on experimental data. On the other hand, over longer panel periods, between 2007 and 2020 for example, age has a significant and negative effect on the time preference score, as is the case in cross-sections.

What about other demographic changes, such as marriage (or entry into a stable union). Cross-sectional data show that married people have a lower preference for the present. This negative correlation can be interpreted in two

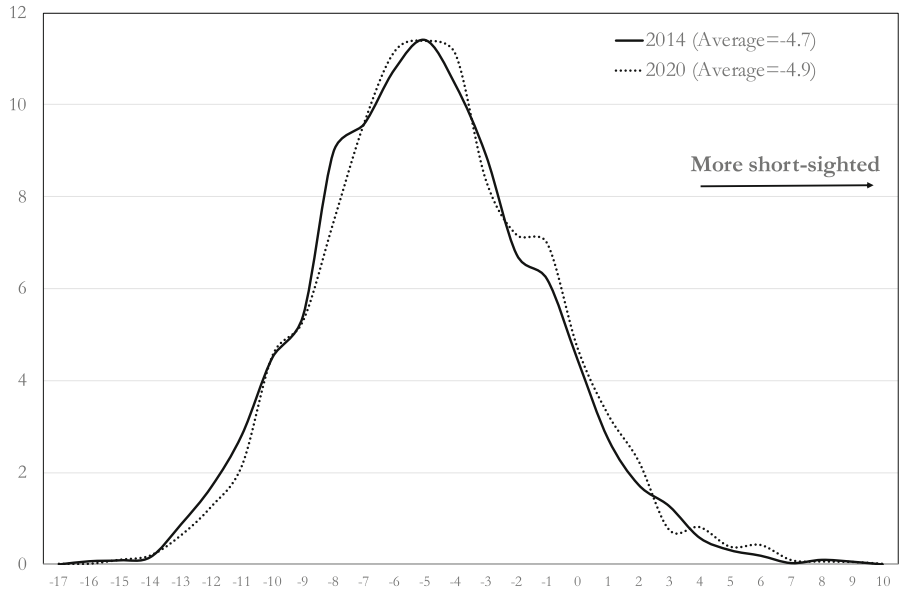


Fig. 1. Stability of time preference: panel data. Source: PATER Panel 2014 and 2020 (N = 1465)

ways: either that individuals with a lower δ rate are more inclined to marry (*spurious dependence*), or that getting married broadens future prospects by reducing δ (*state dependence*). Over longer durations, the second explanation also holds. The PATER panel shows that individuals who married during the period 2007–2020 (134 observations) had a time preference score that fell by 1 point on average (−4.7 vs. −3.7). In addition, the panel econometric study on the overall sample (fixed effects model) confirms a significant negative 'within effect' of unions on the time preference score.

4 Conclusions

This paper extends our previous analysis conducted over the period 2007–2014 on the four first waves of our PATER survey (see Arrondel and Masson 2024). A fifth wave was carried out during the first Covid lockdown. The robust and consistent results obtained from successive waves over the total period 2007–2020, concerning the properties of the time preference score, its determinants and its effects on the amount and composition of wealth, support our existential conception of time preference. The results reveal strong individual heterogeneity and show that time preference is generally stable over a relatively long period, despite a series of economic and health macro shocks. Scores for panellised individuals are highly correlated from one wave to the next, and their variations are largely unexplained, with the exception of age and entry into a stable partnership. Age

effects on time preference are significantly negative, both cross-sectionally and longitudinally, over sufficiently long panel periods.

In line with our existential conception of time preference, which also implies that the latter *may* change at desired or unwanted nodes of existence, we find that getting married significantly decreases the time preference score − an important innovation of this paper. On the other hand, on our possibly still too small samples, other life-cycle events (such as separation or divorce, childbirth or the departure of children from home, retirement or widowhood) have no significant effect, either cross-sectionally or longitudinally, on the time preference score.

At the same time, our scoring method relies on a statistical apparatus that is relatively *heavy*. The time preference score is based on around thirty questions. Simply retaining just a few of these would greatly weaken the previous results and invalidate some of them. However, there is a way out, albeit an imperfect one. Our results show that *the time preference scale*, that is a Likert scale between 0 and 10 (0 referring to "living from day to day" and 10 to a person "who thinks about the future and is far-sighted") is relatively well correlated with the score and preserves many of its properties and results. It therefore represents an acceptable alternative.

References

Akerlof, G.A.: Procrastination and obedience. Am. Econ. Rev. AEA Papers and Proceedings **81**(2), 1–19 (1991)

Arrondel, L., Masson, A.: Why does household demand for shares decline during the crisis? Econ. Stat. **494-495-496**, 155–178 (2017)

Arrondel, L., Masson, A.: Rational Time Preference and the Life Horizon. Opinions & Débats, n° 30, Institut Louis Bachelier, Paris (2024)

Becker, G.S., Murphy, K.M.: A theory of rational addiction. J. Polit. Econ. **96**(4), 675–700 (1988)

Elster, J.: Le laboureur et ses enfants, Editions de Minuit, Paris (1986)

Fisher, I.: The Theory of Interest. Macmillan, New York (1930)

Laibson, D.: Golden eggs and hyperbolic discounting. Q. J. Econ. **112**, 443–477 (1997)

Meier, S., Sprenger, C.D.: Temporal stability of time preferences. Rev. Econ. Stat. **97**(2), 273–286 (2015)

Parfit, D.: Reasons and Persons. Oxford University Press, Oxford (1984)

Strotz, R.H.: Myopia and inconsistency in dynamic utility maximization. Rev. Econ. Stud. **23**, 165–180 (1956)

On a New Perspective in Longevity Risk Management: The Lifetime Shifting

Anna Rita Bacinello, Rosario Maggistro, and Mario Marino(✉)

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,mario.marino}@deams.units.it

Abstract. Longer lives are an achievement and the course of lifespan is increasingly influenced by unobservable risk factors altering the chronological pace of aging. Then, the present work proposes an analytical approach to characterizing the human lifetime based on the concept of non-chronological age. Starting from a chronological Gompertz mortality framework, we define the non-chronological lifespan and characterize it probabilistically by deriving, in closed-form, the expression for the cumulative distribution function, the density of deaths function, and the mortality hazard function. We find that non-chronological death probabilities are a time-dependent affine transformation of chronological death probabilities for a newborn, and we highlight the link between the non-chronological lifetime and the concept of individual frailty in heterogeneous mortality modelling. We believe that our proposal may contribute to shaping a new perspective on longevity risk measurement and management.

Keywords: Lifetime · Non-chronological age · Longevity risk

1 Introduction

In the last decades, the human lifetime has grown continuously and human mortality has shifted to later ages ([8]). Mortality deferment to older ages was empirically observed by investigating changes in the main lifetime indicators expressed in terms of chronological age, such as the mortality hazard, the survival function, and density of deaths function (see, e.g., [6]). While living longer, aka longevity, is a positive achievement at the individual level, it implies significant unexpected financial exposures for governments, annuity providers, and pension schemes, namely 'longevity risk' in actuarial jargon. Such a risk is a compelling matter of interest for both actuaries and policymakers, and both demographic and actuarial literature have seen an enriched focus on mortality modelling and forecasting. Nowadays, thanks to the advances in medical literature, new perspectives on longevity analysis have emerged. In particular, the concept of biological age has been introduced, that is the age indicating how old the human mechanism is at both the cellular and molecular levels (see, e.g., [2]). The biological age may be misaligned with respect to the corresponding chronological age, and it can

M. Corazza et al. (Eds.): MAF 2024, *Mathematical and Statistical Methods for Actuarial Sciences and Finance*, pp. 13–18, 2024.
https://doi.org/10.1007/978-3-031-64273-9_3

reasonably be acknowledged as a key element in analyzing lifespan randomness. The biological age is usually estimated by collecting data concerning physiological and molecular variables for a large sample of people, and, by means of multivariate regression, the sign of statistically-significant regression coefficients leads an increment or a reduction of the corresponding chronological age. In other words, due to biological (and observable) factors, person's age does not move necessarily in lockstep with calendar time and different individuals may age at different rates. Within the actuarial literature, in [3] the meaning and the use of the biological age is discussed for the first time. On one side, this paper highlights that the biological age is a relevant variable to predict the risk of chronic disease and maximize the health span, but not necessarily lifespan; on the other, the presence of a non-chronological age that differs from the chronological one impacts lifespan and the longevity risk measurement. Therefore, by referring to a Gompertz-Makeham mortality framework, [3] paves the way to construct a non-chronological age, namely longevity-risk-adjusted global age (L-RaG), different from the biological one and in contrast to the chronological age. Another type of non-chronological age is defined in [1], namely survivorship-age (s-age), representing the age at which a proportion of a population is still alive. The underlying idea is to invert the relation between the survival function and the chronological age, so that the latter becomes a function of the survival levels. The authors investigate the behaviour of the mortality hazard associated to the s-age, showing that populations experience a similar risk of dying at specific levels of survivorship. The L-RaG and the s-age are outcomes of distinct approaches, but both state the existence of a non-chronological age determined without the use of observable biological factors. Interestingly, we note that this is what happens when frailty-based models are employed in shaping heterogeneous mortality due to unobservable risk factors (see, e.g., [4,5]). Indeed, some biological factors entailing the gap between the non-chronological age and the chronological one may be not directly observable or not available, and, in addition, they imply a mortality differentiation among individuals. Then, the gap between these ages may be assimilated into an unobserved frailty. To some extent, this is also the intuition behind the work in [7]. The authors assume a Generalized Gompertz distribution (GG) for the lifetime and prove that, under specific assumptions, the frailty can be interpreted as a random correction to the chronological age. However, their proposal allows the presence of negative chronological ages. In the vein of the aforementioned literature, the present work aims to characterize the human lifetime taking into account a random shift of the chronological age. More in detail, we primarly consider a chronological age-based mortality by means of the Gompertz model, and then we assume a Generalized Gompertz distribution for the random shift to probabilistically define a non-chronological lifetime. As a result, we provide closed-form expressions for the cumulative distribution function, the density of death function and the mortality hazard under the non-chronological lifetime. Our proposal contributes to the current literature by posing a new modelling perspective concerning the lifespan randomness due to unobservable risk factors, avoiding the possibility of negative lifetimes.

The paper is organized as follows. In Sect. 2 we recall the Gompertz mortality framework and we introduce the GG distribution. In Sect. 3 we develop our proposal by defining the non-chronological lifetime and providing analytical results concerning its distribution. Finally, Sect. 4 poses conclusions.

2 Chronological Lifetime in a Gompertz Framework

Given a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, let T_0 be the random lifetime for a newborn. We assume that T_0 is Gompertz distributed, i.e., $T_0 \sim \mathrm{G}(h, g)$, with cumulative distribution function (cdf) and probability density function (pdf) given, respectively, by

$$F_{T_0}(x) = 1 - \exp\left\{-\frac{h(e^{gx}-1)}{g}\right\}, \qquad f_{T_0}(x) = h\, e^{gx} \exp\left\{-\frac{h(e^{gx}-1)}{g}\right\}, \quad (1)$$

and the following mortality hazard holds

$$\mu(x) = \frac{f_{T_0}(x)}{1 - F_{T_0}(x)} = h\, e^{gx}. \quad (2)$$

Equation (2) represents the well-known Gompertz mortality law (under the chronological age), where the parameter h is the initial mortality level and the parameter g indicates the rate of aging. For any chronological age $x > 0$, the residual random lifetime is defined as $T_x = T_0 - x \,|\, T_0 > x$, and its cdf, pdf and mortality hazard in the Gompertz mortality framework are, respectively,

$$F_{T_x}(t) = 1 - \exp\left\{-\frac{h}{g} e^{gx}(e^{gt}-1)\right\}, \quad (3)$$

$$f_{T_x}(t) = h\, e^{g(x+t)} \exp\left\{-\frac{h}{g} e^{gx}(e^{gt}-1)\right\}, \quad (4)$$

$$\mu(x+t) = h\, e^{g(x+t)}, \quad (5)$$

where $t > 0$. As argued in [7], the lifetime distribution for a newborn can be described in more general terms by adopting the GG distribution. In detail, we say that T_0 has the Generalized Gompertz distribution, $\mathrm{GG}(a, b, c)$, $a \in \mathbb{R}$, $b, c > 0$, if the cdf and the pdf are, respectively, defined as

$$F_{T_0}(x) = 1 - \frac{\Gamma\left(c, \exp\left(\frac{x-a}{b}\right)\right)}{\Gamma(c)}, \quad f_{T_0}(x) = \frac{1}{b\Gamma(c)} \exp\left\{c\,\frac{x-a}{b} - \exp\left(\frac{x-a}{b}\right)\right\}, \quad (6)$$

where $\Gamma(c, w) = \int_w^{+\infty} u^{c-1} e^{-u} du$ is the upper incomplete Gamma function and $\Gamma(c) = \Gamma(c, 0)$ is the complete Gamma function. We notice that cdf and pdf in (6) are defined for $x \in \mathbb{R}$, that is negative lifetimes may occur with positive probability. Despite this drawback, in [7] it is shown that, when the lifetime under a frailty-based model is considered, a GG-frailty defines a random age correction to the chronological lifetime. In the next section, we propose chronological lifetime shifting by using the GG distribution and avoiding negative lifetimes.

3 Shifting the Chronological Lifetime

Let us introduce the non-chronological lifetime $\widetilde{T}$. We assume that for a newborn $\widetilde{T}_0 = T_0$ almost surely, while the residual lifetime can be defined according to the passage of age in a non-chronological manner:

$$\widetilde{T}_x = T_0 - (x + \Delta) \,|\, T_0 > x, \tag{7}$$

where $(x + \Delta)$ is a non-chronological age, being Δ a random shift in width and sign, and with T_0 and Δ stochastically independent. The cdf of (7) is defined in the following Proposition 1.

Proposition 1. *Let the non-chronological lifetime $\widetilde{T}_x$ be defined as in (7), and assume that $\Delta \sim GG(0, b, c)$, with $b, c > 0$ and cdf given by*

$$F_\Delta(\delta) = 1 - \frac{\Gamma\left(c, e^{\frac{\delta}{b}}\right)}{\Gamma(c)}, \quad \delta \in \mathbb{R}. \tag{8}$$

Then, if the chronological lifetime has Gompertz distribution, i.e. $T_0 \sim G(h, g)$, $h, g > 0$, the cdf of the non-chronological lifetime is

$$F_{\widetilde{T}_x}(t) = B(t) + A(t) F_{T_0}(x + t), \quad t > 0, \tag{9}$$

where $F_{T_0}(x + t) = 1 - \exp\left\{-\frac{h}{g}(e^{g(x+t)} - 1)\right\}$, and

$$A(t) = \frac{g^{\frac{c}{gh}-1} b^{\frac{c}{g}-1} h^{\frac{c(h-1)}{hg}} (1 + hb)^{-\frac{c}{g}} \Gamma\left(\frac{c}{b}, e^{-gt}\left(\frac{1+hb}{gb}\right)\right)}{(1 - F_{T_0}(x))\,\Gamma(c)}, \tag{10}$$

$$B(t) = \frac{1}{1 - F_{T_0}(x)} - \frac{\Gamma\left(c, e^{-\frac{t}{b}}\right)}{\Gamma(c)}\left(\frac{2}{1 - F_{T_0}(x)} - 1\right) - A(t). \tag{11}$$

Proof. Since $\widetilde{T}_x = T_0 - (x + \Delta) \,|\, T_0 > x$, then the cdf of the non-chronological lifetime is determined by computing

$$\begin{aligned} F_{\widetilde{T}_x}(t) &= P(T_0 \le x + \Delta + t \,|\, T_0 > x) \\ &= \frac{1}{1 - F_{T_0}(x)} \int_{-t}^{+\infty} P(x < T_0 \le x + \delta + t)\, dF_\Delta(\delta) \\ &= \frac{1}{1 - F_{T_0}(x)} \left\{ \int_{-t}^{+\infty} F_{T_0}(x + \delta + t)\, dF_\Delta(\delta) - F_{T_0}(x)(1 - F_\Delta(-t)) \right\}. \end{aligned} \tag{12}$$

By assuming that $T_0 \sim \mathrm{G}(h, g)$, with $h, g > 0$, the expression of $F_{T_0}(x)$ is the cdf in (1), while the cdf's expression of Δ is given by (8). Then, by substituting in (12), we get

$$\begin{aligned} F_{\widetilde{T}_x}(t) = \exp\left\{\frac{h}{g}(e^{gx} - 1)\right\} - \frac{\Gamma(c, e^{-\frac{t}{b}})}{\Gamma(c)}\left(2\exp\left\{\frac{h}{g}(e^{gx} - 1)\right\} - 1\right) \\ - \frac{\Gamma\left(\frac{c}{b}, e^{-gt}\left(\frac{1+hb}{gb}\right)\right)}{bg\Gamma(c)} \left(\frac{g}{h}\right)^{\frac{c}{gh}} \left(\frac{1 + hb}{hb}\right)^{-\frac{c}{b}} \exp\left\{-\frac{h}{g} e^{gx}(e^{gt} - 1)\right\}. \end{aligned} \tag{13}$$

Due to (1) and (3), it holds that:

$$\exp\left\{\frac{h}{g}(e^{gx}-1)\right\}=\frac{1}{1-F_{T_0}(x)},\quad \exp\left\{-\frac{h}{g}e^{gx}(e^{gt}-1)\right\}=\frac{1-F_{T_0}(x+t)}{1-F_{T_0}(x)}. \tag{14}$$

By substituting (14) in (13), and rearranging the terms, the expressions (9)-(11) follow, completing the proof. □

From Proposition 1, we highlight the following considerations:

- Firstly, (9) provides the probability of death at the non-chronological age $\xi := x + \Delta$, namely ${}_t\tilde{q}_\xi$ by exploiting the actuarial notation, and it differs from the corresponding probability of death at the chronological age x, i.e. ${}_tq_x$. The latter can written as

$${}_tq_x = B + A\,{}_{x+t}q_0, \tag{15}$$

  where ${}_{x+t}q_0$ is the probability of death by the chronological age $(x+t)$ for a newborn, $A=\frac{1}{1-F_{T_0}(x)}$, and $B=-\frac{F_{T_0}(x)}{1-F_{T_0}(x)}$. Looking at (9) and (15), we observe that both the chronological and the non-chronological probabilities of death are affine functions of the probability ${}_{x+t}q_0$. The coefficients of the ${}_tq_x$'s affine transformation are time-invariant, while they become time-dependent (and more complex) when the probability ${}_t\tilde{q}_\xi$ is computed. To some extent, while the chronological probabilities of death are defined in a static way, the non-chronological probabilities stem from a time-dependent adjustment of ${}_{x+t}q_0$;
- The Generalized Gompertz distribution assumption for Δ can be related to the frailty coefficient characterizing the frailty-based mortality models. In particular, a Gamma distribution is usually adopted for the frailty coefficient which is applied, in a multiplicative way, to the population mortality hazard (see, e.g., [4,5]). Then, for all the Gamma realizations in the interval $(0,1)$ the individual mortality hazard is lower than that of the population (lower individual frailty), and the opposite case occurs for realizations in $(1,+\infty)$ (higher individual frailty). For the purposes of our proposal, we highlight that the Gamma and Generalized Gompertz distributions are connected. For instance, $\Delta = b\ln(Y) \sim \mathrm{GG}(0,b,c)$, with $b>0$, if $Y \sim \mathrm{Gamma}(c,1)$, $c>0$. In other words, our proposal supposes a non-chronological lifetime obtained as a frailty-based shift of the chronological lifetime. Then, for every realization $y \in (0,1)$ we attain negative outcomes for Δ, implying a reduction of the chronological age and an increment of the lifetime (i.e. lower frailty); conversely, for every realization $y \in (1,+\infty)$, we have positive values for Δ, a consequent growth of the chronological age and a shortened lifetime (i.e. greater frailty).

Moreover, by differentiating (9), the pdf of the non-chronological lifetime is

$$f_{\widetilde{T}_x(t)} = \frac{d}{dt}F_{\widetilde{T}_x(t)} = C(t) + A(t)f_{T_0}(x+t), \tag{16}$$

where $f_{T_0}(x+t)$ is defined in (1) (with x replaced by $x+t$), and

$$C(t) = B'(t) + A'(t)F_{T_0}(x+t), \tag{17}$$

$$B'(t) = \frac{\exp\left\{-e^{-\frac{t}{b}} - \frac{tc}{b}\right\}\left(1 - \frac{2}{1-F_{T_0}(x)}\right)}{b\Gamma(c)} - A'(t), \tag{18}$$

$$A'(t) = \frac{\exp\left\{-\frac{e^{-gt}(1+bh)+g^2tc}{bg}\right\}\frac{1}{1-F_{T_0}(x)}\left(\frac{1+bh}{b}\right)^{\frac{c(g-b)}{bg}} g^{\frac{c(b-gh)}{ghb}} h^{\frac{c(h-1)}{gh}}}{b\Gamma(c)}. \tag{19}$$

Finally, the non-chronological mortality hazard can be computed as

$$\widetilde{\mu}(x+t) = \frac{f_{\widetilde{T}_x(t)}}{1 - F_{\widetilde{T}_x(t)}}.$$

4 Conclusion

In this work, we have proposed an analytical approach to define a non-chronological lifetime and investigated its main probabilistic features. We have found that non-chronological death probabilities are a time-dependent affine transformation of chronological death probabilities for a newborn. In addition, we have highlighted how the shift between the non-chronological and chronological lifetimes and the concept of individual frailty in heterogeneous mortality models may be related.

References

1. Alvarez, J.-A., Vaupel, J.W.: Mortality as a Function of Survival. Demography **60**(1), 327–342 (2023)
2. Jackson, S.H.D., Weale, M.R., Weale, R.A.: Biological age-what is it and can it be measured? Arch. Gerontol. Geriatr. **36**(2), 103–115 (2003)
3. Milevsky, M.A.: Calibrating Gompertz in reverse: what is your longevity-risk-adjusted global age? Insur. Math. Econ. **92**, 147–161 (2020)
4. Olivieri, A.: Heterogeneity in survival models: applications to pension and life annuities. Belgian Actuarial Bull. **6**, 23–39 (2006)
5. Pitacco, E.: Heterogeneity in mortality: a survey with an actuarial focus. Eur. Actuar. J. **9**(1), 3–30 (2019)
6. Vaupel, J.W., Villavicencio, F., Bergeron-Boucher, M.-P.: Demographic perspectives on the rise of longevity. Proc. Natl. Acad. Sci. **118**, e2019536118 (2021)
7. Willemse, W.J., Kaas, R.: Rational reconstruction of frailty-based mortality models by a generalisation of Gompertz' law of mortality. Insur. Math. Econ. **40**(3), 468–484 (2007)
8. Zuo, W., Jiang, S., Guo, Z., Feldman, M.W., Tuljapurkar, S.: Advancing front of old-age human survival. Proc. Natl. Acad. Sci. **115**, 11209–11214 (2018)

An Application of Beta Binomial GAMLSS for the Estimate of Surrender Rates

Fabio Baione, Davide Biancalana(✉), and Paolo De Angelis

Sapienza University, Rome, Italy
{fabio.baione,davide.biancalana,paolo.deangelis}@uniroma1.it

Abstract. This paper deals with the estimate of surrender rate with explanatory variables by a Generalized Linear Model for Location, Scale, and Shape (GAMLSS) where the response variable is assumed Beta Binomial. In actuarial practice and literature, the Binomial Generalized Linear Model is frequently used to get an estimate of surrender rates per policy count conditional to policy and policyholder features. We suggest a regressive model based on a Beta Binomial assumption of the response variable. Beta Binomial is a discrete random variable that differs from binomial because the probability of success at each of n trials is not fixed, but beta distributed. Beta Binomial random variable is fit to model binomial phenomena where the probability of success is not fixed but is inferred from data. Beta Binomial random variable has greater variance and skewness than a Binomial random variable with the same mean, because in the Beta Binomial approach the uncertainty about what the true probability is, is taken into account. This uncertainty makes values far from mean more plausible. Finally, the Beta Binomial does not belong to the exponential family. For this reason, a GAMLSS model is used to get parameter estimates.

Keywords: Surrender · Lapse · GAMLSS · Beta Binomial

1 Introduction

Surrender is one of the most important expressions of policyholder behavior; the latter has a big influence on asset liability management for most life insurance companies. The goal of this paper is the estimate of surrender rate with explanatory variables by a Beta Binomial Generalized Linear Model for Location, Scale, and Shape (BBGAMLSS). In actuarial practice and literature, the Binomial Generalized Linear Model (BGLM) is used to estimate surrender rates. In [3], BGLM is used to get an estimate of the surrender rate considering economic variables using the logit function and the complementary log-log function. Some recent studies, (see [1]) introduce a two-part model based on the simultaneous use of a BGLM and a Beta Regression (see [2]) in order to get an estimate of the expected cash flows due to lapses considering both the surrender and withdrawal component. Our idea is to get an estimate of surrender probability per

M. Corazza et al. (Eds.): MAF 2024, *Mathematical and Statistical Methods for Actuarial Sciences and Finance*, pp. 19–24, 2024.
https://doi.org/10.1007/978-3-031-64273-9_4

policy count introducing an assumption of Beta Binomial (BB) distribution on the response variable to consider the uncertainty on the surrender probability; in this context, the probability of success can't be assumed fixed as in the toin coss or roll the dice experiment.

BB is a more dispersed (and more asymmetric) random variable (r.v.) than a binomial. This means that when data shows a very dispersed distribution the probability of success could be not fixed but varies around a value. In this case, a BB model could appear more appropriate for the estimate of the probability of success.

Finally, in order to get an estimate of surrender rates conditional to policyholder or policy features we introduce a regression model; in particular, we select a GAMLSS (see [4]) to model a BB response variable. GAMLSS (unlike the GLM) works even if the response variable does not belong to the exponential family.

2 Beta Binomial Random Variable

Let $Y \sim BB(a, b, n)$ be a BB r.v., with parameters $a, b \geq 0$, and n is a positive integer. A BB r.v. Y with parameters a, b and n has probability mass function:

$$f_Y(y) = \frac{\Gamma(y+a) \cdot \Gamma(n-y+b) \cdot \Gamma(a+b) \cdot \Gamma(n+2)}{(n+1) \cdot \Gamma(a+b+n) \cdot \Gamma(a) \cdot \Gamma(b) \cdot \Gamma(y+1) \cdot \Gamma(n-y+1)} \tag{1}$$

A BB r.v. is a binomial r.v. where the probability of success $\mathbf{p}$ is beta distributed with parameters a and b. We can formalize $\mathbf{p} \sim Beta(a, b)$, then $E[\mathbf{p}] = \frac{a}{a+b}$.

Using the notation $\frac{a}{a+b} = p$, we can write the mean variance and skewness (γ) of the BB random variable in a form that is compliant with the binomial case.

$$\begin{cases} E[Y] = n\frac{a}{a+b} = np \\ \sigma^2(Y) = n\frac{ab}{(a+b)^2}\frac{a+b+n}{a+b+1} = np(1-p)\frac{a+b+n}{a+b+1} \\ \gamma(Y) = \frac{(a+b+2n)(b-a)}{a+b+2}\sqrt{\frac{1+a+b}{nab(a+b+n)}} = \frac{1-2p}{\sqrt{np(1-p)}}\frac{a+b+2n}{a+b+2}\sqrt{\frac{1+a+b}{a+b+n}} \end{cases} \tag{2}$$

As one can see, for $n \geq 1$ the BB r.v. has always greater variance and skewness than a binomial r.v. with the same mean.

Assuming a BB distribution for the response variable, our goal is to estimate the number of surrenders (and therefore the probability of surrender) conditional on some covariates.

Surrender cannot be defined as a phenomenon where the probability of the event is fixed for each risk class but reasonably varies around a value. In other words, each policyholder in a single risk class has a different behavior because the regressive model cannot consider all the variables that influence the propensity to surrender.

In order to make a BB regression model, we need to specify the dependency structure between the response variables and the set of covariates $\mathbf{x}_h$. GLM is impossible to use because BB does not belong to the exponential family. Hence, we apply a GAMLSS.

3 A Brief Introduction to GAMLSS

A GAMLSS model assumes a sample of I observations y_i, $i = 1, \ldots, I$ from a random variable Y_i, with probability density function $f\left(y_i|\theta^i\right)$, conditional on $\theta^i = (\theta_{1,i}, \theta_{2,i}, \theta_{3,i}, \theta_{4,i}) = (\mu_i, \sigma_i, \nu_i, \tau_i)$ a vector of four distribution parameters. The parameters are in a regressive relationship with the explanatory variables. Let $(\mu_i, \sigma_i, \nu_i, \tau_i)$ be the distribution parameters.

μ_i is the location parameter, σ_i is the scale parameter and the two remaining give contributions to define skewness and kurtosis, although the model can be generalized to more (or less) than four distribution parameters.

Let $y^T = (y_1, \ldots, y_I)$ be the I length vector of the response variable. Also for $k = 1, 2, 3, 4$, let $g_k(.)$ be known monotonic link functions relating the distribution parameters to explanatory variables by:

$$g_k\left(\theta_k\right) = \eta_k = X_k\beta_k + \sum_{j=1}^{J_k} Z_{jk}\gamma_{jk}, \quad with \quad k = 1, 2, 3, 4. \tag{3}$$

where μ, σ, ν, τ and η_k are vectors of length I, $\beta_k^T = (\beta_{1k}, \beta_{2k}, \ldots, \beta_{m_k k})$ is a parameter vector of length m_k, X_k is a fixed known design matrix of order $I \times m_k$, Z_{jk} is a fixed known $I \times q_{jk}$ design matrix and γ_{jk} is a q_{jx} dimensional random variable which is assumed to be distributed as $\gamma_{jk} \sim N_{q_{jk}}\left(0, G_{jk}^{-1}\right)$, where G_{jk}^{-1} is the (generalized) inverse of a $q_{jk} \times q'_{jk}$ symmetric matrix $G_{jk} = G_{jk}\left(x\lambda_{jk}\right)$, which may depend on a vector of hyperparameters λ_{jk}.

Equation (3) introduces a general definition of GAMLSS without assuming a specific distribution for the response variable.

In this paper, we assume $Y \sim BB(a, b, n)$. BB is compliant with GAMLSS structure, then it is possible to express a and b as a function of μ and σ by the following parametrization:

$$a = \frac{\mu}{\sigma}; \quad b = \frac{1-\mu}{\sigma} \tag{4}$$

4 Some Numerical Results

We set our model on a database from an Italian life insurance company between the years 2009 and 2013. We consider Gender (G) and class of policy duration (CPD) in years as our covariates. In Table 1, we report for each year the observed number of surrenders and number of policies for G, CPD, and for each year of our sample. CPD variable has 5 levels: < 7;$[7, 11)$;$[11, 15)$;$[15, 17)$;≥ 17.

Table 1. Observed number of surrenders by G, CPD and year

G	CPD	2009		2010		2011		2012		2013		Total		
		Surrenders	Policies	Surrenders	Policies	Surrenders	Policies	Surrenders	Policies	Surrenders	Policies	Surrenders	Policies	Surrender rate
F	< 7	503	3,792	454	2,402	316	1,697	774	3,420	1,229	5,624	3,276	16,935	19.34%
F	[7,11)	116	2,435	116	2,736	143	1,710	302	2,011	245	1,881	922	10,773	8.56%
F	[11,15)	0	0	44	954	76	2,229	81	2,085	118	2,005	319	7,273	4.39%
F	[15,17)	0	0	0	0	4	37	12	111	6	85	22	233	9.44%
F	≥ 17	0	0	0	0	0	0	0	0	4	30	4	30	13.33%
M	< 7	443	3,720	463	2,483	275	1,726	618	3,251	1,020	5,507	2,819	16,687	16.89%
M	[7,11)	110	2,475	99	2,615	116	1,751	321	2,087	196	1,950	842	10,878	7.74%
M	[11,15)	0	0	42	1,043	88	2,251	83	2,117	102	2,063	315	7,474	4.21%
M	[15,17)	0	0	0	0	2	47	9	117	3	83	14	247	5.67%
M	≥ 17	0	0	0	0	0	0	0	0	0	39	0	39	0.00%
Total		1,172	12,422	1,218	12,233	1,020	11,448	2,200	15,199	2,923	19,267	8,533	70,569	12.09%

Table 2. BGLM estimates compared with observed means and standard deviations

G	CPD	Estimated mean	Observed mean	Estimated standard deviation	Observed standard deviation
F	< 7	650.23	655.20	22.92	323.13
F	[7,11)	187.31	184.40	13.08	75.60
F	[11,15)	66.94	63.80	7.99	39.61
F	[15,17)	3.74	4.40	1.85	4.45
F	≥ 17	0.37	0.80	0.59	1.60
M	< 7	568.77	563.80	21.72	252.67
M	[7,11)	165.49	168.40	12.37	83.69
M	[11,15)	59.86	63.00	7.58	37.30
M	[15,17)	3.46	2.80	1.79	3.31
M	≥ 17	0.43	0.00	0.63	0.00

Females show a mildly greater propensity to surrender. In order to get an estimate of the number of surrenders, the parameter estimate is made by a BGLM and a BBGAMLSS.

In Table 2 we report the means and the standard deviations of the number of surrenders estimated by BGLM, making a comparison with the observed values.

As one can see, the expected values are very close, but BGLM strongly underestimates the standard deviations. Data show very large dispersion; the latter is due to a probability of success that varies around a value and is not fixed as in the binomial case. This additional uncertainty could be better modeled assuming a BB distribution.

Hence, we get a parameter estimate of a BBGAMLSS model, where random effects are not considered. Furthermore, we choose a logit function as a link for μ and a logarithmic one for σ, in order to get a value included in the interval $(0, 1)$ for μ and a positive value for σ.

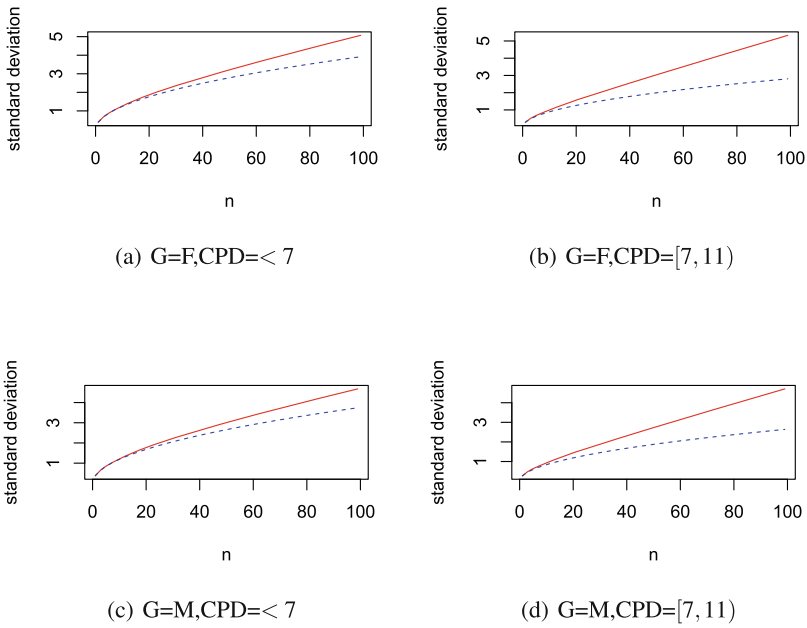
In Table 3 we report the means and the standard deviations of the number of surrenders estimated by BBGAMLSS, making a comparison with the observed values.

Table 3. BBGAMLSS estimates compared with observed means and standard deviations

G	CPD	Estimated mean	Observed mean	Estimated standard deviation	Observed standard deviation
F	< 7	637.75	655.20	114.10	323.13
F	[7,11)	195.22	184.40	99.22	75.60
F	[11,15)	66.65	63.80	12.32	39.61
F	[15,17)	3.57	4.40	1.86	4.45
F	≥ 17	0.38	0.80	0.75	1.60
M	< 7	567.36	563.80	97.60	252.67
M	[7,11)	175.91	168.40	85.86	83.69
M	[11,15)	60.80	63.00	11.20	37.30
M	[15,17)	3.38	2.80	1.80	3.31
M	≥ 17	0.44	0.00	0.83	0.00

Observing Table 2 and Table 3, it is worth noting that even though the two models provide very close estimates of the conditional means, the standard deviations of the conditional distributions are very different.

The BBGAMLSS estimates are closer to observed standard deviations than the BGLM case.



(a) G=F,CPD=< 7 (b) G=F,CPD=$[7, 11)$

(c) G=M,CPD=< 7 (d) G=M,CPD=$[7, 11)$

Fig. 1. Standard deviations for 4 risk classes. Binomial and BB comparison

Finally, we analyze the conditional predicted standard deviations in binomial and BB cases. In Fig. 1, the standard deviations of the two random variables by varying n are reported for 4 risk classes; the risk classes are obtained by combining levels of G and levels < 7 and $[7, 11]$ of CPD. As one can see, the standard deviation increases as n increases in each case and is always greater in the BB case (continuous line) than binomial case (dotted line). The difference is larger as n increases. This is in line with the BB variance formula in Eq. (2).

References

1. Baione, F., Biancalana, D., De Angelis, P.: A two-part beta regression approach for modeling surrenders and withdrawals in a life insurance portfolio. N. Am. Actuar. J. **27**(2), 380–395 (2023)
2. Cribari-Neto, F., Ferrari, S.L.P.: Beta regression for modelling rates and proportions. J. Appl. Stat. **31**(7), 799–815 (2004)
3. Kim, C.: Modeling surrender and lapse rates with economic variables. N. Am. Actuar. J. **9**(4), 56–70 (2005)
4. Rigby, R.A., Stasinopoulos, D.M.: Generalized additive models for location, scale and shape. Appl. Stat. **54**(3), 507–554 (2005)

A Comparison of Beta Regression and Copula Regression for Partial Lapse Rate Estimate

Fabio Baione(✉), Davide Biancalana, and Andrea Santoro

Sapienza University, Rome, Italy
{fabio.baione,davide.biancalana,andrea.santoro}@uniroma1.it

Abstract. In actuarial analysis, it is very useful to analyze the behavior of an interval-bounded random variable, as a percentage, a proportion, or a fraction, conditioned to other explanatory variables. For this kind of variables, considering the presence of bounds, in general in (0,1), the estimate of the conditional mean and/or conditional quantiles is more trivial than other continuous or discrete variables. This work aims to show the application of a copula-based regression model on the study of the percentage lapsed in partially lapsed life policies, demonstrating how this approach can be an effective and powerful tool compared to an alternative standard regression models like Beta Regression.

Keywords: Copula Regression · Copula Quantile Regression · Beta Regression · Pair Copula Construction · Lapse

1 Introduction

The analysis of continuous response variables limited to intervals of finite length is relevant in a wide variety of disciplines to study percentages, proportions, or fractions. Some of the regression approaches for the open interval in $(0, 1)$ are described in Kieschnick & McCullough [7]. An application in the actuarial field of Beta Regression (Ferrari & Cribari Nieto [6]) extended to allow the closed interval is provided by Baione et al. [3]. This work aims to show how a copula-based regression approach can be an alternative tool to support the user in the study of bounded variables.

Starting from the famous Sklar's theorem (Sklar [10]), statistical copulas have played a central role in the analysis of the dependence structure in multiple contexts. One of the key qualities of this approach is the parting of the estimate of marginal distributions from the estimate of the dependence structure. Hence, in a regression context, copulas can provide a valid extension for the study of bounded variables to those already existing in the literature. Inevitably copula regression (CR) and Quantile Copula Regression (CQR) also have their drawbacks - such as the high number of parameters or difficult interpretation of the results particularly in the case of categorical variables. However, CR and CQR

M. Corazza et al. (Eds.): MAF 2024, *Mathematical and Statistical Methods for Actuarial Sciences and Finance*, pp. 25–30, 2024.
https://doi.org/10.1007/978-3-031-64273-9_5

have the advantage of carrying out the analysis of the central value as well as the non-symmetric prediction intervals by estimating the parameters only once, among others.

Before observing the methodological aspects, it is underlined that although it is possible to proceed using n-dimensional copulas (n-copulas) in the case of both discrete and continuous variables - as can be observed in Ahn et al. [2] - in the present work we used the Pair Copula Construction (PCC) which allows you to increase the flexibility of the model using bivariate copulas. For an extensive and exhaustive reference in PCC, CR and CQR readers can refer to Czado [4].

2 Methodological Approach

The PCC approach consists of decomposing the multivariate probability density function (pdf) or probability mass function (pmf), in the discrete case, into the marginal pdf/pmf of the variables and the dependence structure defined by bivariate copulas. This approach allows two main advantages linked to bivariate copulas: the wide range of choices for the bivariate distribution and the flexibility of being able to use different copulas for each pair of variables.

An overview of PCC in the continuous and discrete case is shown in Panagiotelis et al. [8], in this analysis the class of Drawable-Vines or, more commonly, D-Vines (Aas et al. [1]) is used. The D-Vine with $n+1$ variables is made up of a sequence of n trees, in which each node is connected to two edges, except the first and last which are connected to a single edge, furthermore each edge describes a pair-copula density (conditional or not on a subset of the other variables). With this choice, the number of possible D-Vines is $\frac{(n+1)!}{2}$, as it depends exclusively on the initial ordering.

In order to introduce the regression approach with copulas, let a response variable Y and a set of explanatory variables $X_i, i = 1, \ldots, n$ be defined. Applying the probability integral transform we obtain the respective values in the unit interval $F_Y^{-1}(Y) = V$ dependent variable and the set $F_{X_i}^{-1}(X_i) = U_i, i = 1, \ldots, n$ of n explanatory variables. In the presence of both discrete and/or continuous variables, "joined" in pairs by a bivariate copula function C, the joint density of two variables (e.g. U_i and U_j) depends on their nature and similarly occurs for the conditional distribution. To define C it is necessary to explain the h-functions, which correspond to the conditional distributions of the marginals. If both X_i (U_i) and X_j (U_j) are continuous they are defined as

$$h_{i|j}(u_i|u_j) = \frac{\partial C(u_i, u_j)}{\partial u_j}. \tag{1}$$

Indeed, when we consider two discrete variables, defining $u_{i_1}, u_{j_1}, u_{i_2}, u_{j_2} \in [0, 1]$ with $u_{i_1} > u_{i_2}$ and $u_{j_1} > u_{j_2}$ we have:

$$\begin{cases} \hat{h}_{i|j}(u_i|u_{j_1}, u_{j_2}) = \frac{C_{i,j}(u_i, u_{j_1}) - C_{i,j}(u_i, u_{j_2})}{u_{j_1} - u_{j_2}} \\ \hat{h}_{j|i}(u_j|u_{i_1}, u_{i_2}) = \frac{C_{i_j}(u_{i_1}, u_j) - C_{i,j}(u_{i_2}, u_j)}{u_{i_1} - u_{i_2}}. \end{cases} \tag{2}$$

For what concerns statistical regression, Schallhorn et al. [9] shows how to derive the conditional distribution functions iteratively. If we model the joint distribution of $(V, \boldsymbol{U})$ by a D-Vine, with order $V, U_{l_1}, \ldots, U_{l_n}$ with $(l_1, \ldots, l_n)'$ being a permutation of $(1, \ldots, n)$, then we get:

$$C_{V|\boldsymbol{U}}(v|\boldsymbol{u}) = \begin{cases} h_{V|U_{l_n};\boldsymbol{U}_{-l_n}}(C_{V|\boldsymbol{U}_{-l_n}}(v|\boldsymbol{u}_{-l_n})|C_{U_{l_n}|\boldsymbol{U}_{-l_n}}(u_{l_n}|\boldsymbol{u}_{-l_n})) \\ \hat{h}_{V|U_{l_n};\boldsymbol{U}_{-l_n}}(C_{V|\boldsymbol{U}_{-l_n}}(v|\boldsymbol{u}_{-l_n})|C_{U_{l_n}|\boldsymbol{U}_{-l_n}}(u_{l_n}|\boldsymbol{u}_{-l_n}), C_{U_{l_n}|\boldsymbol{U}_{-l_n}}(u_{l_n}^-|\boldsymbol{u}_{-l_n})) \end{cases} \tag{3}$$

with $u_{l_i}^- = F_{X_{l_i}}(x_{l_i}^-) = \lim_{a \to x_{l_i}} F_{X_{l_i}}(a)$.

In order to use copula regression in the PCC approach it is possible to extend the analysis presented in Sungur [12].

The CR function of V with respect to $\boldsymbol{U}$ is defined by the following Stieltjes integral

$$\boldsymbol{E}_C[V|\boldsymbol{U}] = \int_0^1 v dC_{V|\boldsymbol{U}}(v|\boldsymbol{u}) \tag{4}$$

Therefore, it is possible to obtain the CR function of Y with respect to the other variables

$$\boldsymbol{E}[Y|\boldsymbol{X}] = F_Y^{-1}\left(\boldsymbol{E}_C[V|\boldsymbol{U}]\right) \tag{5}$$

with F_Y which corresponds to the distribution function of Y.

Without estimating further parameters, following Schallhorn et al. [9], if we want to observe the conditional quantile of Y with respect to the other n variables

$$\boldsymbol{Q}_\alpha[Y|\boldsymbol{X}] = q_\alpha(\boldsymbol{x}) = F_{Y|X_1,\ldots,X_n}^{-1}(\alpha|\boldsymbol{x}) = F_Y^{-1}(C_{V|\boldsymbol{U}}^{-1}(\alpha|\boldsymbol{u})) \tag{6}$$

where $\alpha \in (0, 1)$ corresponds to the probability level of the quantile.

In the continuous case, it is obvious that the term $C_{V|\boldsymbol{U}}^{-1}(\alpha|\boldsymbol{u})$ is obtained by inverting the h-functions in (1), whereas in the discrete case we need to define the value $\beta \in [0, 1]$ such that:

$$C_{V|\boldsymbol{U}}^{-1}(\alpha|\boldsymbol{u}) = \underset{C_{V|U}(\beta|\boldsymbol{u}) \geq \alpha}{\arg\min} \left(C_{V|\boldsymbol{U}}(\beta|\boldsymbol{u}) - \alpha\right) \tag{7}$$

It is very important to note that in Eq. (6) the marginal distribution of the dependent variable plays a predominant role in the estimation of the conditional quantile.

3 Numerical Application

In the following application, we consider a portfolio of $\ell = 1,625$ (partially) lapsed life insurance policies. The response variable is the percentage of the reserve lapsed by the policyholder in a year ($Lapse\%$). Whereas we consider six explanatory variables, both continuous, discrete and categorical, listed below:

- `RA` - the amount of mathematical reserve at the beginning of the year (continuous);
- `LD` - the lapse duration (discrete): the observations of this variable are in the range between 3 years and 18 years from contract issue;
- `PA` - the amount of premiums paid (ordered categorical): we consider four classes as follows, less than €10,000 (class 1), between €10,000 and €20,000 (class 2), between €20,000 and €30,000 (class 3), over €30,000 (class 4);
- `PT` - the type of premium payment (categorical): single premium (S) or periodic premium (P);
- `Gender` - policyholder gender (categorical): Male (M), Female (F);
- `Age` - policyholder age at inception (discrete): between 14 years and 85 years.

The categorical variables have been pre-processed so that they can be managed as discrete variables. For binary responses this pre-processing does not influence the outcomes.

To carry out the analysis, the *R* package `vinereg` is used, which allows us to obtain the structure of the D-Vine - implementing via the Dißmann algorithm (Dißmann et al. [5]) - to select the underlying copulas and to estimate the corresponding parameters.

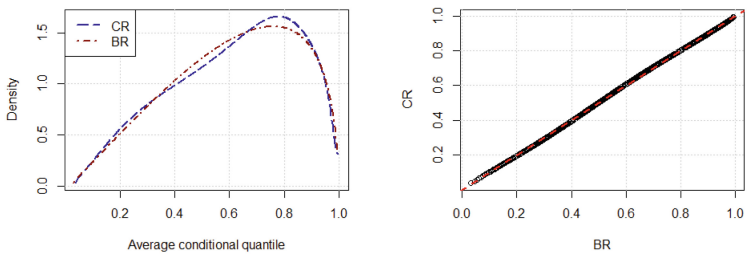
In this application, the estimation of the marginal distributions of each variable is left to the `vinereg` package, which select a kernel density estimate (with bounded support) for continuous variables and a jittered kernel density estimate for discrete variables. The aim is to show how in some contexts the possibility of performing a regression on the dependency structure, without worrying about the underlying distribution, allows this tool to lend itself to different contexts. The model (CR) observed to describe the dependence structure uses exclusively parametric bivariate copulas, therefore kernel distributions such as the Transformation Local Likelihood Copula (see Sumarjaya [11]) are excluded.

The order of variables selected by the D-Vine model is: `Lapse%, PT, RA, PA, Gender, LD, Age`.

In order to demonstrate whether CR can be a suitable and effective tool in the regression analysis of a bounded variable, we compare the CR model with a standard Beta Regression (BR) model (Ferrari & Cribari-Nieto [6]) with a probit link function. By noting that the observed average is 61.56%, the estimated average with CR is 61.66% whereas the value provided by BR is 60.85%. It is worth noting that this outcome can be justified by a different number of parameters used for CR calibration (39) compared to the BR (8).

Then, as previously explained, once the vine structure has been estimated, the same parameters allow determining the quantiles. So, we compare the average distribution of the conditional quantiles of the different observations estimated by the two models.

In Fig. 1 panel (a) we show the density of the conditional response variable obtained as the average, for each probability level in $\boldsymbol{\alpha} = \{0.01, 0.02, \ldots, 0.99\}$, of the conditional quantiles of the each model as: $\sum_{i=1}^{\ell} \boldsymbol{Q}_{\alpha}[Y|\boldsymbol{x}^{(i)}]/\ell$. As one can see the shape of the BR curve is smoother than the CR one as the latter is estimated by a univariate local polynomial kernel density estimators while BR



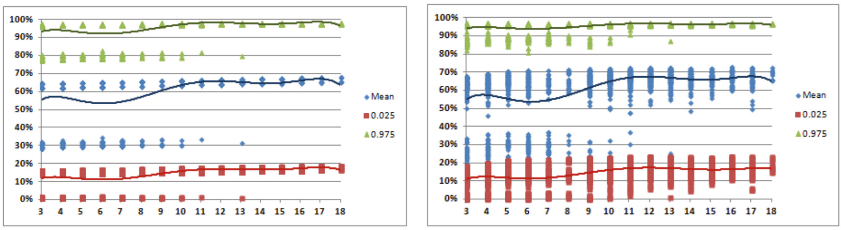
(a) Density plot (b) Q-Q plot

Fig. 1. Comparison between observed and fitted pdf of the conditional response variable - CR (long dash line) and BR (dot dash line) and Q-Q plot graph

is based on a parametric approach. In panel (b) the Q-Q plot of the average conditional quantiles are reported in order to put in evidence the strong likeness of the estimate obtained with the CR and BR.

Finally, to investigate the differences between predictions of CR against BR, we look at the Lapse Duration (`LD`) variable. In Fig. 2 we visualize the influence of `LD` by calculating the mean (circles) and the percentile levels 2.5% (squares) and 97.5% (triangles) for each $\boldsymbol{x}^{(i)}, i = 1, \ldots, \ell$, then we plot it against $x_{ij}, j = 1, \ldots, n$ and add a smooth curve through each point clouds.

It can be seen how the two models provide different realizations albeit with consistent behaviors as remarkable by the smoothed lines. Specifically, the BR shows for each `LD`'s value that lapse rates are clustered into two groups whereas the CR shows more scattered values. The main motivation for these differences is explained by strong influence of the variable premium payment type `PT`, which assumes value P or S, and in the BR model is the most significant variable that influences the partial lapse duration.



(a) BR (b) CR

Fig. 2. Influence of Lapse Duration on the response variable using the regression models.

References

1. Aas, K., Czado, C., Frigessi, A., Bakken, H.: Pair-copula constructions of multiple dependence. Insur. Math. Econ. **44**(2), 182–198 (2009)
2. Ahn, J.Y., Fuchs, S., Oh, R.: A copula transformation in multivariate mixed discrete-continuous models. Fuzzy Sets Syst. **415**, 54–75 (2021)
3. Baione, F., Biancalana, D., De Angelis, P.: A two-part beta regression approach for modeling surrenders and withdrawals in a life insurance portfolio. N. Am. Actuar. J. **27**(2), 380–395 (2023)
4. Czado, C.: Analyzing dependent data with vine copulas. Lecture Notes in Statistics. Springer, Heidelberg (2019)
5. 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)
6. Ferrari, S., Cribari-Neto, F.: Beta regression for modelling rates and proportions. J. Appl. Stat. **31**(7), 799–815 (2004). https://doi.org/10.1080/0266476042000214501
7. Kieschnick, R., McCullough, B.D.: Regression analysis of variates observed on (0, 1): percentages, proportions and fractions. Stat. Model. **3**(3), 193–213 (2003)
8. Panagiotelis, A., Czado, C., Joe, H.: Pair copula constructions for multivariate discrete data. J. Am. Stat. Assoc. **107**(499), 1063–1072 (2012)
9. Schallhorn, N., Kraus, D., Nagler, T., Czado, C.: D-vine quantile regression with discrete variables. arXiv preprint arXiv:1705.08310 (2017)
10. Sklar, M.: Fonctions de repartition an dimensions et leurs marges. Publ. Inst. Statist. Univ. Paris **8**, 229–231 (1959)
11. Sumarjaya, I.W.: A survey of kernel-type estimators for copula and their applications. In: Journal of Physics: Conference Series, vol. 893, no. 1, p. 012027. IOP Publishing (2017)
12. Sungur, E.A.: Some observations on copula regression functions. Commun. Stat.-Theory Methods **34**(9–10), 1967–1978 (2005)

Input Relevance in Multi-Layer Perceptron for Fundraising

Diana Barro[1], Luca Barzanti[2], Marco Corazza[1], and Martina Nardon[1(✉)]

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

[2] Department of Mathematics, University of Bologna, Piazza di Porta San Donato 5, 40126 Bologna, Italy
luca.barzanti@unibo.it

Abstract. In this contribution, we consider a Multi-Layer Perceptron (MLP) methodology for predicting specific gift features, particularly the count of donations and the gift amounts. Moreover, we use Garson's indicator to evaluate the relative importance of the input variables to the output(s) in the MLP model with the aim of enhancing the effectiveness of fundraising campaigns. In the discussed application, the Donors' behaviors are estimated using a simulated dataset that includes individual characteristics and information about donation history.

Keywords: Multi-Layer Perceptron · Input relevance · Garson's indicator · Fundraising Management

1 Introduction

The optimization of a fundraising (FR) campaign, namely the maximization of the estimated global return under budget constraints, relies on the selection of the most promising Donors and requires an efficient use of available information. Accurate estimates of the number of donations, their amounts, and the gift probability are based on Donors' individual characteristics and donation history about past campaigns. These are crucial in evaluating the result of a fundraising campaign. Parametric and non-parametric approaches can be used to estimate the quantities of interest.

Recently, to estimate the gift probability [2] discusses statistical parametric methodologies and suggests modelling the number of gifts as a Poisson random variable with an intensity parameter that depends on Donors' characteristics. A Poisson regression can then be used to estimate the expected number of donations, the probability of gift, and to assign a score to each Donor measuring their propensity to the donation.

The development of non-parametric Machine Learning (ML)-based models for FR is a very recent research stream; see, for example, [4] and [3]. Along this line, [1] proposed a Multi-Layer Perceptron (MLP) to predict the number of

M. Corazza et al. (Eds.): MAF 2024, *Mathematical and Statistical Methods for Actuarial Sciences and Finance*, pp. 31–36, 2024.
https://doi.org/10.1007/978-3-031-64273-9_6

donations and the gift amount and applied it to a simulated dataset of Donors' individual characteristics and donation history.

In this contribution, we extend the analysis carried out in [1], focusing on the relative importance of the input variables in the MLP model with the aim of enhancing the effectiveness of FR campaigns. In particular, in the analysis, we use the Garson's indicator. Section 2 presents the FR model and the dataset. Section 3 introduces the MLP and the input relevance. Section 4 discusses the application, and Sect. 5 concludes.

2 Data Collection and the FR Process

Associations collect in databases (DB) and manage information on Donors, Contacts and results from previous campaigns. Any gift received is associated with the Donor (a person, a company, or other entity), their available individual characteristics, and the gift history (gift events, timing and gift amounts). For large and medium-sized Associations, the information may include quantitative and qualitative features, besides advanced characteristics of the Donors' profile. For smaller ones, a systematic collection of information on Donors is very limited. The availability of data and their quantitative exploitation, together with the expertise of professionals in the field, are crucial elements in determining the success of a campaign.

In [2] and [1], the arrival of a donation, i.e. a 'gift', to an Association is modeled as the outcome of a random variable that can be analyzed according to four different dimensions: the *occurrence* of the donation, represented by a dichotomous variable; the *frequency* as a count variable measuring the number of donations received in a given period of time; the *timing* as a duration variable; and the *amount* of the donation as a positive variable.

Let x_n be the vector that collects selected observable individual characteristics of Donor n, with $n = 1, \ldots, N$. Define z_n as the vector of transformed individual characteristics, where qualitative features are properly transformed into quantitative ones or dummy variables.

Individual profile variables can be divided into personal situation variables (gender, age, number of children, education, place of origin, size of residence town, etc.); economic and financial situation variables (wage, wealth, investments); risk aversion variables (as a proxy of which the number of insurance contracts is taken); and other information such as personal interests, religious involvement, social network, geographical distances and involvement in the campaign subject, among others.

This contribution aims to assess the relevance of these input features in explaining the output elements of an FR campaign, as measured by the count of donations and the gift amounts. To this purpose, we apply and extend the MLP methodology proposed in [1].

Table 1. Some individual characteristics along the Giving Pyramid, with a finer segmentation for the Sporadic and Regular Donors

Donors	Low wealth	Insurance policies ≥ 1	Min gift amount	Max gift amount
Sporadic (sd1)	70%	35%	20	50
Sporadic (sd2)	70%	35%	30	100
Regular (rd1)	40%	65%	50	400
Regular (rd2)	40%	65%	100	500
Large	10%	65%	300	1000

We introduce the simulated[1] DB, which will be used to test the proposed methodology. The dataset is composed by $N = 30\,000$ Donors. The segmentation is as follows: 75% are *Sporadic Donors* (among them, about 25% made only one donation), 19% *Regular Donors*, and 6% *Large Donors*.

Personal profile variables collected are: age and number of children, education[2] (in four categories: Master and Ph.D., Bachelor, High School, other/lower school level), wealth (measured in thousands of euro), risk aversion (measured as numbers of insurance policies signed by the Donor).

The database includes the gift history for each Donor: the number of donations, the average gift amount, and the number of gift requests. Table 1 reports some individual characteristics according to the segmentation in the Giving Pyramid for the data collected in the DB.

3 Basics on MLP and Input Relevance

In this section, we first provide some basics of the supervised ML tool, known as MLP, and then we introduce the approach used to determine the relevance of the input variables to the output one(s).

According to a well-known metaphor, an MLP can be considered as a computational model inspired by the structure and functioning of the biological neural networks that comprise the brain of superior living beings. An MLP is a simple Artificial Neural Network where neurons (or nodes) represent units of computation of the network.

The nodes are organized into layers; typically: an *input layer*, whose nodes (sensors) receive the data from the external environment; one or more hidden

[1] The DB is constructed from experts' knowledge and based on a realistic composition of a set of Donors. The database has already been used in [1].

[2] A categorical variable transformed into values ranging from 1 to 4, assigning 4 to the highest category.

layers, whose nodes carry out the computational tasks[3]; and an *output layer*, whose nodes (devices) release the result of the computation towards the external environment. All the nodes in one layer are fully connected to the nodes in the next one, but not among those within the same layer.

Note that in supervised ML, the MLP is trained on a labeled dataset, meaning that during the phase of parameters estimation, the MLP is presented with a dataset

$$\{(z_{1,n}, \ldots, z_{i,n}, \ldots, z_{I,n};\, o_{1,n}, \ldots, o_{k,n}, \ldots, o_{K,n}), n = 1, \ldots, N\},$$

where $(z_{i,n})_{i=1,\ldots,I}$ is the n-th vector of input features, $(o_{k,n})_{k=1,\ldots,K}$ is the associated vector of output labels, and N is the dimension of the dataset. Pairs of nodes belonging to consecutive layers are associated with weights, namely v_{ij} and w_{jk}, representing the strength of the connections. With reference to the MLP, the hyperparametrization phase provided the best configuration for the number of hidden layers and nodes. In our application, a single-hidden-layer MLP proved to be the best architectural structure identified for making predictions about Donor's behaviors (see [1]).

As for the training phase, the weights are adjusted in an iterative procedure over the training dataset. This stage starts with a random initialization of the weights; in subsequent runs of the MLP, the weights are chosen in order to minimize a suitable error metric based on the distance between the computed and actual outputs. The training process ends when a pre-fixed stopping criterion is satisfied, and then the obtained optimal weights are used in the validation phase.

It is well-known that the design complexity of an MLP, which can be seen as a black box, does not provide a direct interpretation of the obtained weights. In contrast, it is important to assess the impact of the explanatory inputs on the output variable(s).

Our main aim is to study the relevance of inputs to explain the outputs and to obtain relative rankings and ratings of the input features. Over the years, various methods have been proposed in the literature; for a review, see [6] and [8]. In particular, to assess the relative importance of Donors' features (input variables) on the FR campaign results (output variables), we use Garson's indicator (see [5]):

$$G_{ik} = \frac{\sum_{j=1}^{J} \left(|v_{ij} \cdot w_{jk}| \Big/ \sum_{p=1}^{I} |v_{pj}| \right)}{\sum_{q=1}^{I} \sum_{j=1}^{J} \left(|v_{qj} \cdot w_{jk}| \Big/ \sum_{p=1}^{I} |v_{pj}| \right)},$$

where G_{ik} denotes the relevance of the i-th input variable, with $i = 1, \ldots, I$, to the k-the output variable, with $k = 1, \ldots, K$, and J is the number of nodes in the hidden layer. It is noteworthy that Garson's indicator is nonnegative

[3] This represents the "intelligent" part of the computation; where the adjective intelligent means that, under mild assumptions, an MLP "*can approximate virtually any function of interest to any desired degree of accuracy* [...]." (see [7, p. 360].).

and normalized to sum to 1 overall input relevances, thus measuring the relative importance of the inputs but without indicating the sign of the importance itself.

4 Applications and Results

We aim to assess the relevance of the Donors' characteristics on the count of donations and gift amounts. In our investigation, we test the following two MLP-based prediction models:

$$cd = f_{MLP,1}\,(ga, ag, nc, ed, we, ra, co)\,,$$
$$ga = f_{MLP,2}\,(cd, ag, nc, ed, we, ra, co)\,,$$

where cd denotes the count of donations, ga specifies the gift amount, ag, nc, ed, we, ra, and co indicate the Donor's age, number of children, education level, wealth, risk aversion, and number of contacts, respectively (see Sect. 2). In models $f_{MLP,1}$ and $f_{MLP,2}$, the cd and ga forecasts are based on the other $I-1$ inputs.

On the basis of the validation results for these models, detailed in [1], we compute the related Garson's indicators.

In Tables 2a and 2b, we present the results of the analyses of input relevance for the prediction models $f_{MLP,1}$ and $f_{MLP,2}$, respectively. In both tables, in column 1, the input variables are ranked in decreasing order according to their relevance to the output, and in column 2, the values of the Garson's indicator for the same input variables are reported.

Table 2. Results related to the analyses of input relevance to the output

(a)Model $f_{MLP,1}$, with the count of donations (cd) as the output variable

Input variable	G_{i1}
Gift amount (ga)	64.43 %
Risk aversion of the D. (ra)	17.87 %
No. of children of the D. (nc)	7.56 %
Wealth of the D. (we)	5.49 %
Education level of the D. (ed)	1.94 %
No. of contacts of the D. (co)	1.61 %
Age of the D. (ag)	1.11 %

(b)Model $f_{MLP,2}$, with the gift amount (ga) as the output variable

Input variable	G_{i1}
Risk aversion of the D.(ra)	34.45 %
Count of donations (cd)	34.34 %
Wealth of the D. (we)	17.75 %
Education level of the D. (ed)	6.83 %
No. of contacts of the D. (co)	2.90 %
Age of the D. (ag)	2.72 %
No. of children of the D. (ga)	1.02 %

It is worth noting that in $f_{MLP,1}$ the gift amount is the most relevant input factor to predict the count of donations; such a result is coherent with the segmentation of the Giving Pyramid and data reported in Table 1. Whereas, when considering model $f_{MLP,2}$, the count of donations is the second most relevant

input in predicting the gift amount. This highlights a significant mutual dependence between these factors.

Moreover, the (decreasing) order of the input variables according to their relevance presents analogies for both prediction models. It turns out that risk aversion (ra) is highly relevant in both cases, and wealth (we) is more relevant in predicting the amount of donations, but it is also important for the number of donations. These results are also consistent with the data reported in Table 1.

Finally, in both prediction models, the first three to four most relevant input variables collectively contribute to "explain" more than 85% to around 95% of the associated output variable. These values are relative weights that are computed on a number $I - 1$ of available inputs. This result emphasizes the relative importance of these input variables.

5 Concluding Remarks

In this contribution, we considered two MLP-based models to predict the number of donations and the gift amount. Among those proposed in the literature, we applied the Garson's indicator to assess the relative importance of the input variables to the output. The main findings show that it is possible to rank the inputs based on their impact on the output and that a relatively small set of input variables can well predict the variables of interest. These results help identify which inputs are the most relevant for the forecast of the expected gift and can be exploited to implement effective FR campaigns. For future research, the analysis will be extended to consider indicators alternative to Garson's.

References

1. Barro, D., Barzanti, L., Corazza, M., Nardon, M.: Machine Learning and Fundraising: Applications of Artificial Neural Networks. University Ca' Foscari of Venice, Department of Economics Research Paper Series, 33/WP/2023 (2023)
2. Barzanti, L., Nardon, M.: Estimation of the gift probability in fund raising management. In: Corazza, M., Perna, C., Pizzi, C., Sibillo, M. (eds.) Mathematical and Statistical Methods for Actuarial Sciences and Finance. MAF 2022, pp. 70–75. Springer, Cham (2022)
3. Cagala, T., Glogowsky, U., Rincke, J., Strittmatter, A.: Optimal targeting in fundraising: a machine-learning approach. CESifo Working Papers, 9037 (2021)
4. Farrokhvar, L., Ansari, A., Kamali, B.: Predictive models for charitable giving using machine learning techniques. PLoS ONE **13**(10), 1–14 (2021)
5. Garson, G.D.: Interpreting neural-network connection weights. Artif. Intell. Expert **6**(4), 46–51 (1991)
6. Horel, E., Giesecke, K.: Significance tests for neural networks. J. Mach. Learn. Res. **21**(1), 1–29 (2020)
7. Hornik, K., Stinchcombe, M., White, H.: Multylayer feedforward networks are universal approximators. Neural Netw. **2**(5), 359–366 (1989)
8. Mandel, F., Barnett, I.: Permutation-based hypothesis testing for neural networks. arXiv:2301.11354v1 (2023)

Art as a Financial Asset in Portfolio Allocation

Diana Barro[1(✉)], Antonella Basso[1], Stefania Funari[2], and Guglielmo Alessandro Visentin[2]

[1] Department of Economics, Ca' Foscari University of Venice, Cannaregio 873, 30121 Venice, Italy
{d.barro,basso}@unive.it
[2] Department of Management, Ca' Foscari University of Venice, Cannaregio 873, 30121 Venice, Italy
{funari,guglielmo.visentin}@unive.it

Abstract. With increasing interest in recent years, investors have devoted attention to the art market as a potential alternative asset class to be included in portfolios. Besides cultural and aesthetic value, investment in art could provide diversification benefits, and a hedge against market volatility and inflation. We use data on the global art auction market to analyze this asset's risk-return characteristics, and to investigate its inclusion in optimal portfolios, based on the classical Markowitz mean-variance and the mean-CVaR optimization models. The results show that art makes an interesting asset, which enters the optimal portfolios.

Keywords: Art investment · Alternative assets · Portfolio allocation · Mean-variance · Mean-CVaR

1 Introduction

In the last decades, uncertainty in financial markets has led many investors to diversify their portfolios with a mix of standard and alternative investments. Among alternative assets, art is drawing considerable attention among investors and wealth managers. Suffice it to say that ultra-high-net-worth individuals hold, on average, around 5% of their portfolios in art and collectibles [1] and 74% of asset managers in 2023 offered art wealth management services [2].

Besides finance professionals, scholars in finance have also shown interest in the art market. Starting with some isolated pioneering studies in the 1970 s and 1980 s such as the contributions of Stein [3] and Baumol [4], who estimated the rate of return on art investment, the literature on art investment has grown considerably, exploring a broad spectrum of financial themes related to the art market, as outlined in Barro et al. [5]. In particular, a number of studies over the years have investigated the role of art in portfolio allocation. Campbell [6] shows that investing a small fraction of wealth in art improves the trade-off

M. Corazza et al. (Eds.): MAF 2024, *Mathematical and Statistical Methods for Actuarial Sciences and Finance*, pp. 37–42, 2024.
https://doi.org/10.1007/978-3-031-64273-9_7

between risk and return of the portfolio, while Korteweg et al. [7] argue that a better diversification is achieved by targeting specific art market segments and Tucker et al. [8] report a relevant percentage allocated to art asset. On the other hand, Worthington and Higgs [9] find that art does not provide any diversification gains, thus it is still unclear whether art should be included in a financial portfolio.

The main aim of this contribution is to investigate the role of art as an alternative asset in portfolio allocation. We consider global art auction market data on an extended and up-to-date time period; this allows us to verify whether the recent increased interest in art investments has led to an expansion of the presence of art in optimal portfolio allocations, with respect to the previous results presented in the literature, obtained with much less recent data.

Section 2 presents the data and the models employed; Sect. 3 reports the results of the portfolio allocation, and Sect. 4 concludes.

2 Data and Methodology

To carry out the analysis, we consider the perspective of an US investor who holds a well diversified portfolio of both standard and alternative investments. The data on financial markets are retrieved from Bloomberg and include: an equity market index, S&P 500 Index (SPX), a bond market index, Bloomberg Barclays U.S. Aggregate Bond Index (LBUSTRUU), the Gold to USD rate (XAUUSD), and a real estate market index, Real Estate FTSE NAREIT all equity REITS index (FNERTR).

The art price data come instead from the *All Art Index* provided by Art Market Research (AMR), a London-based firm specialized in data analysis for collectibles' market. The All Art Index (henceforth AMRAAI) monitors the most important auction houses worldwide, and all artists who sold there at least one artwork in the past 24 months are included in the computation. In detail, AMR employs a 24 months weighted moving average to determine an average price representative of each artist in its database – where smaller weights are associated to older sales – and to compute an aggregate index, all average prices are added up on a monthly basis. The AMRAAI, which is exclusive of buyer's premium, is expressed in pound sterling, and data are available with a monthly frequency, from January 1978.

We use a dataset spanning from January 1978 to December 2022. We convert the AMRAAI to a US dollar-denominated index using the monthly GBP to USD exchange rate and we adjust all the indices for US inflation, using the Consumer Price Index of the Federal Reserve Bank of St. Louis, to account for the change in the purchasing power over this extended time period.

From an analysis of the monthly returns, we find that art returns are affected by several periodic spikes, spikes that are particularly high in May and November, as confirmed by an inspection of the correlogram (not reported here for the sake of brevity). This peculiar behavior is due to the periodicity of the most important auctions, during which the most expensive works of art are auctioned.

Table 1. Returns' descriptive statistics

Statistic	SPX	LBUSTRUU	XAUUSD	FNERTR	AMRAAI
Mean	0.0431	0.0160	0.0155	0.0463	0.0354
Std. dev	0.1015	0.0465	0.1233	0.1151	0.1147
Maximum	0.2523	0.1708	0.4259	0.2949	0.3741
Minimum	–0.3391	–0.1361	–0.2637	–0.4956	–0.2655
Skewness	–0.5825	0.0531	0.6183	–1.1238	0.2321
Kurtosis	4.2811	4.9091	3.8517	7.2441	3.6158
VaR (95%)	0.1326	0.0611	0.1878	0.1559	0.1613
CVaR (95%)	0.2087	0.0940	0.2382	0.2513	0.2073

Table 2. Returns' correlation matrix

	SPX	LBUSTRUU	XAUUSD	FNERTR	AMRAAI
SPX	1				
LBUSTRUU	0.223	1			
XAUUSD	0.038	–0.02	1		
FNERTR	0.667	0.17	0.114	1	
AMRAAI	0.062	–0.076	0.156	–0.002	1

In this research, we focus on investors interested in high-end works of art, which are mainly traded in May and November auctions. We thus construct semi-annual indices based on May and November values, and this allows us to overcome the issues caused by the seasonal behavior of the AMRAAI returns. The descriptive statistics for semi-annual returns are displayed in Table 1, while Table 2 reports the returns' correlation matrix. As we may see, art performs reasonably well, compared to the other asset classes, exhibiting a relatively high average return and a positive skewness, but it has a quite high standard deviation. In addition, Table 2 shows that art is practically uncorrelated with the other asset classes, as found in several previous studies (e.g. Renneboog and Spaenjers [10]), and such a low correlation persists also with a monthly frequency.

To investigate the potential benefits of including art in a multi-asset portfolio, we first resort to the classical mean-variance Markowitz model [11]. The formulation of the model is reported below:

$$\begin{aligned} \min_{\boldsymbol{w}} \quad & \boldsymbol{w}^{'} \boldsymbol{\Sigma} \boldsymbol{w} \\ \text{s.t.} \quad & \boldsymbol{w}^{'} \boldsymbol{\mu} \geq h \\ & \boldsymbol{w}^{'} \boldsymbol{e} = 1 \\ & \boldsymbol{w} \geq \boldsymbol{0} \end{aligned} \tag{1}$$

where $\boldsymbol{w}$ is the vector of the asset weights, $\boldsymbol{\Sigma}$ is the return variance-covariance matrix, $\boldsymbol{\mu}$ is the vector of the return means, h is the portfolio target return fixed by the investor, $\boldsymbol{e}$ is a vector of 1 s, and $\mathbf{0}$ is the null vector.

In addition, in order to take into account the tail risk, we apply also the mean-CVaR model that tackles risk from a different perspective, as formulated in Rockafellar and Uryasev [12]:

$$\begin{aligned} \min_{\boldsymbol{w},u_k,\alpha} \quad & \alpha + \frac{1}{q(1-\beta)} \sum_{k=1}^{q} u_k \\ \text{s.t.} \quad & \boldsymbol{w}^{'} \boldsymbol{r}_k + \alpha + u_k \geq 0, \quad k = 1, ..., q \\ & u_k \geq 0, \quad k = 1, ..., q \\ & \boldsymbol{w}^{'} \boldsymbol{\mu} \geq h \\ & \boldsymbol{w}^{'} \boldsymbol{e} = 1 \\ & \boldsymbol{w} \geq \mathbf{0}, \alpha \in \mathbb{R} \end{aligned} \tag{2}$$

where $\boldsymbol{r}$ is a random vector of returns, α is the VaR of the portfolio with a confidence level β, u_k is an auxiliary variable which is equal to $(-\boldsymbol{w}^{'} \boldsymbol{r}_k - \alpha)^+$, and q is the number of scenarios generated.

3 Empirical Results

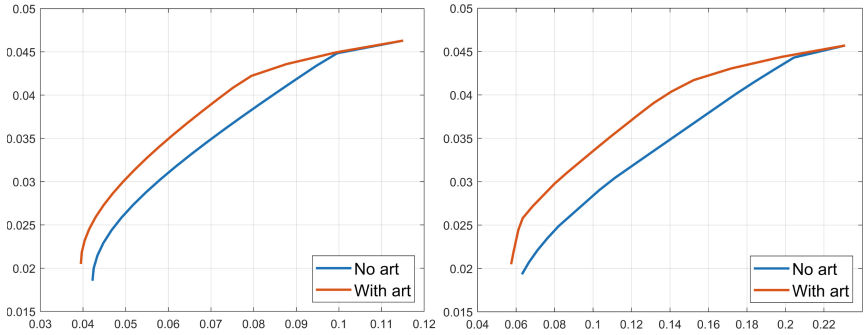
For both optimization models, the efficient frontier is obtained both in the case of portfolios consisting of all five assets considered, including art, and in the case where art is excluded from the portfolio. In the computations, h takes 20 equally spaced values between the return of the minimum variance portfolio ($h = 0.0205$) and the maximum attainable return ($h = 0.0463$).

In the CVaR portfolio optimization, $q = 10,000$ scenarios are generated for the asset returns based on the historical simulation, assuming that the distributions of the returns do not vary over time. The application of the Augmented Dickey-Fuller test, where the order of the model is selected using the Akaike Information Criterion, shows that all the time series of returns of the assets in our portfolio are indeed stationary at the 5% significance level (results are not reported but can be submitted upon request). The confidence level is $\beta = 0.95$.

The results show that investing in art enables the investor to obtain a better portfolio in terms of standard deviation and rate of return (Fig. 1a). Moreover, the results are confirmed for the CVaR model (Fig. 1b), and indeed the better performance seems substantial, more notably with the CVaR model.

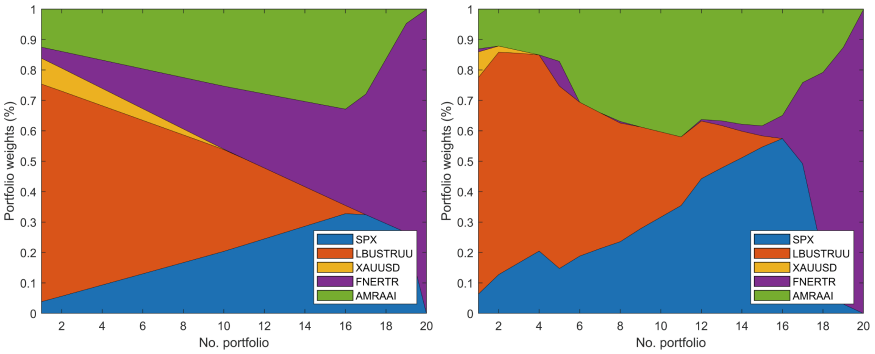
Figure 2 illustrates the composition of the optimal mean-variance and CVaR portfolios that include art. In both cases, the allocation in art is relevant, especially for more aggressive portfolios. For example, in optimal CVaR portfolio no. 10, over 30% of total wealth is allocated to art (see the left panel in Fig. 2).

However, it is unlikely that an institutional investor would allocate such a large proportion to an alternative investment asset. Therefore, we have extended



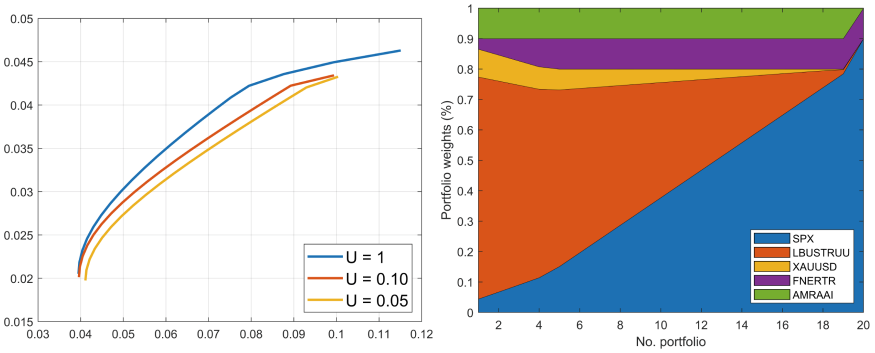
(a) Mean-variance, $h \in [0.0205, 0.0463]$ (b) Mean-CVaR, $h \in [0.0205, 0.0457]$

Fig. 1. Efficient frontiers of portfolios with art and with no art; expected return of the portfolio on the y-axis, risk measure (std. dev. and CVaR) on the x-axis



(a) Mean-variance (b) Mean-CVaR

Fig. 2. Efficient portfolio weights



(a) Efficient frontier as U varies (b) Efficient portfolio weights, $U = 0.10$

Fig. 3. Mean-variance constrained portfolios

our analysis introducing weight restrictions on the alternative asset positions; more precisely, we introduce an upper bound U on the weight of each alternative asset (XAUUSD, FNERTR, AMRAAI). For the sake of brevity, we present only the results for the mean-variance model. Figure 3a shows the behavior of the efficient frontier as the upper bound U varies; Fig. 3b displays the optimal allocation as the target return h varies, for $U = 0.10$. We may notice that the upper bound on the weight for art is always reached, with the exception of the most aggressive portfolio, which is composed only of equity and real estate.

4 Conclusions

The results obtained show that, even including the most recent data for the global art auction market, the optimal portfolios allocate a relevant share to art, in accordance with the results of several previous contributions in the literature, thus confirming that art makes an interesting asset for portfolio diversification purposes. On the other hand, some peculiarities of the art market, such as the low liquidity and the high transaction costs, could be explored more in depth, together with their effects on the portfolio allocation. Further research could also investigate the effects of the inclusion of additional alternative assets, e.g., commodities and cryptocurrencies.

References

1. Knight Frank, The wealth Report. 17th ed. (2023)
2. Deloitte, ArtTactic: Art & Finance Report. 8th ed. (2023)
3. Stein, J.P.: The monetary appreciation of paintings. J. Polit. Econ. **85**(5), 1021–1035 (1977)
4. Baumol, W.J.: Unnatural value - or art investment as floating crap game. Am. Econ. Rev. **76**(2), 10–14 (1986)
5. Barro, D., Basso, A., Funari, S., Visentin, G.A.: A bibliometric analysis of art in financial markets. Department of Management, Ca' Foscari University of Venice. Working Paper no. 05. 1–28 (2023)
6. Campbell, R.A.J.: Art as a financial investment. J. Alternat. Invest. **10**(4), 64–81 (2008)
7. Korteweg, A., Kraeussl, R., Verwijmeren, P.: Does it pay to invest in art? A selection-corrected returns perspective. Rev. Finan. Stud. **29**(4), 1007–1038 (2016)
8. Tucker, M., Hlawischka, W., Pierne, J.: Art as an investment: a portfolio allocation analysis. Manag. Financ. **21**(6), 16–24 (1995)
9. Worthington, A.C., Higgs, H.: Art as an investment: risk, return and portfolio diversification in major painting markets. Account. Finan. **44**(2), 257–271 (2004)
10. Renneboog, L., Spaenjers, C.: Buying beauty: on prices and returns in the art market. Manage. Sci. **59**(1), 36–53 (2013)
11. Markowitz, H.: Portfolio selection. J. Finan. **7**(1), 77–91 (1952)
12. Rockafellar, R.T., Uryasev, S.: Optimization of conditional value-at-risk. J. Risk **2**, 21–42 (2000)

A Robust Sustainability Assessment for SMEs Based on Multicriteria Decision Aiding

Diana Barro, Marco Corazza, and Gianni Filograsso(✉)

Department of Economics, Ca' Foscari University of Venice, Venice, Italy
{d.barro,corazza,gianni.filograsso}@unive.it

Abstract. Developing sustainable business standards for small and medium-sized enterprises (SMEs), entailing targets such as the minimization of carbon emissions, the promotion of equality and transparency in business processes, is key to achieve transition towards a greener economy. In light of the challenges in implementing and embedding sustainability practices in small firms, we consider the problem of evaluating the environmental, social, and governance (ESG) performance of listed European SMEs, w.r.t. different sustainability indicators, using a multicriteria decision analysis (MCDA) approach. Due to lack of agreement over the 'true' ESG metrics of firms discussed in the literature, we depart from the practice of constructing a composite indicator based on a sorting of firms, and we assess the stability of the results when the parameters of the preference model are uncertain. Moreover, with the aim of identifying possible sector-specific differences, the sustainability performance is gauged with reference to five economic sectors, based on a unique dataset of SMEs.

Keywords: ESG · Small and medium-sized enterprises (SMEs) · Multiple criteria Decision Analysis · Sustainability

1 Introduction

The sustainability assessment is becoming nowadays the crux of policies aiming at enforcing minimum Environmental, Social, and Governance (ESG) requirements, as sustainability measures may suffer from inconsistency of the data and how they are reported: existing ESG ratings from prominent data providers tend to disagree significantly on the very definition of sustainability drivers [2]. Such difficulties are compounded by the size of small and medium enterprises, which (i) deal with business data in an unstructured and non-standardised way, especially w.r.t. supply chain management [8], or (ii) may have little interest in managing their ESG performance [7], making the operationalization of sustainability practices difficult.

This contribution investigates the problem of sustainability assessment for small and medium enterprises, by using a multicriteria decision analysis (MCDA)

M. Corazza et al. (Eds.): MAF 2024, *Mathematical and Statistical Methods for Actuarial Sciences and Finance*, pp. 43–48, 2024.
https://doi.org/10.1007/978-3-031-64273-9_8

approach, which allows us to deal flexibly with the multiple, conflicting targets that must be managed when pursuing sustainability. We specifically take care of the size of the involved firms, by selecting twelve indicators ad hoc, which reflect in a parsimonious way many relevant dimensions towards which we expect SMEs should direct their efforts in terms of sustainability goals. In this analysis, rather than constructing a standard composite indicator derived from the aggregation of different dimensions, we evaluate probabilistically the attained rankings, based on rank acceptability and pairwise winning indices, following the stochastic multiacceptability analysis (SMAA) approach [6]. The rank acceptability index gives the probability that, randomly picking a set of parameters, a given firm attains a specific rank, whereas the pairwise winning indices specify the probability that a firm attains a better rank than another. In this way, the outcomes are evaluated in a robust way. The analysis is carried out for five different economic sectors, so as to uncover possible elements where progress on sustainability performance is still uncertain or unstable, with particular attention to best and worst performers within each sector. The rest of the contribution is organized as follows. In Sect. 2 we outline the methodology, then in Sect. 3 we present the data and the empirical application. Section 4 concludes.

2 A Multicriteria Ranking Method for Sustainability Assessment

In what follows, we consider a setting with m alternatives to be assessed w.r.t. n conflicting criteria. The problem is modelled through the MUlticriteria RAnking MEthod (MURAME) approach of [5], which consists in a combination of ELECTRE III and PROMETHEE II outranking methods. We refer to [1] for a detailed discussion of the model, in the context of sustainability assessment. To implement the SMAA, we follow the general framework provided in [6], which is here adapted to the peculiarities of the MURAME approach, in order to construct rank acceptability and pairwise winning indices, allowing to investigate how the ranking of the alternatives varies according to different sets of criteria parameters.

The preferences of a decision maker are integrated in the model by introducing, for any given criterion c_j, with $j = 1, \ldots, n$, an indifference threshold, q_j, a preference threshold p_j and a veto threshold v_j, for which it holds that $q_j \leq p_j \leq v_j$. The main idea of SMAA is to employ Monte Carlo simulation to visit the parameters' spaces, in order to provide the decision makers with values describing the problem, and to perform a parameter stability analysis. Let W and Q, P, V denote the spaces for the feasible weights and the thresholds, respectively, defined as follows:

$$W = \left\{ \mathbf{w} : w_j \geq 0, \sum_{j=1}^{m} w_i = 1 \right\} \qquad Q = \left\{ \mathbf{q} : 0 \leq q_j \leq p_j \right\}$$

$$P = \left\{ \mathbf{p} : q_j \leq p_j \leq v_j \right\} \qquad V = \left\{ \mathbf{v} : p_j \leq v_j \leq 1 \right\}$$

Moreover, let f_w, f_q, f_p and f_v be the probability distributions over these spaces, ranging in the feasible spaces W, Q, P and V. The choice of the distributions depends on whether the decision maker can express definite information about the shape of the distribution of such parameters or not. For simplicity, in our application we assume uniform distributions for each criterion weights and thresholds. To construct a ranking function, which maps the alternative i to a rank h, according to a set of simulated parameters, we perform a Monte Carlo simulation.

We aim to construct the ranking acceptability and the pairwise winning indices that handle possible robustness concerns w.r.t. the stability of the results, so as to give a probabilistic interpretation of the results. We proceed as follows. First, a function $K(i, w, q, p, v) = h$, providing the rank h of a firm i, for a given simulation of parameters, is considered. To derive the two indices, the ranking functions $m_h^i(i, w, q, p, v)$ and $p_h^i(i, k, w, q, p, v)$, taking either value 1 or 0, are constructed:

$$m_h^i(i, w, q, p, v) = \begin{cases} 1, & \text{if} \quad K(i, w, q, p, v) = h \\ 0, & \text{otherwise} \end{cases}$$

$$p_h^i(i, k, w, q, p, v) = \begin{cases} 1, & \text{if} \quad K(i, w, q, p, v) < K(k, w, q, p, v) \\ 0, & \text{otherwise} \end{cases}$$

Essentially, the ranking function $m_h^i(i, w, q, p, v)$ is equal to 1 if the firm i is assigned to a specific rank h; moreover, $p^i(i, k, w, q, p, v)$ is equal to 1 if firm i has a better (lower) rank compared to k. Finally, the rank acceptability index $C_h^i \in [0, 1]$ and the pairwise winning index $P^i \in [0, 1]$ provide a probabilistic classification of alternatives over possible rankings, i.e. respectively the probability that a rank h is assigned to alternative i and the probability that the alternative i is better than alternative k. Mathematically, both are multidimensional integrals over the parameter spaces, which can be estimated numerically via Monte Carlo simulation (see e.g. the formulation in [4]).

3 Application and Discussion of Results

For the empirical analysis, we consider a unique hand-collected dataset of 78 firms for the year 2021, based on a screening we perform to obtain a set of listed SMEs, adequately reporting on sustainability topics. The analysis is performed for firms belonging to Consumer Discretionary (CD), Health Care (HC), Financials (F), Industrials (I) and Information Technology (IT) classes, on a sector-by-sector basis[1]. Other sectors are not included due to lack of data. Furthermore, the selected criteria are aligned to Global Reporting Initiative Standards and include carbon intensity, waste generation intensity, non-renewable electricity

[1] 10 firms belong to CD, 12 to F, 13 to HC, 16 to IT and 27 to I. For a detailed analysis of the dataset, along with information for its replication, we refer the reader to [1].

consumption, water consumption intensity, average training hours, job creation, management diversity by gender, gender pay gap, board diversity by gender, economic value generation and distribution, board independence and the CEO pay ratio. Moreover, assuming that underreporting implies unwillingness or inability to report information, missing data are replaced with the worst value of the sector. Finally, the acceptability and the pairwise winning indices are obtained by performing $n = 100,000$ simulations over the feasible regions specified in Sect. 2.

Table 1. Category acceptability indices by sector.

	Best			
	ESG	E	S	G
CD	0.7224	0.7281	1	0.7283
F	0.7014	1	1	0.9089
HC	0.4739	0.6145	0.6157	0.9997
I	0.9998	1	0.7463	0.9264
IT	0.8708	1	0.5108	0.7853

(a) Probabilities that the best-classified firm of a sector for a combination of ESG or E, S, G criteria does not change its rank.

	Worst			
	ESG	E	S	G
CD	1	1	0.9999	1
F	0.9986	1	1	0.9959
HC	0.9998	1	1	0.9997
I	0.9992	0.9760	1	0.9614
IT	1	0.9312	0.7746	0.5898

(b) Probabilities that the worst-classified firm of a sector for a combination of ESG or E, S, G criteria does not change its rank.

The rank acceptability indices are reported in Tables 1a and 1b, with respect to the ESG performance, along with stand-alone Environmental, Social and Governance components. An important aspect that we observe in the data is that, even when taking into account uncertainty in the evaluation of SME's sustainability performance, the assessments of best and worst firms, i.e. the probabilities that the ranks of the best or the worst firms remain the same across different preferences of a decision maker, such as a policymaker or a credit officer, are robust across all sectors.

When the scores are gauged only w.r.t. specific E, S or G pillars, the outcomes remain stable. Such findings are important when considering the ESG performance of firms, in light of the uncertainty surrounding its measurement, as shown by the lack of agreement between ESG scores from different data providers, with important implications in terms of investor preferences and ultimately, asset prices [3]. Turning to a selection of acceptability and pairwise indices w.r.t. firms in the CD sector, results points to high stability of preferences of good alternatives over poor ones, as shown in Figs. 1 and 2. In the former, we show that little to no variability is observed for most firms, and variation typically occurs within three notches of the ranking. In the latter, the pairwise winning indices are presented. Note that substantial divergence between alternatives emerge in most cases, and in a particular case, a firm is actually preferred over all the alternatives in the dataset. Unreported results[2] for the remaining

[2] Available upon request.

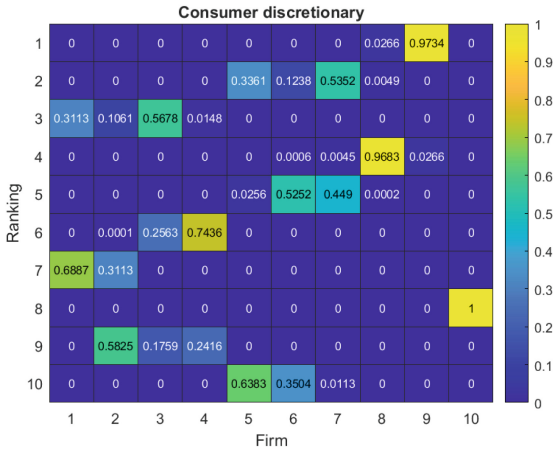


Fig. 1. Category acceptability indices for CD firms. A high percentage implies that the firm is assigned to the corresponding ranking with a high confidence, based on the share of possible parameter values for a given ranking. A low ranking with a high probability should be interpreted as a positive signal about a firm.

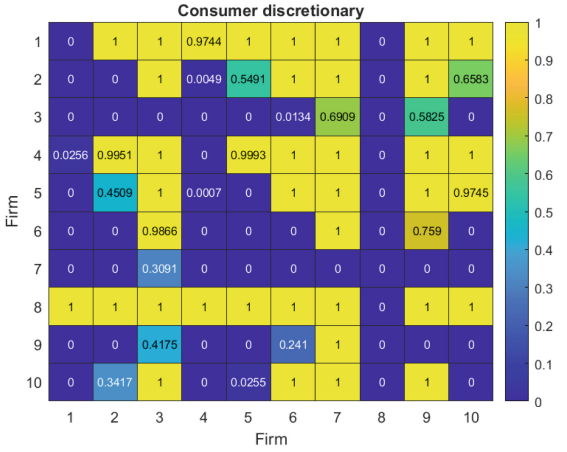


Fig. 2. Pairwise winning indices for CD firms. A high percentage implies that the firm on row i outperforms the firm on column j with a high confidence, based on the share of possible parameter values for a given simulation.

sectors, similarly show that the best firms tend to be picked over poorly performing ones with little uncertainty as well; less accurate assignments for firms with an average performance are observed. Such patterns might suggest that there are no sector-specific ESG reporting practices requiring deeper analysis.

4 Conclusions

In this paper, we have modelled the ESG scoring problem through a combination of an outranking approach and stochastic multiacceptability analysis, to capture the sustainability performance in a probabilistic way. Results confirm the robustness of the methodology, paving the way to more extensive analysis, covering the relationship between ESG performance and financial risk in SMEs.

Disclaimer

This paper was developed within the projects funded by: Next Generation EU - "GRINS - Growing Resilient, INclusive and Sustainable" project (PE0000018), National Recovery and Resilience Plan (NRRP) - PE9 - Mission 4, C2, Intervention 1.3; Next Generation EU - "Just Energy Transition - JET: Stochastic and machine learning methods for the evaluation, mitigation and geographical hedging of involved natural risks (with climate in view)", National Recovery and Resilience Plan (NRRP), Mission 4, C2, Intervention 1.1 - Notice: Progetti di Rilevante Interesse Nazionale (PRIN) 2022, DD No. 104 dated 2/2/2022, P2022XTLM2, CUP J53D23015530001. The views and opinions expressed are only those of the authors and do not necessarily reflect those of the European Union or the European Commission or the Italian Ministry of University and Research. Neither the European Union, the European Commission nor the Italian Ministry of University and Research can be held responsible for them.

References

1. Barro, D., Corazza, M., Filograsso, G.: A ESG rating model for European SMEs using multi-criteria decision aiding, Working Paper 27/2023, Department of Economics, Ca' Foscari University of Venice (2023)
2. Berg, F., Koelbel, J.F., Rigobon, R.: Aggregate confusion: the divergence of ESG ratings. Rev. Finan. **26**(6), 1315–1344 (2022)
3. Billio, M., Costola, M., Hristova, I., Latino, C., Pelizzon, L.: Inside the ESG ratings:(Dis) agreement and performance. Corp. Soc. Responsib. Environ. Manag. **28**(5), 1426–1445 (2021)
4. Gaganis, C., Papadimitri, P., Tasiou, M.: A multicriteria decision support tool for modelling bank credit ratings. Ann. Oper. Res. **306**, 27–56 (2021)
5. Goletsis, Y., Psarras, J., Samouilidis, J.E.: Project ranking in the Armenian energy sector using a multicriteria method for groups. Ann. Oper. Res. **120**, 135–157 (2003)
6. Tervonen, T., Lahdelma, R.: Implementing stochastic multicriteria acceptability analysis. Eur. J. Oper. Res. **178**(2), 500–513 (2007)
7. Trumpp, C., Guenther, T.: Too little or too much? Exploring U-shaped relationships between corporate environmental performance and corporate financial performance. Bus. Strateg. Environ. **26**(1), 49–68 (2017)
8. Tsang, Y.P., Fan, Y., Feng, Z.P.: Bridging the gap: building environmental, social and governance capabilities in small and medium logistics companies. J. Environ. Manage. **338**, 117758 (2023)

Hierarchical Clustering of Time Series with Wasserstein Distance

Alessia Benevento[1], Fabrizio Durante[1], Daniela Gallo[1,2(✉)], and Aurora Gatto[3]

[1] Università del Salento, Lecce, Italy
{alessia.benevento,fabrizio.durante,daniela.gallo}@unisalento.it
[2] Institute for High Performance Computing and Networking, Rende, Italy
daniela.gallo@icar.cnr.it
[3] Free University of Bozen-Bolzano, Bozen-Bolzano, Italy
aurora.gatto@unibz.it

Abstract. Two methods are presented in order to create a dissimilarity measure for random variables. These methods exploit some theoretical and computational advantages of the Wasserstein distance. The dissimilarity measures are hence applied to develop a hierarchical clustering procedure for time series, which are especially helpful in risk analysis.

Keywords: Clustering · Copula · Risk · Wasserstein distance

1 Introduction

In the vast field of machine learning, one often encounters the need to compare probability distributions and measure the dissimilarity between them. In this respect, traditional distance metrics like Kullback-Leibler divergence may fail when dealing with distributions that have different supports or significant overlap. Wasserstein distance, instead, has proved to be able to overcome such limitations [10]. In fact, unlike other distance measures that rely on pointwise differences, the Wasserstein distance measures the minimal effort required to reconfigure the probability mass of one distribution in order to recover the other distribution.

Here, we use the Wasserstein distance $d_{\mathcal{W}_p}$ [13,15] in order to derive two dissimilarity measures among random variables. Specifically, the first dissimilarity measure Δ^1 we propose identifies those variables that have the same distribution. In financial risk management, when X and Y represent the loss distributions, it may identify the losses with similar value-at-risk at different levels. When X and Y represent climate variables collected at different sites, instead, it may be helpful for regionalization purposes [14], i.e. to create regions that have similar climatic behaviour (i.e. similar temperature distribution).

The second dissimilarity measure Δ^2 is instead inspired by recent works on Wasserstein-based measures of association (see, e.g., [7,9]). It measures the distance between the copula associated with (X, Y) and the comonotonic copula M_2 that represents the maximal rank-invariant dependence between X and Y

M. Corazza et al. (Eds.): MAF 2024, *Mathematical and Statistical Methods for Actuarial Sciences and Finance*, pp. 49–54, 2024.
https://doi.org/10.1007/978-3-031-64273-9_9

(see, e.g., [2]). Thus, it serves to quantify whether X and Y are strongly associated regardless of their marginal behavior. In risk analysis, for instance, it may check whether X and Y tend to have jointly large (respectively, small) values, i.e. they are concordant.

The paper is organized as follows. In Sect. 2, we revisit the notion of Wasserstein distance. In Sect. 3, instead, we demonstrate the advantage of the proposed dissimilarity measure to develop agglomerative hierarchical clustering algorithms for time series. An illustration is presented in Sect. 4.

2 The Wasserstein Distance

In order to measure the dissimilarity between probability distributions, we will leverage on the Wasserstein distance between them and the related empirical measures [10,12]. To this end, we recall some basic facts that will be used in this paper (see, e.g., [13,15]).

Given two d–dimensional random vectors $\mathbf{X}$ and $\mathbf{Y}$ with probability laws $\mu_{\mathbf{X}}$ and $\mu_{\mathbf{Y}}$, respectively, their p–Wasserstein distance ($p \geq 1$) can be expressed as

$$d_{\mathcal{W}_p}(\mu_{\mathbf{X}}, \mu_{\mathbf{Y}}) = \inf_{\gamma \in \Gamma(\mu_{\mathbf{X}}, \mu_{\mathbf{Y}})} \left\{ (\mathbb{E}|\mathbf{X} - \mathbf{Y}|^p)^{1/p} : \mathbf{X} \sim \mu_{\mathbf{X}}, \mathbf{Y} \sim \mu_{\mathbf{Y}} \right\}, \quad (1)$$

where $\Gamma(\mu_{\mathbf{X}}, \mu_{\mathbf{Y}})$ denotes the collection of all joint probability laws whose marginals are $\mu_{\mathbf{X}}$ and $\mu_{\mathbf{Y}}$, respectively. It is possible to prove that the above minimization problem has a solution, which is unique if $\mu_{\mathbf{Y}}$ is absolutely continuous with respect to the Lebesgue measure. For one-dimensional random vectors, the distance reduces to the following formula

$$d^p_{\mathcal{W}_p}(\mu_X, \mu_Y) = \int_0^1 |F_X^{-1}(u) - F_Y^{-1}(u)|^p du, \quad (2)$$

where F_X^{-1} and F_Y^{-1} are the quantile functions associated with X and Y, respectively.

Although theoretically appealing, the computation related to the Wasserstein distances may be not available in closed form, up to the Gaussian case. However, the involved probability measures can be approximated by taking consistent empirical versions of the input measures [3,10]. In fact, for discrete measures, the computation of the Wasserstein distance is in principle obtained through the solution of a finite-dimensional linear program; see [12].

Specifically, consider the measures $\mu_{\mathbf{X}}$ and $\mu_{\mathbf{Y}}$ that are only available through a finite number of samples $\{\mathbf{x}_t\}_{t=1,\dots,T}$ and $\{\mathbf{y}_t\}_{t=1,\dots,T}$ in $\mathbb{R}^d$, and all points have the same mass, then

$$\mu_{\mathbf{X}} = \frac{1}{T} \sum_{t=1}^{T} \delta_{\mathbf{x}_t}, \qquad \mu_{\mathbf{Y}} = \frac{1}{T} \sum_{t=1}^{T} \delta_{\mathbf{y}_t},$$

where $\delta_{\mathbf{x}_t}$ is the Dirac distribution at the point $\mathbf{x}_t$. Consider the sample matrices $M_{\mathbf{X}} = (\mathbf{x}_1, \dots, \mathbf{x}_T)^\top \in \mathbb{R}^{T \times d}$ and $M_{\mathbf{Y}} = (\mathbf{y}_1, \dots, \mathbf{y}_T)^\top \in \mathbb{R}^{T \times d}$, respectively. In

this context, the set of all probabilistic couplings between these two measures is then the set of doubly stochastic matrices $\mathcal{P}$ defined as

$$\mathcal{P} = \left\{ \boldsymbol{\gamma} \in \mathbb{R}^{T \times T} | \boldsymbol{\gamma} \mathbf{1}_T = T^{-1} \mathbf{1}_T, \boldsymbol{\gamma}^\top \mathbf{1}_T = T^{-1} \mathbf{1}_T \right\},$$

where $\mathbf{1}_T$ is a T-dimensional vector of ones. Here, we use the fact that all points have the same probability mass. The $d_{\mathcal{W}_p}$ distance between the two measures is:

$$d^p_{\mathcal{W}_p}(\mu_{\mathbf{X}}, \mu_{\mathbf{Y}}) = \min_{\boldsymbol{\gamma} \in \mathcal{P}} \langle \boldsymbol{\gamma}, \mathbf{L} \rangle_F,$$

where $\langle \cdot, \cdot \rangle_F$ is the Frobenius dot product and $\mathbf{L} \geq 0$ is the cost function matrix whose entry is given by $L_{tt'} = \|\mathbf{x}_t - \mathbf{y}_{t'}\|^p$ (see, e.g., [12]).

3 The Methodology

In the framework of copula-based hierarchical clustering [4], we propose two novel dissimilarity measures that can be used for pairwise dissimilarities among all the components of a given random vector.

Given two continuous random variables X and Y with marginals F_X and F_Y, and copula $C_{X,Y}$, the first dissimilarity matrix is given by

$$\Delta^1(X, Y) = d_{\mathcal{W}_p}(X, Y) \tag{3}$$

according to (2). The second dissimilarity measure is instead given by

$$\Delta^2(X, Y) = d_{\mathcal{W}_p}(C_{X,Y}, M_2), \tag{4}$$

where M_2 is the comonotonicity copula that describes the maximal concordance between random variable. Such a dissimilarity has been also suggested in [1].

Notice that, if $Y = \alpha X$ for some $\alpha \neq 0$, then $\Delta^1(X, Y) \neq 0$, since the random variable needs not have the same distribution, while $\Delta^2(X, Y) = 0$, since X and Y are completely dependent. Moreover, for two independent and identically distributed X and Y, it follows that $\Delta^1(X, Y) = 0$, but $\Delta^2(X, Y) \neq 0$. Thus, the two dissimilarity measures describe two different stochastic aspects.

Given a random sample $(x_t, y_t)_{t=1,\ldots,T}$ from (X, Y), the estimation of (3) can be done by relying on the associated discrete measures as explained in Sect. 2. For the estimation of (4), instead, we compute the associated pseudo-observations associated with the random sample [5] and, from them, calculate the associated discrete (copula) measure. As is known, the pseudo-observations are taking values on the grid $\{1/(T+1), \ldots, T/(T+1)\}^2$ where T is the sample size. Here, we prefer to divide by $(T+1)$ in order to be within the boundary of $[0, 1]^2$. Thus, the discrete version of the copula $C_{X,Y}$ can be considered as a matrix in $\mathbb{R}^{T \times 2}$. Moreover, the discrete version of the copula M_2 is obtained from the sample

$$\left\{ \mathbf{m}_t = \left(\frac{t}{T+1}, \frac{t}{T+1} \right) : t = 1, \ldots, T \right\},$$

which describes the comonotonic behaviour between two variables in the copula setting.

4 Illustration: Analysis of the Components of the FTSE MIB

In this section, we present a case study to illustrate the methodology introduced in Sect. 3. Specifically, we are interested in providing a hierarchical clustering of the forty components of FTSE-MIB according to the two introduced dissimilarity measures. As known, such a procedure may be relevant for portfolio diversification (see, e.g., [8] and references therein).

To this end, we consider log-returns of the daily (adjusted) prices of the forty components listed on the FTSE MIB from January 2nd, 2023 to December 29, 2023. 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 253 daily observations for each variable and no missing values. As traditional in risk analysis that focuses on the right-side of the distribution, the relative log-returns are calculated with a negative sign.

First, we consider an ARMA-GARCH copula model (see, e.g., [11]) and estimate the marginals and the copula according to a two-stage procedure as illustrated in [5, section 6.2.3]. Specifically, we estimate an appropriate ARMA(1,1)-GARCH(1,1) model with Student distribution for each time series. It allows us to model the conditional mean and variance taking into account possible time-varying volatility patterns. Moreover, the obtained residuals generally 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 on the pseudo-observations associated with the residuals.

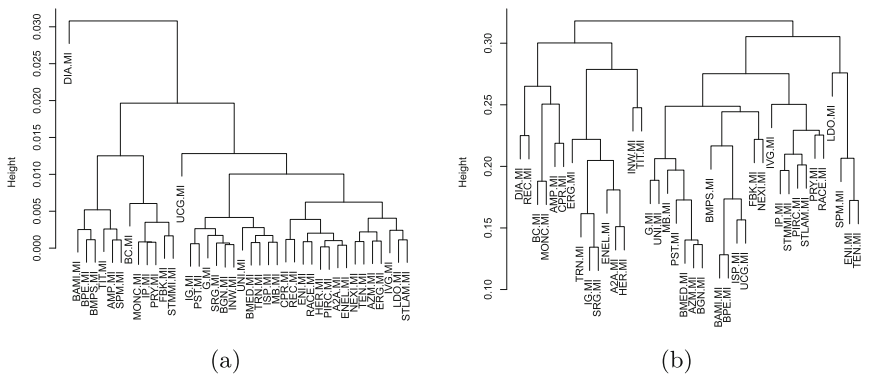


Fig. 1. Dendrograms obtained from agglomerative hierarchical clustering (linkage method: complete) applied to the time series of log-returns for the forty components listed on the FTSE MIB and on: (a) the forecasted one-day ahead distribution and the pairwise dissimilarity Δ^1; (b) the pairwise dissimilarity Δ^2. Period: 02/02/2023–29/12/2023.

In Fig. 1 (panel a), we provide the agglomerative hierarchical clustering (complete linkage) of the time series based on the forecasted one-day ahead distribution of log-returns for each time series and on the pairwise dissimilarity Δ^1. These scenarios may serve to provide a suitable picture of the individual risks. Instead in Fig. 1 (panel b), the agglomerative hierarchical clustering (complete linkage) of the time series based on the pairwise dissimilarity Δ^2 is shown. These two scenarios may serve to provide a suitable picture of concordance among different risks.

In Fig. 2 we adopt, instead, the procedure described in [6] in order to merge both clusters. In such a case, time series belonging to the same group have similar risk profile as well as they tend to comove. These findings should be considered with particular care when a diversified portfolio should be built.

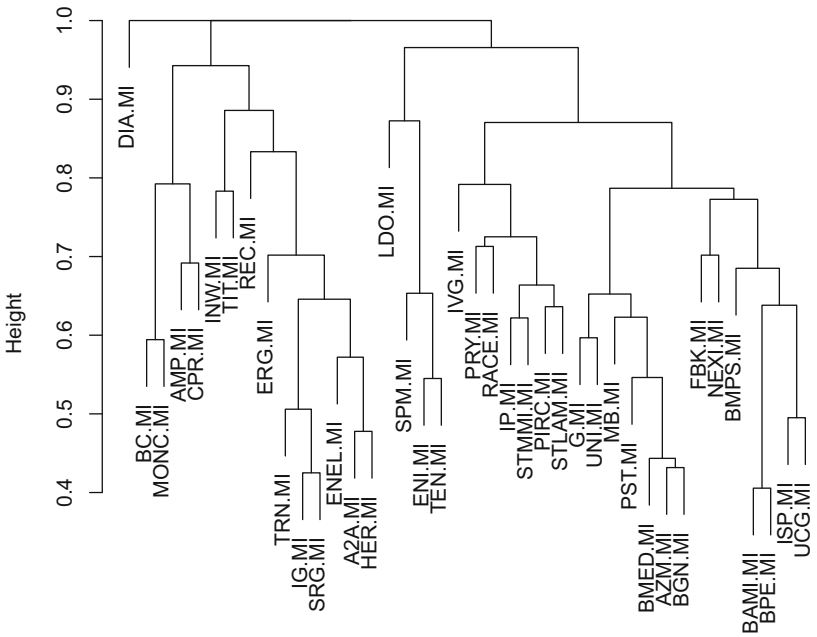


Fig. 2. Dendrogram obtained from merging via the algorithm in [6] the two dendrograms presented in Fig. 1.

Acknowledgments. AB acknowledges the support of Regione Puglia (Italy) via the Programma Regionale "RIPARTI (assegni di RIcerca per riPARTire con le Imprese)" - research project "FIRST: a Framework for Innovation in Risk management to support Territories" (code: c19a5daa).

FD has been supported by MUR-PRIN 2022 PNRR, Project "Stochastic Modeling of Compound Events" (No. P2022KZJTZ) funded by European Union - Next Generation EU.

DG has been supported by the PhD grant "Mathematical and computational models for the representation and dissemination of information" funded by CNR-ICAR.

AG has carried out this study within the Agritech National Research Center and received funding from the European Union - Next Generation EU (Piano Nazionale di Ripresa e Resilienza (PNRR) - Missione 4 Componente 2, Investimento 1.4 - D.D. 1032 17/06/2022, CN00000022).

References

1. Benevento, A., Durante, F.: Wasserstein dissimilarity for copula-based clustering of time series with spatial information. Mathematics **12**(1), 67 (2024)
2. Durante, F., Sempi, C.: Principles of Copula Theory. CRC Press, Boca Raton, FL (2016)
3. Fournier, N., Guillin, A.: On the rate of convergence in Wasserstein distance of the empirical measure. Probab. Theory Relat. Fields **162**(3–4), 707–738 (2015)
4. Fuchs, S., Di Lascio, F.M.L., Durante, F.: Dissimilarity functions for rank-invariant hierarchical clustering of continuous variables. Comput. Statist. Data Anal. **159**, 107201 (2021)
5. Hofert, M., Kojadinovic, I., Mächler, M., Yan, J.: Elements of copula modeling with R. Springer, Cham (2018)
6. Hulot, A., Chiquet, J., Jaffrézic, F., Rigaill, G.: Fast tree aggregation for consensus hierarchical clustering. BMC Bioinform. **21**(1), 1–12 (2020)
7. Marti, G., Andler, S., Nielsen, F., Donnat, P.: Optimal transport vs. Fisher-Rao distance between copulas for clustering multivariate time series. In: 2016 IEEE statistical signal processing workshop (SSP), pp. 1–5. IEEE (2016)
8. Marti, G., Nielsen, F., Bińkowski, M., Donnat, P.: A review of two decades of correlations, hierarchies, networks and clustering in financial markets. In: Nielsen, F. (ed.) Progress in Information Geometry: Theory and Applications, pp. 245–274. Springer International Publishing, Cham (2021)
9. Mordant, G., Segers, J.: Measuring dependence between random vectors via optimal transport. J. Multivariate Anal. **189**, 22 (2022). Id/No 104912
10. Panaretos, V.M., Zemel, Y.: Statistical aspects of Wasserstein distances. Annu. Rev. Stat. Appl. **6**, 405–431 (2019)
11. Patton, A.: A review of copula models for economic time series. J. Multivariate Anal. **110**, 4–18 (2012)
12. Peyré, G., Marti, M.: Computational optimal transport: With applications to data science. Found. Trends® Mach. Learn. **11**(5-6), 355–607 (2019)
13. Santambrogio, F.: Optimal transport for applied mathematicians. Birkäuser, NY **55**(58–63), 94 (2015)
14. Saunders, K.R., Stephenson, A.G., Karoly, D.J.: A regionalisation approach for rainfall based on extremal dependence. Extremes **24**(2), 1386–1999 (2021)
15. Villani, C.: Optimal transport. Old and new, *Grundlehren Math. Wiss.*, vol. 338. Springer, Berlin (2009)

Wind Farm Evaluation Under Real Options Approach

Marta Biancardi, Michele Bufalo, Antonio Di Bari, and Giovanni Villani(✉)

Department of Economics and Finance, University of Bari,
Largo Abbazia S. Scolastica, 70124 Bari, Italy
{marta.biancardi,michele.bufalo,
antonio.dibari,giovanni.villani}@uniba.it

Abstract. The wind farm performance is often characterized by uncertainty since it depends on unstable condition of wind speed and, consequently, on unstable wind energy conversion. This aspect makes the wind projects valuation quite difficult. The Real Options Approach (ROA) represents an adequate methodology to assess wind energy projects. This work applies the ROA by considering a specific stochastic process that would fit for the wind speed modelling, and other typical characteristics of wind projects, such as their multistage nature. We model the wind turbine performance by including three possible scenarios: cut-in speed, rated output speed and cut-out speed. A numerical example is provided to implement our mathematical valuation approach.

Keywords: Wind energy modelling · Real Options Approach · Wind speed uncertainty · Multistage projects

1 Introduction

The electricity production through the wind projects has gained its importance by considering the increasing environmental issues that affect our planet. Increasing generation capacity in liberalized power markets requires the producers to account for future long-term uncertainties. In particular, this is true for wind producers since the inherent intermittency of its energy production depends on wind regimes. Due to the considerable lifetime of these plants and the mentioned uncertainties, traditional discounted cash flow methods used to obtain the NPV of the investment are unsuitable for these projects. In fact, the wind farm performance is often characterized by uncertainty since it depends on unstable condition of wind speed and, consequently, on unstable wind energy conversion ([5]). To take into account the uncertainty, the Real Options Approach (ROA) can be used to develop flexible investment strategies. Depending on the present value in a certain time, the decision maker can decide to execute, wait, or abandon the construction project of a power plant. Previous studies applied the ROA to assess wind energy projects ([1–4]). This work expand the existing literature by providing a ROA that considers a specific stochastic process based on

M. Corazza et al. (Eds.): MAF 2024, *Mathematical and Statistical Methods for Actuarial Sciences and Finance*, pp. 55–60, 2024.
https://doi.org/10.1007/978-3-031-64273-9_10

an adjusted version of the classical mean reversion in order to consider three thresholds applied to wind energy production. The lowest threshold represents the stop of the turbines given the too low wind speed (cut-in speed). Above the cut-in speed threshold, the electric power output rapidly increases up to reach the limit of the electric generator (rated output power). The highest threshold represents the stop of the turbines due to the risk of damage caused by the excessive wind speed. Moreover, the ROA with this specific stochastic process is embedded in a multistage valuation framework that characterized the wind farm investments. A numerical example is also proposed.

2 Stochastic Modelling

This work develops a mathematical model based on the Real Options Approach (ROA) to price the wind projects by modelling the uncertain wind energy production. The ROA is applied by considering a specific stochastic process that would fit for wind speed modelling, and other typical characteristics of wind projects, such as their multistage nature. As shown in [5,6], we model the wind speed as:

$$W_t = s^w(t) + w_t \tag{1}$$

where s^w is a seasonality composition (measured hourly), i.e.:

$$s^w(t) = \beta^{w1} \sin\left(\frac{2\pi t}{8760}\right) + \beta^{w2} \cos\left(\frac{2\pi t}{8760}\right) \tag{2}$$

and w is modelled as a Ornstein-Uhlenbeck process:

$$w_t = \lambda^w(\bar{w} - w_t)dt + \sigma^w dZ_t^w \tag{3}$$

where $\lambda^w > 0$ is the speed of adjustment, $\bar{w}$ is the long run mean of wind speed, σ^w is the wind speed volatility, Z_t^w is the standard Brownian motion and β^{w1}, β^{w2} are two positive parameters of seasonality.

2.1 Power Curve

The power curve of a wind turbine represents the relationship that links the electrical power produced by a wind turbine (WT) with wind speed. Although wind turbine manufacturers provide guaranteed power curves, the performance of a wind turbine can vary depending on weather conditions, as well as the location in which it operates. As a result, a power curve must be created to represent the normal operation of each WT installed at a given site. Wind turbine performance is then monitored and compared to performance under normal operation. We lists the various scenarios that can influence the turbine performance.

– **Cut-in speed (s_L):** When the wind speed is very low, the wind force on the turbine blades is insufficient to rotate them, so the active power produced is zero. As the wind speed increases, the wind turbine begins to rotate and

generate power. The speed at which the turbine begins to rotate and generate power is known as cut-in speed, and it is typically between 3 and 4 m per second.

- **Rated output speed (s_M):** When the wind speed exceeds the cut-in speed, the amount of electric power output rapidly increases. However, when a certain speed is reached, the power output gets to the limit of the electric generator. Such limit is known as the rated output power and the wind speed at which this limit is reached is called the rated output wind speed. The turbine design is intended to limit power to this maximum level at higher wind speed.
- **Cut-out speed (s_H):** When the speed exceeds the rated output wind speed, the forces on the turbine structure increase and the rotor is at risk of being damaged. Consequently, once a certain speed is exceeded, braking is used to stop the rotor. This limit is called the cut-out speed, and when it is exceeded the energy produced returns to zero.

The wind speed to produce energy $\widetilde{W}_t$ is such that: if it is below a threshold $W_t < s_L$, it is equal to zero $\widetilde{W}_t = 0$ as the wind speed is modest to move the blades; if it exceeds the threshold $W_t > s_H$, it is equal to zero $\widetilde{W}_t = 0$ as the gusts of wind are so intense that it requires stopping the blades for safety; if it exceeds the threshold $W_t > s_M$, the wind speed does not improve the rotation of the blades which is already maximum, and so $\widetilde{W}_t = s_M$. We summarize the effective weed speed as:

$$\widetilde{W}_t = s_M \mathbf{1}_{(s_M \leq W_t < s_H)} + W_t \mathbf{1}_{(s_L \leq W_t < s_M)} \tag{4}$$

Once the forecasting of wind speed is performed and validated, the wind turbine power output is then computed according to the following equation, given by A. Betz:

$$P_t = \frac{1}{2} C_p \rho S \widetilde{W}_t^{\,3} \tag{5}$$

Betz explains the impossibility to entirely convert the kinetic energy of a mass of air into mechanical energy. He also discovered that there is an upper limit to the amount of kinetic energy that can be converted. Due to this reason, Betz introduced a parameter, called the power coefficient C_p, which can be calculated as a function of the speed ratio of the wind wake behind the rotor to the speed ahead, and its value is $C_p = 0.593$. Moreover, ρ is the air density, S is the section through which the air mass flows that depends on the diameter D of the wind turbine, and so $S = \frac{\pi D^2}{4}$. Figure 1 describes the relation between the wind speed and the power output of a turbine, where P_R is the rated power of the wind turbine.

3 Real Options Application on Wind Farm Projects

In order to value a wind farm, we hypothesize that electricity price P_e is constant and is determined by Feed-in tariff (FiT) mechanism. The revenue that a wind farm can obtain during its lifetime for each wind turbine are:

$$R_t = P_t \cdot P_e \tag{6}$$

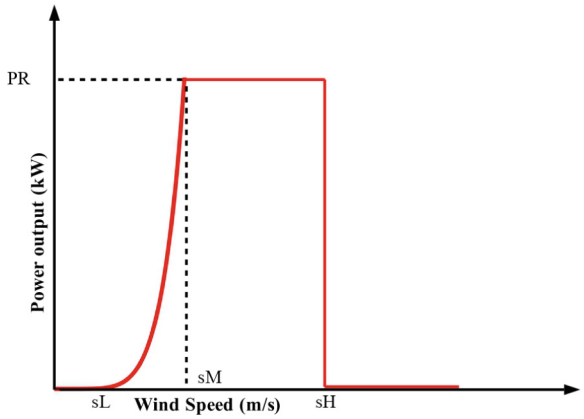


Fig. 1. Wind turbine power output

and therefore, the value of the project V can be obtained by discounting all the revenues generated by selling electricity and subtracting the discounted maintenance costs C (which may have a different timing from the revenues):

$$V = N\left(\sum_{j=0}^{n} R_j e^{-\delta h j} - \sum_{j=0}^{m} C_j e^{-\delta j z}\right) \tag{7}$$

where N is the number of turbine installed in the wind farm, $h = \frac{T_F}{n}$, $z = \frac{T_F}{m}$, n is the number of discretization of revenues, m is the number of discretization of costs, T_F is the technological obsolescence of the wind farm (after T_F there are not revenues). For our applications, we assume to discretize the revenues every hour based on wind speed simulations, while the costs with an annual interval. Finally, we denote as K_0, K_1 and K_2 the investment costs in order to realize and to implement the wind farm (for example costs for concessions K_0, costs of expropriation of areas K_1 and costs of building wind turbines K_2). The project will be realized at maturity T if:

$$s_T = \max[0; V_T - K_2] \tag{8}$$

If we denote as s_{t_1} the value of option at time t_1, with $t_1 < T$, the investment K_1 will be realized if:

$$c_{t_1} = \max[0; s_{t_1} - K_1] \tag{9}$$

and the value of this investment opportunity at initial time $t_0 = 0$ is a compound option c_0. The value of c_0 may be calculated through a Monte Carlo approach, as

$$c_0 = e^{-rt_1}\mathbb{E}[\max(c_{t_1} - K_0, 0)] \tag{10}$$

being r the risk-free interest rate and K_0 the initial investment cost. In particular, since c_t depends on s_t and so, on the stochastic process V_t $(t \in (0,T])$, we need to simulate it. This can be realized thank to a numerical Eulero scheme implemented on a suitable function ϕ of w_t[1], i.e.,

$$\begin{cases} \hat{V}_{t+\Delta} = \phi(\hat{w}_{t+\Delta}) \\ \hat{w}_{t+\Delta} = \hat{w}_t + \lambda^w(\bar{w} - \hat{w}_t)\Delta + \sigma^w\sqrt{\Delta}\,\varepsilon_t \end{cases} \tag{11}$$

where Δ is the time-step and ε_t is a white noise.
Consequently, the real option value (ROV) will be: $ROV = -K_0 + c_0$

4 Numerical Example

For our numerical analysis, we assume to evaluate a wind farm that consists of $N = 10$ wind turbines. We list first of all the technical-natural values and after that the parameters of real option evaluation. Regarding the technical-natural parameters, we assume that $\lambda^w = 3$, $\beta^{w1} = \beta^{w2} = 0.7$; $\bar{w} = 5\,m/s$; $\sigma^w = 5$; $D = 10\,m$; $S = 78.53\,m^2$; $\rho = 1.225\,kg/m^3$; $C_p = 0.593$; $P_R = 30\,kW$; $s_L = 4\,m/s$; $s_M = 9.8\,m/s$; $s_H = 21\,m/s$. Regarding the economic parameters, we assume that $T_F = 5$ years; $t_1 = 1$ year; $T = 3$ years; $P_e = 0.121$ Euro/KWh; $\delta = 0.08$; $C = 1\,000$ Euro (for each turbine); $r = 0.05$; $\sigma_v = 0.60$; $K_0 = 5\,000$ Euro; $K_1 = 40\,000$ Euro; $K_2 = 150\,000$. Based on these assumption, we obtain that each wind turbine produces 173.80 MW (see Figs. 2(a)–2(b)) and multiplied by $N = 10$ we obtain that the entire wind farm produces 1738 MW. So, the value of asset becomes $V = 138\,448$ Euro. Applying formulas (10)–(11), we obtain that $c_0 = 31\,576$ and the real option value $ROV = 26\,576$. Since the ROV is positive, the management should prefer to make the investment of wind farm. The positive result of the ROV means that the investment should be

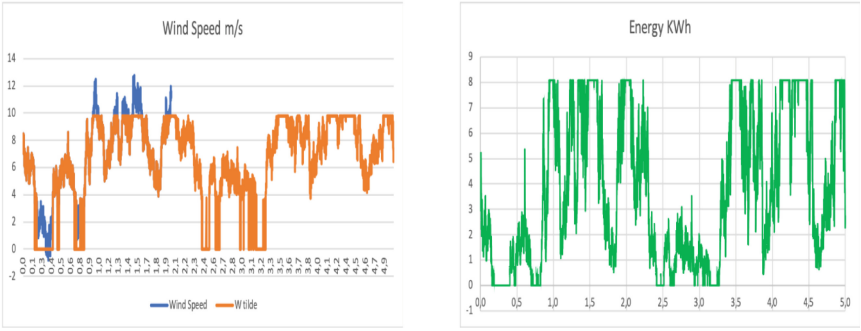


Fig. 2. Wind speed simulation and Energy output obtained from each wind turbine.

[1] Such ϕ is easily obtained by combining Eqs. (1) (4), (5), (6) and (7). We prefer do not write ϕ explicitly for the sake of notation and readability.

pursued since the valuation that includes the uncertainty of project performance is advantageous for investor. If the result had been negative, the investor would have had to reject the project because it could have implied a large possibility of financial loss.

5 Conclusions

In this work, we have analyzed how the uncertainty of wind speed and its prediction can influence the value of a wind project. Starting from the technical characteristics of the turbine, we can determine the level of energy that can be obtained from a wind farm. Subsequently, we have assessed the value of this project through compound real options, considering the different costs that must be incurred sequentially before obtaining the cash flows. Through a numerical case, we estimated the value of this wind farm which turns out to be positive in the context of real options.

Acknowledgments. The Authors acknowledge the financial support from the program MUR PRIN 2022 n. 2022ETEHRM "Stochastic models and techniques for the management of wind farms and power systems".

References

1. Bufalo, M., Di Bari, A., Villani, G.: Multi-stage real option evaluation with double barrier under stochastic volatility and interest rate. Ann. Finance **18**, 247–266 (2022). https://doi.org/10.1007/s10436-021-00403-6
2. Diaz, G., Moreno, B., Coto, José. Gomez-Aleixandre, J.: Valuation of wind power distributed generation by using Longstaff-Schwartz option pricing method. Appl. Energy **145**, 223–233 (2015). https://doi.org/10.1016/j.apenergy.2015.02.046
3. Lee, S.C.: Using real option analysis for highly uncertain technology investments: the case of wind energy technology. Renew. Sustain. Energy Rev. **15**, 4443–4450 (2011). https://doi.org/10.1016/j.rser.2011.07.107
4. Loncar, D., Milovanovic, I., Rakic, B., Radjenovic, T.: Compound real options valuation of renewable energy projects: the case of a wind farm in Serbia. Renew. Sustain. Energy Rev. **75**, 354–367 (2017). https://doi.org/10.1016/j.rser.2016.11.001
5. Muñoz, J.I., Contreras, J., Caamaño, J., Correira, P.F.: A decision-making tool for project investments based on real options: the case of wind power generation. Ann. Oper. Res. **186**, 465–490 (2011). https://doi.org/10.1007/s10479-011-0856-9
6. Nibbering, D., van Buuren, C., Wei W.: Real options valuation of wind energy based on the empirical production uncertainty. Monash Econometrics and Business Statistics Working Papers 19/21, Monash University, Department of Econometrics and Business Statistics (2021)

Fair Volatility in the Fractional Stochastic Regularity Model

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

[1] Sapienza University, 00161 Rome, Italy
sergio.bianchi@uniroma1.it
[2] University of Cassino and Southern Lazio, 03043 Cassino, Italy

Abstract. Within the efficient markets framework, discounted stock prices are typically represented through Brownian martingales. The primary measure for evaluating risk is the volatility of log-returns, under the assumption that higher variability indicates greater associated risk. The theoretical foundation of this claim stems from the characterization of the path regularity of price process through the Lévy characterization theorem of Brownian motion. Since this explanation lacks a financial interpretation when considering more realistic models, such as stochastic volatility models, it is necessary to disentangle volatility and regularity. Replacing volatility by the Hölder regularity provides insights into market deviations from the equilibrium of the martingale model, and - within the Fractional Stochastic Regularity Model - contributes to identify the "fair" volatility aimed by the market.

Keywords: Hölder exponent · Volatility · Fractional Stochastic Regularity Model

1 Introduction

This contribution underscores the crucial distinction between *volatility* and *regularity* for the purpose of characterizing risk in financial dynamics. Volatility quantifies the extent to which data deviate from their mean value, while regularity captures the manner in which data are dispersed. In the framework of paradigmatic Efficient Market Hypothesis (EMH) and the consequent (Brownian) martingale model, the determination of "*how*" data are dispersed is uniquely dictated by the quadratic variation of the process. Nevertheless, challenges emerge when questioning or relaxing this model, prompting the need for a distinct consideration of volatility and regularity and advising against a blanket association of volatility with risk. The introduction of memory, triggered by positive or negative autocorrelation, influences the level of regularity in the sequence and introduces a potential error in the assessment of financial risk based solely on volatility, whose value may not be influenced by autocorrelation.

M. Corazza et al. (Eds.): MAF 2024, *Mathematical and Statistical Methods for Actuarial Sciences and Finance*, pp. 61–66, 2024.
https://doi.org/10.1007/978-3-031-64273-9_11

The need to neatly distinguish between *volatility* and *regularity* also arises from the tendency of literature to support the EMH based on the commonly observed null value of the empirical autocorrelation function of log-returns, a well-known stylized fact. However, a null empirical autocorrelation can also result from a time-varying pointwise regularity of log-returns, a change which is not necessarily detected by volatility. In addition to aiding in the formulation of more realistic models of financial dynamics, the pointwise regularity - owing to its direct connection with the martingale model benchmark - offers insights into market mechanisms that are not captured by volatility alone.

2 Background and Model

Some notions are recalled here which will be combined to show how regularity enhances the informational content of volatility.

Theorem 1. [*Lévy's Characterization Theorem*] *Given the filtered probability space* $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t\geq 0}, \mathbb{P})$*, let* (X_t) *be a local martingale with* $X_0 = 0$*. Then, the following are equivalent:*

(a) $\{X_t\}$ *is standard Brownian motion on the underlying filtered probability space*
(b) $\{X_t\}$ *is continuous and* $\{X_t^2 - t\}$ *is a local martingale*
(c) $\{X_t\}$ *has quadratic variation* $\langle X \rangle_{2,t} = t$*.*

The equivalence between the condition of local martingale of an $\mathcal{F}_t$-Brownian motion and its quadratic variation is fundamental to justify why, in the context of efficient markets, volatility has traditionally served as a risk indicator: according to the Efficient Market Hypothesis (EMH) [6], if the discounted price process behaves as a martingale, then the only governing factor influencing its randomness is determined by the growth of its quadratic variation, proportional to the time interval t. The absence of alternative possibilities makes the adoption of volatility as a risk indicator a natural choice within this framework.

Pointwise Hölder Exponent. Given the continuous real-valued stochastic process $\{Z_t, t \geq 0\}$, its path roughness at any fixed $\tau > 0$ is usually measured through the *pointwise Hölder exponent* at τ. This is defined as [2,3]:

$$\alpha_Z(\tau) := \sup\{\alpha \in [0,1] : \limsup_{r\to 0^+} r^{-\alpha} \mathrm{Osc}_Z(\tau, r) < +\infty\} \tag{1}$$

where, for all real number $r > 0$ small enough,

$$\mathrm{Osc}_Z(\tau, r) := \sup\{|Z_{t'} - Z_{t''}| : (t', t'') \in [\tau - r, \tau + r]^2\}$$

is the oscillation of $\{Z_t\}$ on the circular neighbourhood of τ with radius r. For specific classes of stochastic processes, such as Gaussian processes, the zero-one law implies the existence of a non-random quantity $a_Z(t)$ for which $\mathbb{P}(a_Z(t) = \alpha_Z(t)) = 1$ [2]. When $\{Z_t\}$ is a semimartingale (e.g. Brownian motion), $\alpha_Z = \frac{1}{2}$. Deviations from $\frac{1}{2}$ characterize non-Markovian processes

(whose quintessential example is the well-known fractional Brownian motion, fBm); processes with $\alpha_Z \in \left(\frac{1}{2}, 1\right)$ exhibit excessively high smoothness, while those with $\alpha_Z \in \left(0, \frac{1}{2}\right)$ display insufficient smoothness to satisfy the martingale property. Specifically, the quadratic variation of the process can be proven to be zero if $\alpha_Z > \frac{1}{2}$, and infinite if $\alpha_Z < \frac{1}{2}$.

Multifractional Processes with Random Exponent. When the Hölder exponent is allowed to change through time in a deterministic or stochastic way, a class of stochastic process, named Multifractional Processes with Random Exponent (MPRE), can be defined, subject to some technical constraints [3,8]. A special case of the general MPRE process is

$$K_t^{H,C} = C \int_{-\infty}^{t} \left[(t-s)_+^{H_s - 1/2} - (-s)_+^{H_s - 1/2} \right] dB_s, \tag{2}$$

where C is a scale parameter, $(x)_+ = \max(x, 0)$ and B is the Brownian motion. By introducing a dependence on H_s in the integrand instead of H_t, the integral in Eq. (2) can be formulated in the conventional Itô sense. [8] establish a rescaling limit showing that, for each fixed t, as $h \to 0$,

$$h^{-H_t} \left(K_{t+hr}^{H,C} - K_t^{H,C} \right) \implies C \int_{-\infty}^{r} \left[(r-s)_+^{H_t - 1/2} - (-s)_+^{H_t - 1/2} \right] d\tilde{B}_s \tag{3}$$

where $\tilde{B}_s$ is a Brownian motion independent of H_t. Equation (3), known as Local Asymptotical Self-Similarity property (LASS), states that in the neighborhood of any point t, K_t^H behaves like a fBm with Hurst-Hölder exponent H_t.

(Rough) Fractional Stochastic Volatility Model (RFSV). Introduced by [7] and based on the previous model defined by [5], the Rough Fractional Stochastic Volatility (RFSV) model of the price process S_t reads as:

$$\begin{cases} dS_t = \mu_t S_t dt + S_t \sigma_t dB_t \\ \sigma_t = \exp(X_t) \end{cases} \tag{4}$$

where μ_t is the drift term, B_t is a Brownian motion and X_t is a fractional Ornstein-Uhlenbeck (fOU) process satisfying

$$dX_t = \alpha(m - X_t)dt + \rho dB_t^H, \tag{5}$$

with $m \in \mathbb{R}$, ρ and α positive parameters and with B_t and B_t^H correlated in general. [1] replace the stochastic process in the first line of Eq. (4) by an MPRE driven by a Hölder exponent which follows a proper fOU process related to dX_t by a change of parameters. They show that the stochastic Hölder parameter of the MPRE can replace the log-volatility in the second equation of model (4). This sets up the *Fractional Stochastic Regularity Model* (FSR), defined as

$$\begin{cases} S_t = K_t^{H,C} \\ H_t = m' + \rho' \int_{-\infty}^{t} e^{-\alpha(t-s)} dB_s^H \end{cases} \tag{6}$$

where S_t denotes the log-price of a stock or an index, H_t is the unique pathwise solution of the fOU process of line 2 of (6). Denoting by n the length of the sampled version of MPRE, $m' = -\frac{1}{\log n} \cdot m + \frac{\log C}{\log n}$ and $\rho' = -\frac{1}{\log n} \cdot \rho$. Thus, Eqs. (4) and (6) state that a relation exists between volatility σ_t and regularity H_t when the log-price is modelled by an MPRE and the log-volatility is modeled by a fOU process. In this case, also the Hölder exponent follows a fOU process with parameters which are linear transforms of those used to model the log-volatility. This directly follows from Eq. (3), which entails (see [1]) (Table 1).

$$\log \sigma_{t,n} = \log C - H_t \log n. \tag{7}$$

Figure 1 exhibits the goodness of fit of relation (7) for six global financial indexes: Dow Jones Industrial Average (DJI, USA), Nasdaq Composite (IXIC, USA), Eurostoxx50 (SX5E, Europe), Footsie 100 (UKX, United Kingdom), Hang Seng (HSI, Hong Kong) and Straits Times (STI, Singapore). H_t was estimated as in [1,9].

Table 1. Data set and main statistics of the estimated H_t

	DJI	IXIC	SX5E	UKX	HSI	STI
Start date	1992-01-02	1971-02-05	2000-01-03	1984-01-03	1986-12-31	1987-12-28
End date	2021-12-28	2021-12-28	2021-12-31	2021-12-29	2021-12-29	2022-01-28
# Obs (n)	7,555	12,470	5,730	9,599	5,955	8,409
Mean	0.540	0.512	0.524	0.524	0.512	0.533
St. Dev.	0.0540	0.0549	0.0553	0.0464	0.0498	0.0536
Range	0.308–0.688	0.314–0.670	0.345–0.675	0.329–0.652	0.269–0.622	0.333–0.665

3 Meaning and Financial Interpretation of the Relationship Between Volatility and Regularity

As discussed in the previous section, the Hölder regularity offers insight into the extent to which the process diverges from the martingale property with the baseline value $H_t = 1/2$. Substituting the Hölder exponent for the volatility process is justified by this characterization and brings several advantages:

– Unlike volatility, the Hölder exponent is sensitive to autocorrelation irrespective of the scale parameter. Volatility fails to differentiate between data with high or low correlation when appropriate scale parameters are applied, leading to instances where data with identical volatility levels may exhibit varying degrees of correlation. This incongruity is problematic if volatility is intended to assess financial risk. Conversely, the Hölder parameters of data with differing autocorrelations, are different irrespective of the scale.

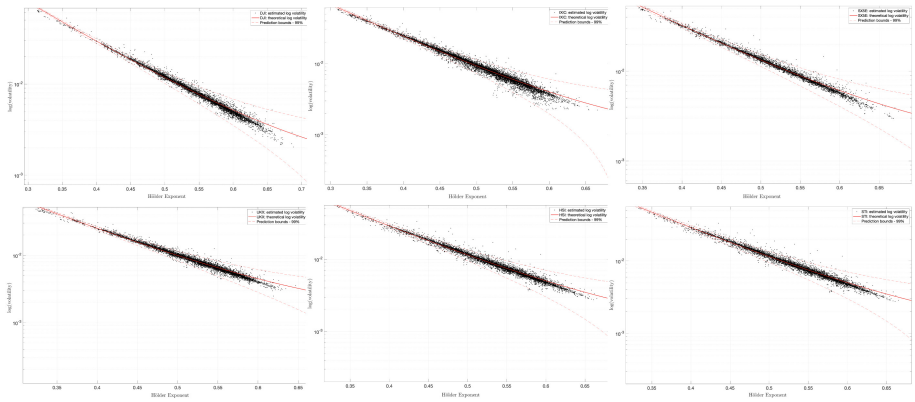


Fig. 1. Realized log volatility versus estimated Hölder exponents (one-day log-changes). X-axis: estimated Hölder exponent; Y-axis: estimated log volatility (black dots), theoretical relation given by Eq. (7) (red line), 99% prediction bounds (dashed red lines). H_t and realized volatility are estimated with a rolling window of 20 trading days.

- Volatility serves as a relative measure, indicating whether a market or asset currently displays more or less variability compared to past periods. However, it does not have the capacity to determine the "optimal" or "fair" level of volatility, one that aligns with an efficient market. In contrast, the Hölder parameter, ranging from 0 to 1, equals 1/2 only when the process aligns with Brownian motion, a key aspect of the Efficient Market Hypothesis (EMH).
- As markets naturally gravitate towards the equilibrium state associated with $H_t = 1/2$ following deviations, the distance $|H_t - 1/2|$ becomes a significant indicator for determining optimal buying or selling times. The dynamics of H_t are expected to exhibit a fluctuating trend around the value 1/2, with the rate of return to this equilibrium increasing as the deviation widens. Essentially, this mechanism provides a stochastic formalization and a theoretically grounded explanation for the commonly known trader adage "*What goes up, must come down.*"
- When financial prices exhibit local behavior resembling a fBm (as seen in processes like MPRE), the relationship between volatility and the Hölder exponent can be expressed through Eq. (7). Consequently, using the Hölder parameter instead of volatility does not result in any loss of information.

Table 2 provides a summary of the relationship between the Hölder exponent and the martingale condition, offering a financial interpretation of this connection [4]. Unlike volatility, the Hölder exponent offers a comprehensive assessment of market dynamics, addressing both the magnitude (*how much*) and character (*how*) of price variability. It provides insights into the deviation from equilibrium, represented by the value $H_t = 1/2$, which acts as a benchmark for a semi-martingale. The pointwise Hölder exponent serves as a descriptor of the prevailing dynamics of the discounted price process at a specific moment, dis-

tinguishing among a *momentum market* (linked to bullish phases or speculative bubbles), a *sideways market* (indicative of directionless efficiency), and a *reversal market* (resulting from rapid buy-and-sell activities, often following significant price adjustments or periods of uncertainty).

Table 2. Financial interpretation of H_t

H_t	Stochastic patterns	Agents' beliefs	Market patterns
$> \frac{1}{2}$	Persistence - Smooth paths - $\langle X \rangle_{2,t} = 0$	New information confirm outstanding position	"Low" volatility - Momentum Overconfidence - Underreaction
$= \frac{1}{2}$	Independence - Martingale - $\langle X \rangle_{2,t} = t$	Information fully incorporated by price	"Normal" volatility - Sideways market - Efficiency
$< \frac{1}{2}$	Mean-reversion - Rough paths - $\langle X \rangle_{2,t} = \infty$	New information disrupt outstanding position	"High" volatility - Reversals - Overreaction

This interpretation suggests that the apparently conflicting paradigms of Rationality and Behavioral Finance can coalesce within a comprehensive framework of bounded rationality, providing a more nuanced understanding of market dynamics. Within this framework, the pointwise Hölder exponent explicitly identifies *when* rationality transitions to irrationality, a shift that volatility fails to capture because of its insensitiveness to changes in the sign and intensity of autocorrelation. In this sense, assuming the FSR model, the *fair volatility* is the value corresponding to the value 1/2 of the Hölder exponent, via relation (7).

References

1. Angelini, D., Bianchi, S.: Nonlinear biases in the roughness of a fractional stochastic regularity model. Chaos, Solitons Fractals **172**, 113550 (2023)
2. Ayache, A.: Continuous gaussian multifractional processes with random pointwise Hölder regularity. J. Theor. Probab. **26**, 72–93 (2013)
3. Ayache, A., Bouly, F.: Moving average multifractional processes with random exponent: lower bounds for local obscillations. Stoch. Process. Appl. **146**, 143–163 (2022)
4. Bianchi, S., Pianese, A.: Time-varying Hurst-Hölder exponents and the dynamics of (in)efficiency in stock markets. Chaos, Solitons Fractals **109**, 64–75 (2018)
5. Comte, F., Renault, E.: Long memory in continuous-time stochastic volatility models. Math. Financ. **8**(4), 291–323 (1998)
6. Fama, E.: Efficient capital markets: a review of theory and empirical work. J. Financ. **25**(2), 383–417 (1970)
7. Gatheral, J., Jaisson, T., Rosenbaum, M.: Volatility is rough. Quant. Finan. **18**(6), 933–949 (2018)
8. Loboda, D., Mies, F., Steland, A.: Regularity of multifractional moving average processes with random Hurst exponent. Stoch. Proces. Appl. **140**, 21–48 (2021)
9. Pianese, A., Bianchi, S., Palazzo, A.: Fast and unbiased estimator of the time-dependent Hurst exponent. Chaos **28**(31102), 1–6 (2018)

The Market Value of Optimal Annuitization and Bequest Motives

Matteo Buttarazzi and Gabriele Stabile(✉)

Dipartimento di Metodi e Modelli per L'Economia, Il Territorio e la Finanza, Sapienza-Università di Roma, Roma, Italy
gabriele.stabile@uniroma1.it

Abstract. Since the seminal contribution of Yaari (1965), who showed that individuals with no bequest motive should convert all their retirement wealth into annuities, a number of papers have analysed the annuitization decision under the so-called *all or nothing* institutional arrangement, where immediate lifetime annuities are purchased just at a one point in time. In this paper, we investigate the effect of linear bequest motives on the annuitization decision for a retired individual who maximizes the market value of future cash flows. Finally, we present numerical examples analyzing optimal annuitization under strong or weak bequest motives.

Keywords: optimal stopping · annuities · bequest motives

1 Introduction

An immediate annuity is an insurance product that pays the annuitant a regular income for as long as he is alive, in exchange for a premium. The annuitization decision has important economic implications because it has a direct effect on the financial resources to support consumption in retirement age. The purchase of an annuity helps individuals to manage the risk of outliving their financial wealth, but it is usually an irreversible transaction, and most annuity contracts impose steep penalties if policyholders want to access their money in the early years of the contract. The natural alternatives to annuitization are the so-called *do-it-yourself* strategies, i.e. the individual asset allocation amongst various financial investment classes. However, it should be taken into account the investment risk as well as the longevity risk to which individuals would be exposed.

Since the seminal paper by Yaari ([10], the study of the annuitization decision has been the subject of a whole research field (see [3–7,9] among others).

This paper would contribute to this literature by investigating to what extent linear bequest motives (see e.g. [6]) affect the annuitization decision. At this aim, we consider an individual whose retirement wealth is invested in a financial fund which eventually must be converted into an annuity. As in [7], we consider two different mortality forces: a *subjective* one, used by the individual to weight

M. Corazza et al. (Eds.): MAF 2024, *Mathematical and Statistical Methods for Actuarial Sciences and Finance*, pp. 67–73, 2024.
https://doi.org/10.1007/978-3-031-64273-9_12

future cashflows (denoted μ), and a *objective* one, used by the insurance company to price the annuity (denoted $\hat{\mu}$). The interplay between these two different mortality forces contributes to some key qualitative aspects of the optimal annuitization decision. Before annuitization the individual's wealth is invested in the financial market, and at the time of an annuity purchase, the entire wealth is converted into a lifetime annuity. The central idea is to compare the value deriving from an immediate annuitization with the value of continuing the investment in the financial market. The optimization criterion pursued by the individual is the maximization of the present value of future expected cash-flows, via the optimal timing of the annuity purchase. In particular, the individual takes explicitly into account the presence of linear bequest motives, with a parameter that measures its strength.

The rest of the paper is organized as follows. In Sect. 2 we introduce the financial and actuarial assumptions and then the optimal annuitization problem. In Sect. 3 we perform its analytical study providing the explicit solutions. In Sect. 4 we present some numerical examples to discuss how the presence of bequest motives affects the annuitization decision.

2 Problem Formulation

In this study we are interested in the portion of the individual's wealth dedicated to retirement needs. Such wealth is invested in a financial fund which eventually will be converted into an annuity. The value $(X_t)_{t\geq 0}$ of the financial fund is modelled by a stochastic process on a filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, \mathbb{P})$. Letting $(B_t)_{t\geq 0}$ be a Brownian motion adapted to $(\mathcal{F}_t)_{t\geq 0}$, the fund's value evolves according to

$$\begin{cases} dX_t^x = (\theta - \alpha)X_t^x dt + \sigma X_t^x dB_t, & t > 0 \\ X_0^x = x \geq 0, \end{cases} \tag{1}$$

where θ is the average continuous return of the financial investment, α is the constant dividend rate and $\sigma > 0$ is the volatility coefficient.

We consider an individual whose age $\eta \geq 0$ is fixed at time 0. At time $t \geq 0$ the individual uses a constant *subjective* mortality force $\mu \in \mathrm{R}_+$ to compute her self-assessed life expectancy ${}_zp_\eta = e^{-\mu z}$, i.e. the subjective probability to survive z years. Furthermore, the probability that the individual dies during the next z years is ${}_zq_{\eta+t} = 1 - {}_z\, p_{\eta+t}$. The insurance company instead relies on a so-called objective survival probability function ${}_z\hat{p}_{\eta+t} = e^{-\hat{\mu} z}$ where $\hat{\mu} \in \mathrm{R}_+$ is the constant *objective* mortality force. The different survival probability functions account for the imperfect information available to the insurer on the individual's risk profile. The value at time $t > 0$ of a unitary life annuity is given by $\hat{a}_{\eta+t} = \int_0^\infty e^{-\hat{\rho} u}\, {}_u\hat{p}_{\eta+t} du$. Here $\hat{\rho}$ is the interest rate guaranteed by the insurer. The individual evaluates the expected present value of a unitary lifetime annuity by using the coefficient $a_{\eta+t} = \int_0^\infty e^{-\rho u}\, {}_up_{\eta+t} du$. In case the annuity is purchased at time t, the constant cash flow paid by the insurer is $P_t = \frac{X_t - K}{\hat{a}_{\eta+t}}$, where the constant K is either a fixed acquisition fee ($K > 0$) or a tax incentive ($K < 0$).

The optimisation criterion is the maximization of the present value of future expected cash-flows. Let $\tau_d : \Omega \to \mathrm{R}_+$ be the residual lifetime of the individual of age η (τ_d is assumed to be independent of the Brownian motion for all $t \geq 0$). Letting τ be the time of the annuity purchase, the value function of the optimisation problem is defined by

$$V(x) = \sup_{\tau \geq 0} \mathbb{E}\left[\int_0^{\tau_d \wedge \tau} e^{-\rho t} \alpha X_t^x dt + \mathbb{1}_{\{\tau_d \leq \tau\}} e^{-\rho \tau_d} \nu X_{\tau_d}^x + P_\tau \int_{\tau_d \wedge \tau}^{\tau_d} e^{-\rho t} dt\right], \quad (2)$$

where ρ is the individual's constant discount rate, $\nu \in [0, 1]$ measures the strength of the bequest motives. Before annuitization, i.e. for $t < \tau$, the individual receives the dividends from the fund at rate α. After annuitization, i.e. for $t > \tau$, she gets the annuity payment at a constant rate P_τ. In case the individual dies before the time of the annuity purchase, i.e. on the event $\{\tau_d \leq \tau\}$, she leaves a bequest equal to her wealth.

The optimisation problem (2) may be rewritten as follows

$$V(x) = \sup_{\tau \geq 0} \mathbb{E}\left[\int_0^{\tau} e^{-(\rho+\mu)t}(\alpha + \mu\nu) X_t^x dt + e^{-(\rho+\mu)\tau} \delta (X_\tau^x - K)\right], \quad (3)$$

with $\delta := \frac{(\hat{\rho}+\hat{\mu})}{(\rho+\mu)}$. To ensure the finiteness of the value function, throughout this paper, we assume that

Assumption 1. $\theta - \alpha - \mu - \rho < 0$

3 Analysis of the Optimal Stopping Problem

To analyze problem (3), we rely on the geometric approach to the optimal stopping problem (see [2]).

Our first task is to put the optimal stopping problem (3) in the form

$$\sup_{\tau} \mathbb{E}\left[e^{-w\tau} G(X_\tau)\right],$$

where w is a discount rate and $G(\cdot)$ is a reward function.

Noticing that $\mathbb{E}\left[\int_0^\infty e^{-(\rho+\mu)t}(\alpha + \mu\nu) X_t^x dt\right] = \beta x$ where $\beta := \frac{\alpha+\mu\nu}{\rho+\alpha+\mu-\theta}$, we may rewrite (3) as follows

$$V(x) = \beta x + \sup_{\tau \geq 0} \mathbb{E}\left[e^{-(\rho+\mu)\tau}\big((\delta - \beta) X_\tau^x - \delta K\big)\right] \quad (4)$$

Therefore, we limit ourself to study the problem

$$v(x) = \sup_{\tau \geq 0} \mathbb{E}\big[e^{-(\rho+\mu)\tau} G(X_\tau^x)\big] \quad (5)$$

where

$$G(x) := (\delta - \beta)x - \delta K. \tag{6}$$

As usual in optimal stopping theory, we let $\mathcal{C} = \{x \in \mathrm{R}_+ : v(x) > G(x)\}$ and $\mathcal{S} = \{x \in \mathrm{R}_+ : v(x) = G(x)\}$ be respectively the so-called continuation and stopping regions since, as long as $X_t \in \mathcal{C}$, it is not optimal to stop the diffusion. Assumption (1) and standard optimal stopping results (see [8, Cor. 2.9, Sect. 2]) guarantee that $\mathcal{C}$ is an open connected set and $\tau_* := \inf\{t \geq 0 : X_t^x \in \mathcal{S}\}$ is optimal for $v(x)$, i.e. the optimal stopping time is the first entry time of X in $\mathcal{S}$. Define the infinitesimal generator $\mathbb{L}$ of X by $(\mathbb{L}u)(x) = \frac{1}{2}\,\sigma^2 x^2 u''(x) + (\theta - \alpha)xu'(x)$, for any $u(\cdot)$ two time continuously differentiable. Then (see e.g. [1], pp. 18–19), there exist two linearly independent, strictly positive solutions of the ordinary differential equation $\mathbb{L}u = (\rho + \mu)u$, i.e. $\psi(x) = x^{\gamma_+}$ and $\phi(x) = x^{\gamma_-}$ where γ_+ and γ_- solve

$$\frac{1}{2}\,\sigma^2\gamma(\gamma - 1) + (\theta - \alpha)\gamma - (\rho + \mu) = 0.$$

Notice that $\gamma_+ > 1$ and $\gamma_- < 0$. As in [2], we define the strictly increasing function $y = F(x) = \frac{\psi(x)}{\phi(x)} = x^{\gamma_+ - \gamma_-}$, together with its inverse function $F^{-1}(y) = y^{\frac{1}{\gamma_+ - \gamma_-}}$, and set

$$\hat{G}(y) := \begin{cases} 0 & \text{if } y = 0, \\ \left(\frac{G}{\phi} \circ F^{-1}\right)(y) & \text{if } y > 0. \end{cases} \tag{7}$$

The following result due to Dayanik and Karatzas (see [2]) relates the convexity of the function $\hat{G}$ to the form of the continuation region and computes the value function.

Theorem 2. *$\hat{G}(y)$ is strictly convex if and only if $(\mathbb{L} - (\rho + \mu))G(x) > 0$. Moreover, let $Q(\cdot)$ be the smallest nonnegative concave function that dominates $\hat{G}(y)$. Then $v(x) = \phi(x)Q(F(x))$.*

In our case,

$$\hat{G}(y) = (\delta - \beta)y^{\frac{1-\gamma_-}{\gamma_+ - \gamma_-}} - \delta K y^{\frac{-\gamma_-}{\gamma_+ - \gamma_-}}, \quad y > 0. \tag{8}$$

Define

$$y_1 = \Big[- \frac{\delta K \gamma_-}{(\delta - \beta)(1 - \gamma_-)} \Big]^{\gamma_+ - \gamma_-},$$

and

$$y_2 = \Big[- \frac{\delta K \gamma_- \gamma_+}{(\delta - \beta)(1 - \gamma_-)(\gamma_+ - 1)} \Big]^{\gamma_+ - \gamma_-}.$$

It is easy to see that $y_1 < y_2$. Depending on the values of the model parameters, we distinguish the following cases

1. **Case** $\delta > \beta$
   (a) If $K \geq 0$ then $\hat{G}'(y) > 0$ for $y > y_1$ and $\hat{G}''(y) < 0$ for $y > y_2$. The smallest nonnegative concave function $Q(y)$ that dominates $\hat{G}(y)$ is

$$Q(y) = \begin{cases} m^* y & \text{if } y < y^*, \\ \hat{G}(y) & \text{if } y \geq y^*, \end{cases}$$

   where m^* and y^* are solutions of the system $\begin{cases} \hat{G}(y) = my \\ \hat{G}'(y) = m. \end{cases}$ In other words, m^* and y^* are such that the line $y = m^* y$ is tangent to $\hat{G}(y)$ at the point y^*. Returning to the variable x, we find that the continuation and stopping regions are respectively $\mathcal{C} = (0, x^*)$ and $\mathcal{S} = [x^*, \infty)$, with $x^* = \frac{\delta K}{(\delta - \beta)} \frac{\gamma_+}{\gamma_+ - 1}$.
   (b) If $K \leq 0$ then $\hat{G}'(y) > 0$ and $\hat{G}''(y) < 0$ for all $y > 0$. Therefore, $\mathcal{S} = [0, \infty)$, i.e. the annuity is immediately purchased whatever is the initial wealth x.
2. **Case** $\delta < \beta$
   (a) If $K \geq 0$ then $\hat{G}'(y) < 0$ and $\hat{G}''(y) > 0$ for all $y > 0$. Therefore, $\mathcal{C} = (0, \infty)$, and it is never optimal to purchase an annuity.
   (b) If $K < 0$ then $\hat{G}'(y) > 0$ for $y < y_1$ and $\hat{G}''(y) > 0$ for $y > y_2$. The smallest nonnegative concave function $Q(y)$ that dominates $\hat{G}(y)$ is

$$Q(y) = \begin{cases} \hat{G}(y) & \text{if } y \leq y_1, \\ \hat{G}(y_1) & \text{if } y \geq y_1. \end{cases} \tag{9}$$

   In other words, for $y \geq y_1$ the function Q is an horizontal line. The continuation and stopping regions are respectively $\mathcal{C} = (x^{**}, \infty)$ and $\mathcal{S} = [0, x^{**}]$, with $x^{**} = F^{-1}(y_1) = \frac{\delta K}{(\delta - \beta)} \frac{\gamma_-}{1 - \gamma_-}$.

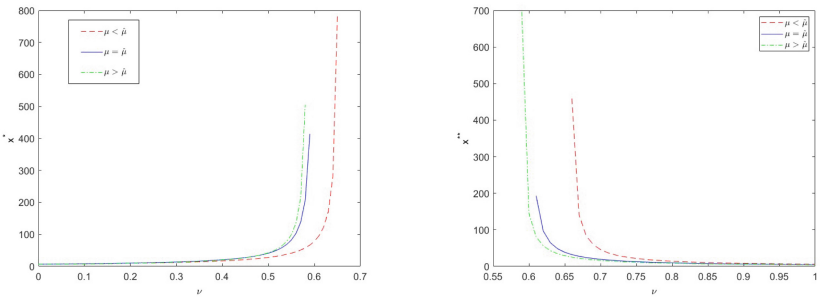
4 Numerical Application

Here we present a numerical application of the results obtained in the previous sections. Fix the following set of parameters $\alpha = 2\%$, $\theta = 7\%$, $\sigma = 6\%$, $\rho = \hat{\rho} = 5\%$. Notice that Assumption 1 is satisfied.

In Fig. 1.(a) we look at case 1.(a) and the optimal stopping threshold x^* between the continuation and stopping regions is plotted when the parameter ν increases, in three different mortality scenarios: $\mu = \frac{1}{20}$, $\hat{\mu} = \frac{1}{18}$, $\mu = \hat{\mu} = \frac{1}{20}$, $\mu = \frac{1}{18}$, $\hat{\mu} = \frac{1}{20}$. Notice that x^* increases as ν increases, that is as the strength of bequest motives increases the continuation region enlarges. If $\mu < \hat{\mu}$ ($> \hat{\mu}$) then the individual believes she is healthier (respectively unhealthier) than the average. For a fixed value of ν, moving from the first to the third scenario x^* increases, and then the continuation region becomes progressively larger.

In Fig. 1.(b) we look at case 2.(b) and the optimal stopping threshold x^{**} is plotted when the parameter ν increases, in the three different scenarios. Notice that, in all scenarios, x^{**} decreases as ν increases, i.e. the continuation region

progressively enlarges as the strength of the bequest motives increases. For a fixed value of ν, moving from the first to the third scenario x^{**} decreases, and then the continuation region becomes progressively larger.



(a) (b)

Fig. 1. The threshold x^* (a) and x^{**} (b) in three different mortality scenarios

Finally, notice that if $K = 0$, then either $\mathcal{S} = [0, \infty)$ or $\mathcal{C} = (0, \infty)$. In particular, letting for example $\mu = \hat{\mu} = \frac{1}{20}$, we find that if $\nu < 0.6$ then $\mathcal{S} = [0, \infty)$, and if $\nu > 0.6$ then $\mathcal{C} = (0, \infty)$. In other words, in case of actuarially fair annuities and zero acquisition fee/tax incentive, if the strength of bequest motives is low enough then the individual immediately purchases the annuity as Yaari ([10]) found. On the other hand, if the strength of bequest motives is high enough then Yaari's result does not hold anymore and the individual never purchases the annuity.

In sum, these numerical examples show that the continuation region enlarges as the strength of bequest motives increases. This means that consistent bequest motives may explain the scarce propensity to purchase annuities.

References

1. Borodin, A.N., Salminen, P.: Handbook of Brownian Motion-Facts and Formulae. Birkhäuser (2002)
2. Dayanik, S., Karatzas, I.: On the optimal stopping problem for one dimensional diffusions. Stochastic Process. Appl. **107**(2), 173–212 (2003)
3. De Angelis, T., Stabile, G.: On the free boundary of an annuity purchase. Fin. Stoch. **23**, 97–137 (2019)
4. Gerrard, R., Højgaard, B., Vigna, E.: Choosing the optimal annuitization time post-retirement. Quant. Finance **12**(7), 1143–1159 (2012)
5. Hainaut, D., Deelstra, G.: Optimal timing for annuitization, based on jump diffusion fund and stochastic mortality. J. Econ. Dyn. Control **44**, 124–146 (2014)
6. Lockwood, L.M.: Bequest Motives and the annuity puzzle. Rev. Econ. Stud. **15**, 226–243 (2012)

7. Milevsky, M.A., Young, V.R.: Annuitization and asset allocation. J. Econ. Dyn. Control **31**, 3138–3177 (2007)
8. Peskir, G., Shiryaev, A.N.: Optimal stopping and free-boundary problems, Lectures in Mathematics, ETH Zürich, Birkhäuser (2006)
9. Stabile, G.: Optimal timing of the annuity purchase: a combined stochastic control and optimal stopping problem. Int. J. Theor. Appl. Finance, **9**(2), 151–170 (2006)
10. Yaari, M.E.: Uncertain lifetime, life insurance and the theory of consumer. Rev. Econ. Stud. **32**, 137–150 (1965)

The Cost of Longevity Risk Transfer by Capital Solution De-risking Strategy

Maria Carannante[1], Valeria D'Amato[2], and Massimiliano Menzietti[3](✉)

[1] Department of Management and Quantitative Sciences, University of Naples Parthenope, Naples, Italy
maria.carannante@collaboratore.uniparthenope.it
[2] MEMOTEF Department, Sapienza University of Rome, Rome, Italy
valeria.damato@uniroma1.it
[3] Department of Economical and Statistical Sciences, University of Salerno, Fisciano, Italy
mmenzietti@unisa.it

Abstract. In this paper, we develop a longevity swap de-risking strategy to mitigate the impact of the longevity risk related to payments that depend on how long individuals are going to live. In order to ensure the development of an effi-cient capital market for longevity risk transfers, the longevity hedge would allow longevity risk to be shared efficiently and fairly between the parties. Our results show that the fixed proportional risk premium that the counter-party requires to take on the longevity risk varies by changing the mortality model adopted to represent the evolution of the longevity of the population underlying the swap and that, as the risk premium changes, the total transfer of longevity risk may become inefficient.

Keywords: longevity risk · de-risking strategy · longevity swap

1 Introduction

Every institutions and governments making payments that depend on how long individuals are going to live face with longevity risk, the risk that individuals live longer than expected. In particular, defined benefit pension plan sponsors, annuity providers are transferring these obligations, to life (re)insurers via insurance and capital solutions such as for instance buy-outs, buy-ins, longevity swaps and so on. Nevertheless, as the demand of longevity risk protection increases, the key question consists in capability of (re)insurance sector to cope with future increasing potential liabilities of the longevity risk exposures [2,4]. The innovative capital market solutions for transferring longevity risk consist in several forms of transactions, each differing in the types of risk transferred and the categories of risk created, including longevity (or survivor) bonds, longevity (or survivor) swaps, mortality (or q-)forward contracts, and reinsurance sidecars (also called strategic reinsurance vehicles). According to Reinsurance Group of

M. Corazza et al. (Eds.): MAF 2024, *Mathematical and Statistical Methods for Actuarial Sciences and Finance*, pp. 74–79, 2024.
https://doi.org/10.1007/978-3-031-64273-9_13

America (RGA) that has completed a new US$1.7 billion longevity risk transfer, the longevity swap arrangement covers roughly 11,000 single premium immediate annuity contracts, transferring the longevity related risk away [7].

A longevity swap transaction is based on periodic fixed payments that are paid to the swap counterparty in exchange for periodic payments according to the difference between the actual and expected pension or annuity mortality experience. When the longevity swap is index-based, the mortality experience is represented by the standardised population cohorts ("index swaps"). De-risking strategies may broaden, above all by considering that risk-mitigation instruments should not involve material basis risk that is intrinsically inherent the longevity hedges. Furthermore, regulatory restrictions could affect the technical forms and options. In particular, the feasibility of the de-risking transactions depends on the appropriate cost of the longevity risk transfer. It is well-known that the degree of the cost-efficient longevity de-risking solutions may stimulate or on the contrary deflate the market's potential for further risk transfers.

In this paper we investigate how the fixed proportional risk premium that the counterparty requires to take on the longevity risk varies with the underlying mortality model adopted. Our results show that, as the risk premium changes, the transfer of longevity risk may become more or less effective.

The remainder of the paper is organized as follows: in Sect. 2 we introduce de-risking strategy based on longevity swap and define the optimization problem that determines the optimal proportion of risk that should be transferred. Section 3 shows the main findings of the numerical application. Section 4 concludes.

2 De-risking Strategies

Let us consider a portfolio of annuitants all aged x_0 at time 0, we define ${}_sp_{x,t}$ the probability that an individual alive at time t, with age x, survives to age $x+s$ at year $t+s$, and ${}_s\hat{p}_{x,t}$ its conditional expected value. We denote with $v = \frac{1}{1+r}$ the discount factor with the discount rate r (assumed to be deterministic), we define the conditional expected value of a life annuity $a(x(t))$ as $\mathbb{E}[a_{\overline{k(x)}}|p_{x,t}, {}_2p_{x,t}, \dots] = \sum_{s=1}^{\omega-x} v^s {}_s\hat{p}_{x,t}$.

Let A_0 and V_0 the asset and portfolio expected liabilities at time 0. The initial unfunded liabilities, UL_0, are given by $V_0 - A_0$. The insurer benefit liability at time t, is the discounted expected value of future benefits, with B_t the total annual benefit in t. B_t is given by the product of the individual benefit b, assumed to be equal for all the insureds, and the number of survived annuitants, n_t: $B_t = b \cdot n_t$. Let J_t be the return on assets from $t-1$ to t at $j(t-1,t)$ rate, this implies that $J_t = A_{t-1} \cdot j(t-1,t)$. Denoting K_t the capital flow in the year t, the portfolio asset is given by:

$$A_t = A_{t-1} + J_t + K_t - B_t \tag{1}$$

while UL_t, without de-risking strategy, is obtained as follows:

$$UL_t = V_t - A_{t-1} - J_t + B_t \tag{2}$$

If a time t the unfolded liabilities are greater than zero, $UL_t > 0$, the insurer experiences a portfolio loss and vice versa. Furthermore, we assume that, the insurer amortizes the unfunded liability year by year, that is $K_t = UL_t \ \ \forall t$.

We consider a de-risking strategy for longevity risk based on a longevity swap (LS), written on living policyholders, with hedge cost HC^{LS}. We denote with $in_t^{LS} = max(K_t^{LS}, 0)$ and $out_t^{LS} = max(-K_t^{LS}, 0)$ the present value of capital inflows and capital outflows, subject to constant penalty factors ψ_1 and ψ_2, respectively. ψ_1 represents the opportunity cost due to the need to increase the capital and ψ_2 the opportunity cost due to lock capital that could have been invested otherwise. The total portfolio cost TPC^{LS} of the strategy is obtained as:

$$TPC^{LS} = HC^{LS} + \sum_{t=1}^{\omega-x} \frac{in_t^{LS}(1+\psi_1) - out_t^{LS}(1+\psi_2)}{(1+r)^t} \tag{3}$$

Denoting with HCF_t^{LS} the hedging cash flows from de-risking strategy, the unfunded liabilities with the de-risking strategy is given by:

$$UL_t^{LS} = V_t - A_{t-1} - J_t + B_t - HCF_t^{LS} \tag{4}$$

We denote with TUL^{LS} the total unfunded liabilities over the entire time horizon of the de-risking strategy based on LS:

$$TUL^{LS} = \sum_{t=1}^{\omega-x} \frac{UL^{LS}}{(1+r)^t} \tag{5}$$

We consider a plain vanilla longevity swap written on n_0 survivors. We define the fixed leg of longevity swap at time t as $b \cdot n_0 \cdot {}_t\hat{p}_x(1+\pi)$ (where π is the fixed proportional risk premium that the counterpart requires to take on longevity risk) and the floating leg as $b \cdot n_0 \cdot {}_tp_x$. At each $t, t = 1, 2, \ldots$ the LS payoff is given by the difference between the floating and the fixed leg: $b \cdot n_0[{}_tp_x - {}_t\hat{p}_x(1+\pi)], \quad t = 1, 2, \ldots$

Setting π so that the swap value is zero at the inception date, the swap price is null, $HP^{LS} = 0$. With a hedging proportion of h^{LS}, the hedging cost is equal to:

$$HC^{LS} = -h^{LS} \cdot b \cdot n_0 \cdot \mathbb{E}\left[\sum_{t=1}^{\omega-x} d(0,t)[{}_tp_x - {}_t\hat{p}_x(1+\pi)]\right] \tag{6}$$

Following [6], the optimal hedge level for the de-risking strategy can be obtained solving an optimisation problem where the insurer aims to minimizing the Conditional Value-at-Risk of the total unfunded liabilities at a fixed confidence level α, $CVaR_\alpha(TUL^{LS})$, with respect to h^{LS}, subjected to the constraint that the total cost does not exceed a fixed amount c. This is formalised in the following non-linear optimization problem:

$$\begin{aligned} &\min_{h^{LS}} CVaR_\alpha[TUL^{LS}] \\ &\quad sub \\ &\quad \mathbb{E}[TPC^{LS}] \leq c \\ &\quad \mathbb{E}[TUL^{LS}] \leq 0 \\ &\quad 0 \leq h^{LS} \leq 1 \end{aligned} \tag{7}$$

As already noted in [6], the de-risking strategy is strongly influenced by its cost (which depends on π). On the other hand, the cost of the LS is also related to the assumptions on the evolution of mortality in terms of both trend and volatility. Therefore, the choice of the mortality model is a determining factor in defining the optimal de-risking strategy. In the following numerical application, we verify how the optimal risk transfer rate h^{LS} is influenced by the choice of the underlying mortality model, and how the existence of an information asymmetry on mortality trends between protection seller (short position on LS) and protection buyer (long position on LS) is crucial.

3 Numerical Application

We consider a portfolio of immediate temporary life annuities (with term $T = 20$), written on a cohort of males all aged 65 at issue ($t = 0$) with $n_0 = 10,000$. For sake of simplicity we assume $b = 1$, so $B_t = n_t$. Expenses and taxes are not considered in the valuation. The single premium, Π_X, is determined according to the Standard Deviation Principle ($\Pi_X = \mathbb{E}[X] + \lambda \cdot SD[X]$). We fix λ at 20%. We consider two different mortality models, the traditional Lee-Carter model [5] (LCA in the following) and the Lee-Carter model including a frailty factor proposed by [3] (denoted with ATFLCA). We estimate LCA and ATFLCA models for English 50–90 aged male population. For LCA model, we used data about death rates and exposures to risk only and we refer to the Human Mortality Database. For ATFLCA model, we also used data relating to the co-morbidity trend in the population, and we referred to the English Longitudinal Study on Ageing (ELSA) [1]. Performing 10,000 simulation, we obtain the evolution of the UL_t, the total unexpected losses, TUL, and the total portfolio costs TPC, without hedging. We then introduce the optimization problems setting a constraints for $\mathbb{E}[TPC]$: the maximum level c for the expected total cost related to strategy j is set in relation to its initial value (without hedging): $c = 0.5 \cdot \mathbb{E}[TPC]$. The following assumption are adopted in the evaluations:

- the initial asset, A_0, are equal to the total portfolio single premium;
- we assume a flat rate of return on asset ($j(t-1,t) = r = 0.02 \ \ \forall t$);
- the penalty factors in the TPC are: $\psi_1 = \psi_2 = 0.2$.

Adopting a demographic technical basis determined through the LCA model (see Table 1 columns 2 and 3), the initial portfolio liabilities are: $V_0 = 168,102$ while the total portfolio single premium, P, are: 168,810. The risk premium of the longevity swap, π, is set at 0.421%. This value was determined consistently with the standard deviation principle adopted for the determination of the single premium. Without hedging the $\mathbb{E}[TUL]$ is negative, denoting an expected profit, but the portfolio is characterized by a positive $CVaR_{99.5\%}[TUL]$ (with an average of the losses beyond the VaR at 99.5% of 8,239.21. The penalties ψ_1 and ψ_2 strongly reduce the profit, but $\mathbb{E}[TPC]$ is still negative. Results show that the optimal strategy minimizing $CVaR_{99.5\%}[TUL^LS]$ is obtained with LS share equal to 46.6%. (partial risk transfer). When a hedging strategy is introduced,

Table 1. Results with no hedging and swap strategies minimizing $CVaR_\alpha[TUL^{LS}]$: LCA scenario (second and third columns) and ATFLCA scenario (forth and fifth columns). Results with swap strategy minimizing $CVaR_\alpha(TUL^{LS})$ in presence of asymmetric information: ATFLCA scenario (sixth column).

	LCA scenario		ATFLCA scenario		Asymm. inf.
	No hedging	Hedging	No hedging	Hedging	Hedging
π^{LS}		0.421%		1.580%	0.421%
h^{LS}	0.0%	46.6%	0.0%	42.9%	100.0%
HC^{LS}	0.00	325.76	0.00	1,122.84	697.29
$\mathbb{E}[TUL]$	–707.71	-381.95	–2,647.99	–1,525.15	–1,950.70
$CVaR_{99.5\%}[TUL]$	8,239.21	4,490.62	29,712.79	17,277.74	–1,183.74
$\mathbb{E}[TPC]$	–96.00	–48.00	–357.91	–178.96	–1,465.93

$\mathbb{E}[TUL]$ is increased (but still negative) while the $CVaR_{99.5\%}[TUL]$ is almost halved. The expected total cost ($\mathbb{E}[TPC]$) is also increased, but still negative.

Adopting a demographic technical basis determined through the ATFLCA model (see Table 1 columns 4 and 5), the initial portfolio liabilities are: $V_0 = 167,595$ while the total portfolio single premium, P, are: 170,242. Without hedging the $\mathbb{E}[TUL]$ is negative, but the portfolio is characterized by a very high $CVaR_{99.5\%}[TUL]$, as a consequence of the high variability of the death probability simulated with the ATFLCA model. As a consequence, the $\mathbb{E}[TPC]$ is strongly increased even if still negative. Results show that the optimal strategy minimizing $CVaR_{99.5\%}[TUL^LS]$ is obtained with LS share equal to 42.9% (partial risk transfer). When a hedging strategy is introduced, $\mathbb{E}[TUL]$ is increased (but still negative) but the $CVaR_{99.5\%}[TUL]$ is drastically reduced. The expected total cost ($\mathbb{E}[TPC]$) is still negative but an half than in the absence of LS.

The last case we consider is one in which the protection seller (short position on LS) and the protection buyer (long position on LS) have different information on the insured population (see Table 1 column 6), so that the former considers appropriate to adopt the LCA model (and price the swap accordingly), while the latter considers the ATFLCA model more appropriate and determines the optimal de-risking strategy accordingly. Results show that the optimal strategy minimizing $CVaR_{99.5\%}[TUL]$ is obtained with LS share equal to 100% (total risk transfer). $\mathbb{E}[TUL]$ is increased but still negative (implying a profit)) but the $CVaR_{99.5\%}[TUL]$ is totally reduced and becomes negative. The expected total cost ($\mathbb{E}[TPC]$) is more negative, which implies an improvement of the annuity provider's position.

4 Conclusions

The aim of this paper is to analyse how de-risking strategies based on longevity swaps are affected by their cost, expressed as a fixed proportional risk premium that the counterparty requires to assume longevity risk, with a focus on the choice of the underlying mortality model. The mortality models considered are the Lee-Carter model and its extension including a frailty factor (ATFLCA model). Our results show that different mortality models imply a different risk assessment of an annuity portfolio. If both counterparties of the longevity swap assume the same mortality model, the cost of the de-risking strategy increases with the portfolio's riskiness and the optimal strategy for the annuity provider (protection buyer) is to transfer only part of the longevity risk to the protection seller, with the share depending on the model adopted. If we assume that the protection seller and the protection buyer have different information about the insured population, so that the former considers it appropriate to adopt the Lee-Carter model (and prices the swap accordingly) while the latter considers the ATFLCA model more appropriate, de-risking is more effective and optimal strategy is achieved through a total transfer of longevity risk.

References

1. Banks, J., et al.: English Longitudinal Study of Ageing: Waves 0-9, 1998-2019. [data collection]. 37th Edition. UK Data Service. SN: 5050 (2021). https://doi.org/10.5255/UKDA-SN-5050-24
2. Blake, D., Cairns, A.J.G., Dowd, K., Kessler, A.R.: Still living with mortality: the longevity risk transfer market after one decade. Br. Actuar. J. **24**(1), 1–80 (2019). https://doi.org/10.1017/s1357321718000314
3. Carannante, M., D'Amato, V., Haberman, S., Menzietti, M.: Frailty-based lee carter family of stochastic mortality models. Qual. Quant. (2023). https://doi.org/10.1007/s11135-023-01786-6
4. Kessler, A.R.: New solutions to an age-old problem: innovative strategies for managing pension and longevity risk. N. Am. Actuar. J. **25**(sup1), S7–S24 (2021). https://doi.org/10.1080/10920277.2019.1672566
5. Lee, R.D., Carter, L.R.: Modeling and forecasting U.S. mortality. J. Am. Stat. Assoc. **87**(419), 659–671 (1992)
6. Lin, Y., MacMinn, R.D., Tian, R.: De-risking defined benefit plans. Insur. Math. Econ. **63**, 52–65 (2015). https://doi.org/10.1016/j.insmatheco.2015.03.028
7. RGA in $1.7bn longevity swap & reinsurance with Western & Southern (2022). http://www.artemis.bm

Cyber Insurance and Risk Assessment: Some Insights on the Insurer Perspective

Maria Francesca Carfora and Albina Orlando(✉)

Istituto per le Applicazioni Del Calcolo "Mauro Picone". Consiglio Nazionale delle Ricerche, Via P. Castellino, 80131 Napoli, Italy
{f.carfora,a.orlando}@na.iac.cnr.it

Abstract. Cyber insurance is a crucial tool for managing risks associated with cyber threats. A challenging task for insurance companies lies in pricing cyber risk. Our study is motivated by the reasonable assumption that firms entering into cyber insurance contracts face diverse cyber threats in terms of types and magnitude. Considering these differences ensures that premiums align with the actual risk exposure of the insured. The study discusses this approach proposing a case study based on the Chronology of Data Breaches provided by the Privacy Rights Clearinghouse.

Keywords: cyber risk · cyber insurance · premium · data breaches

1 Introduction

In contemporary society, our dependence on information systems offers significant opportunities but also brings new risks. Cyber insurance emerges as a key tool for managing these risks. Insurers not only provide the opportunity to relieve insureds from the need to accumulate capital for handling catastrophic events, but they also have the potential to incentivize appropriate cybersecurity measures through premiums and proactive security screening. Over the past five years, the worldwide cyber insurance market has tripled in size and projections indicate a further increase [1]. This industry faces unique challenges and obstacles that are not commonly encountered in traditional insurance markets, such as addressing the correlation of risks, managing the geographical dispersion of risk and dealing with the lack of historical and actuarial data [2]. Several papers extensively examine the relevant literature on cybersecurity insurance, research and practice, in order to draft the current landscape and present the trends, among the others [2,3]. Very insightful contributions concern the possibility of transferring cyber risk through insurance-linked securities (see [5]). A challenging task for insurance companies concerns the pricing of cyber risk even due to the lack of comprehensive data on security breaches and losses. Information on the current industry practices for pricing risks is available in [4]. Remarkable contributions on the topic of insurance policies pricing are given in [6–8].

M. Corazza et al. (Eds.): MAF 2024, *Mathematical and Statistical Methods for Actuarial Sciences and Finance*, pp. 80–86, 2024.
https://doi.org/10.1007/978-3-031-64273-9_14

Our study is motivated by the reasonable assumption that a firm signing a cyber insurance contract faces different cyber threats in terms of types and magnitude. Taking into account these differences, enables more precise pricing of cyber risk and guarantees that premiums align with the actual risk exposure of the insured. This approach allows to mitigate the risks associated with underinsurance or overinsurance, thereby minimizing the potential for premium leakage. The rest of the paper is structured as follows. Section 2 focuses on the pricing methodology. Section 3 discusses a case study and Sect. 4 concludes.

2 Pricing Cyber Risk

The standard assumptions of classical actuarial techniques are not as applicable to price cyber risk. In actuarial mathematics a standard model is the frequency-severity approach, also called collective risk model [9]. Despite the limitations of this approach in the context of the quantification of cyber risk, in any case it can be customized to account for cyber risk-at least as a first approximation. Let us consider a policy covering a given risk. During the *policy year*, a random number N (frequency) of claims will be recorded. Each claim will cause a random loss, L_k k=1, 2, 3,...(severity), to the insured. The insurer will assess the claim amount Y_k for claim k. In case of partial cover we have $Y_k < L_k$, while $Y_k = L_k$ in case of full compensation.

Referring to a single policy, the total payout of the insurer X (aggregate claim amount) during the policy year, is defined as follows: $X = 0$, if $N = 0$ and

$$X = \sum_{k=1}^{N} Y_k, N > 0 \tag{1}$$

The *equivalence premium* (fair premium) Π is given by the expected value of the insurer's payout $E[X]$:

$$\Pi = E[X]. \tag{2}$$

Usually the following assumptions hold: the random variables L_k, k = 1, 2,...,N, are independent of the random number N and the random variables L_k, k = 1, 2,...,N are mutually independent and identically distributed. Typically, insurer adds a safety loading to the fair premium thus obtaining the so-called *net premium*, that is, before loading expenses. Resulting principles are [10]:

the *expected value principle*, $P_{ev} = (1 + \alpha)\Pi$;
the *variance principle*, $P_{var} = \Pi + \alpha Var(X)$;
the *standard deviation principle*, $P_{sd} = \Pi + \alpha\sqrt{Var(X)}$;
the *semistandard deviation principle*, $P_{ssd} = \Pi + \alpha\sqrt{E\{[max(0, X - E(X)]^2\}}$
where $\alpha > 0$ is a constant.

Other premium principles are defined via *utility theory* and incorporate the attitude towards risk of the insurer. One example is the *exponential premium principle*. Given an appropriate constant $\rho > 0$, we have $P_{ut} = \frac{1}{\rho} log[E(e^{\rho \cdot X})]$.

Another principle of the premium assessment is based on the quantiles of the distribution of X and is given by $P_Q(\varepsilon) = F_X^{-1}(1-\varepsilon)$ where F is the distribution function of X and $\varepsilon \in (0,1)$ is the confidence level. It is the quantile of order $(1-\varepsilon)$ of the loss distribution and this means that the insurer wants to get the premium that covers $(1-\varepsilon)\cdot 100\%$ of the possible loss. A reasonable range of the parameter ε is usually from 1% to 5%.

Whatever is the principle to be used for the premium assessment, we need to make realistic assumption on the distribution of both the number of claims N and the claim amounts $Y_1, Y_2, ..., Y_N$. In the following, we consider two possible solutions. In the first case, the insurer takes into account that different claims can be caused by different type of incidents. This approach is described in Sect. 3. The second solution consists in estimating the payout of the insurer considering the distribution of the total aggregate claim amount X.

In order to give an example, in the following we assume that the insurance company prices the risk that a firm may incur financial losses as a result of a data breach which is the main cause of cyber incidents [12].

3 An Illustrative Example

Privacy Rights Clearinghouse (PRC) is a nonprofit organization focusing on data privacy rights and issues. Their Chronology of Data Breaches in the US [11] includes description and type of both the breach and the breached entity, along with the breach severity, measured in terms of the number of breached records, when available. For a detailed description of this dataset see [13]. We restrict our analysis to the more recent data (breaches reported after the 1st of January, 2010) because they could better represent the current cyber threat situation; we also select only breaches with complete information on breach sizes and cause. In some previous studies [13,14] we investigated the causes of data breaches and found significant differences in the distribution of the severity of breaches caused by accidental exposure or inadequate vigilance ("negligent" breaches) with respect to breaches originating from activities that actively targeted private information ("malicious" breaches). Then we decided to model these two distributions separately. In both cases, however, we found that the best fit for the severity is given by a skew-normal distribution. Regarding "negligent" breaches, it is worth specify that many employees are often the weakest link that causes a successful cyber incident [15].

In this illustrative example, we consider a generic organization belonging to the Business typology, that includes financial and banking services, manufacturing, retail. The total amount of registered breaches with full information available for this category in the PRC dataset for the period 2010–2019 includes 732 data. In order to estimate the appropriate premium for this organization, we follow some of the suggestions provided in [16] and build a simulation pipeline

to generate a large number of scenarios suitable to represent the losses distribution. Evidences from the literature and from the available data that guided us in developing this pipeline are the following:

- the probability for an organization of suffering $k \geq 0$ breaches in a year can be modeled by a geometric distribution whose parameter p has been estimated in [16] separately for the three business subcategories. A weighted average of these values results in the value $p = 0.91$;
- historical data on the type of data breaches for a generic Business type organization allows us to estimate the relative frequency of malicious (fm) and negligent (fn) events, so that a simulated event will be of malicious type with probability $fm/(fm + fn)$;
- once the severity distribution (in terms of the volume Y of breached data) has been fitted to the available data, the financial loss L for each breach event can be roughly estimated by a regression model on Y. This formula was originally derived by Jacobs [17] on Ponemon data and then improved by Farkas [16] to better represent extreme events:

$$\log(L) = 9.59 + 0.57 \log(Y). \tag{3}$$

To obtain reliable estimates for the annual losses of a Business type organization, we adopt a Monte-Carlo based simulation approach: once fixed a huge number N of scenarios

- for any $i \leq N$ we simulate a corresponding number of claims n_i by generating a random value from a geometric distribution with parameter p as mentioned before;
- then, in case $n_i > 0$, we generate a random value n_m from a Bernoulli distribution of parameter $fm/(fm + fn)$ to assign n_m events to the malicious category and the remaining $n_i - n_m$ to the negligent category;
- for each event, its severity is generated as a random value from the fitted skew-normal distribution (malicious or negligent) and the related financial loss estimated by the empirical formula (3).

Basing on the different methods described in Sect. 2, premiums can finally be evaluated from the simulated distribution of losses. Cyber insurance commonly distinguishes between "third-party" and "first-party" losses depending on whether they concern external parties or the insured itself. Jacobs transformation estimates both first and third-party losses and we make the same hypothesis. Moreover, we consider the case of full compensation.

4 Results and Conclusive Remarks

We present here some results obtained in the numerical simulations we performed according to the pipeline described in the previous section. As stated in Sect. 2, we also consider the alternative approach and proceed without splitting the simulated claims into malicious and negligent ones, but simply generate the

severity of each claim as a random value from the skew-normal distribution fitted on all data. In both cases, we generated $N = 1$ million scenarios and repeated the simulation pipeline 10 times, to obtain averaged estimates of the different premiums along with their standard errors.

In the columns of Table 1 we report the values we obtained for the fair premium Π, the expected value premium P_{ev}, the standard deviation premium P_{sd}, the semi-standard deviation P_{ssd}, the exponential premium P_{ut} and the quantile premium P_Q along with their standard errors. In all simulations we chose $\alpha = 0.1$, $\rho = 0.01$, $\varepsilon = 0.05$. The first row shows results of the described pipeline. In the second row, given for comparison, the premiums has been evaluated instead by estimating frequency and severity of the breaches without splitting the malicious and negligent events.

Table 1. Estimates of the annual premiums according to the principles described in Sect. 2 with the proposed methodologies.

	Π	P_{ev}	P_{sd}	P_{ssd}	P_{ut}	P_Q
Premiums	23647 ± 38	26011 ± 42	23648 ± 38	23647 ± 38	24034 ± 41	135382 ± 2245
Premiums(alt)	23429 ± 37	25772 ± 41	23430 ± 37	23430 ± 37	23801 ± 40	124746 ± 2787

First of all we observe that in both cases (first and second row of Table 1) all the estimated premiums based on the equivalence principle (P_{ev}, P_{sd}, and P_{ssd}) exceed the fair premium. This surplus serves as a buffer to offset adverse experiences. Regarding the principles involving standard deviation, the loading is associated with the variability of the loss. Indeed, P_{ut} and P_Q are higher than Π, too. In particular, P_{ut} is defined via *utility theory* and incorporates the attitude towards risk of the insurer. As regards P_Q, that is the quantile of order $(1-\varepsilon)$ of the loss distribution, it is significantly higher than the other premiums; the reason is that the insurer wants a premium that covers $(1-\varepsilon)\cdot 100\%$ of the estimated losses.

Regarding the comparison of the two methods, it can be noticed that premiums obtained by estimating frequency and severity of the breaches without splitting the malicious and negligent events (*Premiums(alt)*, second row of Table 1), are lower than the ones obtained with the methodology described in Sect. 3. Although in this specific example the difference is minimal, this is indicative of the fact that if the insurer estimates the premium without considering the different types of cyber incidents, it could incur an underestimation of the premium itself. This difference becomes more evident in reference to the quantile premium P_Q. This means that extreme events could be significantly underestimated by the insurer. In future research, we will test the methodology on other datasets allowing us to consider a richer range of types of incidents and make more extensive comparisons.

Acknowledgments. This work was partially supported by project SERICS (PE00000014) under the MUR National Recovery and Resilience Plan funded by the European Union - NextGenerationEU and performed under the auspicies of the GNCS of INdAM.

References

1. SwissRe Institute: what you need to know about the cyber insurance https://market.swissre.com/risk-knowledge/advancing-societal-benefits-digitalisation/about-cyber-insurance-market.html
2. Tsohou, A., Diamantopoulou, V., Gritzalis, S., Lambrinoudakis, C.: Cyber insurance: state of the art, trends and future directions. Int. J. Inf. Secur. **22**(3), 737–748 (2023). https://doi.org/10.1007/s10207-023-00660-8
3. Marotta, A., Martinelli, F., Nanni, S., Orlando, A., Yautsiukhin, A.: Cyber-insurance survey. Comput. Sci. Rev. **24**, 35–61 (2017). https://doi.org/10.1016/j.cosrev.2017.01.001
4. Romanosky, S., Ablon, L., Kuehn, A., Jones, T.: Content analysis of cyber insurance policies: how do carriers price cyber risk? J. Cybersecur. **5**(1), 1–19 (2018). https://doi.org/10.1093/cybsec/tyz002
5. Braun, A., Eling, M., Jaenicke, C.: Cyber insurance-linked securities. ASTIN Bull. **53**(3), 684–705 (2023). https://doi.org/10.1017/asb.2023.22
6. Awiszus, K., Knispel, T., Penner, I., Svindland, G., Vob, A., Weber, S.: Modeling and pricing cyber insurance. Eur. Actuar. J. **13**, 1–53 (2023). https://doi.org/10.1007/s13385-023-00341-9
7. Zeller, G., Scherer, M.: Risk mitigation services in cyber insurance: optimal contract design and price structure. Geneva Pap. Risk Insur. Issues Pract **48**, 502–547 (2023). https://doi.org/10.1057/s41288-023-00289-7
8. Maochao, X., Lei, H.: Cybersecurity insurance: modeling and Pricing. N. Am. Actuar. J. **23**, 220–249 (2019). https://doi.org/10.1080/10920277.2019.1566076
9. Olivieri, A., Pitacco, E.: Introduction to Insurance Mathematics: Technical and Financial Features of Risk Transfers. Springer International Publishing, EAA Series (2015)
10. Embrechts, P.: Actuarial versus financial pricing of insurance. J. Risk Finance **1**(4), 17–26 (2000). https://doi.org/10.1108/eb043451
11. Privacy Rights Clearinghouse. Chronology of Data Breaches (2018). https://www.privacyrights.org/data-breaches
12. Allianz Global Corporate & Specialty(AGCS). Allianz Risk Barometer 2019. Top Business Risks for 2019. https://www.agcs.allianz.com/content/dam/onemarketing/agcs/agcs/reports/Allianz-Risk-Barometer-2019.pdf
13. Carfora, M.F., Orlando, A.: Some remarks on malicious and negligent data breach distribution estimates. Computation 12 art. number **208** (2022). https://doi.org/10.3390/computation10120208
14. Carfora, M.F., Orlando, A.: Cyber risk: estimates for malicious and negligent breaches distributions. In: Mathematical and Statistical Methods for Actuarial Sciences and Finance, Springer International Publishing (2022). https://doi.org/10.1007/978-3-030-99638-3_23
15. OECD.:Enhancing the role of insurance in cyber risk management (2017). https://doi.org/10.1787/9789264282148-en

16. Farkas, S., Lopez, O., Maud, T.: Cyber claim analysis using generalized pareto regression trees with applications to insurance. Insur. Math. Econ. **98**, 92–105 (2021). https://doi.org/10.1016/j.insmatheco.2021.02.009
17. Jacobs, J.: Analyzing Ponemon cost of data breach (2014). datadrivensecurity.info/blog/posts/2014/Dec/ponemon/

Machine Learning for ESG Rating Classification: An Integrated Replicable Model with Financial and Systemic Risk Parameters

Rosella Castellano[1(✉)], Federico Cini[2], and Annalisa Ferrari[3]

[1] Department of Law and Economics, University of Roma UnitelmaSapienza, Piazza Sassari 4, 00161 Rome, Italy
rosella.castellano@unitelmasapienza.it

[2] Phd School in Social Sciences and Economics, Sapienza University of Rome, Piazzale Aldo Moro 5, 00185 Rome, Italy
federico.cini@uniroma1.it

[3] Department of Law and Digital Society, University of Roma UnitelmaSapienza, Piazza Sassari 4, 00161 Rome, Italy
annalisa.ferrari@unitelmasapienza.it

Abstract. The concept of Sustainability has been identified as a key factor in investment strategies for over a decade. Due to empirical evidence suggesting better risk-return profiles for sustainable investments than non-sustainable ones (under similar conditions), investors consider ESG ratings essential information for investment choices. Despite persistent inconsistencies and methodological uncertainties, new risk measures are perceived as useful in identifying risks associated with the Environmental, Social, and Governance pillars, both individually and collectively. This study aims to assess whether a selected set of financial statement variables and a dynamic measure of systemic risk observed at time t can provide useful information for identifying the ESG rating class of a company at time $t + 1$. To test this hypothesis, we use companies in the EuroStoxx 600 index for the period 2016–2021 and apply a Machine Learning (ML) model. Using a Random Forest (RF) classification model, we estimate the ESG rating class at time $t+1$ with unprecedented accuracy. This agile and parsimonious model can provide valuable information to sustainable investors for making strategic investment decisions.

Keywords: Machine Learning · Random Forest · Classification Models · ESG Ratings

1 Introduction

Environmental, Social, and Governance (ESG) ratings offer an assessment of a company or financial instrument's sustainability profile or characteristics, including its exposure to sustainability risks and its impact on society and/or the environment. The importance of sustainability for businesses has sparked a multitude

M. Corazza et al. (Eds.): MAF 2024, *Mathematical and Statistical Methods for Actuarial Sciences and Finance*, pp. 87–92, 2024.
https://doi.org/10.1007/978-3-031-64273-9_15

of applications across various sectors (see [2] and reference therein). For stakeholders, ESG rating services have become the backbone of responsible investing [12]. ESG ratings are associated with various economically and financially relevant effects: higher intensity of sustainability and compliance with ESG criteria appears to be correlated with lower capital costs [16], fewer capital constraints [6,9], lower analyst forecast errors [21], and lower stock price volatility primarily as a result of disclosure on ESG information [7].

The debate among scholars regarding the informational component of ESG ratings concerning price and return performance is far from concluded. Company-specific ESG classifications are associated with lower stock price synchronicity, causing prices to fluctuate differently due to company-specific information[23]. ESG ratings play an important informative role for investors, and ESG disclosure leads to more efficient stock prices [17]. The authors also suggest that the effect of ESG disclosure on the informative component of stock prices is not uniform across pillars, being more sensitive to the social (S) component. According to [18], ESG ratings can effectively predict investment outcomes, given a high positive relationship between ESG and stock returns.

However, transparency in the ESG assessment process is lacking, as rating agencies use proprietary models with limited disclosure of computation details [10]. Given the varied and multifaceted nature of ESG approaches employed by different providers, often resulting in non-convergent outcomes [3], the importance of identifying specific, replicable, and convergent metrics becomes clear [22]. This emphasis on simplification is particularly useful as sustainable investors can benefit from a clearer view, consequently making informed decisions based on relevant data directly linked to corporate performance and ESG objectives.

Considering the ESG rating class as a strategic tool for the sustainable investor, it becomes advantageous to identify metrics capable of determining the ESG rating class within specific time intervals. In [15], through Rough Set Analysis (RSA), it is showed that corporate performance measures are effective in identifying the ESG rating class. In [18], it is verified that financial and accounting indicators (Total Investments, Nominal Yield, Price Earning, EPS, PS, PB, NPM, Current Ratio, ROA, and ROE) can accurately predict the ESG rating class. In [4], it is highlighted that the predictive accuracy of models depends on the rating size under analysis, and that the E and S dimensions of ratings respond more effectively than the G dimension.

This study aims to explore the feasibility of predicting the ESG rating class at time $t + 1$ based on financial data and market dynamics observed at time t. It proposes a parsimonious, agile, and replicable Machine Learning (ML) model that leverage information directly or indirectly available in a company's financial statements. Unlike current rating agency algorithms, which are complex and not fully disclosed, the proposed model seeks to offer a more transparent and accessible alternative by utilizing readily available information. By focusing on few variables, the model aims to accurately estimate future rating classes, providing investors with a lightweight and adaptable decision-making tool. The literature review underscores the efficacy of ML techniques in identifying patterns and

complex relationships in data [13], thereby enabling a more precise assessment of ESG performance and the detection of trends or anomalies. This paper introduces a novel approach by utilizing classification Random Forest models instead of regression ones, which have demonstrated superior performance. Additionally, it integrates a dynamic measure of systemic risk, represented by Conditional Beta (BETA), as an independent variable.

2 Materials and Methods

In this study, companies included in the STOXX Europe 600 index during the period 2016–2021 are analyzed. For each company, daily stock prices and ESG scores, which are divided into the individual pillars of Environment (E), Social (S), and Governance (G) (EiKon), as well as annual financial statement data (EiKon), are considered. Among them: ROI, ROE, CAPEX, EBITDA, etc.

A preliminary data cleaning phase was necessary. Financial data for each company were normalized relative to its value. Additionally, the data source for ESG scores is Thomson Reuters, which expresses scores within the range [0, 100] and converts these scores into textual format that we use as labels for the classification model implemented here. The estimation of BETA values was conducted by analyzing the conditional correlations between the daily stock prices of each security and the STOXX Europe 600 index [14].

The Random Forest (RF) classification model presented in this study consists of a set of autonomous decision trees, and their estimates are aggregated using a majority voting strategy [20]. With this approach, it is possible to predict the ESG rating class of a certain company in the year following the reference year of the input data. To overcome the problem of overfitting typical of decision trees, the RF algorithm builds the forest using numerous trees generated using the Bootstrap Aggregating (Bagging) technique. Each tree in the forest is trained on a subset of the original dataset and considers only a subset of the available independent variables, generating complementary classifiers [5].

In the process of developing classification models, several predictive algorithms were implemented. Among the considered ones, the RF classifier emerged as the most performing, with more robust results compared to RF regression, K-Nearest-Neighbour (KNN), and Support Vector Machine (SVM). In Sect. 3, only the performances of the RF classifier and KNN are reported as the latter can be considered a traditional benchmark model.

To measure the predictive power of the presented models, the main metric chosen is the accuracy function, which measures, on a scale from 0 to 1, the number of correct predictions out of the total formulated. In the case of multi-class predictions like those under consideration, the function is calculated for each class. In addition, metrics such as precision, recall, and F1-score were considered [19].

3 Results and Conclusions

Models were trained on two different sets of independent variables: (i) ESG rating (including individual Pillars E, S, G), financial and accounting data, as well as BETA; (ii) financial and accounting data, as well as BETA.

Table 1. Macro-class estimation: results obtained with ESG pillars, financial statement and Conditional BETA variables.

Class	Random Forest (RF)			K-Nearest Neighbors (KNN)		
	Precision	Recall	F1-score	Precision	Recall	F1-score
A	0.84	0.81	0.83	0.76	0.79	0.77
B	0.77	0.82	0.80	0.71	0.72	0.72
C	0.79	0.65	0.71	0.67	0.52	0.59
D	0.00	0.00	0.00	0.00	0.00	0.00
Accuracy			0.80			0.73
Macro avg	0.80	0.76	0.78	0.71	0.68	0.69
Weighted avg	0.81	0.80	0.80	0.73	0.73	0.73

Table 2. Macro-class estimation: results obtained with financial statement and Conditional BETA variables.

Class	Random Forest (RF)			K-Nearest Neighbors (KNN)		
	Precision	Recall	F1-score	Precision	Recall	F1-score
A	0.66	0.59	0.62	0.53	0.55	0.54
B	0.70	0.77	0.67	0.55	0.58	0.57
C	0.00	0.00	0.00	0.33	0.17	0.23
D	0.00	0.00	0.00	0.00	0.00	0.00
Accuracy			0.62			0.53
Macro avg	0.42	0.45	0.43	0.47	0.43	0.44
Weighted avg	0.57	0.62	0.59	0.52	0.53	0.54

Models executions with the above-described sets of independent variables (i) and (ii) were aimed at identifying the permanence or migration at time $t+1$ from the observed ESG macro-class (A, B, C, D) at time t. Tables 1 and 2 describe the results obtained in estimating each macro-class from the RF classification model and the KNN control on the two sets of independent variables (i) and (ii).

The superior accuracy of the RF model execution with set (i) shows that considering the dynamics of the ESG rating and individual Pillars, it is possible to offer investors a more accurate tool for predicting the class in which the security will fall on the next year.

The high accuracy achieved with this model allowed proceeding with the execution of the RF classification algorithm to also verify the possibility of predicting the ESG sub-class (A+, A, A−, B+, B, B−, C+, C, C−, D). Regarding the RF model, precision reaches maximum levels in the A+ and B+ sub-classes. The descending trend highlighted seems to suggest a hierarchical distribution of information and values of independent variables for the sub-classes themselves. In other words, for higher ratings (macro-class A), more precise indications are obtained independently of the model. This observation seems to confirm that the intrinsic value of the ESG rating is a function of its level.

During the model building process, we observed the percentage feature importance assigned to each independent variable. This is particularly relevant because higher importance corresponds to a greater influence of the feature on the identification outcome. Our observation corroborates findings from the literature regarding the significance of the informational content of financial data [15] and underscores the importance of the systemic risk measure introduced in this study for predicting the ESG class.

Among the most influential financial indicators, Total Equity (TOTEQUITY) stands out. It provides information about a company's financial strength, signaling its ability to invest in sustainable initiatives, social and environmental responsibility, and corporate governance management. This indicator also offers informative content regarding long-term financial stability and, consequently, the company's resilience. This aspect is directly related to the informational content offered by BETA that, being a dynamic measure of systemic risk, can reflect the greater or lower ability of the company to react to market instability phases. These aspects are relevant for the ESG rating, as highlighted in [1], and [8]. ROE and EBITDA are essential indicators for sustainable investors as they offer insights into a company's financial and operational sustainability, along with its long-term profit generation capabilities [7]. On the other hand, CAPEX, reflecting the company's investments, can also mirror those in environmentally friendly technologies. In synergy with Intangibles (INT), CAPEX can reflect investments in Research and Development and, more generally, in innovation. This latter aspect emphasizes the high weight assumed by the dynamics of the "S" Pillar which considers how a company manages and engages with its personnel, the communities in which it operates, and other stakeholders from a social perspective [11].

References

1. Aevoae, G.M., Andrieş, A.M., Ongena, S., Sprincean, N.: ESG and systemic risk. Appl. Econ. **55**(27), 3085–3109 (2023)
2. Andria, J., Tollo, G., Pesenti, R.: A fuzzy evaluation of tourism sustainability. Bus. Consum. Anal. New Ideas 911–932 (2019)
3. Berg, F., Koelbel, J.F., Rigobon, R.: Aggregate confusion: the divergence of ESG ratings. Rev. Financ. **26**(6), 1315–1344 (2022)
4. Boucher, C., Le Lann, W., Matton, S., Tokpavi, S.: Backtesting ESG ratings. Technical report, Orleans Economics Laboratory/Laboratoire d'Economie d'Orleans (2022)

5. Breiman, L.: Random forests. Mach. Learn. **45**(1), 5–32 (2001). https://doi.org/10.1023/a:1010933404324
6. Cheng, B., Ioannou, I., Serafeim, G.: Corporate social responsibility and access to finance. Strateg. Manag. J. **35**(1), 1–23 (2014)
7. Christensen, D.M., Serafeim, G., Sikochi, A.: Why is corporate virtue in the eye of the beholder? The case of ESG ratings. Account. Rev. **97**(1), 147–175 (2022)
8. Ciciretti, R., Dalò, A., Dam, L.: The contributions of betas versus characteristics to the ESG premium. J. Empir. Financ. **71**, 104–124 (2023)
9. Cornell, B.: ESG preferences, risk and return. Eur. Financ. Manag. **27**(1), 12–19 (2021)
10. Del Vitto, A., Marazzina, D., Stocco, D.: ESG ratings explainability through machine learning techniques. Ann. Oper. Res. 1–30 (2023)
11. Demers, E., Hendrikse, J., Joos, P., Lev, B.: ESG did not immunize stocks during the Covid-19 crisis, but investments in intangible assets did. J. Bus. Financ. Account. **48**(3–4), 433–462 (2021)
12. Dimson, E., Marsh, P., Staunton, M.: Divergent ESG ratings. J. Portfolio Manag. **47**(1), 75–87 (2020)
13. D'Amato, V., D'Ecclesia, R., Levantesi, S.: ESG score prediction through random forest algorithm. CMS **19**(2), 347–373 (2022)
14. Engle, R.: Dynamic conditional correlation: a simple class of multivariate generalized autoregressive conditional heteroskedasticity models. J. Bus. Econ. Stat. **20**(3), 339–350 (2002)
15. García, F., González-Bueno, J., Guijarro, F., Oliver, J.: Forecasting the environmental, social, and governance rating of firms by using corporate financial performance variables: a rough set approach. Sustainability **12**(8), 3324 (2020)
16. Gonçalves, T.C., Dias, J., Barros, V.: Sustainability performance and the cost of capital. Int. J. Financ. Stud. **10**(3), 63 (2022)
17. Kolahgar, S., Schiehll, E.: Disclosing what matters, financially material ESG information and stock price informativeness. In: Financial Management Association 50th Annual Meeting (FMA 2020) (2020)
18. Lee, O., Joo, H., Choi, H., Cheon, M.: Proposing an integrated approach to analyzing ESG data via machine learning and deep learning algorithms. Sustainability **14**(14), 8745 (2022)
19. Lever, J.: Classification evaluation: it is important to understand both what a classification metric expresses and what it hides. Nat. Methods **13**(8), 603–605 (2016)
20. Liaw, A., Wiener, M., et al.: Classification and regression by random forest. R News **2**(3), 18–22 (2002)
21. Muslu, V., Mutlu, S., Radhakrishnan, S., Tsang, A.: Corporate social responsibility report narratives and analyst forecast accuracy. J. Bus. Ethics **154**, 1119–1142 (2019)
22. Serafeim, G., Yoon, A.S.: Understanding the business relevance of ESG issues. J. Financ. Reporting **7**(2), 207–212 (2022)
23. Welch, K., Yoon, A.: Do high-ability managers choose ESG projects that create shareholder value? Evidence from employee opinions. Rev. Account. Stud. 1–28 (2022)

PSO for the Sharpe Ratio in a Financial Trading System Based on Technical Analysis

Marco Corazza[1], Claudio Pizzi[1], and Andrea Marchioni[2(✉)]

[1] Ca' Foscari University of Venice, Venice, Italy
{corazza,pizzic}@unive.it
[2] University of Tuscia, Viterbo, Italy
andrea.marchioni@unitus.it

Abstract. We consider four technical indicators widely used in financial practice to determine the optimal signal aggregation, trading rule definition, and indicator setting using the Particle Swarm Optimization metaheuristic applied to an important financial fitness function, that is the Sharpe Ratio. We experiment our trading system to the Italian index FTSE MIB and to a set of financial stocks belonging to the FTSE MIB over a multi-year period for training and testing. We generally achieve superior out-of-sample performance, using a standard technical analysis system as a benchmark.

Keywords: Trading system · Particle Swarm Optimization/PSO · Sharpe Ratio

1 Introduction

This paper proposes an algorithmic trading system based on Technical Analysis (TA) indicators with optimization of the signal aggregation, the trading rule definition, and the indicator setting. The parameters of the indicators, the trading rules, and the signal weights are the inputs of the system, and the fitness function to be maximized is the well-known Sharpe Ratio over the trading period. It is a complex global optimization problem, which we address by the Particle Swarm Optimization (PSO) metaheuristic (Kennedy and Eberhart in [5]), an approximate bio-inspired numerical optimizer.

TA indicators depend on one or more parameters (mainly time windows), generally assumed equal to standard values in the financial practice. The trading rules are functions whose inputs are the indicators; these functions generate signals based on market prices.

Usually, trading systems jointly consider a plurality of indicators, whose trading signals are aggregated through weighting, permitting a more informed decision-making (e.g., Corazza et al. in [2]). We consider and aggregate four popular standard indicators, that are the Exponential Moving Average (MA), the

M. Corazza et al. (Eds.): MAF 2024, *Mathematical and Statistical Methods for Actuarial Sciences and Finance*, pp. 93–98, 2024.
https://doi.org/10.1007/978-3-031-64273-9_16

Relative Strength Index (RSI), the Moving Average Convergence/Divergence (MACD), and the Bollinger Bands (BB).

The existing studies have a main limitation: they only optimize a single category of parameters at a time, whether it be the indicators setting, or the trading rules definition, or the signal aggregation. For the first time in literature, Corazza et al. in [3] proposed to simultaneously optimize the three categories of parameters of the trading system. In that work, the fitness function to be maximized was the net capital at the end of the trading period, regardless of the risk level of the trading strategy. In the present work, we apply that proposal of simultaneous optimization to a different fitness function, that is the Sharpe Ratio.

The influence diagram of our trading system, from parameters to the fitness function, can be depicted through three intermediate levels: indicators, signals, and the aggregated signal. More precisely, the parameters of the indicators influence the computation of the indicators. Each indicator generates its own signal through its trading rule based on the trading rules parameters. The signals are aggregated into an overall signal using the weights of the indicators. Finally, the Sharpe Ratio – which depends on buying, selling, and holding decisions derived from the aggregated signal – is maximized via PSO by appropriately optimizing the three categories of parameters.

The remainder of this paper is organized as follows. The next section is devoted to describe the methodology used in this work. Section 3 presents the out-of-sample results of our optimized trading system. Some final remarks conclude the paper.

2 Methodology

2.1 Parametrization of the Trading System

Our purpose is to optimize the parametrization of indicators, trading rules, and signal aggregation for a total of 23 parameters. Table 1 collects the parameters and describes their main features. For a description of the indicators, trading rules, and signal aggregation, the reader is referred to Corazza et al. in [3].

The trading system with the standard values of the parameters (reported in the last column of Table 1) serves as a benchmark for our optimized trading system.

2.2 Constrained Optimization of the Sharpe Ratio

As performance measures, we consider the fitness function Sharpe Ratio over the trading period T, which is a risk-adjusted performance ratio. For its computation, we need to determine the net daily return $e(t)$, depending on: the strategy $s(\cdot)$; the stock price variation $P(t)/P(t-1)$; and the trading fee δ in the case of a strategy change:

$$e(t) = s(t-1)\ln\left(P(t)/P(t-1)\right) - \delta\left|s(t) - s(t-1)\right|,\ t = 1, \ldots, T. \qquad (1)$$

Table 1. Parameters of the standard trading system

Parameter	Symbol	Indicator	Area	Standard value
Time window for computing MA	w_{ma}	MA	Indicator	12
Minimum period of validity of MA rule	d_{ma}	MA	Rule	1
Time window for computing RSI	w_{rsi}	RSI	Indicator	26
Threshold of RSI for entering in buy	t_{rsi}^{l,en_b}	RSI	Rule	30
Threshold of RSI for entering in sell	t_{rsi}^{h,en_s}	RSI	Rule	70
Threshold of RSI for exiting from buy	t_{rsi}^{h,ex_b}	RSI	Rule	70
Threshold of RSI for exiting from sell	t_{rsi}^{l,ex_s}	RSI	Rule	30
Short time window for computing MACD line	w_{macd}^{short}	MACD	Indicator	12
Long time window for computing MACD line	w_{macd}^{long}	MACD	Indicator	26
Time window for computing MACD signal line	w_{macd}^{signal}	MACD	Indicator	9
Minimum period of validity of MACD rule	d_{macd}	MACD	Rule	1
Time window for computing the moving average for BB	w_{bb}^{ma}	BB	Indicator	26
Time window for computing the standard deviation for BB	w_{bb}^{std}	BB	Indicator	26
(Positive) number of standard deviation for the upper BB	t_{bb}^{u}	BB	Indicator	2
(Positive) number of standard deviation for the lower BB	t_{bb}^{l}	BB	Indicator	2
Number of standard deviation for exiting from sell	t_{bb}^{c,ex_s}	BB	Rule	0
Number of standard deviation for exiting from buy	t_{bb}^{c,ex_b}	BB	Rule	0
Weight of MA signal	θ_{ma}	MA	Signal aggregation	0.25
Weight of RSI signal	θ_{rsi}	RSI	Signal aggregation	0.25
Weight of MACD signal	θ_{macd}	MACD	Signal aggregation	0.25
Weight of BB signal	θ_{bb}	BB	Signal aggregation	0.25
Threshold for the aggregated signal for entering in buy	t_{as}^{b}	–	Signal aggregation	+1/3
Threshold for the aggregated signal for entering in sell	t_{as}^{s}	–	Signal aggregation	−1/3

The fitness function to be maximized is the Sharpe Ratio at the end of the trading period, $\rho = SR(T)$, under several constraints related to the parameters of the system. The global optimization problems can be rewritten as the following constrained maximization of ρ:

$$\max_{\chi \in X} \rho$$
$$\text{s.t.} \begin{cases} w_{ma},\, d_{ma},\, w_{rsi},\, w_{macd}^{short},\, w_{macd}^{long},\, w_{macd}^{signal},\, d_{macd},\, w_{bb}^{ma},\, w_{bb}^{std} \in N^{+} \\ t_{rsi}^{l,en_b} \geq 0,\, t_{rsi}^{h,en_s} \leq 100,\, t_{rsi}^{l,en_b} \leq t_{rsi}^{l,ex_s} \leq t_{rsi}^{h,en_s},\, t_{rsi}^{l,en_b} \leq t_{rsi}^{h,ex_b} \leq t_{rsi}^{h,en_s} \\ w_{macd}^{long} > w_{macd}^{short} \\ w_{bb}^{std} \geq 2,\, t_{bb}^{u} \geq 0,\, t_{bb}^{l} \geq 0,\, -t_{bb}^{l} \leq t_{bb}^{c,ex_s} \leq t_{bb}^{u},\, -t_{bb}^{l} \leq t_{bb}^{c,ex_b} \leq t_{bb}^{u} \\ \theta_{ma} \geq 0,\, \theta_{rsi} \geq 0,\, \theta_{macd} \geq 0,\, \theta_{bb} \geq 0,\, \theta_{ma} + \theta_{rsi} + \theta_{macd} + \theta_{bb} = 1 \\ t_{as}^{b} > t_{as}^{s} \end{cases} , \tag{2}$$

where X represents the parameter space.

2.3 Particle Swarm Optimization

The constrained global optimization problem (2) is formulated in terms of mixed-integer variables and it is nonlinear and nondifferentiable. Due to these complexities, exact solution algorithms are still sought in literature and we need to use an approximate solution method. Therefore, we consider to apply a metaheuristic and we choose PSO for its exploration and exploitation capabilities (Kennedy and Eberhart in [5], Wakasa and al in [7]).

Standard PSO is a solver for global unconstrained optimization problems, whereas our optimization problem is a global constrained mixed-integer one.

Consequently, we appropriately adapt the standard PSO for managing these specificities.

For dealing with integer variables, we follow a widespread approach in literature, that is the truncation method proposed in Parsopoulos and Vrahatis in [6].

For dealing with the other constraints, we reformulate our optimization problem as an unconstrained one using the exact penalty method described in Fletcher [4] and applied in the financial context in Corazza et al. in [1]. This method permits a correspondence between the optimizer of the original constrained problem and the unconstrained penalized one. The reformulated unconstrained version of the optimization problem is the maximization of the following function $\hat{\rho}$ with penalty parameter ϵ:

$$\begin{aligned}\max_{X\in X} \hat{\rho} = \rho - \frac{1}{\epsilon}\Big[&\max(0, -t_{rsi}^{l,en_b}) + \max(0, t_{rsi}^{h,en_s} - 100) + \max(0, t_{rsi}^{l,en_b} - t_{rsi}^{l,ex_s}) \\ &+ \max(0, t_{rsi}^{l,ex_s} - t_{rsi}^{h,en_s}) + \max(0, t_{rsi}^{l,en_b} - t_{rsi}^{h,ex_b}) + \max(0, t_{rsi}^{h,ex_b} - t_{rsi}^{h,en_s}) \\ &+ \max(0, -w_{macd}^{long} + w_{macd}^{short}) + \max(0, -w_{bb}^{std} + 2) + \max(0, -t_{bb}^{u}) \\ &+ \max(0, -t_{bb}^{l}) + \max(0, -t_{bb}^{l} - t_{bb}^{c,ex_s}) + \max(0, t_{bb}^{c,ex_s} - t_{bb}^{u}) \\ &+ \max(0, -t_{bb}^{l} - t_{bb}^{c,ex_b}) + \max(0, t_{bb}^{c,ex_b} - t_{bb}^{u}) + \max(0, -\theta_{ma}) \\ &+ \max(0, -\theta_{rsi}) + \max(0, -\theta_{macd}) + \max(0, -\theta_{bb}) \\ &+ |\theta_{ma} + \theta_{rsi} + \theta_{macd} + \theta_{bb} - 1| + \max(0, -t_{as}^{b} + t_{as}^{s})\Big];\end{aligned} \tag{3}$$

in our study we use $\epsilon = 10^{-2}$.

3 Applications

Our applications consider the closing prices of the FTSE MIB index and a set of selected stocks belonging to the FTSE MIB at the date of May 31, 2022, and traded on the market starting before January 2, 2007. We select five sectors, that are highly representative of the Italian economy. As results, we apply our methodology to the following stocks: Assicurazioni Generali S.p.A. (sector of insurance); Atlantia S.p.A. (sector of industrial products and services); Enel S.p.A. (sector of public services); Eni S.p.A. (sector of oil and natural gas); Intesa Sanpaolo S.p.A. (sector of banks).

We have conducted three different out-of-sample experiments. First, the trading period is divided into two subperiods, that is a training period and an out-of-sample testing one, but in each experiment a different length for the testing period is considered: 1 stock-month, 2 stock-months, and 3 stock-months, respectively; in Table 2, we provide the start and end dates for each in-sample and out-of-sample subperiod. Then, the trading system is optimized using the metaheuristic PSO over the training subperiod. Finally, the optimized trading system is applied to the out-of-sample testing subperiod.

The analysis is repeated 100 times, and we calculate the average value of the Sharpe Ratio as weel as other quantities of interest. In doing so, we confer a degree of statistical significance to the results, at least to some extent.

Table 2. Start and end dates for each in-sample and out-of-sample subperiod

Experiment	In sample subperiod	Out-of-sample subperiod
1	January 2, 2007 to April 29, 2022	May 2, 2022 to May 31, 2022
2	January 2, 2007 to March 31, 2022	April 1, 2022 to May 31, 2022
3	January 2, 2007 to February 28, 2022	March 1, 2022 to May 31, 2022

The results are collected in Tables 3. As for the structure of the tables, the first column indicates the length of the out-of-sample period; the third and fourth columns report the Sharpe Ratios of the standard system (which obviously are independent from the repetitions) and the mean Sharpe Ratios of the optimized system; the fifth and sixth columns present the annualized returns of the standard system and the mean annualized returns of the optimized system; finally, the last column reports the mean percentage of days in which the equity line of the optimized system is above the equity line of the standard system.

Table 3. Average out-of-sample performance of the optimized trading system over T month(s) for 100 repetitions for each asset

T	Stock	$SR(T)_{st}$	Mean $SR(T)_{PSO}$	$\bar{e}_{st}$	Mean $\bar{e}_{PSO}$	Mean % >
1	Assicurazioni Generali	−2.94	−0.28	−24.64	−6.46	48.76
1	Atlantia	−4.01	−1.87	−22.72	−8.08	71.33
1	Enel	−0.85	−0.27	−17.30	−6.38	41.24
1	Eni	−3.82	0.89	−61.99	7.09	81.90
1	Intesa Sanpaolo	6.86	2.40	111.09	46.64	23.19
1	FTSE MIB	0.35	0.35	2.87	3.29	47.90
2	Assicurazioni Generali	2.30	0.76	22.99	9.20	29.00
2	Atlantia	−3.00	0.47	−12.19	36.23	41.31
2	Enel	−1.66	−0.76	−26.51	−11.13	59.12
2	Eni	−1.89	−0.62	−34.10	−8.83	45.74
2	Intesa Sanpaolo	2.75	0.45	33.45	9.48	37.88
2	FTSE MIB	2.21	−0.64	15.59	−6.80	16.36
3	Assicurazioni Generali	6.15	1.59	74.73	28.47	18.02
3	Atlantia	1.90	1.80	22.74	46.99	32.44
3	Enel	−1.35	−1.03	−27.88	−20.35	51.95
3	Eni	−2.30	−0.72	−46.30	−15.59	72.22
3	Intesa Sanpaolo	2.54	0.74	41.39	16.83	39.11
3	FTSE MIB	1.34	0.82	9.01	12.53	59.71

It is worth highlighting that:

- When considering the 1-month long out-of-sample period, the optimized system draws or wins 5 times out of 6 compared to the standard system; the performances degrade as the length of the out-of-sample period increases. Consequently, our trading system would need to be re-optimized with appropriate frequency;
- Whatever the length of the out-of-sample period, when the optimized system loses against the standard system, the mean values of the annualized returns and of the Sharpe Ratios obtained by the optimized system are always positive, except one case. This could indicate that the PSO works well in the optimization phase also in these cases, but likely paying for the choice of using a unique hyper-parametrization for all stocks;
- In some cases where the optimized system underperforms the standard system, the mean annualized returns of the optimized system are higher than those of the standard system; therefore, the optimized trading system allows an increase in the annualized return, although less than proportional to the increase in the riskiness of the strategy.

4 Concluding Remarks

Our optimized trading system generally leads to superior performance over a standard TA-based trading system for a set of financial stocks belonging to the FTSE MIB on a multi-year horizon for training and testing.

Future researches will focus on: the use of a multi-objective fitness function and multi-objective PSO for constructing an efficient risk-return frontier; the application of the capability of PSO for automatically selecting the indicators.

References

1. Corazza, M., Fasano, G., Gusso, R.: Particle Swarm Optimization with non-smooth penalty reformulation, for a complex portfolio selection problem. Appl. Math. Comput. **224**, 611–624 (2013)
2. Corazza, M., Parpinel, F., Pizzi, C.: Trading system mixed-integer optimization by PSO. In: Corazza, M., Gilli, M., Perna, C., Pizzi, C., Sibillo, M. (eds.) Mathematical and Statistical Methods for Actuarial Sciences and Finance, pp. 161–167. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-78965-7_24
3. Corazza, M., Pizzi, C., Marchioni, A.: A financial trading system with optimized indicator setting, trading rule definition, and signal aggregation through Particle Swarm Optimization. Comput. Manag. Sci. **21**, article number 26 (2024)
4. Fletcher, R.: Practical Methods of Optimization. Wiley, Hoboken (2000)
5. Kennedy, J., Eberhart, R.: Particle swarm optimization. In: Proceedings of ICNN 1995 - International Conference on Neural Networks, vol. 4, pp. 1942–1948 (1995)
6. Parsopoulos, K.E., Vrahatis, M.N.: Recent approaches to global optimization problems through particle swarm optimization. Nat. Comput. **1**(2), 235–306 (2002)
7. Wakasa, Y., Tanaka, K., Nishimura, Y.: Control-theoretic analysis of exploitation and exploration of the PSO algorithm. In: 2010 IEEE International Symposium on Computer-Aided Control System Design (2010)

Actuarial Gains in Life Annuities Due to Declining Health: LTC

J. Iñaki De La Peña[1,2(✉)] and Asier Garayeta[1,2]

[1] Faculty of Economics and Business, University of the Basque Country (UPV/EHU), Agirre Lehendakaria Avenue, 83, 48015 Bilbao, Spain
jinaki.delapena@ehu.es
[2] Consolidated Research Group EJ/GV: IT 1523-22, Bilbao, Spain

Abstract. Life annuities are actuarial products based on technical assumptions, such as mortality. The degeneration of the human body leads, in addition to generating long-term care, to a higher mortality of the dependent than that of the general and insured population. Therefore, the period for receiving this benefit would be shorter. The aim of this paper is to determine the economic impact of the change in the beneficiary's status when receiving this life annuity. It should be stressed that, in the life annuity, the biometric risk is borne by the insurer and that a lower payment expectancy due to the pension beneficiary's change to dependent status entails a benefit, as this gain is not distributed to the beneficiary. A surplus is created by paying out the same benefit. Thus, the use of an appropriate mortality assumption results in a reduction of the mathematical payout provision, which frees up capital and results in a lower solvency capital requirement.

Keywords: Actuarial fairness · life annuities · Long Term Care

1 Introduction

The aim of this paper is to determine the impact of a life annuity when the health status changes to severely dependent/highly dependent with no return. The annuity contract does not include mortality other than that of the general population, however, the change in mortality due to a non-returning health condition may lead to a transformation of the annuity's purpose into a long-term care (LTC) benefit. If the same benefit is paid, the use of an appropriate mortality assumption will reduce the value of the annuity, leading to the release of capital and a lower solvency capital requirement.

This paper makes a breakthrough: it determines the procedure for calculating the surplus for not differentiating mortality in the life annuity. The main contribution lies in quantifying this surplus. On the one hand, as we do not know when the beneficiary becomes severely dependent, this information should be provided by the beneficiary himself/herself. On the other hand, the total effect on the insurer will depend on the demographic composition of its life annuity portfolio.

The next section deals with measurement. The third section includes the methodology for valuation. The next section provides a representative application of the Spanish

M. Corazza et al. (Eds.): MAF 2024, *Mathematical and Statistical Methods for Actuarial Sciences and Finance*, pp. 99–105, 2024.
https://doi.org/10.1007/978-3-031-64273-9_17

market to illustrate this. The discussion and implications of the model are included, and the paper ends with the relevant conclusions.

2 Measuring Economic Impact

The economic effect can be captured with an Actuarial Gain/Loss (AGL) analysis. *AGL* calculates the economic value of the differences between actuarial assumptions and reality, generating a deficit or surplus. It is a common actuarial technique that can be used to analyse both financial assets and biometric liabilities separately or jointly [7] - Fig. 1.

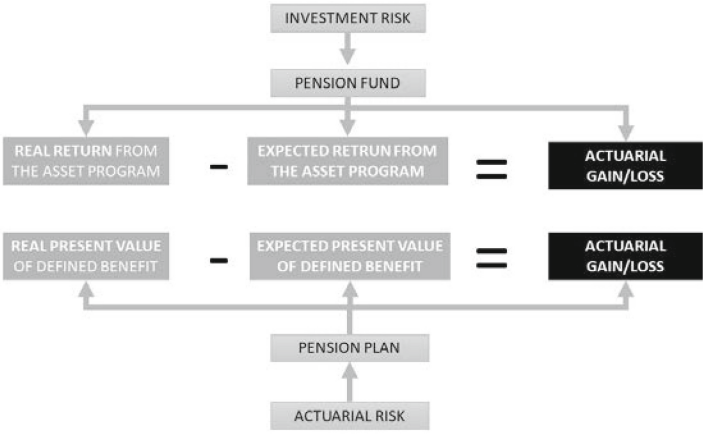


Fig. 1. Main components of Actuarial Gain/Loss. Source: Own elaboration.

The literature analyses actual investment performance on the interest rate assumption of the valuation [1, 3, 4, 14, 21]. It also considers the effect of discretionary choice of actuarial assumptions and their gradual appropriateness. Its simplicity and analytical power have brought it to the forefront of pension information systems [10], and in the amortisation of pension deficits [13, 17, 19].

3 Methodology: Actuarial Gain/Loss

We take as a starting point the actuarial model proposed by [5, 6, 11, 12] contemplating high degrees of dependency without return [4]. The beneficiary's health status changes at age x as he/she becomes severely or highly dependent. Then, under a new survival function and with the initial benefit, the actuarial value of the benefit actually received ($VajR_x$) is given by the expression (1),

$$VajR_x = \int_x^w b_t \cdot e^{-\int_x^{w\,d} \mu_t^m dt} \cdot e^{-\int_x^w \delta(t)dt} \cdot dt \tag{1}$$

Being,

b_t: Benefit function.
${}^{d}\mu_t^m$: Instantaneous mortality rate of a severely or highly dependent person at the t-th instant.
$e^{-\int_x^w {}^{d}\mu_t^m dt}$: Probability of survival of an individual of age x as a function of the instantaneous mortality rate of a dependent person.
$\delta(t)$: instantaneous interest rate.
$e^{-\int_x^w \delta(t)dt}$: Financial discount function up to age x, through the instantaneous interest rate.

The *AGL* is therefore defined as the difference between the actuarial value of the benefit actually received and the benefit expected to receive under the initial assumptions:

$$AGL_x = VajR_{x+1} - E(Vaj)_{x+1} \tag{2}$$

where,

AGL_x: Actuarial Gain/Loss generated at age x by the change in the beneficiary's health status.
$E(Vaj)_{x+1}$: Present value of the benefit expected to be received under the initial health status at age $x + 1$.

It can be determined for each beneficiary in each of the future years in case he/she reaches future ages and becomes a dependent person.

$$VaAGL_x = \sum_{h=x}^{w} AGL_h \cdot {}^{r}q_h^{(d)} \cdot {}_{h-x}^{r}p_x^{(r)} \cdot v^{h-x} \tag{3}$$

$VaAGL_x$: Actuarial value of the *AGL* due to the change in the beneficiary's health, calculated at age x.
${}^{r}q_h^{(d)}$: Probability that a retirement beneficiary at age h will be severely or severely dependent at that age, being exposed to another cause of exit (mortality).
${}_{h-x}^{r}p_x^{(r)}$: Probability that a retirement beneficiary of age x will reach age h without death or change of health status.
v^{h-x}: Financial discount factor from age h to age x.

In the case of the annuity, it is the pension beneficiary who is aware of his or her health status, and if he or she changes to a dependent, only he or she can inform the insurer. Therefore, the incidence of the dependent's mortality is instantaneous from the moment of notification.

4 Dependent Mortality Versus Overall Mortality: Discussion

[12] establishes the life expectancy of an individual in the most severe stages of dependency. [8] start from a general mortality and propose additive displacement on the instantaneous mortality rate. However, [16] indicate that dependents will have an over-mortality that can be expressed by a multiplicative correction - θ - on the mortality probability of the general population:

$${}^{d}q_x^m = \theta \cdot q_x^m \tag{4}$$

This correction can be variable at each age, although [16] indicated that a fixed correction adjusts the mortality of older dependents better than other types of approximations. However, it overestimates mortality at lower ages and underestimates at higher ages. Therefore, it is better to perform an additive adjustment (ε) considering age as an independent variable in a functional form [18].

$$ {}^{d}q_x^m = q_x^m + \varepsilon \quad \text{where} \quad \varepsilon = f(x) \tag{5} $$

As a result, mortality rates are lower at younger ages and increase with the degree of dependency [15]. [20] determined the probability of death for severe dependency; they used general mortality tables and adjusted them to the HID 98–01 statistics for France:

$$ {}^{d}q_x^m = \begin{cases} q_x^m + \frac{\delta}{1+\gamma^{x_i - x}} & \forall x_i < 95 \\ q_x^m \cdot (1+\beta) + \frac{\delta}{1+\gamma^{x_i - x}} & \forall x_i \geq 95 \end{cases} \tag{6} $$

δ: Maximum value to be incorporated as a function of age at which it converges asymptotically.
γ: Slope factor.
x_i: Age of inflection at which the curve changes shape from convex to concave.
β: Multiplicative factor on overall mortality (Table 1).

Table 1. Overmortality factors for severe dependency in Spain. Source: [20].

Factors	Men	Women
δ	0.245	0.165
γ	1.135	1.09
x_i	62.50	58.61
β	0.1142	0.0962

Mortality rates for disabled persons ($^{d}q_x^m$) and invalids ($^{i}q_x^m$) converge to the general population (q_x^m) as age increases (Fig. 2).

As a final result, for both men and women, there is an actuarial gain, i.e. a smaller capital sum needed to guarantee the benefit in the event of a change in health status. The surplus is greater the younger the beneficiary is. The main implication is that, to pay out the same benefit, there is money left over, so a surplus will be generated as it is not distributed to the life annuity beneficiary.

The surplus decreases as the beneficiary becomes severely dependent at older ages, as can be seen in Fig. 3, taking minimum values from the age of 100 onwards in the case of retirement beneficiaries. As for the disability pension, it takes minimum values for men, with a deficit per euro of benefit for women who change their health status to severe dependency from the age of 90.

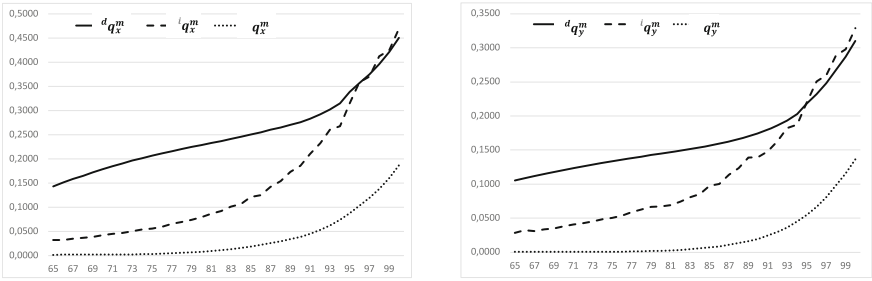


Fig. 2. Mortality differential after retirement age at 65. (**x**) Men; (**y**) Women. Source: Own elaboration. Databases: [2, 9]; Mortality Tables with factors from [20].

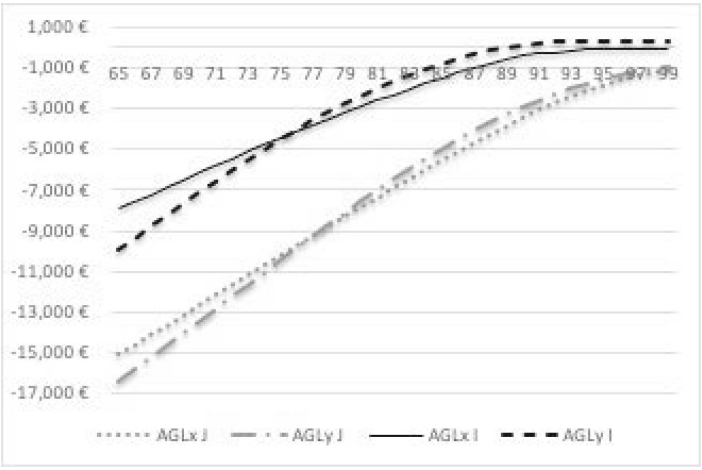


Fig. 3. Evolution of the surplus by change of mortality tables and by euro benefit, according to age and origin (Retirement -J- or Invalidity -I-). Source: Own work.

5 Conclusion

Logically, it is in the insurer's interest that the beneficiary informs the insurer of his or her change in health. This new is compulsory, as the risk status of the insured person changes. However, the beneficiary himself/herself does not perceive a change of benefit after this information, but it is the insurer itself that takes advantage. There is no incentive to update the death risk status information from the beneficiary point of view.

It is necessary to establish a mutually beneficial incentive. So, part of the gain can pass on to the beneficiary, with the mandatory benefit increase to help with LTC. There is an incentive for the beneficiary and partially the insurer also reduces the capital at risk.

References

1. Ashtana, S.: Determinants of funding strategies and actuarial choices for defined-benefit pension plans. Contemp. Account. Res. **16**(1), 39–74 (1999)
2. BAREMO: Law 35/2015, of 22 September, on the reform of the system for the valuation of damages caused to persons in traffic accidents. BOE-A-2015-10197 (2015)
3. Brown, S.: The Impact of Pension Assumptions on Firm Value. Working Paper, Emory University. Atlanta (2006)
4. De la Peña, J.I., Martín, I.: The problem of the surplus due to the dependence in occupational pension schemes. Anales del Instituto de Actuarios Españoles **28**, 30–65 (2022)
5. De la Peña, J.I., Fernández-Ramos, M.C., Garayeta, A., Martín, I.: Transforming private pensions: an actuarial model to face Long-Term Costs. Mathematics **10**(1082), 1–17 (2022)
6. De La Peña, J.I., Fernández-Ramos, M.C., Peña-Miguel, N.: Long Term Care pension benefits coverage via conversion factor based on different mortality rates: more money as age goes on. Interscience J. **43**(1), 9–16 (2018)
7. De La Peña, J.I.: Planes de Previsión Social. Ed. Pirámide, Madrid, Spain (2000)
8. Dickson, D.C.M., Hardy, M.R., Waters, H.R.: Actuarial Mathematics for Life Contingent Risks (International Series on Actuarial Science). 3rd edn. Cambridge University Press, Cambridge (2020)
9. DGSFP - Directorate General of Insurance and Pension Funds-. Resolution of 17 December 2020, of the Directorate General of Insurance and Pension Funds, regarding the mortality and survival tables to be used by insurance and reinsurance entities, and approving the technical guide regarding the supervisory criteria in relation to biometric tables, and on certain recommendations to promote the development of sectoral biometric statistics. BOE-A-2020-17154 (2020)
10. Elkin, J.M.: A method of allocating actuarial gain and losses in a pension fund. In: The Proceedings of Conference of Actuaries in Public Practice, vol. 7, pp. 192–198 (1958)
11. 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
12. Fernández-Ramos, M.C.: Pragmatic solutions in the private sector for the coverage of dependency in Spain. Ph.D. thesis. Universidad del País Vasco, Bilbao Spain. Date: 14/07/2015 (2015)
13. Glaum, M., Keller, T., Street, D.: Discretionary accounting choices: the case of IAS 19 pension accounting. Account. Bus. Res. **48**, 139–170 (2018)
14. Heo, K., Pae, J.: Pension funding regulations and actuarial gains and losses. Aust. Account. Rev. **96**(31), 35–50 (2021)
15. Leung, E.: Projecting the Needs and Costs of Long-Term Care in Australia. Research Paper 110, Centre for Actuarial Studies, University of Melbourne (2003)
16. Macdonald, A., Pritchard, D.: Genetics, Alzheimer's and long-term care insurance. North Am. Actuarial J. **5**(2), 54–78 (2001)
17. Pinto, I., Morais, A.: Pension plans assumptions: the case of discount rate. Account. Res. J. **32**(1), 36–49 (2019)
18. Rickayzen, B.D.: A multi-state model of disability for the United Kingdom: implications for future need for long-term care for the elderly. Br. Actuar. J. **8**, 341–393 (2002)
19. Sahputra, J.H., Hidayat, T.: Motivation on accounting choice of actuarial gain (loss). J. Econ. Bus. Accountancy Ventura **17**(3), 417–428 (2014)

20. Sánchez, E., López, J.M., de Paz, S.: The correction of dependent mortality rates: an application to the Spanish case. Anales del Instituto de Actuarios Españoles **13**, 135–151 (2008)
21. Stadler, C.: Pension accounting choice in Germany: pension discount rate and actuarial gains and losses (2010). http://ssrn.com/abstract=1586117. Accessed 15 Dec 2023

Solvency and Sustainability: Evidence from the Insurance Industry

Rita D'Ecclesia, Alessandro D'Orazio(✉), Susanna Levantesi, and Kevyn Stefanelli

Sapienza University of Rome, Rome, Italy
alessandro.dorazio@uniroma1.it

Abstract. Conducting a financially sustainable business may collide with the constraints imposed by the necessity of the alignment with the Sustainable Development Goals. On the one hand, regulators demand companies to satisfy solvency requirements to operate in the insurance activity. On the other hand, the global transition to a more sustainable economy pushes companies to operate a severe business transformation. We investigate the relationship between the Solvency Ratio and Environmental, Social, and Governance score (ESG) score using an individual fixed effect regression model, including insurance-specific control variables. Our results indicate that an insurance company's ESG commitment increases its solvency level.

Keywords: Solvency · Environmental · social and governance · Insurance

1 Introduction

Sustainability is a primary issue in the agenda of governments, firms, and society. The insurance industry, which plays a fundamental role in the global economy, is deeply related to sustainability [1]. Insurance companies are prominent national and multinational institutional investors involved in sophisticated risk management processes and have a strategic function in helping governments achieve Sustainable Development Goals (SDGs) (for example, incorporating environmental factors into the insurance coverage [5] or dealing with the social sustainability [6]). Although implementing sustainability involves risks and opportunities both on the asset and liability side and from a corporate perspective [4], an increasing number of insurance companies have undertaken a journey to incorporate sustainability into their businesses.

Based on the report of [5], insurance supervisors and regulators are pursuing to understand how sustainable challenges connect to their institutional purposes, first among all, the solvency of insurance companies, which has essential implications for their financial stability. To understand thoroughly the relationship between the solvency of the insurance companies and their environmental, social,

M. Corazza et al. (Eds.): MAF 2024, *Mathematical and Statistical Methods for Actuarial Sciences and Finance*, pp. 106–111, 2024.
https://doi.org/10.1007/978-3-031-64273-9_18

and governance (ESG) rating, we conduct a panel regression analysis to study the impact of ESG on the solvency level of insurance companies.

As observed by [2], the empirical evidence on the effect of ESG on insurance companies is limited in the existing literature. The paper that is more relevant to ours is from [3]. It focused on the relationship between ESG awareness and financial performance, solvency, and size of insurance companies. The authors apply a fractional regression model between the ESG scores and size, profitability, and solvency of large US insurance companies. They find that the most influential variables in undertaking ESG policies are the company's profitability and size, but they also discover a significant association with the solvency ratio. Differently from the paper of [3], we conduct a panel regression model with the solvency ratio as the target variable and ESG score as the independent variable. A list of control variables is also included in the model.

2 The Model

In this paper, we employ a fixed effects (FE) panel regression to assess whether financially sound insurance firms are also aligned with the SDGs. We evaluate the relationship between the Solvency Ratio (SR), chosen as an insurance sustainable indicator, and various business, insurance-specific, and SDGs indicators: (i) the ESG Score (ESG), ranging from 0 to 100, where 100 indicates the most SDGs aligned companies, (ii) Loss Ratio (LR), implied losses over net premium earned (iii), Return on Assets (ROA), net income over total assets, (iv) Market Capitalization (MC), expressed in logarithm, (v) Reserve Ratio (RR), total insurance reserves divided by the total assets. By selecting these variables, we control the relationship between solvency and sustainability for a company's profitability, dimension, and insurance risk level.

The underlying assumptions for using this model are as follows. Focusing on the fiscal year of analysis, we mitigate potential biases arising from inconsistencies in ESG scores assigned to companies across different years. Additionally, we address a potential endogeneity concern derivating from variations in companies' ESG commitment over the years, which may lead to a correlation between the dependent variable and regressors[1]. The model is presented in Eq. 1.

$$Y = X\beta + \alpha_i + \varepsilon, \tag{1}$$

where Y is an $N \times T$ matrix containing observations of the dependent variables, where N is the number of companies and T is the number of sample years. The matrix X is $N \times (K \times T)$, representing observations related to the K regressors. The vector β consists of K coefficients estimated through maximum-likelihood estimation (MLE). The vector α_i contains the MLE fixed-effects parameters of dimension N, and ε is an $N \times T$ matrix containing the error terms.

[1] We conduct statistical tests to assess the validity of these assumptions.

3 Numerical Application

We collect annual data from Refinitiv choosing companies classified as "Insurance" according to the Refinitiv business sector classification (TRBC) from 2013 to 2022. We operate a data-cleaning process removing firms characterized by too many missing values. We obtain a final sample of 723 observations relative to 104 worldwide companies divided into Life (23%), Property and Casualty (38%), and Others (39%) insurers. Around 80% of the insurance companies included in our analysis are based in North America or Europe, but our sample also contains companies from every continent, as reported in Fig. 1. We retrieve ESG data only for 34% of the companies, indicating the lack of popularity in the practice of ESG disclosing within the insurance sector. Also, we do not observe a consistent shift in providing ESG information throughout the years.

We run a panel regression as in Eq. (1) to discern a relationship between financial stability and ESG practice. Table 1 contains the descriptive statistics of the variable used in this analysis. The sample exhibits missing values and wide ranges in the variables. The SR distribution shows fat tails, while the interquartile range is thin. Differently, the ESG values are highly concentrated around the average value and almost symmetrical. The box plots reported in Fig. 2 show that the SR remains almost constant over the years but presents consistent outliers. Even though ESG assumes similar values in the timeframe analyzed, we denote a decrease in the average value comparing 2012 and 2022 data. Given the vast heterogeneity of our data, we use standardized variables to account for the different variables' scales as well as to reduce the influence of outliers. By doing so, we improve both model's interpretability and fit.

The results of the estimation are as follows. The panel regression explains a high portion of SR's total variability (79%), denoting the fact that these variables are relevant to explain its time-varying dynamics. Among the control variables, the linear relationship between SR and both ROA and RR is positive and relevant since their parameter estimates are statistically significant (p-values <0.01 and <0.05, respectively). Thus, increasing the proportion of net income or/and total insurance reserves on total assets improves companies' solvency level, on average. Results indicate a positive and significant relationship between SR and ESG (p-value <0.01), suggesting that an increase in the companies' SDGs alignment determines a decrease in the solvency risk of a company, ceteris paribus (Table 2).

Table 1. Descriptive statistics of the variables used in the analysis, years 2013–2022.

	SR	ESG	LR	ROA	MC	RR
Min	−72.03	5.70	−73.70	−33.82	15.46	0.00
Q1	0.98	34.81	58.50	0.67	21.64	0.48
Q2	2.58	44.58	64.40	1.70	22.69	0.59
Mean	3.58	48.12	61.43	1.97	22.75	0.56
Q3	5.93	61.23	69.65	3.92	24.02	0.67
Max	106.35	94.94	200.10	21.05	28.21	0.90
Skew	0.42	0.40	−1.64	−2.36	0.03	−0.85
Kurt	37.38	2.50	15.58	19.23	3.56	3.72

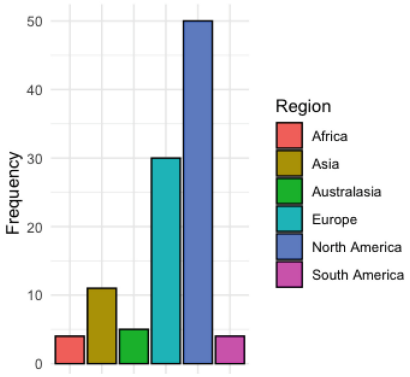


Fig. 1. Barplot of the geographical distribution of the firms' HQ. Data are referred to the final sample of 104 companies.

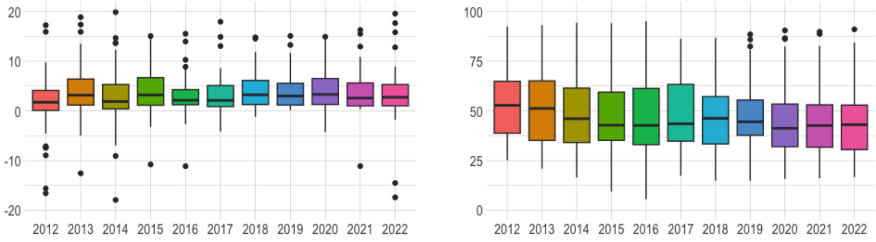


Fig. 2. Box plots of the Solvency Ratio (left panel), and ESG score (right panel), years 2013–2022. Data are referred to the final sample of 104 companies.

Table 2. Model 1 estimates (std errors in brackets) fitted on the final sample of 104 companies. Variables are standardized to enhance the robustness of the results.

Variable	SR
ESG	0.283***
	(0.065)
LR	−0.011
	(0.069)
ROA	0.834***
	(0.053)
MC	0.038
	(0.052)
RR	0.150**
	(0.045)
Constant	0.000
	(0.020)
Obs	723
Companies	104
R-squared	0.79

Note: $^{*}p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$

4 Concluding Remarks

We are interested in exploring the role of firms' ESG commitment in the insurance sectors and its impact on business performance. In particular, we assess whether the solvency requirements are in contrast with the demand for a more sustainable economy.

To the best of our knowledge, this is one of the first works that investigate this topic as researchers have mostly focused on the impact of ESG on the financial sector considered as a whole. Focusing on the sole insurance sector, we contribute to the existing literature by taking into account insurance-specific features.

In this paper, we measure the impact of the ESG score, used as a proxy for the alignment of a company to the Sustainable Development Goals (SDGs), on the solvency level of the worldwide insurance sectors, measured through the Solvency Ratio (SR). We denote that insurance companies often lack adequate disclosure on ESG themes, which implies difficulties retrieving large and reliable datasets.

We use a fixed effects panel regression model that considers the time-constant features of insurance companies and accounts for the time-varying consistency of profitability, insurance-specific, and sustainable variables used in the analysis. The model we chose uses data from 104 worldwide insurance companies observed in the years 2012–2022. Our model explains a large portion of the SR variability and finds a positive and significant relationship with the ESG score. Moreover,

both Return on Assets (ROA) and Reserves Ratio (RR) positively contribute to enhancing companies' solvency levels.

Our results suggest that an insurance company's ESG commitment matters for its solvency level. Henceforth, we will further investigate the actual impact of SDGs alignment in the insurance industry by including other relevant variables and introducing different methodological approaches.

References

1. Aburto Barrera, L.I., Wagner, J.: A systematic literature review on sustainability issues along the value chain in insurance companies and pension funds. Eur. Actuarial J. **13**, 653–701 (2023). https://doi.org/10.1007/s13385-023-00349-1
2. Bressan, S.: Effects from ESG scores on P&C insurance companies. Sustainability **15**, 12644 (2023). https://doi.org/10.3390/su151612644
3. Brogi, M., Cappiello, A., Lagasio, V., Santoboni, F.: Determinants of insurance companies' environmental, social, and governance awareness. Corp. Soc. Responsib. Environ. Manag. **29**(5), 1357–1369 (2022). https://doi.org/10.1002/csr.2274
4. Gatzert, N., Reichel, P., Zitzmann, A.: Sustainability risks & opportunities in the insurance industry. ZVersWiss **109**, 311–331 (2020). https://doi.org/10.1007/s12297-020-00482-w
5. McDaniels, J., Robins, N.I., Bacani, B.: Sustainable insurance the emerging agenda for supervisors and regulators. United Nations Environment Programme (2017)
6. Schanz, K.-U.: The Geneva Association. The role of insurance in promoting social sustainability (2022). https://www.genevaassociation.org/sites/default/files/2022-11/social_sustainability_report.pdf

The Environmental Score and the Financial Statement: A Machine Learning Analysis for Four European Stock Indexes

Rita D'Ecclesia[1], Susanna Levantesi[1(✉)], Gabriella Piscopo[2], and Kevyn Stefanelli[1]

[1] Sapienza University of Rome, Rome, Italy
susanna.levantesi@uniroma1.it
[2] Federico II University of Naples, Naples, Italy

Abstract. Following the principles of a sustainable economy, companies are increasingly adopting business strategies that seek to harmonize profit objectives with their environmental, social, and governance (ESG) policies. The financial sector's growing awareness of climate and environmental risks underscores the necessity for developing sustainable investments that endorse activities with minimal environmental impact. Sustainability, incorporating environmental, social, and governance considerations, is a strategic priority in this paradigm. This study focuses on the environmental risk aspect, encompassing a company's overall environmental impact and potential risks arising from environmental issues. The primary objective is to discern the structural features of listed firms that influence their sustainability levels, as measured by their "E" score. Leveraging balance sheet information from a selection of European listed firms, our investigation aims to reveal potential relationships between corporate financial variables and the E score. To unravel complex, non-linear relationships within one of the most environmentally conscious markets, namely the European market, we employ advanced techniques such as the random forest and gradient-boosting machine algorithms. This approach allows us to deeply understand how financial variables interplay with a firm's environmental sustainability, offering insights into the intricate dynamics shaping sustainable practices in a corporate context.

Keywords: Environmental score · corporate finance · machine learning

1 Introduction

The financial sector's increasing focus on climate change and environmental risks stems from a recognition of the imperative for sustainable investments that bolster productive endeavors while minimizing environmental harm. Environmental risks encompass issues like inefficient energy consumption, deforestation, biodiversity

M. Corazza et al. (Eds.): MAF 2024, *Mathematical and Statistical Methods for Actuarial Sciences and Finance*, pp. 112–118, 2024.
https://doi.org/10.1007/978-3-031-64273-9_19

loss, air and water pollution, and the failure to adapt to climate changes, potentially resulting in fines or lawsuits. A keen awareness of these risks enables the strategic allocation of resources in alignment with the principles of a sustainable economy. In this evolving landscape, the demand for financial investments aligns with sustainability impact objectives. Asset managers play a pivotal role in identifying investments that adhere to sustainability criteria. This paper specifically delves into the Environmental (E) component of the ESG score, a widely used metric for gauging a firm's non-financial performance. Our focus is on exploring how a company's business activities influence its E score value. This contribution offers a distinctive perspective compared to existing literature, which predominantly concentrates on how the ESG score can impact company profitability and financial performance. In chronological order, the literature first focused on the relationship between ESG and financial returns and then moved on to the impact of the ESG score on corporate profitability. As regards the former aspect, [7] observe that companies with higher ESG scores show higher financial returns and suffer during periods of financial crisis. By analyzing a sample of S&P500 companies, [8] finds a positive correlation between ESG scores and firm financial performance and shows that successful shareholder ESG proposals provide positive abnormal returns, while [3] observe an indirect relation between ESG score and the company's cost of capital. Regarding the latter, we refer to [1,6,10,11,14], and [17]. [12] invert the rationale with respect to the cited contributions questioning whether financial performance leads to higher environmental performance. Following this vein, [4] and [5] propose a machine learning (ML) approach to detect if fundamental ratios are predictors of ESG score. The same framework inspires this paper, where, however, the E component is analyzed separately.

2 The Model

In this paper, we propose to implement both the random forest (rf) and the gradient boosting machine (gbm) algorithm (for further details, see [2] and [9]) to Environmental score data to detect potential non-linear relations among variables. The target variable is the Environmental score, $E.score$. The features considered are listed in the following: $Year$ is the current year, DPS is the Dividend per Share, $PtoE$ is the Price to Earnings, ROE is the Return On Equity, ROA is the Return On Asset, SR is the Solvency Ratio, CI is the Carbon Intensity (ratio of total net carbon dioxide emissions in tonnes to revenues million USD), $MCtoE$ is the ratio of Market Capitalization to EBIT, NPM is the Net Profit Margin (ratio of net income to revenues), $EBITm$ is the Earnings before interest and taxes (EBIT) margin (ratio of EBIT to revenues), AT is the Asset Turnover (ratio of revenues to total assets), LR is the Liquidity Ratio (ratio of stock of highly liquid asset to total net cash outflows), and $size$ is the size of the company calculated as the logarithm of the total assets. The relation of the regression among the variables may be expressed by the following expression:

$$E.score = f(Year, DPS, PtoE, ROE, ROA, SR, CI, MCtoE, NPM, EBITm, AT, LR, Size)$$

3 Numerical Application and Concluding Remarks

In this numerical application, we have considered data from four European Stock indexes (CAC, DAX, FTSE, and FTSE MIB), which represent th 70% of the European stock market liquidity, over the period 2012–2022. The annual data are collected from the Refinitiv data source [15]. After the data cleaning process, the final sample includes 265 companies divided as follows: 49.1% FTSE, CAC 18.1%, DAX 18.1%, and FTSE MIB 14.7%. We calculate the Pearson correlation coefficient, r, between pairs of variables and investigate whether it is significantly different from zero based on a t-distribution of the test statistic $t = \frac{r}{\sqrt{1-r^2}}\sqrt{n-2}$, where n is the number of observations in the variables. If the p-value is less than the chosen significance level, the correlation between the variables is significant. Table 1 shows the coefficients r and the results of the t-test for the Pearson correlation coefficient applied to the full dataset. For the sake of brevity, we only report the coefficients among the target variable ($E.score$) and all the other features. For the $E.score$, the most significant (positive) correlation is with $Size$, followed by ROA and AT (in these cases, the correlation coefficient is negative). We also provide the results of the correlation analysis separately for the four European Stock indexes (see Table 2 for the CAC and DAX indexes and Table 3 for the FTSE and FTSE MIB). Although each of the four indices has its specificity, we generally observe that the variable most correlated with $E.score$ is still $size$, and ROA is also highly correlated. We split the dataset into a training set (70% of the data) and a test set (the remaining 30% of the data). We conduct a tuning procedure on the model's parameter space on the

Table 1. Pearson correlation coefficients and correlation tests of $E.score$ with the other features. Correlation coefficients (column **r**); results of the correlation test (column **t**, column **95% CI** for the confidence interval and **p** for the p-value: *: p-value < 0.05, **: p-value < 0.01, ***: p-value < 0.001).

Feature	r	95% CI	t	p
DPS	0.08	$[0.04, 0.13]$	3.72	0.007^{**}
$PtoE$	-0.03	$[-0.07, 0.02]$	-1.18	$>.999$
ROE	-0.07	$[-0.12, -0.03]$	-3.27	0.034^{*}
ROA	-0.21	$[-0.26, -0.17]$	-9.79	$<.001^{***}$
SR	-0.15	$[-0.20, -0.11]$	-6.99	$<.001^{***}$
CI	0.03	$[-0.02, 0.07]$	1.18	$>.999$
$MCtoE$	-0.10	$[-0.14, -0.05]$	-4.38	$<.001^{***}$
NPM	0.01	$[-0.03, 0.05]$	0.45	$>.999$
$EBITm$	-0.03	$[-0.08, 0.01]$	-1.48	$>.999$
AT	-0.17	$[-0.21, -0.13]$	-7.69	$<.001^{***}$
LR	-0.10	$[-0.14, -0.06]$	-4.51	$<.001^{***}$
$Size$	0.57	$[0.54, 0.60]$	31.23	$<.001^{***}$

Table 2. Pearson correlation coefficients and correlation tests of *E.score* with the other features. Correlation coefficients (column **r**); results of the correlation test (column **t**, column **95% CI** for the confidence interval and **p** for the p-value: *: p-value < 0.05, **: p-value < 0.01, ***: p-value < 0.001).

Feature	CAC				DAX			
	r	95% CI	t	p	r	95% CI	t	p
DPS	0.15	[0.05, 0.25]	2.96	0.134	0.09	[−0.02, 0.19]	1.59	>.999
PtoE	0.08	[−0.02, 0.18]	1.59	>.999	−0.04	[−0.14, 0.07]	−0.69	>.999
ROE	−0.12	[−0.22, −0.02]	−2.44	0.556	−0.13	[−0.24, −0.03]	−2.49	0.46
ROA	−0.21	[−0.30, −0.11]	−4.08	0.003^{**}	−0.28	[−0.37, −0.17]	−5.31	$<.001^{***}$
SR	−0.19	[−0.28, −0.09]	−3.75	0.009^{**}	−0.2	[−0.30, −0.10]	−3.79	0.009^{**}
CI	−0.06	[−0.16, 0.04]	−1.25	>.999	0.06	[−0.04, 0.17]	1.14	>.999
MCtoE	−0.02	[−0.12, 0.08]	−0.44	>.999	−0.09	[−0.19, 0.02]	−1.64	>.999
NPM	0.08	[−0.02, 0.18]	1.61	>.999	−0.18	[−0.28, −0.08]	−3.38	0.035^{*}
EBITm	0.06	[−0.04, 0.16]	1.11	>.999	−0.16	[−0.26, −0.06]	−3.00	0.116
AT	−0.28	[−0.37, −0.18]	−5.66	$<.001^{***}$	−0.15	[−0.25, −0.05]	−2.82	0.191
LR	−0.11	[−0.21, −0.01]	−2.15	>.999	−0.21	[−0.31, −0.11]	−4.02	0.004^{**}
Size	0.42	[0.33, 0.50]	9.02	$<.001^{***}$	0.59	[0.51, 0.65]	13.46	$<.001^{***}$

training set using a grid search to find the optimal values of the parameters. The results are obtained from the R packages *randomForest* developed by [13] for the rf algorithm and the *gbm* developed by [16] for the gbm. We conduct the analysis separately for each of the four European Stock indexes considered.

We evaluate the model's ability to predict the E score through two conventional error measures, the root mean square error (RMSE) and the mean absolute percentage error (MAPE). The error values for the train set (70% of the data) and test set (30% of the data) are reported in Table 4. According to an empirical rule, a MAPE value lower than 10% is considered a very good value, while a MAPE value in the range of 10%–20% is considered a good value. Therefore, the performances of our model are good for the indexes considered, except for FTSE in both the train and test samples. Figure 1 shows the importance variable for the four considered indexes. We highlight that the most important variable influencing the E score is the firm size measured as a function of the total asset. The second important variable is the Carbon Intensity, while the performance ratios account for less contribution.

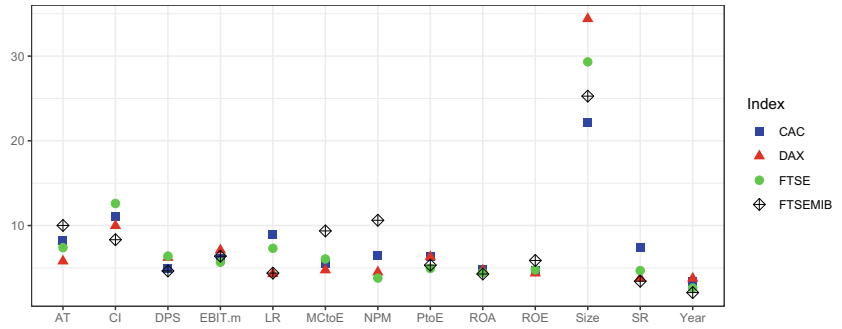


Fig. 1. Relative influence of the variables by stock index. GBM.

Thanks to the machine learning algorithms, we have been able to catch complex and non-linear relations between the features and the *E.score*. This is the case of *CI*, which is linearly uncorrelated with the other features (e.g., Pearson's correlation coefficient between the *E.score* and *CI* is 0.03), but assumes great importance in explaining the *E.score*. This analysis can be helpful for companies that want to pursue environmental sustainability objectives, highlighting the variables on which they can most leverage to obtain high *E.scores*.

Table 3. Pearson correlation coefficients and correlation tests of *E.score* with the other features. Correlation coefficients (column **r**); results of the correlation test (column **t**, column **95% CI** for the confidence interval and **p** for the p-value: *: p-value < 0.05, **: p-value < 0.01, ***: p-value < 0.001).

Feature	FTSE				FTSE MIB			
	r	95% CI	t	p	r	95% CI	t	p
DPS	0.17	[0.11, 0.22]	5.40	$<.001^{***}$	0.14	[0.02, 0.26]	2.21	0.736
PtoE	−0.08	[−0.14, −0.02]	−2.71	0.231	0.07	[−0.06, 0.19]	1.05	>.999
ROE	0.00	[−0.06, 0.06]	0.12	>.999	−0.2	[−0.32, −0.08]	−3.18	0.071
ROA	−0.12	[−0.18, −0.06]	−3.86	0.006^{**}	−0.33	[−0.44, −0.21]	−5.35	$<.001^{***}$
SR	−0.07	[−0.13, −0.01]	−2.39	0.542	−0.21	[−0.33, −0.09]	−3.35	0.041^{*}
CI	0.09	[0.03, 0.15]	2.93	0.136	−0.11	[−0.23, 0.02]	−1.72	>.999
MCtoE	−0.09	[−0.15, −0.03]	−3.03	0.101	−0.28	[−0.39, −0.16]	−4.48	$<.001^{***}$
NPM	0.11	[0.04, 0.17]	3.41	0.031^{*}	−0.36	[−0.47, −0.25]	−6.04	$<.001^{***}$
EBITm	0.02	[−0.04, 0.08]	0.58	>.999	−0.3	[−0.41, −0.18]	−4.80	$<.001^{***}$
AT	−0.15	[−0.21, −0.09]	−4.98	$<.001^{***}$	0.01	[−0.11, 0.14]	0.21	>.999
LR	−0.03	[−0.09, 0.03]	−0.99	>.999	−0.15	[−0.27, −0.02]	−2.32	0.575
Size	0.54	[0.50, 0.58]	20.67	$<.001^{***}$	0.5	[0.40, 0.59]	8.92	$<.001^{***}$

Table 4. RMSE and MAPE by ML model and stock index. The underlined values indicate the best values for the train set, and the doubled underlined ones indicate the best values for the test set.

Index (Country of exchange)	RMSE				MAPE			
	RF		GBM		RF		GBM	
	train	test	train	test	train	test	train	test
CAC (France)	10.89	11.06	11.42	10.66	12.62	13.19	12.92	12.08
DAX (Germany)	9.62	10.96	9.96	10.93	11.55	14.33	12.17	13.71
FTSE (United Kingdom)	13.09	12.29	12.96	11.44	26.96	28.15	24.68	24.33
FTSE MIB (Italy)	13.55	12.15	12.60	11.55	20.98	18.66	18.66	16.11

References

1. Brogi, M., Lagasio, V.: Environmental, social, and governance and company profitability: are financial intermediaries different? Corp. Soc. Responsib. Environ. Manag. **26**(3), 576–587 (2019)
2. Breiman, L.: Random forests. Mach. Learn. **45**, 5–32 (2001)
3. Clark, G.L., Feiner, A., Viehs, M.: From the stockholder to the stakeholder: how sustainability can drive financial outperformance. Available at SSRN 2508281 (2015)
4. D'Amato, V., D'Ecclesia, R.L., Levantesi, S.: Fundamental ratios as predictors of ESG scores: a machine learning approach. Decis. Econ. Finan. **44**, 1087–1110 (2021)
5. D'Amato, V., D'Ecclesia, R.L., Levantesi, S.: ESG score prediction through random forest algorithm. CMS **19**(2), 347–373 (2022)
6. Duque-Grisales, E., Aguilera-Caracuel, J.: Environmental, social and governance (ESG) scores and financial performance of multilatinas: moderating effects of geographic international diversification and financial slack. J. Bus. Ethics **168**(2), 315–334 (2021)
7. Eccles, R.G., Serafeim, G., Seth, D., Ming, C.C.Y.: The performance frontier: innovating for a sustainable strategy: interaction. Harv. Bus. Rev. **91**(7), 17–18 (2013)
8. Flammer, C.: Does corporate social responsibility lead to superior financial performance? A regression discontinuity approach. Manag. Sci. **61**(11), 2549–2568 (2015)
9. Friedman, J.H.: Greedy function approximation: a gradient boosting machine. Ann. Stat. 1189–1232 (2001)
10. Giese, G., Lee, L.E., Melas, D., Nagy, Z., Nishikawa, L.: Foundations of ESG investing: how ESG affects equity valuation, risk, and performance. J. Portfolio Manag. **45**(5), 69–83 (2019)
11. Kim, S., Li, Z.: Understanding the impact of ESG practices in corporate finance. Sustainability **13**(7), 3746 (2021)
12. Laguir, I., Marais, M., El Baz, J., Stekelorum, R.: Reversing the business rationale for environmental commitment in banking: does financial performance lead to higher environmental performance? Manag. Decis. **56**(2), 358–375 (2018)
13. Liaw, A.: Package Randomforest (2018). https://cran.r-project.org/web/packages/randomForest/randomForest.pdf

14. Pellegrini, C.B., Caruso, R., Mehmeti, N.: The impact of ESG scores on cost of equity and firm's profitability. New Challenges Corporate Gov. Theory Pract. **3**, 38–40 (2019)
15. Refinitiv. Environmental, Social, and Governance Scores From Refinitiv (2022). https://www.refinitiv.com/content/dam/marketing/en_us/documents/methodology/refinitiv-esg-scores-methodology.pdf
16. Ridgeway, G.: Generalized Boosted Models: A Guide to the GBM Package (2007). https://cran.r-project.org/web/packages/gbm/gbm.pdf
17. Xie, J., Nozawa, W., Yagi, M., Fujii, H., Managi, S.: Do environmental, social, and governance activities improve corporate financial performance? Bus. Strateg. Environ. **28**(2), 286–300 (2019)

A Combination of NLP and Monte Carlo Technique to Improve Wind Investment Decisions

Antonio Di Bari[1], Luca Grilli[2], Domenico Santoro[1], and Giovanni Villani[1(✉)]

[1] Department of Economics and Finance, University of Bari Aldo Moro, 70124 Bari, BA, Italy
{antonio.dibari,domenico.santoro,giovanni.villani}@uniba.it
[2] Department of Economics, Management and Territory, University of Foggia, 71121 Foggia, FG, Italy
luca.grilli@unifg.it

Abstract. Investment decisions in wind projects can be tough considering the uncertain economic performance depending on the stochastic nature of revenues. Thanks to the recent innovation in Natural Language Processing (NLP), this work tries to present an innovative approach based on the Monte Carlo option pricing model and Sentiment Analysis. Treating it as a financial option, the idea is to price the managerial flexibility of changing investment decisions during the project lifetime depending on the wind investment's profitability. In this way, the Monte Carlo options pricing technique is combined with the sentiment (polarity) score, allowing the modification of transition probabilities from one phase of the investment to another and, consequently, the profitability of the investment.

Keywords: Wind project · Sentiment analysis · Real Options

1 Introduction

Based on financial options theory, the Real Options Approach (ROA) has been used in literature to overcome the limitations of the discounted cash flow approaches, allowing for pricing the managerial flexibility according to the project value evolution [16]. There are many cases of ROA used to evaluate renewable projects [6,10] and through various methodologies [2,4], particularly Monte Carlo (MC) methods [3,9,11,14] highlighting the usefulness of the MC approach in calculating the real options in renewable energy.

This work uses the Monte Carlo approach to price Real Options (RO), focusing on the investment choice in wind farm projects. The main innovation concerns using the MC technique for pricing Real Options with the polarity score to modify the transition probabilities from one phase to the next in investments. This addition helps the decision maker choose whether to continue the investment

M. Corazza et al. (Eds.): MAF 2024, *Mathematical and Statistical Methods for Actuarial Sciences and Finance*, pp. 119–123, 2024.
https://doi.org/10.1007/978-3-031-64273-9_20

based on the information acquired from the external world. Since the polarity score is calculated through the NLP method, this work proposes a combination between Sentiment Analysis and MC options pricing technique. Although previous works combine Options theory with NLP, none of them provides an integrated approach between MC options pricing method with Sentiment Analysis [7]. Moreover, other studies combine Artificial Intelligence (AI) techniques with other mathematical approaches, such as the RO. For example, Rath and Chow [15] propose a scalable Machine Learning framework for RO related to Mobility-on-demand (MOD) services. Or Lazo et al. [12], who used a combination of MC and Genetic Algorithms (GA) to optimize the choice of alternatives in an oil sector investment.

In our paper, the Sentiment Analysis task allows for determining the polarity score of some newspaper articles, making it possible to modify the probability of success/failure in the transition phases.

2 NLP and Sentiment Analysis

NLP models have evolved over time. Especially in recent years, 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. In particular, this new model is an encoder-decoder network that uses the transformers architecture [17] to solve 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 that perform two fundamental tasks: *Masked Language Modeling* (MLM), that randomly mask the 15% of the tokens being in the corpora and the *Next Sentence Prediction* (NSP), that is the ability to predict if two sentences follow each other.

As defined in Di Bari et al. [7], we try to exploit the news polarity score obtained through specific BERT models to improve the transition probabilities. The chosen BERT model is FinBERT [1], which specializes in financial corpora in English. This model was trained on three datasets: TRC2-financial (filtering the TRC2 corpus based on financial keywords), FinancialPhraseBank [13], and

Table 1. FinBERT parameters description [1]

Parameter	Value
Dropout (p)	0.1
Warm-up prop	0.2
Max_Seq length	64 tokens
Learning rate	$2e{-}5$
Loss	0.37
Accuracy	0.86
F1-score	0.84

FiQA Sentiment (a dataset created for the *WWW'18* Conference challenge). As reported by the author, the parameters for implementing this model and the accuracy results after a 10-fold Cross-Validation are reported in Table 1. This model has been implemented with a dropout probability of $p = 0.1$, a warm-up proportion of 0.2, a maximum sequence length of 64 tokens, a learning rate of $2e{-}5$, and a mini-batch size of 64, from which after a 10-fold cross-validation, FinBERT achieves a Loss of 0.37, with an Accuracy of 0.86 and an *F1*-score of 0.84. The output of FinBERT is, for each inserted sentence of an article, a polarity index *[negative, neutral, positive]* accompanied by a numerical score $\gamma \in [-1, 1]$, which average $\bar{\gamma} = \sum \gamma$ represents the global polarity of the news. We can divide γ into $\gamma_- \in [-1, 0)$, representative of sentences with negative polarity, and $\gamma_+ \in [0, 1]$, representative of sentences with positive polarity, hence $\gamma = \gamma_- \cup \gamma_+$. This polarity score can be used to modify the transition probability. As identified in [7] followed the Dias [8] framework, the information revealed $\eta^2 = \bar{\gamma}$. In this way, adjusting the range of variation (through manipulations with logistic function), the revealed success probabilities for positive dependence become [7]:

$$q^+ = q + (1 - q)\sqrt{\gamma_{adj}}, \tag{1}$$

$$q^- = q - q\sqrt{\gamma_{adj}}. \tag{2}$$

3 Monte Carlo Option Pricing Model

We denote as V_t the project present value evolution during the time $t \in [0, T]$. We assume the wind project is characterized by sequential investment costs K_0, K_1, and K_2 that could be related respectively to the costs to initiate the project, the costs to create the wind farm, and the costs for operation activity. The investor holds the operational flexibility to invest in a wind project for a profit $V(T) - K_2$. Otherwise, if $K_2 > V(T)$, the investor should abandon the wind investment to avoid financial losses. Following the classical Black-Scholes model [2], we calculate the project value evolution up to time t_1:

$$V^i(t_1) = V(0)e^{(r - \frac{\sigma^2}{2})t_1 + \sigma\sqrt{t_1}Z} \tag{3}$$

This way, we can compute n possible Monte Carlo iterations of possible revenue scenarios. The $V^i(t_1)$ values will act as the underlying assets of the Black-Scholes formula (4) up to maturity T:

$$e^i_{t_1} = V^i(t_1)N(d_1) - e^{-r(T-t_1)}K_2N(d_2) \tag{4}$$

with d_1 and d_2 in the classical form [2]. This way, we can include positive (q^+) and negative (q^-) information revelation obtained by the sentiment analysis phase to "adjust' the probability of moving on to the next investment phase. The sentiment analysis is relevant to compute q^+ and q^- that represent the probabilities that are adjusted for the parameter $\sqrt{\gamma_{adj}}$ that captures the information from

the newspapers. In fact, this parameter represents the reshaped mean of different polarities calculated through the BERT model on newspaper articles. Thus, we adjust the Black-Scholes formula from t_1 to T by making a kind of expected value between the option multiplied by the success probability of the first phase (p) and the option multiplied by the failure probability of the first phase ($1-p$). Note that the underlying asset in the two scenarios is multiplied respectively by positive (q^+) and negative (q^-) information revelation.

At this point, we can calculate the new payoff of the first stage $t \in [0, t_1]$. It is calculated as a compound option (c) in which the underlying asset is represented by the previous simple option (e). Proceeding backward from t_1 to t_0, and calculating a mean of the number of Monte Carlo simulations $i = 1, 2, ..., n$, the final compound option value that represents the operational flexibility of the wind projects is:

$$\tilde{c}_0^i = E_{\mathbb{Q}}[e^{-r(t_1-t_0)}(e_{t_1} - K_1)^+], \tag{5}$$

whose average gives the $\tilde{c}_0$ value. Following the same logic of [7], the final RO value (ROV) is given by:

$$ROV = -K_0 + p \cdot \tilde{c}_0. \tag{6}$$

A positive value of ROV implies that the wind investment should be pursued because it could generate profits for potential investors.

To implement this approach, the numerical example should consider the characteristics of wind energy production, electricity price, possible incentive mechanism of government, the value of the cost, the success probability of stages, the intensity of information revelation calculated through NLP and other financial parameters such as volatility and risk-free rate. The implementation results should support the investor in investment decision-making in renewable energy projects. If the ROV is positive, the investor should pursue the project; otherwise, he should reject it.

4 Conclusions

This work tries to support decision procedures in wind investment, combining the Monte Carlo options pricing technique and sentiment analysis to allow the investor to assess the wind projects. In this case, the Monte Carlo pricing technique is an embedding of the classical Monte Carlo method and the B&S formula linked under a compound options framework. The Monte Carlo options pricing technique is helpful since it allows us to consider the managerial flexibility of pursuing the investment only if financial profitability conditions exist. At the same time, sentiment analysis makes it possible to incorporate new probability values into the Monte Carlo option pricing model on the transition or not to a subsequent phase of the investment, based on the news that can be extracted from the surrounding environment.

Acknowledgement. The Authors acknowledge the financial support from the program MUR PRIN 2022 n. 2022ETEHRM "Stochastic models and techniques for the management of wind farms and power systems".

References

1. Araci, D.: FinBERT: financial sentiment analysis with pre-trained language models. arXiv:1908.10063v1 (2019)
2. Black, F., Scholes, M.: The pricing of options and corporate liabilities, pp. 3–21
3. Boyle, P.P.: Options: a Monte Carlo approach. J. Financ. Econ. **4**, 323–328 (1977)
4. Cox, J., Ross, S., Rubinstein, M.: Option pricing: a simplified approach. J. Financ. Econ. **7**(3), 229–263 (1979)
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 (Volume 1: Long and Short Papers), pp. 4171–4186. Association for Computational Linguistics (2019)
6. Di Bari, A., Bufalo, M., Villani, G.: Multi-stage real option evaluation with double barrier under stochastic volatility and interest rate. Ann. Financ. **18**, 247–266 (2022)
7. Di Bari, A., Santoro, D., Tarrazon-Rodon, M.A., Villani, G.: The impact of polarity score on real option valuation for multistage projects. Qual. Quant. **58**, 57–76 (2023)
8. Dias, M.A.G.: Real options, learning measures, and Bernoulli revelation processes. Working Paper. Puc-Rio, Presented at 8th Annual International Conference on Real Options, Paris, pp. 1–40 (2005)
9. Fathi-Vajargah, B., Mirzazadeh, M., Ghasemalipour, S.: An efficient Monte Carlo simulation for new uncertain Heston-CIR hybrid model. Soft. Comput. **25**, 8539–8547 (2021)
10. Fernandes, G., Mathé Maia, V., Lima Gomes, L.: Application of real options theory in the evolution of swine biogas storage. Revista de Gestão, Finanças e Contabilidade **5**(2), 5–22 (2005)
11. He, X.J., Lin, S.: A closed-form pricing formula for variance swaps under a stochastic volatility model with a stochastic mean-reversion level. Soft. Comput. **26**, 3939–3946 (2022)
12. Lazo, J.G.L., Pacheco, M.A.C., Vellasco, M.M.B.R.: Real options and genetic algorithms to approach of the optimal decision rule for oil field development under uncertainties. In: Castillo, O., Melin, P., Ross, O.M., Sepúlveda Cruz, R., Pedrycz, W., Kacprzyk, J. (eds.) Theoretical Advances and Applications of Fuzzy Logic and Soft Computing, vol. 42, pp. 445–454. Springer, Heidelberg (2007). https://doi.org/10.1007/978-3-540-72434-6_44
13. 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**, 782–796 (2014)
14. Najafi, P., Talebi, S.: Using real options model based on Monte-Carlo least-squares for economic appraisal of flexibility for electricity generation with VVER-1000 in developing countries. Sustain. Energy Technol. Assess. **47** (2021). https://doi.org/10.1016/j.seta.2021.101508
15. Rath, S., Chow, J.Y.J.: A deep real options policy for sequential service region design and timing. arXiv:2212.14800 (2022)
16. Trigeorgis, L.: Real options and interactions with financial flexibility. Financ. Manag. **22**(3), 202–224 (1993)
17. Vaswani, A., et al.: Attention is all you need. In: Advances in Neural Information Processing Systems, pp. 5998–6008 (2017)

Meeting the Challenges of Longevity: Lifetime Income from Real Estate

Emilia Di Lorenzo[1], Francesco Rania[2(✉)], Marilena Sibillo[3], and Annarita Trotta[2]

[1] Department of Economic and Statistical Sciences, University of Naples Federico II, Monte S.Angelo, 80126 Naples, Italy
diloremi@unina.it

[2] Department of Law, Economics and Sociology, University of Catanzaro Magna Graecia, Loc. Germaneto, 88100 Catanzaro, Italy
{raniaf,trotta}@unicz.it

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

Abstract. The reverse mortgage, in spite of its low uptake, is a tool of protecting the so-called house rich-cash poor persons. It represents an opportunity especially in social realities with a high aging population, pension systems struggling to guarantee acceptable living standards, significant proportion of homeowners with insufficient income to meet daily expenses. The paper is interested in this contract by looking at it from the point of view of the provider/lender who commits to a payment to the borrower/pensioner against the amount that will be made when the house is sold, in which the borrower will continue to live for life. Our focus is on quantifying the risks that impact the contract, due to three main sources of uncertainty. The purpose is to provide formulas for quantifying each of these risks, which are easy to interpret and apply, and useful for proceeding to profitable risk control strategies.

Keywords: Reverse mortgages · solutions for retirement · risk analysis · risk indexes

1 The Reverse Mortgage: The New Way to View Your Home as an Asset

The Reverse Mortgage contract (RM from here on) constitutes an interesting contractual instrument aimed at protecting the so-called house rich-cash poor social group, characterized by modest economic conditions but at the same time owning the home in which they live. In economic-financial realities characterized by a pension system often unable to guarantee acceptable standards of living in old age, as is often the case in the Italian context, such instruments prove to be extremely efficient. The idea behind the Reverse Mortgage is to consider the

M. Corazza et al. (Eds.): MAF 2024, *Mathematical and Statistical Methods for Actuarial Sciences and Finance*, pp. 124–129, 2024.
https://doi.org/10.1007/978-3-031-64273-9_21

house in which an individual lives exactly as an asset. Here briefly the outline of the contractual lines. As in [2], RMs allow (usually elderly) homeowners (borrowers) to receive loans by the lender; such amounts will be repaid through the selling of their homes following their deaths or their moving out of the home for any reason. Such loans may be provided as a lump sum or as a periodic income stream extended over a predefined interval or throughout the owner's lifetime or tenure in the home. Usually, (cf. [2]) a No Negative Equity Guarantee (NNEG) is in the contract: if the balance of the loan exceeds the proceeds of sale of the property, the amount due to the lender is reduced to the second quantity. In fact, for lenders, the future liquidation value of the property will be a future market value, conditioned by the impact of several variables. This uncertainty could lead to conditions that threaten the NNEG condition. The NNEG guarantees that the debt does not exceed the value of the property.

The diffusion of these contracts is not uniform around the world. While strong interest is noted in, for example, the U.S., Canada, and the U.K., in general acceptance in the rest of Europe is lukewarm, if not poor (see [3]). Reasons for such behavior especially in realities where rationally the diffusion should be marked, probably reside in weak and fragmentary information about an instrument with a non-elementary structure (cf. [6]), which fails to "convince" both the potential counterparts, in a general level of financial education that is far from adequate for a developed economy. To this it is likely to assume that from the same institutions potentially interested in the RM contract as lender/provider, there is some resistance, due to the management and control of the many risks that influence the valuations relevant to the definition of the contractual financial characteristics. It is precisely on this point that our work aims to contribute, in order to simplify the quantification of the most relevant risks.

The layout of the paper is set proposing in Sect. 2 the contractual outline, in Sect. 3 the identification and description of the three main sources of risk impacting the contract and finally, in Sect. 4, some formulas proposed as risk measures, useful to the lender to define appropriate strategies to control the risks themselves. Final conclusions close the paper.

2 The Contractual Model

Let's consider a RM issued in $t = 0$. The homeowner (borrower/pensioner) is an individual aged x at the issue time. Again following [2], there are two ways to collect the amounts due as a result of signing the contract: a lump sum LS_0, payable by the lender/provider at the issue time, or its equivalent in the form of a life annuity, against a proportion α of the current market value of his house H_0, after deducting the insurance premium for the NNEG. The NNEG stems from the possible circumstance that the sale value of the house may turn out to be lower than expected. The actuarial value A_0 in $t = 0$ of all the benefits the homeowner/borrower is going to receive coincides with αH_0, both in case of the lump sum and of the annuity. The following equation can be written:

$$LS_0 < A_0 = \alpha H_0 \tag{1}$$

considering the difference $(A_0 - LS_0)$ as the cost of the NNEG to be borne by the borrower. We frame quantitatively the amount V_t payable to the lender by the heirs of the borrower in an instant, here generically denoted by t, $0 < t < \omega - x$, having indicated by ω the extreme age of an individual. If A_t and H_t are respectively the value of the loan and the value of the property at time t, the amount V_t to be borne by the heirs is given by:

$$V_t = A_t - \max(0, A_t - H_t), \text{ with } 0 < t < \omega - x \tag{2}$$

In Eq. (2) we can note a short position in a put option having strike price A_t, that is the actuarial value at time t of the benefits the borrower has to receive; its underlying is the house value H_t, representing the realized house value in the real estate market at that moment.

3 The Main Risk Driver in a Reverse Mortgage

The issue that makes the evaluation of the monetary quantities inherent in the contract particularly complex lies in some basic elements, which are random in nature. The first is the value of the house at the time of its appraisal. There are many micro-and macroeconomic variables that impact real estate market price trends, which makes describing the random variable "house price" particularly challenging. In addition to this, quantities will be valued at a random instant, linked to the death or exit from the house. The description of the phenomenon of mortality and longevity, in particular, attributable to the borrower, will be relevant. Not least, the purely financial aspect plays a particularly relevant role, given the likely wide extension of the contract term. It is therefore a matter of describing the interest rates that will be applied to the actuarial valuations involved in the contract. As clearly described in [5], these three risk drivers can be considered the main vehicles of uncertainty in financial assessments related to RMs.

Schematically, the house price risk depends on the volatility of the real estate market; the lifespan of an individual depends on the longevity risk; the financial risk is linked to the volatility of the financial market.

Some simple considerations can also be made intuitively.

If the borrower's lifespan is longer than expected, the liquidation of the asset through which the lender will recover the money lent will be postponed, exposing that value to the impact of the other two risks for a longer period. So, it is possible to conclude that the longevity risk increases the financial and the house price risk. To this may be added that if the conclusion of the contract is moved forward due to the longevity effect, the probability of having the liquidation value less than the value of the house at that instant, which would mean a capital loss for the lender, could increase.

To frame the three risk component, we use the notation in [2]:

– the longevity risk is described by means of the random survival function, P, used to specify the borrower's survival probability year by year.

– the financial risk is described by means of the interest rate r_s^{RM}, the rate used for loan-related valuations referred to the period s. Denoted by r_s the risk-free rate and π the risk premium, we can write:

$$r_s^{RM} = r_s + \pi$$

– the house price risk is described by means of the rate r_s^{HV}, the rate used for valuations related to real estate value over time and referred to the generic period s, in other words, the rate of appreciation of real estate in the specific market in which the contract is issued.

Again, considering the generic time of valuation t, we have:

$$A_t = \alpha H_0 \prod_{s=1}^{t}(1 + r_s^{RM}) = A_t = \alpha H_0 \prod_{s=1}^{t}(1 + r_s + \pi), \text{ with } 0 < t < \omega - x \quad (3)$$

$$H_t = \alpha H_0 \prod_{s=1}^{t}(1 + r_s^{HV}), \text{ with } 0 < t < \omega - x \quad (4)$$

The financial balance inherent in the contract at the issue time binds the lump sum LS_0 and the sum V_t in the following way:

$$LS_0 = \sum_{t=1}^{\omega - x} V_t \left(\prod_{s=1}^{t}(1 + r_s)^{-1} \right) {}_{t-1/1}q_x \quad (5)$$

where ${}_{t-1/1}q_x$ is the probability that the borrower aged x dies between the ages $(x + t - 1)$ and $(x + t)$. Formula (5) states that LS_0 is exactly the single premium of the contract or, in case of periodic payments, the value in $t = 0$ of the provider/lender's obligations.

In this context, we decide to elect as a synthetic variable representing the financial value of the contract from the point of view of the lender/provider the gain/loss at time t. To describe it we consider the opportunity cost rate r_{CC}:

$$r_{CC} = r_s + \rho$$

in which, as in [2], ρ is an opportune spread with the risk-free interest rate.

Denoting by L_t the gain/loss at time t, we can write:

$$L_t = \left[LS_0 \prod_{s=1}^{t}(1 + r_{CC}) \right] - V_t \quad (6)$$

In formula (6), for what follows, the opportunity cost rate is considered fixed.

The L_t variable represents an important synthetic indicator of the financial structure of the contract and constitutes, in our opinion, the fundamental quantity to control for the lender. For this reason, in what follows, we will focus the analysis of the risks that impact precisely on this quantity.

4 RM Risk Sources and Related Indexes

In this section we propose some formulas, easily interpretable and implementable, usable as risk measures and risk indexes.

From the lender perspective, we deal with the three risk drivers previously presented, precisely the *longevity risk*, the *house price risk* and the *financial risk* represented respectively by the stochastic variables P, r_s^{HV} and r_s^{RM}.

The three sources of risk can be modeled in accordance with the most appropriate theoretical assumptions. For example, as ascertained in the literature, real estate prices can be described through a Geometric Brownian Motion, longevity using the Lee Carter model, and finally the interest rate process with a jump diffusion model, as made in [2].

For this purpose, the well-known variance decomposition formula proposed by [4] and extended as in [1], appears extremely useful.

Starting from the recalled variance decomposition formula, we decompose with respect to the survival model chosen for the actuarial valuations; as shown in [1], proceeding with the decomposition with respect to each of the other two risk sources leads to numerical results that are not significantly different from each other and the one we propose.

We can write:

$$\mathrm{Var}(L_t) = \mathrm{Var}\left(\mathbb{E}\left[L_t/P\right]\right) + \mathbb{E}\left[\mathrm{Var}\left(L_t/P\right)\right] \tag{7}$$

and we can proceed with the further decomposition of the second addend in Eq. (7):

$$\mathbb{E}[\mathrm{Var}(L_t/P)] = \mathbb{E}\left[\mathrm{Var}\left(\mathbb{E}\left[L_t/r^{HV}\right]/P\right)\right] + \mathbb{E}\left[\mathbb{E}\left[\mathrm{Var}\left(L_t/r^{HV}\right)/P\right]\right] \tag{8}$$

In this way, we are able to quantify 3 risk measures, each referring to one of the three sources of risk. In detail:

- $\mathrm{Var}\left(\mathbb{E}\left[L_t/P\right]\right)$ can be considered as a measure of the risk due to uncertainty about longevity predictions. It is a *longevity risk measure* and has the nature of a model risk.
- $\mathbb{E}\left[\mathrm{Var}\left(\mathbb{E}\left[L_t/r^{HV}\right]/P\right)\right]$ can be considered as a measure of the risk due to the impact of the volatility of the real estate market, that is the volatility of the rate of appreciation of real estate. It is a *real estate market risk measure.*
- $\mathbb{E}\left[\mathbb{E}\left[\mathrm{Var}\left(L_t/r^{HV}\right)/P\right]\right]$ can be considered as a measure of the risk due to the volatility in the financial markets, that is the volatility of the rate in mortgage valuations. It is a *financial risk measure.*

These risk measures can be usefully relativized, so as to refer them to each unit of average loss at the moment of evaluation. Indicating with $\mathbb{E}[L_t]$ the expected value of the losses at time t with respect to the variable P, having averaged out the influence of the other two stochastic variables, the following risk indexes are obtained, each referred, respectively, to each of the risk measures above introduced:

$$\frac{\sigma[L_t]}{\mathbb{E}[L_t]} = \begin{cases} \frac{\sqrt{\mathrm{Var}(\mathbb{E}[L_t/P])}}{\mathbb{E}[L_t]} & \text{relative importance of longevity risk} \\ \frac{\sqrt{\mathbb{E}[\mathrm{Var}(\mathbb{E}[L_t/r^{HV}]/P)]}}{\mathbb{E}[L_t]} & \text{relative importance of house price risk} \\ \frac{\sqrt{\mathbb{E}[\mathbb{E}[\mathrm{Var}(L_t/r^{HV})/P]]}}{\mathbb{E}[L_t]} & \text{relative importance of financial risk} \end{cases} \tag{9}$$

5 Conclusions

The paper focuses on the reverse mortgage, that is a tool framed within the silver economy. It places the home at the center, which becomes an asset for the owner. The two counterparts are, precisely, the homeowner, the borrower/pensioner, and the lender/provider, who pays a lump sum or a cash flow for the duration of the contract. The contract ends when the borrower dies or moves out of the home. Throughout the contract term, the borrower retains the right to live there.

The contract, which is widespread in some geographical realities, is struggling to establish itself in countries that, on paper, would have all the characteristics to be places of contract diffusion, such as Italy. Italy is not the only country to show a strong reluctance toward this type of contractual solution: this phenomenon is scarcely widespread in most of Europe, with the exception of the United Kingdom.

This may be due to weak financial education, the objective complexity of the contract, and the fact that various sources of risk, especially for the lender, impact on it. It is on this point that the paper aims to contribute. From the lender's point of view, the three main sources of risk are identified, and their respective measures are proposed through formulas that are easy to interpret and implement. They provide a useful tool for the purpose of refining risk management and control strategies from a corporate perspective.

References

1. Coppola, M., Di Lorenzo, E., Sibillo, M.: Further remarks on risk sources measuring. The case of a life annuity portfolio. J. Actuarial Pract. 1993–2006 **10**, 229–242 (2002)
2. Di Lorenzo, E., Piscopo, G., Sibillo, M., Tizzano, R.: Reverse mortgage and risk profile awareness: proposals for securitization. Appl. Stoch. Model. Bus. Ind. **38**(2), 353–369 (2022)
3. Fornero, E., Rossi, M.C., Brancati, M.C.U.: Explaining why, right or wrong (Italian) households do not like reverse mortgages. J. Pension Econ. Financ. **15**(2), 180–202 (2016)
4. Frees, E.: Relative importance of risk sources in insurance systems. North Am. Actuarial J. **2**(2), 34–52 (1998)
5. Li, J., Kogure, A., Liu, J.: Multivariate risk-neutral pricing of reverse mortgages under the Bayesian framework. Risks **7**(11), 1–12 (2019)
6. Lusardi, A.: Planning for retirement: the importance of financial literacy. Public Policy Aging Rep. **19**(3), 7–13 (2009)

Statistical Approach to Implied Market Inefficiency Estimation

Fabrizio Di Sciorio(✉), Laura Molero González, and Juan E. Trinidad Segovia

Department of Economics and Business, University of Almería, Almería, Spain
fd940@inlumine.ual.es, {lmg172,jetrini}@ual.es

Abstract. This study aims to estimate the information efficiency of financial markets based on the Hurst exponent, with a focus on the S&P 500 index. The approach involves using statistical models to estimate the implied Hurst exponent through the historical series of the VIX (a proxy for implied volatility) with a 30-day time lag. In this way, the traditional backward-type Hurst estimation is reconciled with that derived from the VIX, which represents a forward-looking measure (a proxy for 30-day volatility). The test sample also includes the COVID pandemic period. The results reveal a good fit from ensemble stacking models, with the random forest standing out as the most effective approach in estimating the implied Hurst index.

Keywords: Hurst Exponent · VIX · Inefficient Market · Ensemble Stacking · Regression

1 Introduction

There are different approaches to interpreting the Hurst index [4]. Bianchi [3] considers Hurst to be a measure of relative volatility with respect to the ideal value of $\frac{1}{2}$, where the market is described by a geometric Brownian motion, and thus efficient. Following this approach, the estimation of prospective (implied) Hurst is addressed in the following work as an estimate of expectations on the future informational efficiency of the market. As a first analysis we investigated the relationship between the Hurst exponent (H_t) and implied volatility [7]. The VIX Index, developed by the Chicago Board of Options Exchange (CBOE), is a proxy of implied volatility for the Standard & Poor's 500 Index and is also widely regarded as a measure of turbulence in U.S. and global financial markets [5]. The formula defining VIX is the following [11]:

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

M. Corazza et al. (Eds.): MAF 2024, *Mathematical and Statistical Methods for Actuarial Sciences and Finance*, pp. 130–135, 2024.
https://doi.org/10.1007/978-3-031-64273-9_22

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 $t, O_i\left(K_i, T\right)$ 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 $(T - t)$. Considering a time varying Hurst exponent [1], in a multi-fractional Brownian motion framework (mBm) [9], we built a statistical approach using the VIX with lag t-30 as a predictor. The first step of the analysis is to estimate Hurst. A comprehensive review on the estimation method can be found in Garcin [6]. Following Bianchi and Pianese [2,3] the estimator (AMBE) can be built [2]:

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

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

2 Statistical Models

We consider daily observations from 02/04/2007 to 24/03/2023 and use the natural log-adjusted price to compute the Hurst exponent. For each model we split the dataset into train (60%), validation(20%) and test (20%). We employed a grid search technique with cross fold (n = 5) to find optimal hyperparameters (i.e., number of degree for poly, number of iteration, max depth end learning rate for boosting). The object function chosen for boosting model is the $Huber-loss$.

2.1 ADL

The auto-regressive distributed lag (ADL) model is a dynamic regression model considering both current and past values of explanatory variables to explain current and future values of the dependent variable.The model is as follows:

$$H_t = \beta_0 + \beta_1 \cdot \text{VIX}_{t-30} + \varepsilon \tag{3}$$

2.2 Polynomial Regression

Polynomial regression extends linear regression by introducing polynomial features.The degree of the polynomial (n) is a hyperparameter.

2.3 Support Vector Regression (SVR)

SVR aims to find a function $f(x)$ that predicts y with minimal error. The kernel function ($K(x_i, x_j)$) determines the type of mapping:

$$H_t = f(VIX_{t-30}) = \sum_{i=1}^{n} \alpha_i K(x_i, x) + b$$

where α_i are Lagrange multipliers and b is a bias term. The hyperparameter C controls the trade-off between smoothness and accuracy.

2.4 Decision Tree Regression, Bagging and Boosting

Decision tree regression predicts y by recursively partitioning the input space, where the prediction for each region is the mean of the target values. CatBoost, a boosting algorithm, is renowned for its robustness against overfitting. Its distinctive feature includes the automatic handling of categorical variable encoding and is characterized by robustness, requiring less parameter tuning effort compared to other boosting algorithms, thanks to optimal default settings. Random forest [8] builds an ensemble of decision trees by randomly selecting subsets of the features and samples to train each tree. This helps to reduce over-fitting and improve generalization performance.

2.5 Ensemble Stacking Method

Ensemble stacking is a technique that combines multiple models, by training a meta-model on the output of the individual models (see Odegua [10] and Wu [13]). The meta-model takes the predictions of the base models as input and learns to combine them in a way that maximizes the overall performance. Let M be the number of base models and N be the number of observations in the dataset. Each base model m produces a set of predictions:

$\mathbf{y}^{(m)} = \{y_1^{(m)}, y_2^{(m)}, \ldots, y_N^{(m)}\}$. The stacked dataset for the second-level model can be constructed using the predictions of the base models as input features.

For the meta-model we investigate on different solutions:

- SEA (simple averaging ensemble) of all prediction at level 0
- linear regression
- random forest
- catboost

3 Models Results

The first aspect considered in the analysis is the stationarity of the two time series (VIX and Hurst) in order to avoid spurious correlations in the models. The (Dickey-Fuller) tests carried out show the stationarity hypothesis verified

Table 1. Results of statistical tests

Variable	Test Statistic	Lag Order/df	p-value	Conclusion
H (Time Series)	−5.839	15	0.01	Stationary
VIX (Time Series)	−5.1079	15	0.01	Stationary
Studentized Breusch-Pagan Test	184.39	1	<2.2e−16	Heteroskedasticity
Lag Autocorrelation D-W Statistic	0.9408284	0	0.1179858	No Auto-correlation
Jarque Bera Test	1197.7	2	<2.2e−16	Non-Normality

(Table 1). In addition, the Jarque Bera test indicates a deviation from a Gaussian distribution. The Breusch-Pagan test reveals a significant heteroskedasticity statistic in the residuals of our model, while the Durbin Watson test suggests that this autocorrelation may not be statistically significant.

The results on test set of the previously discussed models are reported below:

Table 2. Models results on test set

Model	max_err	mae	mse	rmse	R^2	mape	medape
lm	0.138	0.032	0.002	0.045	62%	4.925%	4.472%
poly	0.115	0.029	0.001	0.032	68%	4.42%	3.685%
svr	0.149	0.041	0.004	0.055	37%	6.171%	5.622%
dtr	0.121	0.030	0.001	0.032	68%	4.359%	3.469%
SAE	0.114	0.031	0.001	0.032	64%	4.756%	4.105%
ensemb_rf	0.125	0.028	0.001	0.032	75%	4.383%	3.405%
ensemb_lm	0.121	0.030	0.001	0.032	68%	4.35%	3.45%
ensemb_catboost	0.124	0.029	0.002	0.032	67%	4.404%	3.419%

To study the behavior of the prediction error, we introduce the following function:

$$f(H_t, \hat{H}_t) = |H_t - \hat{H}_t| \tag{4}$$

This function allows us to analyze the prediction errors (absolute residuals) between the actual efficiency of the market and the efficiency estimated by the models described previously (Figs. 1, 2).

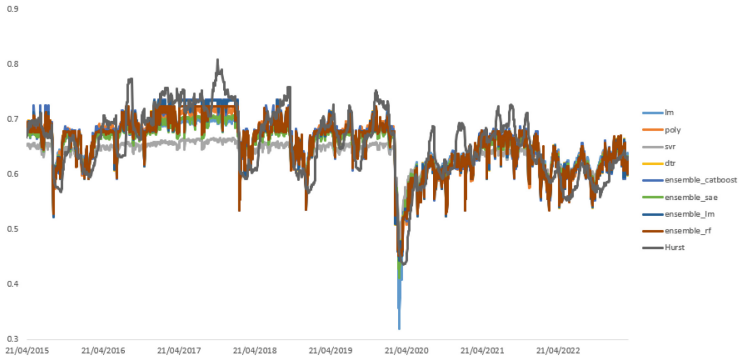


Fig. 1. Estimation of implied Hurst by main models

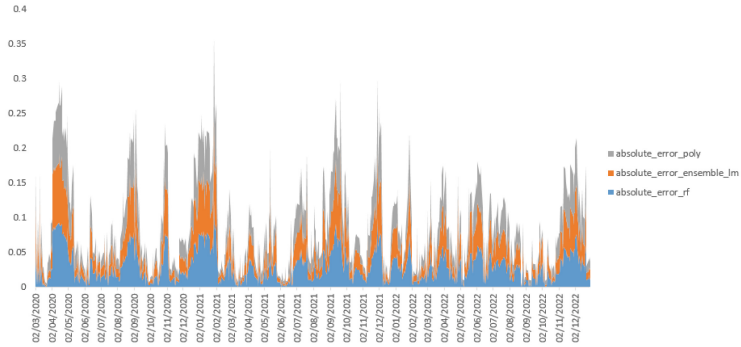


Fig. 2. Absolute difference of implied Hurst estimation during COVID pandemic

4 Conclusions and Further Directions

In conclusion, the financial analysis of Hurst estimation provides valuable insights into the future informational efficiency of the market. The ability to assess the persistence of trends over time through the Hurst parameter proves crucial for anticipating market dynamics and making informed decisions. From a statistical point of view, the ensemble stacking approach, particularly based on random forest, demonstrates superior performance on the test set, see Table 2. These results corroborate existing evidence in the literature regarding stacking/blending approaches [12], emphasizing the robustness and effectiveness of decision tree-based models and such predictive methodologies. Of particular significance is the model's performance during the critical period of the COVID pandemic, where the random forest stands out for having the lowest absolute error. If, on the one hand, we have verified the existence of a link between Hurst and the VIX, just as implied volatility is analyzed, it would be of extreme interest to delve into inefficiency studies. For example, in fractional option pricing theory, investigating the term structure of the Hurst exponent.

References

1. Bianchi, S., Di Sciorio, F., Mattera, R.: Forecasting VIX with Hurst exponent. In: Corazza, M., Perna, C., Pizzi, C., Sibillo, M. (eds.) MAF 2022, pp. 90–95. Springer, Cham (2022). https://doi.org/10.1007/978-3-030-99638-3_15
2. Bianchi, S., Pantanella, A., Pianese, A.: Modeling stock prices by multifractional Brownian motion: an improved estimation of the pointwise regularity. Quantit. Financ. **13**(8), 1317–1330 (2013)
3. Bianchi, S., Pianese, A.: Time-varying Hurst-Hoelder exponents and the dynamics of (in)efficiency in stock markets. Chaos Solitons Fractals **109**(5), 64–75 (2018)
4. Couillard, M., Davison, M.: A comment on measuring the Hurst exponent of financial time series. Phys. A **348**, 404–418 (2005)
5. Degiannakis, S.A.: Forecasting VIX. J. Money Invest. Bank. (4) (2008)
6. Garcin, M.: Estimation of time-dependent Hurst exponents with variational smoothing and application to forecasting foreign exchange rates. Phys. A **483**, 462–479 (2017)
7. Li, H., et al.: Temporal detection of sharp landslide deformation with ensemble-based LSTM-RNNs and Hurst exponent. Geomat. Nat. Haz. Risk **12**(1), 3089–3113 (2021)
8. Luong, C., Dokuchaev, N.: Forecasting of realized volatility with the random forests algorithm. J. Risk Financ. Manage. **11**(4), 61 (2018)
9. Mattera, R., Sciorio, F.D.: Option pricing under multifractional process and long-range dependence. Fluct. Noise Lett. **20**(01), 2150008 (2021)
10. Odegua, R.: An empirical study of ensemble techniques (bagging, boosting and stacking). In: Proceedings of the Conference: Deep Learning IndabaXAt (2019)
11. Ouandlous, A., Barkoulas, J.T., Alhaj-Yaseen, Y.: Persistence and discontinuity in the VIX dynamics. Chaos Solitons Fractals **113**, 333–344 (2018)
12. Shwartz-Ziv, R., Armon, A.: Tabular data: deep learning is not all you need. Inf. Fusion **81**, 84–90 (2022)
13. Wu, D., et al.: On the functional equivalence of TSK fuzzy systems to neural networks, mixture of experts, CART, and stacking ensemble regression. IEEE Trans. Fuzzy Syst. 2570–2580 (2019)

A Tweet Data Analysis for Detecting Emerging Operational Risks

Davide Di Vincenzo[1(✉)], Francesca Greselin[2], Fabio Piacenza[1,2], and Ričardas Zitikis[3]

[1] Group Non Financial Risks, UniCredit S.p.A, Milan, Italy
{Davide.DiVincenzo,Fabio.Piacenza}@unicredit.eu
[2] Department of Statistics and Quantitative Methods, Univ. Milano Bicocca, Milan, Italy
francesca.greselin@unimib.it, f.piacenza1@campus.unimib.it
[3] School of Mathematical and Statistical Sciences, Western University, London, ON, Canada
rzitikis@uwo.ca

Abstract. Operational risk (OpRisk) is emerging as a crucial non financial consideration with widespread implications for financial institutions. Shifting away from traditional regulatory tasks, including data collection, capital requirement calculations, and report generation for managerial decisions, OpRisk functions are now adopting proactive strategies to prevent or mitigate risks. The integration of Artificial Intelligence techniques, increasingly essential for managerial insights, is utilized to glean additional information from data. This study propels the utilization of text analysis techniques in the context of OpRisk. A pioneering dimension involves examining pertinent tweet content from social media X for the continuous monitoring of the evolving risk landscape, aiming to identify early warnings about new types of potentially risky events.

Keywords: clustering · early warning · emerging OpRisks · natural language processing · operational risk · text analysis · tweets

1 Introduction

The OpRisk is related to the risk of losses resulting from frauds, sanctions, physical damage, IT issues, cyberattacks, and errors [1]. International financial institutions have operational risk management functions, to perform regulatory activities, such as loss data collection. They have databases to collect and store the necessary information for each OpRisk event, such as loss amount, reference date, Basel loss event type, and event description.

The present work addresses the application of text analysis techniques to tweet data to observe the trend of OpRisk topics and detect new emerging OpRisks. Text analysis is one of the main tasks of Natural Language Processing (NLP), a branch of Artificial Intelligence (AI). The entire workflow is represented in Fig. 1.

M. Corazza et al. (Eds.): MAF 2024, *Mathematical and Statistical Methods for Actuarial Sciences and Finance*, pp. 136–142, 2024.
https://doi.org/10.1007/978-3-031-64273-9_23

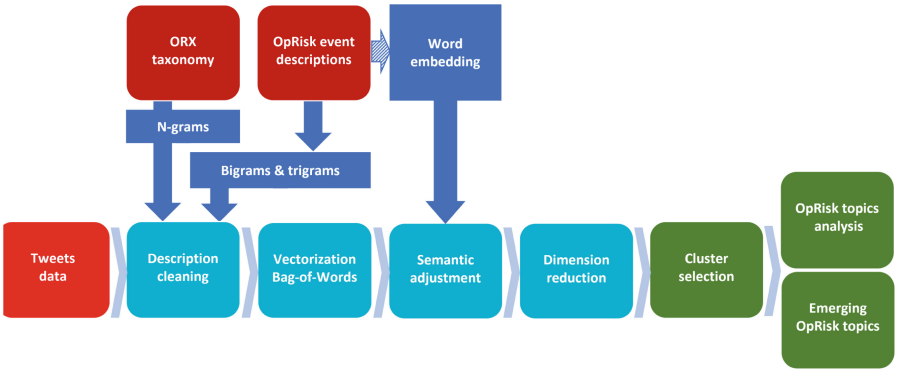


Fig. 1. Workflow for tweet analysis.

2 Tweets Data

Two R scripts based on the package `rtweet` [2] have been written to access and store tweets, and then scheduled to run using the R package `taskscheduleR` [3]:

- Script to extract tweets related to specific keywords (*e.g.*, "fraud", "sanction", "error"), mainly based on the ORX taxonomy [4] scheduled to run every hour. The extraction was performed using the function `search_tweets`.
- Script to extract tweets related to specific accounts for financial news (*e.g.*, Financial Times, Bloomberg, Reuters), scheduled to run every day. The extraction was performed using the function `get_timeline`.

The different schedules of extraction are motivated by the number of tweets expected for each type of search, and by the empirical observation that each extraction cannot exceed around 100,000 tweets. Keyword extraction includes personal account tweets, providing more relevant and earlier information than financial accounts, for particular events (*e.g.*, damages for earthquake or extreme weather conditions). We used a specific API freely available for research activities (after Twitter approval). Since the free API was dismissed when Twitter became X, the available data sets are from May 5^{th} to July 12^{th} for keyword related tweets, and from May 11^{th} to July 11^{th} for account related tweets.

3 Workflow for Tweet Data Analysis

Considering the amount of extracted tweets (around 100,000–150,000 per day), we separately analyze each daily data set. All the steps described within the present section are intended to be applied to each daily data set of tweets.

3.1 Tweet Cleaning

Procedures to clean tweets include the following steps:

- For anonymization, a pre-defined list of most known names, extracted from the R package `gender`, is excluded from tweets.
- Selecting the English-written tweets, setting the functions `search_tweets` and `get_timeline` with the argument `lang='en'`.
- Discarding re-tweets, removing hashtags and web links, ignoring cases, removing punctuations and digits, stop-words, and special characters, reducing words to their lemmas, and removing duplicated tweets.
- Considering n-grams from ORX taxonomy, identifying bigrams-trigrams within OpRisk event descriptions of the financial institution and the tweets based on statistical tests on their frequencies [5].
- Removing all the terms having a frequency lower than 5 within all the tweets.

3.2 Tweet Vectorization and Semantic Adjustment

The tweet data set is transformed into a document-by-term matrix. Each row represents a document (*i.e.*, a tweet), each column represents a term (*i.e.*, a word or an n-gram), and each cell represents the Term Frequency (TF). To consider the semantic similarity between different words, we adopt a word embedding (pre-trained or trained on OpRisk event descriptions) and the related word-similarity matrix similarity (with values higher than 0.8); the value of each "zero" of the document-by-term matrix is updated with the value of the most similar word included in the same row of the matrix, and scaled by the respective word similarity score (refer to [6] for implications and impacts of this adjustment).

3.3 Dimensionality Reduction, Cluster Selection, Topic Analysis, and Emerging Topics Detection

We use UMAP to produce a 2D data representation [7], and seeded LDA to identify topic probability distributions [5] and, consequently, the clustering of the tweets. Seeded LDA enforces specific topics with positive probability only for a set of *seeded tokens*. It requires count data, so it has been applied to the rounded semantic-aware document-by-term matrix and run with 1000 iterations over a burn-in of 500. Seeded topics are based on the ORX taxonomy. Based on seeded LDA results, we assign each document to the cluster related to the topic with the highest probability. In case a document contains one or more seed tokens, the assignment is constrained to the related seeded topics. We observe the tweet daily frequencies for each topic to identify peaks, defined as cases above the 95% quantile of the normal distribution estimated on the time series. Such peaks can represent particular OpRisk events affecting the financial system, the industry, or the governments. OpRisk analysts shall understand which OpRisk events are relevant for their financial institution. To detect emerging OpRisks, we consider five unseeded topics in seeded LDA. Observing the related word

clouds, we can deep dive into the corresponding tweets to understand whether such topics could be early warnings for the financial institution. Early warning systems are extensively used in finance [8]. Furthermore, tweet analysis is often used for predictive purposes [9]. Since OpRisk analysts cannot verify every single tweet, we aim to detect the signal related to many tweets related to a specific topic, including the most frequent words that are represented by word clouds. For each day, it is much easier and faster to look at a few word clouds to spot some OpRisk related words, than reading all the tweets or websites reporting financial news.

4 Application to Tweet Data

We analyse the tweets that have been extracted as explained in Sect. 2 and processed as illustrated in Sect. 3. We apply seeded LDA to each daily data set, considering the seeded topics and the related seed tokens. For each day, we cluster tweets based on the topic with the highest probability (considering the constraint described in Sect. 3.3) and obtain, for each topic, the number of daily tweets. The topic "08_Information_Security" presents a peak (Fig. 2).

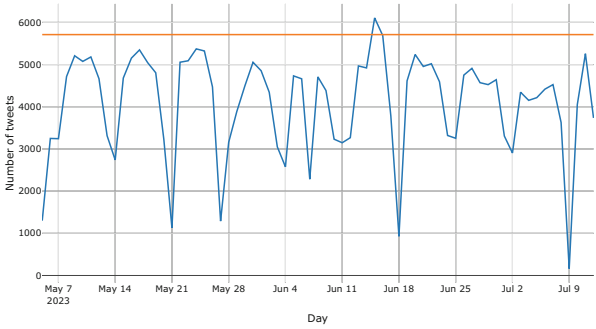


Fig. 2. Number of daily tweets for 08_Information_Security from May 5^{th} to July 12^{th}, where the orange line represents 95% quantile of estimated normal distribution.

The observed peak is related to June 15^{th} and 16^{th}, which report around 1000 more tweets than on other days. We can analyze the UMAP 2D representation including the tweets of June 15^{th}, reported in Fig. 3. Deep diving the interactive version of Figs. 3, it can be observed that several tweets in the clusters related to "08_Information_Security" (*i.e.* the light green ones) are related to a cyberattack on US government agencies [10]. This is a potential early warning for a financial institution, since it is related to a global cyberattack perpetrated by Russian cybercriminals. It could have evolved into a cyber pandemic crisis affecting the financial system. It would have been beneficial for financial institutions to have been informed as soon as possible, to set up the proper prevention initiatives.

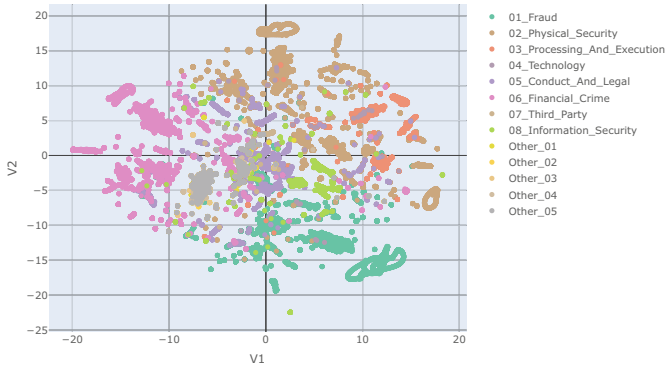


Fig. 3. UMAP 2D representation of tweets related to June 15th 2023.

Peak detection can lead to false negatives. For instance, we detected a peak on June 9th for the topic "08_Financial_Crime" which referred to the alleged bribe accepted by Joe Biden, which is not relevant for financial institutions. To detect emerging new OpRisk topics, it is useful to analyze the word clouds related to the five unseeded topics for each daily result. Inspecting the word cloud of the 4th unseeded topic related to June 15th, the token "severe_thunderstorm_warm" appears in the bottom part of Fig. 4.

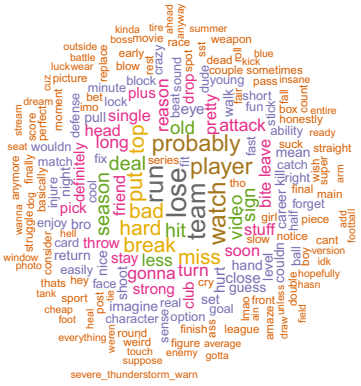


Fig. 4. Word-cloud for the 4th unseeded topic of tweets related to June 15th.

These warnings were related to problematic weather conditions in the Southeast regions of the United States [11]. This early warning could have been relevant for financial institutions having branches in the affected regions, allowing them to prevent or mitigate the damages to employees and buildings.

5 Conclusion

To get early warnings for emerging OpRisk events, we employed recent statistical approaches and models: we utilized seeded LDA for constrained clustering in the analysis of tweets related to OpRisk topics, and UMAP for dimensions reduction. Our refinement of standard text analysis techniques for tweets involved the incorporation of n-grams based on the contemporary ORX taxonomy, as well as relevant bigrams and trigrams. Analyzing daily tweets, a peak related to cyberattacks emerged as a potential early warning for financial institutions. Additionally, the detection of an emerging OpRisk topic concerning severe thunderstorms in Southeast U.S. regions suggests preemptive actions for potential damages. While our proposed framework lays a robust foundation for tweets related to OpRisk topics, further enhancements and extensions are conceivable. Such research directions include:

- Identifying relevant n-grams, with $n > 3$, from OpRisk descriptions and tweets, considering explicitly the multiplicity of the applied statistical tests.
- Incorporating web data, other than tweets, on a wider time window (*e.g.*, social media, such as Threads, and news data providers, such as Talkwalker).
- Considering more advanced techniques to detect significant peaks and trends in the daily number of tweets for each OpRisk related topic. For instance, methods to detect change points can be used [12].
- Applying extensions of LDA (once adapted to include term seeds) able to model the topics dependence, such as Structural Topic Model (STM) [13].
- Extending the analysis to Reputational Risk measurement, since severe OpRisk events (especially, internal frauds) can have a reputational impact, impacting the stock price of the financial institution [14].

References

1. European Parliament and Council of the European Union. Capital Requirement Regulations. Technical report. European Union (2013). https://eur-lex.europa.eu/eli/reg/2013/575/oj
2. Kearney, M.W.: rtweet: collecting and analyzing Twitter data. J. Open Sour. Softw. **4**(42), 1829 (2019). https://doi.org/10.21105/joss.01829, https://joss.theoj.org/papers/10.21105/joss.01829. R package version 0.7.0
3. Wijffels, J., Belmans, O.: taskscheduleR: schedule R scripts and processes with the windows task scheduler. R package version 1.8 (2023). https://CRAN.R-project.org/package=taskscheduleR
4. ORX, Wyman, O.: ORX Reference Taxonomy for operational and non-financial risk – Guidance document. Version 2 (2023). https://orx.org/download/orx-referencetaxonomy
5. Frigau, L., Wu, Q., Banks, D.: Optimizing the JSM program. J. Am. Stat. Assoc. **117**(538), 617–626 (2021). https://doi.org/10.1080/01621459.2021.1978466
6. Di Vincenzo, D., Greselin, F., Piacenza, F., Zitikis, R.: A text analysis of operational risk loss descriptions. J. Oper. Risk **18**(3), pp. 63–90 (2023). https://doi.org/10.21314/JOP.2023.003, https://www.risk.net/journal-of-operational-risk/7957562/a-textanalysis-of-operational-risk-loss-descriptions

7. McInnes, L., Healy, J., Saul, N., Grosberger, L.: UMAP: uniform manifold approximation and projection. J. Open Sour. Softw. **3**(29), 861 (2018). https://doi.org/10.21105/joss.00861 https://doi.org/10.21105/joss.00861
8. Klopotan, I., Zoroja, J., Meško, M.: Early warning system in business, finance, and economics: bibliometric and topic analysis. Int. J. Eng. Bus. Manage. **10** (2018). https://doi.org/10.1177/1847979018797013
9. Cano-Marin, E., Mora-Cantallops, M., Sánchez-Alonso, S.: Twitter as a predictive system: a systematic literature review. J. Bus. Res. **157** (2023). https://www.sciencedirect.com/science/article/pii/S0148296322010268
10. CNN. Exclusive: US government agencies hit in global cyberattack (2023). https://edition.cnn.com/2023/06/15/politics/us-government-hit-cybeattack/index.html
11. National Weather Service. June 15, 2023 Severe Weather & Tornadoes (2023). https://www.weather.gov/cle/event20230615severe
12. Chen, J., Gupta, A.K.: Parametric Statistical Change Point Analysis: With Applications to Genetics, Medicine, and Finance, 2nd edn. Springer, Boston (2012). https://doi.org/10.1007/978-0-8176-4801-5, https://cds.cern.ch/record/1498647
13. Roberts, M.E., Stewart, B.M., Tingley, D., Airoldi, E.M.: The structural topic model and applied social science. In: International Conference on Neural Information Processing (2013). https://api.semanticscholar.org/CorpusID:59893873
14. Perry, J., de Fontnouvelle, P.: Measuring reputational risk: the market reaction to operational loss announcements. In: Working Paper Current Draft: October 2005. Federal Reserve Bank of Boston (2005). https://papers.ssrn.com/sol3/papers.cfm?abstractid=861364

Multipopulation Mortality Modeling with Economic, Environmental and Lifestyle Variables

Matteo Dimai(✉)

Università degli Studi di Trieste, 34127 Trieste, Italy
matteo.dimai@phd.units.it

Abstract. This study introduces an innovative approach to mortality modeling by incorporating external covariates into a multipopulation framework. Unlike traditional extrapolative models, this method acknowledges the multifaceted influences on mortality, including environmental, economic, and lifestyle factors. By integrating these variables into the Lee-Carter framework, the study aims to improve the fit of mortality models, allowing for scenario building for policy planning and risk assessment.

Keywords: multipopulation · mortality · GDP · alcohol · tobacco

1 Introduction

In recent decades, global life expectancy has seen a significant rise, emphasizing the need for accurate mortality forecasts. Traditional models, primarily based on past trends, often fail to capture the dynamic interplay of factors influencing mortality. This study proposes a model that broadens the scope of mortality determinants, incorporating economic, technological, environmental, and lifestyle variables to improve the accuracy and applicability of mortality forecasts.

1.1 Literature Review

Mortality forecasting encompasses three main approaches [9]: extrapolation, which extends current mortality trends into the future; explanation, which models mortality using external variables linked to health outcomes; and expectation, based on expert opinions. While often blended, these methods have distinct focuses. The present study leverages the established link between mortality and economic development [2, 7] to build a multipopulation stochastic mortality model within the Li-Lee framework [5] drawing from the single-population GDP-based model [6] and its multipopulation extension GDP-LL model [1], widening the scope of variables included in the model following other similar, but more limited attempts [3, 4, 8].

Proposal by Antoine Parent (U. Paris 8, OFCE - Sciences Po & CAC - IXXI) of organized session: "Cliometrics, finance and actuarial sciences"; authors: Gilles Gaba, Antoine Parent, Stéphane Loisel, Vincent Touzé, Laurent Gauthier, Matteo Dimai).

M. Corazza et al. (Eds.): MAF 2024, *Mathematical and Statistical Methods for Actuarial Sciences and Finance*, pp. 143–148, 2024.
https://doi.org/10.1007/978-3-031-64273-9_24

2 Methodology

The research aims to formulate an interpretable mortality forecasting model, called MEV (Multipopulation model with External Variables) utilizing various population-level variables. This model, applied to male mortality only for brevity, extends beyond conventional economic indicators like GDP to include fossil fuel use, consumption of fruit and vegetables, alcohol and tobacco, share of adults with raised blood pressure, height of men at age 18, caloric supply per capita and temperature anomalies. The gendered covariates are available for females as well and female mortality, exhibiting the same trends, although with overall lower death rates, especially at younger ages. Data sources include the Human Mortality Database for mortality data and for the external variables WHO, FAO, NCD-RisC, Penn World Tables, HadCRUT4 for temperature data, Wine Research Centre of the University of Adelaide, the International Cigarette Consumption Database and BP Statistical Review of World Energy. The multipopulation model for mortality rates m for country i, age x and year t is as follows:

$$log\left(m_{x,t,i}\right) = A_x + \sum_{j=1}^{J} B_{j,x}K_{j,t} + \sum_{l=1}^{L} c_{l,x}g_{l,t} + a_{x,i} + b_{x,i}k_{t,i} + \epsilon_{x,t,i}. \quad (1)$$

The estimation is adapted from the model in [1]. The J age-period terms $B_{j,x}K_{j,t}$ are estimated at the group level, as well as the L age loadings $c_{l,x}$ and the common age terms A_x. The terms $g_{l,t}$ are linear combinations of O external variables $h_{o,t}$ obtained through singular value decomposition. The $L < O$ principal components which explain each a share of variance greater or equal to a set threshold are retained. After the estimation at the group level, a country-level estimation produces the country intercepts $a_{x,i}$ and the country-specific stationary age-period terms $b_{x,i}k_{t,i}$. In the applications, $J = 0$ or $J = 1$. The model requires additional constraints on the parameters to be identified: most notably, the $g_{l,t}$ terms need to be uncorrelated, which makes the use of principal components a prerequisite.

This work builds upon the author's previous research in single-population models with external variables, of which the multipopulation model is an extension. The criteria for multipopulation grouping have also been derived from the author's previous work on hierarchical clustering of mortality paths through the Hellinger distance, with a dynamic cut-off height between 0.05 and 0.10 for the average distance and using the complete agglomeration method (Table 1).

The model was applied to 27 countries from the Human Mortality Database, organized into seven clusters based on the relative distance of their respective mortality paths.

Table 1. Multipopulation groups, hierarchical clustering Hellinger distance cut-off and main characteristics in the central year of the grouping period (1960–2019).

Countries	Distance cut-off	Life expectancy at birth in 1990	Residual life expectancy at 65 in 1990	Average within-cluster distance
Estonia, Latvia, Lithuania	0,1171	70.249	14.712	0.0565
Bulgaria, Czechia, Hungary, Poland, Slovakia	0,1171	70.721	14.037	0.0558
Denmark, UK, USA	0,0902	75.333	16.480	0.0540
Austria, Belgium, Finland, West and East Germany, Ireland, Portugal	0,0902	75.024	16.015	0.0511
Netherlands, Norway, Sweden	0,0514	77.062	16.942	0.0286
Italy, Spain, Switzerland	0,0514	77.125	17.408	0.0327
Australia, Canada, France	0,0646	76.684	17.403	0.0408

3 Results

3.1 Goodness of Fit

The model's coverage ranges from 1975–2014, varying by cluster, depending on covariate availability. The final year is 2014 except for the cluster with Australia, Canada and France (2010). The starting year is 1996 for the Baltic cluster, 1993 for the Eastern European one and 2000 for the cluster including Austria, Belgium, Finland, West and East Germany, Ireland and Portugal.

Goodness of fit for the MEV model, and to a lesser degree for the GDP-LL model, is less impacted by the number of years used to estimate it than is the case with the Li-Lee model. The residuals in both absolute and percentage terms are presented for a low-mortality cluster (Netherlands, Norway, Sweden) covering the 1975–2014 period: the model appears to fit the data well, without patterns (vertical, horizontal or diagonal lines) or clusters, except for a small cohort effect for WW2 cohorts in the Netherlands (Fig. 1).

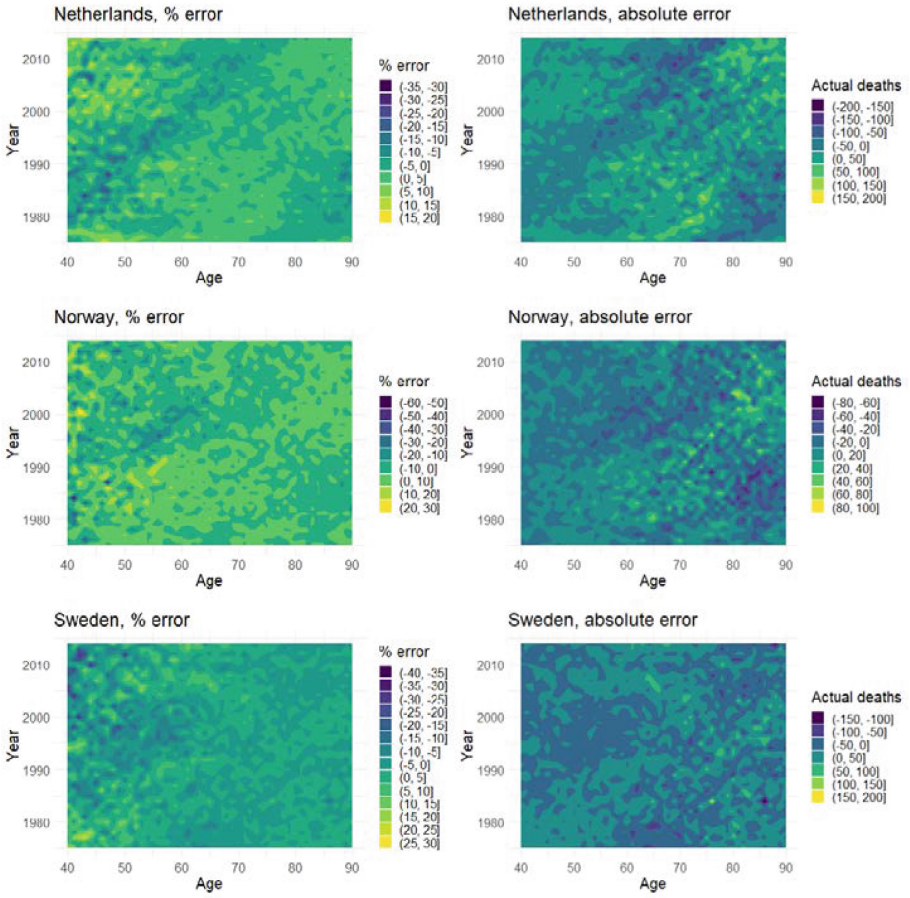


Fig. 1. Residuals by year and age in percentage and absolute terms for a low mortality cluster.

Goodness of fit is assessed using the estimation ratios $R_C(i)$ and $R_{AC}(i)$ [1, 5], which describe how well the common terms and the complete model, respectively, explain variation in the data. The MEV model, with mean $R_C(i)$ and $R_{AC}(i)$ of 0.926 and 0.965, outperforms both the GDP-LL (0.830 and 0.918) and the Lee-Carter (0.808 and 0.888). Additionally, the MEV model is assessed in terms of mean absolute deviation (MAD), mean absolute percentage deviation (MAPE) and Bayesian Information Criterion (BIC) against the GDP-LL model. Out of 33 countries, the MEV shows a lower MAD for 19 countries, a lower MAPE for 20 and a lower BIC for 13. Mean MAPE is 4,01% for the MEV model and 4,03% for the GDP-LL model.

Overall, the MEV model outperforms the GDP-LL model where the latter's fit is not very high and where the Li-Lee fit isn't optimal either, like i.e. in Austria, Bulgaria, Ireland or in the Baltic countries. Otherwise, the two models are similar.

3.2 Stationarity and Cointegration

Time series need to be stationary or cointegrated in order to obtain reliable and meaningful long-term estimates. While the order of integration varies between variables and countries, all variables except temperature anomalies are non-stationary according to at least one between the KPSS and Phillips-Perron test for almost all countries. Testing for cointegration with the k_t time term of the Lee-Carter model reveals at least one cointegration relationship at the 5% level except for Finland and Denmark.

After estimating the multipopulation model, the country-specific k_t residual time trend has been modeled as a stationary AR(1) process in order to ascertain whether the group-level terms account for the stochastic trend in its entirety: residual country-specific time terms need to be stationary in order to ensure coherent forecasts. In no country the estimated autoregressive term was greater or equal to 1, hence the residual country-specific terms are stationary for all countries considered.

4 Discussion

Extrapolative models rely on past trends to estimate future ones without questioning their validity. The inclusion of external variables fits the data better, allows for scenario-based predictions with explicit assumptions and in general turns mortality from an exogenous variable to a possibly endogenous part of more complex models. The multipopulation approach is advantageous for its accuracy, benefiting from shared trends across countries, reducing noise, especially for shorter time frames. The capacity to derive meaningful estimates from just a few years of data proves especially valuable in situations where the relationship between covariates and mortality shifts, as can occur with the onset of new diseases or global catastrophic events. However, it requires comparable and consistent data across all included populations.

This study presents a significant step forward in mortality modeling, offering a versatile tool for scenario planning and policy development. While this model focuses on groups with a common mortality trend, future research should focus on how to model countries with similar covariate effects on mortality, but with divergent trends, obtaining estimates that are coherent conditionally on the values of the covariates.

References

1. Boonen, T.J., Li, H.: Modeling and forecasting mortality with economic growth: a multipopulation approach. Demography **54**, 1921–1946 (2017)
2. Brenner, M.H.: Commentary: economic growth is the basis of mortality rate decline in the 20th century—experience of the United States 1901–2000. Int. J. Epidemiol. **34**, 1214–1221 (2005)
3. Dutton, L., Pantelous, A.A., Seklecka, M.: The impact of economic growth in mortality modelling for selected OECD countries. J. Forecast. **39**, 533–550 (2020)
4. French, D., O'Hare, C.: Forecasting death rates using exogenous determinants. J. Forecast. **33**, 640–650 (2014)
5. Li, N., Lee, R.: Coherent mortality forecasts for a group of populations: an extension of the Lee-Carter method. Demography **43**, 575–594 (2005)

6. Niu, G., Melenberg, B.: Trends in mortality decrease and economic growth. Demography **51**, 1755–1773 (2014)
7. Preston, S.H.: The changing relation between mortality and level of economic development. Popul. Stud. **29**, 231–248 (1975)
8. Seklecka, M., Md. Lazam, N., Pantelous, A.A., O'Hare, C.: Mortality effects of economic fluctuations in selected eurozone countries. J. Forecast. **38**, 39–62 (2019)
9. Stoeldraijer, L., van Duin, C., van Wissen, L., Janssen, F.: Impact of different mortality forecasting methods and explicit assumptions on projected future life expectancy: the case of the Netherlands. Demogr. Res. **29**, 323–354 (2013)

Bayesian Modeling of Mortality in Italian Regions: A Three-Component Approach Incorporating Cohort Effects

Matteo Dimai[1(✉)] and Marek Brabec[2]

[1] Department of Economics, Business, Mathematics and Statistics, University of Trieste, Via Valerio n. 4/1, 34127 Trieste, Italy
matteo.dimai@phd.units.it

[2] Institute of Computer Science, The Czech Academy of Sciences, Pod Vodárenskou věží 271/2, 182 00 Praha 8, Czech Republic
mbrabec@cs.cas.cz

Abstract. The increases in life expectancy over the last decades have strongly impacted the distribution of ages at death. Its parametric estimation can be complicated by cohort effects. Our addresses the issue by extending a recent three-component parametric model to include cohort effects in a Bayesian framework. The model is fit to male mortality data from five diverse Italian regions between 1974 and 2022. Our results demonstrate significant regional variations in mortality, influenced by cohort effects, particularly among cohorts born around World War I. The model effectively captures the evolution of mortality components, with cohort effects markedly improving fit of complex, even multi-modal curves.

Keywords: mortality · skew-normal · Bayesian · Italy · cohort effect

1 Introduction

In mortality studies, multi-component parametric models, such as the classic three-component model [9], typically assume smooth age-at-death distributions. However, this smoothness can be disrupted by cohort effects - unique variations tied to specific birth cohorts [14]. These effects can lead to inaccuracies in traditional models if they are not properly included. Our study aims to improve the fit of these models by incorporating cohort effects into a recent three-component parametric model [15], using a Bayesian approach. This modification helps capture the real-world irregularities in mortality data that are often missed when cohort effects are overlooked.

M. Corazza et al. (Eds.): MAF 2024, *Mathematical and Statistical Methods for Actuarial Sciences and Finance*, pp. 149–153, 2024.
https://doi.org/10.1007/978-3-031-64273-9_25

1.1 Literature Review

Mortality modeling encompasses a diverse range of approaches, reflecting the variety in both the aspects of mortality being modeled, like death rates [8], life expectancy [11] or odds-ratios [6], and the methods employed, from parametric forms [5] to non-parametric estimates [8], intermediate approaches [7] and neural networks [13]. Among those, classical parametric forms continue to be relevant due to their ease of interpretation [15]. Several models have been developed within a Bayesian framework, mainly based on death rates [1, 4] or life expectancy [11].

Some birth cohorts experience higher or lower mortality throughout their lives, leading to cohort effects. These have been recognized and incorporated into various models [10, 12], improving fit and forecasts. However, the integration of cohort effects within Bayesian frameworks remains relatively rare [3]. This gap highlights the innovative aspect of our study, which brings together the Bayesian methodology with a focused consideration of cohort effects.

2 Model and Data

The study analyzes male mortality data from five Italian regions (Lombardia, Lazio, Sicily, Sardinia, and Friuli Venezia Giulia) from 1974 to 2022 for ages 0 to 100. The cohort effect strength, most evident for the 1915–1925 cohorts for which it is modeled, varies substantially between the regions. The model extends the three-component one by Zanotto et al. [15] and applies it in a Bayesian framework. The density distribution of ages at death, unadjusted for cohort effects, is the following mixture, with θ being the vector of parameters, x the age and t the year:

$$f(x;\theta,t) = \zeta_{1,t} \cdot f_I(x;\sigma_I) + \zeta_{2,t} f_{m^*}\left(x;\mu_{m,t},\sigma_{m,t},\gamma_{m,t}\right) + \zeta_{3,t} f_{M^*}\left(x;\mu_{M,t},\sigma_{M,t},\gamma_{M,t}\right). \tag{1}$$

The infant mortality component is a half-Normal with mean 0 and variance σ_I, while the premature mortality component f_{m^*} and the adult mortality component f_{M^*} are Skew-Normals with the centered parametrization [15], both normalized by dividing them by the probability of them assuming a value between 0 and 100. $\zeta \sim Dirichlet(3)$ is a mixing parameter with 2 degrees of freedom. Since the age is a discrete variable, the normalized density is used to approximate the probability of dying at a given age, resulting in faster estimation and no noticeable effect on parameter estimates. The model is then adjusted to account for multiplicative cohort effects α_{t-x}, with $t-x$ being the birth cohort, such that the probability of a person born in year $t-x$ dying at age x in year t is:

$$P(x;\theta,t,\alpha) = \frac{\alpha_{t-x} f(x;\theta,t)}{\sum_{x=0}^{100} \alpha_{t-x} f(x;\theta,t)}. \tag{2}$$

The prior distributions are $\mu_{m,t} \sim N(30, 5)$ and $\mu_{M,t} \sim N(80, 10)$ for the two Skew-Normal means, $\sigma_{m,t}, \sigma_{M,t} \sim Inv.\Gamma(0.001, 0.001)$ for the Skew-Normal variances, $\gamma_{m,t}, \gamma_{M,t} \sim N(0, 1)$ for the skewness parameters, $\zeta_t \sim Dir(1, 1, 8)$ for the mixing parameters and $log(\alpha_{t-x}) \sim U(-0.7, 0.4)$, corresponding to cohort effects from -50% to $+50\%$. The priors have been chosen in order to reduce both the risk of label switching and the impact on posterior distributions.

3 Results

Sampling has been performed in Stan with 4 chains in order to assess convergence, 1,000 warm-up iterations and 1,000 sampling iterations. Cohort effects have been estimated for the 1915–1925 cohorts.

3.1 Goodness of Fit

The age at death curves for the regions studied are frequently difficult to model, with nonlinear effects and, in some cases, even multiple modes. As evidenced in Fig. 1, the inclusion of cohort effects allows for a better fit compared to the three-component model without cohort effects.

Cohort effects vary substantially in strength between regions. For Lombardia, they range from +9,0% (1917) to −8,6% (1920), while in Sardinia the effect is substantially lower, ranging from +3,0% (1916) to −4,0% (1920).

3.2 Convergence

Convergence is achieved for all regions for the premature and adult mortality components, with a Gelman-Rubin statistic $\widehat{R}$ lower than 1,1 for the relevant parameters. For the infant mortality component, convergence is achieved in all 49 years for Lombardia and Sicily. For Friuli Venezia Giulia, Lazio and Sardinia convergence of the mixing parameter ζ_1 is achieved for 48, 49 and 47 years, respectively, and of the variance parameter σ_I for 47, 47 and 44 years. We deem this to be a minor issue. Since infant mortality is a specific component with no risk of label switching, the issue can be easily resolved by either setting the σ_I parameter to a constant [15] or by substituting it with a point probability [2], with a subsequent loss of fit.

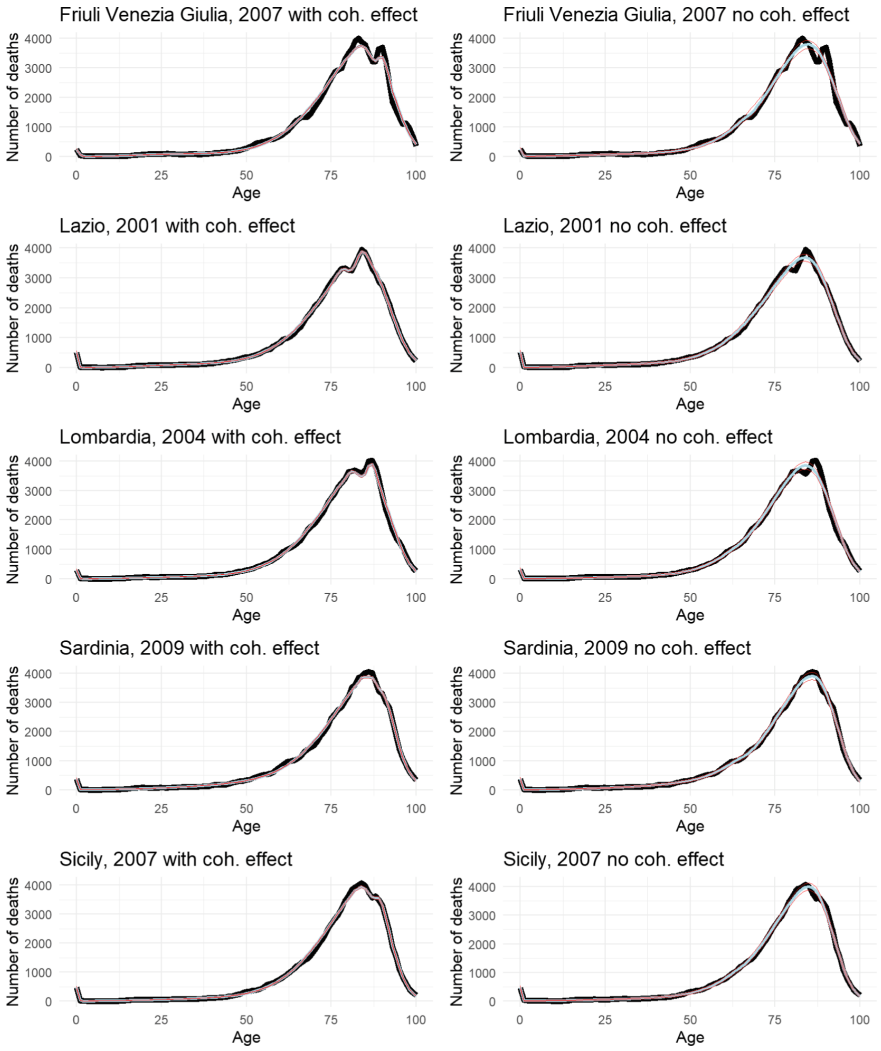


Fig. 1. Actual (black) and fitted (blue) life table deaths, with 95% credible intervals (red) by age and region, selected years, mixture model with cohort effect (left) and without (right)

4 Discussion

The Bayesian three-component model presented shows a good fit to the complex curve of age of death for the Italian regions analyzed. The presence of significant cohort effects has been confirmed, with varying strength across regions, and their inclusion in the model markedly improves fit. As far as we are aware, this is the first time cohort effects have been included in a Bayesian mixture model focused on the age at death. Future developments include leveraging the full power of Bayesian models by calculating quantities from the

posterior that would be impractical to calculate in a MLE setting, like the distribution of mortality indicators such as life expectancy or conditional standard deviation.

References

1. Alexander, M., Zagheni, E., Barbieri, M.: A flexible Bayesian model for estimating subnational mortality. Demography **54**, 2025–2041 (2017)
2. Aliverti, E., Mazzuco, S., Scarpa, B.: Dynamic modelling of mortality via mixtures of skewed distribution functions. J. R. Stat. Soc. Ser. A **185**, 1030–1048 (2022)
3. Fung, M.C., Peters, G.W., Shevchenko, P.V.: Cohort effects in mortality modelling: a Bayesian state-space approach. Ann. Actuarial Sci. **13**, 109–144 (2019)
4. Girosi, F., King, G.: Demographic Forecasting, 1st edn. Princeton University Press, Princeton and Oxford (2008)
5. Gompertz, B.: XXIV. On the nature of the function expressive of the law of human mortality, and on a new mode of determining the value of life contingencies. In a letter to Francis Baily, Esq. FRS &C. Philos. Trans. Roy. Soc. London **115**, 513–583 (1825)
6. Heligman, L., Pollard, J.H.: The age pattern of mortality. J. Inst. Actuaries **107**, 49–80 (1980)
7. Hunt, A., Blake, D.: A general procedure for constructing mortality models. North Am. Actuarial J. **18**, 116–138 (2014)
8. Lee, R.D., Carter, L.R.: Modeling and forecasting US mortality. J. Am. Stat. Assoc. **87**, 659–671 (1992)
9. Pearson, K.: Chances of Death. Edward Arnold, London (1897)
10. Plat, R.: On stochastic mortality modeling. Insur.: Math. Econ. **45**, 393–404 (2009)
11. Raftery, A.E., Chunn, J.L., Gerland, P., Ševčíková, H.: Bayesian probabilistic projections of life expectancy for all countries. Demography **50**, 777–801 (2013)
12. Renshaw, A.E., Haberman, S.: A cohort-based extension to the Lee-Carter model for mortality reduction factors. Insur.: Math. Econ. **38**, 556–570 (2006)
13. Scognamiglio, S.: Calibrating the Lee-Carter and the Poisson Lee-Carter models via neural networks. ASTIN Bull. **52**, 519–561 (2022)
14. Willets, R.C.: The cohort effect: insights and explanations. Br. Actuar. J. **10**, 833–877 (2004)
15. Zanotto, L., Canudas-Romo, V., Mazzuco, S.: A mixture-function mortality model: illustration of the evolution of premature mortality. Eur. J. Popul. **37**, 1–27 (2021)

Forecast Model of the Price of a Product with a Cold Start

Svitlana Drin[1,2(✉)] and Nataliya Shchestyuk[1,2]

[1] School of Business, Department of Statistics, Orebro University, 70182 Orebro, Sweden
svitlana.drin@oru.se

[2] Faculty of Informatics, Department of Mathematics, National University of Kyiv-Mohyla Academy, Kyiv 04070, Ukraine

Abstract. This article presents a comprehensive study on developing a predictive product pricing model using LightGBM, a machine learning method optimized for regression challenges in situations with limited historical data. It begins by detailing the core principles of LightGBM, including gradient descent, and then delves into the method's unique features like Gradient-based One-Side Sampling (GOSS) and Exclusive Feature Bundling (EFB). The model's efficacy is demonstrated through a comparative analysis with XGBoost, highlighting Light-GBM's enhanced efficiency and slight improvement in prediction accuracy. This research offers valuable insights into the application of LightGBM in developing fast and accurate product pricing models, crucial for businesses in the rapidly evolving data landscape.

Keywords: GBM · GBDT · LightGBM · GOSS · EFB · predictive model

1 Introduction

In the current fast-changing business environment, companies face significant challenges in managing their sales strategies and determining the best pricing strategies for their products. One of the most complex aspects of this task is setting the price for a new product that lacks historical sales data.

Developing precise and dependable pricing strategies for new products has given rise to advanced predictive models. These models are essential for companies seeking to determine the best prices without prior experience. Of these models, the LightGBM (Light Gradient Boosting Machine) algorithm is particularly effective in rapidly identifying the optimal forecast. This article employs the LightGBM approach to construct a predictive pricing model for a novel product, designed to address the intricate obstacles of contemporary markets, such as rapidly changing consumer trends and intense market competition.

The LightGBM model stands out due to its efficient handling of large-scale data and its capacity for fast processing, making it particularly suited

M. Corazza et al. (Eds.): MAF 2024, *Mathematical and Statistical Methods for Actuarial Sciences and Finance*, pp. 154–159, 2024.
https://doi.org/10.1007/978-3-031-64273-9_26

for dynamic market conditions. It allows businesses to quickly adapt their pricing strategies in response to real-time market changes, ensuring competitiveness and relevance.

2 Preliminary Theoretical Base

The **Gradient Descent** method is based on the idea that if the function of multiple variables $F(x)$ is defined and differentiable in the vicinity of point a, then $F(x)$ decreases fastest by moving from a in the direction of the negative gradient of F at a, denoted as $-\nabla F(a)$ (where ∇ represents the gradient, a vector of partial derivatives of the function). This implies that if $a_{n+1} = a_n - \gamma \nabla F(a_n)$ for a sufficiently small step size (or learning rate) $\gamma \in R_+$, then $F(a_n) \geq F(a_{n+1})$.

In other words, $\gamma \nabla F(a)$ is subtracted from a because we aim to move against the gradient towards a local minimum. With this understanding, we initiate with x_0 as an assumption for a local minimum of F, considering the sequence $x_0, x_1, x_2, \ldots$ such that $x_{n+1} = x_n - \gamma_n \nabla F(x_n)$, for $n \geq 0$.

As a result, we obtain $F(x_0) \geq F(x_1) \geq F(x_2) \geq \ldots$, and we expect the sequence x_n to converge towards a local minimum. It's worth noting that the step size γ can be adjusted at each iteration.

The **Gradient Boosting Machine** (GBM), a prominent technique in machine learning, is primarily employed for addressing both classification and regression challenges. This method enhances predictive accuracy by integrating several simpler models. Unique to GBM is its iterative model-building approach, a characteristic shared with other boosting techniques. Its versatility is further highlighted by its capability to optimize a wide range of differentiable loss functions.

Often, GBM is implemented in tandem with decision trees, leading to the creation of the Gradient Boosting Decision Trees (GBDT) framework. This integration positions GBM as an adaptive ensemble method variant. In the GBDT context, decision trees function as the primary estimators, a concept extensively discussed in Friedman's foundational paper [1], which delves into the intricacies and applications of gradient boosting in machine learning.

2.1 LightGBM

LightGBM is a type of Gradient Boosting Decision Trees (GBDT) that was developed by a team of researchers at Microsoft in 2016. This model was created to improve upon the popular XGBoost model, which is known for its speed and reliability in multi-class classification projects.

The reason for enhancing XGBoost was to achieve even greater efficiency and faster implementation. The most computationally intensive task in GBDT is the search for optimal split points, which is directly proportional to both the number of features and the number of instances. This leads to speed-related issues when dealing with large datasets. To address this, two new techniques were introduced: Gradient-based One-Side Sampling (GOSS) and Exclusive Feature Bundling (EFB). These techniques aimed to reduce the number of data instances

and functions, thereby mitigating the computational challenges associated with GBDT training on large datasets.

Gradient-Based One-Side Sampling (GOSS)
In GBDT, there is no individual weight for each data instance, but LightGBM that instances with different gradients have varying impacts on information gain calculations. Specifically, instances with higher gradients (less trained samples) exert a greater influence on calculating information gain.

To balance the effect of data distribution, GOSS introduces a constant multiplier for examples with smaller gradients, as presented in the algorithm "Gradient-based One-Side Sampling" [2], compensating for their contribution to the distribution.

More theoretically, GBDT utilizes decision trees to learn functions from the input space χ^s to the gradient space G. Assuming a training dataset of n instances $\{x_1, x_2, \ldots, x_n\}$, where each x_i is a vector of dimension s in χ^s. During each gradient boosting iteration, we compute negative gradients of the loss function with respect to the model's predictions, denoted as $\{g_1, g_2, \ldots, g_n\}$.

Definition. Let O be the training dataset on a fixed node of the decision tree. The variance gain of splitting feature j at point d for this node is defined as

$$V_{j|O}(d) = \frac{1}{n_O}\left(\frac{\left(\sum_{\{x_i \in O: x_{ij} \le d\}} g_i\right)^2}{n^j_{l|O}(d)} + \frac{\left(\sum_{\{x_i \in O: x_{ij} > d\}} g_i\right)^2}{n^j_{r|O}(d)}\right),$$

where $n_O = \sum I[x_i \in O]$, $n^j_{l|O} = \sum I[x_i \in O : x_{ij} \le d]$, and $n^j_{r|O} = \sum I[x_i \in O : x_{ij} > d]$.

In GBDT, the decision tree algorithm chooses $d^*_j = \text{argmax}_d V_j(d)$ for feature j and computes the maximum gain $V_j(d^*_j)$. The data is then split based on feature j^* at point d^*_j into left and right child nodes.

In the novel GOSS algorithm a subset A is first formed by selecting the top $a \times 100\%$ of instances with higher gradients. A random subset B is then sampled from the remaining instances with lower gradients. The instances from subsets $A \cup B$ are split based on the estimated variance reduction $\widetilde{V}_j(d)$.

For a subset A_l, A_r, B_l, and B_r defined as described, the estimated variance reduction is given by:

$$\widetilde{V}_j(d) = \frac{1}{n}\left(\frac{\left(\sum_{x_i \in A_l} g_i + \frac{1-a}{b}\sum_{x_i \in B_l} g_i\right)^2}{n^j_l(d)} + \frac{\left(\sum_{x_i \in A_r} g_i + \frac{1-a}{b}\sum_{x_i \in B_r} g_i\right)^2}{n^j_r(d)}\right),$$

where a and b are constants, and the coefficient $\frac{1-a}{b}$ is used for normalization.

Additionally, the GOSS method is supported by the following theorem:

Theorem. Let $\mathcal{E}(d) = \left|\widetilde{V}_j(d) - V_j(d)\right|$ represent the approximation error in GOSS. With probability at least $1 - \delta$, we have:

$$\mathcal{E}(d) \le C^2_{a,b} \ln\left(\frac{1}{\delta}\right) \cdot \max\left\{\frac{1}{n^j_l(d)}, \frac{1}{n^j_r(d)}\right\} + 2DC_{a,b}\sqrt{\frac{\ln\left(\frac{1}{\delta}\right)}{n}},$$

where $C_{a,b} = \frac{1-a}{\sqrt{b}} \cdot \max_{x_i \in A^c} |g_i|$ and $D = \max\{\bar{g}^j_l(d), \bar{g}^j_r(d)\}$.

Exclusive Feature Bundling (EFB). High-dimensional data often contain numerous features, which can lead to model overfitting. Sparsity in feature space is a common phenomenon.

Sparsity implies that many features are mutually exclusive, meaning they do not have non-zero values simultaneously. This allows us to group them into *"exclusive feature bundles"*. A scanning algorithm enables the construction of histograms for these bundles, instead of individual features, reducing the histogram construction complexity from $O(\#data \times \#feature)$ to $O(\#data \times \#bundle)$, where $\#bundle$ is significantly smaller than $\#feature$.

The EFB algorithm, which is introduced by the algorithm "Merge Exclusive Features" in [2], can consolidate numerous exclusive features into a significantly smaller set of dense features, enabling efficient avoidance of unnecessary computations for zero feature values.

In the study by Dunbray et al. [4], the detailed algorithm integral to the LightGBM method is thoroughly outlined.

3 Practical Implementation

3.1 EDA and Pre-processing

The model was created as a way to solve classification problems. In this article, we will examine how the model works with the problem of forecasting commodity prices without history and for a dataset with non-numerical data.

The dataset is available on the Kaggle website under the name "Mercari Price Suggestion Challenge". Established in 2013, Mercari Inc. is a Japanese company that operates one of the most popular C2C marketplaces in the Japanese market.

The data is already divided into training and testing sets. The dataset comprises the following seven characteristics: "name", "item_condition_id", "brand_name", "category_name", "shipping", "item_description" and "price".

Our initial model looks like this:

$$y = a_0 + a_1 X_1 + \cdots + a_n X_n + \varepsilon$$

where: y is the dependent variable (price), $X_1, \ldots, X_n$ are the independent variables (features), $a_0, a_1, \ldots, a_n$ are the coefficients associated with each independent variable, ε represents the error term, which accounts for the variability in y that is not explained by the model.

During the exploratory analysis and preparation of the data for work, it was chosen log(y) as the dependent variable, because after constructing the histograms of the distribution, it was found that the logarithmic distribution is the closest to the normal one.

Also, the general category was divided into three separate subcategories, missing values were processed. It's worth noting that since most of our characteristics are text, CountVectorizer, TfidfVectorizer were used to convert them to numeric values.

So, now we get a semi-logarithmic eight-factor model

$$log(y) = a_0 + a_1X_1 + a_2X_2 + a_3X_3 + a_4X_4 + a_5X_5 + a_6X_6 + a_7X_7 + a_8X_8 + \varepsilon.$$

3.2 Model Training

In this section, we delve into the practical implementation of the LightGBM framework for price prediction.

The study was conducted in the `Python` environment using the `Params` function. For the 'objective' parameter, we designate 'regression' to align with our regression task. Additionally, we specify the boosting type as 'gbdt', which is the default setting. We include the 'data_sample_strategy' as our GOSS method, which is known for its effectiveness in dealing with large datasets. Furthermore, we activate the 'enable_bundle' option to indicate our utilization of the Exclusive Feature Bundling (EFB) technique. The chosen evaluation metric is 'RMSE', reflecting the Root Mean Squared Error.

This configuration enables us to leverage the strengths of LightGBM for accurate and efficient price prediction.

3.3 Evaluation

To evaluate the performance of the developed model, we will use the metrics R^2, Mean Absolute Error (MAE), and Root Mean Squared Error (RMSE) [3], where the Mean Absolute Error (MAE).

For comparison, we also implemented the XGBoost model. The obtained comparative table presents model evaluations based on training for 1000 iterations using the provided training data (Table 1):

Table 1. Model Performance Comparison

Metric	LGBM	XGBoost
RMSE	0.47667	0.47778
MAE	0.35779	0.35897
R-squared	0.59461	0.59274
Time, s	810.25	2116.69

In order to show that the distribution of the real price and the model are quite similar, we visualized the entire sample, 1000 items and 100 items, which is the most representative. Where real prices are shown in blue, predicted by LightGBM in pink and predicted by XGBoost in green (Fig. 1).

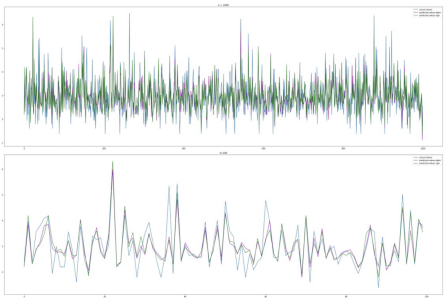


Fig. 1. Distributions of prices

4 Conclusion

Through our work, we have successfully developed a model that can predict the price of a product without a history. Our model is based on the characteristics of the product and data on similar products found in the vicinity, and we utilized the LightGBM method. We used a real data set from Mercari, which is one of the most popular C2C marketplaces in Japan, to showcase the novelty of the LightGBM method. Pseudo-codes were provided to demonstrate how these new algorithms work.

Our findings showed that the LightGBM method provides highly accurate predictions for product prices without history, and is significantly faster compared to other models. This study confirms the efficacy of using the LightGBM method for price prediction. The results of our study can be valuable for companies looking to develop automated recommender systems for a historical commodity price forecasting model.

Acknowledgement. Svitlana Drin acknowledges financial support from the Knowledge Foundation Grant (Dnr: 20220115).

Nataliya Shchestyuk acknowledges financial support from the Knowledge Foundation Grant (Dnr: 20220099).

References

1. Friedman, J.H.: Greedy function approximation: a gradient boosting machine. Ann. Stat. 1189–1232 (2001)
2. Ke, G., et al.: LightGBM: a highly efficient gradient boosting decision tree. In: Advances in Neural Information Processing Systems, vol. 30 (2017)
3. Chicco, D., Warrens, M.J., Jurman, G.: The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation. PeerJ Comput. Sci. **7**, e623 (2021)
4. Dunbray, N., Rane, R., Nimje, S., Katade, J., Mavale, S.: A novel prediction model for diabetes detection using gridsearch and a voting classifier between LightGBM and KNN. In: 2021 2nd Global Conference for Advancement in Technology (GCAT), pp. 1–7. IEEE (2021)

Clustering and Testing Financial Asset Returns Using the Spatial Dynamic Panel Data Model

Giuseppe Feo(✉), Francesco Giordano, Sara Milito, Marcella Niglio, and Maria Lucia Parrella

University of Salerno, 84084 Fisciano, SA, Italy
{gfeo,giordano,smilito,mniglio,mparrella}@unisa.it

Abstract. In this work, we consider a spatio-temporal financial dataset, with $p = 84$ units (a sample of asset returns from the "G7" countries) and $T = 666$ daily observations, from May 2021 to December 2023. Assuming that the p units can be grouped into clusters, we apply a *clusterized spatial dynamic panel data* model and a multiple testing procedure, to investigate the best partition of clusters for the dataset. We show that, among the three candidate partitions considered in our analysis, the best partition is the one based on the asset's economic sector.

Keywords: Clustering · spatio-temporal models · financial returns

1 The Clusterized SDPD Model

Denoting with $\{\mathbf{y}_t, t = 1, 2, \ldots, T\}$ a spatio-temporal dataset – where the vector $\mathbf{y}_t = (y_{t1}, y_{t2}, \ldots, y_{tp})'$ is p-dimensional and includes the observations at time t from p different locations (or spatial units) – we start from the *heterogeneous SDPD* (Spatial Dynamic Panel Data) model, proposed in [1] and given by

$$\begin{aligned}\mathbf{y}_t = D(\boldsymbol{\lambda}_0)\mathbf{W}\mathbf{y}_t + D(\boldsymbol{\lambda}_1)\mathbf{y}_{t-1} + D(\boldsymbol{\lambda}_2)\mathbf{W}\mathbf{y}_{t-1} + D(\boldsymbol{\beta}_1)\mathbf{x}_t^{(1)} + \ldots \\ \ldots + D(\boldsymbol{\beta}_k)\mathbf{x}_t^{(k)} + \mathbf{c} + \boldsymbol{\varepsilon}_t,\end{aligned} \tag{1}$$

where $\mathbf{W}$ is the *spatial matrix*, a known weight matrix reflecting the *distances* between spatial units. The spatial matrix has a zero main diagonal and is usually based on physical distances, to deal with spatial correlation. One can also build matrix $\mathbf{W}$ using distances based on correlations between time series (or some other association measures), instead of physical distances. The i-th equation of this model, for $i = 1, \ldots, p$, is

$$y_{ti} = \lambda_{0i}\mathbf{w}_i'\mathbf{y}_t + \lambda_{1i}y_{t-1,i} + \lambda_{2i}\mathbf{w}_i'\mathbf{y}_{t-1} + \beta_{1i}x_{ti}^{(1)} + \ldots + \beta_{ki}x_{ti}^{(k)} + c_i + \varepsilon_{ti},$$

where the column vector $\mathbf{w}_i$ is the i-th row of $\mathbf{W}$. The parameters of the model are collected in the diagonal matrices $D(\boldsymbol{\lambda}_j)$ and $D(\boldsymbol{\beta}_l)$, with $j = 0, 1, 2$ and

M. Corazza et al. (Eds.): MAF 2024, *Mathematical and Statistical Methods for Actuarial Sciences and Finance*, pp. 160–166, 2024.
https://doi.org/10.1007/978-3-031-64273-9_27

$l = 1, \ldots, k$, where $\boldsymbol{\lambda}_j = (\lambda_{j1}, \ldots, \lambda_{jp})'$ and $\boldsymbol{\beta}_l = (\beta_{l1}, \ldots, \beta_{lp})'$. Note that each location has its own parameters, therefore the model is said *spatially heterogeneous*. A different case is when all locations have the same parameter values, then the Eq. (1) "collapses" into the classic *homogeneous SDPD* model of [4], with $\lambda_{ji} = \lambda_j$ and $\beta_{li} = \beta_l$ for all (i, j, l). Furthermore, the component $D(\boldsymbol{\beta}_l)\mathbf{x}_t^{(l)}$ in Eq. (1) may be included to account for the effects of some covariates on the time series data $\mathbf{y}_t$ (the p-dimensional vector $\mathbf{x}_t^{(l)}$ collects the data observed at time t on the p locations and for a given covariate l, with $l = 1, \ldots, k$). Finally, $\mathbf{c}$ contains the fixed effects while $\boldsymbol{\varepsilon}_t \sim i.i.d.$ with $E(\boldsymbol{\varepsilon}_t) = \mathbf{0}$ and $Var(\boldsymbol{\varepsilon}_t) = \boldsymbol{\Sigma}_\varepsilon$.

When the spatial units are grouped into clusters, a hybrid *SDPD* model can be considered, a cross between the homogeneous and heterogeneous cases. In such a case, the spatial units belonging to the same cluster are expected to be homogeneous. As a consequence, the associated spatial parameters, λ_{ji} and β_{li} for $i = 1, \ldots, p$, are set to be equal within clusters and to change for different clusters. Denote with S the number of clusters and let $\{G_s,\ s = 1, \ldots, S\}$ be a partition of $\{1, \ldots, p\}$, with p_s the number of units in the s-th cluster G_s, so $\sum_{s=1}^{S} p_s = p$. The *clusterized SDPD* model is

$$y_{ti} = \lambda_0^{(s)} \mathbf{w}_i' \mathbf{y}_t + \lambda_1^{(s)} y_{t-1,i} + \lambda_2^{(s)} \mathbf{w}_i' \mathbf{y}_{t-1} + \beta_1^{(s)} x_{ti}^{(1)} + \ldots + \beta_k^{(s)} x_{ti}^{(k)} + c_i + \varepsilon_{ti}, \quad \forall i \in G_s; s = 1, 2, \ldots, S; t = 1, 2, \ldots, T, \tag{2}$$

and it can be estimated by adapting the estimation procedure proposed in [1], as suggested in [2]. When $S = 1$, we obtain the *homogeneous SDPD* whereas, when $S = p$, we obtain the *heterogeneous SDPD*. However, to estimate the *clusterized SDPD* model consistently, it is necessary to know the clustering partition (the number of clusters and their composition), *i.e.* knowing $\{G_s,\ s = 1, \ldots, S\}$. To this aim, [2] have proposed a testing procedure which allows to verify if a candidate partition of clusters, assumed under the null, is suitable to estimate the *clusterized SDPD* model.

In this work, we move along two directions. First of all, we investigate the empirical performance of the testing procedure, proposed in [2], through a simulation study (Sect. 2). Second, we give an application of this testing procedure to analyse a real dataset of financial asset returns and to test the significance of a given partition of clusters for the assets (Sect. 3).

2 A Simulation Study to Investigate the Performance of the Testing Procedure for the Cluster Partition

The idea underlying the method proposed in [3] is based on a testing procedure that compares two setups: a specific *clusterized SDPD* model, assumed under H_0 and an *heterogeneous SDPD* model, assumed under H_1.

The testing procedure is based on the following statistics

$$\widehat{\delta}_{ji}^{(s)} = \widehat{\theta}_{ji} - \widehat{\theta}_j^{(s)} \qquad j = 1, \ldots, 3 + k; \forall i \in G_s; s = 1, \ldots, S; \tag{3}$$

Table 1. Performance indicators of the multiple test calculated over 500 Monte Carlo replications of a *clusterized SDPD* model with $S = 4$ clusters.

	Multiple test for $\boldsymbol{\lambda}_0$ $H_0 : \cap_{i,s} \left\{ \lambda_{0i} = \lambda_0^{(s)} \right\}$			Multiple test for $\boldsymbol{\lambda}_1$ $H_0 : \cap_{i,s} \left\{ \lambda_{1i} = \lambda_1^{(s)} \right\}$			Multiple test for $\boldsymbol{\lambda}_2$ $H_0 : \cap_{i,s} \left\{ \lambda_{2i} = \lambda_2^{(s)} \right\}$			Multiple test for $\boldsymbol{\beta}_1$ $H_0 : \cap_{i,s} \left\{ \beta_{1i} = \beta_1^{(s)} \right\}$		
	False Discovery rate											
$T =$	100	500	1000	100	500	1000	100	500	1000	100	500	1000
$p = 10$	0.024	0.022	0.028	0.044	0.077	0.063	0.038	0.032	0.043	0.020	0.018	0.016
50	0.052	0.018	0.036	0.098	0.056	0.064	0.092	0.026	0.056	0.032	0.046	0.026
100	0.062	0.036	0.064	0.202	0.068	0.054	0.142	0.068	0.076	0.104	0.062	0.044
	Average Power											
$T =$	100	500	1000	100	500	1000	100	500	1000	100	500	1000
$p = 10$	0.458	0.623	0.666	0.444	0.505	0.538	0.324	0.420	0.466	0.728	0.776	0.800
50	0.602	0.746	0.795	0.533	0.646	0.685	0.717	0.835	0.855	0.950	0.991	0.995
100	0.691	0.779	0.814	0.570	0.711	0.752	0.554	0.718	0.769	0.916	0.945	0.952

where $\widehat{\theta}_{ji}$ is the *heterogeneous estimator* of the j-th parameter in the vector $\boldsymbol{\theta}_i = (\lambda_{0i}, \lambda_{1i}, \lambda_{2i}, \beta_{1i}, \ldots, \beta_{ki})'$, while $\widehat{\theta}_j^{(s)}$ is the *clusterized estimator* of the corresponding parameter $\theta_j^{(s)}$, as

$$\widehat{\theta}_j^{(s)} = \frac{1}{p_s} \sum_{i \in G_s} \widehat{\theta}_{ji}, \qquad j = 1, \ldots, 3 + k, \tag{4}$$

that is the average of the heterogeneous estimators for the spatial units in the s-th cluster. Details about these estimators can be found in [3] and [2] and are omitted here to save space. For each $j = 1, \ldots, 3+k$, when the assumed clustering partition holds, the two estimators $\widehat{\theta}_{ji}$ and $\widehat{\theta}_j^{(s)}$ are expected to produce similar results and the statistics $\widehat{\delta}_{ji}^{(s)}$ are expected to be near to zero, for all (i, s). A multiple test (with Bonferroni correction) is then built to test the validity of the null hypothesis H_{0j} (seen as the intersection of p univariate hypotheses $H_{0ji}^{(s)}$)

$$H_{0j} : \cap_{i,s} H_{0ji}^{(s)} \qquad \text{with } H_{0ji}^{(s)} : \{\theta_{ji} = \theta_j^{(s)}\}, \qquad \forall i \in G_s; s = 1, \ldots S, \tag{5}$$

taking into account the familywise error rate (see [3] and [2] for more details).

Here we report some simulation results to confirm the validity of the testing procedure, considering different values of time-series dimension, $p = (10, 50, 100)$, and different sample sizes, $T = (100, 500, 1000)$. The parameter settings of the simulation study are fixed as in [2] (in particular, we considered 500 Monte Carlo replications, we set $S = 4$ clusters and the global size of the test was fixed to $\alpha = 0.10$). The results are shown in Table 1. Note that the four columns in the table refer to the four different spatial parameters of model (2): $\lambda_{0i}^{(s)}$, $\lambda_{1i}^{(s)}$, $\lambda_{2i}^{(s)}$, and $\beta_{1i}^{(s)}$, respectively.

The performance indicators reported in Table 1 are: a) *false discovery rate (FDR)*, which is the proportion of incorrect rejection of the null for at least one of the true univariate hypotheses $H_{0ij}^{(s)}$ (it represents the empirical size of the test, so it should be $FDR \leq \alpha$, at least asymptotically); b) *average power (AP)*, which is the average number of the proportion of correct rejection of the null in each iteration (the benchmark value is one, asymptotically). The values in the table show a good performance for the multiple test, since we have small values for *FDP* and increasing values for *AP*, when T increases. Regarding the *FDR*, it does not exceed $\alpha = 10\%$, as desired, at least for high values of T.

3 Application of the Method to Test the Cluster Partition of Financial Asset Returns

The multiple testing procedure described above can be used in many situations. For example, we might consider the spatial data observed in a given group of countries and we might want to see whether the *SDPD* model parameters are homogeneous within countries and heterogeneous across countries. In such a case, the test is used to verify if the "country partition" is relevant to explain the variability in model parameters. Of course, many other examples of cluster partition can be formulated, based on economic sectors, sex, education level or some other features of the spatial units.

In this work, we consider an application to financial asset returns. In particular, we downloaded from *Yahoo Finance* the stock quotes of 84 financial assets, coming from the G7 countries (Canada, France, Germany, Italy, Japan, United Kingdom and USA). The assets were suitable chosen to represent the main economic sectors of these countries. The names, the correspondent economic sectors and the nationalities of these assets are reported in Table 2. The financial time series range from 5th May 2021 to 29th December 2023, for a total of 666 daily observations. We transformed data into log-returns and, before applying the multiple test, we removed a temporal observation for all the assets if at least one of them contains a missing value for that date. In this way, the final spatio-temporal dataset has $T = 597$ observations.

As widely adopted in the spatial methodology, the spatial matrix $\mathbf{W} = \{w_{i,r}; i, r = 1, 2, \dots p\}$ has been derived by converting the correlation coefficients into distances, as follows:

$$w_{i,r} = \sqrt{2(1 - \rho_{ir})}, \qquad i, r = 1, 2, \dots, p,$$

where ρ_{ir} is the correlation coefficient between the i-th and r-th time series of log-returns. Finally, note that we do not have any exogenous regressors, here, so we consider model (2) without exogenous components. This means that we have to implement the multiple test only three times, once for each given vector of parameters ($\boldsymbol{\lambda}_0$, $\boldsymbol{\lambda}_1$ and $\boldsymbol{\lambda}_2$).

Table 2. Results of the multiple test, for three different cluster partitions assumed under H_0, for the financial dataset described in Sect. 3. Univariate rejections in the multiple test are in grey color, F (= FALSE) is for "rejected" and T (= TRUE) for "not rejected".

	Asset	"Sector" partition				"Country" partition				"Trivial" partition			
1	BCE Inc	T	T	T	Telecommunications	F	T	T	Canada	F	T	T	G7
2	Bank of Nova Scotia	T	T	T	Bank	T	T	T	Canada	T	T	T	G7
3	Cameco Corporation	T	T	T	Energy	T	T	T	Canada	T	T	T	G7
4	Ticray Brands	T	T	T	Health	T	T	T	Canada	T	T	T	G7
5	Royal Bank of Canada	T	T	T	Bank	T	T	T	Canada	F	T	T	**G7**
6	Intact Finantial Corp	T	T	T	Insurance	T	T	T	Canada	F	T	T	G7
7	Gildan	T	T	T	Clothing	T	T	T	Canada	T	T	T	G7
8	Suncor Energy Inc	T	T	T	Oil and Gas	T	T	T	Canada	T	T	T	G7
9	Tourmaline Oil Corp	T	T	T	Oil and Gas	T	T	T	Canada	T	T	T	G7
10	Telus Corporation	T	T	T	Telecommunications	T	T	T	Canada	T	T	T	G7
11	Maple Leaf Food	T	T	T	Food and Beverage	T	T	T	Canada	T	T	T	G7
12	Engie	T	T	T	Energy	T	T	T	France	T	**F**	T	G7
13	Renault	T	T	T	Motors	T	T	T	France	T	T	T	G7
14	Hermes	T	T	T	Clothing	T	T	T	France	T	T	F	G7
15	AXA	T	T	T	Insurance	T	T	T	France	T	T	T	G7
16	Danone	T	T	T	Food and Beverage	T	T	T	France	T	T	T	G7
17	Essilor Luxottica	T	T	T	Health	T	T	T	France	T	T	T	G7
18	Totalenergies	T	T	T	Oil and Gas	T	T	T	France	T	T	T	G7
19	Credit Agricole	T	T	T	Bank	T	T	T	France	T	T	T	G7
20	Orange	T	T	T	Telecommunications	F	T	T	France	F	T	T	G7
21	Societe Generale	T	T	T	Bank	T	T	T	France	T	T	T	G7
22	BNP Paribas	T	T	T	Bank	T	T	T	France	T	T	T	G7
23	Micheline	T	T	T	Tyres	T	T	T	France	T	T	T	G7
24	Hello Fresh	T	T	T	Food and Beverage	T	T	T	Germany	T	T	T	G7
25	BMW	T	T	T	Motors	T	T	T	Germany	T	T	T	G7
26	Volkswagen Group	T	T	T	Motors	T	T	T	Germany	T	T	T	G7
27	Deutsch Bank	T	T	T	Bank	T	T	T	Germany	T	T	T	G7
28	Allianz	T	T	T	Insurance	T	T	T	Germany	T	T	T	G7
29	Adidas	T	T	T	Clothing	T	T	T	Germany	T	T	T	G7
30	E.On	T	T	T	Energy	T	T	T	Germany	T	T	T	G7
31	Continental	T	T	T	Tyres	T	T	T	Germany	T	T	T	G7
32	De Linde Group	T	T	T	Oil and Gas	T	T	T	Germany	T	T	T	G7
33	Munich Re	T	T	T	Insurance	T	T	T	Germany	T	T	T	G7
34	RWE	T	T	T	Energy	T	T	T	Germany	T	T	T	G7
35	Fresenius	T	T	T	Health	T	T	T	Germany	T	T	T	G7
36	Deutsch Telekom	T	T	T	Telecommunications	T	T	T	Germany	F	T	T	G7
37	Poste Italiane	T	T	T	Insurance	T	T	T	Italy	T	T	T	G7
38	Campari	T	T	T	Food and Beverage	T	T	T	Italy	T	T	T	G7
39	Moncler	T	T	T	Clothing	T	T	T	Italy	T	T	T	G7
40	Amplifon	T	T	T	Health	T	T	T	Italy	T	T	F	G7
41	Fineco	T	T	T	Bank	T	T	T	Italy	T	T	T	G7
42	Ferrari	T	T	T	Motors	T	T	T	Italy	T	T	T	G7
43	Unipol	T	T	T	Insurance	T	T	T	Italy	T	T	T	G7
44	Enel	T	T	T	Energy	T	T	T	Italy	T	T	T	G7
45	Generali	T	T	T	Insurance	T	T	T	Italy	T	T	T	G7
46	Telecom	T	T	T	Telecommunications	T	T	T	Italy	T	F	T	G7
47	Eni	T	T	T	Oil and Gas	T	T	T	Italy	T	T	T	G7
48	Intesa San Paolo	T	T	T	Bank	T	T	T	Italy	T	T	T	G7
49	Nissan	T	T	F	Motors	T	T	F	Japan	T	T	F	G7
50	Toyota	T	T	F	Motors	T	T	F	Japan	T	T	F	G7
51	Toray Industrie	F	T	F	Clothing	T	T	F	Japan	F	T	F	G7
52	Japan Tobacco	T	T	F	Food and Beverage	T	T	F	Japan	F	T	T	G7
53	Astellas Pharma	T	T	F	Health	F	T	T	Japan	T	T	T	G7
54	Mizuho Finantial Group	F	T	F	Bank	T	T	F	Japan	F	T	F	G7
55	Sumitomo Mitsui Fin. Grp.	F	T	F	Bank	T	T	F	Japan	F	T	F	G7
56	Chubu Electric Power	T	T	T	Energy	T	T	T	Japan	F	T	T	G7
57	Osaka Gas	F	T	T	Oil and Gas	T	T	T	Japan	F	T	T	G7
58	Tokyo Gas	F	T	T	Oil and Gas	T	T	T	Japan	F	T	T	G7
59	KDDI Corp	T	T	T	Telecommunications	F	T	F	Japan	T	T	T	G7
60	Dai-Ichi Life	F	T	F	Insurance	T	T	F	Japan	F	T	F	G7
61	Burberry Group	T	T	T	Clothing	T	T	T	United Kingdom	T	T	T	G7
62	Smith and Nephew	T	T	T	Health	T	T	T	United Kingdom	T	T	T	G7
63	Rolls-Roice Holdings	T	T	T	Motors	T	T	T	United Kingdom	T	T	T	G7
64	Vodafone Group	T	T	T	Telecommunications	T	T	T	United Kingdom	T	T	T	G7
65	Lloyds Banking plc	T	T	T	Bank	T	T	T	United Kingdom	T	T	T	G7
66	Admiral Group	T	T	T	Insurance	T	T	T	United Kingdom	T	T	T	G7
67	Aviva	T	T	T	Insurance	T	T	T	United Kingdom	T	T	T	G7
68	HSBC Holdings	T	T	T	Bank	T	T	T	United Kingdom	T	T	T	G7
69	Petrofac	T	T	T	Oil and Gas	T	T	T	United Kingdom	T	T	T	G7
70	Royal Dutch Shell B	T	T	T	Oil and Gas	T	T	T	United Kingdom	T	T	T	G7
71	Astrazeneca	T	T	T	Health	T	T	T	United Kingdom	T	T	T	G7
72	Aston Martin Lagonda	T	T	T	Motors	T	T	T	United Kingdom	F	T	T	G7
73	Abbott	T	T	T	Health	T	T	T	USA	T	T	T	G7
74	Tesla	T	T	T	Motors	T	T	T	USA	T	T	T	G7
75	Ford	T	T	T	Motors	T	T	T	USA	T	T	T	G7
76	Exxon Mobil	T	T	T	Oil and Gas	T	T	T	USA	T	T	T	G7
77	Loews Corporation	T	T	T	Insurance	T	T	T	USA	T	T	T	G7
78	Maraton Oil	T	T	T	Oil and Gas	T	T	T	USA	T	T	T	G7
79	Bank of America	T	T	T	Bank	T	T	T	USA	T	T	T	G7
80	JPMorgan	T	T	T	Bank	T	T	T	USA	T	T	T	G7
81	ATandT	T	T	T	Telecommunications	T	T	T	USA	T	T	T	G7
82	Ralph Lauren Corp	T	T	T	Clothing	T	T	T	USA	T	T	T	G7
83	Nike	T	T	T	Clothing	T	T	T	USA	T	T	T	G7
84	PEPSI.Co	F	T	T	Food and Beverage	F	T	T	USA	F	T	T	G7
	Total number of rejections (=F)	**7**	**0**	**8**	**11**	**5**	**0**	**8**	**12**	**15**	**2**	**8**	**21**

In this application, the spatial units are the 84 assets. We considered three different partitions to group them into clusters. The first partition is based on the *economic sectors*: in this case there are 10 different clusters (Bank, Clothing, Energy, Food and Beverage, Health, Insurance, Motors, Oil and Gas, Telecommunications and Tyres). The second partition is based on the *nationality* of the asset, in which case there are 7 clusters (Canada, France, Germany, Italy, Japan, United Kingdom and USA). Finally, the last is the *trivial partition*, which considers a unique cluster of units (denoted by "G7"). We stress that, for each one of the three candidate partitions, we exactly know the composition of each cluster: the multiple test is therefore used to test if that given partition is suitable for the *clusterized SDPD* model to be used on the spatio-temporal dataset.

The results of the analysis are summarized in Table 2. The first column reports the name of the assets. The other three columns report the results of the multiple test for the three (candidate) clustering partitions ("sector", "country" and "trivial", respectively). For each one, we report the decisions of the test for the null hypothesis $H_{0ji}^{(s)}$, which is denoted as F = FALSE (for "rejected") or T = TRUE (for "not rejected"). So, for a given partition, a sequence (T, T, F) in the i-th row of the table means that the null hypothesis has been rejected for the parameter λ_{2i} and not rejected for the parameters λ_{0i} and λ_{1i}. More clearly, this means that asset i-th does not appear to belong to the assigned cluster, at least for the spatial parameter λ_{2i}. To accept the global null hypothesis (= the given cluster partition is correct), we should observe (T,T,T) in any row i of the table (although, normally, a low number of rejections can be tolerated). Now, from the last row of Table 2, we can note that there are 21 rejections for the "trivial" partition, which implies that we must reject the global null in this case and conclude that the *SDPD model* is not homogeneous, that is $S \neq 1$. But there are also 11 rejections for the "sector" partition and 12 rejections for the "country" one, so H_0 should be rejected again. However, most of the rejections occur for the Japanese asset returns, which may be seen as *outliers* with respect to all the cluster partitions. If we ignore the Japanese asset returns, we can conclude that the "sector" partition is the best candidate for the *clusterized SDPD* model since it gives less "rejected" cases (i.e. at least one FALSE in the sequence) among the last three columns of Table 2. In fact, if we remove the Japanese assets, we have 1 rejection (the PEPSI.Co asset in the second column) against 3 for the "country" (the BCE Inc., Orange and PEPSI.Co in the third column) and 11 for the trivial partition respectively.

References

1. Dou, B., Parrella, M.L., Yao, Q.: Generalized Yule-Walker estimation for spatio-temporal models with unknown diagonal coefficients. J. Econom. **194**, 369–382 (2016)
2. Feo, G., Giordano, F., Milito, S., Niglio, M., Parrella, M.L.: Testing clusters of locations in spatial dynamic panel data models. In: Book of Abstracts and Short Papers 14th Scientific Meeting of the Classification and Data Analysis Group, pp. 461–464 (2023)

3. Giordano, F., Niglio, M., Parrella, M.L.: Testing spatial dynamic panel data models with heterogeneous spatial and regression coefficients. J. Time Ser. Anal. (2024) https://doi.org/10.1111/jtsa.12738
4. Lee, L.F., Yu, J.: Estimation of spatial autoregressive panel data models with fixed effects. J. Econom. **154**, 165–185 (2010)

Assessing the Impact of Climate and Environmental News on Financial Markets

Gianna Figà-Talamanca[1(✉)], Andrea Fronzetti Colladon[2], Barbara Guardabascio[1], Marco Patacca[1], and Ludovica Segneri[2]

[1] Department of Economics, University of Perugia, Perugia, Italy
gianna.figatalamanca@unipg.it
[2] Department of Engineering, University of Perugia, Perugia, Italy

Abstract. In this paper, we investigate the impact of climate and environmental attention on the dynamics of returns of two US stock market indexes: the broader S&P 500 and the theme-related S&P 500 Energy.

Keywords: Sentiment analysis · GARCH models · Volatility · Risk measures · Semantic Brand Score

1 Climate and Environmental News Semantic Importance

Climate attention, shaped by a combination of scientific findings, social awareness, and media portrayal, has begun to exert a profound influence on financial markets, mainly in energy and commodity sectors [1]. This change in attention is not merely academic; it is transforming the fundamental dynamics of how these markets operate, are regulated, and are perceived by stakeholders.

The purpose of this paper is to measure the impact of the environmental and climate discourse, extracted from articles appearing in the *New York Times, Los Angeles Times, USA Today, Financial Times, Wall Street Journal*, from January 2014 to November 2022, on the dynamics of returns of two US stock market indexes: the S&P 500 and the S&P 500 Energy. While the significance of sentiment within the stock market has been extensively explored in existing literature, see [2,3,5] and others, our focus lies in delving deeper into the semantic importance of environmental and climatic themes. As a first step, we select the main themes related to climate change and environmental topics and build and list of keywords from glossaries and previous research. Machine learning techniques (Tf-Idf) and word embedding [6] are then applied to the news text corpus to extract other possible words of interest (such as synonyms, hypernyms, hyponyms, and other related terms) that are finally grouped into clusters. A network is built between all the words of the pre-processed text, with a co-occurrence threshold of five words[1]. The selected clusters of keywords are listed in Table 1:

[1] Common text pre-processing routines, such as tokenization, removal of stop-words, and removal of word affixes, known as stemming, are applied to the news corpus before computing the score.

M. Corazza et al. (Eds.): MAF 2024, *Mathematical and Statistical Methods for Actuarial Sciences and Finance*, pp. 167–171, 2024.
https://doi.org/10.1007/978-3-031-64273-9_28

Table 1. Selected keywords for environmental and climate-related words

Cluster	Description
Renewable Energy	Energy sources naturally replenished such as solar wind and hydropower alternatives to fossil fuels
Climate Action	Efforts to reduce greenhouse gas emissions and mitigate the effects of climate change
Climate Human Effects	Effects of climate change caused directly by human actions
Direct Climate Change Effects	General effects of climate change
Waste Negative	Words related to negative waste management
Waste Positive	Words relate to positive waste management
Gas Emission	Gases released in the atmosphere and contributing to global warming
Energy Efficient	Words related to energy efficiency
Green Mobility	Refers to sustainable transportation
Net Zero Emissions	Balance between the amount of greenhouse gases emitted and the amount removed from the atmosphere
Environment&Education	Education topic declined to the concept of environment
Sustainability General	Generic words related to sustainability
Climate Resilience	Generic words related to the concept of climate resilience

The importance score for the specific cluster i is assigned by the Semantic Brand Score (SBS) [4], a measure of semantic importance that combines text mining and network analysis methods. The SBS indicator is based on three dimensions:

- the prevalence (PR), which measures how frequently a cluster of words is used in the text;
- the diversity (DI), which evaluates how heterogeneous and unique the associations to a cluster of words are;
- the connectivity (CO), which highlights how much the concept represented by the cluster of words can bridge connections among other concepts in the text.

The three dimensions are standardized and summed up to finally compute the semantic importance of each cluster under a weekly time frame.

$$SBS(i) = \frac{PR(i) - \overline{PR}}{std(PR)} + \frac{DI(i) - \overline{DI}}{std(DI)} + \frac{CO(i) - \overline{CO}}{std(CO)}, \text{ for } i = 1, 2, \ldots, n, \quad (1)$$

where $\overline{PR}, \overline{DI}, \overline{CO}$ and $std(PR), std(DI), std(CO)$ are, respectively, the average value and standard deviation of each dimension over time. In Fig. 1 we plot the semantic relevance of two of the keyword clusters, namely *Climate human effects*, and *Direct climate change effects*. In this example, all articles appearing in the *Wall Street Journal* and *Financial Times* (over 1500 articles per week on average) are taken into account. From the picture, we can see the value and variability of prevalence, diversity, and connectivity over time, and we are able to catch how the SBS indicator is composed.

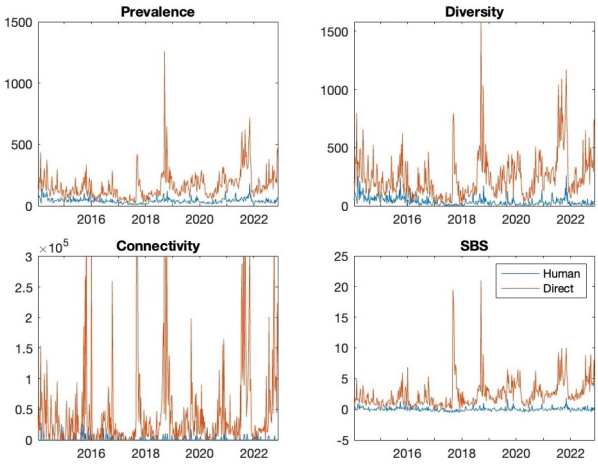


Fig. 1. Semantic relevance of *Climate human effects* and *Direct climate change effects* keywords (Financial news).

2 Model Specification

As a first experiment to verify which theme affects financial markets the most, we fit an ARMAX model for the dynamics of logarithmic returns (Eq. 2) where the error term follows an EGARCHX process (Eq. 3) on two market indexes. Our goal is to check whether the return dynamics is better described when the semantic relevance of a cluster is taken into account as an added exogenous factor.

$$R_t = a + cX_t^r + \epsilon_t,\ t = 1, 2, \ldots, T; \tag{2}$$

where $R_t = \ln \frac{P_t}{P_{t-1}}$ is the logarithmic return, $\epsilon = \{\epsilon_t, t \geq 0\} = \sqrt{h_t}\eta_t$ is the error process with $\eta = \{\eta_t, t \geq 0\}$ Gaussian white noise and h_t is the so-called conditional variance:

$$\log h_t = \alpha_0 + \alpha_1 \eta_{t-1} + \beta_1 \log h_{t-1} + \lambda \left(|\eta_{t-1}| - \mathbb{E}\left[|\eta_{t-1}| \right] \right) + \gamma X_t^h, \tag{3}$$

where $a, b, \alpha_0, \alpha_1, \beta_1, \lambda, \gamma$ are constant parameters.

The explanatory variables $X^r = \{X_t^r, t \geq 0\}$ and $X^h = \{X_t^h, t \geq 0\}$ are represented by the semantic importance score in the mean equation and their first differences in the conditional variance specification. The estimation exercise is finally repeated, considering the mean, rather than the individual, semantic score across the selected clusters.

3 Empirical Results

In this numerical experiment, our objective is to evaluate the impact of weekly discourse on the environment and climate extracted from major US newspapers[2] on both the S&P 500 Index and the S&P 500 Energy Index. The latter includes

Table 2. S&P 500 and S&P 500 Energy parameter estimates

Word	Par	S&P 500		S&P Energy	
		General	Financial	General	Financial
Climate Action	c	0.013**	0.004*	−0.009	0.025
	γ	0.688	1.056	−2.815**	0.950
Climate Human Effects	c	0.003***	−0.001	0.005*	0.004***
	γ	−0.641***	0.118	−0.386*	−0.160
Climate Resilience	c	0.024**	0.005	−0.006	0.036
	γ	−11.414***	−1.697	−6.329	2.996
Direct Climate Effects	c	0.000	0.000	0.000**	0.001***
	γ	−0.006	0.015	−0.013	−0.035
Environment & Education	c	−0.006**	0.008	−0.036	0.022
	γ	−5.174*	0.441	−4.564	−0.738
Energy Efficient	c	−0.005***	0.004	−0.007	0.028
	γ	−2.963	−1.413	−1.350	0.783
Gas Emission	c	0.001	0.001	0.002**	0.003***
	γ	−0.067	0.009	−0.061	−0.001
Green Mobility	c	−0.001	−0.001	−0.004	0.005**
	γ	−0.416**	−0.001	−0.041	−0.049
Net Zero Emissions	c	0.028**	0.005**	0.029	0.033**
	γ	−7.155***	0.111	−4.929*	1.250
Renewable Energy	c	0.000	0.000	0.005	0.005
	γ	−0.188	0.448**	−0.312	0.116
Sustainability General	c	0.006	0.001	0.001	0.011***
	γ	−1.535	0.352	−0.013	−0.145
Waste Negative	c	0.001***	−0.002	0.000	0.001
	γ	−0.400**	0.421***	−0.177	0.311**
Waste Positive	c	0.011*	0.010**	−0.009	0.017*
	γ	−0.106	0.047	0.127	0.408
Mean	c	0.008***	0.003	−0.012	0.036***
	γ	−4.621**	4.508**	−2.338	1.792

Significance codes: ***p-value $<$ 0.001; **0.001 $\leq$ p-value $<$ 0.01; *0.01 $\leq$ p-value $<$ 0.05

[2] New York Times, Los Angeles Times, USA Today, Financial Times, Wall Street Journal.

all companies in the S&P 500 that fall under the energy sector according to the Global Industry Classification Standard (GICS): 30 companies whose combined float-adjusted market capitalization represents more than 5% of the total index S&P 500 and letting, the energy sector holds the eighth position in terms of weightage between the 11 sectors within the S&P 500 index.

The preliminary analysis results are detailed in Table 2. The findings indicate that the S&P 500 Index is primarily influenced by news reported in generalist media, specifically by clusters related to climate action, climate's human effects, environment & education, energy efficiency, net zero emissions and negative waste. In contrast, financial news[3] has a lower impact on the market index. When focusing on the S&P 500 Energy Index, which includes only equities in the Energy Sector, we obtain the opposite results. Here, financial news significantly affects the energy index, particularly clusters related to climate human and direct climate change effects, greenhouse gases, green mobility, net-zero emissions, and sustainability. The estimated parameters, when considering the average score across clusters, confirm that generalist news mainly influences the global market index, while financial news significantly affects the energy sector index.

References

1. Boykoff, M.T.: Who Speaks for the Climate?: Making Sense of Media Reporting on Climate Change. Cambridge University Press, Cambridge (2011)
2. Coqueret, G.: Stock-specific sentiment and return predictability. Quantit. Financ. **20**(9) (2020). https://doi.org/10.1080/14697688.2020.1736314
3. Figà-Talamanca, G., Patacca, M.: An explorative analysis of sentiment impact on S&P 500 components returns, volatility and downside risk. Ann. Oper. Res. 1–23 (2022). https://doi.org/10.1007/s10479-022-05129-w
4. Fronzetti Colladon, A.: The semantic brand score. J. Bus. Res. **88**, 150–160 (2018). https://doi.org/10.1016/j.jbusres.2018.03.026
5. Gu, C., Kurov, A.: Informational role of social media: evidence from Twitter sentiment. J. Bank. Financ. **121**, 105969 (2020). https://doi.org/10.1016/j.jbankfin.2020.105969
6. Pennington, J., Socher, R., Manning, C. D.: GloVe: global vectors for word representation. In: Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP), pp. 1532–1543 (2014)

[3] Articles appeared in the New York Times, Los Angeles Times, and USA Today are referred to as general news whereas those in the Financial Times and Wall Street Journal as financial news.

The Sparsity-Constrained Graphical Lasso

Alessandro Fulci(✉), Sandra Paterlini, and Emanuele Taufer

Department of Economics and Management, University of Trento, Via Inama 5, 38123 Trento, Italy
alessandro.fulci@unitn.it

Abstract. This paper introduces the Sparsity-constrained Graphical Lasso (SCGlasso) for the precision matrix, $\mathbf{\Theta}$, in a multivariate Gaussian framework. The estimator is designed to produce a shrunk estimate of $\mathbf{\Theta}$, while simultaneously imposing a certain degree of sparsity, which is crucial for reconstructing the conditional dependence graph and the partial correlation graph. The proposed method employs an ℓ_1-norm (Glasso) regularization to achieve shrinkage and imposes an ℓ_0-pseudo-norm constraint to ensure sparsity. The proposed approach performs well compared to Glasso on simulated data, also in contexts where the number of variables p exceeds the number of observations n.

Keywords: Gaussian graphical models · Glasso · ℓ_0-constraint · ℓ_1-penalty

1 Introduction

Let $\mathbf{X} \in \mathbb{R}^{n \times p}$ be a matrix of n realizations drawn from p normally distributed random variables with a covariance matrix $\mathbf{\Sigma}$ and a precision matrix $\mathbf{\Theta} = \mathbf{\Sigma}^{-1}$. The objective of this paper is to consider the problem of estimating a simultaneously sparse and shrunk version of $\mathbf{\Theta}$ in a multivariate Gaussian framework by introducing an estimation technique based on the ℓ_1-norm penalty and a constraint on the ℓ_0-pseudo-norm.

Graphical models provide a powerful tool for representing the conditional dependence structure among a set of random variables. A graphical model consists of two main components: a graph $\mathcal{G}(\mathbf{N}, \mathbf{E})$ and a joint distribution $\mathbf{f}$. The vertices (or nodes) of the graph $\mathbf{N} = \{1, ..., p\}$ represent a finite set of random variables, while the set of edges $\mathbf{E}$ indicate the conditional dependencies between pairs of variables. Mathematically, a pair of two edges $(i, j) \in \mathbf{E}$, with $i, j = 1, ..., p$ and $i \neq j$ if and only if $X_i \not\perp\!\!\!\perp X_j | X_c$ where $C = \{k \in \mathbf{N} | k \neq i, j\}$ and $\perp\!\!\!\perp$ indicates stochastic independence between two random variables.

In this context, we focus on undirected Gaussian graphical models, where the joint distribution of random variables follows a multivariate normal distribution. In this particular case, $\mathbf{\Theta}$ allows us to retrieve the dependence structure among each component in the multivariate distribution (see [7,12]), since, given $\theta_{i,j}$, that is the (i, j)-th element of $\mathbf{\Theta}$, $\theta_{i,j} = 0 \iff X_i \perp\!\!\!\perp X_j | X_c$ and, consequently, the edge $(i, j) \in \mathbf{E} \iff \theta_{i,j} \neq 0$.

M. Corazza et al. (Eds.): MAF 2024, *Mathematical and Statistical Methods for Actuarial Sciences and Finance*, pp. 172–178, 2024.
https://doi.org/10.1007/978-3-031-64273-9_29

The purpose of this article is to introduce and analyze through simulations a new possible estimator for the precision matrix $\boldsymbol{\Theta}$. This method is based on both an ℓ_1-norm (Glasso) regularization to produce a shrunk estimate of $\boldsymbol{\Theta}$ and a constraint on the ℓ_0-pseudo-norm to impose sparsity. Nowadays, methods penalizing the norm of $\boldsymbol{\Theta}$ are generally used to estimate sparse and more intuitive models.

Nevertheless, limited attention has been devoted to the implementation of an *a-priori* constraint on the ℓ_0-pesudo-norm of $\hat{\boldsymbol{\Theta}}$ that allows to impose a certain degree of sparsity. Indeed, the importance of sparse modeling techniques (like the ℓ_0-pseudo-norm constraint) is likely crucial for performing inference, especially in those contexts where the number of variables p could vastly exceed the number of observations n, as put forward by [4] for linear regression problems.

The contribution provided by this paper is twofold. Firstly, it introduces a new estimator of $\boldsymbol{\Theta}$ by integrating the ℓ_0-pseudo-norm constraint in the ℓ_1-regularized maximum likelihood estimator. Secondly, we report the results of an extensive comparison with reference to the ℓ_0-unconstrained estimator, obtained through simulations on well-known network structures, such as scale-free and cluster networks. The empirical evidence suggests that, while both Glasso and SCGlasso offer good performances, the SCGlasso outperforms Glasso in the two considered network structures with reference to all sample sizes.

The paper is structured as follows: Sect. 2 defines the estimation algorithm underlying the SCGlasso. Section 3 reports the simulation settings and discusses the results obtained. Section 4 concludes.

2 Sparisity-Constrained Glasso - SCGlasso

We propose a generalization of the estimation problem behind the Glasso that integrates a constraint on the ℓ_0-pseudo-norm. It can be noted that the sparse and shrunk estimation of the matrix $\boldsymbol{\Theta}$ follows a procedure that actually resembles an expanded version of the typical Glasso algorithm (see [1,10]).

With reference to the ℓ_1-norm penalization, the Sparsity-constrained Glasso estimator is obtained by solving the following optimization problem:

$$\hat{\boldsymbol{\Theta}}_{\mathbf{SC}} = \underset{\substack{\boldsymbol{\Theta} \succ 0 \\ \frac{1}{2}\rho_0(\boldsymbol{\Theta}) \leq k}}{\operatorname{argmax}} \; \{ \log(\det(\boldsymbol{\Theta})) - \operatorname{trace}(\mathbf{S}\boldsymbol{\Theta}) - \lambda \rho_1(\boldsymbol{\Theta}) \} \tag{1}$$

where:

- $\rho_0(\boldsymbol{\Theta}) = \sum_{i \neq j}^{p} \mathbb{I}[\theta_{i,j} \neq 0]$ represents the ℓ_0-based quantity, that is a $0-1$ indicator function. Note that, by construction, $\rho_0(\boldsymbol{\Theta}) = 2|E|$, with $|E|$ being equal to the number of edges in the graph uniquely defined by $\boldsymbol{\Theta}$.
- $\rho_1(\boldsymbol{\Theta}) = \sum_{i \neq j}^{p} |\theta_{i,j}|$ is the ℓ_1-norm of the off-diagonal entries of $\boldsymbol{\Theta}$.
- $\mathbf{S}$ is the sample correlation matrix of $\mathbf{X}$.
- λ is the penalty parameter and it used to control the strength of the ℓ_1-penalization.

To solve the problem in (1), it is possible to divide it into two sequential parts, avoiding the non-convexity implicit in the ℓ_0-constraint. Firstly, given the penalty parameter λ, we solve the ℓ_0-unconstrained problem by exploiting the ordinary Glasso algorithm, to infer the structure of $\mathbf{\Theta}$. Exploiting the rules for convex optimization from [5], the sub-gradient equation is:

$$\mathbf{\Theta}^{-1} - \mathbf{S} - \lambda \mathbf{\Psi} = \mathbf{0} \tag{2}$$

With the symmetric matrix $\mathbf{\Psi}$ having diagonal elements zero and $\psi_{i,j} = \text{sign}(\theta_{i,j})$ if $\theta_{i,j} \neq 0$, else $\psi_{i,j} \in [-1, 1]$ if $\theta_{i,j} = 0$. Now it is possible to use the block-wise coordinate descend algorithm proposed by [10]. To do this we consider partitioning all the matrices into one column versus the rest.

$$\mathbf{\Theta} = \begin{bmatrix} \mathbf{\Theta}_{11} & \theta_{12} \\ \theta_{12}^T & \theta_{22} \end{bmatrix}, \quad \mathbf{S} = \begin{bmatrix} \mathbf{S}_{11} & \mathbf{s}_{12} \\ \mathbf{s}_{12}^T & s_{22} \end{bmatrix}, \quad \text{etc.} \tag{3}$$

Denoting by $\mathbf{W}$ the current working version of $\mathbf{\Theta}^{-1}$, with partitions as in (3), the block coordinate descend cycles through each column/row of $\mathbf{W}$, leading to $\mathbf{W}_{11}\beta - \mathbf{s}_{12} - \lambda\psi_{12} = \mathbf{0}$, with $\beta = -\theta_{12}/\theta_{22}$. It can be shown that the system in (2) is equivalent to a modified version of the estimating equations for a Lasso regression (see [15]). This allows us to solve each block-wise step using a modified algorithm for the Lasso, treating each variable as the dependent variable and the other $p - 1$ as the explanatory variables. This approach is similar to what has been previously developed by [14].

From here, it is possible to introduce the second part of the algorithm, by noting that it is always possible to modify the Glasso to have edge-specific penalty parameters $\lambda_{i,j}$. Specifically, if $\lambda_{i,j} = \infty \Rightarrow \hat{\theta}_{ij} = 0$. Consequently, it is possible to: *1)* preliminary use the Glasso to obtain a temporary estimate of $\mathbf{\Theta}$ the does not solve problem (1), *2)* Adopt a greedy rounding procedure. Specifically, apply again the Glasso algorithm, while simultaneously setting to 0 all the edges $\hat{\theta}_{i,j}$ that belong to the set of $[n(n-1)]/2-k$ minimum elements (in absolute value) in the preliminary estimated matrix $\hat{\mathbf{\Theta}}$, to obtain the sparse and shrunk precision matrix $\hat{\mathbf{\Theta}}_{\mathbf{SC}}$. This new algorithm is the Sparsity-constrained Glasso (SCGlasso). In SCGlasso, we are exploiting an approach for variables selection comparable to the Sure Independence Screening, SIS, for linear models [2,8,9].

3 Simulations

We investigate the performance of SCGlasso, compared to the well-established Glasso [10], considering two underlying network structures (scale-free and cluster) with $p = 30$ and $n = (20, 50, 100, 300, 600)$. We rely on the algorithms proposed by [17] and by [16] to generate the adjacency matrix, using the R package *huge* to produce the sparse precision matrices of all the network structures (parameters: $v = 0.3$ and $u = 0.1$). The optimal values of k and λ are selected using Bayesian Information Criterion (BIC) for graphical models (see [11] and [16]). In our application, BIC is minimized trough Brent algorithm (see

[6]), which is derivative-free and performs a combination of golden section search and successive parabolic interpolation.

Two performance measures, the F_1-score and the norm error are used for the comparison of the methods (e.g., [3] and [13]). The F_1-score provides an evaluation of the model selection performance, while the norm error gauges the numerical accuracy of an estimated precision matrix. The norm error is defined as $\eta = \|\hat{\mathbf{\Theta}} - \mathbf{\Theta}\|_2$.

3.1 Simulation Results

We report the comparison of the average performance of the two estimators on the two network configurations, given the optimal parameters' values selected as discussed in the previous section. Results are presented in Table 1. By looking at the model selection performance of the estimators, we observe that:

- **Glasso** (columns 1 and 5): it shows an initial increase in F_1 score followed by a decrease when sample size increases. We notice that this situation is strictly related to the sub-optimality of the ℓ_1-regularization, since when n increases, the optimal λ tends to zero, increasing both the true positive and the false positive rates.
- **SCGlasso** (columns 3 and 7): the model selection performance of this estimator always increases when sample size grows. In fact, by applying the ℓ_0-pseudo-norm constraint on the individual elements of $\hat{\mathbf{\Theta}}$ we are likely allowing the increase in true positive rates and simultaneously constraining the false positive rates.

From the point of view of numerical accuracy, our simulation shows that:

- **Glasso** (columns 2 and 6): unlike its behavior with reference to the F_1 score, η keeps shrinking as sample size increases.
- **SCGlasso** (columns 4 and 8): also from the numerical accuracy point of view, we do not reject the decreasing trend of η along an increase in sample size. The sparsity-constrained estimator also shows a higher accuracy compared to ordinary Glasso for both structures and all sample sizes.

Table 1. Mean values of F_1 scores and norm errors, with standard deviations in brackets, on $B = 100$ Monte Carlo runs for $p = 30$.

Sample size	Scale-Free				Cluster			
	Glasso		SCGlasso		Glasso		SCGlasso	
	F_1	η	F_1	η	F_1	η	F_1	η
$n = 20$	0.113	4.912	**0.217**	**4.151**	0.069	6.947	**0.205**	**5.552**
	(0.009)	(0.013)	(0.008)	(0.033)	(0.007)	(0.057)	(0.006)	(0.049)
$n = 50$	0.395	4.463	**0.463**	**3.421**	0.233	6.462	**0.468**	**4.837**
	(0.017)	(0.023)	(0.008)	(0.043)	(0.015)	(0.066)	(0.007)	(0.044)
$n = 100$	0.614	3.580	**0.667**	**2.284**	0.580	5.309	**0.678**	**3.686**
	(0.008)	(0.034)	(0.008)	(0.031)	(0.008)	(0.033)	(0.006)	(0.062)
$n = 300$	0.584	2.522	**0.898**	**0.958**	0.645	3.570	**0.941**	**1.652**
	(0.006)	(0.022)	(0.006)	(0.019)	(0.004)	(0.039)	(0.003)	(0.022)
$n = 600$	0.526	1.916	**0.937**	**0.691**	0.616	2.630	**0.974**	**1.239**
	(0.005)	(0.019)	(0.004)	(0.013)	(0.003)	(0.027)	(0.003)	(0.009)

4 Conclusion

In this article, we propose SCGlasso, a new method to compute a sparse and shrunk estimate of the precision matrix $\boldsymbol{\Theta}$ in multivariate Gaussian settings. Positive definiteness of the estimated matrix is guaranteed by re-iterating the node-specific version of Glasso. Moreover, SCGlasso algorithm has the crucial advantage of avoiding the non-convexity of the set defined by ℓ_0-based constraint which is essentially formed by the union over all $\binom{\binom{p}{2}}{k}$ possible subsets of k edges.

Results obtained through simulations do not reject the hypothesis of pure ℓ_1-regularization to be sub-optimal with reference to the ℓ_0-constraint, both from the model selection and estimation performance points of view.

Nevertheless, limitations might arise for SCGlasso from two aspects. Firstly, the time complexity of the calibration process is at least of order $O(p^2)$ since BIC should be minimized over λ for each k, with $k \in \mathcal{C} = \{1, 2, ..., \frac{p(p-1)}{2}\}$. For instance, in the case of cluster data structure with p values of 5, 10, and 15, and $n = 50$, computational times (seconds) on a desktop computer with an Intel i7 6-cores processor and $16GB$ of RAM, are the following: 1.97, 17.92, 100.2. In practice, it is possible to assume a certain degree of sparsity in the true graph and iterate BIC minimization only in a subset of $\mathcal{C}$. In our simulation, BIC has been minimized for each $k \in \mathcal{C}^* = \{1, 2, ..., [\frac{p(p-1)}{4}]\}$, (where $[\frac{p(p-1)}{4}]$ denotes the integer part of $\frac{p(p-1)}{4}$) hence assuming, at maximum, the presence of 50% of all possible edges. Secondly, when sample size is small, convergence of SCGlasso might require a computationally relevant number of iterations to reach the optimal λ. To avoid this problem, the parameter space for the implementation of Brent algorithm should be carefully set, and the minimum level of λ should be increased as $n \to p$.

As a possible application on a real-world dataset, we study a numerical example based on [18]. The dataset is made of $p = 39$ expression levels of isoprenoids genes from $n = 118$ samples from the plant Arabidopsis Thaliana. We observe that SCGlasso provides a sparser configuration than Glasso, allowing a better interpretability of the relevant relationships among the 118 genes. Results are available upon request.

Finally, for further research, it is possible to consider that, in general, the Glasso guarantees the support set of $\hat{\boldsymbol{\Theta}}$ to coincide with the support set of $\boldsymbol{\Theta}$, only for a vast number of observations, that is $n = \Omega(d^2 \log(p))$. To mitigate problems connected to the possible lack of oracle property in small samples, it might be convenient to exploit an adaptive version of the Glasso, based on the Adaptive Lasso introduced by [19] which is an oracle procedure for linear models. Adopting this adaptive version of the Glasso in the sparsity-constrained estimator might be appropriate for estimating the true support $\mathcal{A}$ of $\boldsymbol{\Theta}$, nevertheless its tuning procedure might be computationally expensive, since it would imply to minimize BIC over a 3-dimensional parameter space.

References

1. Banerjee, O., Ghaoui, L.E., d'Aspremont, A.: Model selection through sparse maximum likelihood estimation for multivariate Gaussian or binary data. J. Mach. Learn. Res. **9**, 485–516 (2008)
2. Barut, E., Fan, J., Verhassel, A.: Conditional sure independence screening. J. Am. Stat. Assoc. **111**(515), 1266–1277 (2016)
3. Bernardini, D., Paterlini, S., Taufer, E.: New estimation approaches for graphical models with elastic net penalty. Econom. Stat. (2022)
4. Bertsimas, D., Copenhaver, M.S., Mazumder, R.: The trimmed lasso: sparsity and robustness. arXiv preprint arXiv:1708.04527 (2017)
5. Boyd, S., Vandenberghe, L.: Convex Optimization. Cambridge University Press, Cambridge (2004)
6. Brent, R.: Algorithms for Minimization without Derivatives. Prentice-Hall, Englewood Cliffs (1973)
7. Bühlmann, P., Van-de-Geer, S.: Statistics for High Dimensional Data: Methods, Theory and Applications. Springer, Cham (2011). https://doi.org/10.1007/978-3-642-20192-9
8. 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)
9. Fan, J., Song, R.: Sure independence screening in generalized linear models with NP-dimensionality. Ann. Stat. **38**(6), 3567–3604 (2010)
10. Friedman, J., Hastie, T., Tibshirani, R.: Sparse inverse covariance estimation with the graphical lasso. Biostatistics **9**(3), 432–441 (2008)
11. Foygel, R., Drton, M.: Extended Bayesian information criteria for Gaussian graphical models. In: Advances in Neural Information Processing Systems, vol. 23 (2010)
12. Hastie, T., Tibshirani, R., Friedman, J.: The Elements of Statistical Learning: Data Mining, Inference and Prediction. Springer, Cham (2009). https://doi.org/10.1007/978-0-387-84858-7
13. Hastie, T., Tibshirani, R., Wainwright, M.: Statistical Learning with Sparsity: The Lasso and Generalizations. CRC Press, Boca Raton (2016)

14. Meinshausen, N., Bühlmann, P.: High-dimensional graphs and variable selection with the Lasso. Ann. Stat. **34**(3), 1436–1462 (2006)
15. Tibshirani, R.: Regression shrinkage and selection via the lasso. J. R. Stat. Soc. Ser. B Stat. Methodol. **58**(1), 267–288 (1996)
16. Torri, G., Giacometti, R., Paterlini, S.: Robust and sparse banking network estimation. Eur. J. Oper. Res. **270**(1), 51–65 (2018)
17. Watts, D.J., Strogatz, S.H.: Collective dynamics of 'small-world' networks. Nature **393**, 440–442 (1998)
18. Wille, A., et al.: Sparse graphical Gaussian modeling of the isoprenoid gene network in Arabidopsis thaliana. Genome Biol. **5**(11), 1–13 (2004)
19. Zou, H.: The adaptive lasso and its oracle properties. J. Am. Stat. Assoc. **101**(476), 1418–1429 (2006)

Cliometrics and Actuarial Science: New Avenues for Enriching Prospective Mortality Table Construction Models

Kué Gilles Gaba[1(✉)], Stéphane Loisel[2], and Antoine Parent[3]

[1] Laboratoire SAF, Institut de Science Financière et d'Assurances, Université Lyon 1, Lyon, France
kue.gilles.gaba@actuariatech.com
[2] Laboratoire LIRSA, Conservatoire National des Arts et Métiers, Paris, France
stephane.loisel@lecnam.net
[3] Laboratoire d'Économie Dionysien, Université Paris 8, Paris, France
antoine.parent02@univ-paris8.fr

Abstract. In a cliometric context of demographic and epidemiological transitions, we have improved the accuracy of actuarial age-specific mortality projections by develop-ing the multifactor composite PCR-optimal/PLS model. The results show that the proposed model performs well compared to conventional approaches, although variations by age bracket suggest opportunities for improvement.

Keywords: Prospective mortality table · Actuarial mortality forecasting · Multifactorial composite PCR-optimal/PLS model · Cliometrics

1 Introduction

The article has a twofold aim: to fulfil the cliometric objective of measuring History and to develop a prospective modelling of mortality tables. Indeed, the evolution of mortality rates is the counterpart of the historical experiences (epidemics, medical progress, improvement of living conditions, wars, revolutions, etc.) of humanity in different regional realities. The development of a metric of historical dynamics is naturally based on the calculation of life expectancy. The improvement of statistical methods for modelling the deformation of mortality rates over time meets the needs of both actuaries and historians. Floud and Harris (1997) and Fogel (1986) already used anthropometric proxies as a basis for a cliometric approach to mortality trends. We extend this vein of cliometric research in the context of demographic and epidemiological transitions, marked by time lags and disparities in the shape of age-specific mortality rate curves.

This study aims to improve long-term actuarial forecasts. Our innovative approach is based on mathematical generalization and age-specific flexibilization of the conventional models with cliometric inputs. The heterogeneous health

M. Corazza et al. (Eds.): MAF 2024, *Mathematical and Statistical Methods for Actuarial Sciences and Finance*, pp. 179–185, 2024.
https://doi.org/10.1007/978-3-031-64273-9_30

impact of climate change according to age reinforces the argument in favor of model flexibility. The article is structured in two sections: the first one deals with the problems associated with classical models and presents our methodological contributions, while the second one is devoted to the application of the models with sensitivity tests in various contexts.

2 Issues and Methodological Contributions

A mortality table describes the mortality process of a cohort of individuals from birth to death. Prospective tables are built from long-term forecasts of age and sex-specific mortality in a population. Classical internal models (i.e. without reference to exogenous population data) can be grouped into a family called Generalized Age-Period-Cohort (GAPC) models (Villegas *et al.*, 2015). The best-known models are Lee-Carter (LC), Cairns-Blake-Dowd (CBD), Renshaw-Haberman (RH), APC (age-period-cohort), M6, M7, M8 and Plat. LC and CBD are commonly used in the insurance and pension industries.

Throughout this document, we consider a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. The LC model is one of the classical reference models whose performance we will compare with that of the models we propose. In order to propose a generalization, we study and formalize the relationship between the classical LC model and principal component regression (PCR). The LC model focuses on $m_{x,t}$ which represents the central mortality rate between ages x and $x+1$ measured in year t. It is written (Boyer *et al.*, 2015):

$$\ln\left(m_{x,t}\right) = a_x + b_x k_t + \varepsilon_{x,t},$$

where a_x corresponds to the average of the logarithm of age-specific mortality rates calculated over the entire period under consideration, while k_t represents the general trend in mortality levels over time. b_x is the sensitivity to the common factor k_t specific to each age, and the residuals $\varepsilon_{x,t}$ are assumed to be i.i.d., with zero expectation and variance σ_ε^2 .

b_x and k_t are estimated using the first term of the singular value decomposition of the matrix $Z = \ln\left(m_{x,t}\right) - \widehat{a}_x$, (Gaba, 2021)

$$Z = \begin{bmatrix} z_{x_m,t_q} = \ln\left(m_{x_m,t_q}\right) - \widehat{a}_{x_m} & \cdots & z_{x_M,t_q} = \ln\left(m_{x_M,t_q}\right) - \widehat{a}_{x_M} \\ \vdots & \ddots & \vdots \\ z_{x_m,t_Q} = \ln\left(m_{x_m,t_Q}\right) - \widehat{a}_{x_m} & \cdots & z_{x_M,t_Q} = \ln\left(m_{x_M,t_Q}\right) - \widehat{a}_{x_M} \end{bmatrix}.$$

The singular value decomposition of Z is given by:

$$Z = \sum\nolimits_{i\geq 1} \sqrt{\lambda_i} u_i v_i^T,$$

where u_i is the i-th normalized eigenvector of ZZ^T corresponding to the eigenvalue λ_i :

$$ZZ^T u_i = \lambda_i u_i \text{ with } u_i^T u_i = 1.$$

Note that v_i is the i-th normalized eigenvector ($v_i^T v_i = 1$) of $Z^T Z$. This gives us the transfer formulae: $v_i = \frac{1}{\sqrt{\lambda_i}} Z^T u_i$ and $u_i = \frac{1}{\sqrt{\lambda_i}} Z v_i$.

Since the first eigenvalue λ_1 is greater by construction than the others, the best approximation of Z is then given by: $Z \approx \sqrt{\lambda_1} u_1 v_1^T$. By comparing the relationship $Z \approx \widehat{k_t}\widehat{b_x^T}$ and $Z \approx \sqrt{\lambda_1} u_1 v_1^T$ we have: $\widehat{k_t}\widehat{b_x^T} = \sqrt{\lambda_1} u_1 v_1^T$.

Taking into account the orders of $\widehat{k_t}$ and $\widehat{b_x}$ and the constraint ${b_x}^T b_x = 1$, we obtain by identification: $\widehat{b_x} = v_1$ and $\widehat{k_t} = \sqrt{\lambda_1} u_1$.

If we had used the alternative constraint $\sum_{x=x_m}^{x_M} b_x = 1$, we would have obtained the following result: $\widehat{b_x} = \frac{v_1}{\sum_j v_{1j}}$ and $\widehat{k_t} = \sqrt{\lambda_1} \sum_j v_{1j} u_1$ with $\sum_j v_{1j} \neq 0$.

Since the eigenvalues of $(Z^T Z)$ are distinct, each eigenvector v_i is unique (signed):

$z_i^* = Z v_i$ is the i-th principal component of Z.

We had previously established that under the constraint ${b_x}^T b_x = 1$, $\widehat{k_t} = \sqrt{\lambda_1} u_1$ and that $Z v_1 = \sqrt{\lambda_1} u_1$. So we have that $\widehat{k_t} = \sqrt{\lambda_1} u_1 = Z v_1 = z_1^*$. Thus, $\widehat{k_t}$ is equal to the first principal component of Z (z_1^*) and therefore the LC model is a regression (with a constant term) on z_1^*:

$$\ln(m_{x,t}) = a_x + b_x k_t + \varepsilon_{x,t} = a_x + b_x z_1^* + \varepsilon_{x,t}.$$

If we had used the alternative constraint $\sum_{x=x_m}^{x_M} b_x = 1$, we would have obtained the following result: $\widehat{k_t} = \sqrt{\lambda_1} \sum_j v_{1j} u_1 = Z v_1 \sum_j v_{1j} = z_1^* \sum_j v_{1j}$. So in this alternative case, $\widehat{k_t}$ is equal to the first principal component of Z (denoted z_1^*) by one multiplicative factor ($\sum_j v_{1j}$), and so the LC model remains a regression (with a constant term) on z_1^*:

$$\ln(m_{x,t}) = a_x + b_x k_t + \varepsilon_{x,t} = a_x + b_x (z_1^* \sum_j v_{1j}) + \varepsilon_{x,t}.$$

Combining the above demonstrated relationship between the LC model and PCR regression, and the CBD model which targets the *logit* of the probability of death, we formulate a nested PCR regression model with the first k principal components:

$$\text{logit} q_x(t) = a_x + b_{x,1} x_1^*(t) + b_{x,2} x_2^*(t) + b_{x,3} x_3^*(t) + \ldots + b_{x,k} x_k^*(t) + \varepsilon_{x,t},$$

where $q_x(t)$ is the probability of death at age x at time t while a_x is the average $\text{logit} q_x(t)$ for age x. x_i^* is the i-principal component of the matrix $X = \text{logit} q_x(t) - \widehat{a}_x$. Note that k ($k \leq p$) is the number of first principal components selected (by the *forward* method) with $p = M - m + 1$ and that $\varepsilon_{x,t}$ correspond to the residuals of the model. We refer to this model as **Logit-PCR**.

While a single temporal component reflects the demographic transition from the point of view of overall mortality, taking into account the epidemiological and health transition which affects ages in a differentiated way (Meslé & Vallin, 2000), requires age-differentiated temporal components. The aim now is to improve the Logit-PCR model by using one (or more) optimal principal components for each age modeled. We propose some classic methods for selecting explanatory series for the model: adjusted R^2, AIC, BIC or MSEP (which we will use). The model selection procedure is the *stepwise* method (used with adjusted R^2, AIC, BIC) or exhaustive search (which we will use). The new model is now formulated separately for each age:

$$\text{logit}q_x(t) = a_x + \beta^*_{c_1} x^*_{c_1} + \beta^*_{c_2} x^*_{c_2} + \ldots + \beta^*_{c_{k_x}} x^*_{c_{k_x}} + \varepsilon_{x,t},$$

where $x^*_{c_i}$ is a principal component of the matrix $X = \text{logit}q_x(t) - \widehat{a}_x$ (not the i-th principal component), while k_x ($k_x \leq p$) is the number of principal components selected. We refer to this internal model as **Logit-PCR-Optimal** or **PCR-Optimal**.

A limitation of the PCR-Optimal model above is that the calculation of the principal components does not take into account the variable to be explained. We resolve this limitation by using partial least square (PLS) components, which are calculated taking into account their correlations with the variable to be explained. The objective here is to find for each age group x the best k_x PLS components. The PLS components are denoted $l^{(1)}, l^{(2)}, \ldots, l^{(k)}$ linear combinations of the starting variables $l^{(j)} = Xc_j$which are orthogonal to each other and ranked in order of correlation with the variable to be explained. We refer to this new internal model as the **Logit-PLS** or **PLS**:

$$\text{logit}q_x(t) = r_1 l^{(1)} + \ldots + r_{k_x} l^{(k_x)} + \varepsilon_{k_x}.$$

Another limitation of classical mortality models is that they apply the same model class to all mortality rate curves, irrespective of age. To overcome this shortcoming, we introduce the concept of a composite model, allowing different classes to be mixed to model various ages in the population. We propose a **multifactorial composite model (CM)** using PCR-optimal or PLS models or their mean. Drawing inspiration from Barigou *et al.* (2021), the modeling is carried out in two iterations, enabling a more accurate and flexible approach to mortality rate modeling.

First iteration: for each age, we make an *a posteriori* choice of its best model class. The estimation history ranges from t_0 to $t_1$while forecasts are made from $t_1 + 1$ to t_2.

Second iteration: for each age, we use *a priori* its best model class chosen in iteration 1 to estimate its unbiased performance. The estimation history ranges from t_0 to t_2 while forecasts are made for $t_2 + 1$ to t_3. The times $t_0,\ t_1,\ t_2,\ t_3$ are arbitrary cut-off points in the data history.

3 Empirical Results and Sensitivity Tests

Frequent convergence problems have been detected for the APC, RH and M8 models depending on the data used and identification constraints. This finding is made in several publications including Diaz *et al.* (2018), Villegas *et al.* (2015), Hunt *et al.* (2015), Kennes (2017). In addition to the LC and LC Poisson (LC-P) models, the other models tested in our case were the CBD, CBD Log-Poisson (CBD-P), APC, RH, M7 and Plat models. For the APC, M7 and Plat models, we found convergence problems for India and Ecuador, for example. As a result of the above analyses, we will retain the following models as the classic internal models of reference in our empirical work: LC, LC LC-P and CBD-P.

The series modeled is *the probability of female death by age interval*. Data for India and Ecuador are taken from the Human Life-table Database (Max Planck Institute for Demographic Research, 2023), while those for European countries are taken from the Human Mortality Database (2023). Correction for the periods of the two World Wars is made by replacing mortality values with linear interpolations. Some countries are smoothed according to the shape of their mortality curves and the quality of their forecasts during the first iteration: this applies to India (moving average of order 9, as data are very patchy) and Spain (moving average of order 5).

We have studied the performance (and sensitivity) of the CM model in relation to various modeling parameters and usage situations, through a series of tests.

Modeling parameters: choices are made during the first iteration of composite modeling. This involves minimizing the forecast errors (*mean absolute percentage error*, MAPE) to select the best model class for each age bracket, the first year of the learning history and any initial transformation of the modeled series.

Model environments: the countries studied are chosen to vary in terms of geographical location (continent), cohorts and phases in the demographic transition process. Thus, forecasts are made over a 25-year horizon, for the following countries: France, Italy, Spain, India and Ecuador (Table 1).

Looking at the results obtained in different model environments (Table 1), we can make two major observations. Firstly, whatever the country tested and the forecasting horizon (up to 25 years), the CM model systematically outperformed conventional models by a wide margin. Secondly, the superior performance of the CM model is not systematic when we examine the results by age bracket, although the majority of age brackets are always in favor of the CM model. This limitation of the CM model should be borne in mind when using it, and also as an avenue for future improvement.

Robustness test: after reintroducing the mortality rates of the atypical years (1914 to 1918 and 1939 to 1945) of the two World Wars in France, the two main results mentioned above remain valid, although the relative performance of the MC model has been reduced over the 25-year horizon (MAPE 41% lower than the average for conventional models, compared with 55% previously).

Table 1. Model MAPE (%) for five countries via CBD-P, CM, LC and LC-P.

Country	France				Spain				Italy				India				Ecuador			
Model	CBD-P	CM	LC	LC-P	CBD-P	CM	LC	LC-P	CBD-P	CM	LC	LC-P	CBD-P	CM	LC	LC-P	CBD-P	CM	LC	LC-P
Horizon	1994-2018				1994-2018				1993-2017				1990-2014				1995-2019			
1	23,1	**9,7**	29,0	31,6	33,9	**7,6**	18,1	19,8	32,3	**8,4**	27,5	30,7	109,6	**9,3**	15,6	16,6	66,1	**4,4**	8,6	10,2
5	24,1	**10,7**	29,6	32,7	33,7	**7,8**	18,5	20,4	32,8	**9,1**	29,0	32,6	109,3	**9,4**	15,9	17,0	61,1	**7,1**	11,5	13,2
10	25,8	**10,7**	29,6	32,9	33,6	**7,9**	19,5	21,1	33,4	**8,4**	29,6	33,5	109,5	**10,8**	16,8	17,9	55,7	**12,7**	18,8	20,9
15	28,2	**12,0**	29,3	32,8	34,4	**9,3**	20,8	22,1	34,8	**11,3**	30,3	34,6	112,0	**12,7**	19,0	19,9	50,8	**15,0**	22,1	24,3
20	29,9	**13,0**	29,2	32,7	35,8	**11,2**	22,2	23,4	36,2	**14,2**	31,2	35,6	116,9	**15,4**	22,7	23,4	46,9	**17,0**	24,1	26,4
25	31,4	**13,8**	29,0	32,2	37,5	**12,6**	23,8	24,9	37,9	**16,7**	32,3	36,5	122,6	**19,6**	27,8	28,4	44,0	**19,3**	26,3	28,7
Age																				
0	98,6	**9,1**	15,0	66,3	99,6	23,6	29,1	**15,6**	99,6	25,0	**8,7**	55,6	96,5	19,7	20,9	**9,6**	98,4	**5,2**	37,1	42,8
1-4	65,2	**18,6**	45,3	61,2	90,4	**18,5**	54,0	69,6	89,8	**32,5**	68,7	64,6	**59,2**	122,9	117,5	95,2	70,6	**24,8**	49,7	68,8
5-9	**7,7**	20,1	45,9	78,1	69,8	**5,6**	33,3	42,1	70,2	**11,9**	49,4	52,6	77,9	**63,0**	84,0	83,9	**19,1**	32,8	21,6	21,2
10-14	38,2	**9,5**	45,6	54,6	56,0	**2,9**	31,6	34,8	54,6	**24,5**	47,7	50,4	239,8	**13,0**	37,8	49,1	32,2	**25,4**	38,7	42,9
15-19	**9,3**	24,9	35,6	22,2	59,6	**24,6**	26,1	32,9	59,2	**57,0**	60,4	66,5	149,5	**4,4**	18,0	20,2	22,0	**17,7**	37,1	40,7
20-24	**11,0**	**11,0**	36,3	16,8	43,3	**22,5**	28,8	34,9	44,0	**37,2**	62,1	66,9	136,3	**5,6**	14,3	17,5	27,7	**27,5**	34,6	36,6
25-29	34,8	21,9	34,8	**17,8**	26,2	**18,1**	31,1	31,7	**20,8**	37,7	56,6	59,4	182,0	**6,0**	23,8	27,9	47,1	39,1	**35,3**	35,4
30-34	45,3	**19,0**	33,4	24,5	15,8	**10,7**	31,7	30,4	**8,3**	24,4	49,9	50,9	223,9	**5,2**	27,6	34,5	65,5	33,3	**30,7**	31,1
35-39	35,3	**17,7**	31,8	28,2	**12,7**	14,7	33,8	31,9	9,9	**9,1**	40,6	39,8	241,0	**4,0**	27,2	34,9	73,1	**29,2**	29,5	30,3
40-44	24,4	**10,8**	29,0	27,0	**13,8**	16,4	30,0	27,3	10,9	**3,2**	29,0	25,9	235,6	**6,4**	27,7	32,8	69,5	**16,6**	25,0	26,5
45-49	16,5	**14,3**	21,3	17,6	15,9	**15,4**	23,0	19,9	11,3	**2,5**	15,2	10,3	169,0	**7,0**	11,9	13,0	62,4	**14,6**	23,3	24,4
50-54	17,8	**10,0**	15,0	10,9	16,8	**11,6**	16,3	13,4	14,8	**2,6**	8,1	2,8	110,1	**5,7**	8,4	6,0	50,6	**7,7**	17,5	18,1
55-59	25,1	7,2	7,6	**2,8**	20,4	**8,3**	10,5	9,3	24,4	3,5	**1,9**	5,2	61,3	8,3	**2,3**	4,0	39,2	**2,8**	8,8	9,0
60-64	32,2	4,8	**3,1**	4,3	28,7	**5,3**	7,0	6,2	32,7	4,1	**3,9**	9,1	16,5	**2,4**	3,0	**2,4**	27,8	7,0	**3,2**	**3,2**
65-69	34,1	**3,8**	14,0	22,9	31,9	**2,9**	3,1	4,9	36,3	**3,2**	9,0	14,7	11,3	7,4	**4,1**	5,6	18,9	**7,7**	9,2	9,6
70-74	23,9	**10,5**	29,3	38,6	22,2	4,4	**3,5**	6,4	33,2	**3,3**	15,6	21,4	31,1	26,4	**21,7**	23,5	**13,2**	21,8	21,7	22,3
75-79	**14,2**	21,1	49,4	54,0	14,7	**9,0**	12,5	12,9	24,4	**2,9**	21,7	24,9	42,5	25,6	**21,7**	22,8	**11,3**	16,0	24,7	25,2

4 Conclusion

Our aim was to improve long-term actuarial forecasts of age-specific mortality rates, taking into account demographic and epidemiological transitions. We introduced the PCR-optimal or PLS multifactor composite internal model, which generalizes and flexibilizes classical approaches by considering time lags and variations in mortality curves. Empirical results over 25 years and in a variety of contexts, including France, Italy, Spain, India and Ecuador, systematically demonstrate the very good performance of the multifactor composite (CM) model in comparison to conventional models. However, nuances by age group suggest avenues for future improvement.

Further work will be needed to calculate the confidence intervals of the model's forecasts, in particular using a Monte-Carlo approach.

The universality of the demographic transition, which is spread across cohorts of countries, can also be taken into account to improve CM models for emerging and developing countries, via relational cliometric models (i.e., with reference to data from an exogenous population) that we will have to develop.

References

Barigou, K., Goffard, P.-O., Loisel, S., Salhi, Y.: Bayesian model averaging for mortality forecasting using leave-future-out validation. Int. J. Forecast. (2021). https://doi.org/10.1016/j.ijforecast.2022.01.011.hal-03175212v3. In press

Boyer, M., Dorion, C., Stentoft, L.: Les modèles factoriels et la gestion du risque de longévité. L'Actualité Écon. **91**(4), 531–565 (2015)

Floud, R., Harris, B.: Health, height, and welfare: Britain 1700–1980. In: Steckel, R.H., Floud, R. (eds.) Health and Welfare during Industrialisation, pp. 91–126. University of Chicago Press, Chicago (1997)

Fogel, R.: Physical growth as a measure of the economic well-being of populations: the eighteenth and nineteenth centuries. In: Falkner, F., Tanner, J.M. (eds.) Human Growth: A Comprehensive Treatise, vol. 3, 2nd edn., pp. 263–281. Plenum, New York (1986)

Diaz, G., Debón, A., Giner-Bosch, V.: Mortality forecasting in Colombia from abridged life tables by sex. Genus **74**(15) (2018)

Gaba, K.G.: Cliométrie de l'histoire globale des transitions anthropologique, cognitive, institutionnelle, démographique et économique. Thèse de doctorat, Université de Lyon 1 (2021)

Human Mortality Database (2023). https://www.mortality.org/cgi-bin/hmd/hmd_download.php. Accessed 08 Mar 2023

Hunt, A., Villegas, A.M.: Robustness and convergence in the Lee-Carter model with cohort effects. Insur. Math. Econ. **64**, 186–202 (2015)

Kennes, T.: The convergence and robustness of cohort extensions of mortality models. MaRBLe **1**, 36–53 (2017)

Max Planck Institute for Demographic Research. Demographic Data (2023). https://www.demogr.mpg.de/en/research_6120/demographic_data_27/. Accessed 08 Mar 2023

Meslé F., Vallin J.: Transition sanitaire: tendances et perspectives. Méd./Sci. **16**, 116–171 (2000)

Villegas, A.M., Millossovich, P., Kaishev, V.K.: StMoMo: an R package for stochastic mortality modelling. J. Stat. Softw. **66**(3), 1–20 (2015)

How Does Covid-19 Shock Financially Impact the US PAYG Pension Scheme? An Automatic Balance Mechanism Approach

Frédéric Gannon[1], Florence Legros[2], and Vincent Touzé[3(✉)]

[1] Université Le Havre Normandie, France & Sciences Po-OFCE, Paris, France
frederic.gannon@univ-lehavre.fr
[2] ICN-Business School, Nancy, France
florence.legros@icn-artem.com
[3] Sciences Po-OFCE, Paris, France
vincent.touze@sciencespo.fr

Abstract. The COVID-19 pandemic crisis financially impacts the PAYG pension schemes. In the short run, the economic shock induced a decrease in receipts due to the contraction of the GDP. In the medium and long run, the return to the pre-crisis level depends on the ability of the economy to rebound. As to expenses, in the very short term, they do not decrease due to the inertia. However, the fall in contributions is progressively taken into account to calculate the new pensions, which will gradually contract expenditures. The demographic impact is mirrored in a moderate increase in mortality, mainly that of retirees. We develop a macroeconomic model that simulates these changes and is then used to revise the pre-crisis forecasts in pension expenditures and payroll tax receipts. Relying on the US Social Security 75-year forecast, we analyze the impact of the crisis on the financial balance of its pension system. We assess the extent of the additional fiscal adjustments required to restore financial equilibrium through an Automatic Balancing Mechanism (ABM).

Keywords: Pension scheme · Covid-19 · Automatic Balancing Mechanism

1 Introduction

The 2020–2021 pandemic obviously impacts the economy by putting under pressure the public finances and notably the unfunded pension schemes. Feher and Bidegain (2020) and OECD (2020) early warn about the impact of COVID-19 induced recession and mortality on the solvency and adequacy of pension schemes. Tackling the pandemic effects on pension issues is both obvious and urgent. At the individual level, the induced risks are multiple (Rappaport and

M. Corazza et al. (Eds.): MAF 2024, *Mathematical and Statistical Methods for Actuarial Sciences and Finance*, pp. 186–191, 2024.
https://doi.org/10.1007/978-3-031-64273-9_31

Siegel, 2021). From the macroeconomic standpoint, in the short run, the public pension schemes are mainly affected through the receipts channel because the decline in the GPD induces less revenue from payroll tax whereas the pension expenditures show great inertia. In the long run, financial balance is sensitive to the indexation of future pensions to change in GPD and to the permanent impact on revenue. In a recent paper, Munnell and Chen (2021) concludes that "COVID is not a retirement story, but the pre-COVID weaknesses in the retirement remain". Even if the unfunded pension schemes were to be reformed before the pandemic, it can be of interest to assess the future additional adjustments related to the macroeconomic impacts of this exogenous shock.

We develop a simple macroeconomic model to simulate the induced deviation from a benchmark scenario of Social Security forecast. Assessing the issue of financial solvency and the potential need for balancing mechanisms, we use the smooth Automatic Balancing Model (S-ABM) based on an optimal control approach developed by Gannon and al. (2020). We apply the model to the US Social Security and evaluate the induced changes in the automatic adjustment parameters.

A simular study by Fratoni (2022) also relies on an Automatic Balance Mechanism (ABM) to assess the impact of COVID-19 shock. However, his approach differs from ours since he characterizes the change in each demo-economic factors: unemployment rate, wage growth rate, inflation rate, mortality rates and disability inception rates.

Next section details the deviation macroeconomic model and calibrates it to mimic the 75-year US Social Security forecast changes between 2019 Trustees' report (hereafter TR) and the successive reports (from 2020 to 2023). Section 3 assesses the induced changes in the dynamic paths of the S-ABM. Sect. 4 discusses the sensitivity of changes in the parameter values. Last section concludes.

2 Financial Impact of COVID 19: Modelling the Deviation from a Benchmark Scenario

To evaluate the financial impact of COVID-19, we set up a simple deviation model. First, we assume payroll tax receipts are proportional to GDP in the same way as before the health crisis. The GDP, denoted Y_t after COVID shock, is developed as follows depending on the pre-crisis expected value denoted $\hat{Y}_{t-1}$:

$$\frac{Y_t}{\hat{Y}_t} = \left(\frac{Y_{t-1}}{\hat{Y}_{t-1}}\right)^{1-\sigma_Y} \lambda^{t-2020} + \left(1 - \lambda^{t-2020}\right) \left(\frac{Y_{LR}}{\hat{Y}_{LR}}\right) \tag{1}$$

where $\frac{Y_{2020}}{\hat{Y}_{2020}}$ is the initial shock and $\frac{Y_{LR}}{\hat{Y}_{LR}}$ the long-run permanent deviation of GPD with respect to the pre-crisis value. σ_Y is a parameter of convergence to the pre-crisis GDP level and λ an inertia factor of the short-term COVID-19 shock.

The receipts evolve as follows:

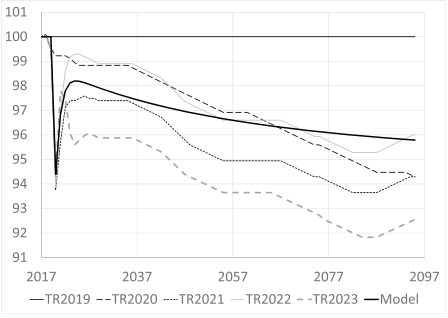
$$REC_t = \frac{\widehat{REC_t}}{\hat{Y}_t} \cdot Y_t \tag{2}$$

where $\widehat{REC_t}$ is the forecasted pre-crisis level of the receipts. In this simulation, the shock on receipts is seen as a pure shock linking their evolution to that of the GDP. In practice, government measures to support activity (financing of partial unemployment, for example) have enabled beneficiary companies to continue to pay salaries and therefore to pay social contributions. This aspect is ignored in our computations.

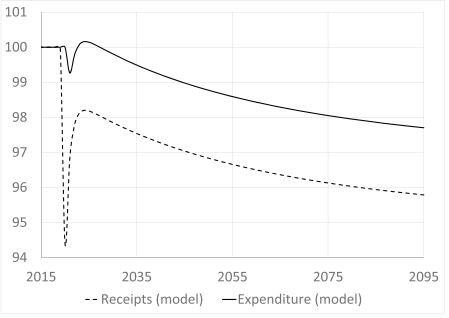
The impact on mortality rates during 2020 and 2021 caused by COVID-19 on the over-75 population has been quantified notably by the WHO over 15% in high income countries. In this paper, we model the impact on the mass of pensions without distinguishing explicitly the number of retirees and the average of pensions. Hence we assumed that the mass of pensions initially planned has strong inertia and evolves as follows:

$$EXP_t = \beta^{t-2020} \cdot \widehat{EXP_t} + \left(1 - \beta^{t-2020}\right) \cdot (1+\mu) \cdot \frac{\widehat{EXP_t}}{\widehat{REC_t}} \cdot REC_t \qquad (3)$$

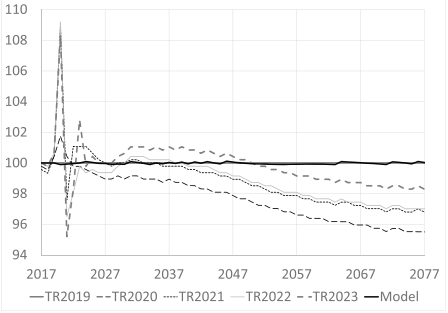
where β captures the inertia factor of pension expenditure and μ is the long-run permanent deviation of the level of expenditures from the pre-shock value.



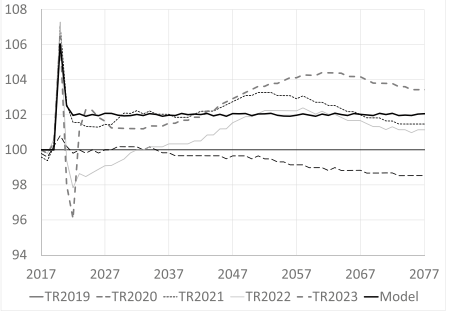
(a) GDP



(b) Receipts and Expenditures



(c) Receipts rate



(d) Expenditures rate

Fig. 1. Deviation from 2019 Trustee report (TR2019) projection.

We assign the following values to the parameters: $\sigma_Y = 0.4$, $\lambda = 0.985$, $\mu = 2\%$ and $\beta = 0.4$ and the following impact of COVID-19 on the GDP $\frac{Y_{2020}}{\tilde{Y}_{2020}} = 94.4\%$ (5.6% decrease in GDP) in the short-run and $\frac{Y_{LR}}{\tilde{Y}_{LR}} = 95\%$ in the long-run. We estimate new forecast values of GDP (Fig. 1a), the receipts (Figs. 1b and 1c) and the expenditures (Figs. 1b and 1d). Our 5-parameter deviation model produces intermediate values between the GDP (resp. expenditure ratio) forecasts from the 2021 and 2022 (resp. 2021 and 2023) reports, compared to the previous 2019 report. From 2021 onwards, the US Social Security Actuarial Office has lowered the share of taxable income in GDP over the long term (see Fig. 1c). This decision is independent of the COVID-19 pandemic. We therefore do not take this change into account. TR2023 includes the new geopolitical context and a lower economic growth path induced by the Russia-Ukraine conflict. We have not calibrated the growth regime to this slightly more pessimistic scenario.

3 COVID 19 Impact: Deviation from the Benchmark Scenario

We use the S-ABM developped in Gannon et al. (2020). The optimal control problem is the following:

$$\begin{cases} \min_{\{A_t,B_t\}_{t=1\dots T}} \sum_{t=2020}^{2094} \frac{1}{(1+\delta)^t} L(A_t, B_t) \\ \sum_{t=2020}^{2094} A_t \frac{REC_t}{\Pi_{j=2020}^{t}(1+r_j)} + F_{2019} = \sum_{t=2020}^{2094} B_t \frac{EXP_t}{\Pi_{j=2020}^{t}(1+r_j)} \end{cases} \tag{4}$$

where $L(A_t, B_t) = (A_t - 1)^2 + (B_t - 1)^2$ is the current loss function to minimize, r_t is the free-risk interest rate, $\delta = 1.5\%$ is the social rate of time preference, EXP_t are the pension expenditures, REC_t are the contribution receipts and F_0 is the buffer fund in period 0. A_t and B_t are the two deformation coefficients modifying respectively the payroll tax rate (receipts) and the pension benefits (expenditures). The optimal values of these coefficients guarantee the financial sustainability of the pension schemes and can be expressed as follows.

(i) Final forecasted adjustment in 2094:

$$\begin{cases} A_{2094} = 1 + \frac{UO_{2019}}{REC_{2094}/\Pi_{i=2020}^{2094}\left(\frac{1+r_i}{1+\delta}\right)} \Big/ \sum_{t=2020}^{2094} \frac{REC_t^2 + EXP_t^2}{\Pi_{i=2020}^{t} \frac{(1+r_i)^2}{1+\delta}} \\ B_{2094} = 1 - \frac{EXP_{2094}}{REC_{2094}} \cdot (1 - A_{2094}) \end{cases} \tag{5}$$

where UO_{2019} is the value of the Unfunded Obligation at the end of 2019.

(ii) Forecasted convergence rule to the final adjustments for $2020 \leq t < 2094$:

$$\begin{cases} (A_t - 1) = \frac{REC_t}{REC_{2094}} \cdot \Pi_{i=t+1}^{2094} \left(\frac{1+r_i}{1+\delta}\right) \cdot (A_{2094} - 1) \\ (B_t - 1) = \frac{EXP_t}{EXP_{2094}} \cdot \Pi_{i=t+1}^{2094} \left(\frac{1+r_i}{1+\delta}\right) \cdot (B_{2094} - 1) \end{cases} \tag{6}$$

Figure 2 gives the change in the trajectories of the adjustment pair (A_t, B_t) with respect to the pre-crisis forecasted adjustments $\left(\hat{A}_t, \hat{B}_t\right)$. The S-ABM suggests the following additional adjustments $\left(A_t - \hat{A}_t, B_t - \hat{B}_t\right)$:

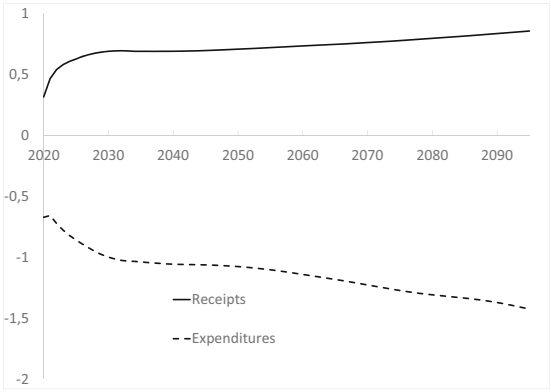


Fig. 2. Change in the Smooth-ABM (A_t: receipts and B_t: expenditures)

1) in the short-run: a +0.3% point increase in payroll tax receipts and a –0.7% point decrease in pension expenditures;
2) in the long-run: a +0.9% point increase in payroll tax receipts and a –1.42% point decrease in pension expenditures.

Table 1. Sensitivity analysis to receipts (GDP) change (in % point change)

Permanent shock					Inertia to initial shock				
$\frac{Y_{LR}}{\hat{Y}_{LR}}$	ΔA_0	ΔA_T	ΔB_0	ΔB_T	λ	ΔA_0	ΔA_T	ΔB_0	ΔB_T
100%	0.19	1.07	–0.53	–1.71	1	0.20	1.08	–0.54	–1.71
97.5%	0.27	0.97	–0.62	–1.58	0.99	0.33	0.86	–0.68	–1.44
95%	0.34	0.87	–0.74	–1.44	0.985	0.34	0.87	–0.74	–1.44
92.5%	0.42	0.77	–0.79	–1.30	0.98	0.36	0.88	–0.72	–1.45
90%	0.50	0.66	–0.88	–1.15	0.97	0.37	0.89	–0.73	–1.46

4 Sensitivity to Deviation Parameters

The model is calibrated to reproduce a (relative) additional expenditure regardless of the possibility of the GDP returning to its pre-crisis level. Also, alternative assumptions on GDP dynamics (Tables 1 and 2) only marginally influence the need for financing after the crisis. Consequently, the long run ratio of GDP deviation $\left(\frac{Y_{LR}}{\hat{Y}_{LR}}\right)$ mainly modifies the slope of the adjustment dynamics. With respect to the benchmark calibration (95%), the higher (resp. lower) its value, the weaker (resp. stronger) the short-run adjustment and the stronger (resp. weaker) the long-run adjustment. The inertia factor to the initial COVID-19 economic shock (λ) has a simular impact. The pre-crisis expenditures inertia

Table 2. Sensitivity analysis to expenditure change (in % point change)

Inertia to pre-crisis level					Permanent excess				
β	ΔA_0	ΔA_T	ΔB_0	ΔB_T	μ	ΔA_0	ΔA_T	ΔB_0	ΔB_T
1	0.69	1.63	–1.10	–2.90	4%	0.77	1.80	–1.19	–3.00
0.75	0.35	0.89	–0.71	–1.46	3%	0.56	1.34	–0.95	–2.22
0.4	0.34	0.87	–0.74	–1.44	2%	0.34	0.87	–0.74	–1.44
0.1	0.34	0.86	–0.70	–1.43	1%	0.12	0.39	–0.45	–0.65
0.0	0.34	0.86	–0.70	–1.43	0.5%	0.07	0.22	–0.39	–0.35

level parameter (β) mainly impacts the solvency of the pension scheme (Tables 1 and 2). The higher (resp. lower) its value, the stronger (the weaker) the short-run adjustment and the long-run adjustment. For small values of β, the variations in adjustments tend to a limit value. The permanent excess parameter (μ) has a similar impact.

5 Conclusion

In this article, we have set up a deviation model of the payroll tax receipts and pension expenditures trajectories after the macroeconomic shock induced by the pandemic. Using an Automatic Balance Mechanism (ABM), we have assessed the additional adjustments required to satisfy the intertemporal financial equilibrium. While these additional adjustments are moderate with respect to the already existing needs before the pandemic shock, the range of the latter mainly depends on the lasting consequences both on the GDP and on the capacity of the pension system to update the mass of pensions to the new level of payroll tax mass. Our computations (and calibration), based on the US OASDI forecasts, show that the COVID-19 shock has a financial permanent impact. In comparison with the pre-COVID unfunded obligations, it seems marginal but not nil. Therefore, COVID-19 is partly "a retirement story".

References

Board of Trustees: 2019-2023 annual reports. Federal Old-Age and Survivors Insurance and Federal Disability Insurance Trust Funds (2019-2023)

Fratoni, L.: The Italian NDC pension system: resilience to the COVID-19 and chances for an LTC coverage with ABMs. Ph.D. thesis. Sapienza, University of Rome. (2022)

Gannon, F., Legros, F., Touzé, V.: Sustainability of pension schemes: building a smooth automatic balance mechanism with an application to the US Social Security. Revue de l'OFCE, n° 170 (2020)

Munnell, A.H., Chen, A.: COVID-19 is not a retirement story. Issue in Brief, Center for Retirement Research, N° 21-4 (2021)

Rappaport, A., Siegel, S.: Impact of COVID-19 on Retirement Risks. Society of Actuaries, Note, April 2021

The Risk of War: An Analysis Combining Real Options and Games

Laurent Gauthier[1,2(✉)]

[1] LED (Université Paris 8), Saint-Denis, France
laurent.o.gauthier@gmail.com
[2] CAC–IXXI (ENS Lyon), Lyon, France

Abstract. Real options have been used to evaluate investment decisions with various structures, but have rarely been put into a game theoretical context. We examine the risk of war over a common resource, whose value is a geometric Brownian Motion, where one country may preemptively appropriate the resource, and the other wage war for it. We derive a closed form expression for the optimal mixed strategy that both countries should follow, and show that the resulting price level that triggers capturing the resource and subsequent war asymptotically follows a power law. The present value of the time until war is triggered also decays slowly asymptotically. In consequence, in spite of both countries' propensity towards preemption, war may be indefinitely delayed.

Keywords: Real options · resource wars · games · mixed strategies

1 Introduction

Considering a resource that may be accessed by two countries in potential conflict, the context is comparable, to some extent, to that of investment decision in a competitive environment —a type of economic problem that has been approached through real options. Real options are indeed useful in evaluating investment or disinvestment opportunities [7], but reflecting strategic behavior in a competitive environment is complex, and requires combining real options and games, an area in which there has been some research [for example 2, 8]. The formal framework for preemption games in continuous time is laid out by [4,5].

In our approach, we consider mixed (randomized) strategies, which offer non-deterministic surprise, an important aspect in war. This requires following a probabilistic approach in studying the real options embedded in the model, because of the added complexity of optimal randomized strategies. This is distinct from usual approaches in real options where the optimal strategy is often directly obtained through a differential equation [7]: we first express the present value of decision variables, and then optimize, as in [1]. The second section presents the model and real option framework. The third section embeds the real options in a strategic framework and shows that the optimal distribution

M. Corazza et al. (Eds.): MAF 2024, *Mathematical and Statistical Methods for Actuarial Sciences and Finance*, pp. 192–197, 2024.
https://doi.org/10.1007/978-3-031-64273-9_32

of the threshold at which one country will open the hostilities by seizing a commonly accessible resource asymptotically follows a power law, and that the time to conflict decays slowly.

2 A Real Option Model for Resource Appropriation

We consider two countries, indexed as i and $-i$. The market price of some resource is given by a stochastic process $(S_t)_{t\geq 0}$, a geometric Brownian Motion solution of $dS_t = \mu dt + \sigma dB_t$. We assume that $\mu < \frac{\sigma^2}{2}$. The Geometric BM is a standard assumption in the real-option literature, and is generally considered a good representation for the value of commodities on which real option-based investment strategies are decided, such as oil or natural gas [6,7]. Each country needs to use this resource and acquires it on the market. The quantity of the resource in question, that must be consumed at each point in time, is a fixed amount ψ_i per unit of time. The total cost of using the resource at time t is therefore $\psi_i S_t$. Decisions are made using an actuarial logic, and both countries use the same discounting rate $\rho > \mu \vee 0$. Hence, for Country i, the expected cost of future resource procurement R^i at time 0 can be written:

$$R^i(S_0) = \mathbb{E}_{S_0}\left[-\int_0^\infty e^{-\rho s}\psi_i S_s ds\right] = \frac{-\psi_i S_0}{\rho - \mu} = -\alpha_i S_0.$$

With our assumptions $\alpha_i > 0$. For a cost C_i, let us assume Country i can appropriate territory where the resource can be found, in which case these resources do not need to be acquired in the market. Then, after this territory is occupied, the procurement cost becomes 0.

If Country i is the only one who could appropriate the resource, then this decision corresponds to a standard real option. As the cost only depends on the resource price S_t, the optimal decision threshold to appropriate the resource must be expressed as a hitting time of the form $T_h = \inf\{u \geq 0 : S_u = h\}$, with $h \geq S_0$. As a function of the threshold h, the total procurement cost, depending on the decision threshold h, can be written:

$$R^i_h(S_0) = \mathbb{E}_{S_0}\left[-\int_0^{T_h} e^{-\rho s}\psi_i S_s ds\right] - C_i \mathbb{E}_{S_0}\left[e^{-\rho T_h}\right]$$

In order to compute $\mathbb{E}_{S_0}\left[-\int_0^{T_h} e^{-\rho s}\psi_i S_s ds\right]$, we simply write it as:

$$\mathbb{E}_{S_0}\left[-\int_0^\infty e^{-\rho s}\psi_i S_s ds\right] + \mathbb{E}_{S_0}\left[\int_{T_h}^\infty e^{-\rho s}\psi_i S_s ds\right] = R^i(S_0) - \mathbb{E}_{S_0}\left[e^{-\rho T_h}\right] R^i(h).$$

We use the strong Markov property of the Brownian Motion applied at the stopping time T_h: the process starting from T_h has the same law as the process starting from h at any point in time. The Laplace transform of a geometric Brownian Motion hitting time is well known [3] and for $h \geq S_0$, $\mathbb{E}_{S_0}\left[e^{-\rho T_h}\right] =$

$\left(\frac{S_0}{h}\right)^{\frac{1}{2}-\frac{\mu}{\sigma^2}+\sqrt{\frac{2\rho}{\sigma^2}+\left(\frac{1}{2}-\frac{\mu}{\sigma^2}\right)^2}}$. We will write $\beta = \frac{1}{2} - \frac{\mu}{\sigma^2} + \sqrt{\frac{2\rho}{\sigma^2} + \left(\frac{1}{2} - \frac{\mu}{\sigma^2}\right)^2}$. Since $\mu < \frac{\sigma^2}{2}$ and $\rho > \mu$, we have $\beta > 1$. Hence, we obtain the value of the potential appropriation strategy $R_h^i(S_0) = -\alpha_i S_0 + \left(\frac{S_0}{h}\right)^\beta (\alpha_i h - C_i)$. This value can be maximized as a function of h to find the optimal strategy. Solving for $\frac{\partial R_h^i(S_t)}{\partial h} = 0$, we obtain: $h_i^* = C_i \frac{\beta}{\alpha_i(\beta-1)}$. The threshold expresses a ratio between the entry cost and the future cost of having to buy the resource.

3 The Option of War in a Strategic Setting

3.1 Impact of Conflict

In a strategic context, Country $-i$ may have already taken possession of the resource, or may decide to attack Country i. We therefore need to make particular assumptions about the impact of the conflict situation on the costs. The cost C_i we considered earlier is only applicable if the territory to be appropriated is not already occupied. If it is occupied, appropriating it requires waging war. We model war as a fixed cost that is applied instantaneously. Costs are suffered by both the attacker and the defender: W_i is the cost for Country i attacking Country $-i$, and D_i is the cost for Country i defending against Country $-i$. After the attacker wages war, both countries will share the resource, with an allocation θ_i for Country i and $\theta_{-i} = 1 - \theta_i$ for Country $-i$. After this war, the country which held the full territory loses a share of its benefit in exploiting the resource, as a consequence.

Assume Country i holds the territory and exploits the resource at time 0. Country $-i$ sets a threshold l_{-i} such that when it is reached, they will wage war. In these conditions, the strategy's value to Country $-i$ can be expressed in a similar way to what we have calculated earlier for $R_h^i(S_0)$:

$$R_{l_{-i}}^{-i}(S_0) = -\alpha_{-i} S_0 + \left(\frac{S_0}{l_{-i}}\right)^\beta (\theta_{-i}\alpha_{-i} l_{-i} - W_{-i}) .$$

From Country i's perspective, who is exploiting the resource alone initially, the expected future costs of procuring the resource is 0 if Country $-i$ does not wage war. If it does, then it becomes:

$$R_{l_{-i}}^{i}(S_0) = -\left(\frac{S_0}{l_{-i}}\right)^\beta (\theta_i \alpha_i l_{-i} + D_i) .$$

We can therefore write the value of the strategies (l_i, l_{-i}) before any of the countries takes the resource, in all generality for Country i as:

$$\begin{aligned} R_{l_i,l_{-i}}^{i}(S_0) =& \mathbb{I}_{l_i<l_{-i}} \left[-\alpha_i S_0 + \left(\frac{S_0}{l_i}\right)^\beta (\alpha_i l_i - C_i) - \left(\frac{S_0}{l_{-i}}\right)^\beta ((1-\theta_i)\alpha_i l_{-i} + D_i) \right] \\ &+ \mathbb{I}_{l_i \geq l_{-i}} \left[-\alpha_i S_0 + \left(\frac{S_0}{l_i}\right)^\beta (\theta_i \alpha_i l_i - W_i) \right] . \end{aligned}$$

Given l_{-i}, one can determine the optimal level for $l_i \geq l_{-i}$ in a deterministic fashion. Indeed, Country i, in this case can find its optimal attack threshold with the only condition that it be greater than l_{-i}. Letting Country $-i$ take over the resource first, results in a value, for Country i, of:

$$-\alpha_i S_0 + \left(\frac{S_0}{l_i}\right)^{\beta} (\theta_i \alpha_i l_i - W_i) .$$

The optimal war threshold without any constraint is therefore $T_i = \frac{W_i \beta}{\theta_i \alpha_i (\beta - 1)}$. However, since we must have $l_i \geq l_{-i}$, the optimal attack takes place when the underlying price reaches $l_{-i} \vee T_i$.

3.2 Game-Theoretical Equilibrium

Each country needs to determine an optimal strategy to set the l thresholds. To find an optimal strategy, we assume that thresholds are drawn from an optimal probability distribution at the Bayes-Nash equilibrium, allowing fully mixed strategies, which is possible [5]. The dominating strategies that consist in preempting as soon as early preemption is as valuable as being a follower [4] may be applicable in economic competition. However, we argue that in the context of war, it is not possible to allow the enemy to know where one's army will be at a discrete point in time, so that the optimal mixed strategy cannot have a compact support. At the optimum, if a player is drawing from the optimal distribution, then whatever pure strategy the other player follows, should result in the same value to them.

We write L_i and L_{-i} the independent random variables corresponding to the optimal strategic choice for both countries, picked over all possible passage times in the future. From the standpoint of Country i, we write the value of these strategies $R^{*,i}_{L_i,L_{-i}}(S_0)$, depending on the relative values of L_i and L_{-i}, as $R^{*,i}_{L_i,L_{-i},\geq}(S_0)$ or $R^{*,i}_{L_i,L_{-i},<}(S_0)$. If $L_i \geq L_{-i}$, then Country i can determine their optimal attack threshold with the only constraint that it be higher than L_{-i}. Therefore, in this case:

$$R^{*,i}_{L_i,L_{-i},\geq}(S_0) = -\alpha_i S_0 + \left(\frac{S_0}{L_{-i} \vee T_i}\right)^{\beta} (\theta_i \alpha_i (L_{-i} \vee T_i) - W_i) .$$

If on the contrary $L_i < L_{-i}$, then i is the first to take over the territory once price L_i is reached, but then Country $-i$ determine their own optimal level, which is symmetrical to the prior case, and equal to: $L_i \vee T_{-i}$. Therefore, in this case:

$$\begin{aligned} R^{*,i}_{L_i,L_{-i},<}(S_0) = &- \alpha_i S_0 - \left(\frac{S_0}{L_i}\right)^{\beta} (\theta_i \alpha_i L_i - C_i) \\ &- \left(\frac{S_0}{L_i \vee T_{-i}}\right)^{\beta} ((1-\theta_i)\alpha_i (L_i \vee T_{-i}) + D_i) . \end{aligned}$$

At the optimum in a Bayes-Nash equilibrium, the strategy followed by $-i$, in setting the distribution of L_{-i}, should make Country i indifferent to any specific deterministic choice of L_i. Therefore, there exists a constant independent of u equal to:

$$\int_{S_0}^{\infty} \mathbb{P}[L_{-i} \in dv] R^{*,i}_{u,v}(S_0) = \int_{S_0}^{u} \mathbb{P}[L_{-i} \in dv] R^{*,i}_{u,v,\geq}(S_0) + \mathbb{P}[L_{-i} \geq u] R^{*,i}_{u,v,<}(S_0).$$

since $R^{*,i}_{u,v,<}(S_0)$ does not depend on v, and $R^{*,i}_{u,v,\geq}(S_0)$ does not depend on u. We write the density $\mathbb{P}[L_{-i} \in dv] = f_{L_{-i}}(v)dv$, which cannot depend on L_i. Taking the derivative of the above equation with respect to u, we obtain:

$$0 = f_{L_{-i}}(u) \left(R^{*,i}_{u,u,\geq}(S_0) - R^{*,i}_{u,v,<}(S_0) \right) + \frac{\partial}{\partial u} R^{*,i}_{u,v,<}(S_0) \mathbb{P}[L_{-i} \geq u].$$

We write $F_{L_{-i}}(u) = \mathbb{P}[L_{-i} \geq u]$, so that $F'_{L_{-i}} = -f_{L_{-i}}$. The equation can be rewritten:

$$F'_{L_{-i}}(u) = F_{L_{-i}}(u) \frac{\frac{\partial}{\partial u} R^{*,i}_{u,v,<}(S_0)}{R^{*,i}_{u,u,\geq}(S_0) - R^{*,i}_{u,v,<}(S_0)}.$$

We have:

$$\begin{aligned}
R^{*,i}_{u,u,\geq}(S_0) =& -\alpha_i S_0 + \left(\frac{S_0}{u \vee T_i} \right)^{\beta} (\theta_i \alpha_i (u \vee T_i) - W_i) \\
R^{*,i}_{u,v,<}(S_0) =& -\alpha_i S_0 - \left(\frac{S_0}{u} \right)^{\beta} (\theta_i \alpha_i u - C_i) \\
& - \left(\frac{S_0}{u \vee T_{-i}} \right)^{\beta} ((1-\theta_i) \alpha_i (u \vee T_{-i}) + D_i).
\end{aligned}$$

If we assume that $S_0 = 1$, without much loss in generality, and if the characteristics of both countries are the same, so that $T_i = T_{-i}$, we obtain the solution:

$$\mathbb{P}[L_{-i} \geq u] = \exp\left(- \int_1^u dx \frac{\theta_i \alpha_i (\beta - 1) - \frac{C_i}{x} + \mathbb{I}_{x<T_i} (1-\theta_i) \alpha_i (\beta - 1) - \mathbb{I}_{x<T_i} \frac{D_i}{x}}{\alpha_i \theta_i x - C_i + x^{\beta} (x \vee T_i)^{-\beta} (\alpha_i (x \vee T_i) + D_i - W_i)} \right).$$

For x large (and greater than T_i in particular), the term under the integral in the exponential can be approximated by $\frac{\theta_i(\beta-1)}{(\theta_i+1)x}$. Hence:

$$\mathbb{P}[L_{-i} \geq u] \sim e^{-\frac{\theta_i(\beta-1)}{\theta_i+1} \ln(u)} = u^{-\frac{\theta_i(\beta-1)}{\theta_i+1}}.$$

This is a fat-tailed distribution, and more specifically a power law. The probability that one would have to wait until an arbitrary large threshold u is reached before a first country occupies the territory is therefore:

$$\mathbb{P}[L_i \geq u \cap L_{-i} \geq u] \sim u^{-\frac{2\theta_i(\beta-1)}{\theta_i+1}}.$$

The rate of decay of this probability is slow, so that there are chances that war may in fact take place in a distant future only. The Laplace transform of the

war-trigger level $L_W = T_i \wedge T_{-i}$ (when a country first takes possession of the resource) hence verifies, for u large:

$$\mathbb{E}\left[e^{-\rho T_{L_W}} \,|L_W \geq u\right] \sim u^{\gamma_i} \int_u^{\infty} dv \gamma_i v^{-\beta-\gamma_i-1} = \frac{u^{-\beta}}{\beta+\gamma_i},$$

where we write $\gamma_i = \frac{2\theta_i(\beta-1)}{\theta_i+1}$. Note that $\mathbb{E}\left[e^{-\rho T_{L_W}} \,|L_W \geq u\right] \sim \frac{1}{\beta+\gamma_i}\mathbb{E}\left[e^{-\rho T_u}\right]$.

Real options, generally used for investment decision, may be exercised at optimal times that are known by the competitor, when in a competitive environment where preemption may be beneficial. Transposing this framework to military conflict, the consequent need for unpredictability changes the nature of the optimal mixed strategy. When both countries act optimally, the time until conflict, asymptotically, decays slowly. Therefore, due to the need for randomization, and in spite of both countries' propensity towards preemption, war may be indefinitely delayed.

References

1. Gauthier, L.: Excursions Height- and Length-Related Stopping Times, and App lication to Finance. Adv. Appl. Probab. **34**(4), 846–868 (2002)
2. Grenadier, S.R.: Option exercise games: the intersection of real options and game theory. J. Appl. Corp Financ. **13**(2), 99–107 (2000). https://doi.org/10.1111/j.1745-6622.2000.tb00057.x
3. Jeanblanc, M., et al.: Mathematical Methods for Financial Markets. Springer, London (2009). https://doi.org/10.1007/978-1-84628-737-4
4. Riedel, F., Steg, J.-H.: Subgame-perfect equilibria in stochastic timing games. J. Math. Econ. **72**, 36–50 (2017). https://doi.org/10.1016/j.jmateco.2017.06.006
5. Touzi, N., Vieille, N.: Continuous-time dynkin games with mixed strategies. SIAM J. Control Optim. **41**(4), 1073–1088 (2002). https://doi.org/10.1137/S0363012900369812
6. Triantis, A., Borison, A.: Real options: state of the practice. J. Appl. Corp Financ. **14**(2), 8–24 (2001). https://doi.org/10.1111/j.1745-6622.2001.tb00327.x
7. Trigeorgis, L.: Real Options: Managerial Flexibility and Strategy in Resource Allocation. MIT Press, Cambridge (1996)
8. Weeds, H.: Strategic delay in a real options model of R&D competition. Rev. Econ. Stud. **69**(3), 729–747 (2002). https://doi.org/10.1111/1467-937X.t01-1-00029

Variable Selection and Asymmetric Links to Predict Credit Card Fraud

Francesco Giordano, Michele La Rocca, Marcella Niglio(✉), and Marialuisa Restaino

Università degli Studi di Salerno, Via Giovanni Paolo II, 132, Fisciano, SA, Italy
{giordano,larocca,mniglio,mlrestaino}@unisa.it

Abstract. Credit card fraud identification is a challenging problem for different reasons: it needs to be suddenly detected; it is based on the use of huge data sets that have to be properly managed; the number of fraudulent transactions is definitely lower than the number of genuine transactions and then, this imbalance requires the use of proper statistical models. Here we discuss how the data reduction, performed through the variable selection, can be combined with the use of Generalized Linear Models with asymmetric link functions which are able to handle imbalanced data. We illustrate how these theoretical results can be used for credit card fraud-detection purposes.

Keywords: credit card fraud · imbalanced data · asymmetric link · variable selection

1 Introduction

Credit card fraud is a business of billions of dollars a year and represents a huge concern for banks. The main aim of banks is to identify a fraud as early as possible and then immediately initiate the procedures to protect both the bank and the customers.

The large amount of data on credit card transactions due to the increasing number of users and services that employ this form of payment makes it difficult not only to manage the data but mainly to identify fraudulent transactions. Luckily, among all transactions, the percentage of frauds is small but unfortunately, it does not make easy their detection and implies a high imbalance between genuine and fraudulent transactions. Further, the evolving habits of the cardholders might change the structure of the transaction data sets, so that the relevant variables having discriminating power against fraudulent transactions may rapidly change.

These themes of data imbalance and variable selection in evaluating credit card fraud have strong statistical consequences and, to the best of our knowledge, they are not jointly addressed in the literature.

M. Corazza et al. (Eds.): MAF 2024, *Mathematical and Statistical Methods for Actuarial Sciences and Finance*, pp. 198–204, 2024.
https://doi.org/10.1007/978-3-031-64273-9_33

Some contributions aiming to detect credit card fraud are based on machine learning approaches (see among the others [1,3,6,8]) and mainly focus the attention on the classification of the transactions or on the implementation of automatic mechanisms to prevent these fraudulent actions.

In the context of credit card transactions, where the database can be very huge and difficult to manage, our contribution aims to face the data reduction problem, through variable selection. When this context is combined with fraud detection another feature is added to the data: the imbalance of the response variable (characterized by a number of zeroes, related to genuine transactions, definitely higher than the number of ones, related to fraudulent transactions), and consequently the presence of a high skewness that needs to be properly accounted for.

Given this background and aiming to jointly treat the variable selection and the data imbalance, in Sect. 2 we first clarify how the imbalance of the response variable can be treated in the presence of binary data; in Sect. 3 we shortly describe the variable selection algorithm that can be applied in this field. Finally, Sect. 4 shows how these results can be applied in the context of credit card fraud detection.

2 Imbalanced Data

Consider a binary response variable Y with distribution belonging to the exponential family and with expectation $E[Y_i] = \mu_i$, for $i = 1, 2, \ldots, n$, where n is the sample size, and let $\mathbf{X} = (X_1, X_2, \ldots, X_p)$ be a vector of p covariates. Consider a monotone and differentiable function $g(\cdot)$:

$$g(\mu_i) = \mathbf{x}_i'\boldsymbol{\beta}, \tag{1}$$

where $\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2, \ldots, \beta_p)$ is the $[(p+1) \times 1]$ vector of parameters and $\boldsymbol{\beta} \in \mathbb{R}^{p+1}$, $\mathbf{x}_i = (1, x_{i1}, x_{i2}, \ldots, x_{ip})'$ is the vector of explanatory variables for unit i.

Equation (1) defines the generalized linear model (GLM) with *link function* function $g(\cdot)$ (see among the others [10]). In our context, Y is a binary response variable having Bernoulli distribution with probability $P(Y_i = y_i) = \pi_i$ if $y_i = 1$ and $1 - \pi_i$ if $y_i = 0$. Furthermore,

$$E[Y_i] = \pi_i = P(Y_i = 1) = F(\mathbf{x}_i'\boldsymbol{\beta}), \tag{2}$$

where $F(\cdot)$ is the cumulative distribution function that, using the notation (1), corresponds to:

$$F(\mathbf{x}_i'\boldsymbol{\beta}) = \pi_i = g^{-1}(\mathbf{x}_i'\boldsymbol{\beta}).$$

In the presence of a balanced, and then symmetric, response variable, the Logit and Probit link functions are the most commonly used. They have the advantage that the maximum likelihood equations are simple and the convergence of the estimators is quite fast. Unfortunately, the imbalance of Y produces

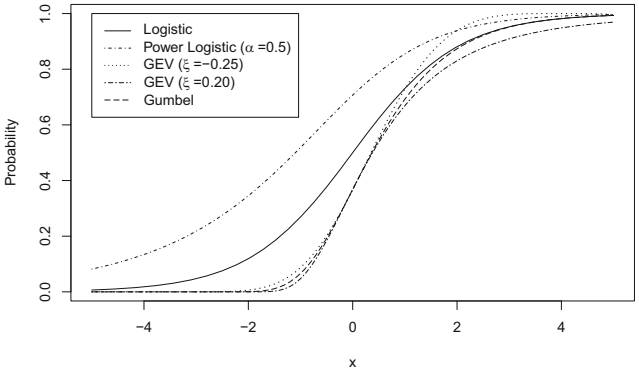


Fig. 1. Distribution function of Logit, GEV, Power Logit, and Gumbel variables.

biased estimates of $\boldsymbol{\beta}$, and therefore asymmetric link functions should be selected (among others see [4]).

For this reason, here, we consider some distributions that allow obtaining asymmetric link functions: the class of *Generalized Extreme Value* (GEV) distributions (as a variant of [12]) and the Power Logistic distribution (originally presented in [7] and largely investigated in [2]). The probability for the GEV distribution is

$$\text{GEV:} \qquad \pi_i = 1 - \exp\{-(1 - \xi \mathbf{x}_i' \boldsymbol{\beta})_+^{-\frac{1}{\xi}}\}, \tag{3}$$

where $(1 - \xi \mathbf{x}_i' \boldsymbol{\beta})_+ = \max\{(1 - \xi \mathbf{x}_i' \boldsymbol{\beta}), 0\}$ and $\xi \in \mathbb{R}$ is the shape parameter, whereas the *Power Logistic distribution* is:

$$\text{Power Logistic:} \qquad \pi_i = \frac{1}{(1 + \exp\{-(\mathbf{x}_i' \boldsymbol{\beta})\})^{\alpha}}, \tag{4}$$

where $\alpha > 0$ is the shape parameter that controls the asymmetry.

Note that both (3) and (4) include the Gumbel distribution. It is obtained when ξ goes to zero [5] and α goes to infinity [11] for GEV and Power Logistic respectively. This implies that the cLogLog link function [10] can be obtained as a particular case of the two previous distributions.

The main problem raised by imbalanced binary data is that their distribution function approaches 0 at a rate different from that approaching 1 and the link function based on the distribution functions (3) and (4) can deal with this problem. To give empirical evidence of this feature, consider the distribution functions in Fig. 1 where the standardized Logistic, Power Logistic (with $\alpha = 0.5$), GEV (with $\xi = \{-0.25, 0.20\}$) and Gumbel are plotted. Comparing the distributions it can be noted that the Power Logistic approaches zero smoothly with respect to all other distributions. The main differences among them can be seen in the tails: the Gumbel and the more general class of GEV distributions differ on the behaviour of the right tail; the Gumbel and Logistic differ on the left tail and in the GEV case (including the Gumbel distribution) the shape parameter ξ mainly affects the behaviour of the right tail.

The bias reduction induced by the use of asymmetric links can be of interest not only to increase the predictive performance of the GLM model (see [4] and [12] in the GEV case) but even to improve the data reduction through variable selection approaches based on the variable ranking, whose order is based on the estimated vector $\hat{\boldsymbol{\beta}}$ (see [9] for the GEV case). Our approach is briefly described in Sect. 3 and then applied to detect credit card fraud in Sect. 4.

3 Variable Selection

We now sketch the variable selection algorithm proposed in [9] that is here extended to a more general class of link functions. It is based on two main steps: the screening step (called GLM-SCR) and the variable selection step (called GLM-VS).

Let $\mathbf{Z}$ be the $(n \times p)$ standardized matrix of covariates $\mathbf{X}$, and let Y be the binary response variable.

GLM-SCR step

For each z_j, with $j = 1, \ldots, p$:

(a) estimate the corresponding parameter β_j maximizing the marginal likelihood $\sum_{i=1}^{n} \ell_j(\boldsymbol{\theta}_j, z_{ij}, y_i)$, where $\boldsymbol{\theta}_j$ includes β_0, β_j and the shape parameter of the link functions (ξ or α in our paper);

(b) rank the standardized covariates included in $\mathbf{Z}$, in non-increasing order with respect to $|\hat{\beta}_j|$ and select the first d covariates (where $d < p$ is a fixed tuning parameter) that will be included in the submodel denoted with $\widehat{\mathcal{M}}_d$.

GLM-VS step

After reducing the dimensionality of data, the variable selection is conducted by identifying the set K of covariates such that each subset I of covariates included in K, has probability $P(I \subseteq \widehat{\mathcal{M}}_d) \geq \pi_{thr}$, where π_{thr} is a fixed threshold, with $0 < \pi_{thr} < 1$. More precisely, the set K is estimated by:

$$\hat{K} = \underset{I \subseteq \widehat{\mathcal{M}}_d}{\arg\max} \{|I| : P(I \subseteq \widehat{\mathcal{M}}_d) \geq \pi_{thr}\}, \tag{5}$$

with $|I|$ the cardinality of I.

To estimate the probability $P(I \subseteq \widehat{\mathcal{M}}_d)$, [9] apply a subsampling procedure whose details are here omitted for brevity (see [9, Sect. 3.2]).

4 Credit Card Fraud Detection

The data set on which the analysis is conducted is a large database available on the Kaggle repository (https://www.kaggle.com/datasets/mlg-ulb/creditcardfraud) composed of 30 variables, 284807 transactions, 492 of which are labeled as fraudulent (0.172% of all transactions).

All variables have been anonymized by the data holder (and referred as "V1", ..., "V28") except for the "Amount" of each transaction and the "Class" (fraudulent = 1; genuine = 0). We have randomly extracted from the data set 20000 transactions guaranteeing that the sample has the same proportion of zeroes and ones in the "Class" variable, as in the full database. This choice has been made to reduce the computational cost of the variable selection which is quite high because of the subsampling used to estimate the probability in (5).

In the first stage of the analysis, we have selected the relevant variables using the algorithm in Sect. 3 and the four link functions: Logistic (LO), GEV, Gumbel (GU), and Power Logistic (PLO). In all cases, the tuning parameter d is fixed equal to 10 whereas the threshold value for the probability $\pi_{thr} = 0.10$. The selected variables are presented in Table 1.

Table 1. Variables selected using the four different link functions

Link	Selected variables
LO	**V11**, **V19**, **V4**, **V2**, V21, V27, V20, V22, V8
GEV	**V11**, **V4**, **V19**, **V2**, V21, V20, V25, V22
GU	**V4**, **V11**, **V19**, V26, V28, V27, **V2**, V25, V13
PLO	**V11**, **V4**, **V2**, **V19**, V21, V13, V15, V27

It can be noted that in all cases V2, V4, V11, and V19 are always selected whereas certain variability can be seen for the selection of the remaining variables. Furthermore, for all link functions V4, V11, and V19 have the highest probability, even if with different order of selection, and in no case the variable "Amount" is considered relevant (giving evidence that the transaction amount cannot be seen as a relevant wake-up call to detect frauds).

In the second stage of the analysis, we have evaluated the predictive ability of the four GLMs. To this aim, we have extracted 5000 further transactions as test set from the database using the same sampling procedure described before. Taking advantage of the variables selected in the previous step, we have estimated the parameters of the four regression models. Then, to evaluate the predictive accuracy of the models, we have computed the Accuracy rate (ACC), True Positive Rate (TPR), and True Negative Rate (TNR), given by:

$$ACC = \frac{TP + TN}{TP + TN + FP + FN}, \quad TPR = \frac{TP}{TP + FN}, \quad TNR = \frac{TN}{TN + FP},$$

with TP = True Positive, TN = True Negtive, FP = False Positive, FN = False Negative. In all cases, the cut-off has been fixed equal to the percentage of ones in the original data set, and the results are presented in Table 2. It clearly shows the superior performance of the GEV and Power Logistic models to predict the fraudulent and the genuine transactions showing the advantage of using asymmetric links when the data are imbalanced.

Table 2. Predictive measures and confusion matrices of the four models

Link	ACC	TPR	TNR			Predicted	
						0	1
LO	79.8%	100.0%	79.8%	Observed	0	3983	1008
					1	0	9
GEV	91.1%	88.9%	91.1%	Observed	0	4545	446
					1	1	8
GU	81.9%	100.0%	81.8%	Observed	0	4084	907
					1	0	9
PLO	99.7%	55.5%	99.7%	Observed	0	4978	13
					1	4	5

References

1. Afriyie, J.K., et al.: A supervised machine learning algorithm for detecting and predicting fraud in credit card transactions. Decis. Anal. J. **6**, 100163 (2023). https://doi.org/10.1016/j.dajour.2023.100163
2. Bazàn, J.L., Torres-Avilé, F., Suzuki, A.K., Louzada, F.: Power and reversal power links for binary regressions: an application for motor insurance policyholders. Appl. Stoch. Models Bus. Ind. **33**, 22–34 (2017). https://doi.org/10.1002/asmb.2215
3. Bhattacharyya, S., Jha, B., Tharakunnel, K., Westland, J.C.: Data mining for credit card fraud: Decis. Support Syst. **50**, 602–613 (2011). https://doi.org/10.1016/j.dss.2010.08.008
4. Calabrese, R., Osmetti, S.: Modelling SME loan defaults as rare events: an application to credit defaults. J. Appl. Stat. **40**, 1172–1188 (2013). https://doi.org/10.1080/02664763.2013.784894
5. Coles, S.: An Introduction to Statistical Modeling of Extreme Values. Springer, London (2001). https://doi.org/10.1007/978-1-4471-3675-0
6. Dal Pozzolo, A., Boracchi, G., Caelen, O., Alippi, C., Bontempi, G.: Credit card fraud detection: a realistic modeling and a novel learning strategy. IEEE Trans. Neural Netw. Learn. Syst. **29**, 3784–3797 (2018). https://doi.org/10.1109/TNNLS.2017.2736643
7. Devidas, M., George, E.O., Zelterman, D.: Generalized logistic models for low-dose response data. Stat. Med. **12**, 881–92 (1993). https://doi.org/10.1002/sim.4780120907
8. Forough, J., Momtazi, S.: Ensemble of deep sequential models for credit card fraud detection. Appl. Soft Comput. **99**, 106883 (2021). https://doi.org/10.1016/j.asoc.2020.106883
9. Giordano, F., Niglio, M., Restaino, M.: A new procedure for variable selection in presence of rare events. J. Oper. Res. Soc. **72**, 1619–1636 (2021). https://doi.org/10.1080/01605682.2020.1740620
10. McCullagh, P., Nelder, J.A.: Generalized Linear Models. Chapmann Hall, London (1989)

11. Shao, Q.: Maximum likelihood estimation for generalized logistic distributions. Commun. Stat. Theory Methods **31**, 1687–1700 (2002). https://doi.org/10.1081/STA-120014908
12. Wang, X., Dey, D.K.: Generalized extreme value regression for binary response data: an application to B2B electronic payments system adoption. Ann. Appl. Stat. **4**, 2000–2023 (2010). https://doi.org/10.1214/10-AOAS354

Partial Hedging of Spread Options with a Given Probability

Betty Guo(✉) and Alexander Melnikov

University of Alberta, Edmonton, AB T6G 2R3, Canada
{betty2,melnikov}@ualberta.ca

Abstract. The paper develops the method of hedging a two-factor diffusion market with a given probability. We construct the maximal perfect hedging set for an option to exchange for another for which a lower bound of option price is achieved. This method is then applied to pricing "pure endowments with a guarantee" equity linked life insurance contracts. The classical approach of pricing such options provides very low gain for investors, the investor may take a given probability of risk in return of a higher gain. Taken into account this argument, the paper develops risk management strategies for this type of insurance and financial mixed instrument.

Keywords: Partial Hedging · Life Insurance · Spread Options

1 Partial Hedging Problem

1.1 Financial Setting

We consider an arbitrage-free two factor diffusion model in a complete market defined on the stochastic basis $(\Omega, \mathcal{F}, \mathcal{F}_t, P), t \geq 0, P \in \mathcal{P}$ and are adapted to the filtration $\mathcal{F}$, generated by two dependent Wiener processes W_t^1 and W_t^2. The capital at time t of an investor is defined by the relation

$$X_t^\pi = \beta_t B_t + \gamma_t^1 S_t^1 + \gamma_t^2 S_t^2 \tag{1}$$

where B_t is the price of a risk-free asset with $B_0 = 1$ and $S_t^i, i = 1, 2$, are risky assets. For simplicity, in this paper we will assume $B_t = 1$, in other words the interest rate is equal to 0 ($r_t = 0$). Furthermore, $\pi = (\pi_t)_{t\geq 0} = (\beta_t, \gamma_t^1, \gamma_t^2)_{t\geq 0}$ is a predictable trading strategy (hedge). By the definition of arbitrage-free complete market, there exists a unique local martingale measure P^* such that the processes $(S_t^i)_{t\geq 0}$ are local martingales with respect to measure P^*. Equivalently, the process $(X_t^\pi)_{t\geq 0}$ is a local martingale with respect to measure P^* for any self-financing hedges. We call the trading strategy $\pi \in SF$ (self-financing), if the capital X_t^π is realized without any inflow or outflow of cash or assets. This is represented by

$$X_t^\pi = X_0^\pi + \int_0^t \gamma_u^1 dS_u^1 + \int_0^t \gamma_u^2 dS_u^2 \tag{2}$$

M. Corazza et al. (Eds.): MAF 2024, *Mathematical and Statistical Methods for Actuarial Sciences and Finance*, pp. 205–210, 2024.
https://doi.org/10.1007/978-3-031-64273-9_34

Let us fix T as the maturity date, and let H represent contingent claims. The classical approach to pricing an European style option requires choosing a hedge $\pi \in SF$ such that

$$P(X_T^\pi \geq H) = 1 \tag{3}$$

The fair price of such option is defined as the minimal initial capital (X_0^π) which is sufficient for (3) to hold.

Definition 1. *Denote $\Pi(x, H)$ as a collection of self-financing hedges where $X_0^\pi = x$ and Eq. (3) holds. Then the fair price of the contingent claim is equal to*

$$\mathbb{C}(H) = \inf\{x : \Pi(x, H) \neq \emptyset\} \tag{4}$$

There exists a hedge $\pi^* \in SF$ such that $\mathbb{C}(H) = X_0^{\pi^*} = E^*[H]$, π^* is called a perfect hedge.

The classical approach of pricing options leads to a very low gain for investors; for European style options this gain is equal to zero by construction of the perfect hedge. Taking this thought into account, we can relax condition (3), so that investors can hedge the contingent claim bearing some risk. Mathematically, this can be written as $P(X_T^\pi \geq H) = 1 - \alpha$, where H is the pay-off at maturity of the contingent claim, and α is a given significance level (Risk level). This approach was considered on general models of a market with one risk-free and one risky asset for pricing and developing minimal hedge of American options by A. Novikov[6]. Our paper will follow a similar approach as the paper by A. Novikov with application to life insurance contracts.

1.2 Construction of Hedge for General Models on European Options

According to the second fundamental theorem of asset pricing, in a complete market there always exists a unique martingale probability measure P^*. Here we introduce and define a new class of hedges $SF(A)$, which is a set of self-financing hedges in relation to a chosen constant A.

$$SF(A) = \{\pi \in SF : X_T^\pi \geq H - A \ a.s.\} \tag{5}$$

where A is a given non-negative constant, and H is the pay-off of the contingent claim. As mentioned in the previous section, to hedge the pay-off perfectly, the value of portfolio at time of maturity X_T^π must be greater than or equal to the pay-off H. In class $SF(A)$, this condition is relaxed by introducing non-negative constant A. The set of possible self-financing hedges is expanded by decreasing the pay-off from H to $H - A$. For this paper we will consider hedges only from the class $SF(A)$. Based on the class of hedges $SF(A)$ we can derive a lower bound of initial investor capital. The lower bound is illustrated in the following lemma.

Lemma 1. *If $\pi \in SF(A)$, then $E^*\left[H - AI\{X_T^\pi < H\}\right] \leq X_0^\pi$.*

We then restrict the set of hedges under consideration further, such that the martingale probability of perfect hedge is equal to $1-\alpha$. The following subset of hedges is created from the class $SF(A)$, with respect to initial capital x, pay-off H, significance level α, and non-negative constant A.

$$\Pi(x, H, \alpha, A) = \{\pi \in (A) : P^*(X_T^\pi \geq H) = 1 - \alpha\} \tag{6}$$

The fair price of an European style option with relation to the new set of hedges is defined as $\mathbb{C}(H, \alpha, A) = \inf\{x : \Pi(x, H, \alpha, A) \neq \emptyset\}$. Under general assumptions, observe that there exists a hedge $\pi^\alpha \in \Pi(x, H, \alpha, A)$ with the initial capital

$$x = X_0^{\pi^\alpha} = E^*[X_T^\pi - AI\{X_T^\pi < H\}] = \mathbb{C}(H) - \alpha A. \tag{7}$$

Based on the above equality we can make the following observation. Lemma 1 indicates $E^*[X_T^\pi - AI\{X_T^\pi < H\}]$ is the lower bound for initial capital x. On the other hand, the fair price is the infimum of all possible initial capitals. Therefore, by Lemma 1 for European options the fair price is equal to $\mathbb{C}(H, \alpha, A) = \mathbb{C}(H) - \alpha A$.

The hedge π^α defined above is constructed using statistical techniques of hypothesis testing, in particular, the Neyman-Pearson Lemma. The given significance level controls the probability of type I error, denoted α in this paper. Suppose there exists an event $E \in \mathcal{F}$ such that $P^*(E) = 1 - \alpha$. In mathematical terms, the likelihood ratio under the two hypotheses are compared. Denote $L_t = L_t(Q) = \frac{dQ}{dP^*}\Big|_{\mathcal{F}_t}$ a Radon-Nikodym derivative, which is considered as the likelihood ratio, of some measure $Q \in \mathcal{P}, Q \neq P^*$. Consider the event to be the perfect hedging set $E = \{L_T \geq \lambda(\alpha)\}$. Consider a test with hypotheses $H_0 : P^*$ and $H_1 : Q$, and let the rejection region be $R = \{L_T < \lambda(\alpha)\}$. For the event E to exists it must be outside of the rejection region. In this case, the implied null hypothesis is that a perfect hedge cannot be achieved, and the alternative hypothesis is that it can hedge pay-off H.

Let $\pi^* = (\gamma_t^{1^*}, \gamma_t^{2^*})$ be the minimal hedge or risk-neutral hedge in the classical pricing problem (when $\alpha = 0$). Let $Y_t = E^*[H]$, then there exists a predictable process $A_t \geq 0, A_0 = 0$ such that $Y_t = M_t - A_t$ where M_t is a martingale with respect to P^* and $Y_t = M_t = H$ almost surely. Rewrite M_t in the form of martingale representation and by assumption of completeness of market, there exists a predictable process γ_t^* such that

$$\begin{aligned} M_t &= M_0 + \int_0^t \gamma_t^{1^*} dS_t^1 + \int_0^t \gamma_t^{1^*} dS_t^2 = Y_0 + \int_0^t \gamma_t^{1^*} dS_t^1 + \int_0^t \gamma_t^{1^*} dS_t^2 \\ &= C(H) + \int_0^t \gamma_t^{1^*} dS_t^1 + \int_0^t \gamma_t^{1^*} dS_t^2 \end{aligned} \tag{8}$$

With all necessary construction complete, we arrive to the final result theorem.

Theorem 1. *Let hedge $\pi^\alpha = (\gamma_t^{1^\alpha}, \gamma_t^{2^\alpha})$ be defined by the following equations*

$$X_t^{\pi^\alpha} = X_0^{\pi^\alpha} + \int_0^t \gamma_t^{1^\alpha} dS_t^1 + \int_0^t \gamma_t^{1^\alpha} dS_t^2 \tag{9}$$

$$X_0^{\pi^\alpha} = C(H) - \alpha A, \quad \gamma_t^{i^\alpha} = \gamma_t^{i^*} - \varphi_t A,\ i = 1, 2 \tag{10}$$

$$m_t = P^*(L_T < \lambda(\alpha)|\mathcal{F}) = \alpha + \int_0^t \varphi_t^1 dS_t^1 + \int_0^t \varphi_t^2 dS_t^2 \quad \textit{(Martingale representation)} \tag{11}$$

Then

$$\alpha = P^*(X_T^{\pi^\alpha} < H), \quad \pi^\alpha \in \Pi(X_0^{\pi^\alpha}, H, \alpha, A), \tag{12}$$

$$\textit{and} \quad C(H, \alpha, A) = X_0^{\pi^\alpha} = C(H) - \alpha A \tag{13}$$

1.3 Extend to Two-Factor Diffusion Model

An exchange option is a derivative security that gives the owner the right but not the obligation to forfeit one unit of a certain risky asset and obtain one unit of the underlying asset in return. The pay-off option of such contingent claim is given as follows.

$$\max\{S_T^1, S_T^2\} \xrightarrow{\text{(it can be reduced to)}} \max\{S_T^1 - S_T^2, 0\} \tag{14}$$

where S_T^1 and S_T^2 are two correlated risky assets following the geometric Brownian motion.

$$dS_t^i = S_t^i(\mu_i dt + \sigma_i dW_t^i) \quad i = 1, 2, \tag{15}$$

W_t^1 and W_t^2 are correlated Wiener processes with correlation $cov(W_t^1, W_t^2) = \rho t$. The martingale probability density of such model is given by

$$Z_T^* = \exp\left(\varphi_1 W_T^1 + \varphi_2 W_T^2 - \frac{1}{2}\sigma_\varphi^2 T\right) \tag{16}$$

where

$$\varphi_1 = \frac{r(\sigma_2 - \sigma_1\rho) + \rho\mu_2\sigma_1 - \mu_1\sigma_2}{\sigma_1\sigma_2(1-\rho^2)}; \quad \varphi_2 = \frac{r(\sigma_1 - \sigma_2\rho) + \rho\mu_1\sigma_2 - \mu_2\sigma_1}{\sigma_1\sigma_2(1-\rho^2)} \tag{17}$$

$$\text{and } \sigma_\varphi^2 = \varphi_1^2 + \varphi_2^2 + 2\rho\varphi_1\varphi_2. \tag{18}$$

Following the steps of the Novikov paper, we can use the Neyman-Pearson Lemma to construct the maximal hedging set. We let $\lambda(\alpha)$ (variable used in construction) be an equation similar to the likelihood ratio but containing a variable (θ_α) dependent on the significance level α.

$$\lambda(\alpha) = \exp\left(\varphi_1\theta_\alpha + \varphi_2\theta_\alpha - \frac{1}{2}\sigma_\varphi^2 T\right) \tag{19}$$

where T is the maturity date. Then we set the likelihood ratio to be greater than and equal to $\lambda(\alpha)$.

$$\{L_T(P) \geq \lambda(\alpha)\} = \left\{\exp\left(\varphi_1 W_T^1 + \varphi_2 W_T^2 - \frac{1}{2}\sigma_\varphi^2 T\right) \geq \exp\left(\varphi_1\theta_\alpha + \varphi_2\theta_\alpha - \frac{1}{2}\sigma_\varphi^2 T\right)\right\}$$
$$= \{\varphi_1 W_T^1 + \varphi_2 W_T^2 - \frac{1}{2}\sigma_\varphi^2 T \geq \varphi_1\theta_\alpha + \varphi_2\theta_\alpha - \frac{1}{2}\sigma_\varphi^2 T\}$$
$$= \{\varphi_1 W_T^1 + \varphi_2 W_T^2 \geq (\varphi_1 + \varphi_2)\theta_\alpha\} \tag{20}$$

The two Wiener processes are normally distributed with the distribution $W_T^i \sim N(0,T)$ for $i = 1, 2$ and $cov(W_T^1, W_T^2) = \rho T$. Suppose we have stochastic differential equation of the form $dX_t = \beta dt + \sigma_1 dW_t^1 + \sigma_2 dW_t^2$. By Levy's characterization of Brownian motion we can come up with some σW_t, where σ is a fixed constant and W_t is another Wiener process. This is illustrated by the stochastic differential equation $dX_t = \beta dt + \sqrt{\sigma_1^2 + 2\sigma_1\sigma_2\rho + \sigma_2^2} dW_t$. In particular, we have,

$$\varphi_1 W_T^1 + \varphi_2 W_T^2 \sim N(0, \sigma^2) \quad \text{where} \quad \sigma^2 = T(\varphi_1^2 + \varphi_2^2 + 2\rho\varphi_1\varphi_2) \tag{21}$$

Let $W_T^* = \varphi_1 W_T^1 + \varphi_2 W_T^2$. We can determine θ_α with the following condition.

$$P(W_T^* \geq (\varphi_1 + \varphi_2)\theta_\alpha) = 1 - \alpha \tag{22}$$

$$P(L_T \geq \lambda(\alpha)) = P\left(\frac{W_T^*}{\sigma} \geq \frac{(\varphi_1 + \varphi_2)\theta_\alpha}{\sigma}\right) = 1 - \alpha \tag{23}$$

Therefore,

$$\frac{(\varphi_1 + \varphi_2)\theta_\alpha}{\sigma} = Z_\alpha \quad \Rightarrow \quad \theta_\alpha = \frac{\sigma Z_\alpha}{\varphi_1 + \varphi_2} \quad \text{where} \quad Z_{1-\alpha} = \Phi^{-1}(1-\alpha) \tag{24}$$

By the previously mentioned theorem, $\mathbb{C}(H, \alpha, A) = \mathbb{C}(H) - \alpha A$, where $\mathbb{C}(H)$ is calculated by Margrabe's formula, $\mathbb{C}(H) = S_0^1 N(d_1) - S_0^2 N(d_2)$.

$$\text{where} \quad d_1 = \ln\left(\frac{S_0^1}{S_0^2}\right) + \frac{\sigma^2}{2}T; \quad d_2 = d_1 - \sigma\sqrt{T}; \quad \text{and} \quad \sigma = \sqrt{\sigma_1^2 + \sigma_2^2 - 2\rho\sigma_1\sigma_2} \tag{25}$$

We have,

$$P(L_T(P) < \lambda(\alpha)|\mathcal{F}_t) = P(W_T^* - W_t^* < (\varphi_1 + \varphi_2)\theta_\alpha - W_t^*)$$
$$= \Phi\left(\frac{(\varphi_1 + \varphi_2)\theta_\alpha - W_t^*}{\sigma_\psi}\right). \tag{26}$$

Applying the two-dimensional Ito's formula, we calculate the correction term (φ_t^i) for the hedge $\gamma_t^{i^*}$ as follows.

$$\varphi_t^i = -\frac{B_t}{S_t^i \sigma_i} \times \frac{1}{\sigma_\psi\sqrt{2\pi}} \exp\left(-\frac{[(\varphi_1^2 + \varphi_2^2)\theta_\alpha - W_t^*]^2}{2\sigma_\psi^2}\right) \tag{27}$$

2 Application to Life Insurance

The insurance contract under discussion is an equity-linked pure endowment contract. In such contracts, the insureds receive the linked equity if they can survive a certain period to maturity. In this paper, we assume there are one risk-free asset and two risky assets. We will consider a diffusion model with two correlated Wiener process in which the risky asset prices are generated from. Let $T(x)$ denote the remaining life time of an individual with current age x at the start of the contract. Then, at maturity, the pay-off of an equity linked pure endowment insurance contract is

$$\max\{S_T^1 - S_T^2, 0\}\mathbb{I}_{\{T(x)>T\}} \tag{28}$$

where T is the maturity time of the contract. Suppose $T(x)$ is defined on the stochastic basis $(\tilde{\Omega}, \tilde{\mathcal{F}}, \tilde{\mathcal{P}})$, the two probability measures P and $\tilde{P}$ are independent. The risk neutral premium of such contract is

$$\mathbb{P} = E^* \times \tilde{E}[\max\{S_T^1 - S_T^2, 0\}\mathbb{I}_{\{T(x)>T\}}] = E^*[H]_T P_x \tag{29}$$

The fair premium for such insurance contract will be less than the price for the exchange option given by Theorem 1. Following the approach of Brennan and Schwartz [1] we have the following equality

$$\mathbb{P} = E^*[H]_T P_x = E^*[H - A\mathbb{I}_{\{X_T^{\pi^\alpha} < H\}}] \tag{30}$$

$$_T p_x = \frac{E^*[H - A\mathbb{I}_{\{X_T^{\pi^\alpha} < H\}}]}{E^*[H]} = \frac{\mathbb{C}(H) - \alpha A}{\mathbb{C}(H)} \tag{31}$$

Remarks. This paper illustrates hedging with a given probability based on a Two-factor Diffusion model with correlated Brownian Motions. This concept can be expanded and explored in terms of Two-factor Jump Diffusion model. Aside from different market models, it can also be applied to efficient hedging with power loss function. Instead of having a given probability, we can hedge with a given expected shortfall. Preliminary results of the two mentioned expansions are achieved.

References

1. Brennan, M.J., Schwartz, E.S.: The pricing of equity-linked life insurance policies with an asset value guarantee. J. Financ. Econ. **3**(3), 195–213 (1976)
2. Follmer, H., Leukert, P.: Quantile hedging. Financ. Stoch. **3**(3), 251–273 (1999)
3. Krutchenko, R.N., Melnikov, A.V.: Quantile hedging for a jump-diffusion finanaicl market model. In: Kohlmann, M., Tang, S. (eds.) Mathematical Finance, pp. 215–229. Birkhäuser, Basel (2001). https://doi.org/10.1007/978-3-0348-8291-0_20
4. Melnikov, A.V.: Efficient hedging of equity-linked life insurance policies. Dokl. Math. **69**(3), 462–464 (2004)
5. Mel'nikov, A.V., Shiryaev, A.N.: Criteria for the absence of arbitrage in the financial market. Front. Pure Appl. Probab. **II**, 121–134 (1996)
6. Novikov, A.: Hedging of options with a given probability. Theory Probab. Appl. **43**(1), 135–143 (1999)

Four Parameter Beta Generalized Mixed Effect Tree and Random Forest for Area Yield Crop Insurance

Dian Kusumaningrum[1,2(✉)], Hari Wijayanto[2(✉)], Anang Kurnia[2], Khairil Anwar Notodiputro[2], and Muhlis Ardiansyah[2,3]

[1] Business Mathematics Program Study, Prasetiya Mulya University, South Jakarta, Indonesia
dian.kusumaningrum@prasetiyamulya.ac.id
[2] Program on Statistics and Data Science - School of Data Science, Mathematics, and Informatics, IPB University, Dramaga, Indonesia
hari@apps.ipb.ac.id
[3] BPS-Statistics of Kotawaringin Timur, South Jakarta, Central Kalimantan, Indonesia

Abstract. Area Yield Index (AYI) was considered as the most potential alternative crop insurance policy in Indonesia. To support this policy, delivering accurate paddy productivity prediction is a must. Thus, we purpose a new flexible Be-ta Four Parameter Generalized Mixed Effect Tree and Random Forest predic-tion model that combines the use of tree regression and random forest with a Bayesian beta four parameter GLMM approach. This model takes into con-sideration that paddy productivity has a bounded minimum and maximum distribution or known as a Beta Four Parameter distribution, variation effect of paddy productivity between areas, and captures complex linear and non-linear relationships in the data. This model was incorporated to design a prototype AYI crop insurance in Central Kalimantan, Indonesia that can be fur-ther developed in other areas. Farmer survey data integrated with processed satellite data was utilized in the process. Results show that high predictive accuracy was achieved in the proposed model. Therefore, beneficial for accu-rately assessing risk, setting fair premiums, reducing adverse selection, effi-ciently allocating resources, and ensuring the long-term sustainability of the paddy crop insurance program.

Keywords: Area Yield Index (AYI) · Beta Four Parameter Distribution (B4P) · General-ized Linear Mixed Model (GLMM) · Beta Four Parameter Generalized Mixed Effect Tree (B4P GMET) · Beta Four Parameter Generalized Mixed Effect Random Forest (B4P GMERF)

1 Introduction

Area Yield Index (AYI) crop insurance policies is a yield or productivity-based in-dex agricultural insurance policy. National Development Planning Agency

M. Corazza et al. (Eds.): MAF 2024, *Mathematical and Statistical Methods for Actuarial Sciences and Finance*, pp. 211–217, 2024.
https://doi.org/10.1007/978-3-031-64273-9_35

(BAPPENAS), MoA, Japan International Cooperation Agency (JICA), and other related stakeholders conducted an AYI pilot study in Karawang and Kendal districts [1]. Reports show that AYI exhibits many positive aspects such as minimal moral hazard and adverse selection. Furthermore, since the indemnity of AYI is determined by average sub district, insurers do not need to conduct loss adjustment surveys. Therefore, AYI has lower administrative costs [1].

However, there are also obstacles needed to be addressed. One of the main issues is lack of reliable historical data required to calculate the benchmark yield (y_c) at an area level (province or district). Next, estimates of monthly or seasonal paddy yields ($\bar{y_i}$) are currently insufficient, particularly at a sub district or village level. As a response, Statistics Indonesia (SI) has developed a monthly paddy productivity estimate based on the Crop Cutting Experiments (CCEs). Nonetheless, the guidelines and data quality need to be evaluated for further applications of AYI [2].

Compensations paid by the insurer to all farmers in the sub area for AYI are based on indemnity formula With this chapter, the preliminaries are over, and we begin the search for periodic solutions to Hamiltonian systems. All this will be done in the convex case; that is, we shall study the boundary-value problem

$$\boldsymbol{Indm} = \max(y_c - \bar{y}_l, 0)\boldsymbol{SI}L_{ij} \tag{1}$$

where L_{ij} is the amount of land cultivated by a farmer j in a sub area i. For simplicity we can set L_{ij} at 1 hectare (Ha). While $\boldsymbol{SI}$ is the sum insured fixed at 6,00,000 IDR per Ha, which is the minimum amount needed to be able to continue farming in the next season if risks occur [3]. Payouts will be made when $\bar{y_l}$ is less than the y_c value.

To calculate indemnity (1), it is apparent that predicting y_c is vital. BAPPENAS has used the current and average historical CCEs data in the pilot study [4]. Meanwhile, other information can also be utilized to enhance predictions accuracy, such as the farmer surveys and satellite data.

Aside from using average historical data for predicting y_c, applying Exponential Smoothing and ARIMA [5] are also popular. Machine Learning and Deep Learning [6] approaches have also been applied. Even though promising, there are several concerns that need to be accounted (1) heterogeneity of paddy productivity among areas, (2) productivity data distributions that are bounded to a certain minimum and maximum value also known as the B4P distribution [7] and (3) occurrence of linear and nonlinear relationships between the response and explanatory variables.

Hence, we propose a new approach and developed a B4P GMET and B4P GMERF prediction models. For this reason, the main objective of this paper is to evaluate the performance of these models and apply it to develop AYI. Our proposed models are described in the next section. Section 3 presents the findings of this study, and Sect. 4 concludes the paper.

2 Developing B4P-GMET and B4P-GMET for Area Yield Index Policy

At first, we have developed a B4P GLMM by using a Bayesian approach [8] that extends the B4P regression model [9]. In this model we assume that $y_c \sim B4P(\alpha, \beta, a, b)$. Where, a is the minimum value, b is the maximum value, α is the location parameter and $\alpha - \beta$ is the scale parameter. Thus, the model can be formulated as

$$\boldsymbol{\eta_i} = \boldsymbol{X_i^T}\boldsymbol{\beta} + \boldsymbol{Z_i^T}\boldsymbol{b_i} \tag{2}$$

The random effect matrix design is denoted by $\boldsymbol{Z_i^T}$ and $\boldsymbol{b_i}$ is the random effect estimators in a particular area or sub areas, where $\boldsymbol{b_i} \sim iidN(0, \sigma_i^2)$. While the fixed effect matrix is shown by $\boldsymbol{X_i^T}$ and the fixed effect estimators of the independent variables used are denoted as $\boldsymbol{\beta}$. When applying a logit link function for the model, then $\boldsymbol{\eta_i} = g\frac{(\text{Mean}[y_c|x_i]-a)}{(b-a)} = g(\mu_i) = log(\frac{\mu_i}{1-\mu_i})$.

To accommodate nonlinear relationships GMET has been developed based on a regression tree and GLMM [10]. The model first assumed a Bernoulli distribution. further we develop a B4P GMET that can be formulated as:

$$\boldsymbol{\eta_i} = f(X_i) + \boldsymbol{Z_i^T}\boldsymbol{b_i} \text{ and } \boldsymbol{b_i} \sim iidN(0, \boldsymbol{\sigma_i^2}) \tag{3}$$

The fixed effect $f(X_i)$ is estimated through CART algorithm and $\boldsymbol{Z_i^T}\boldsymbol{b_i}$ is estimated through the B4P-GLMM. A detail on the algorithms of estimating parameters for GMET can be seen in [10]. Next, GMERF for a Bernoulli distribution was introduced and applied the random forest to estimate the fixed effects parameters for model in Eq. (2) [11]. Such as for GMET, we also continue to developed GMERF for B4P distribution response variables.

The proposed models will be applied to an empirical case study in Central Kali-mantan's first planting season (2020). Paddy productivity was measured by CCE's and there were seven selected explanatory variables from the farmer survey. Current and a four-month period lag of Sentinel 2A satellite data for each plot will also be used (Table 1).

Model selection will be based on Root Mean Square Error (RMSE) and Watanabe-Akaike information criterion (WAIC) values. Through a Bootstrap simulation, the predicted values of the best model will then be accounted to calculate the pure pre-mium and VaR of AYI. The process is summarized as follows:

1. Predict paddy productivity (y_{ij}) for individual farmer's based on the best fit model
2. Resample with replacement 100 bootstrap farmer samples (n) for each sub district
3. Calculate average yield for each sub district ($\bar{y_l}$) with $\bar{y_l} = \frac{1}{n}\sum_{1=1}^{n} y_{ij}$
4. Calculate benchmark yield y_c using average of an area, with $y_c = \frac{1}{N}\sum_{a=1}^{N} \bar{y_l}$. N is the total sub areas with in an area
5. Calculate claim amounts based on formula (1)

Table 1. Predictor Variables used in the Prediction Model

Variable	Variable Name	Unit Measurement/Category
Paddy Productivity	Y	Tons/Ha
Pest Attacks This Year	X_1	1 = Heavy, 2 = Medium, 3 = Light, 4 = Not Affected
Pest Attacks Last Year	X_2	1 = Heavy, 2 = Medium, 3 = Light, 4 = Not Affected
Impact of Climate Change This Year	X_3	1 = Affected, 2 = Not Affected
Impact of Climate Change Last Year	X_4	1 = Affected, 2 = Not Affected
Water Sufficiency This Year	X_5	1 = Not Enough, 2 = Sufficient, 3 = More than Enough
Water Sufficiency Last Year	X_6	1 = Not Enough, 2 = Sufficient, 3 = More than Enough
How to Handle Pest	X_7	0 = No Actions, 1 = Agronomist, 2 = Mechanical, 3 = Biological, 4 = Chemical
Band 4 − Infrared	X_8	Mm
Band 8 − Near Infrared	X_9	Mm
NDVI -Normalized Difference Vegetation Index	X_{10}	Index
KCL	X_{12}	Kg/Ha
Solid Organic Fertilizer	X_{13}	Kg/Ha
Liquid Organic Fertilizer	X_{14}	Kg/Ha

6. Repeat steps 2-6 1.000.000 times and estimate the expected pure premium and VaR
7. Compare the result from simulation with actual conditions

3 Result and Discussion

Empirical data in Central Kalimantan (2020) shows that paddy productivity has a B4P distribution. Estimates of this distribution indicate that the average paddy productivity is estimated to be 3.03 and the variance is 1.070 tons per hectare. The distribution is also skewed to the right. WAIC show that a B4P GLMM is more suitable compared to the B4P GLM. Hence, random effects of a sub district improve model fit and leads to a better un-derstanding of variability of productivity among areas. Prediction wise, Table 2 proofs that the proposed B4P GMERF and B4P GMET has a much better prediction accuracy compared to the other models. Especially, when applying farmer survey and satellite data.

Table 2. Model Evaluation for Developed Models

Model	Variables	WAIC	RMSE
B4P GLM	Farmer Survey Data	–115.000	0.010
	Satellite Data	–91.615	0.011
	Farmer Survey and Satellite Data	–124.000	0.011
	Farmer Survey Data	–202.609	0.029
B4P GLMM	Satellite Data	–187.735	0.028
	Farmer Survey and Satellite Data	–206.745	0.020
	Farmer Survey Data	–	0.003
B4P GMET	Satellite Data	–	0.004
	Farmer Survey and Satellite Data	–	0.026
B4P GMERF	Farmer Survey Data	–	0.003
	Satellite Data	–	0.008
	Faremer Survey and Satellite Data	–	0.010

With regards to AYI, positive results were shown when utilizing satellite data for predicting paddy productivity. Variable importance pointed out that current and lagged infrared and near infrared reflectance along with current NDVI relationship with paddy productivity are apparent. Thus, it can serve as a remedy for current concerns among stakeholders regarding the insufficient availability of accurate and real-time data at both the area and sub-area levels. Other factors that are also considered important from the farmer survey are the severity of pest attacks of the current year and the amount of KCL and Solid Organic Fertilizer used. Lower use of KCL and Solid Organic Fertilizer tend to lead to lower productivity.

Table 3. AYI Premium and VaR based on B4P GMERF Estimates (in Rupiah)

Area	Y_c	Mean of B4P	Pure Premium	$\text{VaR}_{95\%}$
Province	3.03	1.70	640,872.70	1,011,718.10
District 1	2.70	1.02	166,207.50	246,789.23
District 2	3.27	1.80	467,944.70	1,022,576.50

Province and districts were used in the simulation to define a potential area level. If AYI was set at province level, the policy will have a higher pure premium and tail risks but easier to administer. In a region where heterogeneity is apparent, setting a premium at district level is more sufficient. Therefore, preventing high basis risk of farmers and insurers. Nonetheless, administration wise it will be more challenging when in a region there are many districts. Table 3 also shows

the estimated mean of a B4P distribution is much lower than the average value (y_c). Keeping in mind that paddy productivity has a B4P distribution, careful attention should be done in defining y_c. Setting higher average benchmark such as in this study may lead to overestimating the expected losses and cause higher premium for farmers, which may lead to decreased participation in the insurance program itself.

4 Conclusion

This paper introduces B4P GMET and B4P GMERF prediction model applied to pre-dict paddy productivity. This model enhanced the existing models by considering (1) heterogeneity (2) appropriate B4P distributions, and (3) linear and nonlinear condi-tions. This proposed method has shown promising results compared to B4P GLM and B4P GLMM. By calibrating the models to empirical data, extensive Bootstrap studies were performed to estimate the pure premium and VaR of AYI. Here, we concluded that designing AYI at district level is more appropriate when productivity among areas vary. Considerations must also be given in defining the benchmark productivity when there is proof that the distribution of paddy productivity follows a B4P distribu-tion. Last, the use of satellite data in the model has proven a beneficiary, Thus, we encourage further studies in other areas and continuously improve the model.

Acknowledgment. The authors would like to acknowledge research funding support provided by the South East Southeast Asian Regional Center for Graduate Study and Research in Agriculture (SEARCA). We are also very appreciative to Statistics Indonesia (SI) for sharing the empirical data and Prasetiya Mulya University with IPB University for the support and assistance during this study.

References

1. Sanyu Consultants Inc. and Sompo Risk Management Inc.: Area Yield Index Insurance Product Design. National Development Planning Agency (BAPPENAS) and Japan International Cooperation Agency (JICA) (2023)
2. Shynkarenko, I., Shynkarenko, R., Krychevska, L., McConnel, R.: Survey on Sustainable Agricultural Insurance Scheme in Indonesia. National Development Planning Agency (BAPPENAS) and Japan International Cooperation Agency (JICA) (2019)
3. Kusumaningrum, D., Anisa, R., Sutomo, V.A., Tan, K.S.: Alternative area yield index based crop insurance policies in Indonesia. In: Corazza, M., Gilli, M., Perna, C., Pizzi, C., Sibillo, M. (eds.) Mathematical and Statistical Methods for Actuarial Sciences and Finance, LNCS, pp. 285–290. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-78965-7_42
4. Ardiansyah, M., Kurnia, A., Sadik, K., Djuraidah, A., Wijayanto, H.: Numerical prediction of paddy weight of crop cutting survey using generalized Geoadditive linear mixed model. J. Phys. Conf. Ser. **1863**, 1–17 (2021)
5. Skees, J.R., Black, J.R., Barnett, B.J.: Designing and rating an area yield crop insurance contract. Am. J. Agric. Econ. **79**(2), 430–438 (1997)

6. Sun, J., Lai, Z., Di, L., Sun, Z., Tao, J., Shen, Y.: Multilevel deep learning network for county-level corn yield estimation in the U.S. corn belt. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. **13**, 5048–5060 (2020)
7. Hennessy, D.A.: Crop yield skewness and the normal distribution. J. Agric. Resour. Econ. **34**(1), 34–52 (2009)
8. Kusumaningrum, D., Wijayanto, H., Kurnia, A., Notodiputro, K.A., Ardiansyah, M.: Bayesian Estimation of Beta Four Parameter GLMM with Application of Estimating Area Yield Index Crop Insurance Risks and Premium [Unpublished manuscript]
9. Zhou, H., Huang, X.: Bayesian beta regression for bounded responses with unknown supports. J. Comput. Stat. Data Anal. **167** (2022)
10. Fontana, L., Masci, C., Ieva, F., Paganoni, A.M.: Performing learning analytics via generalised mixed-effects trees. MDPI Data J. **6**(74) (2021)
11. Pellagatti, M., Masci, C., Ieva, F., Paganoni, A.M.: Generalized mixed-effects random forest: a flexible approach to predict university student dropout. Stat. Anal. Data Min. ASA Data Sci. J. **14**(3), 241–257 (2021)

Evaluating Forecast Distributions in Neural Network Lee-Carter Type Model for Mortality Rate

Michele La Rocca(✉), Cira Perna, and Marilena Sibillo

Department of Economics and Statistics, University of Salerno, Fisciano, Italy
{larocca,perna,msibillo}@unisa.it

Abstract. In this paper we propose the use of a single hidden layer feed forward artificial neural network as a tool to appropriately capture the nonlinear dynamics of the mortality rates modeled by a Lee-Carter type model. The proposed procedure makes it possible to obtain point forecasts and, by using a bootstrap scheme, the forecast distributions, which allow to take into account the uncertainty of models' predictions. Empirical evidence on Italian data shows a significant improvement contribution of the proposed methodology.

Keywords: Lee-Carter Model · Feed-forward Neural Networks · Bootstrap forecast distributions

1 Introduction

Forecasting human mortality rates is an important problem that is receiving increasing attention especially in economics, demographics, social sciences and, in particular, in actuarial science. In the context of stochastic description of the human mortality, the Lee-Carter model [6] (from here on LC model) is one of the most popular for its simple implementation, its efficiency and its easily interpretable parameters.

In the original LC model, the time series describing the mortality trend over time, is modelled using a random walk model with drift. However, this can lead to obvious limitations, especially in a forecasting framework, due to the presence of any nonlinearities. To solve the problem, there has been growing interest in adopting machine learning techniques, focusing in particular on artificial neural networks (see [10]).

In this paper we explore the use of a single hidden layer feed-forward artificial Neural Networks (NNs) as a forecasting tool to capture the nonlinear dynamics of the mortality rates. The approach is also able to obtain, by using the bootstrap, forecast distributions which allow to evaluate how much uncertainty is associated with each point forecast. Despite their simple structure, NNs face several characteristics which make them valuable and attractive in the forecasting. They are data-driven self-adaptive methods and show good forecasting performance with high accuracy, without suffering the so called "curse of dimensionality". Furthermore, in our opinion, highly complex neural network models cannot be easily justified in the context of mortality rates due to the insufficient data availability, often encountered in these applications.

The paper is organized as follows. In Sect. 2 the LC model is briefly presented. In Sect. 3, the employed NN model is introduced and discussed and the bootstrap scheme

M. Corazza et al. (Eds.): MAF 2024, *Mathematical and Statistical Methods for Actuarial Sciences and Finance*, pp. 218–223, 2024.
https://doi.org/10.1007/978-3-031-64273-9_36

to obtain forecast distribution is presented. In Sect. 4 an application to Italian population is discussed along with some concluding remarks.

2 The Lee Carter Model: Recalls and Remarks

As in [6], the LC model is given by the following form:

$$\ln m_{x,t} = a_x + k_t b_x + \varepsilon_{x,t}$$

where $m_{x,t}$ is the observed central death rate at age x in year t, a_x is the average age-specific pattern of mortality, b_x is a parameter representing the age-specific deviations of mortality with respect to its averaged pattern as k_t varies, k_t being the time index describing the general mortality trend and $\varepsilon_{x,t}$ is the residual term at age x and time t. To avoid identifiability of the parameters, the model requires the following constraints $\sum_x b_x = 1$ and $\sum_x k_x = 0$.

Lee and Carter estimated a_x, b_x and k_t with U.S. mortality data from 1933 to 1987 using least squares. Specifically, they estimate a_x by averaging log-rates over time and b_x and k_t via a singular value decomposition of the residuals. In a second step they adjusted the k_t's so that the observed number of deaths coincide with those estimated. An important feature of the Lee-Carter approach is that, having reduced the time dimension of mortality to a single index k_t, it is possible to use statistical time series methods to model and forecast this index. In their application to U.S. mortality, Lee and Carter discovered that, except for the flu epidemic of 1918, the index behaves like a simple random walk with drift, where:

$$k_t = k_{t-1} + \delta + e_t$$

where δ is the drift and the e_t are independent error terms with variance v. The linearity of the model for k_t could be an obvious limitations, especially in a forecasting framework, due to the presence of any nonlinearities in k_t.

To overcome the problem, machine learning methods and, in particular, artificial neural networks have been recently used. For example, a Recurrent Neural Network with Long Short-Term Memory has been proposed in [9] to model and forecast mortality rates k_t and extended in [7] to derive prediction intervals. The same Neural Network architecture has been used in [8] to obtain a more accurate portrait of the future life expectancy and lifespan disparity. Alternatively, convolutional neural networks, combined with embedding layers, have been proposed in [10]. In the same framework, in [5], a semi-parametric model based on an Autoencoderhas been proposed to capture the nonlinearity features of the survival phenomenon.

3 Feed-Forward Neural Networks

We suppose that k_t is modeled as:

$$k_t = g(\mathbf{x}_{t-1}) + e_t$$

where $g(.)$ is an unknown, possibly nonlinear, function, $\mathbf{x}_{t-1} = (k_{t-1}, \ldots, k_{t-m})$ is a vector of m lagged values of k_t, and e_t are i.i.d. innovations with mean zero and finite variance.
The function $g(.)$ is approximated with a single input, single layer $NN(r,s)$ defined as:

$$f_s(\mathbf{x};\theta) = \sum_{j=1}^{s} c_j \psi\left(\mathbf{w}_j'\mathbf{x} + w_{j0}\right) + c_0 \tag{1}$$

with $\theta' = (c_0, c_1, \ldots, c_s, \mathbf{w}_1', \ldots, \mathbf{w}_s', w_{10}, \ldots, w_{s0})$, parameter vector of dimension $s(r+2)$, where s is the hidden layer size, $\{\mathbf{w}_j, j = 1, \ldots, s\}$ are the weight vectors of the connections between the input layer and the hidden layer; $\{c_j,\ j = 1, \ldots, s\}$ are the weights of the link between the hidden layer and the output neuron; $\{w_{j0}, j = 1, \ldots, s\}$ are the bias terms of the hidden neurons; $\psi(\cdot)$ is a proper chosen activation function for the hidden neurons. On the neural network, it is assumed that the activation function $\psi(\cdot)$ is a continuous squashing function with $\psi(\cdot) \in \mathscr{C}^2(\mathbb{R})$ and that the hidden layer size is such that $s = s(T) = O\left(\sqrt{T/\log T}\right)$, with T the length of time series.

Once the hidden layer size s and the r explanatory variables have been fixed, the weights of the network can be estimated as:

$$\hat{\theta} = \arg\min_{\theta} \sum_{t=m+1}^{T} \mathscr{L}\left(k_t, f_s(\mathbf{x}_{t-1};\theta)\right) + \frac{\lambda}{2}\|\theta\|^2 \tag{2}$$

where $\mathscr{L}$ is an appropriate loss function, f_s is a NN model with s neurons in the hidden layer used to forecast k_{T+1}, $\|\cdot\|$ is the L^2-norm and λ is a regularization parameter which forces the weights to decay towards zero and is usually fixed by cross-validation.

The proposed procedure allows to obtain point forecasts k_{T+h} with $h > 0$ and, using the bootstrap, the distributions of the forecasts, useful for taking into account the uncertainty of the model forecasts. In particular, we implement a standard bootstrap scheme from the residuals of the estimated $NN(r,s)$. The approach is in the spirit of neural network sieve bootstrap introduced in [1] and, consequently, it is asymptotically justified and, in small samples, outperforms alternative bootstrap schemes when the data generation process is non-linear [2]. The procedure is implemented as in Alghoritm 1

4 Application to Real Data and Concluding Remarks

In this section, we introduce NNs in the classical scheme of the LC model. In particular, our aim is to evaluate the advantages of bootstrap forecast intervals based on NN to improve the predictive ability of the LC model. We use the mortality datasets for the Italian population, both for males and females, for the years 1872–2019, collected from the website of Human Mortality Database (https://www.mortality.org/). The observations of the years 1872–1986 have been used as a training set, whereas the remaining observations (from 1987 to 2019) have been used as a testing set for evaluating the accuracy of the forecast results. The first step is to estimate the LC model parameters a_x, b_x and k_t using a singular value decomposition. The extracted time series k_t, whose plot is reported in Fig. 1, represents the initial base for the analysis. We have firstly verified if non linear features are present in k_t. To this aim, the Teraesvirta Neural Network

Algorithm 1. Bootstrap forecast distribution

Require: Fix the forecast horizon h, the number of the bootstrap runs B, the values of r and s.

1: Let $\hat{\theta}$ be the estimate of the neural network model NN(r,s) by using (2).
2: Let $\hat{F}(x)$ the ECDF of the centered residuals from the estimated neural network.
3: **for** b from 1 to B **do**
4: Let $\varepsilon_{T+h}^{(b)}$ be a random draw from $\hat{F}(x)$
5: Compute the future bootstrap observations as $\hat{k}_{T+h}^{(b)} = f(\hat{\mathbf{x}}_{T+h-1}^{(b)}, \hat{\theta}) + \varepsilon_{T+h}^{(b)}$ where $\hat{\mathbf{x}}_{T+h-1}^{(b)} = (\hat{k}_{T+h-1}^{(b)}, \ldots, \hat{k}_{T+h-m}^{(b)})$ with $\hat{k}_t^{(b)} = k_t$ if $t \leq T$.
6: **end for**
7: Given the set of B forecasts $\left\{\hat{k}_{T+h}^{(b)}, b = 1, \ldots, B\right\}$, compute the ECDF $\hat{F}_{T+h}^*$, which can be used to estimate the unknown distribution F_{T+h} of k_{T+h}
8: The $1-\alpha$ forecast interval for k_{T+h} is given by $\left[\hat{Q}^*(\alpha/2), \quad \hat{Q}^*(1-\alpha/2)\right]$ where $\hat{Q}^*(.)$ is the quantile function associated to the estimated bootstrap distribution $\hat{F}_{T+h}^*$.

Test for neglected nonlinearity, reported in Table 1, shows a clear rejection of linearity both for the female and male Italian data.

Following [7], as a benchmark, we have used an $ARIMA(p,d,q)$ model to forecast k_t, with orders p and q fixed by using the AIC information criterion and the differencing order d has been set equal to one, to obtain a stationary-in-mean time series. The best specification has been an $ARIMA(1,1,0)$ for females and an $ARIMA(0,1,0)$ for males, both with drift whose parameters have been estimated using maximum likelihood. Note that the Teraesvirta Neural Network Test, reported in Table 1, shows the non linearity of the residuals of both the models and, as a consequence, the inability of linear models to capture the nonlinear dynamics of the series.

The $ARIMA(p,d,q)$ performance has been then compared to a $NN(r,s)$. The $NN(r,s)$ has been implemented by using Eq. (1) with $\mathbf{x}_t$ including as explanatory variables m lagged values of k_t and the trend; that is $\mathbf{x}_t = (k_{t-1}, \ldots, k_{t-m}, t)$. The value of m has been determined by using the auto distance correlation function which measures the temporal dependence structure of a non linear time-series [12]. The hidden layer size s has been determined by using cross validation. The best specification has been a $NN(4,5)$ for females and a $NN(6,5)$ for males, both with a trend explanatory variable. The weights have been estimated by using (2) in which $\mathscr{L}(.)$ is the squared loss function and the regularization parameter has been fixed equal to 0.001. The Terasvirta test in Table 1 clearly shows the linearity of the residuals of both the NN models.

To compare the performance of the $NN(r,s)$ model against the best $ARIMA(p,d,q)$ in the testing set and to measure the forecasting ability of the two models, we have evaluated the point forecast accuracy by using the Root Mean Square Error (RMSE) and the Mean Absolute Error (MAE). The forecast distributions have been calculated by using the procedure illustrated in Algorithm 1 and the accuracy is evaluated through the Winkler score (WS) [11], the Quantile Score (QS) [4] and the Continuous Ranked Probability Score (CRPS) [3]. All computations have been implemented in R (version 4.3.1).

The results for female and male Italian population are reported in Table 2. Note that for females, the neural network model and the ARIMA model produce very similar

point forecasts, with a slight improvement of the latter compared to the former. However, when evaluating forecast distributions, a better performance of the neural network is evident: all accuracy scores of the neural network are significantly lower than those of the ARIMA model. For males, the neural network model has better accuracy of both point forecasts and forecast distribution. The results are confirmed by looking at Fig. 1 in which are reported, for both *ARIMA* and *NN* models, the point forecasts along with the $(1-\alpha)$ forecast intervals, with $1-\alpha = 0.80$ and $1-\alpha = 0.95$. It is evident that, in both cases, NN models produce narrower forecast intervals. Our results seem in line with the evidence reported, for example, in [7], in which a more complex Neural Network has been implemented.

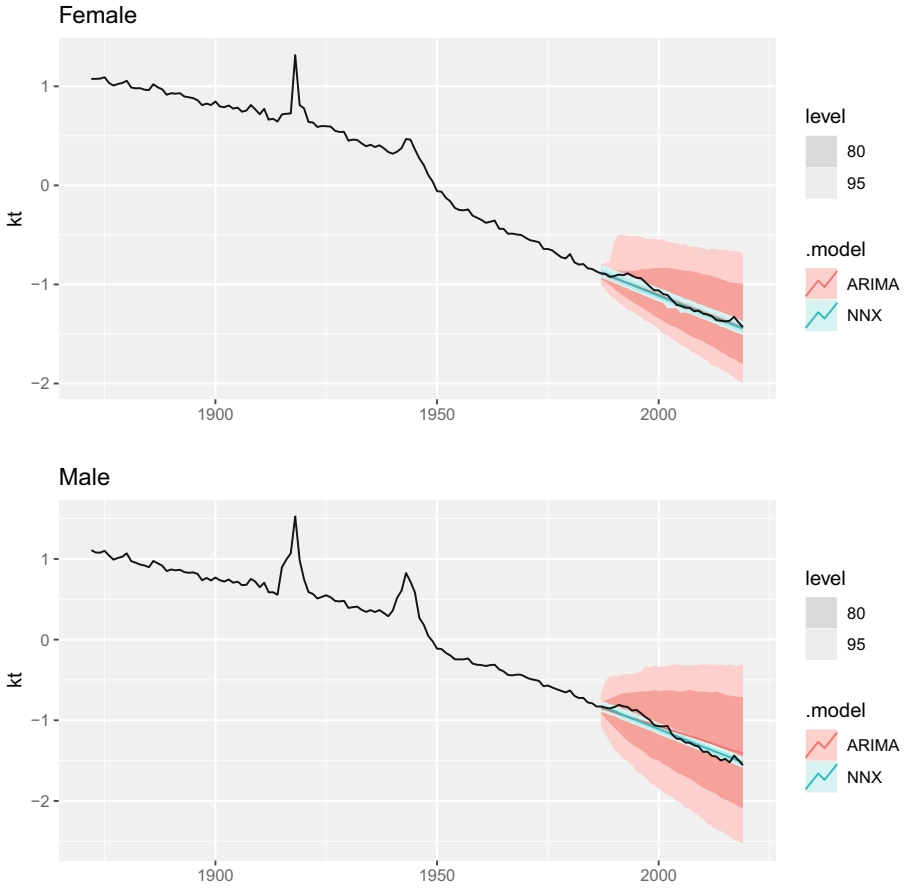


Fig. 1. Observed and forecasted values of k_t for female (on the top) and male Italian population (on the bottom).

Table 1. Test statistics and p-values for the Teräsvirta linearity test for the Italian population.

		Original series	ARIMA Residuals	NN Residuals
Female	Statistic	27.999	33.777	3.997
	pvalue	0.000	0.000	0.136
Male	Statistic	42.306	42.593	3.033
	pvalue	0.000	0.000	0.220

Table 2. Forecast accuracy measures for the Italian population.

		RMSE	MAE	WS	QS	CRPS
Female	ARIMA	0.0449	0.0337	0.9284	0.0532	0.0528
	NN	0.0473	0.0348	0.2243	0.0278	0.0276
Male	ARIMA	0.0988	0.0868	1.5612	0.0907	0.0898
	NN	0.0637	0.0527	0.4458	0.0409	0.0406

References

1. Giordano, F., La Rocca, M., Perna, C.: Forecasting nonlinear time series with neural network sieve bootstrap. Comput. Stat. Data Anal. **51**(8), 3871–3884 (2007)
2. Giordano, F., La Rocca, M., Perna, C.: Properties of the neural network sieve bootstrap. J. Nonparametr. Stat. **23**(3), 803–817 (2011)
3. Gneiting, T., Katzfuss, M.: Probabilistic forecasting. Annu. Rev. Stat. Appl. **1**, 125–151 (2014)
4. Hyndman, R.J., Athanasopoulos, G.: Forecasting: Principles and Practice, 3rd edn, vol. 23, no. 2. Monash University, Australia (2018)
5. La Rocca, M., Perna, C., Sibillo, M., Vignes, A.: Forecasting mortality with autoencoders: an application to Italian mortality data. In: Esposito, A., Faundez-Zanuy, M., Morabito, F.C., Pasero, E. (eds.) Applications of Artificial Intelligence and Neural Systems to Data Science, vol. 360. Springer, Singapore (2023). https://doi.org/10.1007/978-981-99-3592-5_17
6. Lee, R.D., Carter, L.R.: Modeling and forecasting US mortality. J. Am. Stat. Assoc. **87**(419), 659–671 (1992)
7. Marino, M., Levantesi, S., Nigri, A.: A neural approach to improve the Lee-Carter mortality density forecasts. North Am. Actuarial J. **27**(1), 148–165 (2023)
8. Nigri, A., Levantesi, S., Marino, M.: Life expectancy and lifespan disparity forecasting: a long short-term memory approach. Scand. Actuar. J. **2**, 110–133 (2021)
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. Actuarial J. 572–598 (2021)
11. Winkler, R.L.: A decision-theoretic approach to interval estimation. J. Am. Stat. Assoc. **67**(337), 187–191 (1972)
12. Zhou, Z.: Measuring nonlinear dependence in time-series, a distance correlation approach. J. Time Ser. Anal. **33**(3), 438–457 (2012)

Some Evidence Regarding Stock Markets and the Brexit

Diego Attilio Mancuso(✉)

Università Cattolica S.C., 20123 Milano, Italy
diego.mancuso@unicatt.it

Abstract. This work focuses on the changes among three stock exchange markets after the Brexit referendum. The markets and their indexes are the British FTSE100, the German index DAX40 and the American index S&P500. Using some nested GARCH and OLS models, it is shown that the volatility of the European indices reduced and the FTSE index decorrelated both with DAX and the S&P in a significant way after the referendum.

Keywords: Brexit · changing point analysis · GARCH models

1 Introduction

On the 23rd of June 2016 the referendum about the UK exit from the EU (or the Brexit referendum) can be considered as a breaking point for both the economic and the social path of Great Britain. The result seemed to be unexpected by financial markets considering their immediate reaction. Despite the full consequences of the referendum involve the medium term, this paper focuses on the short–term effects shown by the UK stock exchange index or the *FTSE100*. To achieve the goal, the period between the 2nd of January 2014 and the 31st of January 2020 was splitted into two sets according to the referendum date. I remind that the 1st of February 2020 was the date the UK stopped to be an EU member State and the Withdrawn Agreement entered into force. Two additional stock exchange indices were considered. They are the German *DAX40* and the U.S. *S&P500* indexes to represent the EU and American stock markets. In the next paragraph the volatility and the correlations of their log–returns are analyzed.

2 Time Series Analysis

2.1 Volatility

Although standard deviations of log – returns were lower in the period following the referendum than the prior, their naïve ratios are assumed as not correct to assess the changes into volatility cause to the conditional heteroskedasticity of the time series. GARCH models are considered more appropriate for the achievement and their results

M. Corazza et al. (Eds.): MAF 2024, *Mathematical and Statistical Methods for Actuarial Sciences and Finance*, pp. 224–228, 2024.
https://doi.org/10.1007/978-3-031-64273-9_37

for the full period (2/1/2014–31/1/2020) are reported below together with their log–likelihood. A traditional notation is adopted so h_t and r_t are respectively the conditional variance and the log–return at time t and $l(\theta)$ is the log– likelihood.

$$FTSE100 \quad h_t = 0.051 + 0.146r_{t-1}^2 + 0.781h_{t-1} \quad l(\theta) = -363.944$$

$$S\&P500 \quad h_t = 0.042 + 0.182r_{t-1}^2 + 0.758h_{t-1} \quad l(\theta) = -261.352$$

$$DAX40 \quad h_t = 0.026 + 0.076r_{t-1}^2 + 0.902h_{t-1} \quad l(\theta) = -805.24$$

The order of these models is optimal according to the likelihood ratio test with a significance level set to 0.05 and the *Schwartz* (*Bayesian*) *information criterium*.

Splitting the dataset before and after the referendum and applying different GARCH models to each of the subsets, leads to reject the unique specification for the *FTSE* and the *DAX* indices, but it doesn't for the *S&P500* according to the previous criteria. The updated optimal equations for the volatility of the *FTSE* and the *DAX* are below.

FTSE100

$$\begin{array}{ll} \textit{before referendum} & h_t = 0.03 + 0.15r_{t-1}^2 + 0.82h_{t-1} \\ \textit{after referendum} & h_t = 0.12 + 0.15r_{t-1}^2 + 0.61h_{t-1} \end{array} \quad l(\theta) = -353.53$$

DAX40

$$\begin{array}{ll} \textit{before referendum} & h_t = 0.03 + 0.08r_{t-1}^2 + 0.90h_{t-1} \\ \textit{after referendum} & h_t = 0.09 + 0.08r_{t-1}^2 + 0.79h_{t-1} \end{array} \quad l(\theta) = -793.75$$

By a well–known relationship linking parameters of the GARCH model and the unconditional variance of the process, a significant drop in the volatility can be stated both for the British and the German indexes after the referendum. This relationship is below.

$$V(r_t) = \frac{\omega}{1 - \sum_{i=1}^{p} \alpha_i - \sum_{i=1}^{q} \beta_i}$$

The standard deviations implied by models and the related empirical measures are reported in the next table for each of the periods (Table 1).

Table 1. Standard deviations.

Before referendum	FTSE100	S&P 500	DAX40
Expected by GARCH	1.24	0.84	1.66
Empirical measure	0.97	0.86	1.31
After referendum	FTSE100	S&P 500	DAX40
Expected by GARCH	0.71	0.84	0.85
Empirical measure	0.72	0.79	0.89

2.2 Stock Market Correlations

To analyze relationships among the markets, three sets of regressions were run. In every regression, the full series of log – returns of one index was used as a dependent variable and the log–returns of other indexes were used as independent variables. Every combination of splitted rather than full time series was considered, and the best model was selected according to the likelihood ratio test with a significance level set to 0.05 and the *Schwartz* (*Bayesian*) *information criterium.*

The regressions using no splitted independent variables according to the referendum date were always rejected as the optimal choice. Below the best regression models.

$$FTSE = 0.21\ \ S\&P_{before} + 0.09\ S\&P_{after} + 0.54\ DAX$$

$$DAX = 1.05\ FTSE_{before} + 0.75\ FTSE_{after} + 0.13\ S\&P_{before} + 0.29\ S\&P_{after}$$

$$S\&P = 0.38\ FTSE_{before} + 0.15\ FTSE_{after} + 0.12\ DAX_{before} + 0.38\ DAX_{after}$$

The interpretation of the results is quite clear. First, the correlations among the three markets changed after the referendum. Second, the influence of the *FTSE100* on the *DAX40* index weakened but the influence of the *DAX40* on the *FTSE100* appears to remain stable. Third, the result of the referendum strengthened the connections between the *DAX40* and the *S&P500* rather than the *FTSE100* and the *S&P500.*

Upon completion, marginal and partial correlation matrices are reported considering the two periods (Tables 2 and 3).

Table 2. Marginal correlation matrices before and after the referendum.

Before referendum	FTSE100	S&P 500	DAX40
FTSE100	1		
S&P 500	0.58	1	
DAX40	0.83	0.53	1
After referendum	FTSE100	S&P 500	DAX40
FTSE100	1		
S&P 500	0.45	1	
DAX40	0.72	0.53	1

Table 3. Partial correlation matrices before and after the referendum.

Before referendum	FTSE100	S&P 500	DAX40
FTSE100	1		
S&P 500	0.28	1	
DAX40	0.75	0.12	1
After referendum	FTSE100	S&P 500	DAX40
FTSE100	1		
S&P 500	0.11	1	
DAX40	0.64	0.33	1

Moreover, the descriptive statistic R^2 for all the regressions are in the table below. Again, they are computed prior and posterior the referendum date (Table 4).

Table 4. R^2 before and after the referendum

R^2	FTSE100	S&P 500	DAX40
Before referendum	0.71	0.34	0.69
After referendum	0.53	0.29	0.58

Finally, the overall performances for each index are reported below (Table 5).

Table 5. Overall performance in percentage

	FTSE100	S&P 500	DAX40
Before referendum	−6.09	14.33	7.38
After referendum	18.69	58.31	35.84
Whole period	7.96	74.51	35.91

Regarding the last table, it is worth of mention that the gap between the performances of the *DAX* and the *FTSE* is similar in the two periods: 0.13 before the Brexit referendum and 0.17 after it. What it changed was the way these gaps occurred due to the break in the correlation between the markets. A suchlike statement holds in comparing the *S&P* and the *FTSE* performances. Clearly, the last descriptive statistics are confirming the results of the inferential analysis of log–returns.

3 Conclusions

After the Brexit referendum a significant decorrelation of the *FTSE100* from the other indexes is observed together a drop in the volatility of the two European stock markets. Maybe against the common expectation, the direct correlation between the German and the American indexes strengthened after the referendum and that one involving the British and the American indexes weakened.

Portfolio Volatility Contributions of Risk Factors in the Presence of Risk Factors Multi-collinearity

Andrea Mecchina[1(✉)], Enrico Regolin[2], Nicola Torelli[3], and Luca Bortolussi[1]

[1] Department of Mathematics, Informatics and Geosciences, University of Trieste, Trieste, Italy
andrea.mecchina@phd.units.it
[2] Generali Asset Management SGR S.p.A., Trieste, Italy
[3] Department of Economics, Business, Mathematics and Statistics, University of Trieste, Trieste, Italy

Abstract. Attributing the volatility of a portfolio to some exogenous risk factors which are not directly invested in by the portfolio may be a topic of interest to asset managers. Without any restriction on the nature of risk factors, we must take into account that their returns may exhibit strong correlations. Risk factor returns multi-collinearity causes severe problems in estimating their portfolio volatility contributions. In order to solve this issue, we propose a risk attributing pipeline that applies an orthogonalisation algorithm to risk factor returns. Most importantly, the risk factors interpretability is preserved, in the sense that the orthogonalised risk factors are the ones attaining the least Frobenius norm of the matrix of deviations from the original risk factors.

Keywords: risk contributions · multi-collinearity · orthogonalisation · interpretability

1 Motivation

Identifying the different risk sources of a portfolio may be relevant to the work of an asset manager. However, assessing risk contributions of risk factors which a portfolio does not directly invest into requires to be handled with some care.

Considering a set of n assets $\{\mathcal{A}_1, \ldots, \mathcal{A}_n\}$, a portfolio in this investible universe is specified by a vector $w \in \mathbb{R}^n$, such that $\sum_{i=1}^n w_i = 1$. We take $w_i \geq 0$ $\forall i \in 1, \ldots, n$, thus not allowing short selling. We denote by $R \in \mathbb{R}^n$ the vector of asset returns at time t and by $\Sigma \in \mathbb{R}^{n \times n}$ their covariance matrix. We measure the risk of a portfolio by the volatility of its returns, i.e. $\sigma = \sqrt{w^T \Sigma w}$.

We consider a set of m risk factors $\{\mathcal{F}_1, \ldots, \mathcal{F}_m\}$, without any restriction on their nature, and we denote by $F \in \mathbb{R}^m$ the vector of their returns at time t and by $\Omega \in \mathbb{R}^{m \times m}$ their covariance matrix. Following Alexander [1], we assume the linear factor model

$$R = \alpha + BF + \varepsilon \tag{1}$$

M. Corazza et al. (Eds.): MAF 2024, *Mathematical and Statistical Methods for Actuarial Sciences and Finance*, pp. 229–234, 2024.
https://doi.org/10.1007/978-3-031-64273-9_38

for the asset returns at time t, where $\alpha \in \mathbb{R}^n$ is the vector of intercepts and $\varepsilon \in \mathbb{R}^n$ is the vector of residuals at time t. Allocating σ to the m risk factors requires estimating the $n \times m$ factor loadings matrix B.

A possible way to estimate the matrix B is by the ordinary least squares method. Anyway, as discussed by Hastie et al. [3], the presence of risk factor returns multi-collinearity may pose severe issues while trying to disentangle the separate effects of distinct risk factors on portfolio volatility.

Without any restriction on risk factors themselves, the presence of multi-collinearity is perfectly admissible and needs to be addressed properly in order to identify meaningful and interpretable portfolio volatility contributions.

2 Methodology

In order to illustrate our pipeline to allocate volatility to risk factors, we recall the theoretical results our method is built upon.

2.1 Risk Factors Volatility Contributions

Following Cazalet and Roncalli [2] and denoting by B^+ the Moore-Penrose inverse of the factor loadings matrix B, the risk contribution to the portfolio volatility σ of the j-th risk factor, where $j = 1, \ldots, m$, is defined as

$$\mathcal{RC}(\mathcal{F}_j) = \frac{(B^T w)_j (B^+ \Sigma w)_j}{\sigma} \,. \tag{2}$$

In the case of $m < n$, the risk allocation procedure gives rise to $n-m$ residual risk factors. These residual factors carry no meaningful financial interpretation, but we attribute to them the risk contributions

$$\mathcal{RC}(\tilde{\mathcal{F}}_k) = \frac{(\tilde{A} w)_k (\tilde{A} \Sigma w)_k}{\sigma} \,, \tag{3}$$

where $k = 1, \ldots, n-m$ and $\tilde{A} = \mathrm{Ker}(B^T)^T$ is an $n \times (n-m)$ matrix which, transposed, spans the left null space of B. The residual risk contributions vanish when $m \geq n$. As shown by Roncalli and Weisang [6], the residual risk contributions are necessary to satisfy the allocation principle

$$\sigma = \sum_{j=1}^{n} \mathcal{RC}(\mathcal{F}_j) + \sum_{k=1}^{n-m} \mathcal{RC}(\tilde{\mathcal{F}}_k) \,. \tag{4}$$

Remark 1. Despite portfolio volatility being strictly non-negative, risk factors volatility contributions could take negative values. When this happens, the volatility of the portfolio is *reduced* by the exposures to these risk factors.

2.2 Orthogonalisation Procedure

We apply the orthogonalisation algorithm proposed by Klein and Chow [4]. This procedure is *egalitarian*, due to its output not depending on the particular order into which risk factors are processed. A set of orthogonalised risk factor returns is obtained as a linear transformation of the original risk factor returns, namely

$$F^{\perp} = FS \, . \tag{5}$$

We denote $M = (T-1)\Omega$, where T is number of time observations, i.e. the number of rows of F. We refer to O as the $m \times m$ matrix whose columns are the eigenvectors of M and to $D^{-1/2}$ as the $m \times m$ diagonal matrix of the reciprocal square roots of the eigenvalues of M. The transformation matrix is defined as

$$S = OD^{-1/2}O \, . \tag{6}$$

Despite the epitome *orthogonalised*, the $T \times m$ matrix $F^{\perp}$ is *not* an orthogonal matrix per se, as $(F^{\perp})^T F^{\perp}$ it not the $m \times m$ identity matrix. We abuse this notation because $F^{\perp}$, obtained by adding constant terms to orthogonal time series, still results in uncorrelated time series.

Remark 2. The S transformation minimises the overall difference between the original and the orthogonalised risk factor returns for all the Schatten p-norms and, in particular, for the Frobenius norm when $p = 2$.

3 Case Study

To illustrate the effectiveness of the proposed methodology we present a case study on real data for assets, portfolio weights and risk factors.

3.1 Data Retrieval

We choose a set of $n = 31$ investible assets and we build on them a balanced portfolio, such that roughly 60% of our capital is allocated to fixed income assets and the remaining 40% to equity assets.

Remark 3. The assets are selected to be of interest to an investment company and thus they cannot be openly disclosed. However, such assets cover a large majority of financial instruments, diversified by sector and geographical area, and they are reasonable constituents of a strategic asset allocation, a diversified long-term portfolio strategy with periodic rebalancing.

We also choose $m = 5$ risk factors. They are mostly related to the US economy, despite our portfolio being globally diversified. However, we believe these risk factors to be broad enough to provide an exhaustive description of a global balanced portfolio such as ours. We label them as:

- Rate: an index tracking the performance of US treasury bills with a remaining term to maturity between 7 and 10 years;
- Corporate: an index tracking the performance of US investment grade-rated corporate bonds with a remaining term to maturity between 7 and 10 years;
- Market: an index tracking the performance of stocks from large and mid cap segments of the US market;
- Commodity: an index tracking the performance of global futures contracts on physical commodities;
- Currency: the euro against US dollar pair.

Both asset and risk factor series are retrieved on a daily basis, excluding non-business days, from 2012-12-31 to 2023-12-01. Returns series are obtained as percentage changes of the original series of total returns.

3.2 Empirical Results

To enforce a *dynamic* evolution of risk contributions, we choose a rolling window of 22 days, i.e. a month of business days, and we apply (5) to orthogonalise risk factor returns in each window. Risk factors contributions to volatility are computed in every window by (2) and (3). The matrix Σ is estimated by the shrinkage method proposed by Ledoit and Wolf [5], while for the matrix Ω we resort to the sample covariance matrix.

The most critical issue in providing an interpretable risk allocation comes from the fact that Rate and Corporate returns exhibit a very high correlation, exceeding 0.9 for more than 86% of the 2588 rolling correlations evaluated.

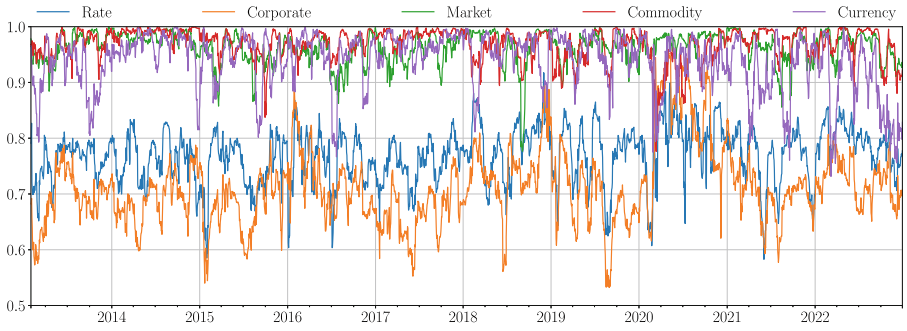


Fig. 1. Cross correlations between orthogonalised risk factor returns and their original counterparts over time. The rolling window is set to 22 business days

The rolling cross correlations between the orthogonalised risk factor returns and their original counterparts are displayed in Fig. 1. Due to their strong multicollinearity, the Rate and Corporate factors are those most modified by the orthogonalisation algorithm, thus exhibiting the lowest cross correlations. The other factors higher cross correlations indicate that they are not strongly multicollinear either among themselves or with the Rate and Corporate factors.

3.3 Results Interpretability

Figure 2(a) displays the annualised risk factors contributions to the volatility of our balanced portfolio. Annualisation is obtained by scaling by $\sqrt{260}$, taking 260 as the number of business days in a year. The contributions in Fig. 2(b) are normalised by dividing them by the volatility of the portfolio.

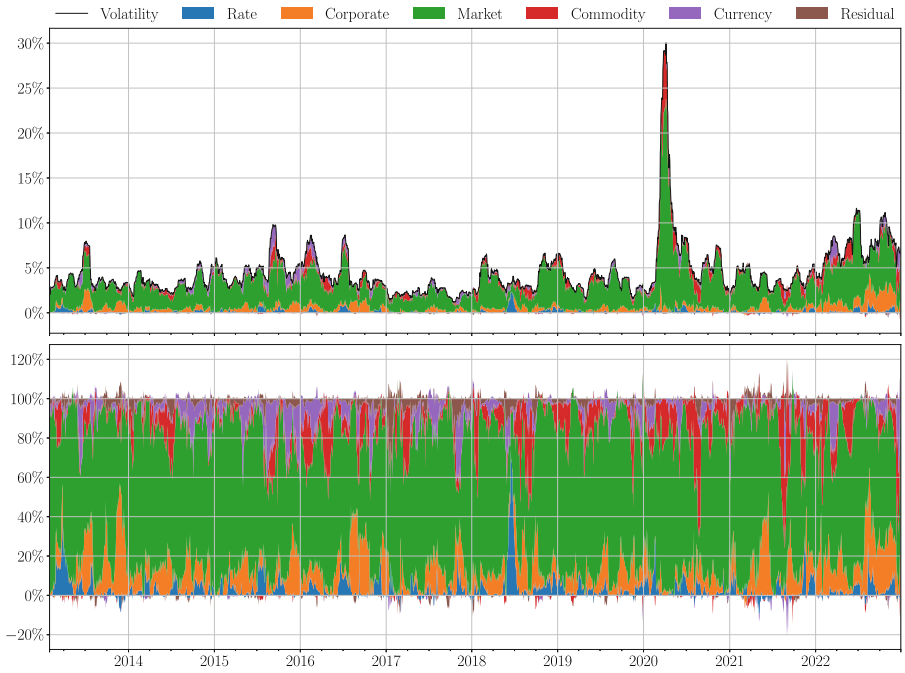


Fig. 2. Balanced portfolio risk factors volatility contributions over time. The rolling window is set to 22 business days. **a.** Risk factors contributions cross-sectionally add up to the portfolio volatility (*solid line*). **b.** Normalised risk factors contributions

Figure 2 allows us to attribute volatility spikes to turmoil periods and major economics events. For the sake of illustration, we comment on some of them:

- Q2 2013: the Fed taper tandrum coincides with a Corporate risk surge;
- Q3 2015: the Chinese stock market turbulence coincides with a Commodity and Currency risks surge;
- Q1 2016: the petrol prices plunge coincides with a Commodity risk surge;
- Q1 2018: the Volmageddon coincides with a Market risk surge;
- Q1 2020: the outbreak of the COVID-19 pandemic coincides with a Market risk surge;
- Q2 2020: the crude oil futures price turning negative coincides with a Commodity risk surge, although partially hidden in Fig. 2 due to the aftermath of the COVID-19 pandemic;

- Q3 2022: the global interest rate normalisation coincides with a Corporate risk surge.

4 Conclusions

In this article we developed an egalitarian pipeline to attribute the volatility of a portfolio to some exogenous risk factors in an interpretable way. The effectiveness of our methodology is illustrated by a case study on a balanced portfolio, and we believe our setting to be general enough to be of help for the different needs of an asset manager.

We also benchmarked our methodology against other techniques specifically designed to handle multi-collinearity between predictors in a linear regression context, such as Ridge or LASSO regressions, as discussed by Hastie et al. [3]. These methods did not prove completely successful because, despite shrinking toward zero the regression coefficients of multi-collinear risk factors, they gave rise to extreme and opposite volatility contributions, due to the lack of an explicit disentanglement of the returns of strongly correlated risk factors.

One future development of this work could be trying to extend the aforementioned Ridge or LASSO regularisation techniques by plugging into their loss functions some ad hoc penalty terms. Moreover, it could interesting to extend our methodology to risk measures other than portfolio volatility. Eventually, we could even compare some nonlinear models in place of (1), and try to automatically detect changes in the underlying risk dependence structure.

Notes and Comments. The authors are thankful to Luca Colussa, from Generali Asset Management SGR S.p.A., Trieste, Italy, for its invaluable contribution and for the fruitful discussions that made this work possible.

References

1. Alexander, C.: Market Risk Analysis, Boxset. Wiley, Hoboken (2009)
2. Cazalet, Z., Roncalli, T.: Facts and fantasies about factor investing. Available at SSRN 2524547 (2014)
3. Hastie, T., Tibshirani, R., Friedman, J.H., Friedman, J.H.: The Elements of Statistical Learning: Data Mining, Inference, and Prediction, vol. 2. Springer, New York (2009). https://doi.org/10.1007/978-0-387-84858-7
4. Klein, R.F., Chow, V.K.: Orthogonalized factors and systematic risk decomposition. Q. Rev. Econ. Finance **53**(2), 175–187 (2013)
5. Ledoit, O., Wolf, M.: Improved estimation of the covariance matrix of stock returns with an application to portfolio selection. J. Empir. Financ. **10**(5), 603–621 (2003)
6. Roncalli, T., Weisang, G.: Risk parity portfolios with risk factors. Quant. Finance **16**(3), 377–388 (2016)

Insurance Premium Implied by Rank Dependence and Probability Distortion

Martina Nardon(✉)

Department of Economics, University Ca' Foscari of Venice,
Cannaregio 873, 30121 Venezia, Italy
mnardon@unive.it

Abstract. The zero utility premium principle is generalized under the Cumulative Prospect Theory. Risk attitude and loss aversion are shaped via a utility or a value function, and probabilities of ranked results are replaced by decision weights. Transformation of objective probabilities models probabilistic risk perception. Some properties of the behavioral premium principle are presented.

We then discuss an application making specific assumptions about the value function, the probability distortion, and the distribution of the claim. In particular, we study the impact of loss aversion on the premium.

Keywords: Insurance Premium Principles · Zero Utility Principle · Cumulative Prospect Theory

1 Introduction

We discuss a premium principle under the Cumulative Prospect Theory (CPT) [19] which generalizes the *zero utility principle* ([4] and [5]), extending previous work of [13].

A number of contributions study insurance premium principles and risk measures under Rank-Dependent Utility Theory (RDU) [15] and Dual Utility Theory (DU) [23]. Distorted probabilities were introduced by [21] in the definition of a premium principle based on a proportional hazard transform of the decumulative distribution function of the insurance risk. [22] applies distortion operators in order to price financial and insurance risks; the approach is more related to the DU. [20] consider different probability weighting functions for gains and losses, with linear utility. Along the same line of research, one may also include the contributions [1,2,9,17], and [3]. [10] introduces a zero utility principle under RDU (see also [18] and [7]). [11] extend the equivalent premium principle in a CPT framework discussing some special cases with linear and exponential utility functions and the properties of the related premium; [12] study iterative conditions of the premium principle defined in [11]. Also [16] apply CPT in order to study the optimality of insurance from the viewpoint of the insured maximizing their prospect value subject to a proportional premium principle adopted by the insurer. More recently, [24] derive a CPT-based shortfall risk measure.

M. Corazza et al. (Eds.): MAF 2024, *Mathematical and Statistical Methods for Actuarial Sciences and Finance*, pp. 235–241, 2024.
https://doi.org/10.1007/978-3-031-64273-9_39

The paper is organized as follows. Section 2 introduces the behavioral premium principles and their properties. Section 3 discusses an application assuming particular functional forms for the value function, the probability weighting function and the continuous distribution of the claim. Section 4 concludes.

2 Behavioral Premium Principles

The *equivalent utility principle* (see [5]) defines a premium which makes the insurer, with a positive initial wealth W, indifferent before and after having accepted the risk, X. The premium, P, is the solution (if it exists) to

$$u(W) = \mathbb{E}[u(W + P - X)], \tag{1}$$

where u denotes the utility function satisfying the usual conditions. The loss severity, X, can be modeled by a non-negative random variable.

When $W = 0$, or the utility function is defined with respect to the a reference point which is set equal to the status quo $\hat{u}(x) = u(W + x)$ (see [10]), one refers to the *zero utility principle* (see [4], p. 86).

Problem (1) has been solved in the literature assuming, in particular, linear, CARA and CRRA utility functions. Further extensions of the premium principle (1) consider the introduction of probability distortion.

A weighting (or distortion) function w is a strictly increasing function which maps the probability interval $[0, 1]$ into $[0, 1]$, with $w(0) = 0$, and $w(1) = 1$. For an arbitrary random variable X, the generalized expected value can be defined by the following Choquet integrals:

$$\mathbb{E}_w(X) = \int_{-\infty}^{0} \left(w\left(\mathbb{P}(X > x)\right) - 1\right) dx + \int_{0}^{+\infty} w\left(\mathbb{P}(X > x)\right) dx, \tag{2}$$

and

$$\mathbb{E}_{w^+w^-}(X) = \int_{-\infty}^{0} \left(w^-\left(\mathbb{P}(X > x)\right) - 1\right) dx + \int_{0}^{+\infty} w^+\left(\mathbb{P}(X > x)\right) dx, \tag{3}$$

when different distortion functions are applied to probabilities of positive and negative results. In particular, when X is a non negative random variable, one has

$$\mathbb{E}_w(X) = \int_{0}^{+\infty} w\left(\mathbb{P}(X > x)\right) dx = \int_{0}^{+\infty} w\left(S_X(x)\right) dx, \tag{4}$$

where $S_X(x) = 1 - F_X(x)$ is the survival function.

When risk attitudes are shaped via a utility function, u, as in RDU, or a value function, v, as in CPT, then

$$\mathbb{E}_w(u(X)) = \int_{0}^{+\infty} w\left(\mathbb{P}(u(X) > y)\right) dy, \tag{5}$$

and

$$\mathbb{E}_{w^+w^-}(v(X)) = \int_{-\infty}^{0} \left(w^-\left(\mathbb{P}(v^-(X) > y)\right) - 1\right) dy + \int_{0}^{+\infty} w^+\left(\mathbb{P}(v^+(X) > y)\right) dy, \tag{6}$$

extend the EU value under RDU and CPT, respectively, for a convenient random variable X.

In particular, when a linear utility function is considered in problem $u(W) = \mathbb{E}_w(u(W+P-X))$, the *distortion premium principle* (with w convex, see [21]) results:

$$P = \mathbb{E}_{\overline{w}}(X)\,, \tag{7}$$

where $\overline{w}$ is the dual probability weighting function[1]. Note that, when w is convex, $\overline{w}$ is concave.

As another special case, let $u = (1-e^{-ax})/a$, with $a > 0$, and consider condition

$$u(W) = \int_0^{+\infty} w\left(\mathbb{P}(u(W+P-X) > y)\right) dy;$$

then the *generalized exponential premium principle* under RDU arises (see [10]):

$$P = \frac{1}{a} \ln \mathbb{E}_{\overline{w}}\left(e^{aX}\right). \tag{8}$$

2.1 Premium Principle Implied by CPT

Differently from EU theory, in CPT individuals are risk averse when considering gains and risk seeking with respect to losses, and they are loss averse. Results are measured relative to a *reference point* rather than in terms of final wealth. The random result $P-X$ in (1) can be positive or negative and is evaluated through a value function, v, which is continuous and strictly increasing, with $v(0) = 0$, and non differentiable at the reference point. Objective probabilities are replaced by decision weights, which are differences in transformed, through a weighting function w, cumulative probabilities of losses and counter-cumulative probabilities of gains.

The zero utility principle (1) can be generalized under CPT as follows:

$$v(0) = \mathbb{E}_{w^+ w^-}[v(P-X)]\,. \tag{9}$$

Let the loss severity X be modeled by a non-negative continuous random variable, with cumulative distribution F_X and density f_X, then the premium P for insuring X is implicitly defined by the condition

$$\begin{aligned} 0 = &\int_0^P v^+(P-x)\,\psi^+[F_X(x)]\,f_X(x)\,dx + \\ &+ \int_P^{+\infty} v^-(P-x)\,\psi^-[1-F_X(x)]\,f_X(x)\,dx, \end{aligned} \tag{10}$$

where $\psi(p) = w'(p)$ is the derivative of the probability weighting function. Condition (10) defines the *zero prospect value premium principle* based on CPT for continuous random variables (see [13]). When $v(x) = x$, and probabilities are not distorted, $w(p) = p$, then the behavioral premium is equal to the equivalence premium, $P = \mathbb{E}(X)$.

The premium principle implied by CPT satisfies the following properties (see [13] for a proof):

[1] The dual probability weighting function $\overline{w}$ is defined as $\overline{w}(p) = 1 - w(1-p)$, with $\overline{w}'(p) = w'(1-p)$. Observe also that $\mathbb{E}_w(-X) = -\mathbb{E}_{\overline{w}}(X)$.

No unjustified safety (or risk) loading: $P(a) = a$, for all constants a;

Non-excessive loading: $P(X) \leq \sup(X)$;

Translation invariance: $P(X + a) = P(X) + a$, for all a.

Considering general value and distortion functions, positive scale invariance ($P(aX) = aP(X)$, for $a > 0$) does not hold for the premium defined by (10) (see also [11]). This property holds under RDU if and only if the value function is linear. Additivity for independent risks, additivity for comonotonic risks, sub-additivity, stop-loss order preserving are studied and proved under RDU (see [6–8, 10], and [11]) for a class of functions including linear and exponential utility, with some restrictions on the value function and on the shape of the probability weighting function.

3 An Application with Exponential Value Function

Let us consider the following exponential value function:

$$v(x) = \begin{cases} v^+ = (1 - e^{-ax})/a & x \geq 0 \\ v^- = \lambda \left(e^{bx} - 1\right)/b & x < 0, \end{cases} \tag{11}$$

where $\lambda > 1$ is the loss aversion parameter, $a > 0$, and $b > 0$. When v is substituted in (10), one obtains the condition:

$$0 = \int_0^P \frac{1 - e^{-a(P-x)}}{a} \psi^+[F_X(x)] f_X(x)\, dx + \int_P^{+\infty} \lambda \frac{e^{b(P-x)} - 1}{b} \psi^-[1 - F_X(x)] f_X(x)\, dx. \tag{12}$$

Consider the two-parameter compound-invariant probability weighting function suggested by Prelec [14]:

$$w(p) = e^{-\delta(-\ln p)^\gamma}, \tag{13}$$

with $\delta > 0$, $\gamma > 0$. When $\gamma < 1$ (with some restrictions on δ) the function has an inverse-S shape. It's derivative is $\psi(p) = \delta\gamma \frac{1}{p}(-\ln p)^{\gamma-1} e^{-\delta(-\ln p)^\gamma}$.

A problem studied in [13], Sect. 6, is extended in two ways: random results $P - X$ are evaluated in an aggregated framing, and a more flexible two-parameters weighting function is adopted.

Assume that the random variable X has a Weibull distribution with parameters $\alpha > 0$ and $\beta > 0$, with

$$f_X(x) = \frac{\alpha}{\beta}\left(\frac{x}{\beta}\right)^{\alpha-1} e^{-\left(\frac{x}{\beta}\right)^\alpha}, \qquad F_X(x) = 1 - e^{-\left(\frac{x}{\beta}\right)^\alpha}.$$

One obtains: $w(1 - F_X(x)) = e^{-\delta\left(\frac{x}{\beta}\right)^{\alpha\gamma}}$, and $\psi(1 - F_X(x)) = \delta\gamma\left(\frac{x}{\beta}\right)^{\alpha(\gamma-1)} e^{\delta\left(\left(\frac{x}{\beta}\right)^\alpha - \left(\frac{x}{\beta}\right)^{\alpha\gamma}\right)}$.

We present some numerical results, based on Prelec's probability weighting function, and assuming for X a Weibull distribution with parameters $\alpha = 5$ and $\beta = 100$. We

have $\mathbb{E}(X) = 91.8169$ and $Var(X) = 4.4230$. Let us consider $a = b$ in the value function (11). Problem (12) can be solved numerically for P (convenient truncation in the second integral is also required). It is worth noting that, when a approaches zero, with $\lambda = 1$, and $w(p) = p$, the premium tends to $P = \mathbb{E}(X) = 91.8169$.

Results are reported in Tables 1 and 2. In particular, in Table 1 only the effect of the value function is computed (moderate values of the parameters are considered), with no probability distortion. P increases with loss aversion, λ, and risk aversion, a. Table 2 shows the results when full form of CPT is adopted. Similar results are obtained with other parameter settings. It is worth noting that loss aversion plays an important role. When λ is close to one, and the shape of probability distortion implies that medium and high probabilities of mid-rank results are under-weighted, combined with the Weibull distribution[2], it may happen that the $P < \mathbb{E}(X)$.

Table 1. Resulting premiums under CPT when $X \sim Wei(\alpha, \beta)$, with $\alpha = 5$, $\beta = 100$, $\mathbb{E}(X) = 91.8169$; with exponential value function (11), assuming $a = b$, letting parameters a and λ vary; assuming no probability distortion ($\gamma^+ = \gamma^- = 1$, and $\delta^+ = \delta^- = 1$)

	a				
λ	0.02	0.04	0.06	0.08	0.1
1	92.0969	92.3013	92.4493	92.5568	92.6355
1.25	94.1007	94.4140	94.6536	94.8380	94.9816
1.5	95.7191	96.1175	96.4284	96.6727	96.8668
1.75	97.0729	97.5404	97.9091	98.2017	98.4365
2	98.2342	98.7593	99.1759	99.5086	99.7771

Table 2. Resulting premiums under CPT when $X \sim Wei(\alpha, \beta)$, with $\alpha = 5$, $\beta = 100$; with value function (11), assuming $a = b$; adopting Prelec's distortion function, with $\gamma^+ = 0.5$, $\gamma^- = 0.6$, $\delta^+ = 1.05$, and $\delta^- = 0.95$)

	a				
λ	0.02	0.04	0.06	0.08	0.1
1.25	92.7326	93.7565	94.3928	94.7937	95.0571
1.5	95.7133	96.8809	97.6059	98.0676	98.3760
1.75	98.1868	99.4642	100.2574	100.7665	101.1103
2	100.2956	101.6608	102.5088	103.0564	103.4294

[2] With an inverse-S shape probability weighting function, decision weights assigned to extreme events are higher than objective probabilities. Nevertheless, in the application, a Weibull distribution is used; alternative distributions with heavier tails, such as the log-normal one, could be considered.

4 Concluding Remarks

Prospect Theory attracts increasing interest in the insurance theory literature and it seems promising for its potential to explain a range of behaviors (diminishing marginal utility, loss aversion, and probability weighting). We defined a premium principle under CPT extending the zero utility principle. In the application, some preliminary results are presented; the obtained behavioral premiums are higher than the equivalent premium, resulting in an implicit loading. As further extensions, fixed-percentage and fixed-amount deductibles can be introduced in the analysis.

References

1. Balbás, A., Garrido, J., Mayoral, S.: Properties of distortion risk measures. Methodol. Comput. Appl. Probab. **11**, 385–399 (2009)
2. Belles-Sampera, J., Merigó, J.M., Guillén, M., Santolino, M.: The connection between distortion risk measures and ordered weighted averaging operators. Insur. Math. Econ. **52**(2), 411–420 (2013)
3. Belles-Sampera, J., Guillén, M., Santolino, M.: The use of flexible quantile-based measures in risk assessment. Commun. Stat. Theory Methods **45**(6), 1670–1681 (2016)
4. Bühlmann, H.: Mathematical Methods in Risk Theory. Springer, Heidelberg (1970). https://doi.org/10.1007/978-3-540-30711-2
5. Gerber, H.U.: An Introduction to Mathematical Risk Theory. S.S. Huebner Foundation for Insurance, University of Pennsylvania, Philadelphia (1979)
6. Gerber, H.U.: On additive principles of zero utility. Insur. Math. Econ. **4**, 249–251 (1985)
7. Goovaerts, M.J., Kaas, R., Laeven, R.J.A.: A note on additive risk measures in rank-dependent utility. Insur. Math. Econ. **47**(2), 187–189 (2010)
8. Goovaerts, M.J., Kaas, R., Laeven, R.J.A., Tang, Q.: A comonotonic image of independence for additive risk measures. Insur. Math. Econ. **35**(3), 581–594 (2004)
9. Hamada, M., Sherris, M.: Contingent claim pricing using probability distortion operators: methods from insurance risk pricing and their relationship to financial theory. Appl. Math. Finance **10**, 19–47 (2003)
10. Heilpern, S.: A rank-dependent generalization of zero utility principle. Insur. Math. Econ. **33**(1), 67–73 (2003)
11. Kaluszka, M., Krzeszowiec, M.: Pricing insurance contracts under cumulative prospect theory. Insur. Math. Econ. **50**(1), 159–166 (2012)
12. Kaluszka, M., Krzeszowiec, M.: On iterative premium calculation principles under cumulative prospect theory. Insur. Math. Econ. **52**(3), 435–440 (2013)
13. Nardon, M., Pianca, P.: Behavioral premium principles. Decis. Econ. Finan. **42**, 229–257 (2019)
14. Prelec, D.: The probability weighting function. Econometrica **66**, 497–527 (1998)
15. Quiggin, J.: Generalized Expected Utility Theory: The Rank-Dependent Model. Springer, Dordrecht (1993). https://doi.org/10.1007/978-94-011-2182-8
16. Sung, K.C.J., Yam, S.C.P., Yung, S.P., Zhou, J.H.: Behavioral optimal insurance. Insur. Math. Econ. **49**, 418–428 (2011)
17. Tsanakas, A.: To split or not to split: capital allocation with convex risk measures. Insur. Math. Econ. **44**, 268–277 (2009)
18. Tsanakas, A., Desli, E.: Risk measures and theories of choice. Br. Actuarial J. **9**(4), 959–991 (2003)

19. Tversky, A., Kahneman, D.: Advances in prospect theory: cumulative representation of the uncertainty. J. Risk Uncertain. **5**, 297–323 (1992)
20. van der Hoek, J., Sherris, M.: A class of non-expected utility risk measures and implications for asset allocations. Insur. Math. Econ. **28**(1), 69–82 (2001)
21. Wang, S.S.: Premium calculation by transforming the layer premium density. ASTIN Bull. **26**(1), 71–92 (1996)
22. Wang, S.S.: A class of distortion operators for pricing financial and insurance risks. J. Risk Insur. **67**(1), 15–36 (2000)
23. Yaari, M.: The dual theory of choice under risk. Econometrica **55**(1), 95–115 (1987)
24. Zhang, S., Xu, H.: Insurance premium-based shortfall risk measure induced by cumulative prospect theory. Comp. Manag. Sci. **19**, 703–738 (2022)

Disclosing the Reserving Process in Life Insurance Through Equivalent Periodic Fees

Annamaria Olivieri(✉)

Department of Economics and Management, University of Parma, Parma, Italy
annamaria.olivieri@unipr.it

Abstract. In the liability-driven life business (i.e., with fixed or participating benefits), the understanding of the reserving process may be out the reach of the policyholder, due to the actuarial principles on which the valuation is based. In particular, the coherence between the costs described under policy conditions and the amount accumulated into the policy reserve may be lost. When the life policy is mainly viewed as an investment by the policyholder, alternative investments may be considered (perhaps erroneously) more convenient and convincing, as their dynamics appears more intuitive. We show that, based on the information available to the policyholder, the reserve dynamics can be reinterpreted in terms of equivalent periodic fees charged to the value of the policyholder's investment. Most policyholders should understand the meaning of a periodic fee, thus gaining greater awareness of the protection provided by life policies, as well as better insights on their costs.

Keywords: Prospective reserve · Embedded periodic fee · Cost of mutuality · Participating life policies · Participating annuities · Liability-driven business · Financial literacy

1 Introduction

In insurance, and in life insurance in particular, premiums are collected on average before the benefits are paid, to ensure that, apart from adverse circumstances, mutuality is cross-subsidised with the money paid by members of the insurance pool, and not with insurer's capital. As a result, the insurer is in a debt position, which is protected by a mandatory fund, the so-called policy reserve.

The policy reserve is assessed based on actuarial principles, usually following a prospective approach, i.e. as the expected present value of future benefits net of future premiums, including a (safety or risk) margin meant to improve the ability of the insurer to meet its obligations even in unfavourable situations. The specific calculation rule of the policy reserve, and the margin in particular, depends on regulation (e.g., local GAAPs, IFRS, Solvency 2). See [7] for a general discussion. The calculation of reserves is described in many actuarial textbooks. See, for example, [2–4].

M. Corazza et al. (Eds.): MAF 2024, *Mathematical and Statistical Methods for Actuarial Sciences and Finance*, pp. 242–247, 2024.
https://doi.org/10.1007/978-3-031-64273-9_40

When the insurance policy is designed for providing not just a protection against a life-contingent event, but also an investment opportunity, the amount of the reserve is regularly disclosed to the policyholder, as from his/her viewpoint it represents the value of his/her investment. However, the understanding of its dynamics may be out of the reach of the individual, unless (s)he holds an actuarial background. In particular, the coherence between the costs described under policy conditions and the amount accumulated into the policy reserve may be lost.

This is not the case of the asset-driven life business, i.e. when benefits are unit-linked. In this kind of policies, by constructions the reserve takes the value of the policy assets, to which periodic fees are explicitly charged to cover the cost of any guarantee or rider benefit. See, for example, [3]. Regardless of the policyholder's ability to assess whether the fees are fair or not, this approach makes the reserving process rather intuitive. Further, fees are applied in a way that is similar to financial products, to which very often individuals compare life policies, as in their view they can represent a convenient alternative.

Whatever the rule actually adopted for its calculation, also in the case of the liability-driven business the reserve dynamics can be reinterpreted in terms of equivalent periodic fees implicitly charged to the policyholder's investment. This is based on information usually accessible to the policyholder. Analysing the dynamics of the periodic fees could improve awareness of the protection provided by the insurance policy, and better insights into the costs charged by the insurer, including loadings. For their part, insurers can find opportunities for product innovations, if policyholders are better aware of their features and costs. On the other hand, if policyholders are better able to estimate loadings, insurers may need to revise their pricing parameters. It is then important for them to be able to justify the size and dynamics of the equivalent periodic fees consistently with the features of the guarantees they are providing.

In this paper we illustrate how to assess equivalent periodic fees, and provide some simple examples for participating endowments and annuities. Although the concept is straightforward, it has the potential to contribute to the financial literacy competencies of individuals. This is an important issue in the context of the financial well being of individuals, in the agenda of many policy makers; see, for example, [5,6]. Our research aim is to check which insights can be gained by individuals from the analysis of equivalent periodic fees, with particular regard to their time profile. To address the individual's perspective, we adopt a simplified framework; in particular, the setting is deterministic and discrete-time.

The paper is organized as follows. In Sect. 2 we define the equivalent periodic fees, and introduce a possible splitting into mutuality and loading components. In Sect. 3 we analyse their time profile in some numerical examples. Finally, in Sect. 4, we close the paper by outlining some further investigations.

2 Equivalent Periodic Fees and Components

In this Section we interpret the dynamics of the policy reserve in the perspective of the policyholder, based on the information available to him/her. Considering

an individual with a basic (or not too much advanced) financial education, we perform our analysis in a discrete-time setting. For brevity, we disregard expenses or asset management fees (which are common to other financial products), and we only focus on the investment return and cost of mutuality (the latter being typical of insurance products). Since the policyholder may consider the value of the policy reserve as the value of his/her investment, we will also call it policy account value (a usual terminology in the asset-driven business).

We consider a general structure of a life policy, allowing us to obtain as particular cases either an endowment insurance or a life annuity. We denote with: π_{t-1} the net premium paid at time $t-1$; $B_t^{[\mathrm{D}]}$ the benefit paid at time t in case of death during $(t-1,t)$; b_t the survival benefit paid at time t if the insured is alive. An endowment insurance is represented by setting a maturity m and: $B_t^{[\mathrm{D}]} > 0$ for $t = 1, 2, \ldots, m$; $b_t > 0$ for $t = m$ and $b_t = 0$ for all other times. A standard annuity is represented by setting: $\pi_{t-1} > 0$ for $t - 1 = 0$ and $\pi_{t-1} = 0$ for all other times; $b_t > 0$ for $t = 1, 2, \ldots$. A life annuity with capital protection can be obtained by setting $B_t^{[\mathrm{D}]} > 0$ for $t = 1, 2, \ldots, m$, where m is a finite time. We assume that a traditional pricing approach is adopted. This means that an actuarial balance between net premiums and benefits is assessed by setting a conservative discount rate (the so-called technical interest rate, i'), and a conservative mortality assumption (that we will represent in terms of a sequence of mortality rates q'_{x+t-1}, $t = 1, 2, \ldots$). We denote with x the entry age of the insured, i.e. his/her age at time 0. Adoption of conservative assumptions in the pricing parameters results in a loading embedded into net premiums, where the loading represents the insurer's expected profit.

For a participating policy, whatever the specific rule adopted by the insurer for assessing the value of the reserve, at each policy anniversary the policy reserve is credited an interest amount, based on the return of the assets backing the reserve, net of the cost of mutuality.

For a policy in-force at time t, the dynamics of the policy reserve between time $t-1$ and time t can be synthetically described as follows:

$$V_t = (V_{t-1} + \pi_{t-1}) \cdot (1 + i_t^{[\mathrm{net}]}) - b_t \ , \tag{1}$$

where $i_t^{[\mathrm{net}]}$ is a measure of the net annual return to the policyholder. Let i_t be the return on the assets backing the reserve (in the participating business, this information is usually disclosed to policyholders). The difference $i_t - i_t^{[\mathrm{net}]}$ is a synthetic measure of the cost of mutuality and loadings.

To split $i_t - i_t^{[\mathrm{net}]}$ into a mutuality and loading component, we need a mortality/longevity index. This information can be obtained from national mortality statistics, or from indexes available in the market of mortality/longevity-linked securities (see [1] for an overview). We assume that from the mortality index it is possible to obtain an estimate of the mortality rate in a reference population, that we denote as $q_{x+t-1}^{[\mathrm{ref}]}$ at age $x+t-1$. In the life insurance business, mutuality arises from the insurer having to pay the benefit $B_t^{[\mathrm{D}]}$ in case of death, while the value that would be cumulated in the policyholder's investment would be

$V_t + b_t$ (i.e., the policy account value before the payment of the survival benefit). The cost of mutuality is then proportional to the amount $B_t^{[\mathrm{D}]} - V_t - b_t$, i.e. the additional death benefit (in respect of the investment value). We describe the cost of mutuality in the following way:

$$(V_{t-1} + \pi_{t-1}) \cdot \xi_t^{[\mathrm{mut}]} = (B_t^{[\mathrm{D}]} - V_t - b_t) \cdot q_{x+t-1}^{[\mathrm{ref}]} \ , \tag{2}$$

where $\xi_t^{[\mathrm{mut}]}$ is the fee for mutuality. We note that the mortality rate realized in the insurer's portfolio, that we denote as $q_{x+t-1}^{[\mathrm{ptf}]}$, can be other than $q_{x+t-1}^{[\mathrm{ref}]}$. This means that an assessment based on the mortality index can over- or underestimate the cost of mutuality really reported by the insurer.

Finally, the fee for loadings can be assessed as follows:

$$\xi_t^{[\mathrm{load}]} = i_t - i_t^{[\mathrm{net}]} - \xi_t^{[\mathrm{mut}]} \ . \tag{3}$$

Again we note that $\xi_t^{[\mathrm{load}]}$ may not coincide with the loading actually charged by the insurer, due to the difference between $q_{x+t-1}^{[\mathrm{ptf}]}$ and $q_{x+t-1}^{[\mathrm{ref}]}$.

The fee for loadings can be further split into a financial and mortality component, if the reserving basis is known. For brevity, we do not discuss this aspect.

3 Numerical Implementation

In this Section we discuss some examples, in a deterministic setting. In particular, we examine the equivalent fees emerging under a given trajectory of the investment return and the mortality index, for participating endowments and annuities.

We first consider a standard endowment insurance; see Table 1, listing the equivalent fees and components for a selection of years. Pricing and other policy parameters are reported in the caption of the table. We point out that the revaluation rate of the reserve implies a minimum guaranteed annual return equal to 2%. If the realized investment return is below the minimum, the insurer incurs into a loss. The mortality rates recorded in the reference population are listed in terms of their ratio to the pricing mortality rates. A ratio higher (lower) than 1 means a loss (profit) to the insurer; however, we recall that the real loss (profit) reported by the insurer depends on the mortality rate realized in its pool, that can be other than in the reference population. The estimated size of the insurer's profit or loss is reflected in the value of the fee for loadings, which can also be negative (in case of large loss). From the point of view of the policyholder, it is interesting to understand how much of the realized return i_t is absorbed by the cost of mutuality. Such a cost depends on the realized mortality, but also on the additional amount in case of death, which is defined by the policy design. If the policyholder prefers to pay less for mutuality, (s)he should choose policies with a lower additional death benefit amount (accepting, however, to lose part of the insurance protection). To illustrate this, we consider alternatively annual level or single recurrent premiums. In the latter case, the benefit amount is progressively

Table 1. Endowment participating insurance. Entry age: 50. Duration: 10 years. Annual premium amount: 1,000.00 euro. Pricing interest rate: $i' = 2\%$. Pricing life table: Period, population life table (mortality rates q'_{x+t-1}). Annual revaluation rate of the reserve: $r_t = \max\left\{\frac{0.95 \cdot i_t - i'}{1+i'}, 0\right\}$.

t	i_t	$\frac{q^{[\text{ref}]}_{x+t-1}}{q'_{x+t-1}}$	Level premiums[a]			Single recurrent premiums[b]		
			$i_t^{[\text{net}]}$	$\xi_t^{[\text{mut}]}$	$\xi_t^{[\text{load}]}$	$i_t^{[\text{net}]}$	$\xi_t^{[\text{mut}]}$	$\xi_t^{[\text{load}]}$
1	2.256%	0.9196	0.171%	1.811%	0.274%	2.105%	0.035%	0.116%
3	2.342%	1.1112	1.600%	0.692%	0.050%	2.190%	0.039%	0.114%
5	1.923%	1.1935	1.679%	0.383%	−0.139%	1.970%	0.036%	−0.083%
7	2.262%	0.8573	1.982%	0.142%	0.138%	2.127%	0.018%	0.117%
9	2.518%	1.1311	2.339%	0.058%	0.121%	2.384%	0.008%	0.127%

[a]: Initial benefit amount 11,018.89 euro.
[b]: Initial benefit amount (first premium) 1,216.06 euro.

cumulated, resulting in lower death benefit amounts before reaching maturity. Indeed, the cost of mutuality is lower in the case of single recurrent premiums.

Table 2 quotes the equivalent periodic fees for a life annuity, for a selection of years. We first note that the fee for mutuality takes negative value, as in the case of life annuities mutuality implies a mortality credit to survivors. The size of such a credit becomes higher as the individual ages, highlighting the advantages of an annuity in respect of purely financial post-retirement arrangements, especially at older ages; this is a very well-known feature to actuaries and insurers, but possibly misunderstood by individuals. It is also interesting to note the increasing path of the net return to the policyholder, mainly originated by the mortality credits. We consider both the case of a standard life annuity and a life annuity with capital protection. In the latter case, the presence of the death benefit clearly reduces the size of mortality credits. Other comments are in line with what discussed for endowments.

4 Further Investigations

In this paper we suggest an intuitive way to describe the life insurance reserving process to the individual, which is based on the information accessible to the policyholder. Such information include primarily the policy reserve amount and the return on the assets backing the reserve, that are usually disclosed by the insurer. Further, if information about the mortality realized in the insurer's portfolio are not available, the use of a mortality/longevity index can allow to obtain proxy estimates. Based on this set of information, the reserving (and ultimately the pricing) process can be reinterpreted in terms of equivalent periodic fees. In practice, this is a simple way to reinterpret the concept of actuarial fairness, making it more accessible to an individual not able to perform complex assessments. Such a reinterpretation can possibly suggest areas for improvement

Table 2. Participating annuity. Entry age: 60. Initial benefit amount: 100.00 euro. Pricing interest rate: $i' = 2\%$. Pricing life table: Projected, cohort, selected life table (mortality rates q'_{x+t-1}). Annual revaluation rate of the reserve: $r_t = \max\left\{\frac{0.95 \cdot i_t - i'}{1+i'}, 0\right\}$.

t	i_t	$\frac{q^{[\text{ref}]}_{x+t-1}}{q'_{x+t-1}}$	Standard annuity[a]			Annuity with capital protection[b]		
			$i_t^{[\text{net}]}$	$\xi_t^{[\text{mut}]}$	$\xi_t^{[\text{load}]}$	$i_t^{[\text{net}]}$	$\xi_t^{[\text{mut}]}$	$\xi_t^{[\text{load}]}$
1	2.071%	1.1497	2.556%	−0.640%	0.155%	2.011%	−0.013%	0.073%
5	2.443%	1.0543	3.066%	−0.785%	0.163%	2.400%	−0.086%	0.129%
10	2.288%	1.1804	3.400%	−1.448%	0.336%	2.479%	−0.362%	0.171%
15	2.372%	1.0767	4.436%	−2.351%	0.286%	4.436%	−2.351%	0.286%
20	2.267%	1.0432	6.675%	−4.717%	0.309%	6.675%	−4.717%	0.309%
25	1.949%	1.0782	11.351%	−10.082%	0.680%	11.351%	−10.082%	0.680%

[a]: Only annuity benefit. Initial capital amount (single premium) 1,838.92 euro.
[b]: Annuity benefit and death benefit. The death benefit is given by the Initial capital amount (single premium) net of annuity benefits paid prior to death, if death occurs prior to age 75. Initial capital amount (single premium) 1,937.44 euro.

of the pricing process or the policy design, or more simply of their description, which may be particularly useful in the case of complex products, such as post-retirement ones. We plan to develop further assessments with regard to mortality/longevity-linked annuity benefits, and to extend the assessment in a stochastic framework to get a more comprehensive representation of the possible paths of the embedded fees in different scenarios. We plan also to discuss more in depth the implications of an overestimate of the insurer's profit, arising from the use of proxy information about mortality (or about other aspects not addressed in this paper).

References

1. Blake, D., Cairns, A.J., Kallestrup-Lamb, M., Rangvid, J.: Longevity risk and capital markets: the 2021–22 update. J. Demogr. Econ. **89**(3), 299–312 (2023)
2. Bowers, N.L., Gerber, H.U., Hickman, J.C., Jones, D.A., Nesbitt, C.J.: Actuarial Mathematics. The Society of Actuaries, Schaumburg (1997)
3. Dickson, D.C.M., Hardy, M.R., Waters, H.R.: Actuarial Mathematics for Life Contingent Risks, 3rd edn. Cambridge University Press, Cambridge (2020)
4. Gerber, H.U.: Life Insurance Mathematics. Springer, Heidelberg (1995)
5. Lusardi, A., Mitchell, O.S.: The importance of financial literacy: opening a new field. J. Econ. Perspect. **37**(4), 137–154 (2023)
6. Lusardi, A., Messy, F.A.: The importance of financial literacy and its impact on financial wellbeing. J. Financ. Literacy Wellbeing **1**(1), 1–11 (2023)
7. Olivieri, A., Pitacco, E.: Introduction to Insurance Mathematics. Technical and Financial Features of Risk Transfers, 2nd edn. Springer, Heidelberg (2015)

Multi-model Forecasting for Finance

Daniel Jader Pellattiero(✉) and Antonio Candelieri(✉)

University of Studies of Milano-Bicocca, 20126 Milan, Italy
danieljaderpellattiero@outlook.com, antonio.candelieri@unimib.it

Abstract. This paper presents a novel approach for forecasting stock prices. Specifically, the approach consists of an ensemble of various deep learning models, namely "multi-model". Each deep learning model produces its own forecast, then all the forecasts are combined into a unique one, according to different strategies and depending on different error metrics. The final forecasts provided by the multi-model have resulted in more reliable predictions than those provided by the individual deep learning models.

Keywords: forecasting · stock markets · deep learning · neural networks · recurrent neural networks · wavelet thresholding · CEEMDAN

1 Introduction

Forecasting stock prices is a challenging topic largely addressed via statistical methods as well as Machine Learning and, more recently, Deep Learning. In this paper, we present a forecasting system consisting of an ensemble of different Deep Learning models, from recent literature, that we have named "multi-model". In finance, ensembles of models recently proved to perform better than individual ones [4]. The remainder of this paper is organized as follows: Sect. 2 describes the selection process of the publications on which this paper relies. Section 3 is based on the implementation of the models that compose the multi-model. Section 4 briefly outlines the technologies used. Section 5 is focused on the stock market data used. Section 6 reports the experimental results obtained.

2 Methodological Approach

Each model belonging to the multi-model has been drawn from a pool of publications representing the current state of the art of deep learning techniques in stock forecasting. It has been considered crucial that each model could detect different features starting from the same initial dataset. The desired diversification among the models have been achieved by choosing publications that proposed different processes for feature selection and/or extraction: some models perform

M. Corazza et al. (Eds.): MAF 2024, *Mathematical and Statistical Methods for Actuarial Sciences and Finance*, pp. 248–254, 2024.
https://doi.org/10.1007/978-3-031-64273-9_41

this operation by leveraging specific algorithms, others by employing embedded neural networks for this purpose. The publications selected for the four implemented models are Tsang et al. [5], Cao et al. [1], Livieris et al. [3] and Eapen et al. [2]. Each model has been implemented from scratch following its reference paper and, in case, slightly modified to address specific implementation issues.

3 Models

3.1 Model No. 1: WSAEs-LSTM

The first model implemented was inspired by the work of Tsang et al. [5], in which a deep learning framework is proposed, characterized by: a min-max normalization of the price time series, a denoising phase performed by means of the Wavelet thresholding algorithm, feature extraction mediated by stacked autoencoders, and finally the prediction of the next-day price by means of a recurrent neural network (i.e., a Long Short-Term Memory network).

Data Preparation: Time series normalization is carried out by using the MinMaxScaler class by the Scikit-learn library. The denoising phase is accomplished by using the open-source library named *wtdenoise*, developed by Xie Xinyang[1], which provides the soft thresholding method SURE-Shrink. Wavelet thresholding is performed by choosing the Symlet-8 (sym8) wave as the mother wavelet.

Model Architecture: As already mentioned, two neural networks are employed to produce the next-day price forecast. The first feed-forward type neural network is a stacked autoencoder composed of a total of 7 layers: 3 mirrored between the encoder and decoder plus the bottleneck layer. The three layers of the encoder (decoder) are composed of 19, 15 and 12 units (flipped) respectively, while the bottleneck layer has 10 units. L2-class layer weight regularizers having a penalty factor of $1e^{-2}$ are also applied to the autoencoder layers via the *activity_regularizer* parameter provided by the Keras API, which allows a penalty to be applied to the output of the individual layer. The Exponential Linear Unit (ELU) is chosen as the activation function for each layer. The second recurrent type neural network employed is built on 2 LSTM layers of 10 and 1 units respectively. The hyperbolic tangent is the activation function of the first layer, while the sigmoid function is that of the second. A coefficient of 0.2 is assigned to the dropout and recurrent dropout parameters in both layers. Of the two neural networks mentioned above, the autoencoder is trained first and then split into its two components, the encoder and the decoder: the former is placed before the LSTM layers, whereas the latter is placed after them; thus, the final model is obtained. It is important to remark that having incorporated the recurrent neural network between the encoder and the decoder, and having it trained while keeping the autoencoder components frozen, this has allowed to demonstrate how remarkably the network was able to learn effectively by making predictions directly from the latent representation produced by the encoder.

[1] https://github.com/courageface/wavelet-denoising.

3.2 Model No. 2: CEEMDAN-LSTM

The second model was based on the publication of Cao et al. [1], in which they combine the CEEMDAN (Complete Ensemble Empirical Mode Decomposition with Adaptive Noise) algorithm, used for feature extraction from the historical price series, with a recurrent neural network composed of LSTM layers.

Data Preparation: From the given time series, an arbitrary number of IMFs (Intrinsic Mode Functions) and a residual, respectively denoted as $C_i(t), i \in \{1, \ldots, M\}$ and $R_M(t)$, are obtained by decomposition, using the CEEMDAN algorithm implementation offered by the PyEMD library. These components are then normalized via MinMaxScaler to be given as input to the model.

Model Architecture: The model consists of 3 LSTM layers of 128, 64 and 32 units respectively, having the hyperbolic tangent as activation function and null coefficients for standard and recurrent dropout parameters. Two densely connected layers of 16 and 1 units follow, having an unspecified linear activation function. This structure is replicated for each of the CEEMDAN-generated components, then the next-day price prediction is obtained as follows:

$$\tilde{S}(t) = \sum_{i=1}^{M} \tilde{C}_i(t) + \tilde{R_M}(t),\ t \in \{1, \ldots, L\} \tag{1}$$

with L the length of the original time series, $\tilde{C}_i(t)$ the series predicted by each IMF, $\tilde{R_M}(t)$ the predicted series of the residue and $\tilde{S}(t)$ the final prediction.

3.3 Model No. 3: CNN-LSTM

The third model was developed starting from the publication of Livieris et al. [3]. It proposes a model that comprises a convolutional neural network, aimed at feature extraction over a min-max normalized time series, a recurrent neural network to handle temporal dependencies and finally a densely connected network to provide the final next-day forecast.

Model Architecture: Feature extraction of time series is carried out by means of 2 one-dimensional convolutional layers (Conv1D) of 64 and 128 filters of size (2,) each, having a default *strides* parameter of 1 and the rectified linear unit (ReLU) as activation function. These two layers are followed by a one-dimensional max-pooling layer of analogous size. A 200-unit LSTM layer is placed after the convolutional component, with a hyperbolic tangent as its activation function and null coefficients for standard and recurrent dropout parameters. Eventually there are 2 densely connected layers of 32 and 1 unit respectively, having an unspecified linear activation.

3.4 Model No. 4: MP-CNN-BDLSTM

The fourth model was developed from the publication by Eapen et al. [2], in which an architecture built from three pipelines is proposed. Given the same min-max normalized time series, each pipeline converges to different weights and from the combination of the generated forecasts the next-day price is obtained.

Model Architecture: Each pipeline comes with 2 one-dimensional convolutional layers of 128 and 256 filters of size (2,), with *strides* parameter equal to 1 and the ReLU as activation function; subsequently there is a one-dimensional max-pooling layer with a pool size equal to 2. Consequently, there is a Keras bidirectional layer that encapsulates a 400-unit LSTM layer, having the hyperbolic tangent (sigmoid) as the (recurrent) activation function, with standard and recurrent dropout coefficients of 0.2. Furthermore, there are two densely connected layers of 32 and 1 units, with an unspecified linear activation function trailing the recurrent component. Ultimately there is a common Keras concatenation layer for all the pipelines, which takes the last dense layer of each of them as input and provides its output to a final 1-unit dense layer, also having an unspecified linear activation function.

4 Computational Settings

All the experiments were run on two computers having the following characteristics. To guarantee replicability of the experiments, we made our code available.[2]

- System No. 1: Windows 10 22H2, Intel Core i7-4790K 4.00 GHz, NVIDIA GTX 970 4 GB, 16 GB DDR3 1867 MHz
- System No. 2: Windows 11 22H2, AMD Ryzen 7 4800H 2.90 GHz, NVIDIA RTX 2060 16 GB, 16 GB DDR4 3200 MHz

Code was written in Python 3.11.5 and Deep learning models have been implemented using the Keras framework with Tensorflow 2.0 as its backend.

5 Stock Markets Data

5.1 Data Sources

In order to train and test the models belonging to the multi-model, two datasets were adopted that differ considerably in terms of type, granularity and amount of data; these profound differences are mainly due to their accessibility, as they are open-source on one side, proprietary on the other. The former source of market data was Yahoo! Finance, from which the public OHLCV data of the KOSPI (코스피지수) was drawn, while the latter one was Borsa Italiana S.p.A., which provided us the historical time series of trades occurred in the Limit Order Book of ten financial instruments traded in the Euronext Milan market, respectively: Intesa Sanpaolo, Enel, Ferrari, Campari, Nexi, Prysmian, Leonardo, Azimut, Moncler and Poste Italiane.

[2] https://github.com/danieljaderpellattiero/multi-model-forecasting-for-finance.

5.2 Datasets

For the reserved source of data provided by Borsa Italiana S.p.A., it has been chosen an observation period between 1st September 2022 and 24th March 2023. It has been decided to use only one trading day for each financial instrument, because of the vastness of data generated in the market during intra-day trading. Within the trading day then, three test runs have been crafted based on the trading time slots and split into the training, validation and test set by adopting an 85-10-5% ratio.

Regarding Yahoo! Finance data, it has been decided to craft three test runs with a 70-20-10% ratio, consisting of a 3-year training set, a 10-month validation set and a 5-month test set each. The beginning date of the first test run is 1st January 2015, with a progressive 5-month slippage from that date. It has been decided to observe only the adjusted closing price.

6 Experimental Results

The three strategies to combine individual forecasts via multi-model are briefly summarized.

The first strategy, which can be thought of as a democratic approach among models, consists into averaging the predictions provided by each model.

The second strategy can be considered as a "statically" weighted approach because it performs a weighted average of the forecasts provided by the models, where each model-specific weight depends (inversely) on the error computed on the test set, used to assess the forecasting performances.

The last strategy can be regarded as a "dynamically" weighted approach. It differs from the previous ones because the weight assigned to each model now depends on the input. Specifically, for each new sequence given as input, the most akin sequence within the training and validation set is selected. Once identified, the error metric measured on the model for that specific sequence is taken as the weight for the prediction made by the model on the new input sequence. By doing so, one can estimate the expected precision of the predictions provided by each model of the ensemble, possibly giving higher relevance to the best performing model for that specific input.

Table 1 and Table 2 report four error metrics for the individual models and the three ensemble strategies, on the KOSPI market data and the portfolio of 10 stocks, respectively. It is clear how merging the forecasts, especially when performed considering the performance of the models, yields more accurate results.

Table 1. Prediction error measures (on the test set) for the KOSPI dataset

Model	MAE	MSE	RMSE	MAPE
M1	0.0752	0.0089	0.0944	0.0085
M2	0.0967	0.0116	0.1078	0.0109
M3	0.0700	0.0076	0.0870	0.0079
M4	0.0933	0.0136	0.1168	0.0106
MM_1	0.0685	0.0069	0.0833	0.0078
$MM_{2,MAPE}$	0.0687	0.0071	0.0843	0.0078
MM_3	0.0643	0.0070	0.0837	0.0075

Table 2. Prediction error measures (on the test set) for the dataset consisting of a portfolio of 10 stocks (average and standard deviation over the 10 stocks)

Model	MAE	MSE	RMSE	MAPE
M1	0.1047 (0.1227)	0.0290 (0.0546)	0.1257 (0.1209)	0.0002 (0.0003)
M2	0.0241 (0.0144)	0.0011 (0.0010)	0.0306 (0.0123)	0.0001 (<0.0001)
M3	0.0475 (0.0386)	0.0051 (0.0057)	0.0645 (0.0320)	0.0001 (0.0001)
M4	0.0798 (0.0831)	0.0151 (0.0269)	0.0982 (0.0776)	0.0002 (0.0002)
MM_1	0.0520 (0.0399)	0.0054 (0.0050)	0.0658 (0.0344)	0.0001 (0.0001)
$MM_{2,MAPE}$	0.0230 (0.0182)	0.0015 (0.0014)	0.0352 (0.0157)	0.0001 (<0.0001)
MM_3	0.0194 (0.0130)	0.0010 (0.0008)	0.0292 (0.0111)	<0.0001 (<0.0001)

7 Conclusions

This paper empirically proves that combining different deep learning models, individually and separately trained for predicting stock prices according to different features and algorithms, leads to more robust and precise predictions. Moreover, among the three ensemble strategies proposed, results show that weighted voting methods can reduce prediction error, especially when a "dynamic" (aka input-dependent) weighting schema is adopted.

References

1. Cao, J., Li, Z., Li, J.: Financial time series forecasting model based on CEEMDAN and LSTM. Phys. A: Stat. Mech. Appl. **519**, 127–139 (2019). https://doi.org/10.1016/j.physa.2018.11.061
2. Eapen, J., Bein, D., Verma, A., 2019. Novel deep learning model with CNN and bi-directional LSTM for improved stock market index prediction. In: 2019 IEEE 9th Annual Computing and Communication Workshop and Conference (CCWC), pp. 0264—0270. https://doi.org/10.1109/CCWC.2019.8666592

3. Livieris, I.E., Pintelas, E., Pintelas, P.: A CNN-LSTM model for gold price time-series forecasting. Neural Comput. Appl. **32**, 17351–17360 (2020). https://doi.org/10.1007/s00521-020-04867-x
4. Rather, A.M.: A new method of ensemble learning: case of cryptocurrency price prediction. Knowl. Inf. Syst. **65**(3), 1179–1197 (2023)
5. Tsang, G., Deng, J., Xie, X.: Recurrent neural networks for financial time-series modelling. In: 2018 24th International Conference on Pattern Recognition (ICPR), pp. 892–897 (2018). https://doi.org/10.1109/ICPR.2018.8545666

Using the Gompertz Distribution to Explore the Impact of Increasing Life Expectancy on the Old-Age Dependency Ratio

Peter Pflaumer(✉)

Department of Statistics, Technical University of Dortmund, Dortmund, Germany
peter.pflaumer@tu-dortmund.de

Abstract. The old-age dependency ratios are indicators of the number of elderly people who are generally economically inactive compared to the number of people of working age. They significantly affect the financial burden of social public pension schemes, making it essential to analyze the influence of mortality on this ratio. In this paper, the Gompertz model is used to investigate the effect of mortality and fertility on the old-age dependency ratio, with a focus on the impact of changes in life expectancy. Elasticity formulas are derived to analyze this effect, and the results indicate that an increase in life expectancy leads to a considerable rise in the old-age dependency ratio.

Keywords: Actuarial Models · Gompertz Distribution · Social Security · Life Table

1 Introduction

The old-age dependency ratio is an important demographic indicator that reflects the proportion of elderly people who are not in the labor force and dependent on those who are working. It has become important in analyzing the financial burden of social pension insurance, as it indicates how many potential retirees a potential worker has to support. Its development significantly affects the financial burden of social pension insurance, making it essential to analyze the influence of mortality on the old-age dependency ratio. The Gompertz model is a suitable model for analyzing this influence since it provides a good approximation for low-mortality life tables. According to the United Nations [7], the old-age dependency ratio is projected to increase significantly in the coming decades. By 2050, it is expected to reach 37% globally, meaning that there will be nearly four elderly people for every ten people of working age. This increase is primarily due to the aging of the baby boomer generation and declining birth rates in many countries.

The Gompertz model is a well-known model of demography that was proposed by Benjamin Gompertz [1] in 1825. It states that the mortality intensity exponentially increases with age in adulthood. It has been much applied in life table analysis and in insurance mathematics using various modifications. Due to declining children and youth

M. Corazza et al. (Eds.): MAF 2024, *Mathematical and Statistical Methods for Actuarial Sciences and Finance*, pp. 255–260, 2024.
https://doi.org/10.1007/978-3-031-64273-9_42

mortality, it has again become essential in order to describe "modern" life tables with low mortality. The model allows us to fully describe the present and future life tables in industrialized countries using only two parameters that are easy to estimate from data.

2 Analyzing the Effects of Life Expectancy

An increase in the old-age dependency ratio causes an increase in the premium, when the pensions are constant, or a decrease in the pensions, when the premiums stay constant, all else being equal, assuming, for example, that there is no change in the population growth rate or the age classes used to define the old-age dependency ratio. Therefore, it is important to analytically analyze the influence of mortality on the old-age dependency ratio. This will be done hereafter with the Gompertz model since it provides a good approximation for low-mortality life tables (see Pollard [6]). Especially the effect of changes in the life expectancy on the ratio is analyzed. An increase in the life expectancy will undoubtedly raise the dependency ratio, but by how much?

The old-age dependency ratio is a demographic indicator that measures the number of elderly individuals (as defined by us, aged 60 and older) relative to the working-age population (as defined by us, those aged 20 to 60); often we find the age 65 instead of 60:

$$OADR = \frac{\int_{60}^{\infty} l(x)dx}{\int_{20}^{60} l(x)dx} = \frac{\int_{60}^{\infty} \exp\left(-e^{k\cdot(x-m)}\right)dx}{\int_{20}^{60} \exp\left(-e^{k\cdot(x-m)}\right)dx} = \frac{e_{60}\cdot l_{60}}{e_{20}\cdot l_{20} - e_{60}\cdot l_{60}} = \frac{1}{\frac{e_{20}\cdot l_{20}}{e_{60}\cdot l_{60}} - 1},$$

where $l(x) = \exp\left(e^{-k\cdot m} - e^{k\cdot(x-m)}\right) \approx \exp\left(-e^{k\cdot(x-m)}\right)$ is the survivor function of the Gompertz distribution with m>>k>0; m is the modal value and k is the growth rate of the exponential force of mortality function. For a detailed presentation of the Gompertz distribution, see, e.g., Pollard [5, 6], Carriere [2, 3].

The mean or life expectancy at birth is

$\mu = e_0 = m - \frac{\gamma}{k}$ with $\gamma = 0.57722....$ The Euler-Mascheroni constant.

The variance is $\sigma^2 \approx \frac{\pi^2}{6} \cdot \frac{1}{k^2}$.

Thus, the reciprocal value of k can be regarded as a dispersion parameter.

Typical values for low mortality life tables fall generally within the range of 85 to 90 for m, and 0.09 to 0.11 for k. For instance, using the German life table 2019/2021 for females with a life expectancy of 83.4 years, a fit with the Gompertz distribution yields values of m = 89.04 and k = 0.1143. This results in an estimated life expectancy of 83.95 years. The use of only two parameters of the Gompertz distribution is sufficient to describe the entire life table and obtain good approximations for the life table parameters of empirical tables.

The life expectancy at age x can be approximated by.

$e(x) = -\frac{\frac{\gamma+k\cdot(x-m)-\exp(k\cdot(x-m))}{k}}{\exp\left(e^{-k\cdot m}-e^{k\cdot(x-m)}\right)}$ (cf. Carriere [2, 3]).

Substituting the values of the life expectancy and the survivor function at age x in the transformed formula of the old-age dependency ratio, provides a good approximation of

the ratio when the modal age m is greater than 70:

$$OADR_1 \approx \frac{1}{\frac{\gamma - k\cdot(m-20) - \exp(-k\cdot(m-20))}{\gamma - k\cdot(m-60) - \exp(-k\cdot(m-60))} - 1}.$$

If the modal value m is still increasing, then the ratio finally tends to.

$$OADR_{hat} = \frac{1}{\frac{\gamma - k\cdot(m-20)}{\gamma - k\cdot(m-60)} - 1} = \frac{m - \frac{\gamma}{k} - 60}{40} = \frac{e_0 - 60}{40}.$$

Figure 1 displays a 3-dimensional plot of the OADRs, which depend on both k and e(0). The OADR is primarily determined by life expectancy and is less affected by k. An increase in life expectancy leads to a considerable rise in OADR, while increasing k only slightly decreases OADR.

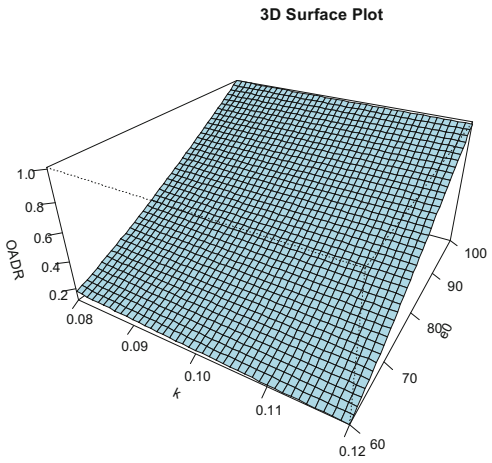


Fig. 1. Surface plot of OADR as a function of e(0) and k

Table 1 presents the old-age dependency ratios for a fixed value of k = 0.1, calculated through numerical integration. The difference between the exact and approximate values is negligible when the modal ages are high. Moreover, even the simple approximation formula provides satisfactory results for very old ages.

To analyze the influence of a change in the life expectancy on the old-age dependency ratio, elasticities are computed. Elasticity is the ratio of the percent change in one variable to the percent change in another variable. Mathematically, elasticity is defined as

$$\varepsilon(OADR, e_0) = \frac{dOADR}{de_0} \cdot \frac{e_0}{OADR}.$$

Using the above approximation formula leads to a rather complicated formula

$$\varepsilon_1(OADR_1, e_0) =$$
$$= \frac{k \cdot e^{\gamma + e_0 \cdot k}\left(40 \cdot k \cdot e^{\gamma + e_0 \cdot k} - e^{20 \cdot k}\left(e^{40 \cdot k}(e_0 \cdot k - 20k + 1) - e_0 \cdot k + 60k - 1\right)\right)}{\left(40 \cdot k \cdot e^{\gamma + e_0 \cdot k} + e^{20 \cdot k}\left(1 - e^{40 \cdot k}\right)\right) \cdot \left(k \cdot e^{\gamma + e_0 \cdot k}(e_0 - 60) + e^{60 \cdot k}\right)} \cdot e_0$$

Table 1. Old-age dependency ratios and elasticities: Exact and approximate values (k = 0.1)

e0	m	OADR	OADR1	OADRhat	Elast	Elasthat
65	70.8	0.220	0.229	0.125	5.18	13
70	75.8	0.314	0.318	0.25	4.42	7
75	80.8	0.418	0.419	0.375	3.86	5
80	85.8	0.528	0.529	0.5	3.43	4
85	90.8	0.644	0.644	0.625	3.09	3.4
90	95.8	0.762	0.762	0.75	2.83	3
95	100.8	0.883	0.883	0.875	2.62	2.71
100	105.8	1.005	1.005	1	2.44	2.5

where $e_0 = m - \frac{\gamma}{k}$.

An easy-to-use elasticity formula is obtained using the simple approximation

$$\varepsilon_{hat}(OADR_{hat}, e_0) = \frac{d\left(\frac{e_0-60}{40}\right)}{de_0} \cdot \frac{e_0}{\frac{e_0-60}{40}} = \frac{e_0}{e_0 - 60}.$$

The elasticity provides insight into the proportional change in the old-age dependency ratio (OADR) in response to a 1 percent increase in the life expectancy. Specifically, at age 85 (e(0) = 85), it indicates that such an increase in life expectancy leads to approximately a 3 percent rise in the OADR, as shown in Table 1.

Figure 2 and Table 1 illustrate the elasticities (elast) of the old-age dependency ratio (OADR) as a function of life expectancy.

3 Stable Populations and Low Fertility

A population with an unchanging age structure and a fixed rate of increase is called a stable population (see, e.g., Keyfitz [4]). To determine the old-age dependency ratio in a stable population with a growth rate of r, one must compute

$$OADR(r) = \frac{\int_{60}^{\infty} e^{-r \cdot x} l(x) dx}{\int_{20}^{60} e^{-r \cdot x} l(x) dx}.$$

This expression can be approximated based on Keyfitz [4] by

$OADR(r) \approx OADR(0) \cdot e^{-T \cdot r}$, where T is the difference between the mean age of the two generations in the age classes 20 to 60 and 60 to ω.

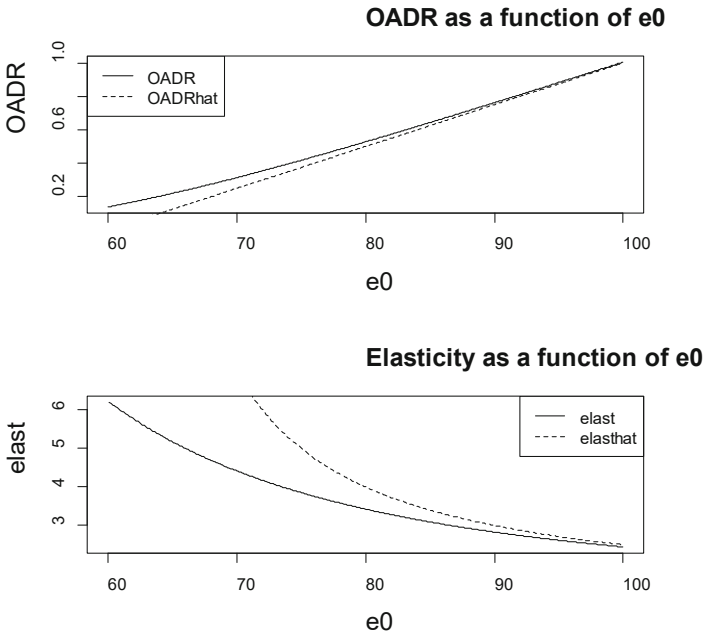


Fig. 2. OADR and Elasticity as a function of e(0) with k = 0.1; dotted lines represent approximations with the simple formulas

Using the simple approximation of the old-age dependency ratio, for example, results in the following elasticity formula:

$$\varepsilon(OADR, e_0) \approx \frac{e_0}{e_0 - 60} \cdot e^{-T \cdot r}.$$

The elasticities are now dependent on the growth rate r and the life expectancy e(0). In a stable population with a positive (negative) growth rate, the elasticities are lower (higher) than in a stationary population by the factor $e^{-T \cdot r}$. For example, at life expectancy e(0) = 85, the elasticity is initially 3.09 (see Table 1). If we consider a negative growth rate due to low fertility of -1% and a generation difference of 30 years (T = 30), the factor $e^{-T \cdot r}$ is approximately 1.35, resulting in an increased elasticity of about 4 under these specific demographic conditions and generation difference.

4 Conclusion

The paper emphasizes the importance of studying the profound impact of an aging population on social welfare policies and the economy. A mere 1% increase in life expectancy translates to approximately a 3% rise in the old-age dependency ratio within a stationary population exhibiting low mortality. This effect is even more pronounced in stable populations with low fertility rates, falling below replacement levels. As the population ages persistently, it is imperative for policymakers to effectively address

the requirements of both the elderly and the younger generations, thereby fostering sustainable economic growth and upholding social welfare. Neglecting this balance could entail consequential economic and social repercussions for generations to come.

References

1. Gompertz, B.: On the nature of the function expressive of the law of human mortality. Philos. Trans. Royal Soc. London Ser. A **115**, 513–585 (1825)
2. Carriere, J.F.: Parametric models for life tables. Trans. Soc. Actuaries **44**, 77–99 (1992)
3. Carriere, J.F.: An investigation of the Gompertz law of mortality. Actuarial Res. Clearing House **2**, 161–177 (1994)
4. Keyfitz, N.: Applied Mathematical Demography, 2nd edn. Wiley, New York (1977)
5. Pollard, J.: An old tool – modern applications. Actuarial Studies and Demography. Research Paper No. 001/98 (1998)
6. Pollard, J.H.: Fun with Gompertz. Genus **XLVII** (n.1–2), 1–20 (1991)
7. United Nations, Department of Economic and Social Affairs, Population Division: World Population Prospects 2019: Highlights (ST/ESA/SER.A/423). New York (2019). https://population.un.org/wpp/Publications/Files/WPP2019_10KeyFindings.pdf

Challenges in Cyber Risk Insurance

Marco Pirra(✉)

University of Calabria, Rende, CS, Italy
marco.pirra@unical.it

Abstract. This paper illustrates the main challenges identified in the literature on cyber risk assessment in order to provide possible solutions for the reduction of the gap between supply and demand of cyber insurance. The aim is to contribute to a better understanding in quantifying, managing and pricing cyber risk by means of: a deeper awareness of the economic consequences they produce; the introduction and validation of new actuarial techniques to allow insurers a more efficient management of this class of risks; the design of innovative insurance contracts and alternative ways of risk transfers to reduce the costs of insurance premiums and mitigate the economic impacts.

Keywords: cyber insurance · risk management · parametric contracts

1 Introduction

The cyber insurance market, despite its growth and increasing relevance, faces several limitations and challenges. These limitations stem from the nature of cyber risks, market dynamics and the evolving landscape of technology and regulations. Explicitly:

- rapidly evolving cyber threats: one of the most significant challenges is the rapid evolution of cyber threats. Unlike traditional insurance domains, where risks are relatively stable and predictable over time, cyber risks change frequently, often outpacing the ability of insurers to understand and price them effectively. This unpredictability, as highlighted in [13], makes it difficult for insurers to assess risk accurately and set premiums.
- historical data: the cyber insurance industry suffers from a lack of extensive historical data, which is crucial for traditional actuarial models. This limitation, as discussed in [6], makes it challenging to predict the frequency and severity of cyber incidents accurately. As cyber threats continuously evolve, historical data might not accurately represent future risks and the need for dynamic models that can adapt to new trends and threats emerges.
- aggregation risk and systemic events: the interconnectedness of technology means that a single cyber event can impact multiple entities simultaneously. This aggregation risk, as explored in [23,25], poses a significant challenge for insurers who may face massive, correlated claims. The potential for systemic events, such as widespread ransomware attacks, exacerbates this risk.

M. Corazza et al. (Eds.): MAF 2024, *Mathematical and Statistical Methods for Actuarial Sciences and Finance*, pp. 261–266, 2024.
https://doi.org/10.1007/978-3-031-64273-9_43

- coverage ambiguities and policy standardization: the cyber insurance market is characterized by a lack of standardization in policy terms and coverage. As pointed out in [16,22], this results in ambiguities and misunderstandings about what is covered, leading to disputes during claim settlements. The diversity in policy wordings and coverage limits complicates the understanding for both insurers and insured.
- regulatory and legal challenges: the regulatory landscape for cyber risks is constantly evolving. Differences in regional and national regulations, as analyzed in [17], create additional complexities for insurers who must navigate varying compliance requirements across jurisdictions. Moreover legal precedents in cyber insurance claims are still developing, adding to the uncertainty.
- affordability and coverage limits: the rising costs of premiums, partly due to the increasing frequency and severity of cyber attacks, are making cyber insurance less affordable for small and medium-sized enterprises (SMEs). Additionally, insurers often impose strict coverage limits to mitigate their risk exposure, as observed in [3], which may result in inadequate coverage for significant cyber events.
- quantifying intangible losses: cyber incidents often result in intangible losses like reputational damage or loss of customer trust, which are difficult to quantify. Actuarial models in this field must therefore incorporate both tangible losses, like data recovery costs, and intangible losses, as discussed in [11]. They propose methodologies for estimating the financial impact of these non-physical damages
- emergence of artificial intelligence and machine learning: recent advancements in artificial intelligence (AI) and machine learning (ML) are being leveraged to enhance cyber risk actuarial models. These technologies allow for the analysis of large datasets and identification of patterns and correlations that traditional methods might miss. [14] provides insights into how AI and ML are revolutionizing this field, allowing for more accurate and up-to-date risk assessments.
- cybersecurity complacency: there is a concern that the availability of cyber insurance might lead to complacency in cybersecurity practices among insured entities. This moral hazard, as discussed in the literature (i.e. see [18]), suggests that companies might invest less in proactive cybersecurity measures, relying instead on insurance to cover potential losses.

While cyber insurance is a vital tool in managing cyber risk, the market faces significant challenges in terms of risk assessment, policy standardization, regulatory compliance, and the ever-evolving nature of cyber threats. A range of actuarial models have been proposed to address the complex and dynamic nature of cyber risk. Zeller, [24], and Herath, [12], both present innovative models, with Zeller focusing on marked point processes and Herath on copula methodology. Bhme, [5], and Eling, [8], emphasize the need for a differentiated view of cyber risk and the importance of human behavior in driving this risk. Liu, [15], and Awiszus, [2], further explore the interdependence of cyber risks and the need for more sophisticated modeling approaches. Eling,[9], and Avanzi, [1], highlight the

challenges of data scarcity and the need for actuarial principles in valuing cyber insurance contracts. These studies collectively underscore the need for robust and flexible actuarial models to effectively manage cyber risk. This work contributes to the literature on the management of cyber risks proposing a possible solution for the reduction of the gap between supply and demand of cyber insurance. The rest of the paper is organized as follows: Sect. 2 illustrates the framework and the methodology; Sect. 3 discusses a case study and draws conclusions.

2 Framework and Approach Proposed

This section outlines the key elements of a framework intended as a possible way of mitigating cyber risks: zero-inflated models to capture the dynamics of the events and parametric contracts to attract the demand of protection.

Zero-Inflated Models. Zero-inflated models, such as the zero-inflated Poisson distribution ([21]), and zero-inflated negative binomial models ([20]), have been widely used in various fields to address the issue of excessive zeros in count data. These models have been applied to diverse areas, including environmental studies, software fault prediction, and product quality monitoring. The use of these models has been shown to improve predictive quality and provide valuable insights into underlying processes. Furthermore, the development of new models, such as the zero-inflated Hermite distribution and the bivariate zero-inflated negative binomial regression model, [21], has expanded the range of applications for zero-inflated models. To the best of our knowledge these models have never been used in actuarial applications. Zero-inflation in the cyber risk assessment has the purpose of capturing the fact that, in some cases, the institution victim of the cyber attack does not report the event. To better capture the dynamics of data breaches and to improve loss estimates and risk pricing the baseline model proposed in [7] has been extended by adding some Markov-Switching features. The calibration of the model has been performed on data provided by the Privacy Rights Clearinghouse (PRC), [19]. The PRC is the largest and most extensive public dataset, that includes information on total breached records, location and date of the incident, the entity level and an incident description. The breach size has been transformed into a loss amount in accordance with [10]. A second dataset was obtained from the Breach Level Index Data Breach Database a centralized, global database of data breaches with calculations of their severity based on multiple factors.

Parametric Insurance. In accordance with [4], a paper in which the authors emphasize the importance of resilience in cyber risk management, advocating for strategies that not only prevent attacks but also ensure quick recovery and minimal impact in the event of a breach, a possible tool intriguing in its application to cyber risk management is parametric insurance, an innovative approach in the insurance industry, that offers a distinct mechanism for risk transfer. Unlike traditional indemnity-based insurance that compensates for actual losses

incurred, parametric insurance operates on pre-agreed triggers, such as the magnitude of a cyber-attack or specific threat indicators. This model provides a more rapid and objective claim settlement process, reducing the lengthy and often complex assessment procedures typical in conventional cyber insurance policies. The potential for applying parametric insurance to cyber risk lies in its ability to offer swift financial support post-incident, crucial for businesses in mitigating the immediate fallout of a cyber-attack. This is particularly relevant in scenarios like widespread data breaches or DDoS attacks, where the quantification of losses can be challenging and time-sensitive. Additionally, parametric insurance could align well with industries that are highly susceptible to specific, measurable cyber threats, providing a tailored risk management tool. The simplicity and speed of this model not only enhance its attractiveness for businesses seeking clear, straightforward coverage but also encourage proactive risk management practices. However, the success of parametric insurance in the cyber domain hinges on the accurate identification and calibration of triggers, a task that requires deep cybersecurity expertise and robust data analytics capabilities. Risk managers who seek to take control of their risk can use parametric insurance to make certain uninsurable risks insurable and complement their traditional indemnity programs. Parametric insurance can help organizations match insurance capital to their specific risk profiles, providing more liquidity exactly when they may need it the most. The parameter or index that acts as a proxy for the actual sustained loss is an important determinant for the success of a parametric product. Any mismatch between the actual loss and parametric payout that arises due to issues in the parametric structure is termed as basis risk. The presence of basis risk reduces the effectiveness of parametric insurance as a risk management tool, creates a perception of parametric insurance being a poor substitute for indemnity insurance, and prevents its widespread adoption. A possible insurance payout could be based on a standard indemnity per lost or stolen record, as follows:

$$I = f\left(i_N \mid x, \mathrm{Tr}, Ex\right) = x \times f(x) = \begin{cases} 1 \text{ if } i_N \leq \mathrm{Tr} \\ \frac{i_N - \mathrm{Tr}}{Ex - \mathrm{Tr}} \text{ if } \mathrm{Tr} < i_N \leq Ex \\ 0 \text{ if } i_N \geq Ex \end{cases} \tag{1}$$

where i_N represents the realised value of the underlying parameter (i.e. number of compromised records), x is the sum insured, Tr is the trigger and Ex is the exit below wich indeminity seizes. The value of the indemnity can be parametrized to the size of the breach in order to mitigate moral hazards.

3 Practical Implications and Conclusions

Some preliminary results are summarised in Table 1, that reports relevant risk measures for different scenarios: case A constant indemnity, case B indemnity decreases as the size of the breach increases, case C decreasing indemnity, trigger checked per month (annually in A and B).

Table 1. Risk Measures Quantification, values in USD

	Expected Loss	StDev	Min	Max	VaR99.5%	ES99.5%
Case A	1,517,108	383,963	709,846	6,681,873	3,491,097	4,468,927
Case B	655,883	130,487	359,246	2,342,271	1,305,110	1,636,666
Case C	517,455	43,299	354,915	665,156	615,633	625,786

The evidence of the numbers indicate that whatever the principle is the one used to compute an insurance premium, either considering the standard deviation or other risk measures, the parametric coverage leads to a reduction in the costs and therefore an offer of more affordable prices. Determining accurate parametric triggers and establishing appropriate payout structures require careful calibration and further analyses and investigation, taking into account the complex and ever-evolving cyber threat landscape to ensure the models remain effective. The payout proposed can be considered as a first responder for small and medium size entities against cashflow shortages and reduced revenue immediately following a cyber event. It can contribute to the reduction of the gap between supply and demand of cyber insurance.

Acknowledgments. *This work has been partially supported by the AFIR-ERM Section of the International Actuarial Association, research grant "Cyber and technology risk: Managing exposures, possibilities, and pricing of new cyber risk insurance products".*

References

1. Avanzi, B., Tan, X., Taylor, G., Wong, B. Cyber insurance risk: reporting delays, third-party cyber events, and changes in reporting propensity–an analysis using data breaches published by US State Attorneys General. arXiv preprint arXiv:2310.04786 (2023)
2. Awiszus, K., Knispel, T., Penner, I., Svindland, G., Voß, A., Weber, S.: Modeling and pricing cyber insurance: idiosyncratic, systematic, and systemic risks. Eur. Actuar. J. 1–53 (2023)
3. Baker, T., Shortland, A.: Insurance and enterprise: cyber insurance for ransomware. Geneva Pap. Risk Insur.-Issues Pract. **48**, 275–99 (2023)
4. Bodeau, D., Graubart, R.: Cyber resiliency design principles. The MITRE Corporation (2017)
5. Böhme, R., Laube, S., Riek, M.: A fundamental approach to cyber risk analysis. Variance **12**(2), 161–185 (2019)
6. Cremer, F., Sheehan, B., Fortmann, M., et al.: Cyber risk and cybersecurity: a systematic review of data availability. Geneva Pap. Risk Insur. **47**(3), 1–39 (2022)
7. De Giovanni, D., Leccadito, A., Pirra, M.: On the determinants of data breaches: a cointegration analysis. Decis. Econ. Finan. **44**, 141–160 (2021)
8. Eling, M., Wirfs, J. H. Modelling and management of cyber risk. In: International Actuarial Association Life Section (2015)

9. Eling, M.: Cyber risk research in business and actuarial science. Eur. Actuar. J. **10**(2), 303–333 (2020)
10. Farkas, S., Lopez, O., Thomas, M.: Cyber claim analysis using generalized Pareto regression trees with applications to insurance. Insur. Math. Econ. **98**, 92–105 (2021)
11. Gatzert, N., Schubert, M.: Cyber risk management in the US banking and insurance industry: a textual and empirical analysis of determinants and value. J. Risk Insur. **89**(3), 725–763 (2022)
12. Herath, H.S., Herath, T.C.: Copula-based actuarial model for pricing cyber-insurance policies. Insur. Mark. Compan. **2**(1), 7–20 (2011)
13. Javaheri, D., Fahmideh, M., Chizari, H., Lalbakhsh, P., Hur, J.: Cybersecurity threats in FinTech: a systematic review. Expert Syst. Appl. 122697 (2023)
14. Jena, T., Shankar, A., Singhdeo, A.: Harnessing machine learning for effective cyber security classifiers. Asian J. Res. Comput. Sci. **16**(4), 453–464 (2023)
15. Liu, Z., Wei, W., Wang, L., Ten, C.W., Rho, Y.: An actuarial framework for power system reliability considering cybersecurity threats. IEEE Trans. Power Syst. **36**(2), 851–864 (2020)
16. Marotta, A., Martinelli, F., Nanni, S., Orlando, A., Yautsiukhin, A.: Cyber-insurance survey. Comput. Sci. Rev. **24**, 35–61 (2017)
17. Meagher, H., Dhirani, L.L.: Cyber-resilience, principles, and practices. In: Cybersecurity Vigilance and Security Engineering of Internet of Everything, pp. 57–74. Springer, Cham (2023)
18. Nwankpa, J.K., Datta, P.M.: Remote vigilance: the roles of cyber awareness and cybersecurity policies among remote workers. Comput. Secur. **130**, 103266 (2023)
19. Privacy Rights Clearinghouse Data breach chronology database. https://privacyrights.myshopify.com/products/data-breach-chronology-data-set. Accessed Dec 2023
20. Phang, Y., Loh, E.: Zero inflated models for overdispersed count data. Int. J. Health Med. Eng. **7**(8), 1331–1333 (2013)
21. Satheesh Kumar, C., Ramachandran, R.: On some aspects of a zero-inflated overdispersed model and its applications. J. Appl. Stat. **47**(3), 506–523 (2020)
22. Tsohou, A., Diamantopoulou, V., Gritzalis, S., et al.: Cyber insurance: state of the art, trends and future directions. Int. J. Inf. Secur. **22**, 737–748 (2023)
23. Welburn, J.W., Strong, A.M.: Systemic cyber risk and aggregate impacts. Risk Anal. **42**(8), 1606–1622 (2022). PMID: 33594708
24. Zeller, G., Scherer, M.: A comprehensive model for cyber risk based on marked point processes and its application to insurance. Eur. Actuar. J. **12**, 33–85 (2022)
25. Zeller, G., Scherer, M.: Is accumulation risk in cyber systematically underestimated? SSRN 4353098 (2023)

Identifying Graphical Configurations in Technical Analysis Using Machine Learning

Claudio Pizzi[1(✉)] and Matteo Munini[2]

[1] Ca' Foscari University of Venice, Venice, Italy
pizzic@unive.it
[2] Venice, Italy

Abstract. In this paper, we enhance Leigh's procedure (Leigh 2002a) for identifying the bull-flag configuration within a specified timeframe. We improve the template construction method by introducing more flexibility and eliminating predefined weight choices. The optimization of parameters to maximize annualized return is achieved using a modified fireworks algorithm. Additionally, in our approach, we introduce a signal generator to enhance model robustness and account for trader risk attitudes. The approach is flexible, as it is suitable for other graphical configurations. Our proposal demonstrates superior performance compared to standard settings.

Keywords: Technical analysis · Bull-flag configuration · Fireworks optimization

1 Introduction

Forecasting the stock market has always intrigued traders, investors, and researchers. Technical analysis is an approach employed to analyze and understand the financial market, seeking trading opportunities. Practitioners in this discipline scrutinize financial time series to identify well-known patterns and signals, utilizing various technical indicators and graphical configurations. In contrast to the weak form of the Efficient Market Hypothesis, which posits the impossibility of predicting the market solely based on price features, technical analysis studies past data, such as price and volume, to forecast future price variations. This paper proposes a trading system based on 'charting,' a subset of technical analysis focusing on analyzing charts for specific patterns in price time series. The method presented centers on the identification of the bull-flag pattern, consisting of a consolidation part representing a downtrend and a breakout capturing the shift in the price's trend. A durable uptrend period is anticipated when a bull-flag configuration is formed, and the trading strategy leverages this

M. Munini—Independent author.

M. Corazza et al. (Eds.): MAF 2024, *Mathematical and Statistical Methods for Actuarial Sciences and Finance*, pp. 267–273, 2024.
https://doi.org/10.1007/978-3-031-64273-9_44

pattern to execute 'buy-sell' operations. The paper demonstrates the profitability of this approach.

The remainder of this paper is organized as follows. The next section is devoted to a short review of literature and Sect. 3 describes the methodology used in this work. Section 4 presents the application of our proposal on a stock market index. Some final remarks conclude the paper.

2 Literature Review

The papers [2] and [3] introduced a template-based matching method for detecting the bull-flag configuration. A bull flag is a chart pattern resembling a flag shape with the masts on either side, indicating consolidation within a trend. It forms from price fluctuations in a narrow range, preceding and following sharp rises or declines. The pattern consists of a consolidation (flag) followed by a breakout, representing a sharp upward movement. The method involves using a k by k matrix W of weights such that $w_{i,j} \leq 1$, where an appropriate assignment of weights enables the representation of the bull-flag pattern. Leigh used $k = 10$ but is possible consider different value for k. However, the authors did not provide a method for defining the weights, introducing subjectivity into the process. Nevertheless, the skewed-v-shape of the bull-flag can be accurately represented by appropriately assigning unit weights, as illustrated in Fig. 1a [2].

1	0	-1	-1	-1	-1	-1	-1	-1	0
1	1	0	-1	-1	-1	-1	-1	-1	0
1	1	1	0	-1	-1	-1	-1	0	1
1	1	1	1	0	-1	-1	-1	0	1
0	1	1	1	1	0	0	0	1	1
0	0	1	1	1	1	0	0	1	1
-1	0	0	1	1	1	1	1	1	1
-1	-1	0	0	1	1	1	1	1	0
-1	-1	-1	-1	0	1	1	1	0	-2
-1	-1	-1	-1	-1	0	1	1	-2	-3
Consolidation								Breakout	

a

	col.1	col.2	col.3	col.4	col.5	col.6	col.7	col.8	col.9	col.10
row.1	α_1	α_2	α_3	α_4	α_5	α_6	α_7	β_7	β_5	β_3
row.2	1	α_1	α_2	α_3	α_4	α_5	α_6	β_6	β_4	β_2
row.3	1	1	α_1	α_2	α_3	α_4	α_5	β_5	β_3	β_1
row.4	α_1	1	1	α_1	α_2	α_3	α_4	β_4	β_2	1
row.5	α_2	α_1	1	1	α_1	α_2	α_3	β_3	β_1	1
row.6	α_3	α_2	α_1	1	1	α_1	α_2	β_2	1	1
row.7	α_4	α_3	α_2	α_1	1	1	α_1	β_1	1	1
row.8	α_5	α_4	α_3	α_2	α_1	1	1	1	1	β_1
row.9	α_6	α_5	α_4	α_3	α_2	α_1	1	1	β_1	β_2
row.10	α_7	α_6	α_5	α_4	α_3	α_2	α_1	β_1	β_2	β_3
	Consolidation							Breakout		

b

Fig. 1. Bull Flag a) template presented in Leigh et al. (2002a), b) our proposal

To compute an index measuring the similarity between time series and the bull-flag pattern, a moving window of width (dw) is chosen, representing the number of observations. The dw observations are divided into k equal subintervals, maintaining chronological order. A k by k data matrix D is then defined, where each column represents one time sub-interval, and the rows represent price level intervals. For each cell in the data matrix, the number of observations falling into it is counted and divided by the number of observations

assigned to the corresponding column, resulting in a data matrix filled with real numbers in the range [0, 1]. The similarity index is computed using the formula:

$$_{dw}Fit_i = \sum_{c=1}^{k}\sum_{r=1}^{k} W_{rc}^{(i)} * D_{rc}^{(i)} \qquad i = dw+1, \ldots, t$$

where t is the length of the time series, dw is the width of the moving window, and i identifies the last observation in the moving window.

A buy signal may be generated when the similarity index exceed an $(1-\alpha)$ quantile [2] and [3], or when it is greater than a threshold value [6]. The sell signal is generated after a constant interval of fh days. In recent papers [1,5] various template grids for the bull-flag pattern along with new trading rules were introduced. Additionally, Wang and Chan [6] formalized a different method for computing weights in the template matrix. However, a common drawback among these methods is the subjectivity of the template, specifically the position of the 1 s and the weights within the matrix. Furthermore, the values of dw and fh have not been optimized.

3 Methodology

3.1 The Template Matrix

The aim of this paper is to provide more flexibility to the Leigh approach, allowing the weights to be chosen through a data-driven approach. The starting point is the shape of the bull-flag graphical configuration proposed by Leigh et al. [2], which is represented by the 1 s in Fig. 1a. For the remaining cells, the weights must be estimated. We consider two sets of parameters, denoted as α and β (Fig. 1b). α is the vector of weights for the consolidation part of the bull-flag, while β is the vector for the breakout part.

We need different αs and βs to penalize in a different way the cells that are far from those containing 1; theoretically, these weights must satisfy the following conditions:

$$1 > \alpha_1 > \ldots > \alpha_k > \ldots > \alpha_9$$
$$1 > \beta_1 > \ldots > \beta_k > \ldots > \beta_9$$

Following the approach in [6] a signal is generated when the threshold is exceeded. Since the bull-flag configuration may be confused with a down-trend pattern, to minimize false signals, we combine information from two similarity indices. The first is calculated with respect to the bull-flag pattern, and the second is computed with respect to the down-trend pattern:

$$\begin{cases} {}_kFit_{(bf)} > ths_{bf} & \text{with} \quad ths_{bf} \in \Re \\ {}_kFit_{(bf)} > \ {}_kFit_{dt} \cdot ths_{dt} & \text{with} \quad ths_{dt} > 1 \end{cases}$$

where ths_{bf} is the bull-flag threshold and ths_{dt} is the down-trend threshold. The signal is generated if and only if both conditions are met.

The methodology introduced in this section depends on several parameters: weights (αs, βs), thresholds (ths_{bl} and ths_{dt}), window width (dw) and forecast horizon (fh). Different combinations of parameters generate different returns, so the problem is how estimate all the parameters to obtain the best performance from the trading system. The problem could not be solved analytically and we estimate the optimal set of parameters by the fireworks optimisation algorithm [4]. The trading system's performance can be assessed using various metrics. In this paper, we focus on evaluating the return at the conclusion of the trading period.

3.2 The FireWorks Algorithm

The Fireworks Algorithm (FA) is an optimisation procedure that searches for the best solution by mimicking the behaviour of fireworks. The explosions of fireworks, along with the sparks they generate, allow the algorithm to explore and exploit the solution space effectively.

The implemented Fireworks Algorithm in this paper is as follows:

1. Choose N starting positions in the parameters space [4].
2. M sparks are generated at each of N locations. Each spark defines a set of parameter values, allowing for the evaluation of the objective function.
3. Evaluate all the sparks, and the best N are retained as new locations. Repeat steps 2–3 for t times.
4. Generate M sparks at N locations by increasing the amplitude of the explosion compared to the previous step, allowing the algorithm to explore the parametric space extensively. Return to step 3
5. Repeat the step 4 for k times.
6. From the final N location select the best.

Usually the FA is used to find the minimum of a function, so in our case the objective function used by the optimizer is:

$$-R(\alpha, \beta, dw, fh, ths_{bl}, ths_{dt}) + P(\alpha, \beta) \cdot \partial$$

where the function R is the cumulative trading gain at the end of the trading period, while P is a penalization function for the constraints not respected by the solution of algorithm and ∂ is a multiplier chosen by the operator:

$$P(\alpha, \beta) = \sum_{i=1}^{7} I_{\alpha_i} \rho_{\alpha_i} + 2 \cdot \sum_{i=1}^{8} I_{\beta_i} \rho_{\beta_i}$$

where ρ is the percentage of a weights in the bull-flag template that belong to the same set (α or β). I_i is the number of constraints violation. A characteristic of the FA is that it allows to define a lower and upper bound for each parameter. In this way, the parametrization of the template can be controlled, and the model is flexible to the trader's needs or will.

4 Application and Results

The model was tested on various financial time series, including both market indices and stocks. In this section, we present, for brevity, only the results for a market index, namely the NASDAQ Composite index (IXIC). However, similar results were obtained for other market indices and stocks.

Table 1. Rolling window scheme used in the application.

	Training set	Operative Window
1^{st} operation	From 1^{st} day to 2800^{th}	From 2801^{st} day to 3000^{th}
2^{st} operation	From 201^{st} day to 3200^{th}	From 3201^{st} day to 3400^{th}
3^{st} operation	From 401^{st} day to 3400^{th}	From 3401^{st} day to 3600^{th}
...	...	...

Regarding FA settings, we opted for wide parameter intervals to allow the model flexibility based on the data. No prior assumptions or operation strategies were imposed, except for the constraint on the relationship between the alpha and beta parameters, as previously explained.

We analyzed the daily closing quotes of the NASDAQ Composite index spanning from March 2, 2006, to April 22, 2023. Employing a rolling window with a width of 3000 observations, which are divided into two parts: the training set, consisting of the first 2800 observations, and the validation set, comprising the last 200 observations, as illustrated in Table 1. The model undergoes re-training for each time window, allowing it to operate out-of-sample from April 17, 2017, to March 22, 2023. Buy-sell operations on IXIC are represented in Fig. 2a, while cumulative returns are depicted in Figure 2b.

The gross return at the end of the out-of-sample period is approximately 45%. Notably, the model exhibited robustness in its buy-sell decisions, effectively mitigating losses during the global crisis that unfolded between 2021 and 2023.

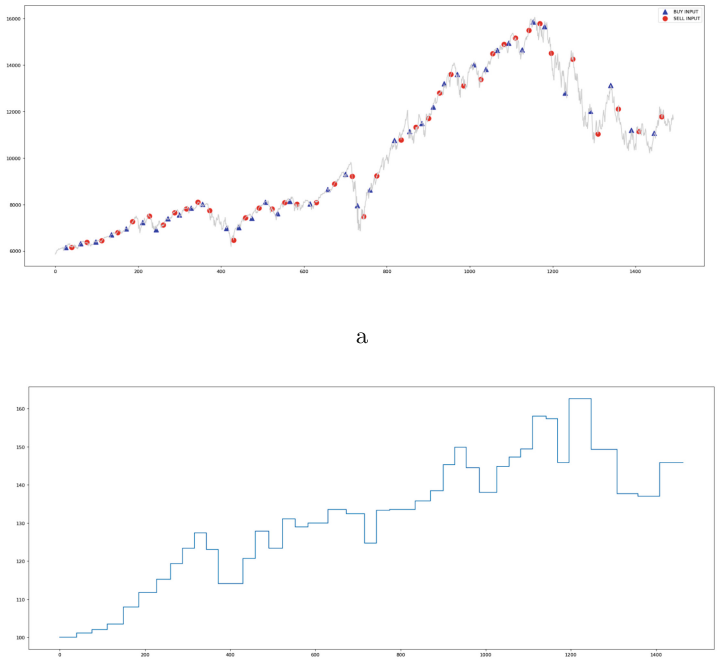


Fig. 2. a) Operations *out-of-sample* for IXIC from 17/04/2017 to 22/03/2023, b) IXIC cumulative returns from 17/04/2017 to 22/03/2023 (below)

5 Conclusions

This paper introduces a novel template matching method that enhances flexibility by employing a parameter grid instead of predefined values. Data-driven weight selection minimizes subjectivity. The results are intriguing but limited to a specific pattern, making it inadequate for an independent automatic trading algorithm. To improve and develop such an algorithm, each technical analysis graphical configuration from the literature should be transformed into a template matrix, creating a library of templates. The proposed procedure would then be applied to the entire library in parallel, selecting the one with the best fit and its corresponding "buy-sell" strategy.

References

1. Cervelló-Royo, R., Guijarro, F., Michniuk, K.: Stock market trading rule based on pattern recognition and technical analysis: forecasting DJIA index with intraday data. Expert Syst. Appl. **42**, 5963–5975 (2015)

2. Leigh, W., Modani, N., Purvis, R., Roberts, T.: Stock market trading rule discovery using technical charting heuristics. Expert Syst. Appl. **23**, 155–159 (2002)
3. Leigh, W., Purvis, R., Ragusa, J.M.: Forecasting the NYSE composite index with technical analysis, pattern recognizer, neural network and genetic algorithm: a case study in romantic decision support. Decis. Supp. Syst. **32**, 361–377 (2002)
4. Tan, Y., Zhu, Y.: Fireworks algorithm for optimization. In: Advances in Swarm Intelligence: First International Conference, ICSI 2010, Beijing, China, 12–15 June 2010, Proceedings, Part I, vol. 1, pp. 355–364. Springer, Heidelberg (2010)
5. Wang, J.-L., Chan, S.-H.: Stock market trading rule discovery using pattern recognition and technical analysis. Expert Syst. Appl. **33**, 304–315 (2007)
6. Wang, J.-L., Chan, S.-H.: Trading rule discovery in the US stock market: an empirical study. Expert Syst. Appl. **36**, 5450–5455 (2009)

A Portfolio's Common Causal Conditional Risk-Neutral PDE

Alejandro Rodriguez Dominguez[1,2(✉)]

[1] Department of Quantitative Research, Miralta Finance Bank S.A., Madrid, Spain
arodriguez@miraltabank.com
[2] Department of Computer Science, University of Reading, Whiteknights House, Reading, UK

Abstract. Portfolio's optimal drivers for diversification are common causes of the constituents' correlations. A closed-form formula for the conditional probability of the portfolio given its optimal common drivers is presented, with each pair constituent-common driver joint distribution modelled by Gaussian copulas. A conditional risk-neutral PDE is obtained for this conditional probability as a system of copulas' PDEs, allowing for dynamical risk management of a portfolio as shown in the experiments. Implied conditional portfolio volatilities and implied weights are new risk metrics that can be dynamically monitored from the PDEs or obtained from their solution.

Keywords: causality · Gaussian copula · partial differential equations · portfolio management · risk-neutral measure

1 Common Causal Conditional Risk-Neutral PDE

It has been proven that optimal portfolio drivers' for diversification must be the common causes of portfolio constituents' correlations [2]. Reichenbach's Common Cause Principle (RCCP) provides a set of independent conditions that variables need to satisfy to be a common cause of a probabilistic correlation [1]. Assume that $(\Omega, \mathcal{F}, P)$ is a standard probability space representing a financial market. A portfolio consists of n financial assets $a_t = [a_{1t}, \ldots, a_{nt}]$ at time t with respective weights $\boldsymbol{w}$. The optimal common causal drivers for the portfolio are selected based on the Commonality Principle [2]. Specifically, the subset of m optimal common causal drivers for a portfolio are selected so that the Reichembach's Common Cause Principle independent conditions have the highest probability [1], from a set of drivers' candidates M belonging to Ω with $M >> m$. The subset of optimal portfolio common drivers $\boldsymbol{D} = \boldsymbol{D_{t-1}} = [D_{t-1,1}, \ldots, D_{t-1,m}]$ assuming for simplicity a lag of 1, satisfying the RCCP conditions, make the portfolio constituents conditional on $\boldsymbol{D}$, independent:

$$P(p_t|\boldsymbol{D_{t-1}}) = P\left(\sum_{i=1}^{n} w_i a_{it} \middle| \boldsymbol{D_{t-1}}\right) = \sum_{i=1}^{n}\sum_{j=1}^{m} w_i D_{t-1,j} P\left(a_{it}|D_{t-1,j}\right) \quad \forall t \quad (1)$$

M. Corazza et al. (Eds.): MAF 2024, *Mathematical and Statistical Methods for Actuarial Sciences and Finance*, pp. 274–279, 2024.
https://doi.org/10.1007/978-3-031-64273-9_45

Jeffrey conditionalization, $P\left(a_{it}|D_{t-1,1} \equiv D_{t-1,1}, \ldots, D_{t-1,m} \equiv D_{t-1,m}\right) = \sum_{j=1}^{m} D_{t-1,j} P(a_{it}|D_{t-1,j})$, $\forall t, \forall i = 1, \ldots n$, is applied in the final step of (1). The joint probabilities of each constituent, with respect to each common driver, follow bivariate distributions that can be modelled with copulas. In the case of a Gaussian copula, with $a_i = a_{it}$, $D_j = D_{t-1,j}$, for any particular t, the density is:

$$C\left(a_i, D_j\right) = \frac{1}{\sqrt{1-\rho_{ij}^2}} exp\left(-\frac{\rho_{ij}^2\left(\Phi^{-1}\left(a_i\right)^2 + \Phi^{-1}\left(D_j\right)^2\right) - 2\rho_{ij}\Phi^{-1}\left(a_i\right)\Phi^{-1}\left(D_j\right)}{2\left(1-\rho_{ij}^2\right)}\right) \tag{2}$$

The conditional probability is given by:

$$P(a_i \leq a_i \mid D_j = D_j) = \frac{\partial}{\partial D_j} C(a_i, D_j)\Big|_{\left(F_{a_i}(a_i), F_{D_j}(D_j)\right)} \tag{3}$$

Applying (2) and (3) to (1), $\forall t$:

$$P(p_t|\boldsymbol{D_{t-1}}) = \sum_{i=1}^{n}\sum_{j=1}^{m} w_i D_{t-1,j} \frac{\partial}{\partial D_{t-1,j}} C(F_{a_{it}}(a_{it}), F_{D_{t-1,j}}(D_{t-1,j})) \tag{4}$$

Using (4), and by computing the partial derivatives of C from (2):

$$P(p|\boldsymbol{D}) = \sum_{i=1}^{n}\sum_{j=1}^{m} \frac{-w_i D_j \exp\left(\frac{-\rho_{a_i D_j}{}^2\left(x_1{}^2 + x_2{}^2\right) - 2\rho_{a_i D_j} x_1 x_2}{2\left(1-\rho_{a_i D_j}{}^2\right)}\right) 2\rho_{a_i D_j}{}^2 \frac{x_2}{\Phi'(x_2)} - 2\rho_{a_i D_j} \frac{x_1}{\Phi'(x_2)}}{2\left(1-\rho_{a_i D_j}{}^2\right)^{3/2}} \tag{5}$$

with $p = p_t$, $\boldsymbol{D} = \boldsymbol{D_{t-1}}$, $a_i = a_{it}$, $D_j = D_{t-1,j}$, $x_1 = \Phi^{-1}(a_i)$ and $x_2 = \Phi^{-1}(D_j)$. In matrix form (5) is given by:

$$P(p|\boldsymbol{D}) = -\boldsymbol{w^T \Pi D} \tag{6}$$

with $\boldsymbol{w} = \begin{bmatrix} w_1 \\ \vdots \\ w_n \end{bmatrix}$, $\boldsymbol{\Pi}_{(nxm)} = \begin{bmatrix} \frac{\partial C(a_1,D_1)}{\partial D_1} & \cdots & \frac{\partial C(a_1,D_m)}{\partial D_m} \\ \vdots & \ddots & \vdots \\ \frac{\partial C(a_n,D_1)}{\partial D_1} & \cdots & \frac{\partial C(a_n,D_m)}{\partial D_m} \end{bmatrix}$ and $\boldsymbol{D}_{(mxn)} = \begin{bmatrix} D_1 & \ldots & D_1 \\ \vdots & \ddots & \vdots \\ D_m & \ldots & D_m \end{bmatrix}$. Deriving (6) with respect to t, the following PDE is obtained:

$$\frac{\partial\ P(p|\boldsymbol{D})}{\partial t} dt = -\boldsymbol{w^T}\left[\left(\frac{\partial \boldsymbol{\Pi}}{\partial \boldsymbol{a}}\frac{\partial \boldsymbol{a}}{\partial t} + \frac{\partial \boldsymbol{\Pi}}{\partial \boldsymbol{D}}\frac{\partial \boldsymbol{D}}{\partial t} + \frac{\partial \boldsymbol{\Pi}}{\partial \boldsymbol{\rho}}\frac{\partial \boldsymbol{\rho}}{\partial t}\right)\boldsymbol{D} + \boldsymbol{\Pi}\frac{\partial \boldsymbol{D}}{\partial t}\right] dt \tag{7}$$

which can be expressed in terms of $P(p|\boldsymbol{D})$ by using (6) as:

$$\left[\frac{\partial\ P(p|\boldsymbol{D})}{\partial t}+\frac{\partial\ P(p|\boldsymbol{D})}{\partial p}\frac{\partial p}{\partial t}+\frac{\partial\ P(p|\boldsymbol{D})}{\partial \boldsymbol{D}}\frac{\partial \boldsymbol{D}}{\partial t}+\left(\frac{\partial\ P(p|\boldsymbol{D})}{\partial \boldsymbol{\rho}}-\boldsymbol{w^T}\,\boldsymbol{\Pi}\frac{\partial \boldsymbol{D}}{\partial \boldsymbol{\rho}}\right)\frac{\partial \boldsymbol{\rho}}{\partial t}\right]dt=0 \quad (8)$$

and the partial derivatives can also be obtained from (6) as $\frac{\partial\ P(p|\boldsymbol{D})}{\partial p} = \boldsymbol{w^T}\left(\frac{\partial \boldsymbol{\Pi}}{\partial \boldsymbol{a}}\boldsymbol{D}\right)$, $\frac{\partial\ P(p|\boldsymbol{D})}{\partial \boldsymbol{D}} = \boldsymbol{w^T}\left(\frac{\partial \boldsymbol{\Pi}}{\partial \boldsymbol{D}}\boldsymbol{D}+\boldsymbol{\Pi}\right)$ and $\frac{\partial\ P(p|\boldsymbol{D})}{\partial \boldsymbol{\rho}} = \boldsymbol{w^T}\left(\frac{\partial \boldsymbol{\Pi}}{\partial \boldsymbol{\rho}}\boldsymbol{D}+\boldsymbol{\Pi}\frac{\partial \boldsymbol{D}}{\partial \boldsymbol{\rho}}\right)$. Portfolio dynamics $\frac{\partial p}{\partial t}$ are given by the Ito process $dp_t = \mu_p dt + \sigma_p dW_{p_t}$ with σ_p the volatility. The common drivers' dynamics $\frac{\partial \boldsymbol{D}}{\partial t}$ follow the m-dimensional Ito process $\boldsymbol{dD_t} = \boldsymbol{\mu_D}dt + \boldsymbol{\sigma_D}d\boldsymbol{W_{D_t}}$, with $\boldsymbol{\sigma_D}$ the square-root of the variance-covariance matrix of the common drivers, and $\boldsymbol{W_{D_t}}$ a m-dimensional correlated Brownian motion. By applying Ito's lemma to $P(p|\boldsymbol{D})$:

$$dP(p|\boldsymbol{D}) = \left[\frac{\partial P(p|\boldsymbol{D})}{\partial t}+\frac{1}{2}\left(\frac{\partial^2 P(p|\boldsymbol{D})}{\partial p^2}\sigma_p^2+\frac{\partial^2 P(p|\boldsymbol{D})}{\partial \boldsymbol{D}^2}{\boldsymbol{\sigma_D}}^2\right)+\frac{\partial^2 P(p|\boldsymbol{D})}{\partial p\partial \boldsymbol{D}}\sigma_p\boldsymbol{\sigma_D}\right]dt+\frac{\partial P(p|\boldsymbol{D})}{\partial p}dp_t+\frac{\partial P(p|\boldsymbol{D})}{\partial \boldsymbol{D}}d\mathbf{D_t} \quad (9)$$

By substituting the PDE given in (8) into the Ito's lemma derivation of $P(p|\boldsymbol{D})$ in (9):

$$dP(p|\boldsymbol{D}) = \left[\frac{1}{2}\left(\frac{\partial^2 P(p|\boldsymbol{D})}{\partial p^2}\sigma_p^2+\frac{\partial^2 P(p|\boldsymbol{D})}{\partial \boldsymbol{D}^2}{\boldsymbol{\sigma_D}}^2\right)+\frac{\partial^2 P(p|\boldsymbol{D})}{\partial p\partial \boldsymbol{D}}\sigma_p\boldsymbol{\sigma_D}\right]dt - \left(\frac{P(p|\boldsymbol{D})}{\boldsymbol{D}}\frac{\partial\ \boldsymbol{D}}{\partial \boldsymbol{\rho}}+\frac{\partial\ P(p|\boldsymbol{D})}{\partial \boldsymbol{\rho}}\right)d\boldsymbol{\rho_t} \quad (10)$$

with $\boldsymbol{\rho_t}$ a $(n*m)$-dimensional vector containing all the correlations between portfolio constituents and the common drivers (not a correlation matrix). These correlations' dynamics must follow Ito processes of any suitable kind, which is irrelevant for this work as the expressions become simplified, but including a $(n*m)$-dimensional Brownian motion. For simplicity, assuming it follows a $(n*m)$-dimensional geometric Brownian motion $\frac{\partial \boldsymbol{\rho}}{\partial t} = d\boldsymbol{\rho_t} = \boldsymbol{\mu_\rho}\boldsymbol{\rho_t}dt + \boldsymbol{\sigma_\rho}\boldsymbol{\rho_t}d\boldsymbol{W_\rho^t}$, substituting in (10):

$$dP(p|\boldsymbol{D}) = \left[\frac{1}{2}\left(\frac{\partial^2 P(p|\boldsymbol{D})}{\partial p^2}\sigma_p^2+\frac{\partial^2 P(p|\boldsymbol{D})}{\partial \mathbf{D}^2}{\boldsymbol{\sigma_D}}^2\right)+\frac{\partial^2 P(p|\boldsymbol{D})}{\partial p\partial \boldsymbol{D}}\sigma_p\boldsymbol{\sigma_D} - \left(\frac{P(p|\boldsymbol{D})}{\boldsymbol{D}}\frac{\partial\ \boldsymbol{D}}{\partial \boldsymbol{\rho}}+\frac{\partial\ P(p|\boldsymbol{D})}{\partial \boldsymbol{\rho}}\right)\boldsymbol{\mu_\rho}\boldsymbol{\rho_t}\right]dt - \left(\frac{P(p|\boldsymbol{D})}{\boldsymbol{D}}\frac{\partial\ \boldsymbol{D}}{\partial \boldsymbol{\rho}}+\frac{\partial\ P(p|\boldsymbol{D})}{\partial \boldsymbol{\rho}}\right)\boldsymbol{\sigma_\rho}\boldsymbol{\rho_t}d\boldsymbol{W_\rho^t} \quad (11)$$

The common causal conditional risk-neutral PDE is given by equating the drift part in (11) to the drift of the common causal drivers' processes' times the conditional probability of the portfolio given these set of common causal drivers. This is due to the fact that, when portfolio constituents are projected in a common causal conditional probability space, they become independent and the portfolio risk is diversified, therefore the common causal conditional probability space becomes conditionally risk-neutral. The return of the portfolio conditional

on the common drivers is proportional to the common drivers' drift term plus an error term Δ accounting for portfolio's common causal drivers' error selection:

$$\left[\frac{1}{2}\left(\frac{\partial^2 P(p|\boldsymbol{D})}{\partial p^2}\sigma_p^2 + \frac{\partial^2 P(p|\boldsymbol{D})}{\partial \mathbf{D}^2}{\boldsymbol{\sigma}_{\boldsymbol{D}}}^2\right) + \frac{\partial^2 P(p|\boldsymbol{D})}{\partial p \partial \boldsymbol{D}}\sigma_p \boldsymbol{\sigma}_{\boldsymbol{D}}\right] dt = [\boldsymbol{\mu}_{\boldsymbol{D}}\, P(p|\boldsymbol{D}) + \boldsymbol{\Delta}]\, dt \quad (12)$$

and there is one condition, given by the following PDE, such that the PDE in (11) does not have a Brownian motion component, making it conditional risk-neutral:

$$\frac{P(p|\boldsymbol{D})}{\boldsymbol{D}}\frac{\partial \boldsymbol{D}}{\partial \boldsymbol{\rho}} = -\frac{\partial P(p|\boldsymbol{D})}{\partial \boldsymbol{\rho}} \quad (13)$$

Parameter $\boldsymbol{\Delta}$ in (12) measures the deviations from conditional risk neutrality due to unknown confounders or not having selected all optimal common causal drivers for the portfolio. We assume now for simplicity it is zero. It can work as a loss function for a learning/calibration problem in which the optimal set of common drivers are found when $\boldsymbol{\Delta}$ is minimum. Additionally, it can be used as a signal that the common drivers' set needs to be revised. If the partial derivatives computed from (6) are substituted in (13):

$$\frac{-\boldsymbol{w}^{\boldsymbol{T}}\boldsymbol{\Pi}\boldsymbol{D}}{\boldsymbol{D}}\frac{\partial \boldsymbol{D}}{\partial \boldsymbol{\rho}} = -\boldsymbol{w}^{\boldsymbol{T}}\left(\frac{\partial \boldsymbol{\Pi}}{\partial \boldsymbol{\rho}}\boldsymbol{D} + \boldsymbol{\Pi}\frac{\partial \boldsymbol{D}}{\partial \boldsymbol{\rho}}\right) \quad (14)$$

which simplifies to:

$$\boldsymbol{\Pi}\frac{\partial \boldsymbol{D}}{\partial \boldsymbol{\rho}} = \left(\frac{\partial \boldsymbol{\Pi}}{\partial \boldsymbol{\rho}}\boldsymbol{D} + \boldsymbol{\Pi}\frac{\partial \boldsymbol{D}}{\partial \boldsymbol{\rho}}\right) \quad (15)$$

This gives another condition for the PDE, $\frac{\partial \boldsymbol{\Pi}}{\partial \rho}\boldsymbol{D} = 0$, which can be derived analytically from $\boldsymbol{\Pi}$ in (6), and consists of a system of n equations. From now on, the common causal conditional risk-neutral PDE is expressed in terms of the vector of portfolio constituents' returns, $\boldsymbol{a}$, instead of the portfolio's return p in (12). The partial derivatives of $P(p|\boldsymbol{D})$ in (12) are computed using the expression in (6):

$$\frac{\partial^2\ P(p|\boldsymbol{D})}{\partial p^2} = \boldsymbol{w}^{\boldsymbol{T}}\left(\frac{\partial^2 \boldsymbol{\Pi}}{\partial \boldsymbol{a}^2}\boldsymbol{D} + \frac{\partial \boldsymbol{\Pi}}{\partial \boldsymbol{a}}\frac{\partial \boldsymbol{D}}{\partial \boldsymbol{a}} + \frac{\partial \boldsymbol{\Pi}}{\partial \boldsymbol{a}}\frac{\partial \boldsymbol{D}}{\partial \boldsymbol{a}} + \boldsymbol{\Pi}\frac{\partial^2 \boldsymbol{D}}{\partial \boldsymbol{a}^2}\right) = \boldsymbol{w}^{\boldsymbol{T}}\left(\frac{\partial^2 \boldsymbol{\Pi}}{\partial \boldsymbol{a}^2}\boldsymbol{D}\right) \quad (16)$$

$$\frac{\partial^2\ P(p|\boldsymbol{D})}{\partial \boldsymbol{D}^2} = \boldsymbol{w}^{\boldsymbol{T}}\left(\frac{\partial^2 \boldsymbol{\Pi}}{\partial \boldsymbol{D}^2}\boldsymbol{D} + 2\frac{\partial \boldsymbol{\Pi}}{\partial \boldsymbol{D}}\right); \frac{\partial^2\ P(p|\boldsymbol{D})}{\partial \boldsymbol{\rho}^2} = \boldsymbol{w}^{\boldsymbol{T}}\left(\frac{\partial \boldsymbol{\Pi}}{\partial \boldsymbol{\rho}}\frac{\partial \boldsymbol{D}}{\partial \boldsymbol{\rho}} + \boldsymbol{\Pi}\frac{\partial^2 \boldsymbol{D}}{\partial \boldsymbol{\rho}^2}\right) \quad (17)$$

$$\frac{\partial^2 P(p|\boldsymbol{D})}{\partial p \partial \boldsymbol{D}} = \boldsymbol{w}^{\boldsymbol{T}}\left(\frac{\partial \boldsymbol{\Pi}}{\partial \boldsymbol{a}} + \frac{\partial \boldsymbol{\Pi}}{\partial \boldsymbol{a}\partial \boldsymbol{D}}\boldsymbol{D}\right) \quad (18)$$

substituting (16–18) in (12), and assuming $\boldsymbol{\Delta} = 0$:

$$\begin{aligned}&\boldsymbol{w^T}\left[\frac{1}{2}\left(\left(\frac{\partial^2 \boldsymbol{\Pi}}{\partial \boldsymbol{a}^2}\boldsymbol{D}\right)\boldsymbol{\sigma_p}^2+\left(\frac{\partial^2 \boldsymbol{\Pi}}{\partial \boldsymbol{D}^2}\boldsymbol{D}+2\frac{\partial \boldsymbol{\Pi}}{\partial \boldsymbol{D}}\right)\boldsymbol{\sigma_D}^2\right)+\left(\frac{\partial \boldsymbol{\Pi}}{\partial \boldsymbol{a}}+\frac{\partial \boldsymbol{\Pi}}{\partial \boldsymbol{a}\partial \boldsymbol{D}}\boldsymbol{D}\right)\boldsymbol{\sigma_p}\boldsymbol{\sigma_D}\right]\\&=-\boldsymbol{\mu_D}\,\boldsymbol{w^T}\,\boldsymbol{\Pi}\,\boldsymbol{D}\end{aligned} \tag{19}$$

with the portfolio variance $\boldsymbol{\sigma_p}^T\boldsymbol{\sigma_p} = \boldsymbol{w}^T\boldsymbol{\Sigma_p}\boldsymbol{w} = \boldsymbol{\sigma_p}^2$ now expressed as a diagonal variance-covariance matrix of independent portfolio constituents, and having factorized (19) by $\boldsymbol{w}^T$, getting, $\boldsymbol{\Sigma_p}\boldsymbol{w} = \begin{bmatrix}\sigma_1^2 & \dots & 0\\ \vdots & \ddots & \vdots\\ 0 & \dots & \sigma_n^2\end{bmatrix}\boldsymbol{w}$. A PDE that does not depend on the weights of the portfolio, except from the term $\boldsymbol{\Sigma_p}\boldsymbol{w}$, is obtained:

$$\begin{aligned}&\frac{1}{2}\left(\left(\frac{\partial^2 \boldsymbol{\Pi}}{\partial \boldsymbol{a}^2}\boldsymbol{D}\right)\boldsymbol{\Sigma_p}\boldsymbol{w}+\left(\frac{\partial^2 \boldsymbol{\Pi}}{\partial \boldsymbol{D}^2}\boldsymbol{D}+2\frac{\partial \boldsymbol{\Pi}}{\partial \boldsymbol{D}}\right)\boldsymbol{\sigma_D}^2\right)+\left(\frac{\partial \boldsymbol{\Pi}}{\partial \boldsymbol{a}}+\frac{\partial \boldsymbol{\Pi}}{\partial \boldsymbol{a}\partial \boldsymbol{D}}\boldsymbol{D}\right)\boldsymbol{\Sigma_p}\boldsymbol{w}\boldsymbol{\sigma_D}\\&+\boldsymbol{\mu_D}^T\boldsymbol{D}\boldsymbol{\Pi}=0\end{aligned} \tag{20}$$

with $\boldsymbol{\sigma_D}^2 = \boldsymbol{\Sigma_D} = Cov(D_{t-1}^k, D_{t-1}^q) = e^{(\mu_{Dk}+\mu_{Dq})(t-1)}(e^{\rho_{Dk,Dq}\sigma_{Dk}\sigma_{Dq}t}-1)$, $\boldsymbol{\Sigma_p} = Cov(a_t^v, a_t^z) = e^{(\mu_{av}+\mu_{az})t}(e^{\sigma_{av}\sigma_{az}t}-1)$, $\forall k,q = 1,\dots,m; v,z = 1,\dots,n$, and $Cov(a_t^v, a_t^z) = 0$ if $v \neq z$. Expressions are given for the partial derivatives in (20), which can be computed with Kronecker products and substituted back in (20). One case is shown due to space constraints:

$$\begin{aligned}&\boldsymbol{D}^T\frac{\partial^2 \boldsymbol{\Pi}}{\partial \boldsymbol{a}^2}\boldsymbol{\Sigma_p}\boldsymbol{w}\\&=\boldsymbol{D}^T_{(nxm)}\left(\left(\frac{\partial}{\partial a_1},\dots,\frac{\partial}{\partial a_n}\right)\oplus\left(\frac{\partial}{\partial a_1},\dots,\frac{\partial}{\partial a_n}\right)\oplus\boldsymbol{\Pi}_{(nxm)}\right)_{(nx(n^2m))}\boldsymbol{\Sigma_p}\boldsymbol{w}_{(nxnx1)}\\&=\boldsymbol{D}^T_{(nxm)}\begin{bmatrix}\frac{\partial^3 C(a_1,\mathrm{D}_1)}{\partial \mathrm{D}_1\partial a_1^2} & \dots 0\dots & \frac{\partial^3 C(a_n,\mathrm{D}_1)}{\partial \mathrm{D}_1\partial a_n^2}\\ \vdots & \ddots\ \vdots\ddots & \vdots\\ \frac{\partial^3 C(a_1,\mathrm{D}_m)}{\partial \mathrm{D}_m\partial a_1^2} & \dots 0\dots & \frac{\partial^3 C(a_n,\mathrm{D}_m)}{\partial \mathrm{D}_m\partial a_n^2}\end{bmatrix}_{(mx(mn^2)}\boldsymbol{\Sigma_p}\boldsymbol{w}_{(n^2x1)}\end{aligned} \tag{21}$$

Notes and Comments. A system of copulas' PDEs for each pair of constituent/common driver is obtained. Each PDE can be used for online risk management at constituent level, as seen in Figs. 1, 2, 3 and 4. Copulas' PDEs are computed from data on a rolling window basis. Deviations from theoretical values are shown in these figures for two constituents and portfolio's common drivers. Large deviations are due to important events such as earning reports (From 18/03/2008 earnings reports, GS shares rise more than 30%), idiosyncratic news (From 28/10/2008 the Volkswagen scandal caused GS shares to plunge), and systematic market moves (30/10/2008, Dax rose 2,44% in one day), all from Fig. 3. Figures 1 and 3 show that deviations remain close to zero except for punctual events, so that the conditional probabilities are almost always described accurately by the PDEs. All copulas' PDEs deviations add to the deviations from conditional risk neutrality in the PDE system expressed as $\boldsymbol{\Delta}$ in (12). If deviations are added up over a period as in Fig. 2, risk-hedging information is obtained

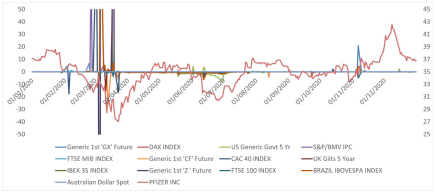


Fig. 1. Copulas' PDEs values for Phyzer - 2020

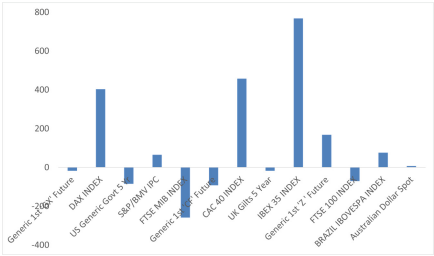


Fig. 2. Sum of values for 2020 (Phyzer)

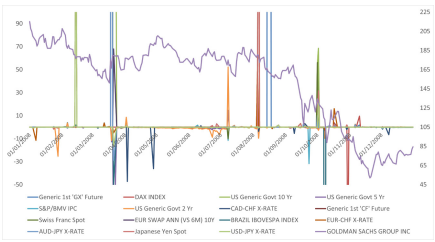


Fig. 3. Copulas' PDEs values for Goldman Sachs (GS) - 2008

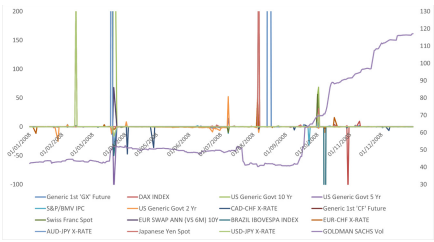


Fig. 4. GS 2008 Volatility

and can be included in sensitivity-based methods for portfolio optimization with respect to common causal drivers [2], as an extra component of the sensitivities for tail-risk management (large deviations). Experiments use equal-weight portfolios, but can be selected to minimize the deviations, left as future work. Large deviations are related to increasing volatility (Fig. 4). Implied market hedges are obtained after solving the weights in the PDE system. With the system PDEs solution, which is left for future work, implied weights based on the conditional risk-neutral common causal measure could be obtained for longer horizons.

Acknowledgement. Thanks to Hugo Valle Valcarcel for helping in the numerical computation of the derivatives and to Miralta Finance Bank S.A. for providing datasets.

References

1. Reichenbach, H.: The direction of time. Dover Publications, Mineola. Edited by Maria Reichenbach (1956)
2. Dominguez, R.: Alejandro,: Portfolio optimization based on neural networks sensitivities from assets dynamics respect common drivers. Mach. Learn. Appl. **11**, 100477 (2023). ISSN 2666–8270

A Structural Credit Risk Model with Default Contagion

Bud Schiphorst[1,2(✉)], Michel Mandjes[1,3], Peter Spreij[1,4], and Erik Winands[1,2]

[1] Korteweg-de Vries Institute for Mathematics, University of Amsterdam, Amsterdam, The Netherlands
B.Schiphorst@uva.nl
[2] Rabobank, Utrecht, The Netherlands
[3] Mathematical Institute, Leiden University, Leiden, The Netherlands
[4] IMAPP, Radboud University, Nijmegen, The Netherlands

Abstract. Structural threshold models are common industry practice for modelling portfolio credit risk, but often only consider default dependence via underlying common factors. We consider a structural model extension that allows for additionally incorporating default contagion effects. A simulation study illustrates that ignoring default contagion effects may lead to significant underestimation of portfolio tail risk. As a key contribution, we propose a procedure for estimating default contagion parameters from historical default probability data.

Keywords: portfolio credit risk · default contagion · structural model

1 Introduction

The dependence between default events of obligors is a key aspect of portfolio credit risk management. A common approach in practice is to use a structural threshold model (see [5, p. 465]) in which a default event of an obligor is triggered by a latent value process reaching some threshold. The value processes of different obligors are then often assumed to be conditionally independent given underlying common factors, such as macroeconomic or industry-specific risk drivers.

Another important form of dependence may however arise due to default contagion effects, in which an increase in default risk of one obligor directly causes an increase in default risk of another obligor. In corporate parent-subsidiary relationships, for example, increased default risk can propagate from a parent company to a subsidiary. As another example, increased default risk of a sovereign issuer may propagate to entities operating in the same country.

We propose a structural threshold model that incorporates both indirect default dependence via underlying common factors, as well as direct default contagion effects. The model specifically allows for the special case where the default of one obligor guarantees the default of another, but also allows default risk to partially propagate from or to multiple different obligors. As a key contribution, we outline a procedure to estimate the contagion parameters from default

M. Corazza et al. (Eds.): MAF 2024, *Mathematical and Statistical Methods for Actuarial Sciences and Finance*, pp. 280–285, 2024.
https://doi.org/10.1007/978-3-031-64273-9_46

probability data. Once calibrated, the model can be easily used for simulation of portfolio losses, similar to the structural threshold models used in practice. The combination of these desirable properties distinguishes the model from previous proposed default contagion extensions of the threshold model (see e.g. [1,3,6])[1].

Based on a simulation study, we illustrate that ignoring default contagion effects may cause significant underestimation of portfolio tail risk. This risk is relatively well captured by using estimated default contagion parameters.

2 Structural Model with Default Contagion

We consider a credit portfolio with obligors indexed by $\mathcal{I} = \{1, \ldots, N\}$ that may default only at the end of some specified time horizon T, e.g. in 1 year. Following a structural approach, we consider a (latent) joint value process $\mathbf{V} = [V_i]_{i\in\mathcal{I}}$ and say that an obligor has defaulted when its value process is non-positive:

$$\{\text{Default of obligor } i \text{ at time } T\} := \{V_i(T) \leq 0\}\,, \ \forall i \in \mathcal{I}. \tag{1}$$

An increase in default risk, is thus represented by a decrease in value.

The joint value process $\mathbf{V}$ specifies both marginal default probabilities as well as default dependence between obligors, including possible default contagion effects. As is common in portfolio credit risk modelling, we focus on modelling the dependence between defaults. To simulate the distribution of 1-year portfolio losses, we therefore assume that the vector of marginal 1-year default probabilities $\hat{\mathbf{PD}} = [\hat{\mathrm{PD}}_i]_{i\in\mathcal{I}}$ is given. For the estimation of default contagion parameters, we assume the availability of the historical time-series $\{\hat{\mathbf{PD}}(t_m)\}_{m=0}^{M}$.

In practice, (estimated) default probabilities may for example be provided by rating agencies or internal bank models. Alternatively, default probabilities may be inferred from market data such as credit spreads. We assume that default contagion effects are already incorporated into these default probabilities.

We additionally assume that we already know the binary structure of the default contagion dependence: i.e. between which obligors there exists such direct dependence. This binary structure is indeed obvious in many practical applications[2]. The strength of the dependencies remains to be estimated.

2.1 Base Model

We first consider $\mathbf{V} = \mathbf{Y}$, where $\mathbf{Y} = [Y_i]_{i\in\mathcal{I}}$ represents the intrinsic value process of all obligors and is assumed to be a correlated N-dimensional Brownian motion:

$$\mathrm{d}\mathbf{Y}(t) = \mathbf{\Sigma}^{\frac{1}{2}}\mathrm{d}\mathbf{B}(t), \ \mathbf{Y}(0) = \mathbf{y} \in \mathbb{R}^N, \tag{2}$$

where $\mathbf{B}$ is an N-dimensional Wiener process and $\mathbf{\Sigma}$ is a correlation matrix[3].

[1] In particular, earlier proposed calibration of default contagion parameters often relies on expert input or on strong ad hoc assumptions.

[2] For example in corporate parent-subsidiary relationships. Alternatively, the binary dependence structure may be identified using a network-based approach, see e.g. [2].

[3] That is, the process is scaled such that $\Sigma_{ii} = 1$ for all $i \in \mathcal{I}$.

In the base model, the marginal default probability for $i \in \mathcal{I}$ can be computed as

$$\mathrm{PD}_i = \mathbb{P}\left[Y_i(T) \leq 0 \;\middle|\; Y_i(0) = y_i\right] = \Phi\left(-\frac{y_i}{\sqrt{T}}\right), \tag{3}$$

where Φ is the standard Gaussian CDF. Conversely, given a default probability $\hat{\mathrm{PD}}_i \in (0,1)$, we can solve for the corresponding starting point $\hat{y}_i$ to obtain

$$\hat{y}_i = -\sqrt{T} \cdot \Phi^{-1}\left(\hat{\mathrm{PD}}_i\right). \tag{4}$$

We note that the base model effectively corresponds to a Gaussian threshold model, similar to multi-factor extensions of the Merton model that are popular in industry [5, p. 430]. We assume here that the correlation matrix $\mathbf{\Sigma}$ is known or has already been calibrated by using such a model. The distribution of losses can then be simulated when the default probability vector $\hat{\mathbf{PD}}$ is also given.

2.2 Default Contagion Extension

We propose a structural extension of the base model in which default risk can propagate via a parent structure represented by a weights matrix $\mathbf{W} = [W_{ij}]_{i,j\in\mathcal{I}}$ with non-negative entries and row-sums equal to 1. We say $j \in \mathcal{I}$ is a parent of child $i \in \mathcal{I}$ when $W_{Ij} > 0$ and $i \neq j$. We say that obligor i has no parents when $W_{ii} = 1$. We assume the existence of a partition $\mathcal{I} = \{\mathcal{I}_P, \mathcal{I}_C\}$, where all obligors without parents are in $\mathcal{I}_P$ and the child obligors in $\mathcal{I}_C$ only have parents in $\mathcal{I}_P$[4].

In the proposed extended model, the value $\mathbf{V}$ is the (element-wise) minimum of the intrinsic value $\mathbf{Y}$ and the propagated value $\mathbf{X} = [X_i]_{i\in\mathcal{I}}$. The intrinsic value $\mathbf{Y}$ is modelled as in (2) and the propagated value $\mathbf{X}$ is modelled as a convex combination of the intrinsic value and the value of parents:

$$\mathbf{V}(t) = \min\left[\mathbf{Y}(t), \mathbf{X}(t)\right], \;\; \mathbf{X}(t) = \mathbf{W}\mathbf{Y}(t). \tag{5}$$

A decrease in value V_i of an obligor i may be caused by either a decrease in intrinsic value Y_i or by a decrease in value of its parents propagated via X_i. Similarly, the default event as defined in (1) can be triggered by a decrease in either intrinsic or propagated value:

$$\{\text{Default of obligor } i \text{ at time } T\} = \{Y_i(T) \leq 0\} \cup \{X_i(T) \leq 0\}, \;\; \forall i \in \mathcal{I}. \tag{6}$$

We highlight two special cases that can be captured by this model:

1. In the special case that no obligor has parents, i.e. $\mathcal{I} = \mathcal{I}_P$ and $\mathbf{W} = \mathbf{I}$, the model reduces to the base model. This is because each obligor $i \in \mathcal{I}_P$ without parents has $W_{Ii} = 1$ and therefore value $V_i = \min[Y_i, Y_i] = Y_i$.
2. If obligor i has one parent j with weight $W_{ij} = 1$, then $V_i \leq X_i = Y_j = V_j$, so that a default event of parent j implies a default event of obligor i.

[4] This imposes a restriction where it is not allowed for parents to have also parents themselves and is similar to the primary-secondary structure assumed in [4].

In the extended model, we have

$$\begin{bmatrix}\mathbf{Y}(T)\\ \mathbf{X}(T)\end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix}\mathbf{y}\\ \mathbf{W}\mathbf{y}\end{bmatrix},\ T\cdot\begin{bmatrix}\boldsymbol{\Sigma} & \boldsymbol{\Sigma}\mathbf{W}^T\\ \mathbf{W}\boldsymbol{\Sigma} & \mathbf{W}\boldsymbol{\Sigma}\mathbf{W}^T\end{bmatrix}\right), \tag{7}$$

so that the marginal default probability of an obligor $i \in \mathcal{I}$ can be computed as

$$\begin{aligned}\mathrm{PD}_i\big(\mathbf{y};\mathbf{W}\big) &= \mathbb{P}\left[Y_i(T)\le 0 \ \text{ or } X_i(T)\le 0 \,\middle|\, \mathbf{Y}(0)=\mathbf{y}\right]\\ &= \Phi\left(-\frac{y_i}{\sqrt{T}}\right)+\Phi\left(-\frac{x_i}{\sigma_{X_i}\sqrt{T}}\right)-\Phi_2^{\rho_i}\left(-\frac{y_i}{\sqrt{T}},\ -\frac{x_i}{\sigma_{X_i}\sqrt{T}}\right),\end{aligned} \tag{8}$$

where $x_i = [\mathbf{W}\mathbf{y}]_i$, $\sigma_{X_i} = \sqrt{[\mathbf{W}\boldsymbol{\Sigma}\mathbf{W}^T]_{ii}}$ and $\Phi_2^{\rho_i}(\cdot,\cdot)$ is the bi-variate Gaussian CDF with correlation $\rho_i = [\mathbf{W}\boldsymbol{\Sigma}]_{ii}/\sigma_{X_i}$. For obligor $i \in \mathcal{I}_P$ without parents, Eq. (8) indeed simplifies to as in the base model in Eq. (3).

Given a weights matrix $\mathbf{W}$, correlation matrix $\boldsymbol{\Sigma}$ and default probability vector $\hat{\mathbf{PD}}$, we want to find the corresponding starting point $\hat{\mathbf{y}}$ that satisfies $\mathrm{PD}_i\big(\hat{\mathbf{y}};\mathbf{W}\big) = \hat{\mathrm{PD}}_i$ for all $i \in \mathcal{I}$. We can first find $\hat{\mathbf{y}}_P := [\hat{y}_i]_{i\in\mathcal{I}_P}$ by using the result in (4). Then given $\hat{\mathbf{y}}_P$, we can numerically solve for the remaining $\hat{\mathbf{y}}_C := [\hat{y}_i]_{i\in\mathcal{I}_C}$ by using the result in (8). The latter is possible because in (8) PD_i is strictly decreasing in y_i and has range equal to $(0,1)$ when $W_{ii} > 0$.[5]

The distribution of portfolio losses can be simulated when the weights matrix $\mathbf{W}$, correlation matrix $\boldsymbol{\Sigma}$ and default probability vector $\hat{\mathbf{PD}}$ are estimated or given. This is analogous to the base model, where $\mathbf{W} = \mathbf{I}$.

3 Estimation of the Weights Matrix W

Intuitively, we propose to estimate the weights matrix $\mathbf{W}$ by using that the value processes of the obligors should have correlation $\boldsymbol{\Sigma}$ after filtering out default contagion effects. We therefore consider a moment condition on the normalized discrete-time increments of the intrinsic value process $\mathbf{Y}$ as defined in (2):

$$\mathbb{E}\left[\mathbf{Z}(t_m)\mathbf{Z}(t_m)^T\right] = \boldsymbol{\Sigma}, \quad \text{for } \mathbf{Z}(t_m) := \frac{\mathbf{Y}(t_m)-\mathbf{Y}(t_{m-1})}{\sqrt{t_m - t_{m-1}}}. \tag{9}$$

Equation (9) implies the following moment condition for the subcomponents $\mathbf{Z}_C := [Z_i]_{i\in\mathcal{I}_C}, \mathbf{Z}_P := [Z_i]_{I\in\mathcal{I}_P}$ of obligors with and without parents respectively:

$$\mathbb{E}\left[\mathbf{Z}_C(t_m)\mathbf{Z}_P(t_m)^T\right] = \boldsymbol{\Sigma}_{CP}, \ \text{where} \ \ \boldsymbol{\Sigma}_{CP} := [\Sigma_{ij}]_{i\in\mathcal{I}_C, j\in\mathcal{I}_P}. \tag{10}$$

Although the increments $\{\mathbf{Z}(t_m)\}_{m=1}^M$ are not directly observed, we can use $\{\hat{\mathbf{PD}}(t_m)\}_{m=0}^M$ to compute the inferred increments $\{\hat{\mathbf{Z}}(t_m)\}_{m=1}^M$ for a given $\mathbf{W}$:

$$\hat{\mathbf{Z}}(t_m;\mathbf{W}) := \frac{\hat{\mathbf{y}}(t_m)-\hat{\mathbf{y}}(t_{m-1})}{\sqrt{t_m - t_{m-1}}}, \text{where} \ \ \mathbf{PD}\big(\hat{\mathbf{y}}(t_m),\mathbf{W}\big) \equiv \hat{\mathbf{PD}}(t_m). \tag{11}$$

[5] Note that PD_i has lower bound $\lim_{y_i\to\infty}\mathrm{PD}_i(\mathbf{y};\mathbf{W}) = \Phi\big(-\frac{x_i}{\sqrt{T}\sigma_{X_i}}\big)$ when $W_{ii} = 0$, since the probability of default triggered by propagated value X_i is then independent of intrinsic value y_i. For example, we have $\mathrm{PD}_i > \mathrm{PD}_j$ when $W_{ij} = 1$.

We can thus estimate $\mathbf{W}$ by (numerically) minimizing the loss function $L(\mathbf{W})$, which is based on the Frobenius norm of the difference between $\mathbf{\Sigma}_{CP}$ and the sample correlation of the inferred value increments:

$$L(\mathbf{W}) := \left\| \mathrm{Corr}\left(\hat{\mathbf{Z}}_C(t_m;\mathbf{W}), \hat{\mathbf{Z}}_P(t_m)\right) - \mathbf{\Sigma}_{CP} \right\|_F, \text{ for } \begin{array}{l} \hat{\mathbf{Z}}_C := [\hat{Z}_i]_{i\in\mathcal{I}_C} \\ \hat{\mathbf{Z}}_P := [\hat{Z}_i]_{i\in\mathcal{I}_P} \end{array}. \quad (12)$$

We note two advantages of the loss function in (12) for numerical optimization. First, each row $\mathbf{W}_i := [W_{ij}]_{j\in\mathcal{I}}$ can be estimated separately for each $i \in \mathcal{I}_C$, as each row affects a different element of $\hat{\mathbf{Z}}_C$. So, if each child only has a few parents, only a few weights have to be estimated per row. Second, we again note that the weights corresponding to obligors without parents $i \in \mathcal{I}_P$ are equal to $W_{ii} = 1$. This means that the increments $\hat{\mathbf{Z}}_P$ have to be computed only once.

4 Numerical Example and Possible Extensions

To illustrate the possible impact of default contagion effects on aggregate portfolio risk, we consider a numerical example. We first simulate portfolio losses using the extended model in (5) for a randomly drawn weights matrix $\mathbf{W}_{\text{true}}$. We then also simulate losses with estimated weights $\mathbf{W}_{\text{est}}$ and naive weights $\mathbf{W}_{\text{naive}} = \mathbf{I}$, to assess the impact on the expected losses and Value-at-Risk (VaR).

We consider a portfolio with $N = 800$ obligors, partitioned into a set of obligors without parents $\mathcal{I}_P = \{1, \ldots, 400\}$ and a set of obligors $\mathcal{I}_C = \{401, \ldots, 800\}$ that all have the same two parents $\{1, 2\}$. We choose an exposure of 100 for all obligors and default probabilities of 0.005, 0.008 for the obligors in $\mathcal{I}_P$ and $\mathcal{I}_C$ respectively. We also choose a correlation matrix $\mathbf{\Sigma}$ with off-diagonal entries equal to 0.3, corresponding to a 1-factor model with equal factor loadings for all obligors. Finally, for each child $i \in \mathcal{I}_C$, we fix randomly drawn weights to both parents in $\{1, 2\}$, independent of all other obligors. The weights are drawn from uniform distributions, such that $W_{i1} \sim \mathrm{Unif}(0, 1)$ and $W_{i2} \sim \mathrm{Unif}(0, 1 - W_{i1})$.

For estimating $\mathbf{W}$, we simulate $M = 250$ daily observations of the default probability vector, corresponding to roughly 1 year of trading days. We then minimize the loss function in (12) by an initial grid search, which is followed by a numerical gradient-based optimization. This combination mitigates the issue of possible local minima and is not too computationally intensive given that only two parameters have to be estimated per row of $\mathbf{W}$.

Figure 1 shows the simulated (relative) expected losses (EL) and VaR for different quantiles for the three different weights matrices, based on 10^7 simulations. For easier visual comparison, all figures are relative to the corresponding figure for the true weights $\mathbf{W}_{\text{true}}$. The results show that the simulated expected losses are very close for all three weights matrices. This is as expected, because the expected losses depend only on the marginal default probabilities to which the model is calibrated by construction. The results also show that the VaR at higher quantiles is significantly underestimated when ignoring default contagion effects (i.e. using when $\mathbf{W}_{\text{naive}}$). In contrast, the results indicate that using the estimated $\mathbf{W}_{\text{est}}$ allows for estimating the VaR reasonably well.

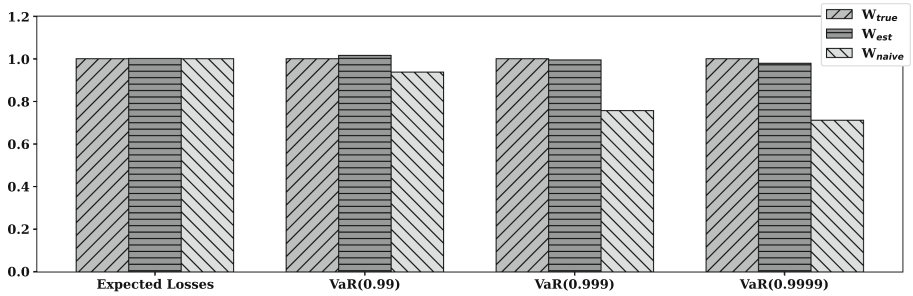


Fig. 1. Simulated expected losses and VaR(q) for $q = 0.99, 0.999, 0.9999$

We note that although the aggregate portfolio risk is estimated reasonable well, the estimation of the weights is relatively volatile on an obligor level. Also, although daily historical default probability data may be inferred from market data such as credit spreads, data provided by rating agencies or internal bank models will likely be less frequent. It may therefore be worthwhile to further improve the efficiency of the proposed estimation procedure: for example by explicitly incorporating data from the underlying factor model. Nevertheless, we have illustrated that it is possible to estimate default contagion parameters from historical default probability data, overcoming the reliance on expert input.

An interesting extension of the proposed model would be to allow parents to also have parents themselves. In that case, chain-like propagation of default contagion effects can be incorporated. In such an extension, the propagated value in (5) could for example become $\mathbf{X}(t) = \mathbf{WV}(t)$. However, the results derived in e.g. (7) and (8) would then take a less simple form. More generally, other functional forms can be considered for the default contagion model extension. For example, using a maximum instead of a minimum in (5) could represent the case where one obligor acts as a guarantor for another obligor.

References

1. Anagnostou, I., Sourabh, S., Kandhai, D.: Incorporating contagion in portfolio credit risk models using network theory. Complexity **2018**, 1–15 (2018)
2. Cimini, G., Mastrandrea, R., Squartini, T.: Reconstructing Networks. Cambridge University Press, Cambridge (2021)
3. Egloff, D., Leippold, M., Vanini, P.: A simple model of credit contagion. J. Bank. Financ. **31**(8), 2475–2492 (2007)
4. Jarrow, R.A., Yu, F.: Counterparty risk and the pricing of defaultable securities. J. Financ. **56**(5), 1765–1799 (2001)
5. McNeil, A.J., Frey, R., Embrechts, P.: Quantitative Risk Management: Concepts, Techniques and Tools-Revised Edition. Princeton University Press, Princeton (2015)
6. Neu, P., Kühn, R.: Credit risk enhancement in a network of interdependent firms. Phys. A **342**(3–4), 639–655 (2004)

Risk Evaluating for Subdiffusive Option Price Model with Gamma Subordinator

Nataliya Shchestyuk[1,2(✉)], Svitlana Drin[1,2], and Serhii Tyshchenko[2]

[1] School of Business, Orebro University, 70182 Orebro, Sweden
nataliya.shchestyuk@oru.se
[2] Department of Mathematics, National University of Kyiv-Mohyla Academy, Kyiv 04070, Ukraine

Abstract. The article focuses on Value-at-risk measuring for options in situations characterized by the lack of liquidity when the underlying stock price has motionless periods. A similar behavior can be observed in physical systems exhibiting sub-diffusion. In the considered sub-diffusive model, the bond movement and stock process are time-changed by the stochastic clock with gamma subordinator. In the model, the two techniques for option pricing were considered. The first very common approach for the time-changed model is to find option prices as the discounted expected payoff under the risk-neutral measure. The second technique for option pricing is based on a fractional version of what is called Dupire's equation. The Value-at-Risk evaluating procedure for the proposed model was discussed and we show that this procedure is based on the Fractional Fokker-Planck equation (FFPE).

Keywords: Option pricing · subdiffusion · Value-at-risk · Gamma subordinator

1 Introduction

In recent decades subdiffusive processes are getting increasing attention. These stochastic processes are usually used in statistical physics to model anomalous diffusion phenomena with trapping events (see [4,5,8], and others). In financial markets with illiquid assets, we often see similar anomalous subdiffusion, when relatively long periods without any trading are observed. This feature is common for crisis periods that negatively affect financial activity or for emerging markets in which the number of participants, and thus the number of transactions, is rather low.

Therefore, the physical models of subdiffusion can be successfully applied to describe financial data and model the dynamics of the financial market, including option pricing. See for example paper [2], where option pricing was proposed in the fractional jump-diffusion model, papers [6,10] for Black-Scholes model, and [7], for Bachelier model in the subdiffusive regime. The theory in these papers was detailed for inverse α - stable [2,6,7], inverse tempered stable [7], inverted inverse Gaussian [10], and inverted Poisson processes [2].

M. Corazza et al. (Eds.): MAF 2024, *Mathematical and Statistical Methods for Actuarial Sciences and Finance*, pp. 286–291, 2024.
https://doi.org/10.1007/978-3-031-64273-9_47

In the paper, we consider the geometric Brownian Motion (BM) model in a subdiffusion regime and propose a procedure for evaluating value-at-risk in the studied model, time-changed by a Gamma subordinator. From this perspective, our study closes a gap that is of particular interest to investors.

The paper is organized as follows. In the second section, we present our model and describe the dynamics of the bond and underlying risk assets in the subdiffusive framework with gamma subordinator. Also, we discuss two option pricing techniques for our model. The third section is devoted to risk measuring in this model. We consider the Fractional Fokker-Planck equation (FFPE) for gamma subordinator, which describes the probability density function $w(t)$ of the sub-diffusive studied stock process and we discuss how it can be used for value-at-risk (VaR) measuring.

2 Subdiffusive GBM Model with Gamma Subordinator

Assume that the market consists of at least one riskless asset, usually called bond B_t, a risky asset with price S_t, usually called the stock, and one derivative security, usually called call option, or put option, which will have a certain payoff at a specified date in the future, depending on the values taken by the stock up to that date.

Sub-diffusion occurs if we replace the calendar time t with some stochastic process H_t, where H_t is called the inverse subordinator and means stochastic clock, operation time. For our model, we replace the calendar time t with H_t in risk-free bond motion and in classical GBM.

Then the time-changed risk-free bond has a value at time t equal to:

$$\frac{dB_{H_t}}{B_{H_t}} = rdH_t, B_0 = 1, \tag{1}$$

and the movement of the underlying risk assets S_t follows a subdiffusive geometric Brownian motion (GBM):

$$\frac{dS_{H_t}}{S_{H_t}} = \left(\mu + \frac{\sigma^2}{2}\right) dH_t + \sigma dB_{H_t}, \quad t > 0, \tag{2}$$

with solution

$$S_{H_t} = S_0 e^{\mu H_t + \sigma B_{H_t}}, \quad t > 0. \tag{3}$$

In formula (3) the standard diffusive process S_{H_t} is time-changed by some stochastic process H_t, which is called the inverse subordinator ("hitting time"). The inverse subordinator H_t is defined as

$$H_t = inf\left(\tau > 0 : G_\tau \geq t)\right), \tag{4}$$

and interpreted as the first time at which G_t hits the barrier t. Thus it is time of the first reach a certain price, which may not change for some time. The definition (4) of the inverse subordinator is based on the use of some other random process called a subordinator G_t.

The subordinator G_t is generally a non-decreasing stochastic process with stationary independent increments (Lévy process), taking value in $R+$ and having Laplace transform in the following form:

$$E\left(e^{-uG_t}\right) = e^{-t\Psi(u)}, \tag{5}$$

where $\Psi(u)$ is called Lévy exponent. The subordinator G_t is often called the "waiting" time. In this paper, we consider the Gamma process as a subordinator. It is reasonable because the gamma process is a pure-jump increasing Lévy process and thus it is a good candidate for construction waiting time.

The gamma subordinator $G_t(a, c)$, $t \geq 0$, is an Lévy process with independent gamma-distributed increments, i.e. with the Lévy measure

$$\tilde{\nu}(dx) = c\frac{e^{-ax}}{x} I_{x>0} dx \tag{6}$$

and the Laplace transform given by:

$$E\left(e^{-uG_t}\right) = \left(\frac{1}{1+ua}\right)^{ct}, a > 0, c > 0. \tag{7}$$

The main component of the model is option pricing. In this paper, we discuss two techniques for option evaluation. From paper [1,2,6,7] we know that the subdiffusive model (1, 2) is arbitrage-free, and incomplete and the fair price of the European call option with expiry date T and strike price K can be found as the discounted expected payoff under some risk-neutral measure:

$$C_H\left(S, K, T, \sigma\right) = \int_0^\infty C\left(S, K, x, \sigma\right) h_\Psi(x, T) dx \tag{8}$$

Here, $h_\Psi(x, T)$ is the PDF of $H(T)$ for the subordinator with Lévy exponent Ψ and $C(S, K, T, \sigma)$ is given by classical Black-Scholes formula. In the above Eq. (8) we can evaluate the subdiffusive call price $C(\cdot)$ by computing the integral numerically. An alternative consists of calculating the price by Monte Carlo simulations.

The second technique for option pricing in the considered model is based on a fractional version of what is called Dupire's equation. The fractional Dupire's equation (PIDE) was proposed by [2]. This equation is presented in a very general form and valid for all invertible Lévy subordinators. In PIDE the derivative concerning time was replaced by a convolution-type derivative, called Dzerbayshan-Caputo (D-C) derivative. The Dzerbayshan-Caputo (D-C) derivative depends upon the chosen kind of subordinator and Bernstein functions $f(.)$. From this PIDE in the Black and Scholes (B-S) regime, when the Brownian volatility is constant and there are no jumps, the fraction Dupire PIDE can be rewritten as

$${}^f DC_H(T, k) = -r\frac{\partial}{\partial k} C_H(T, k) + \frac{\sigma^2}{2}\frac{\partial^2}{\partial k^2} C_H(T, k), \tag{9}$$

where $k = lnK$ and

$$^{f}Du(t) = b\frac{d}{dt}u(t) + \int_0^t \frac{\partial}{\partial t}u(t-s)\nu(s)ds, \tag{10}$$

where f is Bernstein function, which admits a similar representation to the Laplace exponent of Lévy process

$$f(x) = a + bx + \int_0^{+\infty}(1 - e^{-sz})\tilde{\nu}(dz), \tag{11}$$

and $(a, b, \tilde{\nu})$ is the Lévy triplet of the Bernstein function. For gamma Lévy subordinator the Lévy triplet of the Bernstein function is $(0, 0, \tilde{\nu})$. The Lévy measure $\tilde{\nu}(s)$ is defined by (6). The tail of the Lévy measure is given by

$$\nu(s)ds = \left(a + \int_s^{+\infty}\tilde{\nu}(dx)\right)ds = ds\int_s^{+\infty}\tilde{\nu}dx = ds\int_s^{+\infty}c\frac{e^{-ax}}{x}dx. \tag{12}$$

Thus, the D–C derivative for the option pricing equation (9) can be found as

$$^{f}Du(t) = c\int_0^t\int_0^{\infty}\frac{\partial}{\partial t}u(t-s)\frac{e^{-ax}}{x}ds. \tag{13}$$

3 Risk Measuring for Subdiffusion

The value-at-risk is a quite useful tool for investors and can be used for understanding the past and making medium-term and strategic decisions for the future. On the other side, we can apply VaR for the checking model performance. For this, we can use the most important criterion of a risk management system, namely to check if the regulatory requirements are fulfilled.

VaR can be defined as α-quantile of the profit (loss) function.

Let $(\Omega, \mathcal{F}, P)$ be the probability space. The value-at-risk of level α, $0 < \alpha \leq 1$ is a probability functional, defined as α-quantile of the profit (loss) function $Y \in L(\Omega)$:

$$VaR_\alpha(Y) = W^{-1}(\alpha) = inf\{\, y \in R : \alpha \leq W(Y)\}\,, \tag{14}$$

where W is the distribution function of Y, W^{-1} is the quantile function of α, $0 < \alpha \leq 1$.

Let the time horizon coincide with the time to maturity, then the loss (profit) function of the call option with strike price K is

$$Y = Y(S) = |S - K|^{+} - c_0, \tag{15}$$

Then the value-at-risk of level α, $0 < \alpha \leq 1$ for random variable Y is

$$VaR_\alpha(Y) = VaR_\alpha(|S - K|^{+}) - c_0 \tag{16}$$

due to the translation-equivariant property of probability functional VaR.

The cumulative distribution function (CDF) for $Y(S) = |S - K|^+$ is given in [9]:

$$W_Y(y) = \begin{cases} \int_{-\infty}^{y} w_S(u + K)\, du, & y \geq 0 \\ 0, & y < 0. \end{cases} \tag{17}$$

Thus, if the time horizon coincides with the time to maturity, the value-at-risk of level α, $0 < \alpha \leq 1$ one can find as (16–17), where $w(t)$ is a solution of the celebrated Fractional Fokker-Planck equation (see for example [8]):

$$\frac{\partial w}{\partial t} = \Phi_t[-\mu \frac{\partial}{\partial x} + \frac{\sigma^2}{2} \frac{\partial^2}{\partial x^2}]w(x,t). \tag{18}$$

In this equation Φ_t is the integro-differential operator defined as

$$\Phi_t f(t) = \frac{d}{dt} \int_0^t M(t-y) f(y) dy,$$

with the memory kernel $M(t)$ defined via its Laplace transform

$$\tilde{M}(u) = \int_0^\infty e^{-ut} M(t) dt = \frac{1}{\Psi(u)}.$$

The Levy exponent for the gamma process is given by

$$\Psi_{gamma}(u) = c \log(1 + ua), \tag{19}$$

what implicates that the memory kernel $M(t)$ can be expressed as:

$$M(t) = L^{-1}\left(\frac{1}{c \log(1 + ua)}\right).$$

where $L^{-1}(f)$ is the inverse Laplace transform of the $f(t)$ function. J. Janczura and A. Wylomanska in [4] found a formula for the memory kernel $M(t)$ in the form

$$M(t) = \frac{e^{-t/a}}{c} \int_0^\infty \frac{t^{y-1}}{a^y \Gamma(y)} dy. \tag{20}$$

Thus, the formula (18) allows us to find, at least in some particular cases of parameters a, c closed-form formulas for the PDF of the sub-diffusive studied stock process. Approximated solutions $w(t)$ of (18) can be derived by the finite element method for FFPE (see for example [3]) or by the Monte Carlo techniques based on the simulation algorithm of the time-changed stock process (see the section above). Thus, the possibility of numerical computing probability density function $w(t)$ for the sub-diffusive studied stock process (with gamma subordinator) opens the way to evaluate value-at-risk (VaR) in this model.

Funding Information. Nataliya Shchestyuk acknowledges financial support from the project "Portfolio management for illiquid markets" (Dnr: 20220099) funded by the Knowledge Foundation. Svitlana Drin acknowledges financial support from the Knowledge Foundation Grant (Dnr: 20220115).

References

1. Casteli, F., Leonenko, N., Shchestyuk, N.: Student-like models for risky asset with dependence. Stochast. Anal. Appl. Ser. **35**, 452–464 (2017)
2. Donatien, H., Leonenko, N.N.: Option pricing in illiquid markets: a fractional jump-diffusion approach. J. Comput. Appl. Math. Ser. **381**, 112995 (2021)
3. Deng, W.: Finite element method for the space and time fractional Fokker-Planck equation. SIAM J. Numer. Anal. **47**, 204–226 (2008)
4. Janczura, J., Wylomańska, A.: Anomalous diffusion models: different types of subordinator distribution. ACTA Phys. Polonica B Ser. **43**, 1001–1016 (2012)
5. Kumar, A., Wylomańska, A., Poloczański, R., Sundar, S.: Fractional Brownian motion time-changed by gamma and inverse gamma process. Phys. A-Stat. Mech. Appl. **468**, 648–667 (2017)
6. Magdziarz, M.: Black-Scholes formula in subdiffusive regime. J. Stat. Phys. **136**, 553–564 (2009)
7. Magdziarz, M., Orzeł, S., Weron, A.: Option pricing in subdiffusive Bachelier model. J. Stat. Phys. **145**, 187–202 (2011)
8. Metzler, R., Klafter, J.: The random walk's guide to anomalous diffusion: a fractional dynamics approach. Phys. Rep. Ser. **339**(1), 1–77 (2000)
9. Shchestyuk, N., Tyshchenko, S.: Option pricing and stochastic optimization. In: Malyarenko, A., Ni, Y., Rancic, M., Silvestrov, S. (eds.) SPAS 2019. Springer Proceedings in Mathematics and Statistics, vol. 408, pp. 651–669. Springer, Cham (2022). https://doi.org/10.1007/978-3-031-17820-7_28
10. Shchestyuk, N., Tyshchenko, S.: Subdiffusive option price model with inverse Gaussian subordinator. Working Paper 1, School of Business, Orebro University, Sweden (2024)

A New Value-Based Investing Strategy for Portfolio Selection Which Outclasses the Benchmark

Giannicola Simari(✉)

Master in Risk Management, Pisa University, Pisa, Italy
gnsimari@gmail.com

Abstract. This paper proposes a new fundamental analysis-based strategy to build a remunerative stock portfolio. We believe that the value investing paradigm, applied with consistency and automatically without any external interference, can constitute a competitive advantage for investors. The procedure works by managing the information coming from financial statements into six filtering criteria aimed at evaluating profitability, financial condition and price convenience. As case studies, we consider three separate portfolio selections from the S&P 500 (2000–2017), the STOXX Europe 600 (2002–2017) and the S&P 100 (2001–2017) benchmarked against a passive strategy represented by the Index. The criteria proposed, invariant and irrespective of economics conditions and financial markets forecasts are able to select, ex-ante, stocks producing significantly better results than the benchmark for all the timelines considered.

Keywords: Portfolio selection · Value-based investing · Fundamental analysis · Investment Behavior

1 Introduction

Benjamin Graham is known as the father of value investing, a methodology to select assets securities that, based on fundamental analysis, appear undervalued in comparison with their intrinsic value (Graham and Dodd, 1934; Graham 1949). The crucial issue is the determination of the intrinsic value of each asset, given that this quantity is neither observable nor can it be computed in a deterministic way. Regarding this point, Oppenheimer (1984) observes that a suitable choice of stock selection rules can produce higher returns than a passive strategy. Grantham (2004) comes with seven selecting criteria, with qualitative as well as quantitative nature and Graham and Zweig (2004) suggest that it is wise not to overpay for a stock investment. Greenblatt (2005) confides in simple security selection criteria and rules-based disciplined investing strategy. Gray & Carlisle (2012) retain the quantitative approach provides practical insights into an investment strategy that eliminate behavioral errors. Kahneman (2013) demontrates the illusion of skill in investing. He explores how behavioral and cognitive psychological theories with conventional finance provide explanations for irrational behaviors. Frazzini, Kabiller and

M. Corazza et al. (Eds.): MAF 2024, *Mathematical and Statistical Methods for Actuarial Sciences and Finance*, pp. 292–296, 2024.
https://doi.org/10.1007/978-3-031-64273-9_48

Pedersen (2018) postulate that Warren Buffett's Berkshire Hathaway performance is the result of stock selection. Referring to such prescriptions, we propose a new way to process financial statements data to select a remunerative stock portfolio. The Graham pillars of intrinsic value evaluation, concentrated diversification and buying within the margin of safety, are the guiding criteria employed in this work.

In the section "Methodology" we explain the screening rules and the portfolio construction. In the section "Results" we report the main results of our analysis and in the last section "Conclusion" we prove the premises of the paper: the importance of selecting stocks which are profitable, in sound financial condition, and bought at a convenient price.

2 Methodology

2.1 Screening Rules and Portfolio Construction

The stocks entering each portfolio are selected using six criteria concerning the following financial statement areas: Profitability, Sound Financial Condition and Price Convenience.

Profitability provides information about the firm ability to generate earnings in the future. Sound Financial Condition is related to the company capacity to overcome economic cycles. Price convenience provides a margin of safety, meaning that purchasing a stock at a price sufficiently lower than its intrinsic value creates a buffer for an investment to be made at a relatively low level of risk.

We clarify that our selection strategy is not affected by some typical selection bias issues. More specifically:

a) the *forward estimation bias*: it takes places when predicted variables are used in place of historical ones. We use only historical values;
b) *the look ahead bias*: is the bias coming from data unknown at the time of the analysis (for example a revision relative to an accounting field which is published after a few months). We use only the first published data, without considering revisions or restatements;
c) the *micro & small cap sample bias*: this selection bias may raise concerns about the lack of liquidity. We focus our analysis on three major stock indices.

2.2 Two Separate Strategies

For the S&P 500 and the STOXX Europe 600 we adopt an identical approach. As far as the stock selection, the six criteria listed in the above section are applied sequentially: if a stock does not pass one criterion it is eliminated and is not checked by the successive ones. At the end, we select a maximum of ten assets (a number which is considered a good compromise between risk diversification and return concentration). If more pass the six filtering rules, only the best ten are considered; otherwise, only those satisfying all six criteria are taken into account. The K($\leq$10) stocks selected in the portfolio are equally weighted[1].

[1] All data used in the analysis are taken from Bloomberg. We processed more than 1,500,000 financial data for the S&P 500 Index, more than 1,800,000 for the STOXX Europe 600 and more than 500.000 for the SPX 100, excluding stocks with missing accounting records.

For the S&P 100 Index, we adopt a different strategy. We compute the six criteria described above for all index components, ranking them from the best to the worst for each criterion. We therefore combine all such ranks into a single score by adding them, for example if a stock is the 20th in the first criterion, 30th in the second, 15th in the third, 10th in the fourth, 5th in the fifth and 9th in the last one, its overall score is 89. This different screening approach allows for have a larger number of stocks and to answer to a frequent financial institutions requirement to invest in leading companies by market capitalization. At the end, we select a maximum of thirty assets and the K($\leq$30) stocks are equally weighted.

2.3 Strategy Evaluation

We compute the portfolio returns during a generic time period [t − 1, t] as

$$R_t = \sum_{i=1}^{K} w_i r_{i,t}$$

and to summarize the return of the strategy after n periods we use the compound annual growth rate

$$CAGR(t_0, t_n) = \left(\frac{V(t_n)}{V(t_0)} \right)^{\frac{1}{t_n - t_0}} - 1$$

We considered two different timelines:

Yearly Timeline. After the first portfolio choice (done for example in January 2003), the stock selection is updated each year (January 2004, January 2005 and so on), keeping the portfolio unchanged between two consecutive yearly revisions.

Monthly Timeline[2]**.** The strategy is the same as the previous one, with the difference that the portfolio is updated monthly (instead of yearly).

We benchmark the active strategy using a paired difference test for the null hypothesis that the period differences between the portfolio returns produced by the active and the passive strategy (buying the index), has a zero mean. The test statistic is

$$\frac{\overline{d}}{se(\overline{d})}$$

where $\overline{d}$ is the mean of d_t's and its standard error, $se(\overline{d})$, is estimated according to Newey and West (1987) to take into account the possible autocorrelation of the d_t's differences. Under the null, the test statistic in the above equation is asymptotically distributed as a standard normal.

[2] The important difference between the two timelines lies in the number of observations available for the statistical check: the yearly one (which is applied to all case studies) produces as many observations as the number of years considered; the monthly one (which is used only in the S&P 500 and the STOXX Europe 600 case studies) gives a number of observations equal to 12 × number of years.

3 Results

The S&P 500 Index: The main results of the analysis involving the S&P 500 are shown in Table 1. The p-values evidence that the proposed value investing strategy produces returns significantly higher than the benchmark irrespective of the month where the investment began. The monthly timeline evidences an average difference above 16%; the yearly timelines provide mean differences ranging from 12% to 17% (the only exception is October with 8.7%).

The higher variability shown by the monthly one is explained by the fact that the yearly timeline tends to dampen the variability with respect to returns computed over shorter periods.

The column Sign Diff. reports the number of times, in percentage, the active strategy produces a return higher than the passive one for the whole period.

Table 2 reports the CAGR value of the yearly timeline, coming from the investment started in the corresponding month. The active selection gives average returns considerably higher than the passive strategy for all the months.

The S&P 100 Index: The results about the strategy implemented on the S&P 100 Index are shown in Table 7. There is a lower mean difference between the active strategy and the benchmark, ranging from 5.4% to 6.9%, in comparison with the methodology used in the other two case studies. We easily explain this with the fact that we invested in a broader number of stocks (30 against 10) taken from a lower set of components: we expect that selecting $30/100 = 30\%$ of components tends to produce results more similar to the index than taking $10/500 = 2\%$. The general outcome is, however, the same: the active selection gives returns higher than the passive strategy irrespective of the month when an investment starts. In addition, the inclusion of a larger number of components contributes to decrease the volatility of the selected portfolio, making it quite close to the whole portfolio.

Table 2 reports the CAGR value of the yearly timeline.

4 Conclusions

In this paper we present a new way to manage the information coming from financials to compose a remunerative stock portfolio. The results coming from applying the active strategy benchmarked against a passive strategy, represented by the Index, confirm the premises of the paper: the importance of selecting stocks which are profitable, in sound financial condition, and bought at a convenient price. The criteria proposed, invariant and irrespective of economics conditions and financial markets forecasts are able to select, ex-ante, stocks producing significantly better results than the benchmark for all the timelines considered. We tested the selection strategy proposed only on the S&P 500, the STOXX Europe 600 and the S&P 100 indices for one fundamental reason, the financial statements of the companies included in these indices are timely and reliable, a necessary condition to implement the methodology.

References

Graham, B.: The Intelligent Investor. Harper & Brothers (1949)
Graham, B.: The Intelligent Investor. Harper & Row (1973)
Graham, B., Dodd, D.: Security Analysis. McGraw-Hill (1934)
Graham, B., Dodd, D.: Security Analysis, 3rd edn. McGraw-Hill (1951)
Graham, B., Zweig, J.: The Intelligent Investor, Revised edn. (4th edn.). Harper Business Essentials, New York (2004)
Grantham, J.: The Case for Quality—The Danger of Junk. GMO White Paper (2004)
Gray, W.R., Carlisle, T.E.: Quantitative Value: A Practitioner's Guide to Automating Intelligent Investment and Eliminating Behavioral Errors. Wiley, Hoboken (2012)
Frazzini, A., Kabiller, D., Pedersen, L.H.: Buffett's alpha. Financ. Anal. J. **74**(4), 35–55 (2018)
Greenblatt, J.: The Little Book That Beats the Market (2005)
Kahneman, D.: Thinking, Fast and Slow (2013)
Kanuri, S., McLeod, R.W.: Sustainable competitive advantage and stock performance: the case for wide moat stocks. Appl. Econ. **48**(52), 5117–5127 (2016)
Meehl, P.: Clinical versus statistical prediction (1996)
Newey, W.K., West, K.D.: A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. Econometrica **55**, 703–708 (1987)
Oppenheimer, H.R.: A test of Ben Graham's stock selection criteria. Financ. Anal. J. **40**, 68–74 (1984)

On the Effect of Pension Expectations and Financial Literacy on Pension Planning: A Preliminary Investigation for the Italian Population

Rosaria Simone(✉) and Mariarosaria Coppola

University of Naples Federico II, 80138 Naples, Italy
{rosaria.simone,m.coppola}@unina.it

Abstract. Pension reforms are on the agenda of several governments worldwide, especially those experiencing a serious longevity risk, like Italy, due to the combination of ageing of the population and declining fertility rates. As a result, younger generations will have to cope with late pension age and possibly lower pension incomes, and individuals may opt to subscribe private pensions to sustain their expectations, in terms of retirement age and pension benefits. Propensity to private pension planning depends heavily on financial literacy, as highlighted in the literature (see [3–5], among others). In this context, for the Italian population we propose to resort to model-based regression trees [8] to highlight individuals' features that entail different effects of pension expectations and financial literacy on propensity to pension planning.

Keywords: Model-based Regression Trees · Logistic regression · Propensity to pension planning · Financial Literacy · Pension expectations

1 Motivation and Methodology

The goal of this preliminary investigation is the analysis of the effect that expectation of pensions (in terms of both age and income) and financial literacy have on retirement planning in Italy. With respect to the state of the art, our approach is meant to challenge the ideas and the results of [2] concerning the UK population, to the case of Italy, by means of an original statistical analysis able to investigate which individuals' features possibly modify the alleged effects that pension expectations and financial literacy have on private pension plan participation. To pursue this goal, we resort to the setting of model-based regression trees [6,8], since it allows to disclose if and how a given maintained model fits and explains the relationship between a response and a set of predictors, given the values of partitioning variables that are selected automatically by the procedure according to some optimization criteria. The result is a tree where each node, corresponding to a given set of observations, is characterized by a locally optimal specification of the maintained model, thus enhancing its interpretation.

M. Corazza et al. (Eds.): MAF 2024, *Mathematical and Statistical Methods for Actuarial Sciences and Finance*, pp. 297–302, 2024.
https://doi.org/10.1007/978-3-031-64273-9_49

2 Investigating Pension Planning in Italy

To focus on the Italian population, we resort to a selected set of variables collected within the Survey of Household Income and Wealth (SHIW) [1] run by the Bank of Italy in 2020, for a total of $n = 2955$ observations corresponding to a set of employed or self-employed individuals. In particular, we will consider the following variables:

- `pens_ins`: responses to the binary question: *'Have you recently subscribed a private pension insurance to supplement your future pension income?'*
- `exp_income`: responses to the question: *'What percentage of the current income do you expect to be represented by your pension income?'*
- `exp_age`: responses to the question: *'What do you expect will be your retirement age?'*
- `finlit`: variable counting the number of correct responses to the 3 financial literacy questions provided by the respondent, mimicking the construction of the so-called Lusardi-Mitchell score [5].

Our preliminary investigation will consist of separate logistic regression analyses to study the effects that pension expectations, with respect both to income and age, exert on pension planning in terms of subscription of private pensions. Similarly, we will adopt the same methodology to investigate the effect that financial literacy has on pension planning. The logics of model-based trees will be exploited by assuming as maintained models the logistic regressions[1]:

1. $\texttt{pens\_ins}_i \sim \texttt{exp\_income}_i$ (see Subsect. 2.1);
2. $\texttt{pens\_ins}_i \sim \texttt{exp\_age}_i$ (see Subsect. 2.2);
3. $\texttt{pens\_ins}_i \sim \texttt{finlit}_i$ (see Subsect. 2.3).

The choice of growing a tree for each predictor is meant to identify its individual effect on the response, conditionally to partitioning variables. Comparative results with a tree, based on the logistic regression of the response against both expectation variables and financial literacy indicators, will be discussed in the concluding section[2].

2.1 The Effect of Expected Pension Income on Pension Planning

Figure 1 displays the resulting tree based on the logistic regression $\texttt{pens\_ins}_i \sim \texttt{exp\_income}_i$. The main differences is found between self-employed and not-self employed individuals. Among the latter, being or not home owners makes a difference. Further, for home owners there is a difference in model parameters on the basis of the geographical area of residence (the discriminating factor is residence in the South of Italy or not). Estimated parameters of the maintained model are reported in Table 1: coefficients that correspond to significant effects at the 5% level are highlighted in bold.

[1] For short, we will use the notation $Y_i \sim X_i$ to denote that we run a regression of the response Y_i against predictor X_i.

[2] Hereafter this *omnibus* tree will not be presented for the sake of brevity.

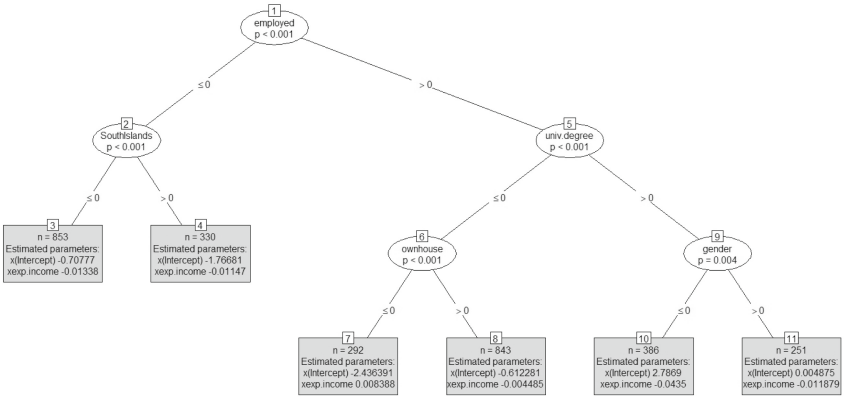


Fig. 1. Model-based tree for the logistic regression $\texttt{pens\_ins}_i \sim \texttt{exp.income}_i$

Table 1. Regression coefficients estimated locally for the logistic regression $\texttt{pens\_ins}_i \sim \texttt{exp.income}_i$

	Leaves (Terminal Nodes)					
	3	4	7	8	10	11
Intercept	−0.708	−1.767	−2.436	−0.612	2.787	0.005
exp.income	**−0.013**	−0.011	0.008	−0.004	**−0.043**	−0.012

2.2 The Effect of Expected Pension Age on Pension Planning

Figure 2 displays the resulting model-based tree based on the regression $\texttt{pens\_ins}_i \sim \texttt{exp\_age}_i$. The main differences is found between self-employed and not-self employed individuals. Among the latter, having or not a university degree slightly modifies the maintained model for individuals belonging to the boomer generation. It turns out that relevant gender differences in the effect that expected pension age exerts on pension planning pertain only to employed individuals belonging to the age generation X having a university degree (nodes 14 and 15).

Estimated parameters of the maintained model, for each tree leaf, are reported in Table 2: coefficients that correspond to significant effects at the 5% level are highlighted in bold.

2.3 The Effect of Financial Literacy on Pension Planning

As a final step in our investigation, we apply the logistic-based regression trees for the regression $\texttt{pens\_ins}_i \sim \texttt{finlit}_i$ in order to disclose locally the effect of financial literacy on pension planning [7]. Results indicate that, overall, higher financial literacy scores induce higher propensity to pension planning, but to

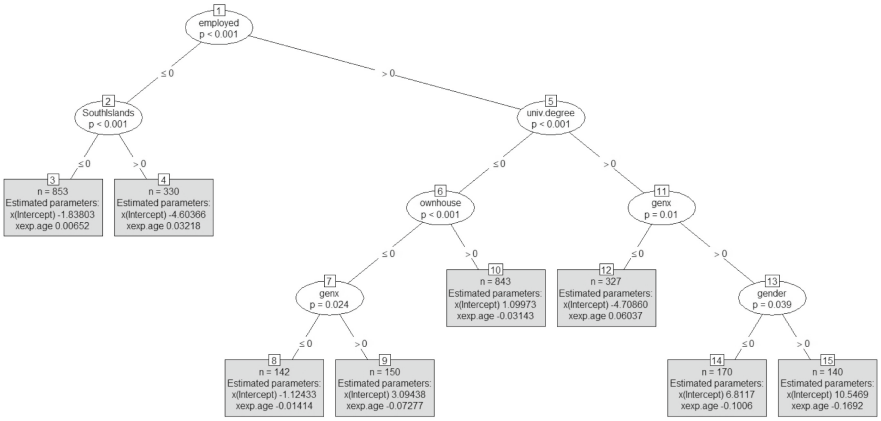


Fig. 2. Model-based tree for the logistic regression $\texttt{pens\_ins}_i \sim \texttt{exp\_age}_i$

Table 2. Regression coefficients estimated locally for the logistic regression $\texttt{pens\_ins}_i \sim \texttt{exp\_age}_i$

	Leaves (Terminal nodes)							
	3	4	8	9	10	12	14	15
Intercept	−1.838	−4.604	−1.124	3.094	1.100	−4.709	6.812	10.547
exp.age	0.007	0.032	−0.014	−0.073	−0.031	0.060	**−0.101**	**−0.169**

different extent. The effect of financial literacy on the chosen indicator of pension planning is statistically significant (at the 5% level) only for the response profiles corresponding to nodes 2,6,9,10,15. In this respect, it is worth noticing that, for employed people owning a house, having a university degree entails statistically significant variations of the model parameters differently across Italy: in the South and Islands, indeed, financial literacy has a significant effect on pension plan participation only for employed individuals that are home owners belonging to the generation X (node 15) (Table 3 and Fig. 3) .

Table 3. Locally Estimated coefficients for the model-based trees based on the logistic regression $\texttt{pens\_ins}_i \sim \texttt{finlint}_i$

	Leaves (Terminal nodes)							
	3	4	6	9	10	12	14	15
x(Intercept)	−2.658	−3.211	−2.991	−1.617	−1.160	−2.043	−1.715	−3.592
finlit	**0.515**	0.358	**0.709**	**0.390**	**0.409**	0.297	0.298	**1.287**

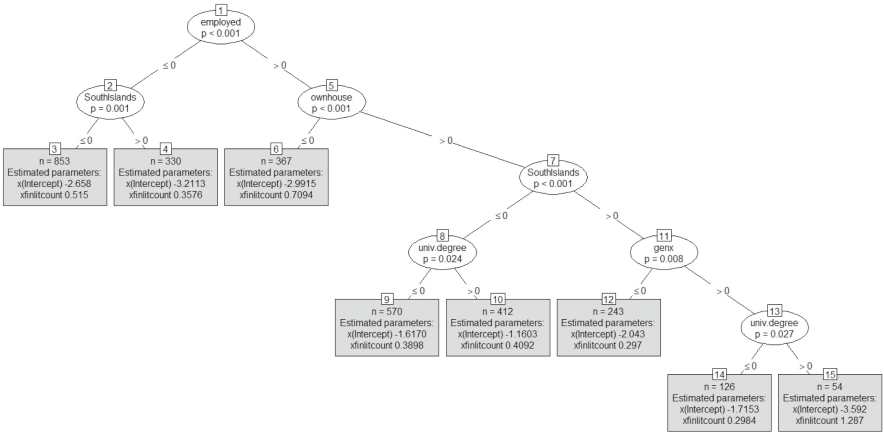


Fig. 3. Model-based tree for the logistic regression $\texttt{pens\_ins}_i \sim \texttt{finlit}_i$

Table 4. Goodness of fit of the fitted logistic regression trees

	Predictors				
	`exp.income`	`exp.age`	`exp.income` + `exp.age`	`finlit`	`exp.income` + `exp.age` + `finlit`
Loglik	−1523.545	−1542.483	−1517.688	−1497.508	**−1476.673**
BIC	3182.942	3268.765	3219.174	**3178.815**	3185.092

3 Concluding Remarks

In conclusion, it is important to establish the advantages of the proposed approach with respect to more standard ones. As a benchmark, consider that `BIC` = 3205.356 and `loglik` = −1586.696 for the logistic regression of the response using expectation variables and financial literacy indicators as predictors on the whole dataset. Table 4 reports the overall log-likelihood and BIC values of the fitted trees; for completeness, we report also results for the model-based tree based on the logistic regression with only expectation variables as predictors, as well as both expectation variables and financial literacy indicators. Overall, results clearly indicate advantages in goodness of fit and explanatory power entailed by the setting of model-based trees with respect to standard regression. In terms of covariates effects, it turns out that expectations on pension income are a stronger leverage to boost private pension planning than expectations on retirement age, yet financial literacy itself provides the best trade-off between model-complexity and goodness of fit. As a consequence, promoting financial literacy with respect to pensions should be considered a key strategy for the sustainability of reforms for National Social Security Systems.

The preliminary investigation presented here will be completed with the assessment of prediction performances: to this aim, ensemble tree methods, as

random forests - in both parametric and non-parametric settings - should be considered as competitive methods.

References

1. Faiella, I., Gambacorta, R.: The weighting process in the SHIW, Bank of Italy. Economic Working Paper No. 636 (2007)
2. Farrar, S., Moizer, J., Lean, J., Hyde, M.: Gender, financial literacy, and pre-retirement planning in the UK. J. Women Ageing **31**(4), 319–339 (2018)
3. Fornero, E., Monticone, C.: Financial literacy and pension plan participation in Italy. J. Pension Econ. Financ. **10**(4), 547–564 (2011)
4. Kaiser, T., Lusardi, A., Menkhoff, L., Urban, C.: Financial education affects financial knowledge and downstream behaviors. J. Financ. Econ. (2021). https://doi.org/10.1016/j.jfineco.2021.09.022
5. Lusardi, A., Mitchell, O.S.: How ordinary consumers make complex economic decisions: financial literacy and retirement readiness. Working paper No. 15350 of National Bureau of Economic Research (2009). https://doi.org/10.3386/w15350
6. Hothorn, T., Zeileis, A.: partykit: a modular toolkit for recursive partytioning in R. J. Mach. Learn. Res. **16**, 3905–3909 (2015)
7. Von Rooij, M.C.J., Lusardi, A., Alessie, R.J.H.: Financial Literacy, Retirement Planning and Household Wealth. Econ. J. **122**, 449–478 (2012)
8. Zeileis, A., Hothorn, T., Hornik, K.: Model-based recursive partitioning. J. Comput. Graph. Stat. **17**(2), 492–514 (2008)

Author Index

M. Corazza et al. (Eds.): MAF 2024, *Mathematical and Statistical Methods for Actuarial Sciences and Finance*, pp. 303–304, 2024.
https://doi.org/10.1007/978-3-031-64273-9