Marco Corazza
María Durbán
Aurea Grané
Cira Perna
Marilena Sibillo
Editors

Mathematical and Statistical Methods for Actuarial Sciences and Finance

MAF 2018

Springer

Mathematical and Statistical Methods for Actuarial Sciences and Finance

Marco Corazza • María Durbán • Aurea Grané •
Cira Perna • Marilena Sibillo
Editors

Mathematical and Statistical Methods for Actuarial Sciences and Finance

MAF 2018

Springer

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

María Durbán
Dept. of Statistics
Universidad Carlos III de Madrid
Getafe
Madrid, Spain

Aurea Grané
Dept. of Statistics
Universidad Carlos III de Madrid
Getafe
Madrid, Spain

Cira Perna
Dept. of Economics and Statistics
University of Salerno
Fisciano
Salerno, Italy

Marilena Sibillo
Dept. of Economics and Statistics
University of Salerno
Fisciano
Salerno, Italy

ISBN 978-3-319-89823-0 ISBN 978-3-319-89824-7 (eBook)
https://doi.org/10.1007/978-3-319-89824-7

Library of Congress Control Number: 2018906933

Preface

This volume gathers selected short papers presented at the international conference Mathematical and Statistical Methods for Actuarial Sciences and Finance (MAF), held at Universidad Carlos III de Madrid (Spain) on 4–6 April 2018.

The contributions highlight new ideas on mathematical and statistical methods in actuarial sciences and finance. The interaction between mathematicians, statisticians and econometricians working in insurance and finance is a very fruitful field that yields unique theoretical models and practical applications, as well as new insights in the discussion of problems of national and international interest.

The book covers a wide variety of subjects in actuarial science and financial fields, all of which are treated in light of the cooperation between the three quantitative approaches. The topics are: actuarial models; analysis of high-frequency financial data; behavioural finance; carbon and green finance; credit risk methods and models; dynamic optimization in finance; financial econometrics; forecasting of dynamical actuarial and financial phenomena; fund performance evaluation; insurance portfolio risk analysis; interest rate models; longevity risk; machine learning and soft computing in finance; management in insurance business; models and methods for financial time series analysis; models for financial derivatives; multivariate techniques for financial markets analysis; optimization in insurance; pricing; probability in actuarial sciences, insurance and finance; real world finance; risk management; solvency analysis; sovereign risk; static and dynamic portfolio selection and management; and trading systems.

The MAF Conference was first held in Salerno in 2004 and is organized biennially. Previous meetings were held in Salerno (2004, 2006, 2010, 2014), Venice (2008, 2012) and Paris (2016).

Madrid, Spain
February 2018

Aurea Grané
María Durbán

Contents

About the Editors

Marco Corazza has a PhD in "Mathematics for the Analysis of Financial Markets" and is an associate professor at the Department of Economics of the Ca' Foscari University of Venice (Italy). His main research interests include static and dynamic portfolio management theories; trading system models; machine learning applications in finance; bio-inspired optimization techniques; multi-criteria methods for economic decision support; port scheduling models and algorithms; and non-standard probability distributions in finance. He has participated in several research projects, at both the national and international level, and is the author/coauthor of one hundred and twenty scientific publications, some of which have received national and international awards. He is also Editor-in-Chief of the international scientific journal "Mathematical Methods in Economics and Finance", and is a member of the scientific committees of several conferences and of some private companies. His combines his academic activities with consulting services.

María Durbán is a professor of Statistics at Universidad Carlos III de Madrid (Spain). Her main areas of research are non-parametric regression, smooth mixed models and regression models for spatio-temporal data. She has numerous publications in these topics and their application in areas such as epidemiology, economics, and environmental sciences. She has been part of many scientific committees of international conferences.

Aurea Grané is a professor of Statistics at Universidad Carlos III de Madrid (Spain). Her research interests are mainly in goodness-of-fit, multivariate techniques for mixed-type data, functional data analysis and she has published numerous papers on these topics in international journals. She has been a member of several scientific committees of international conferences, and was co-director of the Master in Quantitative Techniques for the Insurance Sector and vice-director of the Department of Statistics at Universidad Carlos III de Madrid.

Cira Perna is a professor of Statistics and head of the Department of Economics and Statistics, University of Salerno (Italy). Her research mainly focuses on non-linear time series, artificial neural network models and resampling techniques, and she has published numerous papers on these topics in national and international journals. She has been a member of several scientific committees of national and international conferences.

Marilena Sibillo is a professor of Mathematical Methods for Economics, Finance and Actuarial Sciences at the University of Salerno (Italy). She has several international editing engagements and is the author of over a hundred publications. Her research interests are mainly in longevity risk in life contracts, de-risking strategies, personal pension products and mortality forecasting.

The Effect of Rating Contingent Guidelines and Regulation Around Credit Rating News

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

Abstract This paper investigates the effect of rating-based portfolio restrictions that many institutional investors face on the trading of their bond portfolios. Particularly, we explore how credit rating downgrades affect to bondholders that are subject to such rating-based constrains in the US corporate bond market. We go beyond the well-documented investment grade (IG) threshold by analyzing downgrades crossing boundaries usually used in investment policy guidelines. We state that the informativeness of rating downgrades will be different according to whether they imply crossing investment-policy thresholds or not. We analyze corporate bond data from the TRACE dataset to test our main hypothesis and find a clear response around the announcement date consistent with portfolio adjustments made by institutions in their fulfillment of investment requirements for riskier assets.

Keywords Credit rating · Investment guidelines · Corporate bonds

1 Introduction

The vast majority of institutional investors are subject to credit rating restrictions, either in their investment decisions or in the holding of deteriorated rating securities. One example of regulatory restrictions based on ratings is the National Association of Insurance Commissioners (NAIC)'s risk-based capital system for insurance companies. This framework is based on the credit ratings provided by the most relevant

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

A. Díaz and A. Escribano (✉)
University of Castilla-La Mancha, Albacete, Spain
e-mail: antonio.diaz@uclm.es; ana.escribano@uclm.es

M. D. Robles
University Complutense of Madrid and ICAE, Madrid, Spain
e-mail: mdrobles@ccee.ucm.es

M. Corazza et al. (eds.), *Mathematical and Statistical Methods for Actuarial Sciences and Finance*, https://doi.org/10.1007/978-3-319-89824-7_1

credit rating agencies (CRAs), and establishes six credit quality designations of bonds that correspond to different rating scales (from AAA to A, BBB, BB, B, CCC, and below CCC). The system imposes higher capital requirements to those investment portfolios that contain bonds rated in the lower credit quality categories. Other example of non-regulatory restrictions can be found in the internal investment procedures from some portfolio managers. They use a wide variety of rating-based guidelines that potentially affect their portfolio investments decisions. In these cases, the investment policy statements provide restrictions to the investment process depending on the credit rating that bear the target assets.

The prior literature has focused in the analysis of the well-documented investment grade (IG) threshold, by analyzing the effects of credit rating changes (CRCs) crossing the IG frontier on the investors' portfolios subject to restrictions in the holding of junk bonds. Some examples are the papers of [4] or [1], who analyze the effects of fallen angels that may lead to the fire sales and/or price-pressure effects.

In this paper, we explore how credit rating downgrades (CRDs) crossing boundaries usually used in investment policy guidelines affect to bondholders that are subject to such investment constrains. We state that their responses to bond rating adjustments that lead to jumps across rating thresholds, should be different than those downgrades that do not involve them. We compare the effect of downgrades with implications in the boundaries investment guidelines with their reaction to other type of downgrades, such as downgrades across the IG threshold (fallen angels) and downgrades without regulatory implications ("standard" downgrades). In addition, we control for CRDs that cross any one of the buckets set on the NAIC's system.

We study a comprehensive sample of 2082 CRCs in the whole US corporate bond market, using transaction data from the Trade Reporting and Compliance Engine (TRACE) dataset and rating information from the Fixed Income Securities Database (FISD) during the period from July 2002 to December 2014.

Our findings shows a clear effect around downgrades that could have implications in the investment guidelines used by many portfolio managers. Accordingly, we highlight a clear policy implication of our results. Moreover, our paper relates to those strands of the literature that documents the forced-selling phenomenon and price pressure effect on the corporate bond market (e.g., [1, 4]). Besides, to the literature that studies the information content of CRCs (e.g., [6, 7]).

The flow of our paper is organized as follows: Sect. 2 introduces rating-based regulation and the main hypotheses. Section 3 presents the data and the analysis used. Section 4 shows the main results. Finally, Sect. 5 concludes.

2 Rating-Based Investment Guidelines

Most of the portfolio managers use a wide variety of rating-based guidelines that potentially affect their portfolio investments and performance tracking. Among other goals, these investment policy (IP) statements serve as a policy guide and

provide discipline to the investment process. In this line, [3] observe that both a threshold of A and the IG cut-off of BBB/Baa are usually used in the US fund managers guidelines (80% and 88% respectively). Moreover, [5] list a number of rating-based regulations beyond the IG cutoff. For instance, the AA/Aa rating has been used in the mortgage-backed security market, the SEC Rule 2a-7 stated that money market mutual funds are required to limit investments rated less than A+/A1, or the Department of Labor instituted a regulation limiting pension fund investment in asset-backed securities rated A or better. Similar guidelines are used by banks, bondholders, and other fiduciary agents, although these thresholds have received little attention.

Our main hypotheses is that all CRDs crossing boundaries usually used in IP guidelines will carry rigged a stronger reaction than other downgrades that do not involve any constraints. Most portfolio managers use a wide variety of rating-based guidelines that potentially affect portfolio investments and performance tracking. The credit deterioration of bonds included in their portfolios will induce them to unwind positions to fulfill their investment guidelines, although they are not forced to sell. This response should be stronger than the market reaction to CRDs that do not involve any rating constraints, since the consequences of this lasts downgrades are mainly of an economic risk nature.

3 Data and Analysis Implementation

We use two main sources of data. First, the TRACE database that collects data of corporate bond transactions in the secondary market and includes, among others, information about trading prices and volumes. Second, the FISD dataset that includes information about the main features of the securities as well as rating history information. After filtering and matching both datasets, we select only those CRDs that correspond to straight bonds and that meet certain imposed criteria about trading frequency. We end up with a sample of 2082 CRCs involving 1250 bonds from 245 issuers, covering the period from July 2002 to December 2014.

To test the hypothesis in Sect. 2, we run different OLS regressions where the dependent variable is either the abnormal value of the transaction price (PR), the yield spread (YS), the trading volume (TV) or the number of trades (NT). We control for jumps across NAIC's buckets[1] in addition to other features such as the final rating, the jump size, the CRA, the negative watch-lists, the coupon or the age.

[1] NAIC is a dummy variable equal to 1 if the initial and the final rating are in different NAIC's buckets and 0 else. Around 70% of the CRCs do not imply a jump between NAIC's buckets.

Table 1 Results from the OLS regressions

Variable	PR	PR	YS	YS	TV	TV	NT	NT
c	0.036**	0.047***	0.422	0.606*	2.040***	1.858***	0.465**	0.480**
NAIC		0.057***		−0.122		−1.067***		−0.385**
Policy		−0.052***		1.182***		1.068**		0.801***
Final rating	0.000	0.003	0.012	−0.107***	−0.051	−0.123**	0.011	−0.059**
Jump size	−0.018**	−0.030***	0.316*	0.321**	0.252	0.477**	0.084	0.157***
CRA	−0.042***	−0.032***	0.247	0.158	0.547**	0.366	0.367***	0.275**
Watch negative	−0.033***	−0.038***	0.830***	0.778***	0.360	0.456	0.399***	0.409***
Coupon	−0.004*	−0.005**	−0.174***	−0.140*	−0.174*	−0.151	−0.079**	−0.059*
Age	0.038***	0.037***	0.561	0.474	0.076	0.086	−0.048	−0.079
Adj. R-squared	0.094	0.131	0.039	0.055	0.008	0.015	0.024	0.041

***, ** and * indicates significance at 1%, 5% and 10% levels respectively

4 Results

The main results are displayed in Table 1. The dependent variable in each model is the abnormal value of the analyzed variable within the period of working days [0,5], i.e. during 1 week after the announcement day.[2] We can observe that when controlling for rating-based constraints (*Policy* variable) the explanatory power of the models increases since the adjusted R^2 is larger in all models. Moreover, the coefficients for this variable present the expected signs and are highly significant in all models. As preliminary results, these findings suggests that trading activity seems to be trigger by rating-based restrictions.

5 Concluding Remarks

Rating-based regulation and constraints seems to trigger abnormal trading activity around CRCs in the US corporate bond market. According to our findings, we highlight a clear policy implication of our results. A finer granularity both in the rated-based contingent guidelines and in the NAIC's capital requirement buckets should mitigate market overreactions and fire sales among investors.

Acknowledgements This work was supported by the Spanish Ministerio de Ciencia y Tecnología (ECO2015-68367-R and ECO2015-67305-P, ECO2014-59664-P) and Junta de Comunidades de Castilla-La Mancha (PEII-2014-019-P). Any errors are solely the responsibility of the authors.

References

1. Ambrose, B.W., Cai, N., Helwege, J.: Forced selling of fallen angels. J. Fixed Income **18**, 72–85 (2012)
2. Bessembinder, H., Kahle, K.M., Maxwell, W.F., Xu, D.: Measuring abnormal bond performance. Rev. Financ. Stud. **22**(10), 4219–4258 (2009)
3. Cantor, R., Gwilym, O.A., Thomas, S.H.: The use of credit ratings in investment management in the US and Europe. J. Fixed Income **17**(2), 13–26 (2007)
4. Ellul, A., Jotikasthira, C., Lundblad, C.T.: Regulatory pressure and fire sales in the corporate bond markets. J. Financ. Econ. **101**(3), 596–620 (2011)
5. Kisgen, D., Strahan, P.: Do regulations based on credit ratings affect a firm's cost of capital? Rev. Financ. Stud. **23**(12), 4324–4347 (2010)
6. May, A.D.: The impact of bond rating changes on corporate bond prices: new evidence from the over-the-counter market. J. Bank. Financ. **34**(11), 2822–2836 (2010)
7. Steiner, M., Heinke, V.G.: Event study concerning international bond price effects of credit rating actions. Int. J. Financ. Econ. **6**, 139–157 (2001)

[2] Abnormal values are computed comparing the observed values with the values of each variable during the control period that we set on the window of days [−41,−21] prior to the event day (t = 0). According to [2], the use of the firm-specific past history as a benchmark is a good alternative although it could be less powerful. Creating a matching portfolio of stable rating bonds with characteristics as similar as possible to the re-rated bond is almost infeasible in corporate bond markets due to infrequent trading observed even in the most liquid case such as the U.S. market.

Practical Problems with Tests of Cointegration Rank with Strong Persistence and Heavy-Tailed Errors

Niklas Ahlgren and Paul Catani

Abstract Financial time series have several distinguishing features which are of concern in tests of cointegration. An example considered in this paper is testing the approximate non-arbitrage relation between the credit default swap (CDS) price and bond spread. We show that strong persistence and very high persistence in volatility are stylised features of cointegrated systems of CDS prices and bond spreads. There is empirical support that the distribution of the errors is heavy-tailed with infinite fourth moment. Tests for cointegration have low power under such conditions. The asymptotic and bootstrap tests are unreliable if the errors are heavy-tailed with infinite fourth moment. Monte Carlo simulations indicate that the wild bootstrap (WB) test may be justified with heavy-tailed errors which do not have finite fourth moment. The tests are applied to CDS prices and bond spreads of US and European investment-grade firms.

Keywords ARCH · Cointegration · Credit default swap · Heavy tails · Wild bootstrap

1 Introduction

Financial time series have several distinguishing features which are of concern in tests of cointegration. An example considered in this paper is the approximate non-arbitrage relation between the credit default swap (CDS) price and bond spread (see [1]). The paper investigates the power of asymptotic, bootstrap and wild bootstrap (WB) tests of cointegration rank with strong persistence, very high persistence in volatility and heavy-tailed errors. The tests are applied to CDS prices and bond spreads of US and European investment-grade firms.

N. Ahlgren (✉) · P. Catani
Hanken School of Economics, Helsinki, Finland
e-mail: niklas.ahlgren@hanken.fi; paul.catani@hanken.fi

M. Corazza et al. (eds.), *Mathematical and Statistical Methods for Actuarial Sciences and Finance*, https://doi.org/10.1007/978-3-319-89824-7_2

2 Power of Tests in the Heteroskedastic VAR Model with Heavy-Tailed Errors

We consider the p-dimensional heteroskedastic cointegrated VAR model of Cavaliere et al. [3]:

$$\Delta \mathbf{y}_t = \alpha\beta' \mathbf{y}_{t-1} + \sum_{i=1}^{k-1} \boldsymbol{\Gamma}_i \Delta \mathbf{y}_{t-i} + \alpha\rho' D_t + \phi d_t + \varepsilon_t, \quad t = 1, \ldots, T. \tag{1}$$

The deterministic variables satisfy $D_t = 1$ and $d_t = 0$. The errors $\{\varepsilon_t\}$ are a martingale difference sequence with respect to the filtration $\mathscr{F}_{t-1} = \sigma(\varepsilon_{t-1}, \varepsilon_{t-2}, \ldots)$, and satisfy (1) the global homoskedasticity condition

$$\frac{1}{T} \sum_{t=1}^{T} \mathrm{E}(\varepsilon_t \varepsilon_t' | \mathscr{F}_{t-1}) \to \boldsymbol{\Sigma}, \quad \text{in probability,} \tag{2}$$

and (2) $\mathrm{E} \|\varepsilon_t\|^4 \leq K < \infty$, where $\|\cdot\|$ denotes the norm.

The tests that we consider are the asymptotic $Q_{r,T}$ test [4], bootstrap $Q_{r,T}^{*B}$ test and WB $Q_{r,T}^{*WB}$ test of cointegration rank r. We use the bootstrap algorithms of Cavaliere et al. [2, 3] which recursively generate bootstrap and WB observations.

Figure 1 shows the simulated power functions for $p = 2$ and $T = 1000$ in the VAR(2) model. The nominal level is 5%. The data-generation process (DGP)

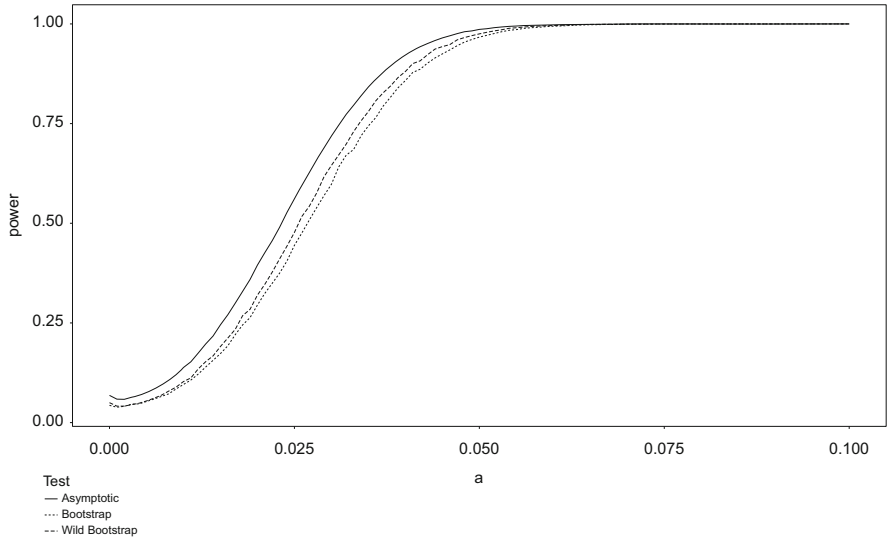


Fig. 1 Simulated power functions for $p = 2$, $T = 1000$, CCC-GARCH(1, 1) errors with $a_{11} = a_{22} = 0.08$, $b_{11} = b_{22} = 0.9$ and conditional correlation coefficient 0

is a VAR(1) model with $\boldsymbol{\alpha} = (-a, 0)'$ and $\boldsymbol{\beta} = (1, 0)'$. The model for the errors is a constant conditional correlation generalised autoregressive conditional heteroskedasticity (CCC-GARCH) model with ARCH parameters $a_{11} = a_{22} = 0.08$, GARCH parameters $b_{11} = b_{22} = 0.9$ and conditional correlation coefficient 0. The CCC-GARCH process is characterised by very high persistence in volatility ($a_{ii}+b_{ii} = 0.98, i = 1, 2$); satisfies the stationarity condition (the largest eigenvalue is 0.98); satisfies the fourth moment condition (the largest eigenvalue is 0.9732). The simulated size ($a = 0$) of $Q_{0,T}$ is 6.8%, the simulated size of $Q_{0,T}^{*B}$ is 4.3% and the simulated size of $Q_{0,T}^{*WB}$ is 5.0%. The simulated powers when the largest stationary root is 0.98 ($a = -0.02$) are 39.5%, 29.5% and 32.0%, respectively; when it is 0.97 are 71.9%, 59.8% and 64.5%, respectively; and when it is 0.95 are 98.6%, 96.7% and 97.5%, respectively. Figure 2 shows the simulated power functions with ARCH parameters $a_{11} = 0.175$, $a_{22} = 0.35$, GARCH parameters $b_{11} = 0.824$, $b_{22} = 0.5$ and conditional correlation coefficient 0. The CCC-GARCH process is characterised by very high persistence in volatility in the first equation ($a_{11} + b_{11} = 0.999$), moderate persistence in the second equation ($a_{22} + b_{22} = 0.85$) and large ARCH parameters; satisfies the stationarity condition (the largest eigenvalue is 0.999); does not satisfy the fourth moment condition (the largest eigenvalue is 1.0593). The asymptotic $Q_{0,T}$ test and bootstrap $Q_{0,T}^{*B}$ test are oversized if the condition for the existence of the fourth moment of the errors is not satisfied. The simulated size of the former is 12.6% and the latter 9.2%. The violation does not have an effect on the size of the WB $Q_{0,T}^{*WB}$ test. The simulated size of $Q_{0,T}^{*WB}$ is 5.0%. The simulated powers

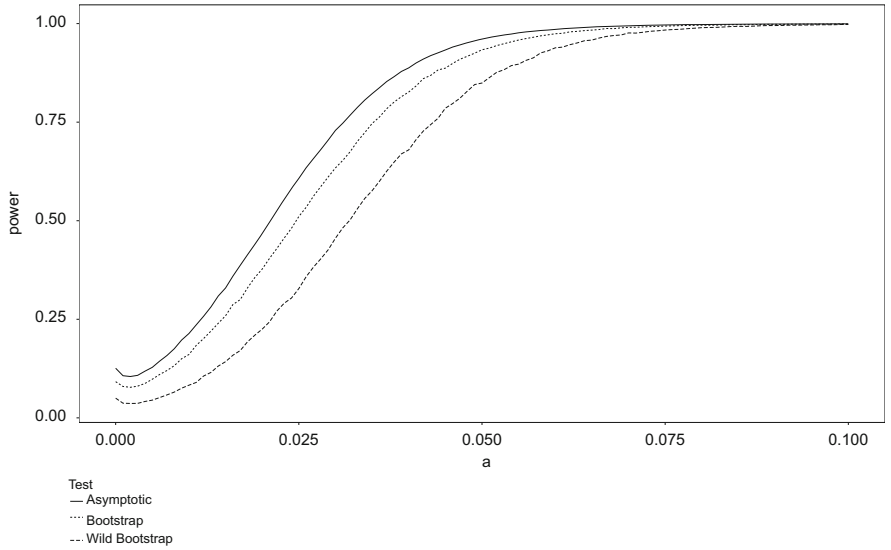


Fig. 2 Simulated power functions for $p = 2$, $T = 1000$, CCC-GARCH(1, 1) errors with $a_{11} = 0.175$, $a_{22} = 0.35$, $b_{11} = 0.824$, $b_{22} = 0.5$ and conditional correlation coefficient 0

Table 1 WB tests of cointegration between the CDS prices and bond spreads[a]

	$Q_{0,T}^{*WB}$	$Q_{1,T}^{*WB}$	$Q_{0,T}^{*WB}$	$Q_{1,T}^{*WB}$	$Q_{0,T}^{*WB}$	$Q_{1,T}^{*WB}$
Company	1.1.2010–31.8.2016		1.1.2010–1.5.2013		2.5.2013–31.8.2016	
Bank of America	0.028	0.491	0.104	0.466	0.341	0.241
Barclays Bank	0.692	0.597	0.149	0.573	0.708	0.850
Citigroup	0.016	0.795	0.105	0.897	0.001	0.431
Deutsche Telekom	0.388	0.911	0.386	0.710	0.829	0.558
Goldman Sachs	0.039	0.770	0.140	0.637	0.189	0.460
Morgan Stanley	0.000	0.212	0.000	0.181	0.056	0.564
Telefonica	0.004	0.837	0.009	0.401	0.607	0.785
Vodafone	0.001	0.108	0.022	0.254	0.258	0.501

[a]The table shows the p-values of the tests

of $Q_{0,T}$, $Q_{0,T}^{*B}$ and $Q_{0,T}^{*WB}$ when the largest stationary root is 0.98 are 46.6%, 37.7% and 22.5%, respectively; when it is 0.97 are 72.9%, 63.4% and 45.6%, respectively; and when it is 0.95 are 96.1%, 93.4% and 84.9%, respectively.

3 Empirical Results

The tests are applied to daily observations from 2010 to 2016 on the CDS price and bond spread of US and European investment-grade firms. The empirical results support strong persistence and very high persistence in volatility. Hill estimates for the standardised residuals from the VAR models indicate that the distribution of the errors has heavy tails with finite variance $\mathrm{E}|\varepsilon_{it}^2| < \infty$ but infinite fourth moment $\mathrm{E}|\varepsilon_{it}^4| = \infty$, $i = 1, 2$. We therefore use the WB test. The WB tests reported in Table 1 accept cointegration for most firms in the full sample period ($T = 1739$). The evidence for cointegration is weak in sub-sample periods from 2010 to 2013 ($T = 869$) and 2013 to 2016 ($T = 870$).

4 Conclusions

Tests of cointegration rank have low power in cointegrated systems with strong persistence and very high persistence in volatility. Obtaining high power requires more than 1000 observations, or more than 4 years of daily observations. The asymptotic and bootstrap tests are unreliable if the errors are heavy-tailed with infinite fourth moment. Monte Carlo simulations indicate that the WB test may be justified with heavy-tailed errors which do not have finite fourth moment.

References

1. Blanco, R., Brennan, S., Marsh, I.W.: An empirical analysis of the dynamic relation between investment-grade bonds and credit default swaps. J. Finance **60**, 2255–2281 (2005)
2. Cavaliere, G., Rahbek, A., Taylor, A.M.R.: Bootstrap determination of the co-integration rank in vector autoregressive models. Econometrica **80**, 1721–1740 (2012)
3. Cavaliere, G., Rahbek, A., Taylor, A.M.R.: Bootstrap determination of the co-integration rank in heteroskedastic VAR models. Econom. Rev. **33**, 606–650 (2014)
4. Johansen, S.: Likelihood-Based Inference in Cointegrated Vector Autoregressive Models. Oxford University Press, Oxford (1996)

Inference in a Non-Homogeneous Vasicek Type Model

Giuseppina Albano and Virginia Giorno

Abstract In the paper we propose a stochastic model, based on a Vasicek non-homogeneous diffusion process, in which the trend coefficient and the volatility are deterministic time-dependent functions. The stochastic inference based on discrete sampling in time is established using a methodology based on the moments of the stochastic process. In order to evaluate the goodness of the proposed methodology a simulation study is discussed.

Keywords Non-homogeneous diffusion · Conditional moments · Estimation procedure

1 Introduction and Background

Models based on stochastic diffusion processes are widely used in economic and financial fields [4, 5, 8]. In particular, the Vasicek process seems to be appropriate to describe no-arbitrage interest rates for which the long-run equilibrium value and the volatility are generally dependent on time. It is solution of the following Stochastic Differential Equation (SDE):

$$dY(t) = [-aY(t) + b]dt + \sigma dB(t), \qquad Y(0) = y_0, \tag{1}$$

where $\sigma > 0$ and $a, b, y_0 \in \mathbb{R}$. In (1) and throughout the paper $B(t)$ is a standard Wiener process. The model (1) with $b = 0$ was originally proposed by Ornstein and Uhlenbeck (OU) in the physical context to describe the velocity of a particle moving in a fluid under the influence of friction and then it was generalized by Vasicek to model loan interest rates. For $a > 0$, $Y(t)$ is a *mean reverting* process, i.e. it tends to oscillate around some equilibrium state. Another interesting property of the OU

G. Albano (✉) · V. Giorno
University of Salerno, Fisciano, SA, Italy
e-mail: pialbano@unisa.it; giorno@unisa.it

M. Corazza et al. (eds.), *Mathematical and Statistical Methods for Actuarial Sciences and Finance*, https://doi.org/10.1007/978-3-319-89824-7_3

process is that, contrary to the Brownian motion, it is a process with finite variance for all $t \geq 0$. In finance modeling σ can be interpreted as the volatility, b/a is the long-run equilibrium value of $Y(t)$ whereas a represents the speed of reversion. The classical model (1) has been extended to the non-homogeneous case in which the drift term depends on the time via certain deterministic functions [7].

In the present paper we propose a generalization of the model (1) in which the drift term and the infinitesimal variance generally depend on time. This is the case of dynamics of suitable economic quantities (as, for example, consumptions, emissions of CO_2, interest rate) that can be subject to the effect of time-dependent exogenous factors, i.e. expectations concerning the future impact of monetary policies, and expected trends in other macroeconomic variables. The problem of estimating the unknown drift term and the volatility function of the model is of great interest in many scientific fields both in theory and in practice. Whereas for the model (1) the parameters estimation is easily obtained via the ML method, the problem of estimating the general time depending trends of the generalized model is, to the knowledge of the authors, not yet still solved. In [7], an ML estimation of the non-linear trend coefficient is provided when it depends on suitable exogenous variables. The aim of our paper is to develop a statistical methodology able to fit the time-dependent drift and volatility functions in the general case, also when no exogenous variables are observable.

The layout of the paper is the following. In Sect. 2 we introduce the model and we give the main probabilistic characteristics of the process such as its transition probability density, the trend functions and the conditional moments generating function. Then, in Sect. 3 we focus on the stochastic inference based on discrete sampling in time and we provide a methodology for fitting the unknown functions. Finally, in Sect. 4, in order to evaluate the goodness of the proposed methodology, a simulation study is discussed.

2 The Model

Let $\{X(t),\ t \in [t_0, T]\}$ be a process defined in $\mathbb{R}$ described via the SDE:

$$dX(t) = [-aX(t) + b(t)]dt + \sigma(t)dB(t), \qquad \mathbb{P}[X(0) = x_0] = 1 \tag{2}$$

where $a \in \mathbb{R}$, $b(t)$ and $\sigma(t) > 0$ are continuous deterministic functions. Equation (2) is able to describe no-arbitrage interest rates for which the long-run equilibrium value and the volatility are generally dependent on time. The transition probability density function (pdf) of $X(t)$, $f(x, t|x_0, t_0)$, is solution of Kolmogorov and of Fokker-Plank equations and it satisfies the initial delta condition, so $f(x, t|y, \tau)$ is a normal pdf characterized by mean and variance

$$M(t|y,\tau) = ye^{-a(t-\tau)} + \int_\tau^t b(\theta)e^{-a(t-\theta)}d\theta, \quad V(t|\tau) = \int_\tau^t \sigma^2(\theta)e^{-2a(t-\theta)}d\theta. \tag{3}$$

3 Fitting the Model

In this section we propose an estimation procedure for the unknown functions $b(t)$ and $\sigma(t)$ in (2). Similar methods are also been proposed in different contexts and for various models (cf. [1–3, 6, 9]). In the present study we assume that the parameter a in (2) is known. We note that from (3) we obtain:

$$b(t) = \frac{dM(t \mid y, \tau)}{dt} + aM(t \mid y, \tau), \qquad \sigma^2(t) = 2aV(t \mid \tau) + \frac{dV(t \mid \tau)}{dt}. \quad (4)$$

Now let us consider a discrete sampling of the process (2) based on d sample paths for the times $t_{ij}, i = 1, \ldots, d$ and $j = 1, \ldots, n$, with $t_{i1} = t_1$ $(i = 1, \ldots, d)$. We observe the values $\{x_{ij}\}_{i=1,\ldots,d;\, j=1,\ldots,n}$ of the variables $X(t_{ij})$ constituting the sample for the inferential study. Starting from the Eq. (4) we suggest the following procedure for the estimation of the functions $b(t)$ and $\sigma^2(t)$:

- From the data $\{x_{ij}\}$, $i = 1, \ldots, d$, and $j = 1, \ldots, n$ we obtain the sample mean μ_j and the sample variance v_j:

$$\mu_j = \frac{1}{d}\sum_{i=1}^{d} x_{ij}, \qquad v_j = \frac{1}{d-1}\sum_{i=1}^{d}(x_{ij} - \mu_j)^2. \quad (5)$$

- Recalling (4) and setting $\Delta_j = t_j - t_{j-1}$ for $j = 2, 3, \ldots, n$ and $b_1 = \sigma_1^2 = 0$, we consider the following fitting points for $b(t)$ and $\sigma^2(t)$

$$b_j = \frac{\mu_j - \mu_{j-1}}{\Delta_j} + a\mu_j, \qquad \sigma_j^2 = 2\,a\,v_j + \frac{v_j - v_{j-1}}{\Delta_j}. \quad (6)$$

- Finally, by interpolating the points b_j $(j = 1, \ldots, n)$ and σ_j^2 $(j = 1, \ldots, n)$ we obtain the functions $\widehat{b(t)}$ and $\widehat{\sigma^2(t)}$ as estimators of $b(t)$ and $\sigma^2(t)$, respectively.

We point out that in (6) we use a simple approximation of the derivatives given by the difference quotient, this choice can be satisfactory for $\Delta_j \to 0$. Moreover, the interpolation can be substitute by other parametric or non parametric regression methods.

4 A Simulation Study

In order to evaluate the goodness of the proposed procedure we present a simulation study. We assume that the parameter $a = 5$ is known and then we consider two cases: Case 1: $b(t) = 0, \sigma(t) \equiv \sigma^2 = 0.01$, and Case 2: $b(t) = 0.1\sin(0.1\,t)$,

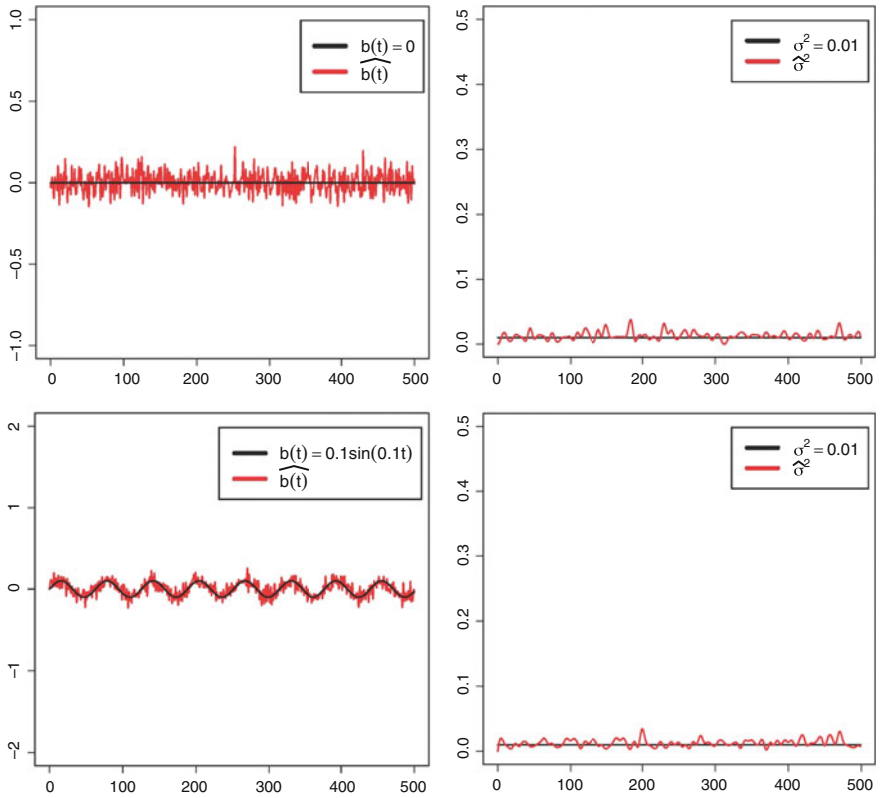


Fig. 1 Functions $b(t)$ (left) and $\sigma^2(t)$ (right) and their estimates are shown for the Case 1 (up) and Case 2 (down)

$\sigma(t) \equiv \sigma^2 = 0.01$. For any cases we simulate 50 sample-paths of the process $X(t)$ with $t_0 = 0$ and $x_0 = 0$. The sample paths include 500 observations of the process starting from $t_1 = t_0 = 0$ with $t_i - t_{i-1} = 1$. The results obtained by using the proposed procedure are shown in Fig. 1. We note explicitly that in the considered cases the proposed procedure is able to capture the trend of the unknown functions.

References

1. Albano, G., Giorno, V., Román-Pomán, P., Torres-Ruiz, F.: Inference on a stochastic two-compartment model in tumor growth. Comp. Stat. Data Anal. **56**, 1723–1736 (2012)
2. Albano, G., Giorno, V., Román, P., Román, S., Torres, F.: Estimating and determining the effect of a therapy on tumor dynamics by means of a modified Gompertz diffusion process. J. Theor. Biol. **364**, 206–219 (2015)

3. Albano, G., Giorno, V., Roman-Roman, P., Torres-Ruiz, F.: On a non-homogeneous Gompertz-type diffusion process: inference and first passage time. In: Lecture Notes in Computer Science, vol. 10672, pp. 47–54. Springer (2018)
4. Chen, W., Xu, L., Zhu, S.P.: Stock loan valuation under a stochastic interest rate model. Comput. Math. Appl. **70**, 1757–1771 (2015)
5. Di Lorenzo, E., Orlando, A., Sibillo, M.: Stochastic model for loan interest rates. Banks and Bank Syst. **8**, 94–99 (2013)
6. Giorno, V., Román, P., Spina, S., Torres, F.: Estimating a non-homogeneous Gompertz process with jumps as model of tumor dynamics. Comput. Stat. Data Anal. **107**, 18–31 (2017)
7. Gutiérrez, R., Nafidi, A., Pascual, A.: Detection, modelling and estimation of non-linear trends by using a non-homogeneous Vasicek stochastic diffusion. Application to CO_2 emissions in Morocco. Stoch. Environ. Res. Risk Assess. **26**, 533–543 (2012)
8. Hull, J.: Options, Futures and Other Derivatives. Pearson, London (2000)
9. Spina, S., Giorno, V., Román, P., Torres, F.: A stochastic model of cancer growth subject to an intermittent treatment with combined effects. Bull. Math. Biol. **77**(11), 2711–2736 (2014)

Small Sample Analysis in Diffusion Processes: A Simulation Study

Giuseppina Albano, Michele La Rocca, and Cira Perna

Abstract In this paper we analyse small sample properties of the ML estimation procedure in Vasicek and CIR models. In particular, we consider short time series, with a length between 20 and 100, typically values observed in finance and insurance contexts. We perform a simulation study in order to investigate which properties of the parameter estimators remain still valid.

Keywords Vasicek process · CIR process · Boostrap resampling

1 Introduction

Diffusion processes are commonly used in a lot of fields ranking from economics to biology, from genetics to engineering. In particular they are able to model stochastic phenomena, such as dynamics of financial securities and short-term loan rates (see, for example, [1] and [5]). Asymptotic properties of ML estimators for the parameters of these models have been analysed in [6], in which also a parametric bootstrap procedure to reduce the bias of the drift estimates has been proposed. However, in many applications data are yearly or quarterly observed, so in the estimation of the parameter the asymptotic condition means to observe the phenomenon for a very long period and most likely such kinds of time series present structural breaks.

In this paper we consider Vasicek and CIR models, mostly employed processes in insurance for the valuation of life insurance contracts and in finance for modeling short-term interest rates (see, for example, [2] and [3]). In particular, we consider short time series, with a length between 20 and 100, tipically values observed in these contexts. We perform a simulation study in order to investigate which properties of the parameter estimator remain still valid.

G. Albano (✉) · M. La Rocca · C. Perna
University of Salerno, Fisciano, Italy
e-mail: pialbano@unisa.it

M. Corazza et al. (eds.), *Mathematical and Statistical Methods for Actuarial Sciences and Finance*, https://doi.org/10.1007/978-3-319-89824-7_4

The paper is organized as follows. Section 2 briefly review the processes, the ML estimators and the bootstrap procedure for reducing the bias of the drift estimates. Section 3 shows a simulation experiment in which the bias and the RMSE are evaluated for different sample sizes, and some concluding remarks, including some guidelines for practitioners, are also given.

2 ML Estimation and Bootstrap Correction

A family of diffusion processes for the interest rates dynamics consists of the following linear drift processes:

$$dX_t = k(\alpha - X_t)dt + \sigma X_t^{\rho} dB_t, \tag{1}$$

where α, k and ψ are unknown parameters. Vasicek and CIR models are obtained from (1) setting $\rho = 0$ and $\rho = \frac{1}{2}$ respectively. Let $\boldsymbol{\theta} = (k, \alpha, \sigma)$ be the vector of the unknown parameters. The MLE for the Vasicek model can be obtained explicitly [6]. Instead, for CIR process explicit expression of the MLEs for θ is not attainable, so suitable approximations of its SDE can be used in order to obtain a pseudo log-likelihood and a related pseudo-MLE [4]. Expansions to the bias and variance of estimators for the Vasicek and CIR processes have been developed in [6]. In particular, the bias expansion of estimators for the Vasicek and CIR process reveals that the bias of the k estimator is of a larger order of magnitude than the bias of estimators for the parameter α and the diffusion parameter σ^2. Moreover the variances of the estimators for the two drift parameters k and α are of larger order than that of the estimator of σ^2. This explains why estimation of k incurs more bias than the other parameters and why the drift parameter (k and α) estimates are more variable than that of the diffusion parameter σ^2. The problem increases when the process has a lack of dynamics, which happens when k is small; in this case the ML estimator for k can incur relative bias of more than 200% even when the time series length is greater than 120. To this aim, a parametric bootstrap procedure for bias correction for general diffusion processes can be adopted [6]. Both theoretical and empirical analysis show that the proposed bias correction effectively reduces the bias without inflating the variance. Precisely, let $\{x_1, \ldots, x_n\}$ be the time series observed at equidistant time-points $i\delta$, $i = 1, \ldots, n$, and let $\hat{\theta} = (\hat{k}, \hat{\alpha}, \hat{\sigma}^2)$ be a mean square consistent estimator of $\theta = (k, \alpha, \sigma^2)$. The bootstrap procedure consists of the following steps:

1. Generate $\{X_t^*\}_{t=1}^n$ with the same δ from $dX_t = \hat{k}(\hat{\alpha} - X_t)dt + \hat{\sigma} X_t^{\rho} dB_t$;
2. Obtain a new estimator $\hat{\theta}^* = (\hat{k}^*, \hat{\alpha}^*, \hat{\sigma}^{2*})$ from the sample path by applying the same estimation procedure as $\hat{\theta}$;
3. Repeat step 1 and 2 N_B number of times and obtain $\hat{\theta}^{*,1}, \ldots, \hat{\theta}^{*,N_B}$;
4. let $\bar{\hat{\theta}}^* = N_B^{-1} \sum_{b=1}^{N_B} \hat{\theta}^{*,b}$, the bootstrap estimator is $\hat{\theta}_B = 2\hat{\theta} - \bar{\hat{\theta}}^*$.

3 Simulation Experiment and Results

In our simulation the parameters for Vasicek process are $\theta = (k, \alpha, \sigma^2) = (0.858, 0.0891, 0.00219)$ (Model V1), $\theta = (0.215, 0.0891, 0.0005)$ (Model V2), $\theta = (0.140, 0.0891, 0.0003)$ (Model V3). For CIR process $\theta = (k, \alpha, \sigma^2) = (0.892, 0.09, 0.033)$ (Model C1), $\theta = (0.223, 0.09, 0.008)$ (Model C2), $\theta = (0.148, 0.09, 0.005)$ (Model C3). In models V3 and C3 the autoregressive coefficient of the discrete time model is 0.99, so it is near to the unit root case. All the simulations are based on 5000 simulations and 1999 bootstrap resamples. Further, we choose $\delta = 1$ that corresponds to yearly observations. The sample size n was 20, 30, 50, 100, 200. Figure 1 shows the relative bias of $\hat{k}$ and its RMSE by using bootstrap correction (on the left) and for ML estimator (on the right). We can observe that the relative bias for ML estimator is very high for $n = 20, 30$ and 50, being greater that 200% for $n = 20$. The bootstrap correction is very efficient in correct the bias but the increase in RMSE of the estimator is greater as much as smaller is the bias in the ML estimator. Concerning the parameter α in the Vasicek model, we observe that the bootstrap correction even performs a worse relative bias and RMSE with respect to the MLE (Fig. 2). Analogous results are obtained for CIR model, as shown in Figs. 3 and 4. Other simulations, not shown here for the sake of brevity, lead us to conclude that, when n is small the bias in the estimators involved in diffusion processes can be really strong, the relative bias till 200% for small samples. Still, the relative bias is worse when the coefficients are near to the non stationarity case. Moreover, the bootstrap procedure is enough efficient in correcting the bias also for very small sample but the increase in RMSE of the estimator is

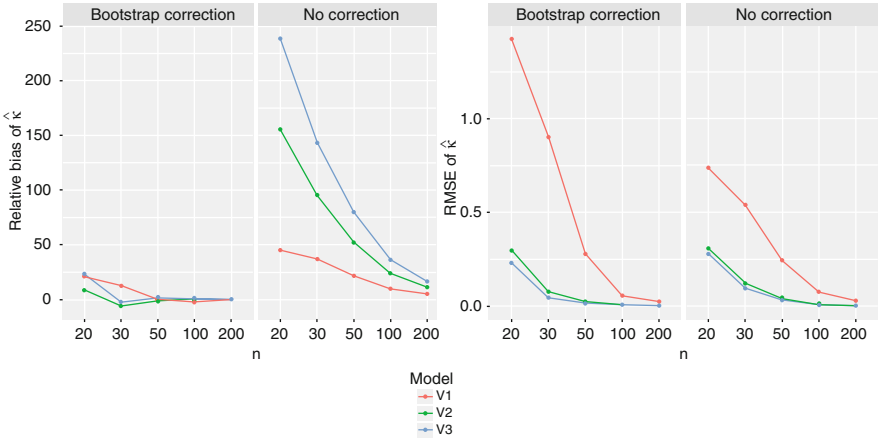


Fig. 1 Vasicek model: relative bias (left) and RMSE (right) of $\hat{k}$ by bootstrap and by MLE

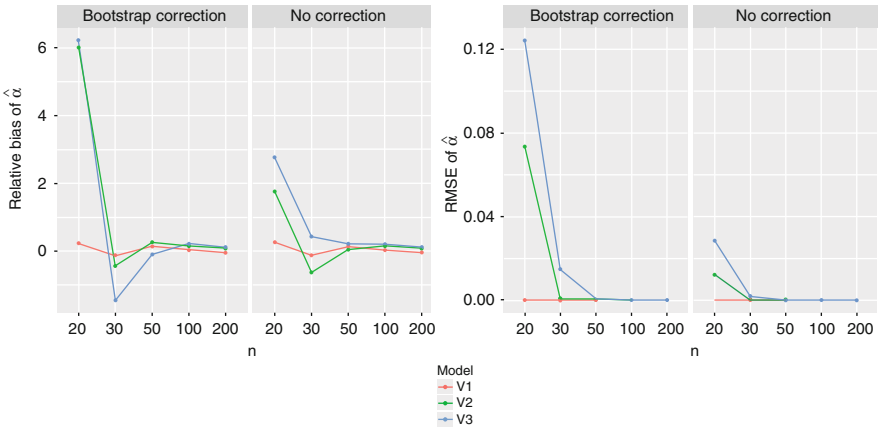


Fig. 2 Vasicek model: relative bias (left) and RMSE (right) of $\hat{k}$ by bootstrap and by MLE

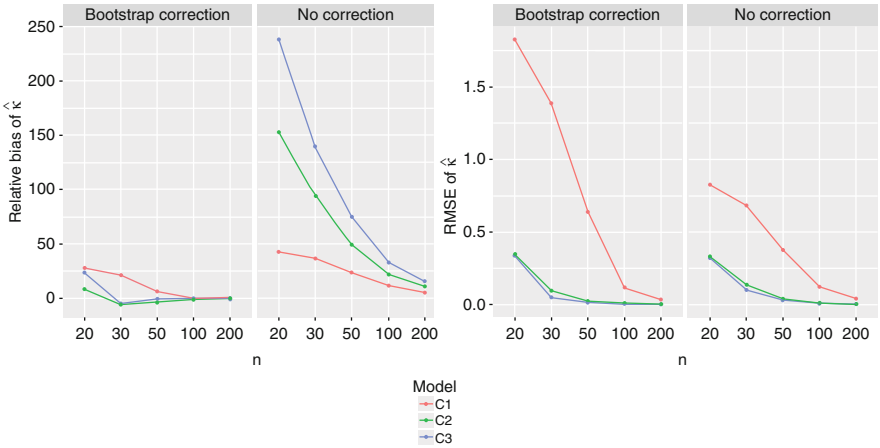


Fig. 3 CIR model: relative bias (left) and RMSE (right) of $\hat{k}$ by bootstrap and by MLE

greater as much as smaller is the bias in the ML estimator. Also instability problems in the estimations can occur when we are near the non stationarity case and when n is small (20–30). Finally, for $n = 20, 30, 50$ the sample distribution of the estimators are asymmetric, so we cannot use tests or confidence intervals based on Normal distribution.

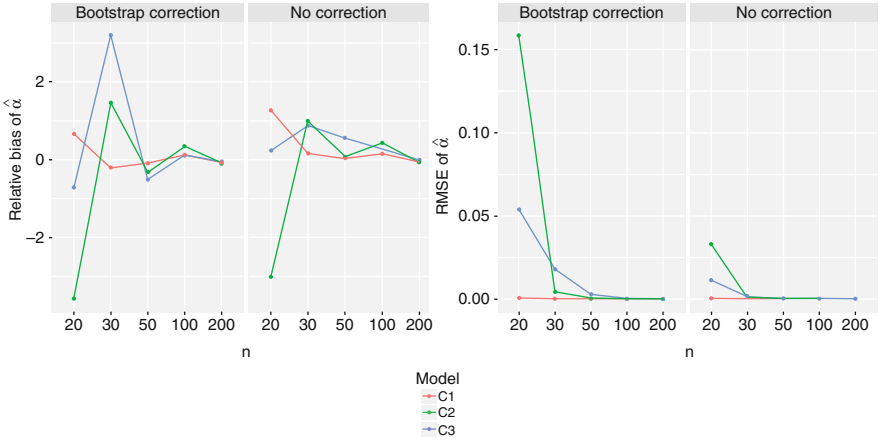


Fig. 4 CIR model: relative bias (left) and RMSE (right) of $\hat{\alpha}$ by bootstrap and by MLE

References

1. Chan, K., Karolyi, A.G., Longstaff, F.A., Sanders, A.B.: An empirical comparison of alternative models of the short-term interest rate. J. Financ. **47**, 1209–1227 (1992)
2. Chen, W., Xu, L., Zhu, S.P.: Stock loan valuation under a stochastic interest rate model. Comput. Math. Appl. **70**, 1757–1771 (2015)
3. Di Lorenzo, E., Orlando, A., Sibillo, M.: A stochastic model for loan interest rates. Banks Bank Syst. **8** (4), 94–99 (2013)
4. Nowman, K.: Gaussian estimation of single-factor continuous time models of the term structure of interest rates. J. Financ. **52**, 1695–1706 (1997)
5. Sundaresan, S.M.: Continuous-time methods in finance: a review and an assessment. J. Financ. **55**(4), 1569–1622 (2000)
6. Tang, C.Y., Chen, S.X.: Parameter estimation and bias correction for diffusion processes. J. Economet. **149**(1), 65–81 (2009)

Using Deepest Dependency Paths to Enhance Life Expectancy Estimation

Irene Albarrán-Lozano, Pablo J. Alonso-González, and Aurea Grané

Abstract The aim of this work is to estimate life expectancy free of dependency (LEFD) using categorical data and individual dependency trajectories that are obtained using the whole medical history concerning the dependency situation of each individual from birth up to 2008, contained in database EDAD 2008. In particular, we estimate LEFD in several scenarios attending to age, gender, proximity-group and dependency degree. Proximity-groups are established according to an L^2-type distance from the dependency trajectories to a central trend within each age-gender group, using functional data techniques.

Keywords Cox Regression · Dependency · Disability · Functional data

1 Introduction

Dependency, that is, lack of autonomy in performing basic ADL can be seen as a consequence of the process of gradual aging. In Europe in general and in Spain in particular this phenomenon represents a problem with economic, political and social implications. The prevalence of dependency in the population, as well as its intensity and evolution over the course of a person's life are issues of greatest importance that should be addressed.

The aim of this work is to estimate life expectancy free of dependency (LEFD), that is, the expected number of years that a person can live free of this contingency based on mortality and morbidity conditions. The evolution of dependency in the Spanish population is studied through a pseudo panel constructed from

I. Albarrán-Lozano (✉) · A. Grané
Statistics Department, Universidad Carlos III de Madrid, Getafe, Spain
e-mail: irene.albarran@uc3m.es; aurea.grane@uc3m.es

P. J. Alonso-González
Statistics Department, Universidad de Alcalá, Alcalá de Henares, Spain
e-mail: pablo.alonsog@uah.es

M. Corazza et al. (eds.), *Mathematical and Statistical Methods for Actuarial Sciences and Finance*, https://doi.org/10.1007/978-3-319-89824-7_5

EDAD 2008 [1, 3], in the lack of longitudinal studies or the possibility to link different cross-sectional surveys.

EDAD 2008 (Survey on Disability, Personal Autonomy and Dependency Situations 2008, undertaken by the Spanish Statistical Office—INE) is the most recent Spanish survey about disability and was the first Spanish survey that used the internationally accepted measures established by the 'International classification of functioning, disability and health'.[1] It was also the first time that the survey included information useful for studying the dependency phenomenon, such as the average hours per week of special care received by the dependent person. Moreover, EDAD 2008 contains, among many other variables, the ages at which each person in the sample has suffered a change in his/her health condition related to disability. Therefore, applying the Spanish legislation (Act 39/2006 and Royal Decree 504/2007) it is possible to recover the individual historical event data and, thus, to obtain the individual dependency evolution from birth up to 2008.

The sample selected for the present study is formed by 7446 individuals and represents 2.35% of the Spanish population in 2008, that is more than 1 million people (325,253 men and 773,079 women). Each individual in the sample has a weight reflecting the population group that represents. These weights have been taken into account in all the computations of this paper.

2 Methodology

The main contribution of this paper is the estimation of LEFD not only by gender or dependency degree (I-moderate, II-severe, III-major), but also by partitioning the individuals in homogeneous groups with a similar dependency evolution.

For all the analysis performed in this article the first age interval is $[50, 60)$ and the last one $[90, \infty)$. Moreover, we are particularly interested in those people with a dependency score of zero at the age of 30, and from now on, they will be grouped in ten age-gender intervals (five groups per gender) according to their current age at 2008.

For each one of these ten groups we compute the deepest curve in terms of modified band depth [4] using the `roahd` package in R by Tarabelloni et al. [5]. As an example, in Fig. 1 we depict the dependency trajectories with the corresponding deepest curve for several age-gender groups, where we observe that at the same age the first score value reached by these deepest curves is lower for women than for men. The deepest curve of a sample in terms of modified band depth has been

[1] In 2001, the World Health Organization (WHO 2011) [6] established a framework for measuring health and disability at both individual and population levels, which was known as the 'International classification of functioning (ICF), disability and health'. The ICF tries to establish a consensus in its understanding, by establishing a difference between the basic activities of living daily (ADL) and the instrumental ADL. The basic activities are defined as those activities which are essential for an independent life.

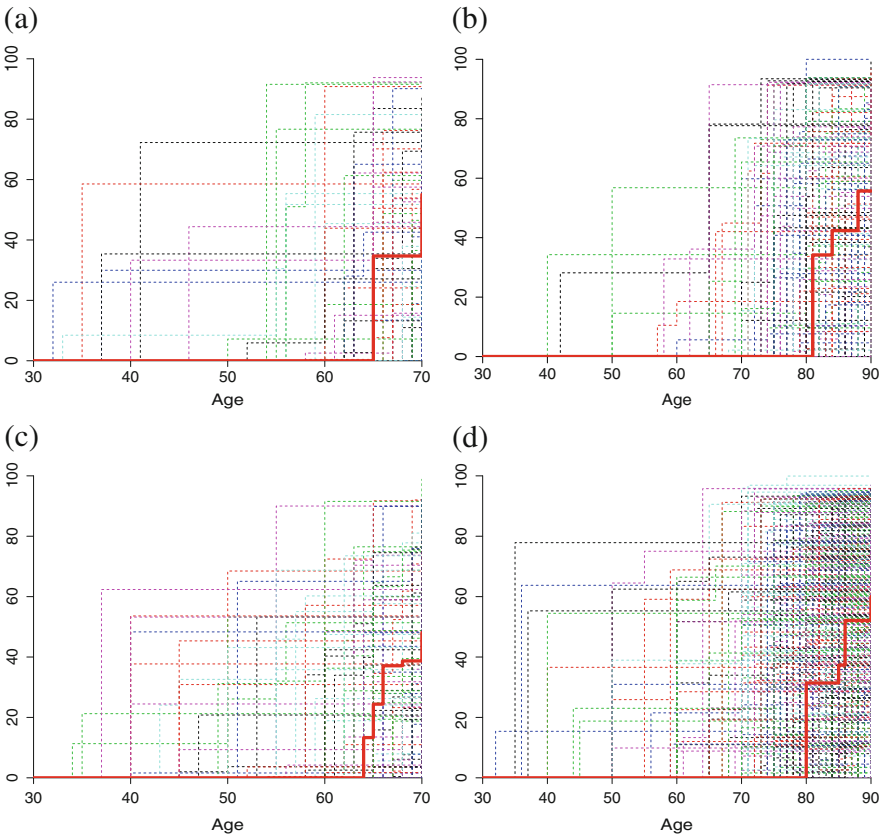


Fig. 1 Dependency trajectories for men and women with their corresponding deepest curves (in bold red). (**a**) Men at 70. (**b**) Men at 90. (**c**) Women at 70. (**d**) Women at 90

proven to be an accurate and robust estimator of the central pattern of a sample of curves in the time warping model [2].

Then, within each group we compute the L^2-distance of each trajectory to the corresponding deepest one multiplied by 1 (or -1) if the trajectory is most of the time above (or below) the deepest curve. This yields a numerical summary for each one of the trajectories, d_j, that can be used to establish different patterns. As we will see later, these patterns will exhibit quite different life expectancies. In particular, negative values of d_j correspond to trajectories below the deepest curve and, therefore, to individuals with lower dependency scores than those of the central trend of their age-gender groups. In fact, the best situations are expected for the left-tail values of d_j. On the other hand, positive values of d_j correspond to trajectories above the deepest curve and, hence, to individuals with higher dependency scores than those of the central trend of their age-gender groups. In this case, the worst situations are expected for large values of d_j. Since each set of values $\{d_j < 0\}$ and $\{d_j \geq 0\}$ has a different meaning, we propose to compute the LEFD for the groups

of individuals established by the quartiles in each set, yielding to eight groups for each gender, that we call proximity-groups.

Once these proximity-groups are established, we use Cox regression model to obtain the 'survival' probabilities (in fact, the staying free of dependency probability at a given age given that a person is alive at that age). Marginal probabilities are calculated by multiplying these estimates by survival probabilities given by the Spanish disabled pensioners' mortality table. Finally, we obtain the LEFD for Spanish population within homogenous groups considering age, gender and dependency degree. As far as we know, this is the first time that the dependency evolution is used to estimate life expectancy.

3 Results

Table 1 contains the estimated LEFD for men and women at three particular ages jointly with the LEFD calculated without taking into account the partition by proximity-groups (rows global LEFD for men and global LEFD for women). We may remind that these LEFD estimations are computed from survey EDAD 2008, that contains only dependent people. Therefore, they must interpreted as the 'at least' expected numbers of years free of dependency. Nevertheless, the methodology that we propose in this paper is not restricted to the database.

In Table 1 we observe that the variance of LEFD increases with age and tends to decrease with dependency degree. In general, the variance is greater for women. We also observe that the global LEFD by gender, calculated without taking into account the partition by proximity-groups, is far from any of the LEFD values estimated by proximity-groups.

In order to highlight these findings, in Fig. 2 we depict the evolution of LEFD along age for two scenarios. For the one hand, the expected number of years for people free of any dependency degree and, on the other hand, for people free of major dependency degree. In each panel, the lower threshold corresponds to the LEFD estimation for people in the worst dependency situation (Proximity-group IV, $\{d_j \geq 0\}$), whereas the upper threshold reflects the LEFD estimation for people in the best dependency situation (Proximity-group IV, $\{d_j < 0\}$). Finally, the black line corresponds to global LEFD estimation.

Figure 2 reinforces the conclusions derived from Table 1 and, additionally, we observe that LEFD decreasing rate is higher (and more abrupt) for men than for women. The huge difference between the global LFED and the two thresholds may suggest that the global LEFD estimation is not be representative of the Spanish dependent population, not even for those individuals within the most central proximity-groups, that is, for those that are the nearest to the corresponding central trend. From economic and demographic points of view, this is a relevant finding, since the expected dependent population would demand care services (health care, pensions and other services) that should be covered and related expenditures should be financed.

Table 1 Summary of LEFD for three particular ages

	Men	30 years old			50 years old			70 years old		
	Degree	I	II	III	I	II	III	I	II	III
$\{d_j < 0\}$	Proximity-group IV	28.234	32.320	32.416	14.117	19.501	19.623	4.417	8.424	8.648
	Proximity-group III	28.616	33.167	35.395	14.701	20.618	23.567	4.040	9.022	11.706
	Proximity-group II	25.664	31.426	35.186	11.352	18.389	23.290	3.065	6.745	11.234
	Proximity-group I	24.069	32.908	35.648	9.424	20.275	23.902	3.004	8.899	12.120
$\{d_j \geq 0\}$	Proximity-group I	21.848	26.600	35.375	7.378	12.228	23.541	1.838	3.711	11.812
	Proximity-group II	20.792	25.213	31.548	6.524	10.375	18.474	1.346	2.857	7.400
	Proximity-group III	14.762	22.884	31.427	4.695	8.647	18.659	0.878	2.127	7.425
	Proximity-group IV	12.191	13.509	20.194	2.426	3.183	7.497	0.359	0.600	2.337
Global LEFD for men		18.733	23.614	31.868	6.475	10.429	19.391	1.556	3.195	8.583
	Women	30 years old			50 years old			70 years old		
	Degree	I	II	III	I	II	III	I	II	III
$\{d_j < 0\}$	Proximity-group IV	32.805	34.370	35.099	20.195	22.210	23.175	8.187	9.780	11.110
	Proximity-group III	31.714	33.353	35.264	18.694	20.864	23.394	6.465	8.783	11.589
	Proximity-group II	30.578	33.549	35.614	17.215	21.141	23.860	4.753	8.973	12.116
	Proximity-group I	26.172	33.743	–	11.557	21.380	–	2.343	8.699	–
$\{d_j \geq 0\}$	Proximity-group I	23.145	27.523	35.166	8.766	13.904	23.264	1.664	4.017	11.102
	Proximity-group II	23.527	28.316	33.279	9.340	14.536	20.765	1.938	4.283	8.204
	Proximity-group III	18.077	26.109	31.882	6.543	11.858	18.958	1.190	2.745	6.992
	Proximity-group IV	13.852	14.923	21.651	3.288	3.909	9.114	0.579	0.759	2.757
Global LEFD for women		21.601	26.657	32.860	8.781	13.623	20.577	2.092	4.337	9.045

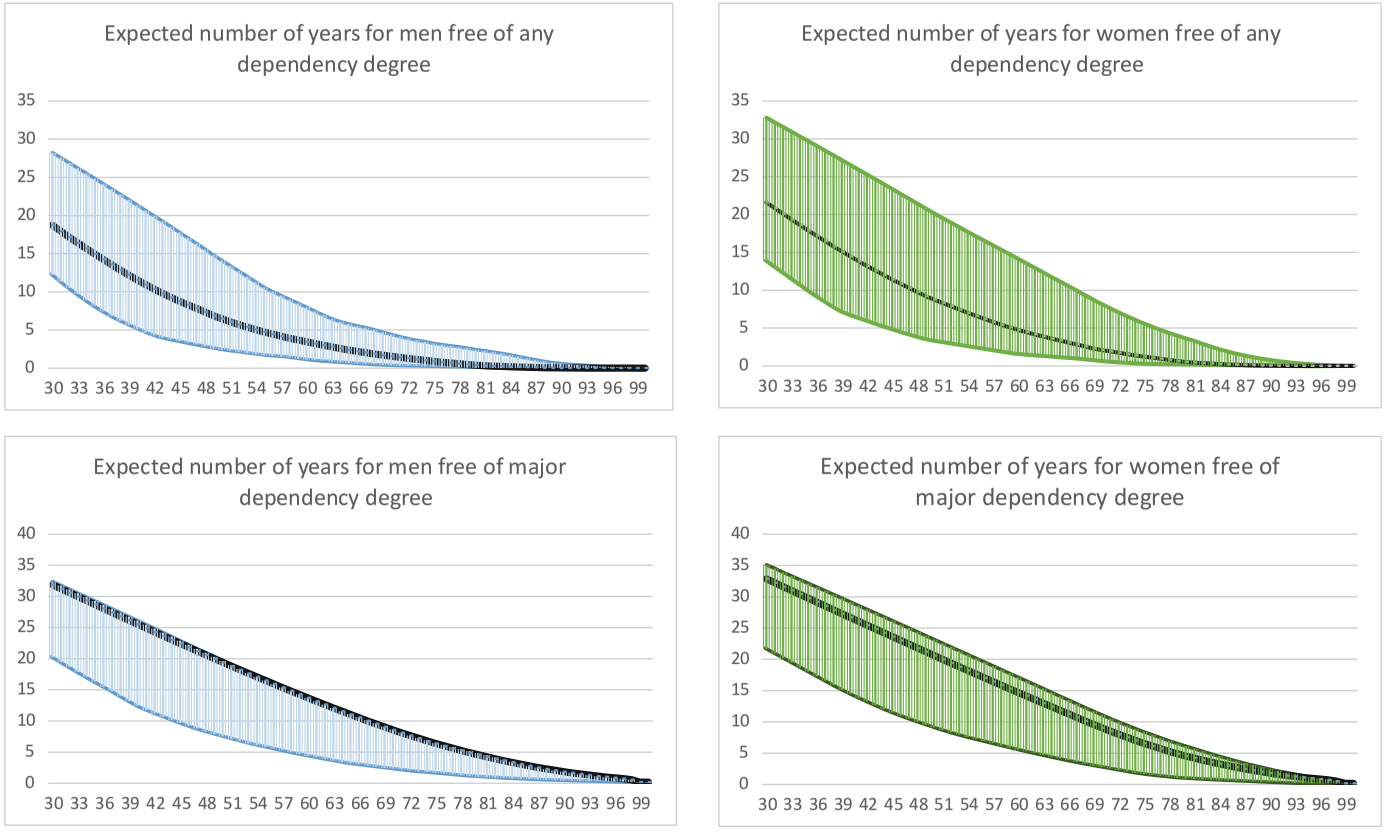


Fig. 2 Evolution of LEFD along age by gender. Two scenarios

Acknowledgements The authors would like thank their former colleague Ana Arribas-Gil for her useful comments and suggestions. Financial support from research project MTM2014-56535-R by the Spanish Ministry of Economy and Competitiveness.

References

1. Albarrán-Lozano, I., Alonso-González, P., Arribas-Gil, A.: Dependence evolution in the Spanish disabled population: a functional data analysis approach. J. R. Stat. Soc. A **180**(2), 657–677, (2017)
2. Arribas-Gil, A., Romo, J.: Robust depth-based estimation in the time warping model. Biostatistics **13**, 398–414 (2012)
3. EDAD 2008: Encuesta sobre Discapacidad, Autonomía personal y Situaciones de Dependencia, Metodología. Ed. Subdirección General de Estadísticas Sociales Sectoriales (INE), Madrid, España (2010)
4. López-Pintado, S., Romo, J.: On the concept of depth for functional data. J. Am. Stat. Assoc. **114**, 486–503 (2009)
5. Tarabelloni, N., Arribas-Gil, A., Ieva, F., Paganoni, A., Romo, J.: ROAHD: Robust analysis of high dimensional data. Package on CRAN. Version 1.0.4. Published 2016-07-06
6. WHO: Global Health and Aging. World Health Organization. US National Institute of Aging, Geneva (2011)

The Optimal Investment and Consumption for Financial Markets Generated by the Spread of Risky Assets for the Power Utility

Sahar Albosaily and Serguei Pergamenshchikov

Abstract We consider a spread financial market. We construct the optimal consumption/investment strategy for the power utility function. We study the Hamilton–Jacobi–Bellman (HJB) equation by the Feynman–Kac (FK) representation. We study the numeric approximation and we establish the convergence rate.

Keywords Optimality · Feynman–Kac mapping · Hamilton–Jacobi–Bellman equation · Itô formula · Brownian motion · Ornstein–Uhlenbeck process · Stochastic processes · Financial market · Spread market

1 Market Model

Let $(\Omega, \mathscr{F}_T, (\mathscr{F}_t)_{0\le t\le T}, \mathbf{P})$ be a standard filtered probability space with $(\mathscr{F}_t)_{0\le t\le T}$ adapted Wiener processes $(W_t)_{0\le t\le T}$. Our financial market consists of one *riskless bond* $(\check{S}_t)_{0\le t\le T}$ and *risky spread stocks* $(S_t)_{0\le t\le T}$ governed by the following equations:

$$\begin{cases} \mathrm{d}\check{S}_t = r\,\check{S}_t\,\mathrm{d}t\,, & \check{S}_0 = 1\,, \\ \mathrm{d}S_t = -\kappa\, S_t\,\mathrm{d}t + \sigma\,\mathrm{d}W_t\,, & S_0 > 0\,. \end{cases} \tag{1}$$

S. Albosaily (✉)
Laboratoire de Mathématiques Raphael Salem, UMR 6085 CNRS - Université de Rouen, Saint-Étienne-du-Rouvray, France

University of Hail, Ha'il, Saudi Arabia
e-mail: sahar.albosaily@etu.univ-rouen.fr

S. Pergamenshchikov
Laboratoire de Mathématiques Raphael Salem, UMR 6085 CNRS - Université de Rouen, Saint-Étienne-du-Rouvray, France

International Laboratory of Statistics of Stochastic Processes and Quantitative Finance of National Research Tomsk State University, Tomsk, Russia
e-mail: Serge.Pergamenshchikov@univ-rouen.fr

M. Corazza et al. (eds.), *Mathematical and Statistical Methods for Actuarial Sciences and Finance*, https://doi.org/10.1007/978-3-319-89824-7_6

Here $r > 0$ is the interest rate, $\kappa > 0$ is the market constant parameter from $\mathbb{R}$ and $\sigma > 0$ is the market volatility. We consider the wealth process corresponding to the investment $\alpha = (\alpha_t)_{0\le t\le T}$ and the consumption $c = (c_t)_{0\le t\le T}$ strategies as the following

$$\mathrm{d}X_t^{\upsilon} = (rX_t^{\upsilon} - \kappa_1\alpha_t S_t)\,\mathrm{d}t + \alpha_t\sigma\,\mathrm{d}W_t - c_t\,\mathrm{d}t\,, \tag{2}$$

where $\kappa_1 = \kappa + r > 0$.

Definition 1 The financial strategy $\upsilon = (\upsilon_t)_{0\le t\le T}$ is called admissible if this process is adapted and the Eq. (2) has a unique strong nonnegative solution.

We denote by $\mathscr{V}$ the set of all admissible financial strategies. For initial endowment $x > 0$, admissible strategy υ in $\mathscr{V}$ and state process $\varsigma_t = (X_t^{\upsilon}, S_t)$, we introduce for $0 < \gamma < 1$ the following cost function

$$J(\varsigma, t, \upsilon) := \mathbf{E}_{\varsigma,t}\left(\int_t^T c_t^{\gamma}\,\mathrm{d}t + (X_T^{\upsilon})^{\gamma}\right), \tag{3}$$

where $\mathbf{E}_{\varsigma,t}$ is the conditional expectation with respect to $\varsigma_t = \varsigma = (x, s)$. We set $J(\varsigma\,,\upsilon) = J(\varsigma\,,0\,,\upsilon)$. Our goal is to maximize the cost function (3), i.e. $\sup_{\upsilon\in\mathscr{V}} J(\varsigma\,,\upsilon)$. To do this we use the dynamical programming method so, we need to study the problem (3) for any $0 \le t \le T$.

2 Stochastic Programming Method

Denoting by $\varsigma_t = (X_t, S_t)$, we can rewrite Eqs. (1) and (2) as,

$$\mathrm{d}\varsigma_t = a(\varsigma_t, \upsilon_t)\,\mathrm{d}t + b(\varsigma_t, \upsilon_t)\,\mathrm{d}W_t\,,$$

where

$$a(\varsigma, \mathbf{u}) = \begin{pmatrix} rx - \kappa_1\alpha s - c \\ -\kappa s \end{pmatrix}, \quad b(\varsigma, \mathbf{u}) = \begin{pmatrix} \alpha\sigma \\ \sigma \end{pmatrix} \quad \text{and} \quad \mathbf{u} = (\alpha, c)\,.$$

We introduce the Hamilton function, for any

$$q = \begin{pmatrix} q_1 \\ q_2 \end{pmatrix}, \quad M = \begin{pmatrix} M_{11} & M_{12} \\ M_{21} & M_{22} \end{pmatrix}, \quad \text{and} \quad 0 \le t \le T\,,$$

we set,

$$H(\varsigma, t, q, M) := \sup_{\mathbf{u}\in\Theta} H_0(\varsigma, t, q, M, \mathbf{u}), \qquad \Theta \in \mathbb{R}\times\mathbb{R}_+,$$

where $H_0(\varsigma, t, q, M, \mathbf{u}) := a'(\varsigma, t, \mathbf{u})q + \mathbf{tr}[bb'(\varsigma, t, \mathbf{u})M]/2 + c^\gamma$, where the prime $'$ here denotes the transposition. In order to find the solution to the value function we need to solve the Hamilton–Jacobi–Bellman equation which is given by

$$\begin{cases} z_t(\varsigma, t) + H(\varsigma, t, \partial z(\varsigma, t), \partial^2 z(\varsigma, t)) = 0, & t \in [0, T], \\ z(\varsigma, T) = x^\gamma, & \varsigma \in \mathbb{R}^2, \end{cases} \tag{4}$$

where

$$\partial z(\varsigma, t) = \begin{pmatrix} z_x \\ z_s \end{pmatrix} \quad \text{and} \quad \partial^2 z(\varsigma, t) = \begin{pmatrix} z_{xx} & z_{xs} \\ z_{sx} & z_{ss} \end{pmatrix}.$$

In this case the Hamilton-Jacobi-Bellman equation is in the following form

$$\begin{aligned} & z_t(\varsigma, t) + \frac{1}{2}\frac{(\sigma^2 z_{xs} - \kappa_1 s z_x)^2}{\sigma^2 |z_{xx}|} + \frac{\sigma^2 z_{ss}}{2} + rxz_x - \kappa s z_s \\ & + (1 - \gamma)\Big(\frac{z_x}{\gamma}\Big)^{\frac{\gamma}{\gamma - 1}} = 0, \end{aligned} \tag{5}$$

where $z(\varsigma, T) = x^\gamma$. To study this equation we use the following form for the solution

$$z(x, s, t) = x^\gamma U(s, t) \quad \text{and} \quad U(s, t) = \exp\left\{\frac{s^2}{2} g(t) + Y(s, t)\right\}, \tag{6}$$

for some function $g(.) > 0$, and

$$\begin{cases} Y_t(s, t) + \frac{1}{2}\sigma^2 Y_{ss}(s, t) + s g_1(t) Y_s(s, t) + \Psi_Y(s, t) = 0, \\ Y(s, T) = 0. \end{cases} \tag{7}$$

As we will see later, that the Eq. (7) has a solution in $\mathbf{C}^{2,0}(\mathbb{R} \times [0, T])$ which can be represented as a fixed point for the Feynman–Kac mapping

$$Y(s, t) = \mathbf{E} \int_t^T \Psi_Y(\eta_u^{s,t}, u)\, \mathrm{d}u = \mathscr{L}_Y(s, t). \tag{8}$$

Using the solution of Eq. (6), we define the functions

$$\begin{aligned} \check{\alpha}_0(\varsigma, t) &= \frac{\kappa_1 s z_x(\varsigma, t)}{\sigma^2 z_{xx}(\varsigma, t)} - \frac{z_{xs}(\varsigma, t)}{z_{xx}(\varsigma, t)} = \check{\beta}(s, t) x, \\ \check{c}_0(\varsigma, t) &= \left(\frac{z_x(\varsigma, t)}{\gamma}\right)^{\frac{1}{\gamma - 1}} = G(s, t) x, \end{aligned} \tag{9}$$

where $\check{\beta}(s, t) = \big(s g(t) + Y_s(s, t) - \kappa_1 s/\sigma^2\big)/(1 - \gamma)$.

Now we set the following stochastic equation to define the optimal wealth process, i.e., we set

$$\mathrm{d}X_t^* = a^*(t)\, X_t^*\, \mathrm{d}t + b^*(t)\, X_t^*\, \mathrm{d}W_t\,, \tag{10}$$

where $a^*(t) = r - \kappa_1\, S_t\, \check{\beta}(S_t, t) - \check{G}(S_t, t)$ and $b^*(t) = \sigma\, \check{\beta}(S_t, t)\,.$

Using the stochastic differential equation (10) we define the optimal strategies:

$$\alpha_t^* \;=\; \check{\alpha}_0(\varsigma_t^*\,, t) \quad \text{and} \quad c_t^* \;=\; \check{c}_0(\varsigma_t^*\,, t)\,, \tag{11}$$

where $\varsigma_t^* = (X_t^*\,, S_t)\,,\ X_t^*$ is defined in (10).

Remark 1 Note that the main difference in the Hamilton–Jacobi–Bellman equation (5) from the one in [2] is the last nonlinear term as we see we can not use the solution method from [2].

3 Main Results

First we study the Hamilton–Jacobi–Bellman equation.

Theorem 1 *There exists $T_0 > 0$ such that for all $0 \le T \le T_0$, then Eq. (4) is the solution defined by (6), where Y is the unique solution of (7) in $\mathscr{X}$ and is the fixed point for the Feynman–Kac mapping, i.e., $Y = \mathscr{L}_Y$.*

Theorem 2 *Assume that $0 \le T \le T_0$, Then the optimal value of $J(t, \varsigma, \upsilon)$ is given by*

$$\max_{\upsilon \in \mathscr{V}} J(\varsigma, t, \upsilon) \;=\; J(\varsigma, t, \upsilon^*) \;= x^{\gamma} U(s, t)\,,$$

where the optimal control $\upsilon^* = (\alpha^*, c^*)$ for all $0 \le t \le T$ is given in (9) with the function Y defined in (8). The optimal wealth process $(X_t^*)_{0 \le t \le T}$ is the solution to(10).

Similarly to [1], we study the HJB equation through the FK representation. Therefore, let us now define the approximation sequence $(h_n)_{n \ge 1}$, for h as $h_0 = 0$, and for $n \ge 1$, as $h_n = \mathscr{L}_{h_{n-1}}$. In the following theorems we show that the approximation sequence goes to the fixed function h, i.e. $h = \mathscr{L}_h$

Theorem 3 *For some $\delta > 0$, the approximation*

$$||h - h_n|| \le O(n^{-\delta n}) \quad as \quad n \to \infty.$$

Remark 2 Note that the convergence rate is super geometrical.

Now we define the approximation. We set

$$\alpha_n^*(\varsigma\, t) = \check{\beta}_n(s,t)\, x \quad \text{and} \quad c_n^*(\varsigma, t) = G_n(s,t)\, x$$

where $\check{\beta}_n(s,t) = \Big(s\, g(t) \, + \partial h_n(s,t)/\partial s \, - \kappa_1 s/\sigma^2\Big)/(1-\gamma)\,.$

Theorem 4 *There exists $\delta > 0$ such that as $n \to \infty$,*

$$\sup_{\varsigma, 0\le t\le T} \Big(\big|\alpha^*(\varsigma,t) \, - \, \alpha_n^*(\varsigma,t)\big| + \big|c^*(\varsigma,t) \, - \, c_n^*(\varsigma,t)\big|\Big) \, \le \, O(n^{\,-\,\delta n})\,.$$

Remark 3 As it is seen from Theorem 1 the approximation scheme for the Hamilton–Jacobi–Bellman equation implies the approximation for the optimal strategy with the some super geometrical rate. i.e. more rapid than any geometrical ones.

Remark 4 Indeed the amount T_0 can be calculated in explicit form.

References

1. Berdjane, B., Pergamenchtchikov, S.M.: Optimal consumption and investment for markets with randoms coefficients. Finance Stochast. **17**(2), 419–446 (2013)
2. Boguslavsky, M., Boguslavskaya, E.: Arbitrage under power. Risk **17**, 69–73 (2004)

Combining Multivariate Volatility Models

Alessandra Amendola, Manuela Braione, Vincenzo Candila, and Giuseppe Storti

Abstract Forecasting conditional covariance matrices of returns involves a variety of modeling options. First, the choice between models based on daily or intradaily returns. Examples of the former are the Multivariate GARCH (MGARCH) models while models fitted to Realized Covariance (RC) matrices are examples of the latter. A second option, strictly related to the RC matrices, is given by the identification of the frequency at which the intradaily returns are observed. A third option concerns the proper estimation method able to guarantee unbiased parameter estimates even for large (MGARCH) models. Thus, dealing with all these modeling options is not always straightforward. A possible solution is the combination of volatility forecasts. The aim of this work is to present a forecast combination strategy in which the combined models are selected by the Model Confidence Set (MCS) procedure, implemented under two economic loss functions (LFs).

Keywords Multivariate volatility · Model confidence set · Realized covariances · Forecast combination

1 Introduction

Conditional covariance matrices of returns can be estimated by means of different approaches. Among these, MGARCH models [6], based on daily returns, represent one of most prominent example. Moreover, in more recent years the research has also focused on the development of unbiased and efficient realized estimators of conditional covariance matrices [5]. Specifications like the Conditional Autoregressive Wishart (CAW) model [11] directly use RC measures to obtain covariance

A. Amendola · V. Candila (✉) · G. Storti
Department of Economics and Statistics, University of Salerno, Fisciano, Italy
e-mail: alamendola@unisa.it; vcandila@unisa.it; storti@unisa.it

M. Braione
SOSE spa, Roma, Italy
e-mail: mbraione@sose.it

M. Corazza et al. (eds.), *Mathematical and Statistical Methods for Actuarial Sciences and Finance*, https://doi.org/10.1007/978-3-319-89824-7_7

forecasts. When the focus is on the RC matrices, an additional issue arises: the choice of the frequency at which the intradaily returns are observed. In fact, practitioners dealing with RC models have to carefully choose the optimal intradaily frequency, giving the best bias/variance trade-off [4, 17]. Finally, as regards the estimation method, also the choice between Quasi Maximum Likelihood Estimator (QMLE) or Composite QMLE (cQMLE) still represents an open issue [16].

Disentangling simultaneously all these issues could be seriously problematic. A possible solution is the combination of volatility forecasts [3] coming from different models, characterized by different estimation procedures and information observed at different frequencies. Thus, this work aims at investigating the profitability of a forecasting strategy based on the combination of multivariate volatility models, under economic LFs. More in detail, the proposed combination strategy relies on the MCS [13] procedure, through which several MGARCH and RC-based models, estimated by QMLE or cQMLE, are assessed, with respect to two (univariate) economic LFs and four different portfolio compositions. Models entering the so-called *training* MCS are then averaged in order to obtain the combined predictor. Subsequently, the so-called *evaluation* MCS assesses the forecasting performances of MCS-based combined predictor with respect to those of the whole model universe augmented with the simple equally weighted average of forecasts from all the candidate models.

2 MCS Combination Strategy

Let $\boldsymbol{r}_t$ be a $k \times 1$ vector of daily-log returns for k assets at time t. Moreover, let $H_{i,t}$ be a $(k \times k)$ matrix for the volatility model i, with $i = 1, \ldots, M$. Thus, the matrices $H_{i,t}$ are the forecasts of the covariance matrix of $\boldsymbol{r}_t$ conditionally on the information set I_{t-1}, generated by each candidate model i.

Let H_t^{Comb} denote the combined predictor based on the available M candidate models. The proposed combination rule defines the combined prediction H_t^{Comb} as equally averaging all the models belonging to the *training* MCS. Let H_t^{All}, the natural benchmark of H_t^{Comb}, denote the equally weighted combined predictor, obtained by averaging $H_{i,t}$. Both H_t^{All} and H_t^{Comb} as well as all the model universe are evaluated in the *evaluation* MCS. The MCS-based combination strategy is carried out under two economic LFs: the Caporin [7] and Lopez [14] LFs. Both LFs are univariate and belong to the Value-at-Risk (VaR) framework [2].

In order to fairly assess the predictive performances of H_t^{Comb}, H_t^{All} and each single candidate models, we only consider portfolios with constant weights. Namely, in addition to an Equally Weighted (EW) allocation strategy, we also consider three alternative unequally weighted allocation strategies, labelled as R1, R2 and R3, assigning different weights to the economic sectors composing our dataset.

3 Empirical Analysis

The performance of our proposed strategy is evaluated by analysing a portfolio of $k = 12$ stocks listed on the New York Stock Exchange and $M = 24$ different specifications, of which eight are MGARCH and sixteen are RC-based models.

Table 1 shows the frequency at which each specification enters the MCS over the training period. Regardless of the economic LF adopted, under EW and R1 weights, the BEKK specification almost always enters the MCS. Instead, under R2 and R3 portfolio configurations, the model included in the MCS with the highest frequency is the DECO.

The results of the *evaluation* MCS are illustrated in Table 2, where for brevity, only the MCS p-values of H_t^{Comb} and H_t^{All} are reported. First of all, we note that the combined predictor always enters the MCS, independently of the considered vector of weights. Secondly, and more importantly, our principal competitor H_t^{All} has a substantially worse performance since it does not enter the MCS in four out of the eight cases considered.

To conclude, for data and period under consideration, H_t^{Comb} always belongs to the *evaluation* MCS, regardless of the LFs adopted or the portfolio composition.

Table 1 Frequency of models entering the *training* MCS

Weights	EW	R1	R2	R3	EW	R1	R2	R3
Loss	Caporin LF				Lopez LF			
CAW_5	0.000	0.077	0.115	0.115	0.192	0.077	0.077	0.077
CAW-C_5	0.000	0.000	0.000	0.077	0.077	0.077	0.000	0.077
RRM_5	0.000	0.000	0.000	0.000	0.000	0.038	0.000	0.038
RMC_5	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
CAW_10	0.000	0.115	0.115	0.115	0.269	0.192	0.038	0.077
CAW-C_10	0.000	0.000	0.000	0.077	0.385	0.077	0.000	0.077
RRM_10	0.000	0.000	0.000	0.000	0.192	0.000	0.000	0.000
RMC_10	0.000	0.000	0.000	0.000	0.231	0.000	0.000	0.000
CAW_15	0.000	0.115	0.115	0.115	0.500	0.192	0.038	0.077
CAW-C_15	0.000	0.038	0.000	0.038	0.269	0.077	0.000	0.038
RRM_15	0.000	0.000	0.000	0.000	0.038	0.000	0.000	0.000
RMC_15	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
CAW_30	0.000	0.115	0.000	0.077	0.231	0.192	0.000	0.038
CAW-C_30	0.000	0.038	0.000	0.000	0.077	0.077	0.038	0.000
RRM_30	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
RMC_30	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
BEKK-C	0.000	0.000	0.000	0.000	0.538	0.423	0.000	0.077

(continued)

Table 1 (continued)

Weights	EW	R1	R2	R3	EW	R1	R2	R3
Loss	Caporin LF				Lopez LF			
BEKK	**0.923**	**0.846**	0.423	0.462	**1.000**	**1.000**	0.500	0.538
cDCC-C	0.308	0.231	0.000	0.000	0.538	0.385	0.000	0.077
cDCC	0.308	0.231	0.000	0.000	0.538	0.385	0.000	0.077
DCC-HR	0.231	0.115	0.000	0.000	0.538	0.385	0.000	0.077
DECO	0.538	0.000	**0.885**	**0.846**	0.538	0.423	**0.962**	**0.808**
RM	0.000	0.000	0.000	0.000	0.231	0.192	0.000	0.000
MC	0.000	0.000	0.000	0.000	0.385	0.192	0.000	0.000
#(MCS)	2.308	1.923	1.654	1.923	6.769	4.385	1.654	2.154

Notes: A model's label followed by "δ" means that model has been estimated using RC matrices built from intradaily returns computed over an interval of length equal to $\delta = 5, 10, 15, 30$ min. A model's label followed by "-C" means that model has been estimated by means of the cQMLE. RRM and RMC stand for realized RiskMetrics [15] and realized Moving Covariance estimator. Models based on daily log-returns are: the BEKK model [10], the corrected DCC (cDCC) [1], the DCC [8] specification as proposed by Hafner and Reznikova [12] (DCC-HR), the Dynamic Equicorrelation (DECO) of [9], the RiskMetrics (RM) and the Moving Covariance (MC) models. #(MCS) is the size of the MCS obtained under a given loss function, on average. The significance level of the tests of the MCS procedure is $\alpha = 0.25$. The period under consideration ranges from July 2, 1997 to July 18, 2008 for a total of 2744 trading days. Models entering the MCS with the highest frequency are indicated in bold

Table 2 *Evaluation* MCS P-values

Weights	EW	R1	R2	R3	EW	R1	R2	R3
Loss	Caporin LF				Lopez LF			
H_t^{Comb}	**1.000**	**1.000**	**0.447**	**0.398**	**1.000**	**0.930**	**0.253**	**1.000**
H_t^{All}	**0.948**	**0.935**	0.156	0.148	**0.725**	**0.702**	0.014	0.123

Notes: The significance level of the tests of the MCS procedure is $\alpha = 0.25$. Bold p-values denote the statistically inclusion within the *evaluation* MCS

References

1. Aielli, G.P.: Dynamic conditional correlation: on properties and estimation. J. Bus. Econ. Stat. **31**(3), 282–299 (2013)
2. Amendola, A., Candila, V.: Evaluation of volatility predictions in a VaR framework. Quant. Finance **16**(5), 695–709 (2016)
3. Amendola, A., Storti, G.: Model uncertainty and forecast combination in high-dimensional multivariate volatility prediction. J. Forecast. **34**(2), 83–91 (2015)
4. Bandi, F.M., Russell, J.R.: Microstructure noise, realized variance, and optimal sampling. Rev. Econ. Stud. **75**(2), 339–369 (2008)
5. Barndorff-Nielsen, O., Shephard, N.: Econometric analysis of realized covariation: high frequency based covariance, regression, and correlation in financial economics. Econometrica **72**(3), 885–925 (2004)
6. Bauwens, L., Laurent, S., Rombouts, J.V.K.: Multivariate GARCH models: a survey. J. Appl. Economet. **21**(1), 79–109 (2006)

7. Caporin, M.: Evaluating value-at-risk measures in the presence of long memory conditional volatility. J. Risk **10**(3), 79–110 (2008)
8. Engle, R.F.: Dynamic conditional correlation: a simple class of multivariate generalized autoregressive conditional heteroskedasticity models. J. Bus. Econ. Stat. **20**(3), 339–350 (2002)
9. Engle, R.F., Kelly, B.T.: Dynamic equicorrelation. J. Bus. Econ. Stat. **30**(2), 212–228 (2012)
10. Engle, R.F., Kroner, K.F.: Multivariate simultaneous generalized ARCH. Econ. Theory **11**(01), 122–150 (1995)
11. Golosnoy, V., Gribisch, B., Liesenfeld, R.: The conditional autoregressive Wishart model for multivariate stock market volatility. J. Economet. **167**(1), 211–223 (2012)
12. Hafner, C.M., Reznikova, O.: On the estimation of dynamic conditional correlation models. Comput. Stat. Data Anal. **56**(11), 3533–3545 (2012)
13. Hansen, P.R., Lunde, A., Nason, J.M.: The model confidence set. Econometrica **79**(2), 453–497 (2011)
14. Lopez, J.A.: Evaluating the predictive accuracy of volatility models. J. Forecast. **20**(2), 87–109 (2001)
15. J.P. Morgan Guaranty Trust Company: RiskMetrics-Technical Document (1996)
16. Pakel, C., Shephard, D., Sheppard, K., Engle, R.F.: Fitting vast dimensional time-varying covariance models. Technical report, New York University (2014)
17. Zhang, L.: Estimating covariation: epps effect, microstructure noise. J. Economet. **160**(1), 33–47 (2011)

Automatic Detection and Imputation of Outliers in Electricity Price Time Series

Ilaria Lucrezia Amerise

Abstract In high frequency time series of electricity prices, one frequently observes a feature which is common for most electricity markets, namely sudden extreme prices. The present study relates to a method for automatically determining and replacing outliers. The core of our method is the construction of a reference time series through the rolling decomposition into trend-cycle and seasonal components of the original time series. Deviations of residuals above a given threshold indicate anomalies which are replaced with a more reasonable alternative price.

Keywords Time series · Anomalies · Electricity markets

1 Introduction

Electricity prices time series are characterized by spikes, that is, significant peaks and jumps. The "commonness" of aberrant prices, however, does not absolve us from trying to use more effective albeit more invasive treatment of outliers. On the one hand, even if the removal of legitimate data points in a time series could be accepted as a permissible practice, the number of values that could be suspected of being anomalous in the time series discussed in this study, is too large to justify their exclusion. Electricity price peaks and/or jumps, in fact, need not be treated as enemies, because they are very important for energy market participants. On the other hand, the choice of the statistical method to employ should not be dependent on such extremes unless they contain information that we cannot afford not to consider.

The main purpose of this paper is to present a new method to detect and impute outliers observed in time series of electricity prices. The essence of the procedure is a rolling trigonometric-polynomial regression, which is used to build a reference

I. L. Amerise (✉)
Department of Economics, Statistics and Finance, University of Calabria, Rende, CS, Italy
e-mail: ilaria.amerise@unical.it

M. Corazza et al. (eds.), *Mathematical and Statistical Methods for Actuarial Sciences and Finance*, https://doi.org/10.1007/978-3-319-89824-7_8

time series (that is, a smoothed version of the investigated time series). The price at the time point in which the difference between observed and reference time series is, in absolute terms, greater than a prefixed threshold, is considered anomalous and replaced with the corresponding price of the reference time series.

The paper is organized as follows. In Sect. 2, we present our basic method. The case studies in Sect. 3 serve to exemplify how the method can be used as a data preparation tool to ensure that statistical modeling is supported by valid data. In particular, data for this article consist of historical data on hourly zonal prices traded at the day ahead Italian energy market.

2 Time Series Outlier Detection and Imputation

Let us suppose that the time series electricity price (if necessary, log-transformed) is generated from

$$P_t = x\,(t) + y\,(t) + e_t, \qquad t = 1, 2, \cdots, n\ , \tag{1}$$

where $x(t)$ is the sum of orthogonal polynomials in t of degree $1, 2, \cdots, m_1$.

$$x\,(t) = \sum_{j=1}^{m_1} \beta_j g_j\,(t)\,, \quad \text{with} \quad \sum_{t=1}^{n} g_j\,(t)\, g_i\,(t) = 0,\ i \neq j\ . \tag{2}$$

The term $x\,(t)$ reflects the influence of events having a medium or long-term impact on the time series level. The positive integer m_1 is a tuning factor of the procedure. The term $y(t)$ expresses the seasonal fluctuations as a sum of harmonics

$$y\,(t) = \sum_{j=1}^{m_2} \sum_{i=1}^{m_{3,j}} \gamma_{j,i} \cos\left(i\omega_j t\right) + \delta_{j,i} \sin\left(i\omega_j t\right)\ , \tag{3}$$

where m_2 is the number of seasonal periods and $m_{3,j}$, $j = 1, 2, \cdots, m_2$ is the number of harmonics for each seasonal period; $\omega_j = 2\pi/s_j$ is the first Fourier coefficient, with s_j being the number of seasons into the j-th periodic component. Note that, because of collinearity one trigonometric term of the last harmonic of each seasonality is omitted. We assume that the seasonalities are known and specified. There are methods that allow uncovering of cycles in a time-series, but lie beyond the scope of the present study. In this regard, we found very useful the chi-square periodogram method proposed by [4].

Finally, e_t is a white noise process (i.e. mutually independent and identically distributed random variables), which includes anything else in the time series and t is the day of measurement. In the following, (1) will be referred to as the basic model. This same approach has been followed by [3] in the function *tsoutliers* of the *R* package *forecast*. However, the volatility and the length of the time series usually

observed for the electricity prices is rather long, even in the cases of a separate treatment for each hour of the day. We believe, therefore, that (1) is not suitable to represent the entire time series and, consequently, an alternative method is needed.

In this study, the time series is subdivided into a set of overlapping time points. Let ν be an odd integer that represents the number of sub-series into which the time series is divided. The value of ν is also a tuning factor of our method. In the initial step, the sub-series contains prices for day 1 through ν. The coefficients $\boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\delta}$ of the basic model are estimated by ordinary least squares (OLS) and the fitted values within the range of the first sub-series are obtained. In the successive step, a window of ν contiguous prices is moved along the time series, producing a sequence containing prices for day 2 through day $(\nu + 1)$. The basic model is fitted to the new sub-series and another set of fitted values are obtained. The procedure is iterated until the basic model achieves the last sub-series covering the time points $(n - \nu + 1), \cdots, n$. The *Berliner Verfahren* (Berlin procedure), discussed in [1] is based on these premises.

At the end of the iterations, we have accumulated a variable number of fitted values for each observed price. The sequence of repetitions is as follows: $1, 2, \cdots, \nu - 1, \nu, \cdots, \nu$, where ν is replicated for $n - 2(\nu - 1)$ times, $\nu - 1, \nu - 2, \cdots, 1$. Here we use the median of the replicated fitted values to estimate the individual price of the reference time series: $\widehat{P}_1, \widehat{P}_2, \cdots, \widehat{P}_n$. Let $\widehat{e}_t = P_t - \widehat{P}_t,\ t = 1, 2, \cdots, n$ be the time series of the estimated residuals. These residuals are labelled as outliers if they lie outside the range $M - k * \text{mad}\,(\widehat{e}_t)\,,\ M + k * \text{mad}\,(\widehat{e}_t)$, where M is the median of the residuals and mad() is the median of absolute deviations from the median. When $k = 1.4826$ mad() is equal to standard deviation of a Gaussian distribution. The filter becomes more tolerant as k increases, allowing more spikes and/or jumps to be accepted as regular values. Decreasing k makes the filter more invasive, declaring more prices to be local outliers and replacing them with the corresponding price in the reference time series. Of course, k is another tuning factor.

3 Results and Discussion

The data for this article consist of historical data on hourly zonal price traded at the day ahead Italian energy market. Because of transmission capacity constraints, Italy is partitioned into six zones or macro-regions: North, Centre-North, Centre-South, South, Sardinia, Sicily with a separate price for each zone. More details in [2].

Given the strong seasonalities present both in the intra-day and inter-day dynamics and the large amount of data which is generally available, we propose to treat each hour of the day as a separate time series and hence we will analyze 24 different daily time series, one for each hour of the day (and for each zone). The time series analyzed goes from 1am on Monday, 07/01/2013 to 24pm on Sunday, 26/02/2017 and hence cover a total of 24 hourly prices in 1148 days for six zones.

To assess the potentiality of our procedure, we compare the results of the routine *tsoutliers* with those obtained by the method proposed in this study, which

we call diats (detection and imputation of anomalies in time series) defined by $\nu = \lceil 2n/\log_2(n) \rceil$ (plus one if the result is even), which for $n = 1148$, implies $\nu = 227$. The other tuning factor is $m_1 = 5$ (we used a higher degree with respect of the traditional third-degree polynomial model to take more into account the oscillatory bevavior of the prices and, consequently, favor a better local fit. Furthermore, $m_2 = 7, m_3 = 4, k = 4.5$. We apply the two routines to $24 \times 6 = 144$ daily time series of hourly zonal prices in Italy.

Our findings indicate that *diats* is broadly more aggressive than *tsoutliers* since reject more prices in more than two-third of all time series. In fact, the average number of rejections is $20/1148$ for the former and $18/148$ for the latter. The two procedures replace the same prices with an average of $9/1148$ time points. Figure 1 illustrates the relationship by means of a scatter plot with loess regression curves. We have ascertained that for time series of moderate length, the calculations necessary to implement *diats* are heavier than with *tsoutliers*, but remain compatible with the hardware resources generally available. Nonetheless, for big data time series, it could be necessary to intervene on the window size and on the number of

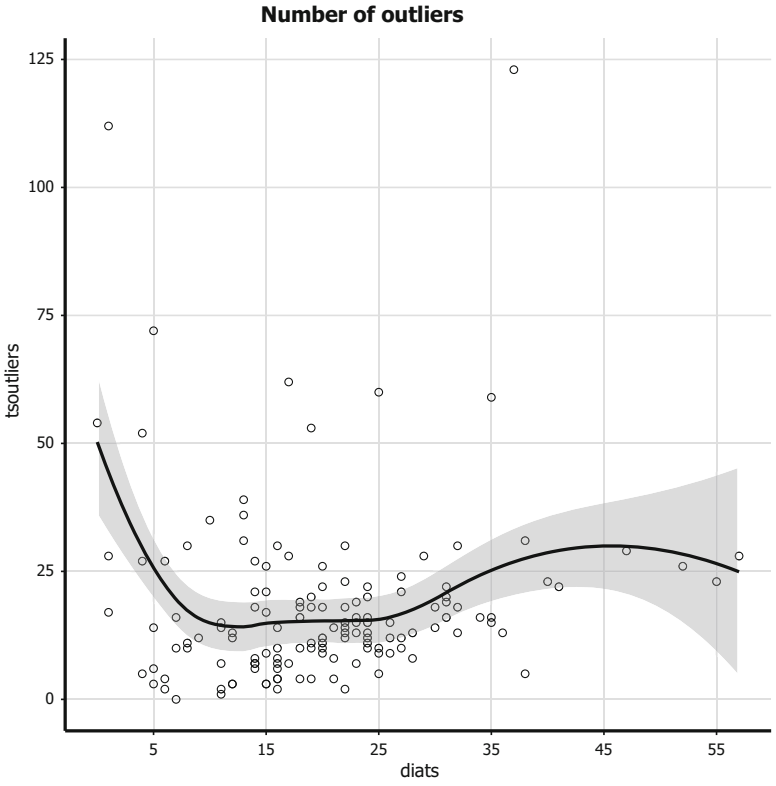


Fig. 1 Comparison between two methods for dealing with outliers

values to be changed from one window to another. Finally, it must be observed that the type of routine discussed in this study are far from perfect, but the experiments performed on real and simulated data with Box-Jenkins sarimax processes, confirm that, not only both *diats* and *tsoutliers* confer beneficial performance characteristics to the models, but also limit the interferences, being based on a different approach.

References

1. Cieplik, U.: BV4.1 Methodology and User-Friendly Software for Decomposing Economic Time Series. German Federal Statistical Office, Berlin (2006)
2. Gianfreda, A., Grossi, L.: Forecasting Italian electricity zonal prices with exogenous variables. Energy Econ. **34**, 2228–2239 (2012)
3. Hyndman, R.J., Khandakar, Y.: Automatic time series forecasting: the forecast package for R. J. Stat. Softw. **26**, 1–22 (2008)
4. Sokolove, P.G., Bushel, W.N.: The chi square periodogram: its utility for analysis of circadian rhythms. J. Theor. Biol. **72**, 131–160 (1978)

Bayesian Factorization Machines for Risk Management and Robust Decision Making

Pablo Angulo, Víctor Gallego, David Gómez-Ullate, and Pablo Suárez-García

Abstract When considering different allocations of the marketing budget of a firm, some predictions, that correspond to scenarios similar to others observed in the past, can be made with more confidence than others, that correspond to more innovative strategies. Selecting a few relevant features of the predicted probability distribution leads to a *multi-objective optimization* problem, and the Pareto front contains the most interesting media plans. Using expected return and standard deviation we get the familiar two moment decision model, but other problem specific additional objectives can be incorporated. *Factorization Machines*, initially introduced for recommendation systems, but later used also for regression, are a good choice for incorporating interaction terms into the model, since they can effectively exploit the sparse nature of typical datasets found in econometrics.

Keywords Data-driven decision making · Decision-support systems · Bayesian regression · Factorization machines · Multi-objective optimization · Risk metrics

1 Introduction

In the marketing industry, a batch of advertising slots is bought on a yearly basis, as this allows better pricing from the media retailer. Our problem is to assist a one-time decision that involves distributing a fixed advertising budget over a 1 year period.

P. Angulo (✉)
ETSIN, UPM, Madrid, Spain
e-mail: pablo.angulo@upm.es

V. Gallego · D. Gómez-Ullate
ICMAT, Madrid, Spain
e-mail: victor.gallego@icmat.es; david.gomez-ullate@icmat.es

P. Suárez-García
Depto. Física Teórica, Facultad de Física, UCM, Madrid, Spain
e-mail: pasuarez@fis.ucm.es

M. Corazza et al. (eds.), *Mathematical and Statistical Methods for Actuarial Sciences and Finance*, https://doi.org/10.1007/978-3-319-89824-7_9

On each week, advertising budget can go into several *channels*, such as TV, Radio or Out-of-Home. A *strategy* is a choice of investment on each week and channel.

The decision should be based on historic data: a time series of N observations $y_t \in \mathbb{R}^R$ for time periods $t = 1, ..., N$, plus, for each time period, a set of predictor variables x_t^i that media specialists believe to be correlated with sales. For one particular customer, a chain of fast food restaurants, y_t^r are the total sales of restaurant $r = 1 \ldots R$ in week $t = 1 \ldots N$, and the predictor variables represent the effect of climate, sport events, special holidays, socioeconomic indicators such as unemployment or inflation, and of course the investments in advertisements during that week. We split the predictor variables into two sets:

- the variables that we cannot control: weather, events, economic indicators, etc... Some of them are real valued (unemployment, mean temperature, ...), while others are binary (Christmas, Easter, major sport event, ...). The holiday type and events variables are *sparse*, since most of them are zero at any given week.
- the variables that we control, i.e. the variables that specify an investment strategy. All of them are real and positive.

We know the values of some of the predictor variables in the first set with certainty (events and holidays variables), while for others we only have probabilistic forecasts (weather and socioeconomic variables). We are allowed to fix the investment strategy, given some constraints such as total budget.

In [2], we predict sales for the next week, as a function of investments, with the knowledge of all previous weeks, and adapt the control variables each week. This setup is now being used at media analytics company Annalect, who ordered this study.

2 Prediction

We are given a set of N observations, $\{(x_1, y_1), ..., (x_N, y_N)\}$, where $x_t \in X$ is the feature vector of the t-th week and $y_t \in \mathbb{R}^R$ is the target: the sales y_t^j at each of the $j = 1 \ldots R$ restaurants and each time $t = N + 1 \ldots N + t$.

Factorization Machines [6] use a quadratic function where the matrix for the quadratic part has rank at most k:

$$g(x) := w_0 + \sum_{i=1}^{p} w_i x_i + \sum_{i=1}^{p} \sum_{j=i+1}^{p} \langle v_i, v_j \rangle x_i x_j \tag{1}$$

- w_0 is the global bias.
- w_i captures the effect of the i-th feature.
- $\langle v_i, v_j \rangle$ captures the interaction effect between features i and j, but using one latent factor $v_i \in \mathbb{R}^k$ per feature.

FMs requires $(k+1)p$ parameters, while the full-rank second order version needs $\sim p^2/2$. This is a critical aspect of FMs making it suitable for fitting small size datasets that arise in many business contexts. Factorization machines in particular, and factorization models in general, have been widely employed in tasks such as recommender systems or ad click prediction [1, 4, 6], problems characterised by the prevalence of outrageously big datasets. We show that these models may also be helpful for other kind of datasets in which observations are scarce and there are sparse blocks in the data matrix X.

We add another block of predictor variables: x^r are binary, and there is one for each restaurant $r = 1 \dots R$, so that exactly one of them takes the value 1 at any data point. The model identifies each restaurant with its mean w_i and its feature vector $v_i \in \mathbb{R}^k$, and this forces the model to generalize.

In *Bayesian Parametric Regression*, the mean of the target variable is a deterministic function of the predictor variables, but the function depends on a few unknown parameters. We assume that the function belongs to the FM family and the distribution is a normal with fixed variance (that we estimate later).

$$y = w_0 + \sum_{i=1}^{p} w_i x_i + \sum_{i=1}^{p} \sum_{j=i}^{p} \langle v_i, v_j \rangle x_i x_j + \varepsilon, \quad \varepsilon \sim \mathcal{N}(0, \sigma^2) \tag{2}$$

In other words, the likelihood of y conditioned to the model parameters is

$$p(y|w_0, w_i, v_i) \sim \exp\left(\left(\frac{y - w_0 + \sum_{i=1}^{p} w_i x_i + \sum_{i=1}^{p} \sum_{j=i}^{p} \langle v_i, v_j \rangle x_i x_j}{\sigma}\right)^2\right) \tag{3}$$

Since the parameters are unknown, we model our uncertainty about them with a prior probability distribution, and Bayes theorem gives the posterior belief about the model parameters. For fixed values of the predictor variables and the model parameters, (2) gives the sales of the restaurants. We integrate our posterior probability measure over the set of parameters and we also integrate over the distributions of the predictor variables that we don't known with certainty. If we add the sales of all the restaurants, we get a probability distribution for a single real number.

A straightforward application of Bayes theorem leads to untractable integrals, so we use the *Markov Chain Monte Carlo* (MCMC) method [3, ch 12]. MCMC replaces the probability measure by a representative sample that is obtained by performing a random walk on the feature space X, but one that is modulated by a multiple of the posterior density, which can be computed as the product of the prior distribution and the likelihood.

The optimal decision problem relies heavily on our ability to make good forecasts of the sales in the future, not only of its expectation but also of the variance and other features of its probability distribution. With the aim of measuring the quality of our predictions, we follow the standard procedure of splitting the data set into a training

and a test set. The first one is used to learn the posterior probability distribution whereas the test set is employed to estimate the performance of the model.

We can compare the posterior mean with the observed sales to get a first measure of the quality of the predictions, but we must also calibrate our estimations of the variance. For any week and restaurant in the test set, our prediction is a different probability distribution, and we only get one sales value for each such distribution. In order to solve this, we apply the probability integral transform, that takes any probability measure into the uniform distribution in the interval [0, 1]. We get in this way a sample that we can compare to the [0, 1]-uniform distribution, both for measuring goodness-of-fit and for selecting hyperparameters.

3 Multiobjective Optimization

With this predictive model we consider a multiobjective optimization problem over the control variables. The typical choice for multiobjective optimization function in financial settings is maximizing expected sales while minimizing expected variance, but other problem specific additional objectives can be incorporated.

For the chain of fast food restaurants, we added the *"restaurants at risk"* metric (RAR). An innovative strategy might increase total sales by increasing sales in a few big restaurants, but at the same time disappoint many restaurant owners, who believe that the chosen strategy harms their restaurant in particular.

The RAR is the expected number of restaurants whose sales figure with the new plan will be below the 5% quantile of the base plan. The RAR is not zero (but 5%!) if the new plan is actually the same as the base plan.

In order to find the Pareto frontier, we use the technique of *scalarization*, in which the different objectives are combined into a single function in different ways. There are many alternative methods [5], but the simplest weighted sum method was good enough: maximize a linear combination of the multiple objectives with different weights, and vary the weights to get new points in the Pareto frontier.

In the end, the outcome of our model is a representative set of Pareto optimal investment strategies for the set of objective functions. The human decision maker can choose among this small set of concrete strategies, which is more convenient than elicitation of the full utility function.

References

1. Freudenthaler, C., Schmidt-Thieme, L., Rendle, S.: Bayesian factorization machines. In: Workshop on Sparse Representation and Low-Rank Approximation. Neural Information Processing Systems (NIPS), Granada (2011)
2. Gallego, V., Angulo, P., Suárez-Garcia, P., Gómez-Ullate, D.: Sales forecasting and risk management under uncertainty in the media industry (2018). http://arxiv.org/abs/1801.03050

3. Gelman, A., Carlin, J.B., Stern, H.S., Dunson, D.B., Vehtari, A., Rubin, D.B.: Bayesian Data Analysis, 3rd edn. CRC Press, Boca Raton (2014)
4. Juan, Y., Zhuang, Y., Chin, W.S., Lin, C.J.: Field-aware factorization machines for CTR prediction. In: Proceedings of the 10th ACM Conference on Recommender Systems, pp. 43–50. ACM, New York (2016)
5. Marler, R.T., Arora, J.S.: Survey of multi-objective optimization methods for engineering. Struct. Multidiscip. Optim. (2004). https://doi.org/10.1007/s00158-003-0368-6
6. Rendle, S.: Factorization machines. In: Proceedings of the 2010 IEEE International Conference on Data Mining, pp. 995–1000. IEEE Computer Society, Washington, D.C. (2010)

Improving Lee-Carter Forecasting: Methodology and Some Results

Giovanna Apicella, Michel M. Dacorogna, Emilia Di Lorenzo, and Marilena Sibillo

Abstract The aim of the paper is to improve the Lee-Carter model performance developing a methodology able to refine its predictive accuracy. Considering relevant information the discrepancies between the real data and the Lee-Carter outputs, we model a measure of the fitting errors as a Cox-Ingersoll-Ross process. A new LC model is derived, called mLC. We apply the results over a fixed prediction span and with respect to the mortality data relating to the Italian females aged 18 and 65, chosen as examples of the model application. Through the backtesting procedure within a static framework, the model mLC proves itself to outperform the LC model.

Keywords Backtesting methods · Cox-Ingersoll-Ross process · Lee-Carter model · Out-of-sample forecasting performance

1 Introduction and Literature

Longevity and its consequences are among the most crucial concerns for actuaries. As stressed in [9], in a number of actuarial calculations, especially those regarding pensions, life annuities and generally speaking the insurance product in case of life, allowing for future mortality trends is required. In order to avoid underestimation

G. Apicella (✉)
University of Rome "La Sapienza", Rome, Italy
e-mail: giovanna.apicella@uniroma1.it

M. M. Dacorogna
DEAR-Consulting, Zürich, Switzerland
e-mail: michel@dacorogna.ch

E. Di Lorenzo
University of Naples Federico II, Naples, Italy
e-mail: diloremi@unina.it

M. Sibillo
University of Salerno, Fisciano (SA), Italy
e-mail: msibillo@unisa.it

M. Corazza et al. (eds.), *Mathematical and Statistical Methods for Actuarial Sciences and Finance*, https://doi.org/10.1007/978-3-319-89824-7_10

of the relevant liabilities, the insurance company (or the pension plan) must adopt an appropriate forecast of future mortality, which should account for the most important features of past mortality trends. Furthermore, basing on what experienced during the twentieth century, when longevity increased much more consistently than it was expected to, it is very important to be able to catch the speed of the mortality decreasing. Forecasting mortality is the subject matter for an ongoing study among actuaries and demographers and, as a result, a number of stochastic mortality projection models have been developed over time (cf. [4]). According to [6], a "good" model should produce forecasts that perform well out-of-sample when evaluated using appropriate forecast evaluation or backtesting methods, as well as provide good fits to the historical data and plausible forecasts ex ante.

This paper aims at providing a further insight into the methodological approach for improving the predictive accuracy of the existing mortality projection models, proposed in [1]. In this paper the basic idea is to dynamically model the fitting errors of a survival function chosen as the baseline, by a Cox- Ingersoll-Ross process (cf. [5]). This approach gives rise to a new version of the baseline, profiting by the past errors information captured by the CIR multiplicative role. In the same paper, in the particular case of the Cairns-Black-Dowd model (cf. [3]), the methodology is developed and the survival model performance analyzed comparing it with the modified one. The new version of the CBD model proves to be a parsimonious model, providing better results, in terms of predictive accuracy, than the CBD model itself. Aim of this paper is to test the attitude of the corrective methodology developed in [1] in adjusting the forecasting performance of the Lee-Carter (LC) mortality projection model (cf. [8] and [2]), in the case of the Italian females and the time horizon [1906–2012]. Our data are taken from the HMD [7].

For brevity, after reviewing the mathematical framework and the empirical methodology in Sect. 2, in Sect. 3 we show and remark on two graphical comparisons, chosen as examples of applications, between the central projections provided by the Lee-Carter model, both with and without our correction, plotted against the realized death rates. In Sect. 4, conclusions are given.

2 Mathematical Framework and Empirical Methodology

In this Section, we recall the key idea illustrated in [1] and explain how, in the modelling framework set out in this paper, "adjusted" projections can be obtained, starting from the baseline provided by the Lee-Carter model. Our aim is to compare the forecast of mortality against the realized one, by using a static backtesting procedure. Accordingly, we split the sample of available reliable mortality data into two fixed time intervals: the "lookback" window (cf. [6]), from 1906 up to 1977, and the "lookforward" window, from 1978 up to 2012. We denote by x the age and by t the calendar year when the age is measured.

1. Over the lookback window, we consider the rv $Y_{x,t}$, describing the following measure of the fitting error of any selected mortality model:

$$Y_{x,t} = B_{x,t}/\mu_{x,t} \tag{1}$$

   where $B_{x,t}$ is the observed central death rate and $\mu_{x,t}$ is the baseline provided by the mortality model for describing the same rate. Throughout the lookback window, we model $Y_{x,t}$ as a Cox-Ingersoll-Ross process and look at the optimal parameters coming from its calibration over the same time horizon.
2. Over the lookforward window, we construct the "adjusted" projections $\tilde{B}_{x,t}$, in a multiplicative way, as follows:

$$\tilde{B}_{x,t} = \tilde{Y}_{x,t}\tilde{\mu}_{x,t} \tag{2}$$

   where $\tilde{\mu}_{x,t}$ is the forecast coming from the mortality model (baseline) and $\tilde{Y}_{x,t}$ is the "best estimate" of the correction factor $Y_{x,t}$.

In the context of this paper, $\tilde{B}_{x,t}$ is the outcome of the "mLC" model. Such a "new" model results from the combination of the deterministic forecasting output of the Lee-Carter model and the stochastic correction CIR factor $Y_{x,t}$, whose estimate, over the prediction time span, is $\tilde{Y}_{x,t}$.

3 Graphical Assessment of the Predictive Accuracy of the "mLC" Model

In this Section, we perform a graphical analysis, for assessing the quality of the forecasts and gaining insight into their dynamics. In Fig. 1, for providing meaningful examples of application, we display, over the lookforward window [1978, 2012], for ages 18 and 65: the realized death rates, $B_{\ddot{x},t}$ (the gray line); the LC projection of death rates, $\dot{m}_{\ddot{x},t}$ (the continuous black line); the "mLC" projection of death rates, $\breve{m}_{\ddot{x},t}$ (the dotted black line). As to age 18 (graph on the left), the LC model steadily predicts lower death rates than the realized ones, whereas this happens with the mLC model only from 1988 on. Except for 1978, the mLC projection turns out to be closer to the observed death rates than the LC projection; using the mLC model enables us to reduce the Root Mean Square Error of the LC model itself by 33%. As to age 65 (graph on the right), the mLC projection is almost indistinguishable from the LC one; indeed, the CIR correction lets us earn only 0.2% in terms of predictive accuracy. In this case, due to the very good fit of the LC model to the observed data over the lookback period, the predicted long-term mean of the CIR process is very close to 1; therefore, $\tilde{Y}_{65,t}$ does not play any significant role in Eq. (2).

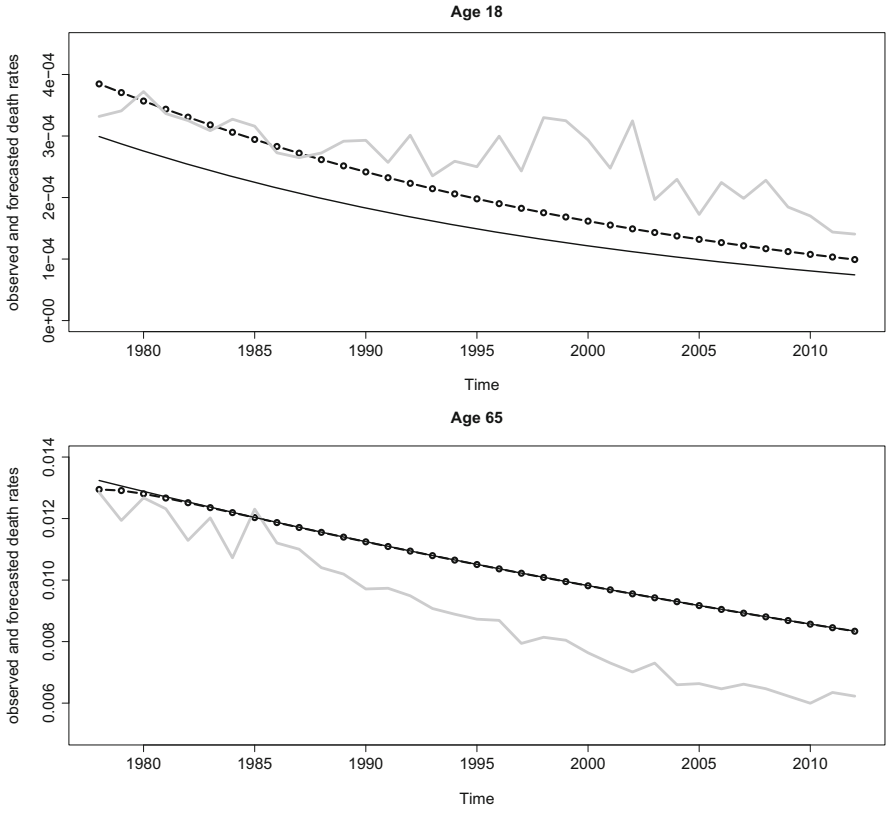


Fig. 1 Age 18 and 65. Observed and forecasted death rates over the lookforward window [1978–2012]. Gray line: observed death rates; continuous black line: LC central projection; dotted black line: "mLC" central projection

4 Concluding Remarks

The aim of this paper is to investigate on the capability of the corrective methodology proposed in [1] to increase the predictive accuracy of the LC model. Under a static approach for backtesting both the LC and mLC model, our empirical analyses, performed with respect to the Italian females aged 18 and 65, show that the CIR correction factor, $Y_{18,t}$, is able to drive the deterministic baseline towards the realized phenomenon over the lookforward window, although with a moderate speed, while $Y_{65,t}$ affects the baseline only very slightly. It is thus worth advancing the research by deepening the role of the CIR factor as mortality model benchmark and, also, by using dynamic out-of-sample validation methods, in order to see if the parameters of the CIR process adapt to the changes in the underlying mortality data, thus providing up-to-date information to exploit in the prediction phase.

References

1. Apicella, G., Dacorogna, M.M., Di Lorenzo, E., Sibillo, M.: Using interest rate models to improve mortality forecast. https://ssrn.com/abstract=3070891. https://doi.org/10.13140/RG.2.2.25183.56484 (2017)
2. Brouhns, N., Denuit, M., Vermunt, J.K.: A Poisson log-bilinear approach to the construction of projected lifetables. Insur. Math. Econ. **31**(3), 373–393 (2002)
3. Cairns, A.J.G., Blake, D., Dowd, K.: A two-factor model for stochastic mortality with parameter uncertainty: theory and calibration. J. Risk Insur. **73**(4), 355–367 (2006)
4. Cairns, A.J.G., Blake, D., Dowd, K., Coughlan, G.D., Epstein, D., Ong, A., Balevich, I.: A quantitative comparison of stochastic mortality models using data from England and Wales and the United States. N. Am. Actuar. J. **13**(1), 1–35 (2009)
5. Cox, J.C., Ingersoll, E.J., Ross, S.A.: A theory of the term structure of interest rates. Econometrica **53**(2), 385–407 (1985)
6. Dowd, K., Cairns, A.J.G., Blake, D., Coughlan, G.D., Epstein, D., Khalaf-Allah, M.: Backtesting stochastic mortality models: an ex post evaluation of multiperiod-ahead density forecasts. N. Am. Actuar. J. **14**(3), 281–298 (2013)
7. Human Mortality Database: University of California, Berkeley (USA), and Max Planck Institute for Demographic Research (Germany). http://www.mortality.org or http://www.humanmortality.de (2016) (data downloaded on October 20, 2016)
8. Lee, R.D., Carter, L.R.: Modelling and forecasting U.S. mortality. J. Am. Stat. Assoc. **87**(14), 659–675 (1992)
9. Pitacco, E., Denuit, M., Haberman, S., Olivieri, A.: Modelling Longevity Dynamics for Pensions and Annuity Business. Oxford University Press, Oxford (2009)

The Bank Tailored Integrated Rating

Daniela Arzu, Marcella Lucchetta, and Guido Massimiliano Mantovani

Abstract We develop a banks specific integrated rating, tailored incorporating the various heterogeneity dimensions characterizing financial institutions (see Mantovani et al., Int Res J Appl Finance IV:458–489, 2013 and Mantovani et al., J Bus Econ Finance 3:18–49, 2014 regarding the heterogeneity risk analysis in corporate firms), named bank tailored integrated rating (BTIR). The approach is inherently coherent with the challenging frontier of forecasting tail risk in financial markets (De Nicolò and Lucchetta, J Appl Econ 32(1):159–170, 2017) since it considers the downside risk in the theoretical framework. The innovation consists in using the integrated rating (IR) with the pre-selection of the variables through a statistical procedure that takes into account the characteristics of risk and greater heterogeneity of the banks. A Vector Autoregressive Model (VAR) is only a first simple application proposal.

Keywords Bank tailored integrated rating · Banks' heterogeneity · Financial cycle

1 Motivation and Methodology

The capital regulatory policies imposed on banking institutions, increasingly reveal the need to consider the heterogeneity of regulated entities and, at the same time, to avoid obvious errors above or under assessment of the risks inherent in the various business models of modern banks. The corporate performance literature

D. Arzu (✉)
Department of Management, Ca' Foscari University of Venice, Venice, Italy
e-mail: daniela.arzu@unive.it

M. Lucchetta
Department of Economics, Ca' Foscari University of Venice, Venice, Italy

G. M. Mantovani
International University of Monaco, Fontvieille, Monaco

M. Corazza et al. (eds.), *Mathematical and Statistical Methods for Actuarial Sciences and Finance*, https://doi.org/10.1007/978-3-319-89824-7_11

introduces the Lintner's model [1] as an alternative approach to appraise firms and their performance, through the companies' asset-side capability management in the long term. The analysis is useful to understand whether there is an appropriate allocation of financial resources, in line with the goodness of the performance and it is important to assess company pay-out and managerial rents as in Lambrecht and Myers [2]. However, Leibowitz and Henriksson [3] noted that it is important to consider a shortfall approach that looks more on a "confident equivalent", rather that the Lintner's certainty equivalent, which is a minimum threshold that may be overpassed, according to a certain confidence percentage. Determining either the threshold and the confidence is up to the investor, even before choosing the investment. Indeed, in banking analysis the downside risk is particularly important since "tail risk" is considered an important component in financial market analysis as underlined in De Nicolò and Lucchetta [4]. The cited current literature on risk assessment concentrates on corporate firms and the "tail risk" analysis is mainly oriented to macroeconomic risk measures. This paper fills these gaps and contributes to the identification of a synthetic indicator of company performance and long-term creditworthiness, which is also able to take into consideration the investor's risk aversion and the downside risk component: the "bank tailored integrated rating" (BTIR). This need arises from studies on rating modelling in order to make easier the implementation and use of the results within banking organizations. Indeed, it must be ensured that the indicator has three characteristics: (i) scientifically reliable and (ii) comprehensible to customers, finally (iii) consistent with the credit policies adopted. The indicator is inspired by the Integrated Rating methodology [5–7].

2 Stylized Mathematical Approach

In order to start, we have run a panel regression with components suitable for banks and understand whether the main model might be thought for banks:

$$Y_{it} = \beta_0 + \beta_1 X_{1it} + \cdots + \beta_n X_{nit} + \epsilon_i$$

Where Y_{it} is a banks' performance indicator and βs are banks' health characteristics (Appendix).

We hypothesized to transform the indicator, through a logistic transformation, deriving from the logistic function, which is the better fitting methodology into the whole model. The logistic transformation allows us to have an indicator included in a range between 1 and −1 and a unit standardize and concave curvature. However, it is possible to investigate to detect a multiplicative constant in the exponential component, which changes the degree of curvature of the function, going to change the degree of discrimination of the data set, compared to more extreme values.

$$f(x) = \frac{L}{1 + e^{-k(x-x_0)}}$$

for all the real values of x with codomain $[0, L > 0]$, with inflection point in x_0 and with slope $k > 0$.

The logistics transformation allows us not to overestimate xs that have a better performance, than expectations and do not underestimate xs that are in line with expectations. This effect can be regulated by a multiplicative constant in the exponential component and allows to determine a degree of convexity/concavity that can adapt to the needs. This proxies the differences in risks attitude of the institutions. In conclusion, the bank specific integrate rating project, here detailed, focus our research on the development of a mathematical/econometric method that allows us to identify the best algorithm, to determine a correct degree of convexity and concavity (and therefore, consequently, the correct degree of risk aversion of the investor), which can be dynamic and adaptable, consequently to heterogeneous banks. To take into account the characteristics of risk and greater heterogeneity of the banks, we propose a challenge procedure that employs a Vector Autoregressive Model (VAR) to preselect the relevant banks' variables.

$$VAR\ of\ order\ p : y_t = c + \Phi_1 y_{t-1} + \cdots + \Phi_p y_{t-p} + \epsilon_t$$

Where y_t is a banks' performance indicator, Φ_i: (N*N) $\forall i$ are the other banks' health indicators.

We have chosen the model VAR because it is very simple to implement the selection of important variables using a large number of variable vectors. Furthermore, the VAR procedure makes it possible to recognize at system level the components of systemic risk that would otherwise be ignored without such a process. This further step allows us to design our "bank tailored integrated rating" (BTIR). The approach is inherently coherent with the challenging frontier of forecasting tail risk in financial markets.

3 Summaries and Future Developments

The current development of ever-increasing banking regulations requires the study and the development of increasingly precise rating methods that take into account the increasing heterogeneity of banks and the presence of systemic risk, in addition to ongoing contagion relations between financial institutions. Also, the traditional and simple capital regulatory policies imposed on banking institutions, increasingly reveal the need to consider the heterogeneity of regulated entities and, at

the same time, to avoid obvious errors above or under assessment of the risks inherent in the various business models of modern banks. Our work considers the extension of the integrated rating (IR) procedure, used primarily for non-financial companies, developing the "bank tailored integrated rating" (BTIR). The approach is inherently coherent with the challenging frontier of forecasting tail risk in financial markets [4] since it considers the downside risk in the theoretical framework. The innovation consists in using the integrated rating (IR) with the pre-selection of the variables through a statistical procedure that takes into account the characteristics of risk and greater heterogeneity of the banks. In this first proposal, we use a simple VAR. However, our innovative procedure may include, in the future, more sophisticated pre-selection of variables such as CoVARs. This work requires testing whether a more sophisticated pre-selection model is better than a traditional VAR. In fact, for simplicity, we believe that starting with a simple methodology is the first step of research. Our BTIR makes possible to adapt the rating procedures to all banks, even that showing very different characteristics. In fact, the VAR allows to pre-select and to evaluate markets with high systemic risk, avoiding errors due to general market conditions that may differ from country to country. In conclusion, our BTIR opens the door to a new research line to innovative ideas for the development of increasingly accurate ratings for banks embedding the needs of macro- and micro-prudential policies.

Appendix

Where Bank performance indicator is

$$Decomposed\ ROE = \frac{Pre-Tax\ Profit}{Op.Income} * \frac{Tot.Assets}{Equity} * \frac{Net\ Revenue}{Tot.Assets} * \frac{Op.Income}{Net\ Revenue}$$

(i) is Asset Quality; (ii) Capital Ratios; (iii) Operations Ratios; (iv) Liquidity Ratios; (v) Structure Ratio.

		Coefficients	St. errors
	Intercept	21.1780 ∗ ∗∗	1.5268
i	NPL/Gross loans	−0.1550 ∗ ∗∗	0.0353
	NPL/Tot. assets	1.1367 ∗ ∗∗	0.0707
	NCO/Avg gross loans	0.2954 ∗ ∗∗	0.0872
	NCO/Net Inc. bef. Ln Lss Prov.	−0.0028 ∗ ∗∗	0.0008
	Impaired loans/Equity	−0.0789 ∗ ∗∗	0.0038
ii	Equity/Net loans	−0.0625 ∗ ∗∗	0.0176
	Equity/Tot. liabilities	4.5427 ∗ ∗∗	0.3062
iii	Profit margin	0.1510 ∗ ∗∗	0.0115
	Net Int. Rev./Avg ass.	−0.3912∗	0.1613
	Non Int. Exp. Avg Ass.	1.2895 ∗ ∗∗	0.2909
	Pre-Tax Op. Inc./Avg ass.	3.2931 ∗ ∗∗	0.4963
	ROA	6.9032 ∗ ∗∗	0.5271
	Cost to income	−0.0355†	0.0196
iv	Recurring earning power	−1.3532∗	0.6156
	Net loans/Tot. assets	−0.0544 ∗ ∗∗	0.0107
v	Solvency	−6.8087 ∗ ∗∗	0.4127

Total sum of squares: 90,259
Residual sum of squares: 6964.1
R-squared: 0.92284
Adj. R-squared: 0.92193
F-statistic: 1055.44 on 16 and 1345 DF, p-value: <2.22e-16
Signif. codes: ***0.001; **0.01; *0.05; †0.1

References

1. Lintner, J.: The valuation of risk assets and the selection of risky investments in stock portfolios and capital budgets. Rev. Econ. Stat. **47**, 13–37 (1965)
2. Lambrecht, B.M., Myers, S.C.: A Lintner model of payout and managerial rents. J. Financ. **67**(5), 1761–1810 (2012)
3. Leibowitz, M.L., Henriksson, R.D.: Portfolio optimization with shortfall constraints: a confidence-limit approach to managing downside risk. Financ. Anal. J. **45**(2), 34–41 (1989)
4. De Nicolò, G., Lucchetta, M.: Forecasting tail risks. J. Appl. Econ. **32**(1), 159–170 (2017)
5. Mantovani, G.M., Castellan, E.: How to Rate and Score Private Companies? Evidence from the North Eastern Italian Districts (2015). https://papers.ssm.com/abstract=2697090
6. Mantovani, G.M., Daniotti, E., Gurisatti, P.: In search of corporate risk measures to complete financial reporting: the case of the Caldarerie – industry. Int. Res. J. Appl. Finance **IV**, 458–489 (2013)
7. Mantovani, G.M., Gurisatti, P., Corò, G., Mestroni, M.: Toward an integrated rating methodology for corporate risk detection. J. Bus. Econ. Finance **3**, 18–49 (2014). ISSN 2146-7943

A Single Factor Model for Constructing Dynamic Life Tables

David Atance and Eliseo Navarro

Abstract The objective of this paper is to develop a single factor model to construct dynamic life tables. The paper seeks to identify the mortality rate that best explains the global behavior of life tables. Once this key rate is identified, we assume that changes in mortality rates depend linearly on changes in the mortality rate corresponding to the key rate. Next, we proceed to adjust the sensitivities of the changes in mortality rates to changes in the key mortality rate, using non-parametric methods. Assuming that this rate follows a specific ARIMA process it can be used to forecast future mortality rates. The resulting model has a similar structure to the well-known Lee-Carter model but with the advantage that their parameters and variables can be easily identified. Finally, the forecasting ability of the model is tested using out-of-sample data from Spanish experience. The results show that the proposed Single Factor Model significantly outperforms the Lee-Carter model.

Keywords Dynamic life table · Key mortality rate · GLM · Forecasting

1 Single Factor Model

Inspired by literature developed to describe the behavior of the term structure of interest rates and by similarities between life tables (where mortality rates are assumed to depend on age) and the term structure of interest rates (where interest rates depend on the term to maturity), in this paper, we seek to develop a tractable model that can be used to build dynamic life tables. In a model proposed by Elton et al. [3], it is assumed that changes in interest rates with different maturities linearly depend on changes in some key interest rates identified as those that best explain the whole behavior of the term structure of interest rates.

D. Atance (✉) · E. Navarro
University of Alcala, Madrid, Spain
e-mail: david.atance@uah.es; eliseo.navarro@uah.es

M. Corazza et al. (eds.), *Mathematical and Statistical Methods for Actuarial Sciences and Finance*, https://doi.org/10.1007/978-3-319-89824-7_12

In a similar way, the Single Factor Model assumes that the whole life table can be explained by one mortality rate of a particular age, the "key" mortality rate. Particularly, we suppose that:

$$\triangle \ln\left(\hat{q}_{x,t}\right) = \alpha_{x,x^*} + b_{x,x^*} \cdot \left[\triangle \ln\left(\hat{q}_{x^*,t}\right)\right] + \varepsilon_{(x,x^*),t} \tag{1}$$

where:

- $\triangle \ln\left(\hat{q}_{x,t}\right)$ is the variation in the logarithm of the crude[1] mortality rate at age x from year t-1 to year t.
- $\triangle \ln\left(\hat{q}_{x^*,t}\right)$ is the change in the logarithm of the mortality rate at the key age x^* from t-1 to t. This key age will be chosen to maximize the explanatory power of the model.
- α_{x,x^*} is a constant term that captures the general tendency of a reduction (increment) in mortality rates and is assumed to be independent of the behavior of the key mortality rate $\hat{q}_{x^*,t}$. The value of this term may differ from one age to another indicating a differential behavior in the reduction of mortality rates over time.
- b_{x,x^*} is a parameter that describes the sensitivity of the logarithm of the mortality rate at age x to changes in the logarithm of the key morality rate and captures changes in the shape of the mortality curve over time.
- $\varepsilon_{(x,x^*),t}$ is a random error term with zero mean and constant variance $\sigma^2_{\varepsilon,(x,x^*)}$.

We identified the mortality rate with the greatest explanatory power with respect to the whole life table as the mortality rate that maximizes the objective function $\varphi\,(x^*)$ defined as:

$$\max_{x^*} \varphi\left(x^*\right) = \max_{x^*} \sum_{x} R^2_{x,x^*} \cdot \operatorname{var}\left(\triangle \ln\left(\hat{q}_{x,t}\right)\right) \tag{2}$$

1.1 Adjusting a Sensitivity Function to b_{x,x^*}

Once the key age is determined, we can use linear regression techniques to obtain estimates of the parameters α_{x,x^*} and b_{x,x^*}. The second step consists of finding a function $b^*(x)$, $b^*(x)$ to describe the values of $\hat{b}_{x,x^*}$ with two constrains. First, the function must be sufficiently smooth and second, $b^*_{x^*} = 1$.

We propose two different approaches to find the function $b^*(x)$. The first consists of adjusting a parametric function, inspired by Díaz et al. [2]:

$$\hat{b}_{x,x^*} = b^*\,(x) + u_x = \beta_1 \cdot \exp\left[-\beta_2\left(x - x^*\right)^2\right] + (1 - \beta_1) + u_x \tag{3}$$

[1]The model could be implemented using graduated mortality rates. Eventually, we decided to use crude mortality rates to avoid data manipulation.

The second option, was to fit splines to describe the values of $\hat{b}_{x,x^*}$. In this case, two important problems had to be addressed: the number of knots (we apply the criterion proposed by McCulloch [6] and Shea [7]) and to determinate the position of these knots (we select those that minimized the sum squared errors, testing all integer numbers between the maximum and the minimum age considered).

1.2 Forecasting Mortality Rates

The final step in the process of constructing the dynamic life tables consists of developing a methodology to forecast future mortality rates. According to Eq. (1), if we substitute b_{x,x^*} for $b^*(x)$ and rearrange the terms, we obtain:

$$\ln\left(\hat{q}_{x,t}\right) = \ln\left(\hat{q}_{x,t-1}\right) + \alpha_{x,x^*} + b^*(x) \cdot \left[\triangle \ln\left(\hat{q}_{x^*,t}\right)\right] + \eta_{x,t} \tag{4}$$

where:

- $\triangle \ln\left(\hat{q}_{x^*,t}\right)$; represents the change in the logarithm of the mortality rate corresponding to the key age x^* from $t-1$ to t or, alternatively, the relative change in the key mortality rate.
- $\eta_{x,t}$ is an error term with mean zero and variance σ_η^2.

Employing an ARIMA time series to model the behavior of the key mortality rate, and will allows us to forecast the others mortality rates.

2 Lee-Carter (1992) Model

One of the most popular models used to forecast future mortality rates and hence construct dynamic mortality tables was developed and published in a seminal paper by Lee and Carter [5]. This model suggests the adjustment of the central mortality rate $m_{x,t}$ as an exponential function that depends on age x and time t. The classical form of the Lee-Carter model is:

$$m_{x,t} = \exp\left(a_x + b_x k_t + \varepsilon_{x,t}\right) \tag{5}$$

Or equivalently:

$$\ln\left(m_{x,t}\right) = a_x + b_x \cdot k_t + \varepsilon_{x,t} \tag{6}$$

The similarities with our single-factor model are evident. We compared the SFM with a version of the Lee-Carter model [1] where it is assumed that the probability of deaths, $q_{x,t}$ follows a binomial distribution with a logit link and uses GLM to

estimate the model parameters. So, instead of model (6), we assumed that:

$$\ln\left(\frac{q_{x,t}}{1-q_{x,t}}\right) = a_x + b_x \cdot k_t + \varepsilon_{x,t} \tag{7}$$

Predictions of future mortality rates in the Lee-Carter model are based on the adjustment of a time series to the mortality index k_t, following Box-Jenkins methodology. In the original paper, Lee and Carter used an ARIMA(0,1,0). In this paper we apply the *auto.arima* and *forecast* functions of the R package developed by Hyndman [4] to determine the ARIMA model that best fits the data.

3 Comparison Between the Single Factor Model and the Lee-Carter Model

Finally, we compared the Single Factor Model with the Lee-Carter model analyzing the forecasting power of both models. They have been calibrated using data corresponding to the Spanish experience over 1975–2006 period, employing the 2007–2015 period for out-of-sample testing. Ages covered in this study ranges from 18 to 89. Data was obtained from the Instituto Nacional de Estadística (INE, Spanish Statistics National Institute).

The analysis of the out of sample forecasting errors clearly shows a better performance of the SFM over the Lee-Carter model. The results indicate that the SFM produces better results independently of the forecasting horizon employed, although this outcome is much more pronounced in the short term. This better performance of the SFM appears to be concentrated in the range of ages between 18 and 50–60 years old[2] and, in the case of the male population, for people aged 67 and above too.

In summary, the SFM is a very simple and tractable model that is highly effective in forecasting future mortality rates compared with other competing models.

Moreover, the SFM is highly flexible in capturing sudden changes in mortality rates (for instance, the dramatic changes in mortality produced by the AIDS crisis) and can be easily extended to a multifactorial framework.

Its tractability can make of this model a very useful tool for valuing life insurance products and for measuring the risk (notably, longevity risk) inherent in such products.

[2]Fifty in the case of the male population and sixty in the case of the female population.

References

1. Debón, A., Montes, F., Puig, E.: Modelling and forecasting mortality in Spain. Eur. J. Oper. Res. **189**(3), 624–637 (2008)
2. Díaz, A., Merrick, J.J., Navarro, E.: Spanish treasury bond market liquidity and volatility pre- and post-European monetary union. J. Bank. Financ. **30**(4), 1309–1332 (2006)
3. Elton, E.J., Gruber, M.J., Michaely, R.: The structure of spot rates and immunization. J. Financ. **45**(2), 629–642 (1990)
4. Hyndman, R.: Forecast: forecasting functions for time series. R package version 1.11 (2008)
5. Lee, R.D., Carter, L.R.: Modeling and forecasting US mortality. J. Am. Stat. Assoc. **87**(419), 659–671 (1992)
6. McCulloch, J.H.: Measuring the term structure of interest rates. J. Bus. **44**(1), 19–31 (1971)
7. Shea, G.S.: Pitfalls in smoothing interest rate term structure data: equilibrium models and spline approximations. J. Financ. Quant. Anal. **19**(3), 253–269 (1984)

Variable Annuities with State-Dependent Fees

Anna Rita Bacinello and Ivan Zoccolan

Abstract In this paper we consider a variable annuity with guarantees at death and maturity financed through the application of state-dependent fees. We define a general valuation model for them, and propose to apply the LSMC approach in order to analyse the interaction between fee rates, death/maturity guarantees, fee thresholds and surrender penalties under alternative model assumptions and policyholder behaviours. However, special care is needed in the numerical implementation of this approach, due to the shape of the surrender region. We can stem the numerical errors arising in the regression step by using suitable arrangements of the LSMC valuation algorithm.

Keywords Variable annuities · State-dependent fees · Surrender option · LSMC

1 Introduction

Variable annuities are very flexible life insurance contracts that can package living and death benefits with a number of possible guarantees against financial or biometric risks. Typically, a lump sum premium is paid at inception, and is invested in well diversified mutual funds. This initial investment sets up a reference portfolio (*policy account*) and all guarantees are financed by periodical proportional deductions (*fees*) from this account.

Guarantees are often set in such a way that at least the lump sum premium is totally recouped. Then, when the account value is high, the policyholder has an incentive to surrender the contract, stopping to pay high fees for an out-of-the-money guarantee. Conversely, when the account value is low, the policyholder

A. R. Bacinello (✉)
Department of Economics, Business, Mathematics and Statistics 'B. de Finetti', University of Trieste, Trieste, Italy
e-mail: bacinel@units.it

I. Zoccolan
Oracle Italia S.r.l., Cinisello Balsamo, Italy

M. Corazza et al. (eds.), *Mathematical and Statistical Methods for Actuarial Sciences and Finance*, https://doi.org/10.1007/978-3-319-89824-7_13

pays a low fee for an in-the-money guarantee. Summing up, there is an unfair misalignment between costs incurred by the insurer and premiums (fees) to cover them, and a great incentive, for policyholders, to abandon their contracts when they become uneconomical. To eliminate this misalignment and reduce the surrender incentive insurers can adopt a *threshold expense structure*, or *state-dependent fees*, according to which the fees, still proportional to the account value, are paid only if this value is below a given threshold.

In this paper we consider a variable annuity which provides guarantees at death and maturity financed through the application of a state-dependent fee structure, as defined first in [3] and extensively analysed in [4] and [5]. We define a general valuation model for such guarantees, along the lines of [2], and test the application of Least Squares Monte Carlo methods (LSMC), that allow to analyse numerically the interaction between fee rates, death/maturity guarantees, fee thresholds and surrender penalties under alternative model assumptions and policyholder behaviours. In particular, special care is needed when applying these techniques, due to the shape of the surrender region. We can stem the numerical errors arising in the regression step by using suitable arrangements of the LSMC valuation algorithm based on a theoretical result.

The paper is structured as follows. In Sect. 2 we describe the structure of the contract. In Sect. 3 we present our valuation framework. Section 4 is devoted to a discussion of the problems encountered in the numerical implementation of the model and the settlements to overcome them.

2 The Structure of the Contract

Consider a single premium variable annuity contract which provides guarantees at death and maturity. We denote by P the single premium, 0 the time of issuance, T the contract maturity, and assume that the death benefit is paid upon death within the contract maturity. The single premium is invested in a well diversified mutual fund with unit price process S, and the (net) value of the accumulated investments in this fund is referred to as the *policy account value*. We denote by A_t this value at time t. The cost of the guarantees is recouped through the application of a proportional deduction from this account, at a rate denoted by φ (fee rate). However, this deduction is assumed to be made only when the account value is below a given threshold, denoted by β, i.e., we adopt a *state-dependent fee* structure. Of course, in the degenerate case of $\beta = \infty$ (no barrier) we recover a *constant fee* structure.

Both death and maturity benefits contain a guarantee of the *roll-up* type, with the same roll-up rate δ. The death benefit is given by $b_\tau^D = \max\{A_\tau, P\mathrm{e}^{\delta\tau}\}$, $\tau \leq T$, and the survival benefit is $b_T^M = \max\{A_T, P\mathrm{e}^{\delta T}\}$, $\tau > T$, where we have denoted by τ the residual lifetime of the policyholder.

We assume that the contract can be surrendered at any time before maturity, if the insured is alive, and that, in case of surrender at time $\lambda < T \wedge \tau$, the policyholder

receives a cash amount, called surrender value, given by $b_\lambda^S = A_\lambda(1 - p_\lambda)$, where p_λ is a penalty rate, possibly time dependent and such that $0 \leq p_\lambda < 1$ for any λ.

3 Valuation Framework

A key-element in the valuation of the contract from the insurer's point of view is constituted by the behavioral risk. The policyholder, in fact, can choose among a set of possible actions, such as partial or total withdrawal (i.e., surrender), selection of new guarantees, switch between different reference funds, and so on. In particular, in [2] the possible policyholder behaviours are classified, with respect to the aspect concerning partial or total withdrawals, into three categories, characterized by an increasing level of rationality: *static*, *mixed* and *dynamic*. Although in principle partial withdrawals from the account value may be admitted also within the specific contract analysed in this paper, the most relevant valuation approaches for it are the first two, static and mixed.

3.1 The Static Approach

Under this approach it is assumed that the policyholder keeps her contract until its natural termination, that is death or maturity, without making any partial or total withdrawal from her policy account value.

The instantaneous evolution of the account value while the contract is still in force can be formally described as follows:

$$\frac{dA_t}{A_t} = \frac{dS_t}{S_t} - \varphi 1_{\{A_t < \beta\}} dt,$$

where $A_0 = P$ and 1_C denotes the indicator of the event C. Then, the return on the account value is that of the reference fund, adjusted for fees that are applied, according to the fixed rate φ, only when A_t is below the barrier β.

The contract value at time $t < T$, on the set $\{\tau > t\}$, is thus given by

$$V_t = E\left[b_\tau^D \frac{B(t)}{B(\tau)} 1_{\{\tau \leq T\}} + b_T^M \frac{B(t)}{B(T)} 1_{\{\tau > T\}} \middle| \mathscr{F}_t \right],$$

where $B(u) = \mathrm{e}^{\int_0^u r_v dv}$ defines the bank account value accumulated with the risk-free rate r, the expectation is taken under a given risk-neutral measure and the filtration $\mathbb{F} \doteq (\mathscr{F}_t)_{t \geq 0}$ carries knowledge on all financial and biometric variables.

3.2 The Mixed Approach

Under this approach it is assumed that, at any time of contract duration, the policyholder chooses whether or not to exercise the surrender option, and her decision is aimed at maximizing the current value of the contract payoff.

The instantaneous evolution of the account value is the same as in the static approach, while the contract value at time $t < T$, on the set $\{\tau > t,\ \lambda \geq t\}$, is the solution of the following optimal stopping problem: $V_t = \sup_{\lambda \in \mathbb{T}_t} V_t(\lambda)$, where

$$V_t(\lambda) = E\left[b_\tau^D \frac{B(t)}{B(\tau)} 1_{\{\tau \leq T \wedge \lambda\}} + b_T^M \frac{B(t)}{B(T)} 1_{\{\tau > T,\, \lambda \geq T\}} + b_\lambda^S \frac{B(t)}{B(\lambda)} 1_{\{\lambda < \tau \wedge T\}} \middle| \mathscr{F}_t \right]$$

is the contract value given the surrender time λ, and $\mathbb{T}_t$ is the set of stopping times taking values in $[t, +\infty)$.

Note that the contract value V_t can also be expressed as $V_t = \max\{V_t^c,\ b_t^S\}$, with V_t^c denoting the continuation value, given by $V_t^c = \sup_{\lambda \in \mathbb{T}_t^c} V_t(\lambda)$, where $\mathbb{T}_t^c$ is now the set of stopping times taking values in $(t, +\infty)$.

In particular, in [5] it is proven, under the assumption of lognormality for the price process S and deterministic mortality intensity, that surrender is never optimal (i.e., the continuation value is higher than the surrender benefit) if the account value is above the fee threshold. The intuition behind this result is clear: when $A_t \geq \beta$ the guarantees at death and maturity are offered for free, hence there is no incentive for the policyholder to surrender the contract. We are able to generalize this result, just requiring that the discounted price process is a martingale (under the risk-neutral measure) and financial related variables are independent of mortality (see [1]).

Finally, we note that the contract value under the mixed approach is not less than the corresponding value under the static approach (American versus European-style contract), and the difference between them is the surrender option value.

4 Numerical Implementation

The optimal stopping problem giving the contract value under the mixed approach needs to be tackled numerically. In [4] it is claimed that the Least Squares Monte Carlo techniques are unsuitable to solve it, due to the shape of the surrender region, that is like a corridor (even very strict), thus implying too significant numerical errors. Since we believe that the intrinsic flexibility of Monte Carlo methods is a very important feature, we have tested their application to the solution of the problem. Doing this, we have actually verified that a straightforward application of them is a bit problematic, specially for relatively low levels of the fee, i.e., when very likely the surrender incentive has been completely eliminated leading to a valueless surrender option. In these cases, in fact, the contract value under the static approach turns out to be higher than that under the mixed approach, contradicting the

theoretical relation and confirming the claim by Bernard et al. [4]. Given this appears to happen only for low levels of the fee, a possible explanation is that the regression tends to underestimate the continuation value, thus inducing surrender even when this is not the optimal decision. We have observed this behaviour by comparing the residuals plots printed at each regression step for the cases of constant and state-dependent fees. While in the constant case the residuals appeared to be balanced between positive and negative values for all regression steps, in the state-dependent case they tended to shift towards positive values in the last few steps. Since the LSMC algorithm proceeds backward, this means that at the very first surrender decision dates the real continuation values were generally much greater than the predicted ones, leading to earlier and sub-optimal terminations of the contract. Therefore, in the attempt to improve the regression, we have tried several methods, such as changing type and number of basis functions, or using different regression techniques, which however have not brought substantial enhancements. In contrast, the theoretical result mentioned before, according to which the regression step can be skipped when $A_t \geq \beta$, has allowed us to significantly reduce the numerical error.

In the following tables we report some results for the contract value under the mixed approach obtained with Monte Carlo simulation and alternative regression techniques (Least Squares, LS; Generalized Linear Models, GLM; Ridge regression; the Lasso) without skipping the regression step, as well as those obtained with LS by skipping this step (Adj LS), and the values under the static approach (Static). Although, as previously mentioned, we advocate the use of Monte Carlo methods for their flexibility, and hence for the absence of model constraints, in these examples we show results under very simple assumptions, i.e., a constant interest rate, a deterministic mortality intensity and a Geometric Brownian Motion (GBM) for the assets evolution. This is because in this framework we have a benchmark, that is the contract value obtained by using the PDE approach as described in [5]. We refer instead to [1] for a wide range of numerical results under alternative, and rather complex, model assumptions.

In Table 1 we have tried to reproduce some results by MacKay et al. [5], that use the PDE approach to compute the contract values and determine also the fair fee rate, that is a fee rate making the contract value equal to the initial premium. More in detail, we fix a fee rate exactly equal to the fair level reported in [5] for different policyholder ages at inception. The contract parameters are as follows: $P = 100$, $T = 10$, $\delta = 0$, $p_t = 1 - \mathrm{e}^{-0.008(10-t)}$, $\beta = 150$. Moreover, the risk-free rate is $r = 0.03$, the assets volatility is $\sigma = 0.165$, and the mortality intensity follows a Makeham law: $\mu_y = 10^{-4}(1 + 3.5 \cdot 1.075^y)$. The number of simulations is 20,000.

Table 1 Contract values estimated with PDE and Monte Carlo methods

Age	Fee	PDE	LS	GLM	Ridge	Lasso	Adj LS	Static
50	0.0167	100.01	99.22	99.08	98.69	98.98	99.44	99.50
60	0.0179	100.00	99.26	99.17	99.09	98.99	100.04	99.83
70	0.0204	100.01	99.35	99.02	98.88	98.85	99.49	99.45

Table 2 Contract values estimated with PDE and Monte Carlo methods

Fee	PDE	LS	GLM	Ridge	Lasso	Adj LS	Static
2%	113.89	112.07	111.73	112.28	112.17	113.14	113.04
6%	101.82	101.05	101.06	101.11	100.97	101.51	101.39
7%	100.52	99.95	100.02	99.87	99.95	100.01	98.70
9%	99.08	98.27	98.27	98.18	98.19	98.40	95.35

In Table 2 we consider contracts with different fee rates, both under the fair level (contract value higher than P), for which the improvement obtained with the LSMC adjustment is more remarkable, and over. We fix now $T = 15$, $\delta = 2\%$, $\beta = 134.98588$, $\sigma = 0.2$, take a policyholder aged 50 years, a constant surrender penalty of 2%, a Weibull mortality intensity $\mu_y = 10.002 \cdot 88.14778^{-10.002} y^{9.002}$, and the same values as before for P, r and the number of simulations.

References

1. Bacinello, A.R., Zoccolan, I.: Variable Annuities with State-Dependent Fees: Valuation, Numerical Implementation, Comparative Static Analysis and Model Risk. DEAMS Research Paper Series, n.1/2017 (2017). Available online at http://hdl.handle.net/10077/15141
2. Bacinello, A.R., Millossovich, P., Olivieri, A., Pitacco, E.: Variable annuities: a unifying valuation approach. Insur. Math. Econ. **49**, 285–297 (2011)
3. Bae, T., Ko, B.: Pricing maturity guarantee under a refracted Brownian motion. Lobachevskii J. Math. **34**, 234–247 (2013)
4. Bernard, C., Hardy, M.R., MacKay, A.: State-dependent fees for variable annuity guarantees. ASTIN Bull. **44**, 559–585 (2014)
5. MacKay, A., Augustyniak, M., Bernard, C., Hardy, M.R.: Risk management of policyholder behavior in equity-linked life insurance. J. Risk Insur. **84**, 661–690 (2017)

Dynamic Policyholder Behavior and Surrender Option Evaluation for Life Insurance

Fabio Baione, Davide Biancalana, Paolo De Angelis, and Ivan Granito

Abstract Since 2016, Solvency II introduced in the insurance industry sector new capital requirements rules for the evaluation of life insurance liabilities based on VaR risk measure. This paper aims to analyze the effect of the dynamic policyholder behavior on the evaluation of lapse risk of a portfolio of participating life insurance policies.

Keywords Solvency 2 · Policyholder behavior · Life insurance · Surrender option

1 Introduction

The policyholders' behavior is a determining factor for the valuation of technical provisions for a life insurance portfolio. This paper considers the problem of estimating the lapse rates in a portfolio composed by different profiles of policyholders when conditions in the financial market change over time. The aim is to analyze the behavioral economy of policyholder related to financial risk drivers. The literature

F. Baione
Department of Mathematical Sciences, Mathematical Finance and Econometrics, Catholic University, Milan, Italy
e-mail: fabio.baione@unicatt.it

D. Biancalana · I. Granito
Department of Statistical Sciences, Sapienza University of Rome, Rome, Italy
e-mail: davide.biancalana@uniroma1.it; ivan.granito@uniroma1.it

P. De Angelis (✉)
Department of Methods and Models for Economics Territory and Finance, Sapienza University of Rome, Rome, Italy
e-mail: paolo.deangelis@uniroma1.it

M. Corazza et al. (eds.), *Mathematical and Statistical Methods for Actuarial Sciences and Finance*, https://doi.org/10.1007/978-3-319-89824-7_14

proposes different approaches and there are two different best practices in the market:

– One-step model: A GLM Logit model is used to estimate the lapse rates based on policyholder/contract features or individual risk factors (e.g. age, gender, guarantee rate, etc.) and financial variables (e.g. return on investments, market yield). In that case there could be a good data fitting, but outcome could not have a full economic meaning.
– Two-step model: The first step coincides with the one-step model previously depicted and it is performed to estimate lapse rates with respect to the policyholder/contract features. Afterwards, lapse rates derived in step one are corrected by a multiplicative factor (hereafter term "correction factor" or "CF") that considers the difference between the rate of return of a benchmark financial asset and the policy crediting rate; in the following this difference is named "market spread".

The advantage of the second approach is that it offers greater economic consistency of the outcomes given by a set of constraints on the functional relationship between lapse rates and the market spreads. As shown in the following, our contribution is to provide a function for CF more reliable in terms of economic meaning.

2 A Model for the Lapse Rate Estimation According to Policyholder Behavior

In order to model the lapse rates according to a two-step model we consider as basic model the one step model described above, based on the policyholder/contract features. The second step consists on defining the CF as a double sigmoid function dependent on the market spread. Considering the policyholder as an arbitrager, he/she lapses the insurance contract to purchase a new more profitable contract. The model proposed is based on a double sigmoid function [1] which, compared to the other models in the literature, has a greater adaptability in cases where the policyholder's behavior is similar to that described in the following:

- if the spread is around zero, lapse rates do not depend on the market spread and then no correction is applied to the basic model;
- if the spread becomes negative, as a consequence it is fair that lapse rates decrease. When the positive difference between the policy crediting rate and the benchmark rate of return is too large it is necessary to stop an ever-decreasing withdrawals trend, by means of a floor (lower asymptote);
- if the spread becomes positive, as a consequence it is fair that lapse rates increase. When the negative difference between the policy crediting rate and the benchmark rate of return is too large it is necessary to stop an ever-increasing withdrawals trend, by means of a cap (upper asymptote);

Models proposed in the literature [2–6] describe the relationship between market spread and lapse rate by means of increasing monotonous functions limited by two asymptotes (higher and lower) with just one change of concavity.

Instead, the double sigmoid function is monotonous growing between two asymptotes, but may have one or two concavity changes to better fit the data. From a mathematical point of view the model is as follow.

2.1 Step 1

The expected lapse rate (r) of a single generic policyholder characterized by a set of individual risk factors $\boldsymbol{\theta}$, is obtained by a logit GLM calibrated on the portfolio data:

$$r = f(\boldsymbol{\theta}) \tag{1}$$

where $f(\cdot)$ indicates the inverse of the link function of the GLM.

2.2 Step 2

The lapse rate calculated in the first step is corrected by a multiplicative factor modelled with the double sigmoid function. Then it holds:

$$r_t = f(\theta) \cdot q(\Delta_t) \tag{2}$$

where $q(\Delta_t)$ is the CF dependent on the market spread at time t:

$$q(\Delta_t) = \alpha + \beta \cdot \left[\tanh\left(\frac{\Delta_t - c_1}{w_1} \right) \cdot \delta + \tanh\left(\frac{\Delta_t - c_2}{w_2} \right) \cdot \gamma \right] \tag{3}$$

$$\delta, \gamma, w_1, w_2 > 0 \tag{4}$$

where:
$\alpha = \frac{cap+floor}{2}$; $\beta = \frac{cap-floor}{2\cdot(\delta+\gamma)}$; $cap = \frac{ymax}{f(\theta)}$; $floor = \frac{ymin}{f(\theta)}$

- *ymax* represents the upper asymptote as $\Delta_t \to +\infty$ and is based on expert judgment;
- *ymin* represents the lower asymptote as $\Delta_t \to -\infty$ and is based on expert judgment;

In general, the sigmoid function and the double-sigmoid function are expressed by means of the hyperbolic tangent function or exponential function. Lipovetsky [7]

derive the sigmoid function using exponential function and applying a particular parameterization to better fit the data. The double-sigmoid function used in this paper is based on an hyperbolic tangent function with a different parameterization just to consider our specific needs. In the construction of the function we have followed a procedure similar to [7] applied to our context.

We will provide a mathematical derivation of the function $q(\Delta_t)$ and an application to a portfolio of participating life insurance policies, representative of the Italian insurance market.

3 Some Numerical Results

Figure 1 shows a comparison between the one-step and the two-step model, based on a set of sampled dataset composed by monthly lapse rates.

As it can be noted, the two models provide very different results. The one and two step model both show a lapse rate increasing with the market spread. Under the logit model the absence of a functional constraint that imposes a horizontal asymptote, provides a lapse rate that indefinitely increases. It means that increasing the market spread the estimates become gradually less accurate from an economical point of view. However the functional constraint provides estimates in this case less accurate than the GLM ones in the statistical context of goodness of fitting.

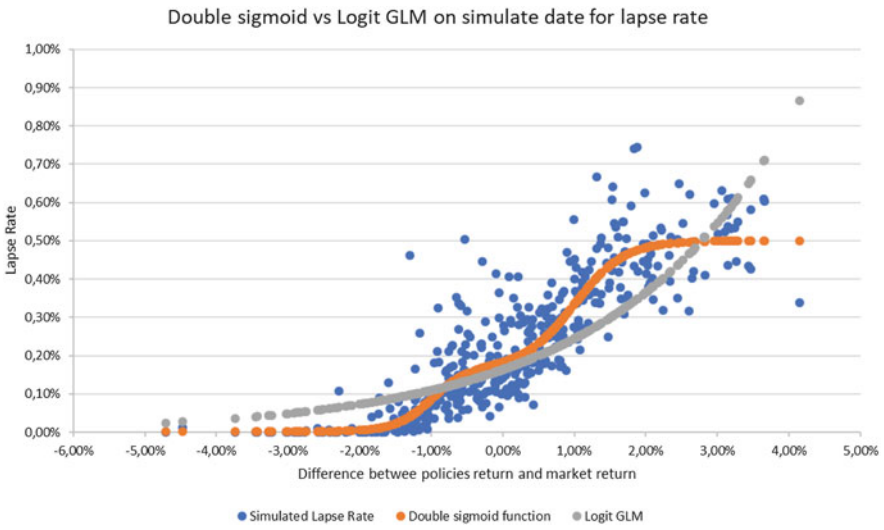


Fig. 1 Lapse rate fitted: comparison between one and two step model

References

1. Roper, D.: Using sigmoid and double-sigmoid function for earth-states transitions. https://www.researchgate.net/publication/242689104Using_Sigmoid_and_Double-SigUsing_Sigmoid_and_Double-Sigmoid_Functions_for_Earth-States_Transitions_1 (2015)
2. MacKay, A., Augustyniak, M., Bernard, C., Hardy, M.R.: Risk management of policyholder behavior in equity-linked life insurance. J. Risk Insur. **84**(2), 661 (2017)
3. Knoller, C., Kraut, G., Schoenmaekers, P.: On the propensity to surrender a variable annuity contract: an empirical analysis of dynamic policyholder behavior. J. Risk Insur. **83**(4), 979 (2016)
4. Kim, C.: Modeling surrender and lapse rates with economic variables. N. Am. Actuar. J. **9**(4), 56–70 (2005)
5. Buchardt, K., Møller, T.: Life insurance cash flows with policyholder behavior. Risks. **3**(3), 2 (2015)
6. Nolte, S., Schneider, J.C.: Don't lapse into temptation: a behavioral explanation for policy surrender. J. Bank. Financ. **79**(1), 12–27 (2017)
7. Lipovetsky, S.: Double logistic curve in regression modeling. J. Appl. Stat. **37**, 1785–1793 (2010). https://doi.org/10.1080/02664760903093633

Classification Ratemaking via Quantile Regression and a Comparison with Generalized Linear Models

Fabio Baione, Davide Biancalana, Paolo De Angelis, and Ivan Granito

Abstract In non-life insurance, it is important to develop a loaded premium for individual risks, as the sum of a pure premium (expected value of loss) and a safety loading or risk margin. In actuarial practice, this process is known as classification ratemaking and is performed usually via Generalized Linear Model. The latter permits an estimate of individual pure premium and safety loading both; however, the goodness of the estimates are strongly related to the compliance of the model assumption with the empirical distribution. In order to investigate the individual pure premium, we introduce an alternative pricing model based on Quantile Regression, to perform a working classification ratemaking with weaker assumptions and, then, more performing for risk margin evaluation.

Keywords Quantile regression · Generalized linear model · Premium principles · Risk margin · Ratemaking

1 Introduction

In the valuation of insurance risks it is standard practice to consider a margin for adverse deviations compared to the expected value. This margin is commonly defined as safety loading or risk margin. Risk margin is largely debated by both practitioners and academic actuaries as there is not a single definition as well as

F. Baione
Department of Mathematical Sciences, Mathematical Finance and Econometrics, Catholic University, Milan, Italy
e-mail: fabio.baione@unicatt.it

D. Biancalana · I. Granito
Department of Statistical Sciences, Sapienza University of Rome, Rome, Italy
e-mail: davide.biancalana@uniroma1.it; ivan.granito@uniroma1.it

P. De Angelis (✉)
Department of Methods and Models for Economics Territory and Finance, Sapienza University of Rome, Rome, Italy
e-mail: paolo.deangelis@uniroma1.it

M. Corazza et al. (eds.), *Mathematical and Statistical Methods for Actuarial Sciences and Finance*, https://doi.org/10.1007/978-3-319-89824-7_15

a unique model for the relative estimation. Under the International Accounting Standard Board (IASB) proposals, the risk margins should be "an explicit and unbiased estimate of the risk margin that market participants require for bearing risk". This view is consistent with the one proposed in the financial reporting of Solvency II [1]. Whatever be the field of application, the assessment of a risk margin could be performed by means of a specific risk measure applied to a loss (or profit) random variable (r.v.). The first use of a risk measure in actuarial science was the development of premium calculation principles [2, 3] and were essentially based on variability measures of the loss such as variance or standard deviation.

In order to investigate a premium principle based on a quantile risk measure, we compare two multivariate regression models: the Generalized Linear Models [4] (GLM) and the Quantile Regression [5] (QR). Our aim is to measure an individual risk margin for each rating class defined by a set of rating factors. We implement a quantile premium principle to this aim to be compliant with an approach widely applied for solvency purposes. Firstly, we consider a traditional two-part model to decompose the cost of claim between the event of the claim and its amount, if the claim occurs. Secondly, we determine the risk margin of the claim's severity per insured by means of a binomial GLM for the frequency and a Gamma GLM for the severity per claimant. As alternative to the latter model we also propose the application of QR to estimate the severity of claims per insured in order to better assess the risk margin value.

2 A Quantile Premium Principle Based on a Two-Part Model

2.1 The Two Part Model for the Individual Risk Model

We consider a non-life insurance portfolio of r insurance contracts. In an individual risk model the loss of the i-th individual risk is modeled by means of the generic r.v. $Y^{(i)} \in \mathrm{Y}$, $i = 1, 2, \ldots, r$. The r.v.s $Y^{(i)}$ are assumed to be independent.

To predict the cost of a claim per insured, it is usual to decompose two-part data: one part for the frequency, indicating whether a claim has occurred or, more generally, the number of claims; another part for the severity, indicating the amount of a claim.

This way of decompose [6] the loss is traditionally term as frequency-severity model in actuarial literature, while in statistics is generally called two-part model [7]. Hereafter, we focus in the following decomposition of the r.v. claim amount per insured:

$$Y^{(i)} = \mathbb{I}_{N^{(i)}} \cdot \tilde{Y}^{(i)} \tag{1}$$

where

- $\mathbb{I}_{N^{(i)}} \sim$ Bernoulli (p_i) is a indicator random variable indicating whether the i-th insured has had at least one claim
- $\tilde{Y}^{(i)} = \left(\sum_{j=1}^{N^{(i)}} Z_j^{(i)} | N^{(i)} > 0\right)$ is a strictly positive random variable representing the total loss amount per claimant, independent from $\mathbb{I}_{N^{(i)}}$. $Z_j^{(i)}$ is a strictly positive random variable representing the loss amount for j-th claim of the i-th insured.

  The cumulative density function (cdf) of $Y^{(i)}$ is expressed as:

$$F_{Y^{(i)}}(y) = p_i + (1 - p_i) \cdot \Pr\left(\tilde{Y}^{(i)} < y\right) \tag{2}$$

2.2 A Quantile Premium Principle Based on a Two-Part Model

In order to assess the pure premium based on the individual risk model previously stated, it is usual to develop a separate analysis on the factors or explanatory variables that influence frequency and severity, due to the their independence and, then, compute their combination to form the pure premium. In non-life actuarial practice, this analysis is widely implemented through multivariate regression techniques such as GLMs, assuming log links for response variables. The advantage of GLM is in the ability to estimate the conditional expected value of the response variable [8], while shows some drawbacks if used to analyze other moments of the probability distribution. However, to extend the analysis to other distribution's moments of $Y^{(i)}$, we suggest as an alternative to GLM to adopt a fat-tailed regression model to assess the conditional quantile of the severity component $\widetilde{Y}^{(i)}$ as QR. Therefore, the pure premium approach we propose is defined by applying a functional π that assigns a non-negative real number to random variables $Y^{(i)}$ representing uncertain pay-offs, in our context the insurance losses, as follows:

$$P_Q = \pi\left(Y^{(i)}\right) = \mathrm{E}\left[\mathbb{I}_{N^{(i)}}\right] \cdot Q_{\theta_i}\left[Y^{(i)}\right] = \mathrm{E}\left[\mathbb{I}_{N^{(i)}}\right] \cdot Q_{\overset{\sim}{\theta}_i}\left[\tilde{Y}^{(i)}\right] \tag{3}$$

where θ_i represents the probability level or security level adopted to calculate the pure premium for the i-th risk assured.

Following Kudryavtsev [9], by setting $F_{Y^{(i)}}\left(Q_{\overset{\sim}{\theta}_i}\left[\tilde{Y}^{(i)}\right]\right) = \theta_i$, by Eq. (2) we have:

$$\theta_i = p_i + (1 - p_i) \cdot \overset{\sim}{\theta}_i \iff \overset{\sim}{\theta}_i = \frac{\theta_i - p_i}{1 - p_i}. \tag{4}$$

The premium loading expressed as a percentage of the expected value, hereafter named as risk margin function, is defined as follows:

$$\rho\left(Y^{(i)}\right) = \frac{\pi\left(Y^{(i)}\right)}{E\left(Y^{(i)}\right)}. \tag{5}$$

With reference to the estimate of the conditional quantile of the r.v. severity per claimant, $Q_{\underset{\theta_i}{\sim}}\left[\overset{\sim(i)}{Y}\right]$, we compare a Gamma GLM with unitary prior weights[1] (i.e. $w_i \equiv 1, \quad i = 1, 2, \ldots, r$) and a QR approach.

The first approach leads to several drawbacks as the risk margin function (5) is constant for each risk profile and the premium loading depends exclusively by the expected value of $Y^{(i)}$: i.e. $P_Q^{GLM} = (1 - p_i) \cdot \rho_{\underset{\theta}{\sim}} \cdot E\left(\overset{\sim(i)}{Y}\right)$.

Instead, the latter approach based on QR leads the risk margin function to be dependent to the risk profile characteristics contained in the row-vector of independent covariates. So, compared to the solution provided with a Gamma GLM the risk function in the QR framework proposed allows to better understand the different variability of the risk profiles considered.

3 Simulation Study

In order to perform a comparison between the quantile approaches previously described we implement the two-part model on three different simulated data set. Data sets are characterized by the same total exposure $r = 100{,}000$. We consider two rating factors (gender and age-class) each one with two levels ("Male" and "Female", "High" and "Low" respectively). Therefore the total exposure is split in four risk classes, $h = 1, 2, 3, 4$. We assume that each risk class has a different exposure but the same number of claimants, say 2500. Hence, we simulate 2500 claims for each h, with a total of 10,000 for each data set k. The main difference between data sets is the assumption used for the cdf of each h-th risk class of the r.v. total claim amount per claimants, in order to perform a sensitivity analysis among risk-classes:

- the first data set ($k = 1$) is sampled by four Gamma with the same shape parameter and different scale parameter; hence the "Coefficient of Variation (CV)" is constant for each class;
- the second data set ($k = 2$) is sampled by four Gamma with the same mean and different CV;

[1] Each insured has unitary exposure and belongs to a risk class defined by means of a set of rating factors.

Table 1 Estimates of pure premium and risk margin function with $\widetilde{\theta}_h = 0.75$

Data Set	h	Exposure	$E(Y^{(h)})$	P_Q		$\rho(Y^{(h)})$		θ_h (%)
			GLM	GLM	QR	GLM (%)	QR (%)	
k = 1	1 (MH)	7500	0.97	1.12	1.12	115.60	115.67	91.67
	2 (FH)	22,500	0.77	0.89	0.90	115.60	115.98	97.22
	3 (ML)	12,500	3.75	4.35	4.35	115.60	115.95	95.00
	4 (FL)	57,500	2.53	2.92	2.94	115.60	116.26	98.91
k = 2	1 (MH)	7500	7.51	8.79	8.03	116.96	106.88	91.67
	2 (FH)	22,500	2.21	2.58	2.47	116.96	112.19	97.22
	3 (ML)	12,500	3.99	4.71	4.71	116.96	118.00	95.00
	4 (FL)	57,500	0.99	1.16	1.23	116.96	123.85	98.91
k = 3	1 (MH)	7500	0.97	1.13	1.03	116.37	106.56	91.67
	2 (FH)	22,500	0.76	0.89	0.86	116.37	111.87	97.22
	3 (ML)	12,500	3.77	4.36	4.36	116.37	115.59	95.00
	4 (FL)	57,500	2.51	2.92	3.05	116.37	121.34	98.91

- the latter data set ($k = 3$) is sampled by four Lognormal, with different parameter for each risk class.

As shown in Table 1 if the sampling distribution are Gamma with same CVs GLM and QR results are quite similar; if these strong assumption are relaxed GLM gamma is not compliant to estimate an individual risk margin function because the value is the same for each risk profile.

References

1. EIOPA: Directive 2009/138/EC of the European parliament of the council 25 November 2009 on the taking-up and pursuit of the business of Insurance and Reinsurance (Solvency-II). https://eiopa.europa.eu
2. Gerber, H.: An introduction to mathematical risk theory. Huebner Foundation, Philadelphia (1979)
3. Olivieri, A., Pitacco, E.: Introduction to Insurance Mathematics: Technical and Financial Features of Risk Transfers. Springer, Berlin (2011)
4. Nelder, J.A., McCullagh, P.: Generalized Linear Models, 2nd edn. Chapman Hall, London (1989)
5. Koenker, R.W., Bassett, G.W.: Regression quantiles. Econometrica. **46**, 33–50 (1978)
6. Bühlmann, H.: Risk Theory. Springer, Berlin (1970)
7. Duan, N., Manning, W.G., Morris, C.N., Newhouse, J.P.: A comparison of alternative models for the demand for medical care. J Bus. Econ Statist. **1**, 115 (1983)
8. Goldburg, M., Khare, A., Tevet, D.: Generalized Linear Models for Insurance Rating Casualty Actuarial Society Monograph n. 5. Casualty Actuarial Society, Arlington (2016)
9. Kudryavtsev, A.A.: Using quantile regression for ratemaking. Insur Math Econ. **45**, 296–304 (2009)

An Empirical Analysis of the Lead Lag Relationship Between CDS and Stock Market

Laura Ballester, Rebeca Fernández, and Ana González-Urteaga

Abstract This paper complements the recent literature providing a thorough research of the lead lag relationship between stock and sovereign CDS markets using a rolling VAR framework. We find that the transmission channel between the credit and stock market exist. This phenomenon is time varying, it seems to be related with the economic cycle and in general, it's more intense in US than in Europe.

1 Introduction

Following the collapse in September 2008 of Lehman Brothers, financial markets experienced tremendous distortions and the importance of the study of risk management, especially sovereign credit risk, has increased. Since this moment and in the context of the European debt crisis, credit spreads rose to unprecedented levels and the literature has focused on the study of the relationships between credit risk and stock markets. The significance of this interconnection has increased in recent years as credit derivatives have been trading in all financial markets, being the credit default swap (CDS hereafter) the most commonly used instrument for transfer credit risk. The present paper follows this line of research. The literature shows a tension when it comes to the question of whether the CDS market leads the stock market or whether the stock market leads the CDS market. Initial studies

L. Ballester (✉)
University of Valencia, Valencia, Spain
e-mail: laura.ballester@uv.es

R. Fernández
MUSAAT, Madrid, Spain

A. González-Urteaga
INARBE (Institute for Advanced Research in Business and Economics), Pamplona, Spain

Public University of Navarre, INARBE (Institute for Advanced Research in Business and Economics), Pamplona, Spain
e-mail: ana.gonzalezu@unavarra.es

M. Corazza et al. (eds.), *Mathematical and Statistical Methods for Actuarial Sciences and Finance*, https://doi.org/10.1007/978-3-319-89824-7_16

have focused on corporate US data with mixed results. [1] provide evidence of information flow from CDSs to equity, which is more significant for entities with a greater number of bank relationships. However, more recently, [4] document the opposite effect. On the other hand, according to [3] the link depends on the credit quality of the underlying reference entity. Finally, [5] shows that the direction of the relationship differ across industry sectors. As a result of the 2008 crisis there is a growing strand of the literature. [6] focus on major US financial institutions under stress, concluding that both markets become more integrated in times of stress and that fast equity price changes lead furious CDS spread. In the wake of the Eurozone sovereign debt crisis, [2] shows that during tranquil times stock market leads transmission, but in financial crises there has been an inverse relationship.

2 Data and Methodology

This paper provides an exhaustive investigation of the interaction between stock indices and credit markets using data for US and 14 developed European countries during 2004–2016. As a novel contribution, we provide evidence of the time variation of the lead lag relationship by a rolling framework, which enables us to analyse the evolution of the effect over time covering both crisis and non-crisis periods. Concretely, we estimate a rolling VAR(p) model with 250-observations time window. For each estimation we save the obtained p-value for both the conventional Granger-causality from CDS to stocks and from stocks to CDS. We improve the conventional estimations, analysing whether the relationships vary through time.

3 Results

To make easier the understanding of results, European countries are allocated into different groups distinguishing between Eurozone and Non Eurozone, and within Eurozone, between Euro-core and Euro-peripheral. Table 1 presents the causality results for all the European countries and US for the full sample (FS), for pre-crisis period (PC, Jan 2004–Jun 2007), the Global Financial Crisis (GFC, Jul 2007–Dec 2009) and the sovereign debt crisis (SDC, Jan 2010–Apr 2016). It presents the percentage of windows where a variable causes the other at 5% of significance. For European countries, we present the mean of the percentages in each case to obtain the average results by monetary zone and for geographical zone inside Eurozone.[1]

[1]The idea is to contrast if there are significant differences segmenting Eurozone between countries more affected by the SDC (Euro-peripheral countries) and the others (Euro-core countries).

Table 1 Granger causality test: stock and CDS returns results

	FS		PC		GFC		SDC	
Zone/Country	− >	< −	− >	< −	− >	< −	− >	< −
Austria	39.06%	4.48%	0.00%	0.00%	68.65%	2.75%	42.86%	6.96%
Belgium	27.07%	11.26%	0.91%	0.00%	45.11%	5.81%	30.34%	17.95%
France	24.81%	13.45%	0.00%	0.00%	11.31%	0.31%	33.76%	20.63%
Germany	26.79%	6.69%	0.00%	1.38%	39.91%	2.45%	32.23%	10.50%
Netherlands	26.60%	2.61%	–	–	50.00%	2.93%	18.32%	2.50%
Euro-core	28.87%	7.70%	0.23%	0.34%	43.00%	2.85%	31.50%	11.71%
Greece	22.33%	2.04%	0.00%	0.00%	61.62%	2.75%	6.55%	3.28%
Ireland	23.88%	14.06%	–	–	62.82%	0.00%	10.07%	19.05%
Italy	30.99%	16.67%	0.00%	0.00%	24.01%	35.63%	46.03%	15.69%
Portugal	39.07%	19.01%	2.95%	0.00%	25.54%	35.17%	58.67%	20.02%
Spain	27.99%	12.39%	–	–	51.81%	2.93%	19.54%	15.75%
Euro-peripheral	28.85%	12.83%	0.98%	0.00%	45.16%	15.29%	28.17%	14.76%
Eurozone	28.86%	10.27%	0.55%	0.20%	44.08%	9.07%	29.84%	13.23%
Denmark	10.74%	1.12%	0.00%	0.34%	19.42%	0.15%	9.22%	1.65%
Norway	13.02%	12.04%	0.00%	0.00%	7.49%	0.00%	20.45%	21.67%
Sweden	23.64%	10.44%	0.00%	0.00%	30.12%	0.46%	30.53%	18.62%
UK	27.46%	6.08%	–	–	97.03%	7.92%	14.59%	5.74%
Non Eurozone	18.71%	7.42%	0.00%	0.11%	38.52%	2.13%	18.70%	11.92%
Europe	25.96%	9.45%	0.39%	0.17%	42.49%	7.09%	26.65%	12.86%
US	44.80%	9.46%	–	–	53.85%	2.45%	43.22%	10.68%

− > and < − indicates that stocks leadership CDS and CDS leadership stocks, respectively

For the FS, we find a bidirectional relationship between stock and CDS markets in Europe with a clear predominance of the stocks leading the CDS. Stock market leads the process in Europe (25.96% vs 9.45%). Regarding Eurozone, the stock market always leads the CDS market with a greater impact of Eurozone (28.86%) with respect to Non Eurozone (18.71%). Inside Eurozone, both geographical zones presents similar contribution. In US, we find a feedback transmission, largest from stocks returns to CDS returns US (44.80% vs 9.46%). In addition, the information transfer is more intense in US (44.80%) than in Europe (25.96%).

It should be pointed out that the contribution percentage varies significantly over time. Before GFC, the causality relationship practically does not exist and therefore we focus in the 2007–2016 results.[2] This result is expected due to the typically absence of assets correlations comovements in stable periods. During the GFC, we find significant relationships with remarkable differences between US and Europe. A clear feedback relationship that varies over in European countries exists (around 40%), although stock returns leads CDS, with a maximum stock contribution in 2009 (71.51%).[3] We observe significate differences between crisis

[2]We don't present results for 2004 and 2016 since we obtain few p-values results.

[3]Results by year are available upon request.

subperiods as well as between European zones. In this period, the stock leadership is strong in the majority of the European countries due to the US financial crisis contagion to Europe, especially in Eurozone with respect to the Non Eurozone countries (44.08% vs 38.52%). Within Eurozone, we observe a similar transmission and magnitude direction in the Euro-peripheral countries (45.16%) and Euro-core countries (43.00%).

During the SDC, the tendency changes in Europe. Stock returns continue leading in Europe (26.65%), but not as evident as before (42.49%). A clear bidirectional causality relationship exist. During this period the causality relationships were weaken in Europe, perhaps due to the adjustments made by European countries, although the stock market continue leading the price discovery effect, except in the Euro-peripheral countries where the CDS market leads the stock market in 2010 and 2011. US results shows the same bidirectional causality relationship that we observe in Europe with a highest transmission from stocks to CDS during the GFC and the SDC (53.85% and 43.22%).

4 Conclusions

We could conclude that the transmission process between the credit and stock market exists being the stock market who anticipates CDS returns. These results are especially interesting to understand the dynamics of risk transmission and the knowledge of price discovery process between markets should take into consideration for portfolio management in order to match assets in optimal portfolios.

Acknowledgements Authors express their gratitude to Fundación Ramón Areces. A. González-Urteaga acknowledges support from ECO2015-67035-P and ECO2016-77631-R. We thank A. Pardo, O.Carchano, J.A. Álvarez and participants at the 15th INFINITI Conference and Mathematical and Statistical Methods for Actuarial Sciences and Finance (MAF 2018) for comments.

References

1. Acharya, V.V., Johnson, T.C.: Insider trading in credit derivatives. J. Financ. Econ. **84**, 110–141 (2007)
2. Asandului, M., Lupu, D., Claudiu, G., Musetescu, R.: Dynamic relations between CDS and stock markets in Eastern European countries. Econ. Comput. Econ. Cybern. Stud. Res. **49**, 151–170 (2015)
3. Fung, H.G., Sierra, G.E., Yau, J., et al.: Are the US stock market and credit default swap market related? Evidence from the CDX indices. J. Altern. Invest. **11**, 43–61 (2008)
4. Hilscher, J., Pollet, J.M., Wilson, M.: Are credit default swaps a sideshow? Evidence that information flows from equity to CDS markets. J. Financ. Quant. Anal. **50**, 543–567 (2015)
5. Narayan, P.K., Sharma, S.S., Thuraisamy, K.S.: An analysis of price discovery from panel data models of CDS and equity returns. J. Bank. Financ. **41**, 167–177 (2014)
6. Trutwein, P., Schiereck, D.: The fast and the furious-stock returns and CDS of financial institutions under stress. J. Int. Financ. Mark. Inst. Money **21**, 157–175 (2011)

Integration of Non-financial Criteria in Equity Investment

Diana Barro

Abstract In recent years the awareness of social, environmental and governance issues associated with investments have drawn relevant interest in the investment industry. Investors are more careful in considering investments that comply with their ethical and moral values, as well as with social impact. Hence, the ethical and social responsibility of investments (SRI) is becoming more popular in the academic literature due to the fact that socially responsible investment provide profitability and social commitment together. In this contribution we discuss the main issues that arise when integrating socially responsible criteria into a financial decision problem.

Keywords Socially responsible investments · Portfolio optimization · ESG

1 Introduction

In [5] the authors interestingly reported a relevant shift in the terminology used to identify the non-financial goals of the investor moving from a widely used 'ethical investment', mainly based on religious beliefs and/or moral values, to a wider concept of 'socially responsible investment' (SRI) in the last years. This change reflects at least two different aspects. The first is a broadening of the values carried by investors and their increased commitment in impact investing and the second is a more positive versus negative perspective in the investment process with preference given to inclusion of virtuous companies versus exclusion of specific industries/companies. This new perspective poses new challenges to the traditional way in which ethical dimension was paired with equity investment, see, for example, [13].

There are many different contributions in the literature that analyse the trade-off between the financial risk-return dimension and ethical or social dimension of the investment, see, for example, [11, 12]. Results in Dupré et al. [7] suggested

D. Barro (✉)
Department of Economics, Ca' Foscari University, Venezia, Italy
e-mail: d.barro@unive.it

M. Corazza et al. (eds.), *Mathematical and Statistical Methods for Actuarial Sciences and Finance*, https://doi.org/10.1007/978-3-319-89824-7_17

that strong social requirements had an impact on the financial performances. Many studies compared the performances of conventional and socially responsible mutual funds and found that there are no statistically significant differences (see, for example, [2, 3, 15]). Different contributions have tried to find a way to incorporate these preferences into the portfolio selection model. Integrating non-financial criteria into financial decision models is non-trivial and it is usually tackled through a multi-step selection procedure. Our analysis aims at showing that this approach may result in suboptimal portfolio compositions when the trade-off between financial and non-financial criteria is not properly discussed.

2 Including Non-financial Criteria

The general approach to the inclusion of SRI criteria, that encompasses many different contributions in the literature, requires three main steps, each featuring specific challenges. First, a preliminary step for the collection of information on social responsibility behaviour of a company is necessary; second, the selection of a subset of suitable companies through positive or negative screening; third, optimal portfolio composition based on specific financial criteria.

The first step is crucial and recently an increasing number of data providers and consulting firms made available ratings and evaluations with respect to different dimensions of the SRI criteria. Unfortunately, the lack of a standard scoring determines also a lack of robustness with reference to the studies on the performance of socially responsible investment when compared with traditional investment. Furthermore, we must remember that data are usually made available by companies on a voluntary basis.

The second step consists in a reduction of the universe of admissible assets on the basis of non-financial criteria, this may cause serious limitations to the diversification potential of the investment and a worsening of the risk-adjusted performance. Furthermore, investors are required to set a priori a desired level of responsibility in order to discriminate acceptable versus non acceptable companies. When the evaluation is carried out across different dimensions this is not trivial and the definition of a proper measure becomes particularly relevant.

Finally, the third step is the portfolio selection based on financial criteria. Among the different contributions in the literature we refer to [4, 8, 14].

Other contributions in the literature extend traditional selection approaches to include further criteria (see, for example [10]). Doumpos et al. [6] discuss the relevance of multi-criteria decision methods and present a review of the MCDM approach for portfolio selection problems. For the extension of MCDM analysis to the inclusion of ethical and socially responsible criteria we refer to [1, 9].

To overcome, at least partially, the main limitations associated with this multi-step paradigm which separates the financial from the non-financial decision phases we propose an approach to integrate financial and non-financial criteria in the selection of a set of assets admissible for investment according to the preferences of the investor.

3 Constrained Tracking Error Approach

We propose a portfolio selection model in which the investor has a tracking goal with respect to a reference portfolio (benchmark) and we model the socially responsible level requirement for the portfolio as a constraint.

We assume a tracking goal can be expressed through a distance measure denoted by $\Phi(y_t, z_t)$ where y_t is the managed portfolio and z_t the reference benchmark. Tracking error portfolio management problem allows to set performance and risk profile references for the investment. The inclusion of a constraint on the dimension of socially responsibility of the investment changes the risk-return profile of the investment and the difference in terms of performance and risk can be easily captured by the tracking error.

We assume that for each asset we can observe a score representing its socially responsible level. The score can be either referred to a single dimension, like for example social or environmental or governance, or it can be an composite index representing a combination of different features. Under independence assumption, portfolio sustainability level is computed as weighted average of the sustainability levels of the assets included. We denote with α_{it} the level of the score α for asset i, $i = 1, \ldots, M$, at time t and the score for the portfolio, α_{yt} can be computed as the weighted average of the scores of the assets included in the portfolio where the vector of weights is the portfolio composition $\alpha_{yt} = \sum_{i=1}^{M} w_{it}\alpha_{it}$.

We denote with γ a social responsibility parameter which sets the minimum desired responsibility level. With $\bar{\alpha_t}$ we denote the social responsibility feature of the benchmark portfolio at time t. The resulting constraints for the tracking error problem can thus be written as $\alpha_{yt} \geq \gamma\bar{\alpha_t}$.

The inclusion of a deterministic constraint corresponds to the strictest requirement and eventually collapse into a preliminary screening approach. The formulation of a two-goals objective function is also possible and the definition of a proper trade-off between tracking and socially responsible level has to be determined according to the preferences of the investor.

In the proposed formulation the target level of the social responsibility score for the portfolio has to be kept along the entire planning horizon and not only reached at the end of the horizon. In this formulation of the problem the social responsibility target is binding and the investor is willing to modify the risk-return profile of the benchmark in order to achieve the desired goal. The choice of the parameter $\gamma > 0$ determines the resulting tracking error but this formulation does not allow to separately control for deviations in return and deviation in risk profile from the benchmark. The inclusion of additional constraints may allow to act specifically on the preferred dimension. We carried out some preliminary out-of-sample tests on real market data using the Eurostoxx50 as benchmark and the ESG disclosure score provided by Bloomberg for the stocks included in the Eurostoxx 50. The obtained results confirm previous findings in the literature, strong social responsibility requirements can only be met sacrificing financial performance.

4 Conclusion

In this contribution we briefly highlight the potential pitfalls associated with the paradigm commonly adopted in the literature for the selection of ethical and/or socially responsible investment and suggest a way to incorporate non-financial criteria into a broadly applied approach to portfolio selection based on a tracking error model.

References

1. Ballestero, E., Bravo, M., Perez-Gladish, B., Arenas-Parra, M., Pla-Santamaria, D.: Socially responsible investment: a multicreiteria approach to portfolio selection combining ethical and financial objectives. Eur. J. Oper. Res. **216**, 487–494 (2012)
2. Bauer, R., Koedijk, K., Otten, R.: International evidence on ethical mutual fund performance and investment style. J. Bank. Financ. **29**, 1761–1767 (2005)
3. Bello, Z.: Socially responsible investing in portfolio diversification. J. Financ. Res. **28**, 41–57 (2005)
4. Bilbao-Terol, A., Arenas-Parra, M., Canal-Fernández. V.: Selection of socially responsible portfolios using goal programming and fuzzy technology. Inf. Sci. **189**, 110–125 (2012)
5. Capelle-Blancard, G., Monjon, S.: Trends in the literature on socially responsible investment: looking for the keys under the Lamppost. Bus. Ethics. **21**(3), 239–250 (2011)
6. Doumpos, M., Zopounidis, C.: Multi-criteria classification methods in financial and banking decisions. Int. Trans. Oper. Res. **9**, 567–581 (2002)
7. Dupre, D., Girerd-Potin, I., Kassoua, R.: Adding an ethical dimension to portfolio management. In: EFMA 2003 Helsinki Meetings (2003)
8. Gasser, S., Rammerstorfer, M., Weinmayer, K.: Markowitz revisited: social portfolio engineering. Eur. J. Oper. Res. **258**, 1181–1190 (2016)
9. Hallerback, W., Ning, H., Soppe, A., Spronk, J.: A framework for managing a portfolio of socially responsible investments. Eur. J. Oper. Res. **153**, 517–529 (2004)
10. Hirschberger, M., Steuer, R., Utz, S., Wimmer, M., Qi, Y.: Computing the nondominated surface for a tri-criterion portfolio selection. Oper. Res. **611**, 69–183 (2013)
11. Ito, Y., Managi, S., Matsuda, A.: Performances of socially responsible investment and environmentally friendly funds. J. Oper. Res. Soc. **64**, 1583–1594 (2013)
12. Kempf, A., Osthoff, P.: The effect of socially responsible investing on portfolio performance. Eur. Financ. Manag. **13**, 908–922 (2004)
13. Renneboog, L., Ter Horst, J., Zhang, C.: Socially responsible investment: institutional aspects, performance, and investor behaviour. J. Bank. Financ. **32**(9), 1723–1742 (2008)
14. Sherwood, M.W., Pollard, J.: The risk-adjusted return potential of integrating ESG strategies into emerging market equities. J. Sustain. Financ. Invest. **8**, 26–44 (2018)
15. Statman, M., Glushkov, D.: Classifying and measuring the performance of socially responsible mutual funds. J. Portfol. Manage. **42**(2), 140–151 (2016)

A Generalized Moving Average Convergence/Divergence for Testing Semi-strong Market Efficiency

Francesco Bartolucci, Alessandro Cardinali, and Fulvia Pennoni

Abstract We propose a generalized version of the moving average converge divergence (MACD) indicator widely employed in the technical analysis and trading of financial markets. By assuming a martingale model with drift for prices, as well as for their transformed values, we propose a test statistic for the local drift and derive its main theoretical properties. The semi-strong market efficiency hypothesis is assessed through a bootstrap test. We conclude by applying the indicator to monitor the crude oil prices over a 6 years period.

Keywords Martingales · Nonparametric bootstrap test · Trading strategies

1 Introduction

We aim to monitor financial asset price series by a generalized version of the moving average convergence/divergence (MACD) trend indicator that is currently employed as technical indicator in trading systems [7]. We use the proposed indicator to test the semi-strong form of market efficiency as defined in his pioneer work by Fama [6]. He states that the price sequence follows a martingale model when, on the basis of the information available up to a given time, the expected returns are equal to zero. If these are positive, a submartingale model results which implies

F. Bartolucci
Department of Economics, University of Perugia, Perugia, Italy
e-mail: francesco.bartolucci@unipg.it

A. Cardinali (✉)
School of Computing and Mathematics, Plymouth University, Plymouth, UK
e-mail: alessandro.cardinali@plymouth.ac.uk

F. Pennoni
Department of Statistics and Quantitative Methods, University of Milano-Bicocca, Milano, Italy
e-mail: fulvia.pennoni@unimib.it

M. Corazza et al. (eds.), *Mathematical and Statistical Methods for Actuarial Sciences and Finance*, https://doi.org/10.1007/978-3-319-89824-7_18

that the trading rules based on the available information "*could not have greater expected profits then a policy of always buying-and-holding the security during the future period...*".

We consider a real valued sequence x_t representing prices at time $t = 1, 2, \ldots$ and, for some strictly increasing and differentiable real function $g(\cdot)$, we define $y_t = g(x_t)$ as the $g-$transformed prices, and $r_t = y_t - y_{t-1}$, with $t = 2, 3, \ldots$, as their first order difference. As notable examples of this framework, setting $g(\cdot) = 1$ implies working with raw prices and their differences, and setting $g(\cdot) = \log(x_t)$ implies working with log-prices and log-returns.

Consider an integer lag $s < t$, and let $\mathbf{r}_{t-s+1:t}$ be the vector of observations r_{t-h} for $h = 0, 1, \ldots, s-1$. We define its probability space as $(\Omega, \mathbb{P}, \mathfrak{F})$ along with the following two assumptions:

Assumption 1 $\mathbf{r}_{t-s+1:t}$ *is a (second-order) stationary vector with joint distribution* $P^s(\xi, \theta)$*, for some bounded* ξ *(mean) and* θ *(standard deviation);*

Assumption 2 $(r_{t-h} - \xi)|\mathfrak{F}_{t-h-1}$ *is a martingale difference (MD) sequence.*

According to the above assumptions, the sequence r_t is unconditionally distributed as a locally stationary process (see [5]) with piecewise constant mean, which becomes stationary (constant mean) in the time intervals $t - s + 1, \ldots, t$, while the (possibly transformed) prices y_t form a sub/super martingale. Therefore, the test

$$H_o : \xi = 0 \qquad \text{vs} \qquad H_a : \xi \neq 0 \tag{1}$$

evaluates the absence of (local) drift in the time interval at issue, with the null hypothesis H_o corresponding to the semi-strong efficient markets hypothesis [6].

2 The Proposed Indicator

We introduce a general definition of the moving average convergence/divergence which is defined as

$$\mathrm{MACD}^*_{t,sf} = \mathrm{M}^*_{t,f} - \mathrm{M}^*_{t,s},$$

where, in general, $\mathrm{M}^*_{t,k} = \sum_{h=0}^{k-1} a^*_{h,k}\ y_{t-h}$ is the weighted average based on k values, with $a^*_{h,k} \geq 0$ for all h and $\sum_{h=0}^{k-1} a^*_{h,k} = 1$. When $s > f$ the statistic $\mathrm{M}^*_{t,s}$ represents a smoother signal than $\mathrm{M}^*_{t,f}$ and their difference is usually adopted to indicate upcoming trends in financial prices. In the following, for space limitations we consider only the simple moving average of k observations $\mathrm{M}^A_{t,k}$ with the weights

determined as $a^A_{h,k} = 1/k$ for $h = 1, \ldots, k$. Therefore, the trend indicator is

$$\text{MACD}^A_{t,sf} = \sum_{h=0}^{f-1} a^A_{h,f}\ y_{t-h} - \sum_{h=0}^{s-1} a^A_{h,s}\ y_{t-h} = \sum_{h=0}^{s-1} c^A_{h,sf} r_{t-h} \tag{2}$$

with $s < f$ and

$$c^A_{h,sf} = \begin{cases} \dfrac{(h+1)}{f} - \dfrac{(h+1)}{s} = \dfrac{(h+1)(f-s)}{sf}, & \text{for } h = 0, \ldots, f-1 \\ 1 - \dfrac{h+1}{s}, & \text{for } h = 0, \ldots, s-1. \end{cases} \tag{3}$$

In general, we show that assumptions 1 and 2 defined in Sect. 1 imply that

$$\mathbb{E}(\text{MACD}^*_{t,sf}) = \xi\ A^*_{sf} = \mu^*_{sf}; \qquad \mathbb{V}(\text{MACD}^*_{t,sf}) = \theta^2\ B^*_{sf} = (\phi^*_{sf})^2,$$

where $A^*_{sf} = \sum_{h=1}^{s-1} c^*_{h,sf}$ and $B^*_{sf} = \sum_{h=1}^{s-1} (c^*_{h,sf})^2$ with $\theta > 0$. Therefore, based on (2) and (3) a martingale difference CLT (see [2]) implies the following convergence in distribution for increasing values of s and $f/s \to 0$

$$\frac{\text{MACD}^*_{t,sf} - \mu^*_{sf}}{\phi^*_{sf}} \xrightarrow{d} N(0, 1). \tag{4}$$

3 Nonparametric Bootstrap Test

We derive out test statistics $\text{T}^*_{t,sf}$ noting that the sample standard deviation of returns $\hat{\theta} \xrightarrow{p} \theta$, which implies that from (4) under the null hypothesis in (1) we have

$$\text{T}^*_{t,sf} = \frac{\text{MACD}^*_{t,sf}}{\hat{\phi}^*_{sf}} \xrightarrow{d} N(0, 1),$$

where $\hat{\phi}^*_{sf} = \hat{\theta}\sqrt{B^*_{sf}}$. In principle our algorithm is similar to the *Bootstat* procedure from [4], although here we focus on mean returns, see also [1]:

1. given the estimate $\hat{\xi}$ we calculate $\text{T}^*_{t,sf}$ we use the centered returns $\hat{r}_t = r_t - \hat{\xi}$, to produce a large number of bootstrap samples (e.g. 500) for $\text{T}^*_{t,sf}$ that we denote as $\text{T}^{**}_{t,sf}$;
2. we reject the null hypothesis if $\text{T}^*_{t,sf} < \text{T}^{**}_{(\frac{\alpha}{2}),sf}$ or $\text{T}^*_{t,sf} > \text{T}^{**}_{(1-\frac{\alpha}{2}),sf}$, where $\text{T}^{**}_{(x),sf}$ is the $100x^{th}$ percentile of the bootstrap distribution $\text{T}^{**}_{t,sf}$.

3.1 Simulations

Using simple moving averages as in (2) we set $T^*_{t,sf} = T^A_{t,sf}$. For $\xi = 0.0, 0.2, 0.4, 0.6$ we assess, respectively, the empirical size and power of our test when the data are simulated from the following innovations: Gaussian i.i.d., Student$-t(10)$ i.i.d., Gaussian $Garch(1,1)$, and Student$-t(10)$ $Garch(1,1)$, where the $Garch(1,1)$ model parameters (see [3]) are set to $\omega = 1e-6, \alpha_1 = 0.1$ and $\beta_1 = 0.8$. We find that, for hypotheses (1), the empirical size and power of the bootstrap test rapidly converges to the nominal levels. Its behaviour for large value of $s \geq 30, 40$ and small values of $f \leq 20$ is much more reliable than the standard test based on the asymptotic Gaussian distribution. This suggests that a long/short equity trading strategy utilizing high-medium frequency data can be based on this test, and in particular on the sign of $T^A_{t,sf}$ that we represent as $\text{sgn}(T^A_{t,sf})$, with active positions only set/maintained under H_a from (1).

4 Empirical Test Results

We apply the proposed strategy based on the $\text{sgn}(T^A_{t,sf})$ trading rules: 1 open or maintain a long position, 0 (under H_o) close any position eventually open or do not open any, and -1 open or maintain a short position. Figure 1 shows the cumulative returns according to the proposed strategy to daily crude oil prices with $s = 30$ and $f = 10$. Data are 1469 daily quotes collected over the 6 year period 2010–2016. We

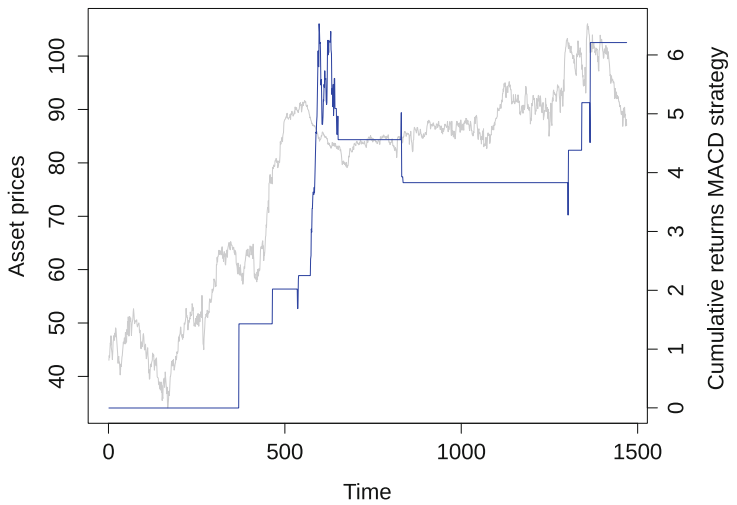


Fig. 1 Left vertical axis: asset prices (Bloomberg®). Right vertical axis: cumulative returns according with the MACD trading strategy based on simple averages

also applied our trading rule to daily quotes of SPY ETF, Eurodollar Forex and the Nasdaq Index finding that the cumulative returns are higher that that we would gain by using the standard Gaussian test.

References

1. Beran, R.: Simulated power functions. Ann. Stat. **14**, 151–173 (1986)
2. Billingsley, P.: Probability and Measure, 3rd edn. Wiley, New York (1995)
3. Bollerslev, T.: Generalized autoregressive conditional heteroskedasticity. J. Econ. **31**, 307–327 (1986)
4. Cardinali, A., Nason, G.P.: Costationarity of locally stationary time series. J. Time Ser. Econ. **2**(2), 1–18, Article 1 (2010)
5. Dahlhaus, R.: Locally stationary processes. In: Handbook of Statistics, vol. 30, pp. 351–412. Elsevier, Amsterdam (2012)
6. Fama, E.: Efficient capital markets: a review of theory and empirical work. J. Finance **25**, 383–417 (1970)
7. Raudys, A., Lenčiauskas, V., Malčius, E.: Moving averages for financial data smoothing. In: International Conference on Information and Software Technologies, pp. 34–45 (2013)

Periodic Autoregressive Models with Multiple Structural Changes by Genetic Algorithms

Francesco Battaglia, Domenico Cucina, and Manuel Rizzo

Abstract We present a model and a computational procedure for dealing with seasonality and regime changes in time series. In this work we are interested in time series which in addition to trend display seasonality in mean, in autocorrelation and in variance. These type of series appears in many areas, including hydrology, meteorology, economics and finance. The seasonality is accounted for by subset PAR modelling, for which each season follows a possibly different Autoregressive model. Levels, trend, autoregressive parameters and residual variances are allowed to change their values at fixed unknown times. The identification of number and location of structural changes, as well as PAR lags indicators, is based on Genetic Algorithms, which are suitable because of high dimensionality of the discrete search space. An application to Italian industrial production index time series is also proposed.

Keywords Time series · Seasonality · Nonstationarity

1 Model Description and Estimation

In this work we are interested in economic time series showing a trend and seasonal fluctuations that are not very stable over time. This may be caused by economic agents who have preferences, technologies, constraints, and expectations which are not constant over the seasons [1]. In these kind of series it may happen that linear seasonal adjustment filters, as in $SARIMA$ models, are not likely to remove the intrinsic seasonality. In that case the residuals from these models show patterns that can be explained by the presence of dynamic periodicity [1]. Recently, economic

F. Battaglia · M. Rizzo (✉)
Department of Statistical Sciences, Sapienza University of Rome, Rome, Italy
e-mail: francesco.battaglia@uniroma1.it; manuel.rizzo@uniroma1.it

D. Cucina
Department of Economics and Statistics, University of Salerno, Fisciano, Italy
e-mail: dcucina@unisa.it

M. Corazza et al. (eds.), *Mathematical and Statistical Methods for Actuarial Sciences and Finance*, https://doi.org/10.1007/978-3-319-89824-7_19

models in which seasonally varying parameters are allowed attracted some attention [2]. The present paper considers PAR models with linear piecewise trends, allowing nonstationarity and seasonality simultaneously. There is empirical evidence of the existence of discontinuities due to structural changes possibly originated by policy changes [3]. Consequently, detecting the existence of structural changes and estimating their number and locations in periodic time series is an important stage. In this paper we propose an automatic procedure based on Genetic Algorithms (GAs) for estimating number and locations of change points.

We refer to a seasonal time series of period s, observed s times a year for N complete years ($s = 12$ for monthly data). This series is possibly divided in a number of regimes up to M, specified by $m = M - 1$ change points $\tau_1, \ldots, \tau_m$ occurring at the end of the year $\tau_j - 1$, and set $\tau_0 = 1$ and $\tau_M = N + 1$. In order to ensure reasonable estimates, it is required that each regime contains at least a minimum number mrl of years, therefore $\tau_j \geq \tau_{j-1} + mrl$ for any regime j. We let $R^j = \{\tau_{j-1}, \tau_{j-1} + 1, \ldots, \tau_j - 1\}$, $j = 1, \ldots, M$, so that if year n belongs to set R^j then observation at time $(n-1)s + k$ is in regime j, where $k = 1, \ldots, s$. For the observation in season k of year n the model is:

$$X_{(n-1)s+k} = a^j + b^j[(n-1)s + k] + \mu_k^j + Y_{(n-1)s+k}, \; n \in R^j, \; 1 \leq k \leq s, 1 \leq j \leq M \tag{1}$$

where process $\{Y_{(n-1)s+k}\}$ is a PAR given by:

$$Y_{(n-1)s+k} = \sum_{i=1}^{p} \phi_i^j(k) Y_{(n-1)s+k-i} + \epsilon_{(n-1)s+k}, \quad n \in R^j, \; 1 \leq k \leq s, 1 \leq j \leq M. \tag{2}$$

Model (1) is a pure structural change model where the trend parameters a^j and b^j depend only on the regime, whereas means μ_k^j are allowed to change also with seasons. The parameters $\phi_i^j(k)$, $i = 1, \ldots, p$, represent the PAR coefficients during season k of the j-th regime, and some of them may be allowed to be constrained to zero, in order to get more parsimonious subset models. The innovations process $\{\epsilon_t\}$ in Eq. (2) corresponds to a periodic white noise, with $E(\epsilon_{(n-1)s+k}) = 0$ and $Var(\epsilon_{(n-1)s+k}) = \sigma_j^2(k) > 0, n \in R^j$.

Our model is characterized by both structural parameters: the changepoints number m, their locations $\tau_1, \tau_2, \ldots, \tau_m$ and PAR lags indicator, which specifies presence or absence of PAR parameters; and by regression parameters: the trend intercepts a^j, the slopes b^j, the seasonal means μ_k^j, the AR parameters $\phi_i^j(k)$ and the innovation variances $\sigma_j^2(k)$. These regression parameters, assuming that structural parameters are given, can be estimated by Ordinary Least Squares.

Identification of this complex model requires only selection of structural parameters. We adopt a procedure based on a GA [4] for optimization. The GA is a nature-inspired metaheuristic algorithm, for which a population of solutions evolves

through iterations by use of so-called genetic operators, until a stopping rule is reached. Out implementation optimizes structural parameters simultaneously by considering an objective function based on NAIC criterion, introduced in [5] for threshold models selection, given by:

$$g = [\sum_{j=1}^{M}\sum_{k=1}^{s} n_{j,k}\log(\hat{\sigma}_j^2(k)) + 2\sum_{j=1}^{M}\sum_{k=1}^{s} P_{j,k}]/(Ns),$$

where $\hat{\sigma}_j^2(k)$ is the model residual variance of series in regime j and season k, $n_{j,k}$ is related sample size, $P_{j,k}$ is related number of parameters. Other options of penalization could be adopted basing on identification criteria literature. The GA objective function to be maximized, named *fitness*, is a scaled exponential transformation of g given by $f = \exp\{-g/\beta\}$, where β is a positive scaling constant.

2 Application and Conclusions

The industrial production index is an important macroeconomic variable since it can reflect business cycle behaviour and changing directions of an underlying trend. Forecasts of this variable are used by many decision makers [1]. Given that decisions sometimes concern time intervals shorter than 1 year, forecasts for the monthly observed industrial production index can be useful. The time series of industrial production index in Italy (Ateco category C) for months 1990.1–2016.12 is displayed in Fig. 1. From this figure it can be observed that the time series shows trend and a non-stable seasonal patterns.

The GA procedure, when applied to the original data, segmented the series in two regimes with one structural change at $\tau_1 = 2009$ and an opposite trend behaviour (see Fig. 1). To examine the effectiveness of subset piecewise PAR models, we compared their performance, in terms of fitness, with that of some other models used for seasonal data. Four types of models were considered in this study and

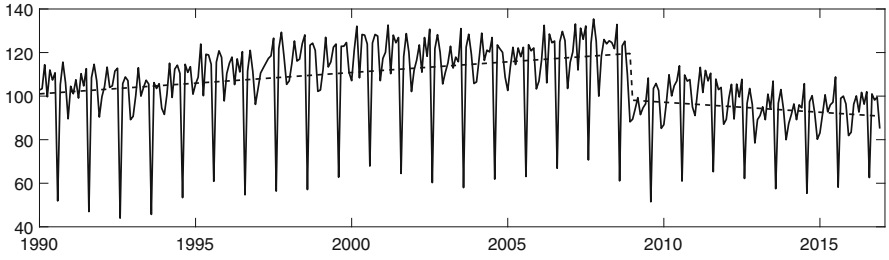


Fig. 1 Italy Industrial Production Index series with estimated trend and changepoint at end 2008

Table 1 Performance of different models

Model	Fitness	Prediction MSE
Subset piecewise $PAR(3)$	0.799	20.22
Complete piecewise $PAR(3)$	0.789	26.64
Piecewise $AR(3)$	0.769	20.37
$SARIMA(2,0,1) \times (0,1,0)_{12}$	0.738	27.11

the corresponding fitness values are reported in the first column of Table 1. The first two models are a complete and subset piecewise PAR, as described above. The third model is a non-periodic $AR(3)$ applied on data after removing the break, the trend and the seasonal means. The fourth model is a more conventional $SARIMA(2,0,1) \times (0,1,0)_{12}$ but estimated separately for each regime. The best model, in terms of fitness, is the subset piecewise PAR, while the complete model has a slightly smaller fitness. The non-periodic piecewise AR reaches an even smaller fitness, while the $SARIMA$ model is markedly less fit.

As far as diagnostic checking is concerned, the Box-Pierce statistics with 12 or 18 lags computed on residuals were not significant at level 0.01 for the piecewise $PAR(3)$ models, and largely significant for the other two models.

We computed also out-of-sample forecasts for the first 9 months of 2017. The mean square prediction errors for the four models are reported in the second column of Table 1. We may conclude that for the industrial production index dataset the proposed procedure outperforms the other models both in terms of fitting and forecasting.

References

1. Franses, P.H.: Periodically integrated subset autoregressions for dutch industrial production and money stock. J. Forecast. **12**, 601–613 (1993)
2. Franses, P.H., Paap, R.: Periodic Time Series Models. Oxford University Press, Oxford (2004)
3. Giordani, P., Kohn, R., van Dijk, D.: A unified approach to nonlinearity, structural change and outliers. J. Economet. **137**, 112–133 (2007)
4. Goldberg, D.E.: Genetic Algorithms in Search, Optimization, and Machine Learning. Addison-Wesley, New York (1989)
5. Tong, H.: Non-linear Time Series. A Dynamical System Approach. Oxford University Press, Oxford (1990)

Mortality Projection Using Bayesian Model Averaging

Andrés Gustavo Benchimol, Juan Miguel Marín Diazaraque, Irene Albarrán Lozano, and Pablo Jesús Alonso-González

Abstract In this paper we propose Bayesian specifications of four of the most widespread models used for mortality projection: Lee-Carter, Renshaw-Haberman, Cairns-Blake-Dowd, and its extension including cohort effects. We introduce the Bayesian model averaging in mortality projection in order to obtain an assembled model considering model uncertainty. We work with Spanish mortality data from the Human Mortality Database, and results suggest that applying this technique yields projections with better properties than those obtained with the individual models considered separately.

Keywords Bayesian model averaging · Cairns-Blake-Dowd model · Lee-Carter model · Longevity · Model uncertainty · Mortality projection · Renshaw-Haberman model

1 Introduction

Human mortality has experienced considerable improvements since the second part of the twentieth century and people have been living more than expected. This

A. G. Benchimol (✉)
Department of Statistics & Actuarial Science, The University of Hong Kong, Pokfulam, Hong Kong
e-mail: benchi@hku.hk

J. M. Marín Diazaraque
Department of Statistics, Universidad Carlos III de Madrid, Getafe, Spain
e-mail: jmmarin@est-econ.uc3m.es

I. Albarrán Lozano
Department of Statistics, Universidad Carlos III de Madrid, Colmenarejo, Spain
e-mail: ialbarra@est-econ.uc3m.es

P. J. Alonso-González
Economics Department, Universidad de Alcalá, Alcalá de Henares, Spain
e-mail: pablo.alonsog@uah.es

M. Corazza et al. (eds.), *Mathematical and Statistical Methods for Actuarial Sciences and Finance*, https://doi.org/10.1007/978-3-319-89824-7_20

positive fact has a side effect, which is a threat to life annuities business and public pensions systems. Thus, the need to resort to suitable models for projecting mortality rates accurately has become crucial for all economic issues linked to the elderly.

When it comes to model selection, statisticians usually choose the 'best' model among a set of models according to some criterion, discarding the remaining ones. This practice ignores model uncertainty, which is transferred to the projections leading to over–confident inferences. A proposal to get over this issue is to assemble models according to some criteria. Thus, the final model turns out to be a mixture that takes the whole set of candidates models into consideration, and it often leads to better projections and more accurate inferences.

In this paper, we work with Bayesian specifications of four of the most widespread models used for mortality projection and assemble them applying a technique called *Bayesian Model Averaging* (BMA) by Hoeting et al. [4]. To the best of our knowledge, the BMA had not been used in this field before.

2 The Bayesian Model Averaging (BMA)

In statistical practice, analysts usually choose a model and work as if data were generated by the selected model, leading to inferences and decisions based on it, without considering the uncertainty associated to the model selection.

Model assembling techniques solve the former issue while considering several models simultaneously.

In particular, the Bayesian model averaging makes inferences based on a weighted average over the model space. In this way, the model uncertainty is also included in both predictions and parameter estimates (see [5] and [3]).

From a practical point of view, the posterior probability for each model can be computed by a MCMC algorithm (see [2]) and the various projections are weighted by the posterior probability in order to obtain the assembled model.

3 Mortality Projection Applying BMA

We worked with four of the most popular models for mortality projection. Let $q_x(t)$ and $m_x(t)$ be the death probability and central death rate at age x in year t, respectively. Table 1 summarizes their specifications.

M1 was the first age-period model, introduced by Lee and Carter [6]. The authors proposed a model for describing the change in mortality as a function of a latent factor evolving throughout time. M2 by Renshaw and Haberman [9] is a generalization of M1 including cohort effects. On the other hand, M5 by Cairns et al. [1] is a two-latent factor model and M6 is its generalization including cohort effects.

The reason for working with these four models is that that M1 is the most extended mortality model, while M3 is the most widespread alternative to M1 model, while M2 and M6 are their respective specifications including cohort effects. It is worth mentioning that there are other widespread approaches based on B-Splines and P-Splines to add smoothness. However, these type of models are suitable for fitting mortality, but not so much for projections, and therefore they were not considered in this proposal.

We worked with Spanish mortality data from the *Human Mortality Database* (see www.mortality.org). We focused on male gender data, a timespan between calendar years 1960 and 2009, and ages between 60 and 100 years old.

Table 1 Mortality projection models

Model	Specification
M1	$\log[m_x(t)] = \beta_x^{(1)} + \beta_x^{(2)}\kappa_t$
M2	$\log[m_x(t)] = \beta_x^{(1)} + \beta_x^{(2)}\kappa_t + \beta_x^{(3)}\gamma_{t-x}$
M5	$\text{logit}[q_x(t)] = \kappa_t^{(1)} + \kappa_t^{(2)}(x - \overline{x})$
M6	$\text{logit}[q_x(t)] = \kappa_t^{(1)} + \kappa_t^{(2)}(x - \overline{x}) + \gamma_{t-x}$

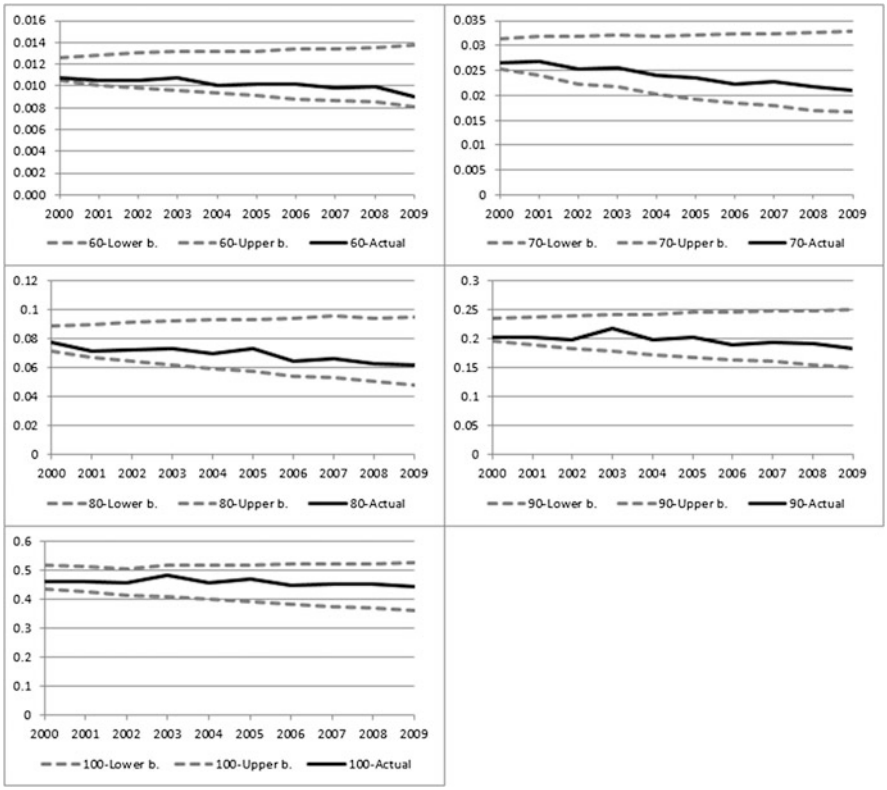


Fig. 1 95% prediction bands and actual mortality trajectories of $m_x(t)$

We first considered Bayesian specifications of the former models. We assumed vaguely informative prior distributions for all the parameters of the four models. For the sake of brevity we do not show the complete mathematical details.

Regarding the assembling process, it was assumed a discrete uniform distribution on the four models weights as prior distribution.

The MCMC algorithm was programmed using `Jags` [7] by means of the package `R2jags` [10] from the `R` project [8].

In order to validate the predictive performance of the assembled model, we trained the model with data from 1960 to 1999 and obtained projections for the central death rates $m_x(t)$ for the period 2000 to 2009 and compared them with the actual rates.

For visualizing the results, Fig. 1 shows both the 95% prediction bands and actual mortality trajectories for ages $x = 60$, 70, 80, 90 and 100. As it can be seen, prediction bands generated by the methodology applied to the training sample contain the actual mortality trajectories observed in the validation sample, which was not the case for the models considered separately for this data set.

4 Final Remarks

We applied the Bayesian model averaging to some of the most extended parametric mortality models in order to consider model uncertainty in mortality projection. Although this model assembling methodology had been applied to different areas, it is the first time that it is used in this field, to the best of our knowledge.

Under this methodology, the resulting projected central deaths rates were accurate and it is proposed to be used by actuaries in life insurance companies and public pension systems.

References

1. Cairns, A., Blake, D., Dowd, K.: A two factor model for stochastic mortality with parameter uncertainty: theory and calibration. J. Risk Insur. **73**(4), 687–718 (2006)
2. Chib, S.: Marginal likelihood from the Gibbs output. J. Am. Stat. Assoc. **432**(90), 1313–1321 (1995)
3. Hoeting, J.A.: Methodology for Bayesian model averaging: an update. International Biometrics Conference Proceedings (2002). http://www.stat.colostate.edu/~jah/papers/ibcbma.pdf
4. Hoeting, J.A., Madigan, D., Raftery, A.E., Volinsky, C.T.: Bayesian model averaging. In: Proceedings of the AAAI Workshop on Integrating Multiple Learned Models, pp. 77–83 (1998)
5. Hoeting, J.A., Madigan, D., Raftery, A.E., Volinsky, C.T.: Bayesian model averaging: a tutorial. Stat. Sci. **14**, 382–401 (1999)
6. Lee, R.D., Carter, L.R.: Modeling and forecasting U.S. mortality. J. Am. Stat. Soc. **87**, 659–675 (1992)
7. Plummer, M., et al.: JAGS: a program for analysis of Bayesian graphical models using Gibbs sampling. In: Proceedings of the Third International Workshop on 'Distributed Statistical Computing' (2003)

8. R Core Team: R: A Language and Environment for Statistical Computing (2012)
9. 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)
10. Su, Y.S., Yajima, M.: R2jags: Using R to Run 'JAGS' (2015)

Robust Time-Varying Undirected Graphs

Mauro Bernardi and Paola Stolfi

Abstract Undirected graphs are useful tools for the analysis of sparse and high-dimensional data sets. In this setting the sparsity helps in reducing the complexity of the model. However, sparse graphs are usually estimated under the Gaussian paradigm thereby leading to estimates that are very sensitive to the presence of outlying observations. In this paper we deal with sparse time-varying undirected graphs, namely sparse graphs whose structure evolves over time. Our contribution is to provide a robustification of these models, in particular we propose a robust estimator which minimises the γ-divergence. We provide an algorithm for the parameter estimation and we investigate the rate of convergence of the proposed estimator.

Keywords Divergence · Kernel methods · Robust methods · Dynamic models · Graphical models

1 Introduction

Sparse methods refer to statistical approaches specifically tailored to deal with estimation of large dimensional models with potentially many more features than observations. Those problems have recently attracted lots of researchers and many contributions have appeared in the literature mainly because of the relevance of the applications in several different fields ranging from econometrics to physics and biology. For an up to date and comprehensive accounting of those recent developments see, e.g., [4]. Despite their practical relevance, most of the recent

M. Bernardi (✉)
Department of Statistical Sciences, University of Padova, Padova, Italy
e-mail: mauro.bernardi@unipd.it

P. Stolfi
Department of Economics, Roma Tre University and Istituto per le Applicazioni del Calcolo "Mauro Picone" - CNR, Rome, Italy
e-mail: paola.stolfi@uniroma3.it

M. Corazza et al. (eds.), *Mathematical and Statistical Methods for Actuarial Sciences and Finance*, https://doi.org/10.1007/978-3-319-89824-7_21

contributions have been confined under the restrictive assumption of independently and identically distributed Gaussian observations, with only few exceptions: [2, 6, 7], and [1]. However, they do not deal with the potentially time-varying covariance matrices that may originate from dependent data. This case has been investigated by [8], where they introduce sparse time-varying undirected graphs, namely, graphs whose structure changes smoothly over time. These models works well in high dimensional setting, that is when lots of parameters need to be estimated and few observations are available. A drawback, however, is their sensitivity to outliers due to the penalised maximum likelihood procedure. The contribution of this paper is to solve this issue by proposing a robustification of the sparse time-varying graphs proposed by [8]. Specifically, following [5], we propose a robust estimator which minimises the γ-divergence between the postulated and theoretical distribution. Furthermore, we provide an algorithm to handle parameter estimation and we investigate the rate of convergence of the covariance matrix and of its inverse.

2 Methodology

Following [8] we consider the dynamic model

$$W_t = W_{t-1} + Z_t, \qquad Z_t \sim \mathsf{N}(0, \Sigma_t), \text{ for } t > 0.$$
$$W_0 \sim \mathsf{N}(0, \Sigma_0).$$

The Lasso-penalised maximum likelihood in the non i.i.d. case is given at time t by

$$\widehat{\Sigma}_n(t) = \operatorname{argmin}_{\Sigma \succ 0} \left\{ \operatorname{tr}\left(\Sigma^{-1} \widehat{S}_n(t) \right) + \ln|\Sigma| + \lambda |\Sigma^{-1}|_1 \right\}, \tag{1}$$

where

$$\widehat{S}_n(t) = \frac{\sum_s r_{st} Z_s Z_s'}{\sum_s r_{st}}, \tag{2}$$

is a weighted covariance matrix with weights $r_{st} = K(|s-t|/h_n)$ given by a symmetric nonnegative kernel over time. In order to get a robust alternative to the estimator in Eq. (1) we follow the approach of [5] and we consider the γ-divergence between the postulated and the true unknown distribution. More precisely, let

$$\ell_\gamma(Z, \Sigma) = -\frac{1}{\gamma} \ln \left\{ \frac{1}{n} \sum_s f(Z_s, \Sigma)^\gamma \right\} + \frac{1}{1+\gamma} \ln \int f(Z, \Sigma)^{1+\gamma} \, dZ, \tag{3}$$

the negative γ-scoring function as implied by the γ-divergence. In the non i.i.d. framework of [8] formalised in Eqs. (1)–(2) we get the following formula for the

negative γ-likelihood function:

$$\ell_\gamma(Z, \Sigma) = -\frac{1}{\gamma}\ln\left\{\frac{1}{n}\sum_s \frac{\exp\left\{-\gamma \operatorname{tr}\left(\Sigma^{-1}\widehat{S}_s\right)\right\}}{(2\pi)^{\frac{\gamma}{2}}|\Sigma|^{\frac{\gamma}{2}}}\right\} + \frac{1}{1+\gamma}\ln\left\{\int \frac{\exp\left\{-(1+\gamma)\operatorname{tr}\left(\Sigma^{-1}Z_t Z_t'\right)\right\}}{(2\pi)^{\frac{1+\gamma}{2}}|\Sigma|^{\frac{1+\gamma}{2}}} dZ_t\right\}. \tag{4}$$

Then the γ-Lasso maximum likelihood is therefore obtained by solving the following minimisation problem:

$$\widehat{\Sigma}_n(t) = \arg\min_{\Sigma \succ 0} \ell_\gamma(\Sigma) + \lambda|\Sigma^{-1}|_1, \qquad \text{for } t = 1, 2, \ldots, n. \tag{5}$$

3 Algorithm

Let $\vartheta_t = \left(\operatorname{vec}(\Sigma_t)'\right)'$ be the vector of parameters at time t, and $\widehat{\vartheta}_t^{(j)}$ be an estimate at the j-th iteration, then denote

$$\omega_s^{(j)} = \frac{f\left(Z_s, \widehat{\vartheta}_t^{(j)}\right)^\gamma}{\sum_k f\left(Z_k, \widehat{\vartheta}_t^{(j)}\right)^\gamma}, \qquad x_s^{(j)} = \sum_k f\left(Z_k, \widehat{\vartheta}_t^{(j)}\right)^\gamma \frac{f\left(Z_s, \widehat{\vartheta}_t^{(j)}\right)^\gamma}{f\left(Z_s, \widehat{\vartheta}_t^{(j)}\right)^\gamma},$$

in such a way that $\omega_s^{(j)} x_s^{(j)} = f\left(Z_s, \widehat{\vartheta}_t^{(j)}\right)^\gamma$. Using the Jensen's inequality to the convex function $y = -\log(x)$, we get

$$\ell_1(\vartheta_t) \leq -\sum_s \omega_s^{(j)} \log\left(f(Z_s, \vartheta)\right) + \kappa = \tilde{\ell}_1(\vartheta_t), \tag{6}$$

where κ is a constant term. Equation (6) is a weighted negative log-likelihood and satisfies the properties of a majorisation function, i.e., then

$$\tilde{\ell}_{\gamma,\lambda}(\vartheta_t) = \tilde{\ell}_1(\vartheta_t) + \ell_2(\vartheta_t) + \lambda|\Sigma^{-1}|_1 \tag{7}$$

From the properties of $\tilde{\ell}_1$ it holds that $\tilde{\ell}_{\gamma,\lambda}$ is monotone decreasing. Following [5] we proved that

$$\tilde{\ell}_1(\vartheta_t) + \ell_2(\vartheta_t) = \frac{1}{2}\operatorname{tr}\left(\Sigma^{-1}\widehat{S}_n^\omega(t)\right) - \log|\Sigma|\left(\frac{1}{2(1+\gamma)}\right) + \tilde{\kappa}, \tag{8}$$

where

$$\widehat{S}_n^{\omega}(t) = \frac{\sum_s \omega_s^{(j)} r_{ts} Z_t Z_t'}{\sum_s r_{ts}}, \tag{9}$$

and $\tilde{\kappa}$ is a constant. Then to find the update $\vartheta_t^{(j+1)}$ we need to solve the following nonconvex minimisation problem

$$\widehat{\vartheta}_t^{(j+1)} = \arg\min_{\vartheta_t} \frac{1}{2}\mathrm{tr}\left(\Sigma^{-1}\widehat{S}_n^{\omega}(t)\right) - \left(\frac{1}{2(1+\gamma)}\right)\log|\Sigma| + \lambda|\Sigma^{-1}|_1, \tag{10}$$

that correspond to graphical lasso, see [3]. We investigated the asymptotic property of the proposed estimator and we get similar rate of convergence of the covariance matrix and of its inverse to that presented in [8].

4 Simulation Study

In order to test the effectiveness of the proposed estimator we replicated the simulation study in [8] by considering contaminated data. In particular we consider a graph that evolves according to a type of Erdos–Rényi random graph model, we start from a graph with 50 edges with certain weights, the vertex changes smoothly and the weights smoothly decay to zero.

References

1. Bernardi, M., Petrella, L., Stolfi, P.: The sparse multivariate method of simulated quantiles (2017). arXiv Preprint. arXiv:1710.03453
2. Finegold, M., Drton, M.: Rejoinder: Robust Bayesian graphical modeling using Dirichlet t-distributions [mr3256053; mr3256054; mr3256055; mr3256052]. Bayesian Anal. **9**(3), 591–596 (2014)
3. Friedman, J., Hastie, T., Tibshirani, R.: Sparse inverse covariance estimation with the graphical lasso. Biostatistics **9**(3), 432–441 (2008)
4. Hastie, T., Tibshirani, R., Wainwright, M.: Statistical Learning with Sparsity: The Lasso and Generalizations. Monographs on Statistics and Applied Probability, vol. 143. CRC Press, Boca Raton (2015)
5. Hirose, K., Fujisawa, H., Sese, J.: Robust sparse Gaussian graphical modeling. J. Multivariate Anal. **161**, 172–190 (2017)
6. Lafferty, J., Liu, H., Wasserman, L.: Sparse nonparametric graphical models. Statist. Sci. **27**(4), 519–537 (2012)
7. Vogel, D., Tyler, D.E.: Robust estimators for nondecomposable elliptical graphical models. Biometrika **101**(4), 865–882 (2014)
8. Zhou, S., Lafferty, J., Wasserman, L.: Time varying undirected graphs. Mach. Learn. **80**(2–3), 295–319 (2010)

Two-Sided Skew and Shape Dynamic Conditional Score Models

Alberto Bernardi and Mauro Bernardi

Abstract In this paper we introduce a family of 2-Sided Skew and Shape distributions that accounts for asymmetry in the tails decay. The proposed distributions account for many of the stylised fact frequently observed in financial time series, except for the time-varying nature of moments of any order. To this aim we extend the model to a dynamic framework by means of the score updating mechanism. The asymptotic theory of the proposed model is derived under mild conditions.

Keywords Dynamic models · Skew and shape distribution

1 Introduction

The financial econometrics literature offers several valid alternatives to model the dynamic evolution of the conditional distribution of financial returns. In a fully parametric setting, following the same arguments presented by Cox et al. [3], the first issue is to choose between the alternative classes of observation driven and parameter driven models. The former can be well represented by the GARCH-type models introduced by Engle [5] and Bollerslev [2], while the latter is typically associated with the class of latent factor models usually represented in state space form; see, e.g., [7, 8]. Recently, the family of observation driven models has been enlarged by the inclusion of the Dynamic Conditional Score (DCS) models, introduced by Creal et al. [4] and Harvey [9]. GAS models are gaining a lot of consideration in many fields of theoretical and applied time-series analysis, because of their ability to use all the information coming from the entire conditional distribution of the observed instead of just using proxies for past conditional moments, as for GARCH-type models, see, e.g., [10]. Furthermore, an additional advantage GAS models share with the whole family of observation driven models is

A. Bernardi (✉) · M. Bernardi
Department of Statistical Sciences, University of Padova, Padova, Italy
e-mail: alberto.bernardi.5@studenti.unipd.it; mauro.bernardi@unipd.it

M. Corazza et al. (eds.), *Mathematical and Statistical Methods for Actuarial Sciences and Finance*, https://doi.org/10.1007/978-3-319-89824-7_22

the availability of a closed form expression for the likelihood function and the ease of evaluation of the updating mechanism.

In this paper we introduce the family of 2-Sided Skew and Shape (2SSS) distributions extending the generating mechanism of [6] to account for an asymmetric tails decay. We also provide analytical closed form expressions for the cumulative density function, the moments up to the fourth-order, the score and the information matrix. Then, we further extend the 2SSS distribution to a dynamic framework by adapting the DCS mechanism. The asymptotic theory of the proposed model is derived under mild conditions.

2 Two-Sided Skew and Shape Distribution

The 2-Sided Skew and Shape distribution (2SSS) is defined by generalised the skewing mechanism introduced by [6]. Denote by $f\left(x, \lambda, \sigma_f, \vartheta_f\right) = \frac{1}{\sigma_f} f\left(\frac{x-\lambda}{\sigma_f}, \vartheta_f\right)$ and $g\left(x, \lambda, \sigma_g, \vartheta_g\right) = \frac{1}{\sigma_g} f\left(\frac{x-\lambda}{\sigma_g}, \vartheta_g\right)$, the unnormalised kernels of two different symmetric distributions, where $\lambda \in \mathbb{R}$ and $\sigma_f, \sigma_g \in \mathbb{R}^+$ are location and scale parameters, respectively, while ϑ_f and ϑ_g are parameters that are specific to the two distributions. We define the 2SSS probability density function constructed of f truncated to $(-\infty, \lambda)$ and g truncated to $[\lambda, \infty)$, as follows:

$$\tilde{f}\left(x; \lambda, \psi, \nu\right) \propto \frac{1}{\xi\psi} f\left(\frac{x-\lambda}{\xi\psi}, \vartheta_f\right) \mathbb{1}_{(-\infty,\lambda)}\left(x\right) + \frac{\xi}{\psi} g\left(\frac{\xi\left(x-\lambda\right)}{\psi}, \vartheta_g\right) \mathbb{1}_{(-\infty,\lambda)}\left(x\right), \tag{1}$$

and the normalising constant that renders Eq. (1) a proper probability density function is

$$C_{\vartheta_f,\vartheta_g} = F\left(0, 0, 1, \vartheta_f\right) + G\left(0, 0, 1, \vartheta_g\right), \tag{2}$$

where capital letters denote the cumulative density functions.

Example 1 Let us consider the case where f is the kernel of the Student–t density and g is the kernel of the Generalised Error distribution:

$$f\left(y; \lambda, \psi, \nu\right) \propto \left(1 + \frac{\left(y-\lambda\right)^2}{\nu\psi^2}\right)^{-\frac{\nu+1}{2}} \mathbb{1}_{(-\infty,\infty)}\left(y\right) \tag{3}$$

$$g\left(y; \lambda, \psi, \beta\right) \propto \exp\left\{-\frac{1}{2}\left[\frac{\left(y-\lambda\right)^2}{\psi^2}\right]^{\frac{\beta}{2}}\right\} \mathbb{1}_{(-\infty,\infty)}\left(y\right), \tag{4}$$

then the normalised 2SSS distribution becomes

$$f_Y(y;\lambda,\psi,\nu,\beta) = \frac{2}{C(\xi,\beta,\nu)\sqrt{\psi^2}}\left(f(y;\lambda,\psi\xi,\nu)\,\mathbb{1}_{(-\infty,\lambda)}(y) + g\left(y_t;\lambda,\frac{\psi}{\xi},\beta\right)\mathbb{1}_{(-\infty,\lambda)}(y)\right), \tag{5}$$

where $C(\xi,\beta,\nu) = C_f(\xi,\nu) + C_g(\xi,\beta)$ is the normalising constant, where

$$C_f(\xi,\nu) = \frac{\Gamma\left(\frac{\nu}{2}\right)\sqrt{\pi\nu\xi^2}}{\Gamma\left(\frac{\nu+1}{2}\right)}, \qquad C_g(\xi,\beta) = \frac{2^{\frac{1}{\beta}+1}\Gamma\left(\frac{1}{\beta}\right)}{\beta\xi}.$$

3 2SSS Dynamic Conditional Score Model

The key feature of GAS models is that the updating mechanism of a time-varying parameter is driven by the score of the conditional density, which becomes the forcing variable for parameters' dynamics.

Formally, we assume that, the random variable y_t at time $t = 1, 2, \ldots, T$, is conditionally distributed according to the filtration $\mathscr{F}_{i,t-1} = \sigma(y_1, y_2, \ldots, y_{t-1})$ as:

$$y_t \mid \mathscr{F}_{t-1} \sim 2\mathscr{SSS}_\star\left(y_t;\mu_t,\sigma_t^2,\xi,\nu,\beta\right), \tag{6}$$

where $2\mathscr{SSS}(y_t;\cdot)$ denotes the Two sided Skew and Shape distribution defined in Sect. 2 with time-varying scale $\sigma_t^2 \in \mathbb{R}^+$, and constant constant location $\mu \in \mathbb{R}$, shape $\nu \in (4,+\infty)$, $\beta \in (1,2)$ and skewness parameter $\xi \in \mathbb{R}^+$. Moreover, we parameterise the Two sided Skew and Shape distribution as in [1] in such a way that the location μ and the scale σ_t^2 coincide with the conditional mean and variance, respectively, i.e., $\mathsf{E}_{t-1}[y_t] = \mu$ and $\mathrm{Var}_{t-1}[y_t] = \sigma_t^2$, where $\mathsf{E}_{t-1}[y_t]$ is a shorthand to denote the expectation of y_t conditional to the past information, i.e., $\mathsf{E}[y_t \mid \mathscr{F}_{t-1}]$. Consequently, the log-density of the Two sided Skew and Shape distribution evaluated in y_t becomes:

$$\ell_t\left(y_t;\sigma_t^2,\boldsymbol{\vartheta}\right) \propto -2\log C(\xi,\beta,\nu) + \frac{1}{2}\log K^2(\xi,\beta,\nu) - \frac{1}{2}\log\left(\sigma^2\right) - \frac{\nu+1}{2}\log\left[1+\frac{(y_t-\lambda)^2}{\nu\psi^2\xi^2}\right]\mathbb{1}_{(-\infty,\lambda)}(y_t) - \frac{1}{2}\left[\frac{(y_t-\lambda)^2\xi^2}{\psi^2}\right]^{\frac{\beta}{2}}\mathbb{1}_{[\lambda,\infty)}(y_t), \tag{7}$$

where $C(\xi,\beta,\nu)$ and $K(\xi,\beta,\nu)$ are constants that depend upon the parameters ξ, ν and β, while λ and ψ are functions of all the parameters and also upon the time-varying parameter σ_t^2. For convenience, we collect the static parameters μ, ξ,

ν and β in the vector $\boldsymbol{\vartheta} = (\mu, \xi, \nu, \beta)' \in \mathbb{R} \times \mathbb{R}^+ \times (4, \infty) \times (1, 2)$. Furthermore, we specify $h : \mathbb{R} \to \mathbb{R}^+$ to be a mapping function, such that $\sigma_t^2 = h\left(\tilde{\sigma}_t^2\right)$, where $\tilde{\sigma}_t^2 \in \mathbb{R}$ is the reparametrisation of σ_t^2 according to the inverse mapping $h^{-1}(\cdot)$. A convenient choice for the function $h(\cdot)$ is:

$$h\left(\tilde{\sigma}_t^2\right) = \sigma_t^2 = \exp\left(\tilde{\sigma}_t^2\right). \tag{8}$$

The DCS(1, 1) dynamic for $\tilde{\sigma}_t^2$ is then:

$$\tilde{\sigma}_{t+1}^2 = \omega + \varphi \tilde{s}_t + \kappa \tilde{\sigma}_t^2, \tag{9}$$

where $\tilde{s}_t = \mathscr{J}\left(\tilde{\sigma}_t^2\right)' \left[\mathscr{H}\left(\boldsymbol{\vartheta}, \sigma_t^2\right)\right]^{-\alpha} \nabla_{\sigma^2}\left(y_t; \boldsymbol{\vartheta}, \sigma_t^2\right)$ is the score of the conditional distribution of y_t with respect to the reparametrised vector of parameters $\tilde{\sigma}_t^2$.

References

1. Bauwens, L., Laurent, S.: A new class of multivariate skew densities, with application to generalized autoregressive conditional heteroscedasticity models. J. Bus. Econ. Stat. **23**(3), 346–354 (2005)
2. Bollerslev, T.: Generalized autoregressive conditional heteroskedasticity. J. Economet. **31**(3), 307–327 (1986)
3. Cox, D.R., Gudmundsson, G., Lindgren, G., Bondesson, L., Harsaae, E., Laake, P., Juselius, K., Lauritzen, S.L.: Statistical analysis of time series: some recent developments [with discussion and reply]. Scand. J. Stat. **8**, 93–115 (1981)
4. Creal, D., Koopman, S.J., Lucas, A.: Generalized autoregressive score models with applications. J. Appl. Economet. **28**(5), 777–795 (2013)
5. Engle, R.F.: Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflation. Econometrica **50**(4), 987–1007 (1982)
6. Fernández, C., Steel, M.F.: On bayesian modeling of fat tails and skewness. J. Am. Stat. Assoc. **93**(441), 359–371 (1998)
7. Harrison, J., West, M.: Bayesian Forecasting and Dynamic Models. Springer, Berlin (1999)
8. Harvey, A.C.: Forecasting, Structural Time Series Models and the Kalman Filter. Cambridge University Press, Cambridge (1990)
9. Harvey, A.C.: Dynamic Models for Volatility and Heavy Tails: With Applications to Financial and Economic Time Series, vol. 52. Cambridge University Press, Cambridge (2013)
10. Koopman, S.J., Lucas, A., Scharth, M.: Predicting time-varying parameters with parameter-driven and observation-driven models. Rev. Econ. Stat. **98**(1), 97–110 (2015)

Sparse Networks Through Regularised Regressions

Mauro Bernardi and Michele Costola

Abstract We propose a Bayesian approach to the problem of variable selection and shrinkage in high dimensional sparse regression models where the regularisation method is an extension of a previous LASSO. The model allows us to include a large number of institutions which improves the identification of the relationship and maintains at the same time the flexibility of the univariate framework. Furthermore, we obtain a weighted directed network since the adjacency matrix is built "row by row" using for each institutions the posterior inclusion probabilities of the other institutions in the system.

Keywords Financial networks · Sparsity · Bayesian inference

1 Introduction

Models with High-dimensional data where the number of parameters is larger than the size dimension represent one of the most prominent research field in econometrics and statistics. The seminal paper of Tibshirani [3] introduced the least absolute shrinkage and selection operator (LASSO), one of the most popular method that can simultaneously perform parameters estimation and selection in regression models. Then, scholars began to develop sparse estimators in high-dimensions. Among the most important shrinkage methods proposed in the literature there are the least angle regression (LARS) of Efron et al. [1], the adaptive LASSO of Zou

M. Bernardi
Department of Statistical Sciences, University of Padova, Roma, Italy

Istituto per le Applicazioni del Calcolo "Mauro Picone" - CNR, Roma, Italy
e-mail: mauro.bernardi@unipd.it

M. Costola (✉)
Research Center SAFE, House of Finance, Goethe University Frankfurt am Main, Frankfurt, Germany
e-mail: costola@safe.uni-frankfurt.de

M. Corazza et al. (eds.), *Mathematical and Statistical Methods for Actuarial Sciences and Finance*, https://doi.org/10.1007/978-3-319-89824-7_23

[5] and the group LASSO of Yuan and Lin [4]. In this paper, we propose a Bayesian approach to the problem of variable selection and shrinkage in high dimensional causal sparse regression models where the regularisation method is an extension of a previous LASSO in a Bayesian framework. The model allows us to extend the pairwise Granger causality in the network estimation by including a large number of institutions which improves the identification of the relationship and maintains at the same time the flexibility of the univariate framework. Furthermore, we obtain a weighted directed network since the adjacency matrix is built "row by row" using for each institutions the posterior inclusion probabilities of the other institutions in the network.

2 The Model

Let $\mathbf{y} = (y_1, y_2, \ldots, y_T)'$ be the vector of observations on the scalar response variable Y, $\mathbf{X} = \left(\mathbf{x}_1', \mathbf{x}_2', \ldots, \mathbf{x}_T'\right)'$ be the $(T \times p)$ matrix of observations on the p covariates, i.e., $\mathbf{x}_{j,t} = \left(x_{j,1}, x_{j,2}, \ldots, x_{j,p}\right)$ while $\mathbf{Z} = \left(\mathbf{z}_1', \mathbf{z}_2', \ldots, \mathbf{z}_T'\right)'$ be the $(T \times q)$ matrix of observations on predetermined variables which may include the lagged values of the endogenous variable Y up to the p-th lag. We consider the following regression model

$$\pi\left(\mathbf{y} \mid \mathbf{X}, \mu, \boldsymbol{\alpha}, \boldsymbol{\beta}, \sigma_\varepsilon^2\right) = \mathsf{N}\left(\mathbf{y} \mid \boldsymbol{\iota}_T \mu + \mathbf{Z}\boldsymbol{\alpha} + \mathbf{X}\boldsymbol{\beta}, \sigma_\varepsilon^2\right), \tag{1}$$

where $\boldsymbol{\iota}_T$ is the $T \times 1$ vector of unit elements, $\mu \in \mathbb{R}$ denotes the parameter related to the intercept of the model, $\boldsymbol{\alpha} = \left(\alpha_1, \alpha_2, \ldots, \alpha_q\right)' \in \mathbb{R}^q$ and $\boldsymbol{\beta} = \left(\beta_1, \beta_2, \ldots, \beta_p\right)' \in \mathbb{R}^p$ are vectors of regression parameters and $\sigma_\varepsilon^2 \in \mathbb{R}^+$ is the scale parameter. Hereafter, we distinguish between the vector of predetermined variable $\mathbf{z}$ and that of covariates $\mathbf{X}$ because of the role they play within the context of Granger causality we consider. Specifically, in what follows, we assume that the parameters corresponding to the predetermined variable cannot be excluded from the regression while those corresponding to the covariates can also be excluded.

2.1 Spike-and-Slab EM

Using standard notation, let $\boldsymbol{\gamma}$ be the p-vector where $\gamma_j = 1$ if the j-th covariate $\mathbf{X}_j$ is included as explanatory variable in the regression model and $\gamma_j = 0$, otherwise. Assuming that $\gamma_j \sim \mathsf{Ber}(\omega)$, the prior distribution for β_j, $j = 1, 2, \ldots, p$ can be written as the mixture

$$\pi\left(\beta_j \mid \tau, \sigma_\varepsilon, \omega\right) = (1 - \omega)\, \delta_0\left(\beta_j\right) + \omega \mathsf{DE}\left(\beta_j \mid \tau, \sigma_\varepsilon\right), \tag{2}$$

where $\delta_0(\beta_j)$ is a point mass at zero and DE denotes the doubly-exponential distribution with probability density function

$$\mathsf{DE}(x \mid \tau, \sigma_\varepsilon) = \frac{\tau}{\sigma_\varepsilon} \exp\left\{-\frac{\tau|x|}{\sigma_\varepsilon}\right\} \mathbb{1}_{(-\infty,\infty)}(x), \tag{3}$$

where $\tau \in \mathbb{R}^+$ acts as the shrinkage parameter in the Lasso framework and σ_ε is the scale parameter. The regression model defined in Eq. (1) with the spike and slab ℓ_1 prior defined in Eq. (2) becomes

$$\pi\left(\mathbf{y} \mid \mathbf{X}, \mu, \boldsymbol{\alpha}, \boldsymbol{\beta}, \sigma_\varepsilon^2\right) = \mathsf{N}\left(\mathbf{y} \mid \boldsymbol{\iota}_T \mu + \mathbf{Z}\boldsymbol{\alpha} + \mathbf{X}\boldsymbol{\beta}, \sigma_\varepsilon^2\right) \tag{4}$$

$$\pi(\mu \mid \tau, \sigma_\varepsilon) = \mathsf{DE}(\mu \mid \tau, \sigma_\varepsilon) \tag{5}$$

$$\pi(\boldsymbol{\alpha} \mid \tau, \sigma_\varepsilon) = \prod_{j=1}^{q} \mathsf{DE}\left(\alpha_j \mid \tau, \sigma_\varepsilon\right) \tag{6}$$

$$\pi(\boldsymbol{\beta} \mid \tau, \sigma_\varepsilon, \omega) = \prod_{j=1}^{p} \left[(1-\omega)\,\delta_0\left(\beta_j\right) + \omega \mathsf{DE}\left(\beta_j \mid \tau, \sigma_\varepsilon\right)\right]. \tag{7}$$

It is worth noting that the prior distributions in Eqs. (5)–(7) allow the corresponding parameters to be always included into the model specification. The definition of the model is completed by the specification of the prior on the remaining parameters $\left(\sigma_\varepsilon^2, \tau, \omega\right)$. The scale parameter σ_ε and the shrinkage parameter τ, as well as the prior inclusion probability ω are parameters that have to be estimated. Common choices for the prior on those parameters are: $\sigma_\varepsilon^2 \sim \mathsf{IG}\left(\sigma_\varepsilon^2 \mid \lambda_\sigma, \eta_\sigma\right)$, $\tau \sim \mathsf{G}(\tau \mid \lambda_\tau, \eta_\tau)$ and $\omega \sim \mathsf{B}e(\omega \mid \lambda_\omega, \eta_\omega)$. where $(\lambda_\sigma, \eta_\sigma, \lambda_\tau, \eta_\tau, \lambda_\omega, \eta_\omega)$ are prior hyperparameters. Hereafter, $\boldsymbol{\vartheta} = \left(\mu, \boldsymbol{\alpha}, \boldsymbol{\beta}, \sigma_\varepsilon^2, \tau, \omega\right)$ collects all the unknown parameters that should be estimated.

The EM algorithm consists of two major steps, one for expectation (E-step) and one for maximisation (M-step), see [2]. At the $(m+1)$-th iteration the EM algorithm proceeds as follows:

(i) **E-step:** computes the conditional expectation of the complete-data log-likelihood given the observed data $\{y_t, \mathbf{z}_t, \mathbf{x}_t\}_{t=1}^T$ and the m-th iteration parameters updates $\boldsymbol{\vartheta}^{(m)}$

$$\mathscr{Q}\left(\boldsymbol{\vartheta}, \boldsymbol{\vartheta}^{(m)}\right) = \mathbb{E}_{\boldsymbol{\vartheta}^{(m)}}\left[\log \mathscr{L}_c(\boldsymbol{\vartheta}) \mid \{y_t, \mathbf{z}_t, \mathbf{x}_t\}_{t=1}^T\right]; \tag{8}$$

(ii) **M-step:** choose $\boldsymbol{\vartheta}^{(m+1)}$ by maximising (8) with respect to $\boldsymbol{\vartheta}$

$$\boldsymbol{\vartheta}^{(m+1)} = \arg\max_{\boldsymbol{\vartheta}} \mathscr{Q}\left(\boldsymbol{\vartheta}, \boldsymbol{\vartheta}^{(m)}\right). \tag{9}$$

3 Application to Network Analysis

We can define a network as a set of nodes $V_t = \{1, 2, \ldots, n_t\}$ and directed edges between nodes. The network can be represented through an n_t-dimensional adjacency matrix A_t, with the element $a_{ijt} = 1$ if there is an edge from i directed to j with $i, j \in V_t$ and 0 otherwise. The matrix A_t represents the weighted network estimated by using the proposed model where the linkages are estimated by the inclusion probability above a given threshold c,

$$A = \begin{bmatrix} 0 & p_{1,2} & \cdots & p_{1,j} & p_{1,n_t} \\ \vdots & \ddots & \cdots & \cdots & \vdots \\ \vdots & \cdots & \ddots & \cdot & \vdots \\ p_{i,1} & \cdots & \cdots & \ddots & p_{i,n_t} \\ p_{n_t,1} & \cdots & \cdots & \cdots & 0 \end{bmatrix} \tag{10}$$

The aim to the analysis is to show that our methodology avoid the over- and mis-identification of the linkages of the pairwise approach. As the reference measure for comparison, we consider the density of the network in each period d_t, defined as

$$d_t = \frac{1}{2n_t(n_t - 1)} \sum_{i=1}^{n_t} \sum_{j=1}^{n_t} a_{ijt}. \tag{11}$$

$t = 1, \ldots, T$. When $(d_t - d_{t-1}) > 0$, there is an increase of system interconnectedness.

Acknowledgements The author acknowledges financial support from the Marie Skłodowska-Curie Actions, European Union, Seventh Framework Program HORIZON 2020 under REA grant agreement n.707070. He also gratefully acknowledges research support from the Research Center SAFE, funded by the State of Hessen initiative for research LOEWE.

References

1. Efron, B., Hastie, T., Johnstone, I., Tibshirani, R.: Least angle regression. Ann. Stat. **32**(2), 407–499 (2004)
2. McLachlan, G., Krishnan, T.: The EM Algorithm and Extensions, vol. 382. Wiley, Hoboken (2007)
3. Tibshirani, R.: Regression shrinkage and selection via the lasso. J. R. Stat. Soc. B **58**, 267–288 (1996)
4. Yuan, M., Lin, Y.: Model selection and estimation in regression with grouped variables. J. R. Stat. Soc. Ser. B (Stat. Methodol.) **68**(1), 49–67 (2006)
5. Zou, H.: The adaptive lasso and its oracle properties. J. Am. Stat. Assoc. **101**(476), 1418–1429 (2006)

Approximate EM Algorithm for Sparse Estimation of Multivariate Location–Scale Mixture of Normals

Mauro Bernardi and Paola Stolfi

Abstract Parameter estimation of distributions with intractable density, such as the Elliptical Stable, often involves high-dimensional integrals requiring numerical integration or approximation. This paper introduces a novel Expectation–Maximisation algorithm for fitting such models that exploits the fast Fourier integration for computing the expectation step. As a further contribution we show that by slightly modifying the objective function, the proposed algorithm also handle sparse estimation of non-Gaussian models. The method is subsequently applied to the problem of selecting the asset within a sparse non-Gaussian portfolio optimisation framework.

Keywords Sparse estimation · Multivariate heavy-tailed distributions · Expectation maximisation · Portfolio optimisation

1 Introduction

The maximum likelihood paradigm for model-based parameter estimation requires that the likelihood function will be analytically available for any value of the parameter space. However, sometimes, finding maximum likelihood estimates appears to be a very hard computational task, either because of the model complexity or because the probability density function is not analytically available. In this paper, we consider multivariate non-Gaussian models, whilst they do not admit a closed form expression for the likelihood function, whose probability density function can be expressed as location–scale mixtures of tractable distributions.

M. Bernardi
Department of Statistical Sciences, University of Padova, Padova, Italy
e-mail: mauro.bernardi@unipd.it

P. Stolfi (✉)
Istituto per le Applicazioni del Calcolo "Mauro Picone" - CNR, Roma, Italy

Department of Economics, Roma Tre University, Rome, Italy
e-mail: paola.stolfi@uniroma3.it

M. Corazza et al. (eds.), *Mathematical and Statistical Methods for Actuarial Sciences and Finance*, https://doi.org/10.1007/978-3-319-89824-7_24

Indeed, it is often the case that the location–scale latent factor has lower dimension than that of the observed vector, thereby allowing a scalable inferential algorithm. Sub-Gaussian models, such as the Elliptical Stable and the Tempered Stable distributions, are examples where our approach can be effectively applied. We propose an Expectation–Maximisation [1] algorithm where the expectation step, i.e., the expected value of the complete-data log-likelihood with respect to the full conditional distribution of the latent factor given the observed data is computed by numerical integration. The advantages of such approach are twofold. First, it delivers a scalable EM algorithm whenever the dimension of the latent factor is lower than that of the observed variable. Second, it often provides the way of getting closed form expressions for the subsequent maximisation step.

As an additional contribution of our work, we handle sparse estimation of large-dimensional non-Gaussian models by adding the ℓ_1 lasso penalty to the M-step showing that penalised maximisation problem correspond to the graphical lasso. This is a very interesting property in large-dimensional setting because sparse estimators reduce the complexity of the model.

As regards applications to real data, we consider the well-known portfolio optimisation problem, introduced by Markowitz [3], under the Value-at-Risk constraint. The original the mean–variance approach relies on quite restrictive conditions about the underlying DGP that are relaxed here by assuming that returns follow a multivariate Stable distribution.

2 Multivariate Model Specifications

The parameter estimation method that we propose can be applied to a general random vector $\mathbf{Y} \in \mathbb{R}^m$ such that $\mathbf{Y} \sim f(\cdot, \vartheta)$ where f can be represented as a location scale mixture of skew elliptical distribution. Without loss of generality here we consider a particular family, the elliptical stable distributions (ESD), in order to make the presentation of the method as clear as possible.

A random vector $\mathbf{Y} \in \mathbb{R}^m$ is elliptically distributed if $\mathbf{Y} =^d \xi + \mathscr{R}\Gamma\mathbf{U}$ where $\xi \in \mathbb{R}^m$ is a vector of location parameters, Γ is a matrix such that $\Omega = \Gamma\Gamma'$ is a $m \times m$ full rank matrix of scale parameters, $\mathbf{U} \in \mathbb{R}^m$ is a random vector uniformly distributed in the unit sphere $\mathbb{S}^{m-1} = \left\{\mathbf{U} \in \mathbb{R}^m : \mathbf{U}'\mathbf{U} = 1\right\}$ and $\mathscr{R}$ is a non-negative random variable stochastically independent of $\mathbf{U}$, called generating variate of $\mathbf{Y}$. If $\mathscr{R} = \sqrt{Z_1}\sqrt{Z_2}$ where $Z_1 \sim \chi^2_m$ and $Z_2 \sim \mathscr{S}_{\frac{\alpha}{2}}(\xi, \omega, \delta)$ is a positive Stable distributed random variable with characteristic exponent equal to $\frac{\alpha}{2}$ for $\alpha \in (0, 2]$, location $\xi = 0$ and scale parameter $\omega = 1$, stochastically independent of χ^2_m, then the random vector $\mathbf{Y}$ has Elliptical Stable distribution, denoted as $\mathbf{Y} \sim \mathscr{ESD}_m(\alpha, \xi, \Omega)$. See [6] for more details on the positive Stable distribution and [5] for the recent developments on multivariate elliptically contoured stable distributions. Except for few cases, $\alpha = 2$ (Gaussian), $\alpha = 1$ (Cauchy) and $\alpha = \frac{1}{2}$ (Lévy), the density function cannot be represented in closed form.

3 Parameters Estimation via FFT–EM Algorithm

The multivariate $\mathbf{y} \sim \mathscr{ESD}_d(\mathbf{y}, \xi, \Omega, \alpha)$ admits the following representation:

$$p(\mathbf{y} \mid \xi, \Omega, \alpha) = C \int_{\mathbb{R}^+} \zeta^{\lambda-1} \exp\left\{-\delta_y \zeta\right\} h(\zeta, \alpha)\, d\zeta, \tag{1}$$

where $h(\zeta, \alpha)$ is the probability density function of the random variable $\zeta \sim \mathscr{S}_{\frac{\alpha}{2}}(0, \bar{\omega}_\alpha, 1)$, $\lambda = \frac{d}{2} + 1$, $\delta_y = \frac{1}{2}(\mathbf{y} - \xi)' \Omega^{-1} (\mathbf{y} - \xi)$ and $C = \frac{1}{(2\pi)^{\frac{d}{2}} |\Omega|^{\frac{1}{2}}}$. Equation (1) integrates a Gamma distribution with respect to a totally skewed Stable distribution, therefore it can be numerically evaluated using the fast Fourier transform method. Augmenting the observations $\{\mathbf{y}_t, t = 1, 2, \ldots, T\}$ with the latent variables $\{\zeta_t, t = 1, 2, \ldots, T\}$ gives the following complete-data log-likelihood:

$$\log \mathscr{L}_c(\xi) \propto -\frac{d}{2} \sum_{t=1}^{T} \log(\zeta_t) - \frac{T}{2} \log|\Omega| - \frac{1}{2} \sum_{t=1}^{T} \zeta_t \varepsilon_t' \Omega^{-1} \varepsilon_t, \tag{2}$$

where $\varepsilon_t = (\mathbf{y}_t - \xi)$ and $\Xi = (\xi, \Omega, \alpha)$ denotes the vector of parameters.

The EM algorithm consists of two major steps, one for expectation (E-step) and one for maximization (M-step), see [4]. At the $(m+1)$-th iteration the EM algorithm proceeds as follows:

E-step: at iteration $(m+1)$, the E-step requires the computation of the so-called $\mathscr{Q}$-function, which calculates the conditional expectation of the complete-data log-likelihood given the observations and the current parameter estimates $\xi^{(m)}$

$$\mathscr{Q}\left(\xi, \xi^{(m)}\right) \propto -\frac{d}{2} \sum_{t=1}^{T} \widehat{\log(\zeta_t)} - \frac{T}{2} \log|\Omega| - \frac{1}{2} \sum_{t=1}^{T} \widehat{\zeta}_t \mathrm{tr}\left(\Omega^{-1} \varepsilon_t \varepsilon_t'\right), \tag{3}$$

where the conditional expectations $\widehat{\zeta}_t^{(m)}$ and $\widehat{\log(\zeta_t)}^{(m)}$ denote the current estimate of the conditional expectation of ζ_t and $\log(\zeta_t)$ given the observation at time T, $\mathbf{y}_{1:T}$, and it is computed using the fast Fourier transform method.

M-step: at iteration $(m+1)$, the M-step maximizes the function $\mathscr{Q}\left(\xi, \xi^{(m)}\right)$ with respect to ξ to determine the next set of parameters $\xi^{(m+1)}$.

In order to introduce sparsity we prove that the $\mathscr{Q}$ function can be factorised as

$$\mathscr{Q}(\Omega, \mathbf{Y}) \propto -\frac{T}{2} \log|\Omega| - \frac{1}{2} \mathrm{tr}\left(\mathbf{S}_{\hat{\zeta}} \Omega^{-1}\right), \tag{4}$$

with $\mathbf{S}_{\hat{\zeta}} = \tilde{\mathbf{Y}}\tilde{\mathbf{Y}}'$, where $\tilde{\mathbf{Y}} = \left[\frac{\mathbf{y}_1-\xi}{\zeta_1} \; \frac{\mathbf{y}_2-\xi}{\zeta_2} \; \dots \; \frac{\mathbf{y}_T-\xi}{\zeta_T}\right]$, is the $d \times T$ matrix of observations $\frac{\mathbf{y}_t-\xi}{\zeta_t}$, $t = 1, 2, \dots, T$ stacked by row. Then by adding the ℓ_1 penalty

$$\mathscr{Q}(\Omega, \mathbf{Y}) \propto -\frac{T}{2}\log|\Omega| - \frac{1}{2}\mathrm{tr}\left(\mathbf{S}_{\hat{\zeta}}\Omega^{-1}\right) - \lambda|\Omega^{-1}|_1 \tag{5}$$

the maximisation problem in the M-step correspond to the graphical lasso, see [2].

4 Portfolio Application

We consider a portfolio allocation problem, where, at each time $t = 1, 2, \dots, T$, the investor's wealth allocation is based on the choice of the vector of optimal portfolio weights $\mathbf{w}_t > 0$ by minimising the following objective function

$$\arg\min_{\mathbf{w}_t} -\mathsf{E}_t\left(\mathbf{w}_t'\mathbf{Y}_{t+1}\right) - \kappa\mathsf{VaR}_t^{\lambda}\left(\mathbf{w}_t'\mathbf{Y}_{t+1}\right), \text{ s.t. } \mathbf{w}_t'\mathbf{1} = 1, \tag{6}$$

where $\mathbf{Y}_t \sim \mathscr{ESD}(\alpha, \xi, \Omega)$, $\mathsf{E}_t\left(\mathbf{w}_t'\mathbf{Y}_{t+1}\right)$ and $\mathsf{VaR}_t^{\lambda}\left(\mathbf{w}_t'\mathbf{Y}_{t+1}\right)$ denote the portfolio's expected return and the portfolio Value-at-Risk at level $\lambda \in (0, 1)$ evaluated at time t for the period $(t, t+1]$, respectively. Here, $\kappa \geq 0$ denotes the investor's risk aversion parameter: the larger κ, the higher is the penalisation for the risk profile of the selected portfolio. The empirical application is structured as follows. We consider a basket of weekly returns of seventeen MSCI European indexes. For each week we estimate the ESD parameters using a rolling windows; for each window, we solve Eq. (6) for the vector of optimal allocations $\mathbf{w}_t$, where the portfolio expected returns and VaR are calculated exploiting the closure property with respect of the linear combination of the ESD. The VaR confidence level is fixed at $\lambda = 0.99$ and several levels of investors' risk aversion are considered. We forecast the one-step ahead conditional returns' distribution over the whole sample period. The sequence of predictive distributions delivered by the competing models, are then used to build the mean–VaR optimal portfolios.

References

1. Dempster, A.P., Laird, N.M., Rubin, D.B.: Maximum likelihood from incomplete data using the em algorithm (with discussion). J. R. Stat. Soc. B **39**, 1–39 (1977)
2. Friedman, J., Hastie, T., Tibshirani, R.: Sparse inverse covariance estimation with the graphical lasso. Biostatistics **9**(3), 432–441 (2008)
3. Markowitz, H.M.: Portfolio selection. J. Financ. **7**, 77–91 (1952)
4. McLachlan, G., Krishnan, T.: The EM Algorithm and Extensions, vol. 382. Wiley, Hoboken (2007)
5. Nolan, J.P.: Multivariate elliptically contoured stable distributions: theory and estimation. Comput. Stat. **28**(5), 2067–2089 (2013)
6. Samorodnitsky, G., Taqqu, M.S.: Stable non-Gaussian random processes. In: Stochastic Modeling. Chapman & Hall, New York (1994). Stochastic models with infinite variance

An Extension of Multidimensional Scaling to Several Distance Matrices, and Its Application to the Italian Banking Sector

Alessandro Berti and Nicola Loperfido

Abstract Multidimensional scaling is an exploratory statistical technique which is widely used for detecting structures in multivariate data. Unfortunately, it relies on a single distance matrix. We propose an extension of multidimensional scaling to several distance matrices which is particularly useful when the latter are roughly proportional to each other. We apply the proposed method to several balance sheet ratios collected from Italian banks. The main empirical finding is that the two major banks are very different from each other and from smaller banks, which are clustered together. It contributes to the current debate on dimension as stabililizer of the banking system.

Keywords Balance sheet ratio · Bank · Multidimensional scaling

1 Introduction

Human vision is naturally apt at detecting patterns from interpoint distances. Unfortunately, this gift is limited to Euclidean distances between points on the line, in the plane and in the space. Multidimensional scaling (MDS) overcomes these difficulties by approximating the interpoint distances with Euclidean distances between numbers, pairs of numbers or triplets of numbers. It is particularly useful when it leads to a reliable graph whose axes are interpretable and where points hint for a clear structure. A detailed description of multidimensional scaling, together with its potential applications, may be found in [4].

A major limitation of MDS is its reliance on the chosen distance matrix. Different distances might lead to different MDS results. For example, some points might resemble a cluster in a MDS graph while being very distant from each other in a MDS graph based on a different distance matrix, albeit computed from the

A. Berti · N. Loperfido (✉)
Università degli Studi di Urbino "Carlo Bo", Dipartimento di Economia, Società e Politica, Urbino, PU, Italy
e-mail: alessandro.berti@uniurb.it; nicola.loperfido@uniurb.it

M. Corazza et al. (eds.), *Mathematical and Statistical Methods for Actuarial Sciences and Finance*, https://doi.org/10.1007/978-3-319-89824-7_25

same data. The problem would not arise with proportional distance matrices, since they give the same information about the data points configuration. Similarly, the problem would be eased if the distance matrices were roughly proportional to each other. We therefore propose to replace several distance matrices which are believed to satisfactorily represent distances between units with a single distance matrix which is as proportional as possible to all of them.

2 Method

Assume that distance between units are adequately summarized by the distance matrices $\Delta_1, \ldots, \Delta_k$. We shall now describe an algorithm for finding another distance matrix which is as close as possible, in the squared norm sense, to matrices proportional to the given distance matrices. We shall make extensive use of the vec operator, which vectorizes the matrix A by stacking its columns on top of each other, so as to obtain the vector $vec(A)$ (see, for example, [7, p. 200]).

First, obtain the matrix Ξ by lining side by side the vectorized distance matrices $\Delta_1, \ldots, \Delta_k$. Second, compute the first singular value λ of Ξ, and the associated right and left singular vectors v and u. Third, assess the approximation by the ratio $\lambda / \|\Xi\|$. Fourth, matricize the left singular vector u into Δ, so that u satisfies the identity $u = vec(\Delta)$. Fifth, use the matrix Δ for the MDS analysis. The following theorem shows that the algorithm leads to an optimal choice of the approximating matrix, and also that it is a valid distance matrix.

Proposition 1 *Let $\Delta_1, \ldots, \Delta_k$ be $n \times n$ distance matrices with positive off-diagonal elements and let Ξ be a matrix whose i-th column is the vectorized Δ_i, for $i = 1, \ldots, k$. Then the first right singular vector v of Ξ might be taken to be positive and the first left singular vector of Ξ is proportional to a valid, vectorized distance matrix Δ. Also, the sum $\|\Delta_1 - c_1 D\|^2 + \ldots + \|\Delta_k - c_k D\|^2$, where $c = (c_1, \ldots, c_k)^T \in \mathbb{R}^k$ and $D = D^T \in \mathbb{R}^n \times \mathbb{R}^n$ attains its minimum only when c and D are proportional to v and Δ, respectively.*

Proof We shall prove the theorem for distance matrices only, the proof for similarity and dissimilarity matrices being very similar. Let $\lambda \in \mathbb{R}_+$ and $u \in \mathbb{R}^{n^2}$ be the first singular value of Ξ and the associated left singular vector. First, consider the product $\Xi^T \Xi = c_1 c_1^T + \ldots + c_{n^2} c_{n^2}^T$, where c_i is the i-th column of Ξ^T. By assumption, $\Delta_1, \ldots, \Delta_k$ are $n \times n$ distance matrices with positive off-diagonal elements, so that all the above summands are positive matrices, but the first, the $(n+1)$-th, ..., and the n^2-th, which are null matrices. Hence $\Xi^T \Xi$ is a matrix with strictly positive elements. By the Perron-Frobenius theorem [7, p. 473] λ is strictly greater than any other singular value of Ξ, and all elements of u have the same sign, which might be taken to be positive without loss of generality. We shall now prove that the first left singular vector u of Ξ is a vectorized, valid distance matrix. First, let $\delta_{ij,h}$ be the element in the i-th row and in the j-th column of the h-th distance matrix, for

$i, j = 1, \ldots, n$ and $h = 1, \ldots, k$. Second, let $c = (c_1, \ldots, c_k)^T = \lambda^{-1} v$. Third, let $\xi_i = vec(\Delta_i)$ be the i-th column of the matrix $\Xi = \{\xi_{ij}\}$. Finally, let δ_{ij} be the element in the i-th row and in the j-th column of the matrix Δ, where $u = vec(\Delta)$. Since (λ, v, u) is the first singular triple of Ξ, we have $\Xi v = \lambda u$, that is $u = c_1\xi_1 + \ldots + c_k\xi_k$. The vector u is a conical combination of vectorized, symmetric matrices with null diagonal elements and positive off-diagonal elements. Hence the same holds true for u. Since Δ_h is a distance matrix, we have $\delta_{ij,h} \leq \delta_{iw,h} + \delta_{wj,h}$ $\xi_{ah} = \delta_{ij,h}$ that is $\xi_{ph} \leq \xi_{qh} + \xi_{rh}$, where $p = (j-1)n + i$, $q = (w-1)n + i$, $r = (j-1)n + w$. Then $u_p = c_1\xi_{p1} + \ldots + c_k\xi_{pk} \leq c_1(\xi_{q1} + \xi_{r1}) + \ldots + c_k(\xi_{qk} + \xi_{rk}) = u_q + u_r$, that is $\delta_{ij} \leq \delta_{iw} + \delta_{wj}$. Since any triple $(\delta_{ij}, \delta_{iw}, \delta_{wj})$ in Δ satisfies the triangle inequality, Δ is a valid distance matrix. The above statements remain valid if the word "distance" is replaced with either "dissimilarity" or "similarity".

3 Analysis

We shall now apply the proposed method to several 2016 balance sheet ratios collected from the twelve largest Italian banks, and interpret the results within the current debate on dimension as stabililizer of the banking system.

The 2008 world financial crisis motivated the regulator to strengthen the capital requirements in order to preserve the stability of the credit system. Political choices have been implemented in the Basel 3 agreements, several EU directives and their application on a national level, as for example the Circ.285 of 2013 of the Bank of Italy. However, the consequent incentive to concentration of the banking system raised some controversial issues in the academic literature. Beck [2], using data from 69 countries, concluded that crises are less likely in more concentrated banking systems. On the other hand, [1, 3] concluded that the quality of the bank performance is independent of its dimension.

We shall consider the 12 major Italian banks. The banks are Unicredit SpA, Intesa SanPaolo SpA, Credem, Credito Valtellinese, BPER, Credito Cooperativo Ravennate e Imolese, BCC di Roma, Cassa Rurale di Rovereto, BCC di Chianti Fiorentino, Banca popolare Valconca, Banca Popolare di Ragusa, Banca popolare di Bari.

The 2016 balance sheet ratios are: ratio of operating income to total equity and liabilities, operating income, total equity and liabilities, funding, equity, ratio of net interest margin to total equity and liabilities, net interest margin, total equity and liabilities, ratio of net commission margin to operating income, net commission margin, operating income, cost-to-income ratio, operating costs, operating income, common equity tier ratio, non-performing loans ratio, non-performing loans, total loans, non-performing loans coverage ratio, write-downs, gross impaired loans, number of employees, total equity and liabilities-to-employees ratio.

We chose to measure the distances between units with the Euclidean, Manhattan and Chess distance. The matrix obtained with the proposed method is quite

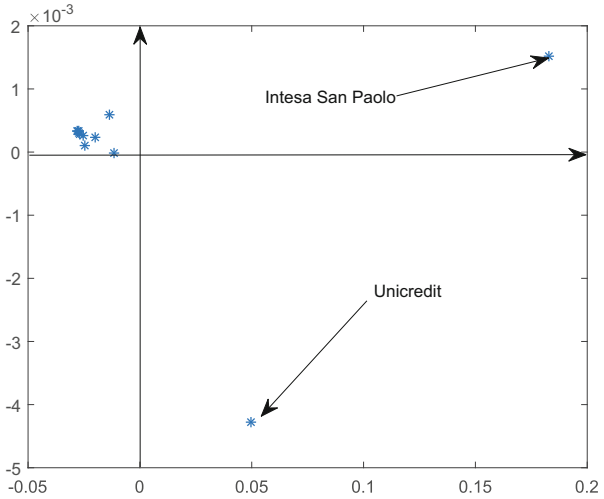


Fig. 1 Scatterplot of the first two MDS scores

proportional to the Euclidean, Manhattan and Chess distance matrices. The approximation error, defined as before, is smaller than *1%*. The first MDS score is highly correlated with bank deposits (99.5%), which is a proxy for size. The second score is highly correlated with the coverage ratio (99.8%), which is a proxy for stability. Hence, the first (second) score measures the relevance of size (stability) in explaining the difference between entities. The scatterplot in Fig. 1 clearly shows that the two major banks are very different from each other and from smaller banks, which are clustered together. More precisely, Unicredit and Intesa San Paolo are the only banks in the fourth and the first quadrant, while all other banks lie in the second quadrant. The third quadrant is empty. This empirical finding hints that large banks might have very different stability requirements. Thus, reducing the problem of financial stability to a dimensional problem might lead to oversimplification. We also used skewness-based projection pursuit to detect interesting data features, as proposed in [5, 6], obtaining similar results (not shown here due to space constraints).

References

1. Allen, F., Gale, D.: Capital adequacy regulation: in search of a rationale. In: Arnott, R., Greenwald, B., Kanbur, R., Nalebuff, B. (eds.) Economics for an Imperfect World: Essays in Honor of Joseph E. Stiglitz, pp. 83–109. MIT Press, Cambridge (2003)
2. Beck, T., Demirgüç-Kunt, A., Levine, R.: Bank concentration, competition, and crises: first results. J. Bank. Financ. **30**, 1581–1603 (2006)
3. Boyd, J., De Nicolò, G., Al Jalal, A.: Bank risk taking and competition revisited: new theory and new evidence. IMF Working Paper No. 06/297 (2007)

4. Izenman, A.J.: Modern Multivariate Statistical Techniques: Regression, Classification, and Manifold Learning. Springer, New York (2008)
5. Loperfido, N.: Skewness and the linear discriminant function. Stat. Prob. Lett. **83**, 93–99 (2013)
6. Loperfido, N.: Skewness-based projection pursuit: a computational approach. Comput. Stat. Data Anal. **120**, 42–57 (2018)
7. Rao, C.R., Rao, M.B.: Matrix Algebra and its Applications to Statistics and Econometrics. World Scientific, Singapore (1998)

Disagreement in Signed Financial Networks

Monica Billio, Roberto Casarin, Michele Costola, and Lorenzo Frattarolo

Abstract We extend the study of rate of convergence to consensus of autonomous agents on an interaction network. In particular, we introduce antagonistic interactions and thus a signed network. This will allow to include the, previously discarded, sign information, in the analysis of disagreement on statistical financial networks.

Keywords Consensus dynamics · Financial networks · Predictability · Connectedness

1 Introduction

Given the threat to financial stability and the real economy, quantifying systemic risk is now investigated by scholars as well as policy makers. More recently, graph theoretic measures and in particular convergence of autonomous agents on the network to a consensus have been involved in the systemic risk measurement such as early warning indicator for banking crises [1]. In this paper, we propose a generalization of disagreement index introduced in [1]. Building on the lifting approach in [6], we are able to apply results in [1] to the lifted dynamics extending their scope to signed networks. This allows to consider a more general consensus dynamics and disagreement with antagonism.

M. Billio · R. Casarin · L. Frattarolo (✉)
Department of Economics, Ca' Foscari University of Venice, Venice, Italy
e-mail: billio@unive.it; r.casarin@unive.it; lorenzo.frattarolo@unive.it

M. Costola
Research Center SAFE, House of Finance, Goethe University, Frankfurt am Main, Germany
e-mail: costola@safe.uni-frankfurt.de

M. Corazza et al. (eds.), *Mathematical and Statistical Methods for Actuarial Sciences and Finance*, https://doi.org/10.1007/978-3-319-89824-7_26

2 Disagreement in Unsigned Directed Network

The relationship between the eigenvalues of the directed Laplacian (Diplacian), introduced in [7], and the rate of convergence of autonomous agents on the network to a consensus was studied in [1]. The application of those techniques to financial networks could be understood as measuring persistence of disagreement that is "magnified when major events occur in financial markets" according to [2]. Our approach considers a limited communication network [8] among agents, approximated by statistical causality relationship between stocks returns as already done in [1].

Consider a graph with adjacency matrix A and elements a_{ij} and with out-degree diagonal matrix D with non zero elements $d_i^{out} = \sum_{j=1}^{n} a_{ij}$. We introduce the following discrete time autonomous agent system:

$$x_{it} = x_{it-1} + \frac{1}{2d_i^{out}} \sum_{j=1}^{n} a_{ij} \left(x_{jt-1} - x_{it-1} \right) \tag{1}$$

Similar systems, introduced by [4], are building blocks in models of belief evolution of bounded rational agents, with a persuasion bias [5].

The model written has the vectorial form :

$$\mathbf{x}_t = \frac{1}{2} I_n + \frac{1}{2} \left(I_n - D^{-1} (D - A) \right) \mathbf{x}_{t-1} = P_L x_{t-1}$$

Where $\mathbf{x}_t = (x_{it}, \ldots, x_{nt})$ is the state vector of the agents, $P = D^{-1}A$ is the transition probability matrix of the Markov chain associated with random walks on G, where the probability of transitioning from vertex i to vertex j, $p_{ij} = a_{ij}/d_i^{out}$ of a random walk starting at i and $P_L = (I_n - P)$ corresponds to the transition matrix of the lazy random walk introduced in [3].

If the graph is strongly connected, P_L is irreducible and aperiodic, according well known results, the system converge to a consensus with group decision value $\boldsymbol{\varphi}' \mathbf{x}_0$. The group decision is conserved by the dynamics:

$$\boldsymbol{\varphi}' \mathbf{x}_t = \boldsymbol{\varphi}' P_L \mathbf{x}_{t-1} \boldsymbol{\varphi}' \mathbf{x}_{t-1} = \alpha.$$

We define the disagreement vector and its law of motion

$$\boldsymbol{\xi}_t = \mathbf{x}_t - \alpha \mathbf{1}$$
$$\boldsymbol{\xi}_t = P_L \boldsymbol{\xi}_{t-1}$$

The disagreement dynamics allows us to study the convergence rate in this directed unsigned case to this decision value. We exploit the theoretical results on lazy random walks on strongly connected directed graphs due to [3] and [7]. In particular

in [7] the Diplacian Γ and its decomposition of in symmetric and asymmetric part is introduced

$$\Gamma = \varphi^{1/2}\,(I - P)\,\varphi^{-1/2},\ \Gamma = L + \Delta,\ L = \frac{\Gamma + \Gamma'}{2},\ \Delta = \frac{\Gamma - \Gamma'}{2}.$$

According to theorem 3 in [1] speed of convergence is expressed in terms of λ_2 the second smallest eigenvalue of L and of the second largest singular value $\sigma_{n-1}\,(I_n - L)$ of $I_n - L$ and the largest singular value $\sigma_n\,(\Delta)$ of the skew-symmetric part of the diplacian Δ, as in the following equation:

$$\|\boldsymbol{\xi_t}\| \leq \exp\left\{\frac{1}{2}\left[\log\left(\frac{\max(\boldsymbol{\varphi})}{\min(\boldsymbol{\varphi})}\right) + \log(\mu)\,t\right]\right\}\|\boldsymbol{\xi_0}\|$$

$$\mu = \frac{3}{4} - \frac{\lambda_2}{2} + \frac{(\sigma_{n-1}\,(I_n - L) + \sigma_n\,(\Delta))^2}{4}.$$

3 The Lifted Dynamics and Disagreement in Signed Networks

In the previous section the a_{ij}'s were non negative. In view of an application to a network, based on a vector autoregression (VAR) or similar methodology, this could result in neglecting an important source of information coming from the coefficients sign. The signed framework that correspond to antagonistic interactions among agents can be transformed in the unsigned one by a clever lifting trick introduced in [6]. Consider a signed network $a_{ij} \in \mathbb{R}$ and define $b_{ij} = \max\left(0, a_{ij}\right)$, $c_{ij} = \max\left(0, -a_{ij}\right)$, $d_i^{|out|} = \sum_{j=1}^{n} \left|a_{ij}\right|$. We study the following dynamics with antagonistic interactions

$$\begin{aligned} x_{it} &= x_{it-1} + \frac{1}{2d_i^{|out|}} \sum_{j=1}^{n} \left|a_{ij}\right| \left(\text{sign}\left(a_{ij}\right) x_{jt-1} - x_{it-1}\right) \\ &= x_{it-1} + \frac{1}{2d_i^{|out|}} \sum_{j=1}^{n} b_{ij} \left(x_{jt-1} - x_{it-1}\right) + \frac{1}{2d_i^{|out|}} \sum_{j=1}^{n} c_{ij} \left(-x_{jt-1} - x_{it-1}\right) \end{aligned}$$

or

$$x_{it} - x_{it-1} = \frac{1}{2d_i^{|out|}} \sum_{j=1}^{n} b_{ij} \left(x_{jt-1} - x_{it-1}\right) + \frac{1}{2d_i^{|out|}} \sum_{j=1}^{n} c_{ij} \left(-x_{jt-1} - x_{it-1}\right)$$

If we call $y_{it} = -x_{it}$ we obtain an analogous law of motion for y_{it}

$$y_{it} - y_{it-1} = \frac{1}{2d_i^{|out|}} \sum_{j=1}^{n} c_{ij} \left(y_{jt-1} - y_{it-1} \right) + \frac{1}{2d_i^{|out|}} \sum_{j=1}^{n} b_{ij} \left(-y_{jt-1} - y_{it-1} \right)$$

The joint dynamics can be written

$$\begin{bmatrix} \mathbf{x}_t \\ \mathbf{y}_t \end{bmatrix} = \frac{1}{2} \begin{bmatrix} I_n & \mathbf{0} \\ \mathbf{0} & I_n \end{bmatrix} - \frac{1}{2} \begin{bmatrix} \left(D^{|out|} \right)^{-1} & \mathbf{0} \\ \mathbf{0} & \left(D^{|out|} \right)^{-1} \end{bmatrix} \begin{bmatrix} B & C \\ C & B \end{bmatrix} \begin{bmatrix} \mathbf{x}_{t-1} \\ \mathbf{y}_{t-1} \end{bmatrix} = P_L \begin{bmatrix} \mathbf{x}_{t-1} \\ \mathbf{y}_{t-1} \end{bmatrix}$$

Analogously to [6] $\mathbf{x}'_t$ is a solution of (2) if and only if $\mathbf{z}_t$ is a solution of the "classical" discrete time consensus system (2). If the solution exist then applying the same methodology as the unsigned case, by defining the lifted transition probability

$$P = \begin{bmatrix} \left(D^{|out|} \right)^{-1} & \mathbf{0} \\ \mathbf{0} & \left(D^{|out|} \right)^{-1} \end{bmatrix} \begin{bmatrix} B & C \\ C & B \end{bmatrix}, \tag{2}$$

the corresponding decision vector and Laplacians we can use (2) to bound the speed of convergence of the lifted dynamics and thus a upper bound for consensus dynamics on a signed directed network.

As discussed for the unsigned case in [1], those results could be readily applied to build a disagreement index that includes the sign information and understand its role in the measurement of systemic risk.

Acknowledgment The author acknowledges financial support from the Marie Skłodowska-Curie Actions, European Union, Seventh Framework Program HORIZON 2020 under REA grant agreement n.707070.

References

1. Billio, M., Casarin, R., Costola, M., Frattarolo, L.: Contagion Dynamics on Financial Networks. Mimeo, New York (2017)
2. Carlin, B.I., Longstaff, F.A., Matoba, K.: Disagreement and asset prices. J. Financ. Econ. **114**(2), 226–238 (2014)
3. Chung, F.: Laplacians and the cheeger inequality for directed graphs. Ann. Comb. **9**(1), 1–19 (2005)
4. DeGroot, M.H.: Reaching a consensus. J. Am. Stat. Assoc. **69**(345), 118–121 (1974)
5. DeMarzo, P.M., Vayanos, D., Zwiebel, J.: Persuasion bias, social influence, and unidimensional opinions. Q. J. Econ. **118**(3), 909–968 (2003)
6. Hendrickx, J.M.: A lifting approach to models of opinion dynamics with antagonisms. In: IEEE 53rd Annual Conference on Decision and Control (CDC), 2014, pp. 2118–2123. IEEE, Piscataway (2014)
7. Li, Y., Zhang, Z.-L.: Digraph laplacian and the degree of asymmetry. Internet Math. **8**(4), 381–401 (2012)
8. Parikh, R., Krasucki, P.: Communication, consensus, and knowledge. J. Econ. Theory **52**(1), 178–189 (1990)

Bayesian Tensor Binary Regression

Monica Billio, Roberto Casarin, and Matteo Iacopini

Abstract In this paper we present a binary regression model with tensor coefficients and present a Bayesian model for inference, able to recover different levels of sparsity of the tensor coefficient. We exploit the CONDECOMP/PARAFAC (CP) representation for the tensor of coefficients in order to reduce the number of parameters and adopt a suitable hierarchical shrinkage prior for inducing sparsity. We propose a MCMC procedure with data augmentation for carrying out the estimation and test the performance of the sampler in small simulated examples.

Keywords Tensor binary regression · Sparsity · Bayesian inference · Binary matrices · Hierarchical shrinkage prior

1 Bayesian Markov Switching Binary Tensor Regression Model

Define a tensor as a generalisation of a matrix into a D-dimensional space, for a remarkable survey on this subject, see [4]. We model each entry of an observed binary matrix $\mathbf{X}_t$ via a zero-inflated logit (see [3]), using a vector of common covariates $\mathbf{z}_t$ as regressors (see [1] for greater details):

$$x_{ij,t}|\rho_t, \mathbf{g}_{ij}(t) \sim \rho\delta_{\{0\}}(x_{ij,t}) + (1-\rho)\delta_{\{d_{ij,t}\}}(x_{ij,t}) \tag{1}$$

$$d_{ij,t} = \mathbb{1}_{\mathbb{R}_+}(x^*_{ij,t}) \tag{2}$$

M. Billio · R. Casarin
Department of Economics, Ca' Foscari University of Venice, Venice, Italy
e-mail: billio@unive.it; r.casarin@unive.it

M. Iacopini (✉)
Department of Economics, Ca' Foscari University of Venice, Venice, Italy

Université Paris 1 - Panthéon-Sorbonne, Paris, France
e-mail: matteo.iacopini@unive.it

M. Corazza et al. (eds.), *Mathematical and Statistical Methods for Actuarial Sciences and Finance*, https://doi.org/10.1007/978-3-319-89824-7_27

$$x^*_{ij,t} = \mathbf{z}'_t \mathbf{g}_{ij} + \varepsilon_{ij,t} \qquad \varepsilon_{ij,t} \overset{iid}{\sim} \text{Logistic}(0, 1)\,. \tag{3}$$

By collecting together all the vectors $\mathbf{g}_{ij}$ in a third order tensor $\mathscr{G} \in \mathbb{R}^{I \times J \times Q}$ we can rewrite Eq. (3) as:

$$\mathbf{X}^*_t = \mathscr{G} \times_4 \mathbf{z}_t + \mathscr{E}_t\,. \tag{4}$$

The symbol $\times_n$ stands for the mode-n product between a tensor and a vector, as defined in [4]. Notice that this model admits a representation as a factor model and a SUR model (see [1]). In order to provide a significant reduction of the number of parameters, we assume a CONDECOM/PARAFAC (CP) decomposition of the tensor of coefficients (more details in [4] and [1]), as follows:

$$\mathscr{G} = \sum_{r=1}^{R} \mathscr{G}^{(r)} = \sum_{r=1}^{R} \boldsymbol{\gamma}_1^{(r)} \circ \boldsymbol{\gamma}_2^{(r)} \circ \boldsymbol{\gamma}_3^{(r)}\,, \tag{5}$$

where the vectors $\boldsymbol{\gamma}_j^{(r)} \in \mathbb{R}^{d_j}$, $j = 1, 2, 3$ are called marginals of the CP representation and R is the CP-rank of the tensor. We assume R known and fixed. This decomposition permits a significant reduction of the number of parameters of the coefficient tensor (from exponential to linear growth in terms of the size of the matrix $\mathbf{X}^*$).

2 Bayesian Inference

We follow the Bayesian approach for inference and the CP representation for the coefficient tensor allows to reduce the problem of specifying a prior distribution on a multi-dimensional tensor, for which few possibilities are available in the literature, to the standard multivariate case. We assume a Beta prior for the mixing probability and a hierarchical global-local shrinkage prior on the marginals of the PARAFAC decomposition (similarly to [2]):

$$\pi(\rho) \sim \mathscr{B}e(a_\rho, b_\rho) \tag{6}$$

$$\pi(\boldsymbol{\gamma}_j^{(r)} | \mathbf{W}, \boldsymbol{\phi}, \tau) \sim \mathscr{N}_{d_j}(\mathbf{0}, \tau \phi_r w_{j,r} \mathbf{I}_{d_j}) \qquad \forall r \quad \forall j \tag{7}$$

$$\pi(w_{j,r} | \lambda) \sim \mathscr{E}xp(\lambda^2/2) \quad \forall j \quad \forall r \tag{8}$$

$$\pi(\tau) \sim \mathscr{G}a(a_\tau, b_\tau) \quad \pi(\boldsymbol{\phi}) \sim \mathscr{D}ir(\boldsymbol{\alpha}_\phi) \quad \pi(\lambda) \sim \mathscr{G}a(a_\lambda, b_\lambda)\,. \tag{9}$$

We use the Pólya-Gamma data augmentation (see [5]) to derive the complete data likelihood of the model is given in the following, while the details of the Gibbs

sampler along with the analytical derivation of the full conditionals are given in [1].

$$L(\mathbf{X}, \mathbf{D}, \boldsymbol{\Omega}|\boldsymbol{\theta}) = \prod_{t=1}^{T}\prod_{i=1}^{I}\prod_{j=1}^{J} \rho^{d_{ij,t}} \cdot \delta_{\{0\}}(x_{ij,t})^{d_{ij,t}} \left(\frac{1-\rho}{2}\right)^{1-d_{ij,t}} \tag{10}$$

$$\cdot \exp\left\{-\frac{\omega_{ij,t}}{2}(\mathbf{z}_t'\mathbf{g}_{ij})^2 + \kappa_{ij,t}(\mathbf{z}_t'\mathbf{g}_{ij})\right\} \cdot p(\omega_{ij,t}) \,.$$

3 Simulation and Application

We simulated a dataset of $T = 100$ couples $\{\mathbf{X}_t, \mathbf{z}_t\}_t$ of size $I = J = 20$ and $Q = 3$, respectively. All marginals have been simulated from their prior, while we have chosen a mixing probability $\rho = 0.8$.

We initialised the marginals of each tensor $\mathscr{G}$ via simulated annealing and run the MCMC algorithm for $N = 2000$ iterations. As an indicator of the goodness of fit of the estimated parameters, we computed the Frobenious norm between the original tensor and the one reconstructed via the posterior of the marginals. The outcome is shown in Fig.1, which reports the posterior distribution and the trace plot. The matricised version of the absolute distance between the estimated and the true tensor (average over iterations) is plotted in Fig. 2. Finally, Fig. 3 shows the posterior distribution and trace plot for the mixing probability. Overall, the estimation procedure performs well in recovering the true value of the main parameters of the model within a reasonable computing time. Mayor details are reported in [1].

We apply the model to US data on land air temperature anomalies, registered for $T = 120$ months on a regularly spaced grid of size 61×22, assuming $L = 2$. We use a constant and the global temperature as covariates. The results are reported

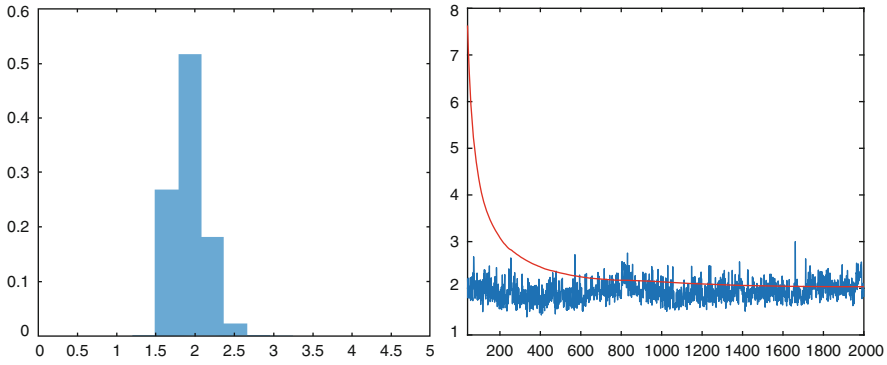


Fig. 1 Quadratic norm of the estimated tensor of coefficients: posterior distribution (*left*) and trace plot (*right*)

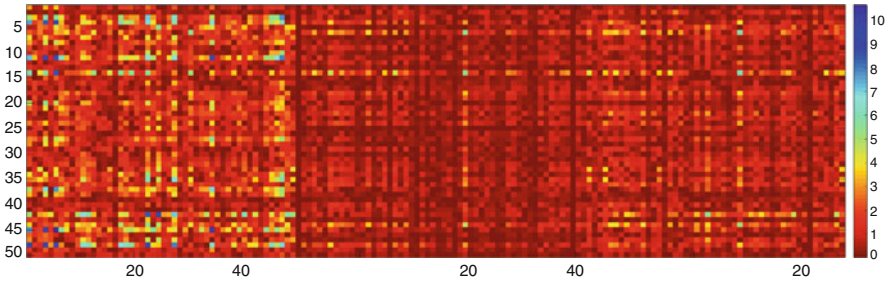


Fig. 2 Mean absolute distance between estimated and true tensor, in matricised representation

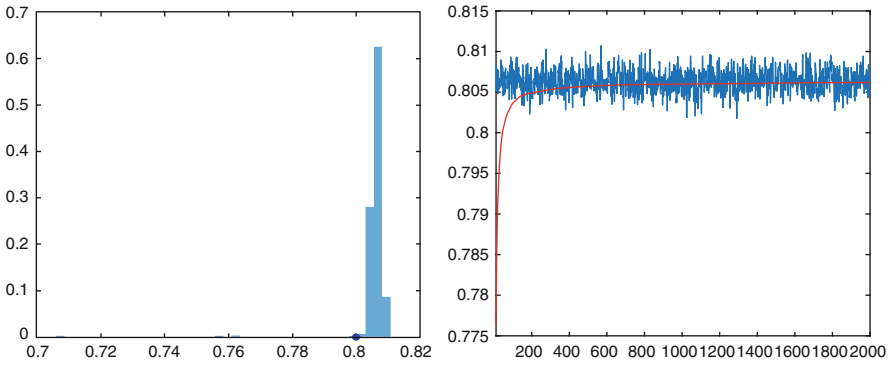


Fig. 3 Mixing probability: posterior distribution (*left*) and trace plot (*right*)

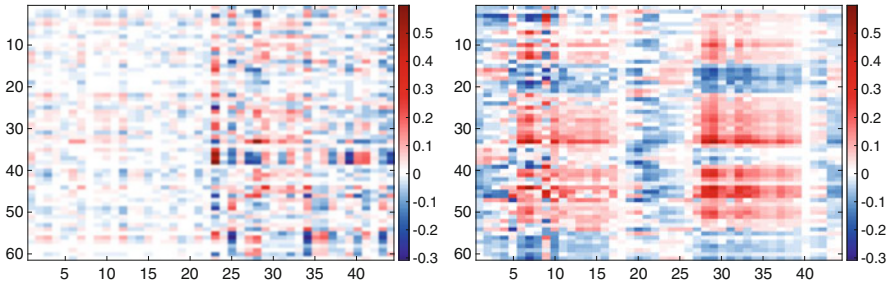


Fig. 4 Estimated coefficient tensor in state 1 (sparse, *left*) and state 2 (dense, *right*)

in Fig. 4, which shows the estimated coefficient tensor in matricized form, for each regime. We find substantial evidence of different effects of the covariates on the probability of an anomaly between the two regimes. In the sparse regime global temperature has heterogeneous effects, whereas in the dense regime both covariates have significant impact with geographical regularities.

4 Conclusions

We presented a statistical framework for modelling of a time series of binary matrices, where the parameters are grouped together into a third order tensor. We specified a zero-inflated logit model for the probability of each entry and adopted the PARAFAC decomposition on the tensor of logit regression coefficients for parsimony. Efficient Bayesian estimation via Gibbs sampler with multiple data augmentation steps has been carried out, showing good performance on small simulated datasets and on real data.

Acknowledgements This research has benefited from the use of the Scientific Computation System of Ca' Foscari University of Venice (SCSCF) for the computational for the implementation of the inferential procedure.

References

1. Billio, M., Casarin, R., Iacopini, M.: Bayesian markov switching tensor regression for time-varying networks (2017). arXiv preprint. arXiv:1711.00097
2. Guhaniyogi, R., Qamar, S., Dunson, D.B.: Bayesian tensor regression. J. Mach. Learn. Res. **18**, 1–31 (2017)
3. Harris, M., Zhao, X.: A zero-inflated ordered probit model, with an application to modelling tobacco consumption. J. Economet. **141**, 1073–1099 (2007)
4. Kolda, T.G., Bader, B.W.: Tensor decompositions and applications. SIAM Rev. **51**, 455–500 (2009)
5. Polson, N., Scott, J., Windle, J.: Bayesian Inference for logistic models using Polya–Gamma latent variables. J. Am. Stat. Assoc. **108**, 1339–1349 (2013)

Bayesian Tensor Regression Models

Monica Billio, Roberto Casarin, and Matteo Iacopini

Abstract In this paper we introduce the literature on regression models with tensor variables and present a Bayesian linear model for inference, under the assumption of sparsity of the tensor coefficient. We exploit the CONDECOMP/PARAFAC (CP) representation for the tensor of coefficients in order to reduce the number of parameters and adopt a suitable hierarchical shrinkage prior for inducing sparsity. We propose a MCMC procedure via Gibbs sampler for carrying out the estimation, discussing the issues related to the initialisation of the vectors of parameters involved in the CP representation.

Keywords Tensor regression · Sparsity · Bayesian inference · Hierarchical shrinkage prior

1 Bayesian Tensor Regression Model

Define a tensor as a generalisation of a matrix into a D-dimensional space, namely: $\mathscr{X} \in \mathbb{R}^{d_1 \times \ldots \times d_D}$, where D is the order of the tensor and d_j is the length of dimension j. Matrices, vectors and scalars are particular cases of tensor variables, of order 2, 1 and 0, respectively. The common operations defined on matrices and vectors in linear algebra can be applied also to tensors via generalisations of their definition. For a remarkable survey on this subject, see [3].

The general tensor linear regression model (see [1] for greater details) we present here can manage covariates and response variables in the form of vectors, matrices

M. Billio · R. Casarin
Department of Economics, Ca' Foscari University of Venice, Venice, Italy
e-mail: billio@unive.it; r.casarin@unive.it

M. Iacopini (✉)
Department of Economics, Ca' Foscari University of Venice, Venice, Italy

Université Paris 1 Panthéon-Sorbonne, Paris, France
e-mail: matteo.iacopini@unive.it

M. Corazza et al. (eds.), *Mathematical and Statistical Methods for Actuarial Sciences and Finance*, https://doi.org/10.1007/978-3-319-89824-7_28

or tensors. It is given by:

$$\mathscr{Y}_t = \mathscr{A} + \mathscr{B} \times_{D+1} \text{vec}\,(\mathscr{X}_t) + \mathscr{C} \times_{D+1} \mathbf{z}_t + \mathscr{D} \times_n W_t + \mathscr{E}_t,\ \mathscr{E}_t \overset{iid}{\sim} \mathscr{N}_{d_1,\ldots,d_D}(0, \Sigma_1, \ldots, \Sigma_D) \tag{1}$$

where the tensor response and errors are given by $\mathscr{Y}_t,\ \mathscr{E}_t \in \mathbb{R}^{d_1 \times \ldots \times d_D}$; while the covariates are $\mathscr{X}_t \in \mathbb{R}^{d_1^X \times \ldots \times d_M^X}$, $W_t \in \mathbb{R}^{d_n \times d_2^W}$ and $\mathbf{z}_t \in \mathbb{R}^{d_z}$. The coefficients are: $\mathscr{A} \in \mathbb{R}^{d_1 \times \ldots \times d_D}$, $\mathscr{B} \in \mathbb{R}^{d_1 \times \ldots \times d_D \times p}$, $\mathscr{C} \in \mathbb{R}^{d_1 \times \ldots \times d_D \times d_z}$, $\mathscr{D} \in \mathbb{R}^{d_1 \times \ldots \times d_{n-1} \times d_2^W \times d_{n+1} \ldots \times d_D}$ where $p = \prod_i d_i^X$. The symbol $\times_n$ stands for the mode-n product between a tensor and a vector, as defined in [3]. This model extends several well-known econometric linear models, among which univariate and multivariate regression, VAR, SUR and Panel VAR models and matrix regression model (see [1] for formal proofs).

We focus on the particular case where both the regressor and the response variables are square matrices of size $k \times k$ and the error term is assumed to be distributed according to a matrix normal distribution:

$$Y_t = \mathscr{B} \times_3 \text{vec}\,(X_t) + E_t \qquad E_t \overset{iid}{\sim} \mathscr{N}_{k,k}(\mathbf{0}, \Sigma_c, \Sigma_r). \tag{2}$$

To significantly reduce the number of parameters we assume a CONDECOM/PARAFAC (CP) representation (more details in [3]) for the tensor. Let the vectors $\boldsymbol{\beta}_j^{(r)} \in \mathbb{R}^{d_j}$, $j = 1, \ldots, D$, also called marginals of the CP representation, and R be the CP-rank of the tensor (assumed known and constant), then:

$$\mathscr{B} = \sum_{r=1}^{R} \mathscr{B}^{(r)} = \sum_{r=1}^{R} \boldsymbol{\beta}_1^{(r)} \circ \ldots \circ \boldsymbol{\beta}_D^{(r)}, \tag{3}$$

2 Bayesian Inference

We follow the Bayesian approach for inference, thus we need to specify a prior distribution for all the parameters of the model. The adoption of the CP representation for the tensor of coefficients is crucial from this point of view, as it allows to reduce the problem of specifying a prior distribution on a multi-dimensional tensor, for which very few possibilities are available in the literature, to the standard multivariate case. Building from [2], we define a prior for each of the CP marginals of the tensor coefficient $\mathscr{B}$ by means of the following hierarchy:

$$\pi(\boldsymbol{\beta}_j^{(r)} | \mathbf{W}, \boldsymbol{\phi}, \tau) \sim \mathscr{N}_{d_j}(\mathbf{0}, \tau \phi_r W_{j,r}) \qquad \forall r \quad \forall j \tag{4}$$

$$\pi(w_{p,j,r}) \sim \mathscr{E}xp(\lambda_{j,r}^2/2) \quad \forall r \quad \forall j \quad \forall p \tag{5}$$

$$\pi(\tau) \sim \mathscr{G}a(a_\tau, b_\tau) \quad \pi(\boldsymbol{\phi}) \sim \mathscr{D}ir(\boldsymbol{\alpha}_\phi) \quad \pi(\lambda_l) \sim \mathscr{G}a(a_\lambda, b_\lambda) \quad \forall j \quad \forall r\,. \tag{6}$$

We complete the prior specification by assuming two Inverse Wishart prior for the covariance matrix of the rows of error term (fixing the other to the identity for identification). Given a sample $(\mathbf{Y}, \mathbf{X}) = \{Y_t, X_t\}_{t=1}^{T}$ and defining $\mathbf{x}_t \coloneqq \text{vec}(X_t)$, the likelihood function of the model (2) is given by (see [1] for details of the Gibbs sampler):

$$L(\mathbf{Y}, \mathbf{X}|\boldsymbol{\theta}) = \prod_{t=1}^{T} (2\pi)^{-\frac{k^2}{2}} |\Sigma_2|^{-\frac{k}{2}} |\Sigma_1|^{-\frac{k}{2}} \tag{7}$$
$$\times \exp\left\{-\frac{1}{2}\Sigma_c^{-1}(Y_t - \mathscr{B} \times_3 \mathbf{x}_t)' \Sigma_1^{-1}(Y_t - \mathscr{B} \times_3 \mathbf{x}_t)\right\}.$$

3 Simulation and Application

We performed a stimulation study by drawing a sample of $T = 100$ couples $\{\mathbf{Y}_t, \mathbf{X}_t\}_t$ of square matrices of size 10. The regressor is built by entry-wise independent AR(1) processes and the errors have been assumed normal with unitary variance.

We initialised the marginals of the tensor $\mathscr{B}$ by simulated annealing and run the Gibbs sampler for $N = 30{,}000$ iterations. Figure 1 shows the estimated coefficient tensor and the absolute value of the estimation error, in matricized form. Figures 2 and 3, respectively, plot the L_2 norm between the original and the estimated tensor and the mean of the vectorised version of the estimated tensor, as indicators of the

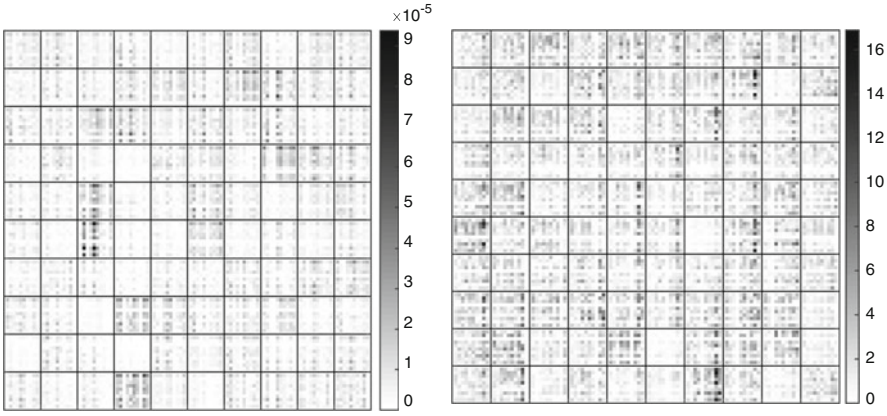


Fig. 1 *Left:* estimated tensor, averaged over all iterations (stacked representation). *Right:* estimation error (absolute value), averaged over all iterations (stacked representation)

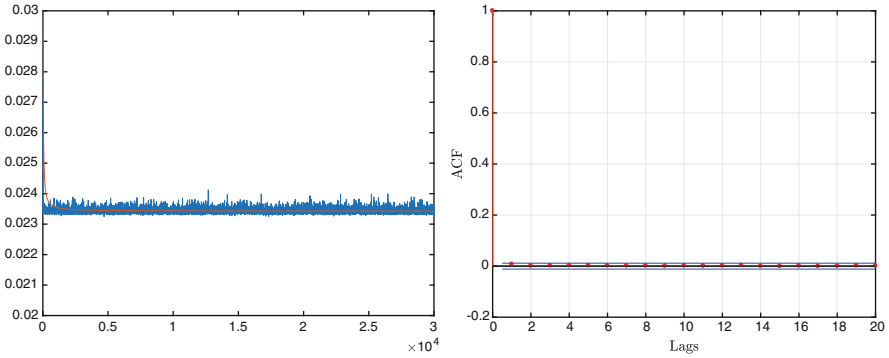


Fig. 2 *Left:* trace plot of the L_2 norm of the difference between the true and the estimated tensor of coefficients. *Right:* corresponding autocorrelation function

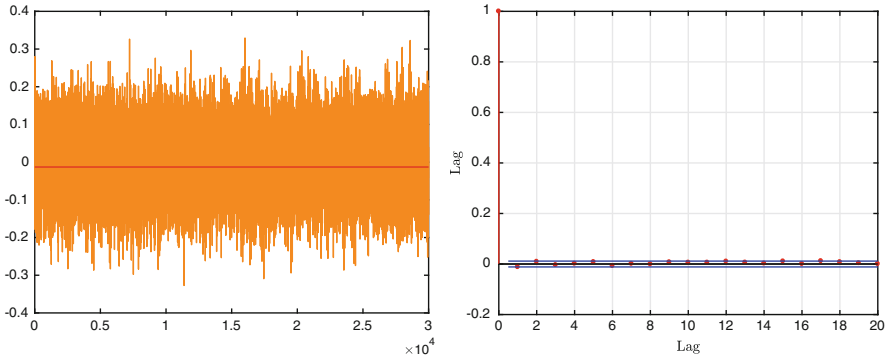


Fig. 3 *Left:* trace plot of the mean of the vectorised tensor of regression coefficients $\text{vec}(\hat{\mathscr{B}})$ (orange) and true value (red). *Right:* corresponding autocorrelation function

goodness of fit. The plots suggest good performance of the proposed sampler in recovering the true value of the parameter. See [1] for more details on the results.

We apply the model to the study of $T = 12$ yearly average temperatures, registered on a 10×10 regular grid (latitude-longitude) in US, with $X_t = Y_{t-1}$. The results are in Fig. 4. The estimated coefficient tensor (10^4 entries) is rather sparse, with some regular patterns indicating that the temperature depends on its past and on the neighbouring locations. The covariance matrices indicate that there is higher dependence along the longitudinal axis than along the latitude.

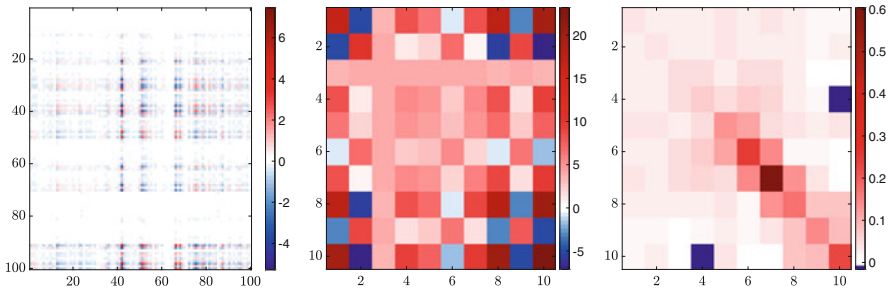


Fig. 4 Mode-3 matricization of the estimated coefficient tensor (*left*); estimated error covariance matrices: Σ_1 (*left*) and Σ_2 (*right*)

4 Conclusions

We propose a matrix linear regression model (a reduced form of a tensor regression) which extends standard econometric models and allows each entry of the covariate to exert a different effect on each entry of the response. The initialisation has been carried out via an efficient implementation of simulated annealing. The accuracy of the model has been successfully tested both on a synthetic and a real dataset, allowing the efficient estimate of large coefficient tensor.

Acknowledgements This research has benefited from the use of the Scientific Computation System of Ca' Foscari University of Venice (SCSCF) for the computational for the implementation of the inferential procedure.

References

1. Billio, M., Casarin, R., Iacopini, M.: Bayesian Dynamic Tensor Regression, arXiv preprint arXiv:1709.09606 (2017)
2. Guhaniyogi, R., Qamar, S., Dunson, D.B.: Bayesian tensor regression. J. Mach. Learn. Res. **18**, 1–31 (2017)
3. Kolda, T.G., Bader, B.W.: Tensor decompositions and applications. SIAM Rev. **51**, 455–500 (2009)

Bayesian Nonparametric Sparse Vector Autoregressive Models

Monica Billio, Roberto Casarin, and Luca Rossini

Abstract Seemingly unrelated regression (SUR) models are useful in studying the interactions among economic variables. In a high dimensional setting, these models require a large number of parameters to be estimated and suffer of inferential problems. To avoid overparametrization issues, we propose a hierarchical Dirichlet process prior (DPP) for SUR models, which allows shrinkage of coefficients toward multiple locations. We propose a two-stage hierarchical prior distribution, where the first stage of the hierarchy consists in a lasso conditionally independent prior of the Normal-Gamma family for the coefficients. The second stage is given by a random mixture distribution, which allows for parameter parsimony through two components: the first is a random Dirac point-mass distribution, which induces sparsity in the coefficients; the second is a DPP, which allows for clustering of the coefficients.

Keywords Bayesian nonparametrics · Bayesian model selection · Shrinkage · Large vector autoregression

1 Introduction

In the last decade, high dimensional models and large datasets have increased their importance in economics (e.g., see [8]). The use of large dataset has been proved to improve the forecasts in large macroeconomic and financial models (see, [1, 3, 5, 9]). For analyzing and better forecasting them, SUR/VAR models have been introduced [11, 12], where the error terms are independent across time, but may have cross-equation contemporaneous correlations. SUR/VAR models require

M. Billio · R. Casarin
University Ca' Foscari of Venice, Venice, Italy
e-mail: billio@unive.it; r.casarin@unive.it

L. Rossini (✉)
Free University of Bozen, Bolzano, Italy
e-mail: luca.rossini@unibz.it

M. Corazza et al. (eds.), *Mathematical and Statistical Methods for Actuarial Sciences and Finance*, https://doi.org/10.1007/978-3-319-89824-7_29

estimation of large number of parameters with few observations. In order to avoid overparametrization, overfitting and dimensionality issues, Bayesian inference and suitable classes of prior distributions have been proposed.

In this paper, a novel Bayesian nonparametric hierarchical prior for multivariate time series is proposed, which allows shrinkage of the SUR/VAR coefficients to multiple locations using a Normal-Gamma distribution with location, scale and shape parameters unknown. In our sparse SUR/VAR (sSUR/sVAR), some SUR/VAR coefficients shrink to zero, due to the shrinking properties of the lasso-type distribution at the first stage of our hierarchical prior, thus improving efficiency of parameters estimation, prediction accuracy and interpretation of the temporal dependence structure in the time series. We use a Bayesian Lasso prior, which allows us to reformulate the SUR/VAR model as a penalized regression problem, in order to determine which SUR/VAR coefficients shrink to zero (see [7, 10]).

As regards to the second stage of the hierarchy, we use a random mixture distribution of the Normal-Gamma hyperparameters, which allows for parameter parsimony through two components. The first component is a random Dirac point-mass distribution, which induces shrinkage for SUR coefficients; the second component is a Dirichlet process hyperprior, which allows for clustering of the SUR/VAR coefficients.

The structure of the paper is as follows. Section 2 introduces the vector autoregressive model. Section 3 describes briefly the Bayesian nonparametric sparse model. Section 4 presents some simulation results for different dimensions. Section 5 concludes.

2 The Vector Autoregressive model

Let $\mathbf{y}_t = (\mathbf{y}'_{1,t}, \ldots, \mathbf{y}'_{N,t})' \in \mathbb{R}^m$ be a vector-valued time series. We consider a VAR model of order p (VAR(p)) as

$$\mathbf{y}_t = \mathbf{b} + \sum_{i=1}^{p} B_i \mathbf{y}_{t-i} + \boldsymbol{\varepsilon}_t, \tag{1}$$

for $t = 1, \ldots, T$, where $\mathbf{y}_t = (y_{1,t}, \ldots, y_{m,t})'$, $\mathbf{b} = (b_1, \ldots, b_m)'$ and B_i is a $(m \times m)$ matrix of coefficients. We assume that $\boldsymbol{\varepsilon}_t = (\varepsilon_{1,t}, \ldots, \varepsilon_{m,t})'$ follows a independent and identically distributed Gaussian distribution $\mathcal{N}_m(\mathbf{0}, \Sigma)$ with mean $\mathbf{0}$ and covariance matrix Σ.

The VAR(p) in (1) can be rewritten in a stacked regression form:

$$\mathbf{y}_t = (I_m \otimes \mathbf{x}'_t)\boldsymbol{\beta} + \boldsymbol{\varepsilon}_t, \tag{2}$$

where $\mathbf{x}_t = (1, y'_{t-1}, \ldots, y'_{t-p})'$ is the vector of predetermined variables, $\boldsymbol{\beta} = \text{vec}(B)$, where $B = (\mathbf{b}, B_1, \ldots, B_p)$, $\otimes$ is the Kronecker product and vec the

column-wise vectorization operator that stacks the columns of a matrix in a column vector.

3 Bayesian Nonparametric Sparse VAR

In this paper we define a hierarchical prior distribution which induces sparsity on the vector of coefficients β. In order to regularize (2) we incorporate a penalty using a lasso prior $f(\boldsymbol{\beta}) = \prod_{j=1}^{r} \mathcal{NG}(\beta_j|0, \gamma, \tau)$, where $\mathcal{NG}(\beta|\mu, \gamma, \tau)$ denotes the normal-gamma distribution with location parameter μ, shape parameter $\gamma > 0$ and scale parameter $\tau > 0$. The normal-gamma distribution induces shrinkage toward the prior mean of μ, but we can extend the lasso model specification by introducing a mixture prior with separate location parameter μ_j^*, separate shape parameter γ_j^* and separate scale parameter τ_j^* such that: $f(\boldsymbol{\beta}) = \prod_{j=1}^{r} \mathcal{NG}(\beta_j|\mu_j^*, \gamma_j^*, \tau_j^*)$. In our paper, we favor the sparsity of the parameters through the use of carefully tailored hyperprior and we use a nonparametric Dirichlet process prior (DPP), which reduces the overfitting problem and the curse of dimensionality by allowing for parameters clustering due to the concentration parameter and the base measure choice.

In our case we define $\boldsymbol{\theta}^* = (\boldsymbol{\mu}^*, \boldsymbol{\gamma}^*, \boldsymbol{\tau}^*)$ as the parameters of the Normal-Gamma distribution, and assume a prior $\mathbb{Q}_l$ for $\boldsymbol{\theta}_{lj}^*$, that is

$$\beta_j \overset{ind}{\sim} \mathcal{NG}(\beta_j|\mu_j^*, \gamma_j^*, \tau_j^*), \tag{3}$$

$$\boldsymbol{\theta}_{lj}^*|\mathbb{Q}_l \overset{i.i.d.}{\sim} \mathbb{Q}_l, \tag{4}$$

for $j = 1, \ldots, r_l$ and $l = 1, \ldots, N$.

Following a construction of the hierarchical prior similar to the one proposed in [4] we define the vector of random measures

$$\begin{aligned} \mathbb{Q}_1(d\boldsymbol{\theta}_1) &= \pi_1 \mathbb{P}_0(d\boldsymbol{\theta}_1) + (1 - \pi_1)\mathbb{P}_1(d\boldsymbol{\theta}_1), \\ &\vdots \\ \mathbb{Q}_N(d\boldsymbol{\theta}_N) &= \pi_N \mathbb{P}_0(d\boldsymbol{\theta}_N) + (1 - \pi_N)\mathbb{P}_N(d\boldsymbol{\theta}_N), \end{aligned} \tag{5}$$

with the same sparse component $\mathbb{P}_0$ in each equation and with the following hierarchical construction as previously explained,

$$\begin{aligned} &\mathbb{P}_0(d\boldsymbol{\theta}) \sim \delta_{\{(0,\gamma_0,\tau_0)\}}(d(\mu, \gamma, \tau)), \\ &\mathbb{P}_l(d\boldsymbol{\theta}) \overset{i.i.d.}{\sim} \mathrm{DP}(\tilde{\alpha}, G_0), \quad l = 1, \ldots, N, \\ &\pi_l \overset{i.i.d.}{\sim} \mathcal{B}e(\pi_l|1, \alpha_l), \quad l = 1, \ldots, N, \end{aligned} \tag{6}$$

$$(\gamma_0, \tau_0) \sim g(\gamma_0, \tau_0 | \nu_0, p_0, s_0, n_0),$$
$$G_0 \sim \mathcal{N}(\mu | c, d) \times g(\gamma, \tau | \nu_1, p_1, s_1, n_1)$$

where $\delta_{\{\boldsymbol{\psi}_0\}}(\boldsymbol{\psi})$ denotes the Dirac measure indicating that the random vector $\boldsymbol{\psi}$ has a degenerate distribution with mass at the location $\boldsymbol{\psi}_0$, and $g(\gamma_0, \tau_0)$ is the conjugate joint prior distribution (see [6]). We apply the Gibbs sampler and the hyperparameters given in [2] for the posterior approximation.

4 Simulation Results

The nonparametric prior presented in Sect. 3 allows for shrinking the SUR coefficients. In order to assess the goodness of the prior we performed a simulation study of our Bayesian nonparametric sparse model. We consider different datasets with sample size $T = 100$ from the VAR model of order 1:

$$\mathbf{y}_t = B\mathbf{y}_{t-1} + \boldsymbol{\varepsilon}_t, \quad \boldsymbol{\varepsilon}_t \overset{i.i.d}{\sim} \mathcal{N}_m(\mathbf{0}, \Sigma) \quad t = 1, \ldots, 100,$$

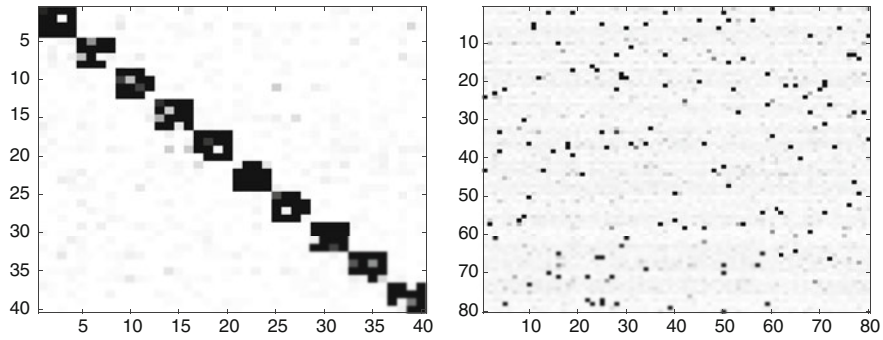
where the dimension of $\mathbf{y}_t$ and of the square matrix of coefficients B can take different values: $m = 20$ (small dimension), $m = 40$ (medium dimension) and $m = 80$ (large dimension). Furthermore, we choose different settings of the matrix B, focusing on a block-diagonal structure with random entries of the blocks:

- the block-diagonal matrix $B = \text{diag}\{B_1, \ldots, B_{m/4}\} \in \mathcal{M}_{(m,m)}$ is generated with blocks B_j ($j = 1, \ldots, m/4$) of (4×4) matrices on the main diagonal:

$$B_j = \begin{pmatrix} b_{11,j} & \ldots & b_{14,j} \\ \vdots & \vdots & \vdots \\ b_{41,j} & \ldots & b_{44,j} \end{pmatrix},$$

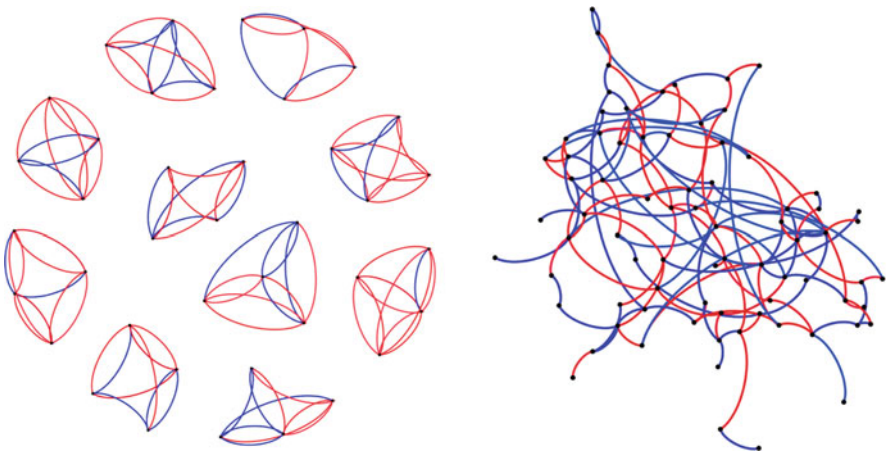
  where the elements are randomly taken from an uniform distribution $\mathcal{U}(-1.4, 1.4)$ and then checked for the weak stationarity condition of the VAR;
- the random matrix B is a (80×80) matrix with 150 elements randomly chosen from an uniform distribution $\mathcal{U}(-1.4, 1.4)$ and then checked for the weak stationarity condition of the VAR.

Figure 1 exhibits the posterior mean of Δ, which shows us the allocation of the coefficients between the two random measures $\mathbb{P}_0$ and $\mathbb{P}_l$. In particular, we have that the white color indicates if the coefficient δ_{ij} is equal to zero (i.e. sparse component), while the black one if the δ_{ij} is equal to one, for nonsparse components. The definition of the pairwise posterior probabilities and of the co-clustering matrix for the atom locations μ allows us to built the weighted networks (see Fig. 2), where the blue edges represent negative weights, while the red ones represent the positive



(a) (b)

Fig. 1 Posterior mean of the matrix of δ for $m = 40$ (**a**) and for $m = 80$ (**b**) with random element



(a) (b)

Fig. 2 Weighted network for $m = 40$ (**a**) and for $m = 80$ (**b**) with random elements, where the blue edges means negative weights and red ones represent positive weights

weights. In each coloured graph the nodes represent the n variables of the VAR model, and a clockwise-oriented edge between two nodes i and j represents a non-null coefficient for the variable $y_{j,t-1}$ in the i-th equation of the VAR.

5 Conclusions

This paper proposes a novel Bayesian nonparametric prior for SUR models, which allows for shrinking SUR coefficients toward multiple locations and for identifying groups of coefficients. We introduce a two-stage hierarchical distribution, which consists in a hierarchical Dirichlet process on the parameters of the Normal-Gamma distribution. The proposed hierarchical prior is used to proposed a Bayesian nonparametric model for SUR models. We provide an efficient Monte Carlo Markov Chain algorithm for the posterior computations and the effectiveness of this algorithm is assessed in simulation exercises. The simulation studies illustrate the good performance of our model with different sample sizes for B and $\mathbf{y}_t$.

References

1. Banbura, M., Giannone, D., Reichlin, L.: Large Bayesian vector autoregressions. J. Appl. Econ. **25**(1), 71–92 (2010)
2. Billio, M., Casarin, R., Rossini, L.: Bayesian nonparametric sparse seemingly unrelated regression model (2017). https://arxiv.org/abs/1608.02740
3. Carriero, A., Clark, T.E., Marcellino, M.: Bayesian VARs: specification choices and forecast accurancy. J. Appl. Econ. **30**(1), 46–73 (2015)
4. Hatjispyros, S.J., Nicoleris, T.N., Walker, S.G.: Dependent mixtures of Dirichlet processes. Comput. Stat. Data Anal. 55(6), 2011–2025 (2011)
5. Koop, G.: Forecasting with medium and large Bayesian VARs. J. Appl. Econ. **28**(2), 177–203 (2013)
6. Miller, R.: Bayesian analysis of the two-parameter gamma distribution. Technometrics **22**(1) (1980)
7. Park, T., Casella, G.: The Bayesian Lasso. J. Am. Stat. Assoc. **103**(482), 681–686 (2008)
8. Scott, S.L., Varian, H.R.: Predicting the present with Bayesian structural time series. Int. J. Math. Model. Numer. Optim. **5**(1–2), 4–23 (2013)
9. Stock, J.H., Watson, M.W.: Generalized shrinkage methods for forecasting using many predictors. J. Bus. Econ. Stat. **30**(4), 481–493 (2012)
10. Tibshirani, R.: Regression shrinkage and selection via the lasso. J. R. Stat. Soc. B **58**(1), 267–288 (1996)
11. Zellner, A.: An efficient method of estimating seemingly unrelated regressions and tests of aggregation bias. J. Am. Stat. Assoc. **57**(298), 500–509 (1962)
12. Zellner, A.: An Introduction to Bayesian Inference in Econometrics. Wiley, New York (1971)

Logistic Classification for New Policyholders Taking into Account Prediction Error

Eva Boj and Teresa Costa

Abstract An expression of the mean squared error, MSE, of prediction for new observations when using logistic regression is showed. First, MSE is approximated by the sum of the process variance and of the estimation variance. The estimation variance can be estimated by applying the delta method and/or by using bootstrap methodology. When using bootstrap, e.g. bootstrap residuals, it is possible to obtain an estimation of the distribution for each predicted value. Confidence intervals can be calculated taking into account the bootstrapped distributions of the predicted new values to help us in the knowledge of their randomness. The general formulas of prediction error (the square root of MSE of prediction), PE, in the cases of the power family of error distributions and of the power family of link functions for generalized linear models were obtained in previous works. Now, the expression of the MSE of prediction for the generalized linear model with Binomial error distribution and logit link function, the logistic regression, is obtained. Its calculus and usefulness are illustrated with real data to solve the problem of Credit Scoring, where policyholders are classified into defaulters and non-defaulters.

Keywords Logistic regression · Mean squared error · Estimation variance · Delta method · Bootstrapping residuals · Credit risk

1 Introduction

The main objective is to estimate prediction error (PE) for new observations in logistic regression. PE is the square root of mean squared error (MSE), which can be approximated by the sum of two components: the process variance and the estimation variance. The estimation variance can be estimated by applying the delta method and/or by using bootstrap methodology. General formulas of PE for

E. Boj · T. Costa (✉)
Department of Mathematics for Economics, Finance and Actuarial Sciences, Faculty of Economics and Business, University of Barcelona, Barcelona, Spain
e-mail: evaboj@ub.edu; tcosta@ub.edu

M. Corazza et al. (eds.), *Mathematical and Statistical Methods for Actuarial Sciences and Finance*, https://doi.org/10.1007/978-3-319-89824-7_30

generalized linear models (GLM) with error distributions and link functions in the power families were deduced in previous works (see [1, 3] and [4]). In the current study the expressions of MSE for the GLM with Binomial error distribution and logit link function, the logistic regression, are obtained.

In credit risk, the observed data are binary, and logistic regression is applied to estimate the probability of default, $\mu_i = \hat{p}_i$. A credit scoring model classifies obligors into defaulters and non-defaulters. If the probability of default is over a cut-off, usually 0.5, then the individual is classified as bad risk; otherwise the individual is classified as good risk. The classification of the model and the current classification can be compared in a contingency table and error rates can be measured taking into account misclassification. To validate scoring models there are measures to evaluate both discrimination and calibration. The discrimination is the model's ability to separate between defaulters and non-defaulters and the calibration is the model's ability to estimate the probabilities of default. The Brier's score (BS), also known as the mean squared error, is a measure of accuracy and thus quantifies the deviation of forecasts probabilities and observed binary outcomes. In addition, some papers suggest the decomposition of the BS in the calibration, the discrimination and other components (see, e.g., [16] and [11]). These measures are calculated using only the actual data. Sometimes, the data are used twice, both to fit the model and to check its accuracy. In the proposed formulations, the idea is to compute the MSE for new cases different from those of observed data. In particular, the aim is the classification of new policyholders taking into account the information given by the PE.

2 Mean Squared Error for New Observations

Assume a GLM (see [13]). Let $\Omega = (\Omega_1, \ldots, \Omega_n)$ be a population of n individuals, let $Y : (Y_1, \ldots, Y_n)^T$ be the random response variable, and $(y_1, \ldots, y_n)^T$ be the observed response variable of size $n \times 1$,let $(w_1, \ldots, w_n)^T$ be (*a priori*) weights of individuals of size $n \times 1$, let $\left(F_1, \ldots, F_p\right)$ be the set of p observed predictors of the linear predictor η, let a link function $\eta = g(\mu)$, and let the mean and the variance of Y:

$$\mu = E[y] \text{ and } Var[y] = \left(\phi / w\right) V(\mu).$$

MSE of prediction, $E\left[\left(y - \hat{\mu}\right)^2\right]$, is the difference between the probability of defaults forecasted and the actual default or non-default outcome for each obligor that is observed. Because y_{n+1} is *a priori* unknown for new observations, an estimation of the MSE is required. The aim then is to calculate $E\left[\left(y_{n+1} - \hat{\mu}_{n+1}\right)^2\right]$

for $i = n + 1$. As it is showed in, e.g., [1, 3, 8, 12, 14] and [4], MSE can be approximated with the sum of the *process variance* and of the *estimation variance*:

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

Using the delta method (see [10, 14] and [15]), the estimation variance can be derived from $Var\left[\hat{\mu}_i\right] \cong \left|\frac{\partial \mu_i}{\partial \eta_i}\right|^2 Var\left[\eta_i\right]$, and then MSE is:

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

If bootstrap methodology, e.g., bootstrapping based on Pearson residuals is applied (see [6] and [7]), an alternative approximation formula of the estimation variance can be used and then MSE is: $E\left[\left(y_i - \hat{y}_i\right)^2\right] \cong \left(\phi / w_i\right) V\left(\mu_i\right) + Var\left[\hat{\mu}_i^{boot}\right]$.

In logistic regression (see, e.g., [13] or [9]) $Y \sim Binomial(\pi)$, $V\left(\mu_i\right) = \mu_i\left(1 - \mu_i\right)$ and $\phi = 1$, and the process variance is $Var\left[\mu_i\right] = \left(1 / w_i\right) \mu_i\left(1 - \mu_i\right)$. And the canonical link is the logit, $\eta_i = \text{logit}\left(\mu_i\right) = \log\left(\frac{\mu_i}{1+\mu_i}\right)$, and $\mu_i = \frac{\exp(\eta_i)}{1+\exp(\eta_i)}$. Developing the previous formulas of MSE in the case of logistic regression the two following approximations are obtained:

- Delta method: $E\left[\left(y_i - \hat{y}_i\right)^2\right] \cong \left(1 / w_i\right) \mu_i\left(1 - \mu_i\right) + \left(\mu_i\left(1 - \mu_i\right)\right)^2 Var\left[\eta_i\right]$.
- Bootstrap methodology: $E\left[\left(y_i - \hat{y}_i\right)^2\right] \cong \left(1 / w_i\right) \mu_i\left(1 - \mu_i\right) + Var\left[\hat{\mu}_i^{boot}\right]$.

3 Application: Credit Scoring

To illustrate the interpretability and the usefulness of PE when using logistic regression, the formulas are applied to solve the credit scoring problem. A real German data set[1] is analyzed. The data consist of 700 examples of creditworthy applicants and 300 examples where credit should not be extended. For each applicant, 24 predictors or risk factors describe credit history, account balances, loan purpose, loan amount, employment status, age and job among others. Risk factors are grouped in four classes and there are 10 continuous variables, 8 qualitative variables and 6 binary variables. In the literature and in previous works different methodologies have been applied to German data (see [2, 17] and [5]). For logistic regression the `stats` function in `glm` package of `R` is used to fit the model and forecast the default probabilities.

[1] http://archive.ics.uci.edu/ml/datasets/Statlog+(German+Credit+Data).

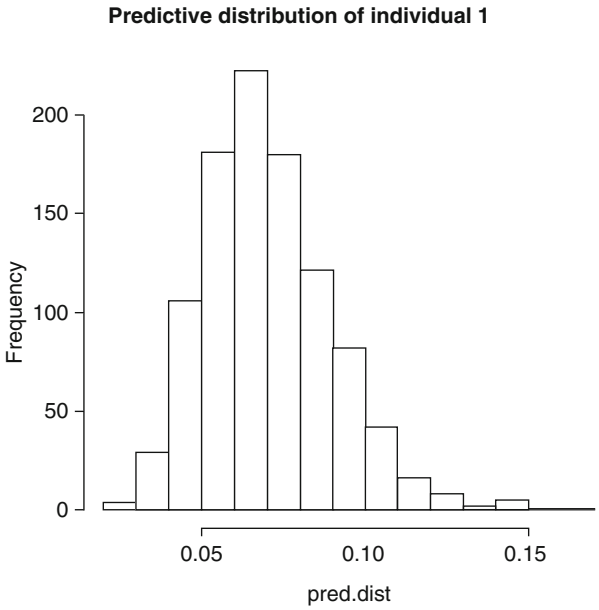


Fig. 1 Predictive distribution of PE for individual 1, a non-defaulter or good risk, $y = 0$, assuming that it doesn't belong to the portfolio (i.e., it is a new random case). Descriptive statistics of the predictive distribution are: mean = 0.07041644; standard deviation = 0.01995177 and coefficient of variation = 28.33397%

The previous formulas for a new individual are applied. Assume that individual 1 (a non-defaulter, good risk) doesn't belong to the portfolio (i.e., it is a new random case), and it must be classified. The estimated probability of default is $\hat{p}_{(1)} = 0.03376495$. PE using the delta method is $\hat{PE}_1 = 0.1814338$. Confidence intervals including a 5% of the PE can be constructed: [0.02467884, 0.04285106]. In the case of bootstrap, PE is $\hat{PE}_1 = 0.1817222$ and confidence intervals can be constructed with the 5th and 95th quantiles of the predictive distribution (see the histogram in Fig. 1 above): [0.04179172, 0.10677569]. It can be observed that, when including PE, all values in the confidence intervals are lower than the cut-off point of 0.5. That is, as a result individual 1 is classified in both cases as non-defaulter or good risk. Therefore, by analyzing the particular cases, the decision maker may choose to grant the credit or not to an applicant.

Acknowledgements Spanish Ministry of Economy and Competitiveness, MTM2014-56535-R.

References

1. Boj, E., Costa, T.: Provisions for claims outstanding, incurred but not reported, with generalized linear models: prediction error formulation by calendar years. Cuadernos de Gestión **17**(2), 157–174 (2017)
2. Boj, E., Claramunt, M.M., Esteve, A., Fortiana, J.: Criterios de selección de modelo en credit scoring, aplicación del análisis discriminante basado en distancias. Anales del Instituto de Actuarios Españoles **15**, 209–230 (2009)
3. Boj, E., Costa, T., Fortiana, J.: Prediction error in distance-based generalized linear models. In: Palumbo, F., Montanari, A., Vichi, M. (eds.) Data Science. Innovative Developments in Data Analysis and Clustering. Studies in Classification, Data Analysis, and Knowledge Organization, vol. 1, pp. 191–204. Springer International Publishing, Berlin (2017)
4. Costa, T., Boj, E.: Prediction error in distance-based generalized linear models. In: Corazza, M., Legros, F., Perna, C., Sibillo, M. (eds.) Mathematical and Statistical Methods for Actuarial Sciences and Finance. MAF 2012, pp. 97–108. Springer International Publishing, Switzerland (2017)
5. Costa, T., Boj. E., Fortiana, J.: Bondad de ajuste y elección del punto de corte en regresión logistica basada en distancias. Aplicación al problema del credit scoring. Anales del Instituto de Actuarios Españoles **18**, 19–40 (2012)
6. Efron, B.: Bootstrap methods: another look at the Jacknife. Ann. Stat. **7**, 1–26 (1979)
7. Efron, B., Tibshirani, J.: An Introduction to the Bootstrap. Chapman and Hall, New York (1998)
8. England, P.D., Verrall, R.J.: Analytic and bootstrap estimates of prediction errors in claims reserving. Insur. Math. Econ. **25**, 281–293 (1999)
9. Hosmer, D.W., Lemeshow, S.: Applied Logistic Regression, 2nd edn. Wiley, New York (2000)
10. Hosmer, D.W., Lemeshow, S., May. S.: Applied Survival Analysis: Regression Modeling of Time-to-Event Data, 2nd edn. Wiley, New Jersey (2008)
11. Joseph, M.P.: A PD validation framework for basel II internal ratings-based systems. Credit Scoring and Credit Control IV (2005)
12. Kaas, R., Goovaerts, M., Dhaene, J., Denuit, M.: Modern Actuarial Risk Theory Using R, 2nd edn. Springer, Heidelberg (2008)
13. McCullagh, P., Nelder, J.A.: Generalized Linear Models, 2nd edn. Chapman and Hall, London (1989)
14. Renshaw, A.E.: On the Second Moment Properties and the Implementation of Certain GLIM Based Stochastic Claims Reserving Models. Actuarial Research Paper, 65. Department of Actuarial Science and Statistics, City University, London (1994)
15. Sánchez-Niubó, A.: Development of statistical methodology to study the incidence of drug use. PhD Thesis, University of Barcelona (2014)
16. Stephenson, D.B., Coelho, C.A.S., Jolliffe, I.T.: Two extra components in the brier score decomposition. Weather Forecast. **23**, 752–757 (2008)
17. West, D.: Neural network credit scoring models. Comput. Oper. Res. **27**, 1131–1152 (2000)

Conditional Quantile-Located VaR

Giovanni Bonaccolto, Massimiliano Caporin, and Sandra Paterlini

Abstract The Conditional Value-at-Risk (CoVaR) has been proposed by Adrian and Brunnermeier (Am Econ Rev 106:1705–1741, 2016) to measure the impact of a company in distress on the Value-at-Risk (VaR) of the financial system. We propose an extension of the CoVaR, that is, the Conditional Quantile-Located VaR (QL-CoVaR), that better deals with tail events, when spillover effects impact the stability of the entire system. In fact, the QL-CoVaR is estimated by assuming that the financial system and the individual companies simultaneously lie in the left tails of their distributions.

Keywords CoVaR · Systemic risk · Regression quantiles

1 Methods

We first introduce the Conditional Value-at-Risk (CoVaR) proposed by Adrian and Brunnermeier [1]. Then, we provide the details about the Conditional Quantile-Located Value-at-Risk (QL-CoVaR). Let y_t and $x_{i,t}$ be the returns of the financial system and of the i-th financial company at time t, respectively, for $i = 1, \ldots, N$ and $t = 1, \ldots, T$. Let $Q_\tau(x_{i,t}|\mathbf{I}_{t-1})$ denotes the τ-th quantile of $x_{i,t}$, for $\tau \in (0, 1)$, conditional to the information set $\mathbf{I}_{t-1}$, where $\mathbf{I}_{t-1} = (y_{t-1}, x_{i,t-1}, m_{t-1})$ with

G. Bonaccolto
University of Enna "Kore", viale delle Olimpiadi, Enna, Italy
e-mail: giovanni.bonaccolto@unikore.it

M. Caporin (✉)
Department of Statistical Sciences, University of Padova, Padova, Italy
e-mail: massimiliano.caporin@unipd.it

S. Paterlini
FACT Department–Finance, EBS Business School, Wiesbaden, Germany

Department of Economics and Management, University of Trento, Trento, Italy
e-mail: sandra.paterlini@ebs.edu

M. Corazza et al. (eds.), *Mathematical and Statistical Methods for Actuarial Sciences and Finance*, https://doi.org/10.1007/978-3-319-89824-7_31

m_{t-1} being a control variable at time $t-1$. Similarly, $Q_\theta(y_t|\mathbf{I}_{t-1}, x_{i,t})$ is the θ-th quantile of y_t conditional to the information set available at $t-1$ as well as to the return of the i-th company observed at time t, for $\theta \in (0,1)$. For simplicity, we set $Q_\tau(x_{i,t}|\mathbf{I}_{t-1}) \equiv Q_\tau(x_{i,t})$ and $Q_\theta(y_t|\mathbf{I}_{t-1}, x_{i,t}) \equiv Q_\theta^{(i)}(y_t)$; θ and τ take low values, typically in the interval $(0, 0.05)$. The CoVaR introduced by Adrian and Brunnermeier [1] is then estimated from the quantile regression models (see [4]):

$$Q_\tau(x_{i,t}) = \alpha_\tau^{(i)} + \beta_\tau^{(i)} m_{t-1}, \tag{1}$$

$$Q_\theta^{(i)}(y_t) = \delta_\theta^{(i)} + \lambda_\theta^{(i)} x_{i,t} + \gamma_\theta^{(i)} m_{t-1}. \tag{2}$$

Let $\widehat{Q}_\tau(x_{i,\tau}) = \widehat{\alpha}_\tau^{(i)} + \widehat{\beta}_\tau^{(i)} m_{t-1}$ be the estimated τ-th quantile of $x_{i,t}$, it is possible to compute the CoVaR at the distress and at the median state of the conditioning company, respectively, as follows:

$$CoVaR_{t,\theta,\tau}^{(i)} = \widehat{\delta}_\theta^{(i)} + \widehat{\lambda}_\theta^{(i)} \widehat{Q}_\tau(x_{i,t}) + \widehat{\gamma}_\theta^{(i)} m_{t-1}, \tag{3}$$

$$CoVaR_{t,\theta,1/2}^{(i)} = \widehat{\delta}_\theta^{(i)} + \widehat{\lambda}_\theta^{(i)} \widehat{Q}_{1/2}(x_{i,t}) + \widehat{\gamma}_\theta^{(i)} m_{t-1}, \tag{4}$$

and compute the ΔCoVaR to quantify the marginal contribution of the i-th company to the systemic risk (see [1]). Note that $CoVaR_{t,\theta,1/2}^{(i)}$ is always parameterized to the median state of the i-th conditioning company. Hence, we can omit the level 1/2 as subscript of the ΔCoVaR measure as follows:

$$\Delta CoVaR_{t,\theta,\tau}^{(i)} = CoVaR_{t,\theta,\tau}^{(i)} - CoVaR_{t,\theta,1/2}^{(i)} = \widehat{\lambda}_\theta^{(i)} \left[\widehat{Q}_\tau(x_{i,t}) - \widehat{Q}_{1/2}(x_{i,t})\right]. \tag{5}$$

For simplicity, we set $\theta = \tau$ and, then, $\Delta CoVaR_{t,\theta,\tau}^{(i)} \equiv \Delta CoVaR_{t,\tau}^{(i)}$. It is important to highlight that the parameters in (2) and the coefficients in (3) are functions of θ only, neglecting the role of τ. Therefore, the estimation process behind (3) depends on $x_{i,t}$ and not on $Q_\tau(x_{i,t})$. In contrast, we estimate the parameters in (2) assuming that the financial system and the i-th company simultaneously lie in the left tails of their distributions. We then take into account the impact exerted by $x_{i,t}$—in the neighbourhood of its τ-th quantile—on $\widehat{Q}_\theta^{(i)}(y_t)$. This allows us to increase the distress degree in the connections between the individual companies and the system to make our risk measure more sensitive to extreme events. The model we propose is defined as follows:

$$Q_{\theta,\tau}^{(i)}(y_t) = \delta_{\theta,\tau}^{(i)} + \lambda_{\theta,\tau}^{(i)} x_{i,t} + \gamma_{\theta,\tau}^{(i)} m_{t-1}, \tag{6}$$

where the parameters have both θ and τ as subscripts, as they depend on the quantiles levels of both y_t and $x_{i,t}$.

In fact, the unknown parameters in (6) are estimated from the following minimization problem:

$$\underset{\delta_{\theta,\tau}^{(i)},\lambda_{\theta,\tau}^{(i)},\gamma_{\theta,\tau}^{(i)}}{\arg\min} \sum_{t=1}^{T} \rho_\theta \left[y_t - \delta_{\theta,\tau}^{(i)} - \lambda_{\theta,\tau}^{(i)} x_{i,t} - \gamma_{\theta,\tau}^{(i)} m_{t-1} \right] K \left(\frac{\widehat{F}_{t|t-1}(x_{i,t}) - \tau}{h} \right), \quad (7)$$

where $\rho_\theta(e) = e(\theta - \mathbf{1}_{\{e<0\}})$ is the asymmetric loss function used in the quantile regression method by Koenker and Bassett [4]; $\mathbf{1}_{\{\cdot\}}$ is an indicator function, taking the value of 1 if the condition in $\{\cdot\}$ is satisfied, the value of 0 otherwise; $K(\cdot)$ is the kernel function, with bandwidth h, whereas $\widehat{F}_{t|t-1}(x_{i,t})$ is the empirical conditional quantile of $x_{i,t}$. Sim and Zhou [5] used a similar approach to estimate the relations in quantiles between oil prices and stock returns.

In contrast to [5], we estimate $\widehat{F}_{t|t-1}(x_{i,t})$ dynamically using a rolling window procedure. For each window, we estimate a large set of $x_{i,t}$ quantiles in the support $\tau \in (0, 1)$ from the quantile regression model (1), using the method proposed by Bondell et al. [2] to ensure the monotonicity of the multiple quantiles for $\tau \in (0, 1)$. Then, we linearly interpolate the set of quantiles to obtain the conditional distribution of $x_{i,t}$ at time t, denoted as $\widehat{F}(x_{i,t}|m_{t-1})$. Finally, we recover $\widehat{F}_{t|t-1}(x_{i,t})$, as the probability level, extrapolated from $\widehat{F}(x_{i,t}|m_{t-1})$, corresponding to the realization $x_{i,t}$. From the method described above, we then compute the QL-CoVaR at the τ-th level: $QL\text{-}CoVaR_{t,\theta,\tau}^{(i)} = \widehat{\delta}_{\theta,\tau}^{(i)} + \widehat{\lambda}_{\theta,\tau}^{(i)} \widehat{Q}_\tau(x_{i,t}) + \widehat{\gamma}_{\theta,\tau}^{(i)} m_{t-1}$, where $\widehat{Q}_\tau(x_{i,t}) = \widehat{\alpha}_\tau^{(i)} + \widehat{\beta}_\tau^{(i)} m_{t-1}$, and define the ΔQL-CoVaR as:

$$\begin{aligned} \Delta QL\text{-}CoVaR_{t,\tau}^{(i)} &= QL\text{-}CoVaR_{t,\theta,\tau}^{(i)} - QL\text{-}CoVaR_{t,\theta,1/2}^{(i)} = \widehat{\delta}_{\theta,\tau}^{(i)} - \widehat{\delta}_{\theta,1/2}^{(i)} \\ &+ \widehat{\lambda}_{\theta,\tau}^{(i)} \left[\widehat{Q}_\tau(x_{i,t}) - \widehat{Q}_{1/2}(x_{i,t}) \right] + (\widehat{\lambda}_{\theta,\tau}^{(i)} - \widehat{\lambda}_{\theta,1/2}^{(i)}) \widehat{Q}_{1/2}(x_{i,t}) \\ &+ (\widehat{\gamma}_{\theta,\tau}^{(i)} - \widehat{\gamma}_{\theta,1/2}^{(i)}) m_{t-1}. \end{aligned} \quad (8)$$

It is important to highlight that $\Delta QL\text{-}CoVaR_{t,\tau}^{(i)}$ includes more components than $\Delta CoVaR_{t,\tau}^{(i)}$ in (5), as the coefficients in (8) also depend on the state of the i-th company. We then have further information about the relationships between the financial system and the individual companies when we focus on the left tails of their distributions. We compute the standard errors of the ΔCoVaR and the ΔQL-CoVaR coefficients using the bootstrap approach (see, e.g., [3]).

2 Empirical Results

We implement the methods discussed in Sect. 1 on the daily returns of 1155 U.S. financial institutions (952 banks and 203 insurance companies) in the period between October 10, 2000 and July 31, 2015, for a total of 3864 days.[1]

[1] The data are recovered from Thomson Reuters Datastream.

We note that some of the companies enter the dataset after October 10, 2000, whereas others exit before July 31, 2015. We estimate the models described in Sect. 1 for each of the financial companies for which we have at least 200 observations, resulting in 1030 companies. We also build an index providing the return of the financial system (y_t) from the returns of the 1155 financial institutions included in our dataset, weighted by their market values. As for m_t, we use the first principal component of variables that are related to bond, equity and real estate markets: (i) the CBOE Volatility Index (VIX); (ii) the liquidity spread (LS), computed as the difference between the three-month collateral repo rate and the three-month bill rate; (iii) the change in the three-month Treasury bill rate (TB); (iv) the change in the slope of the yield curve (YC), computed as the spread between the ten-year Treasury rate and the three-month bill rate; (v) the change in the credit spread between BAA-rated bonds and the Treasury rate (CS), both with the 10 years maturity; (vi) the daily equity market return (EM); (vii) the excess return of the real estate sector over the market return (RE).[2] In particular, we checked that the first principal component (m_t) of the variables listed above explains 96.50% of the variability in the data. We estimate the CoVaR and the QL-CoVaR using two quantile levels—$\theta = \tau = 0.01$ and $\theta = \tau = 0.05$. As for the estimation of the QL-CoVaR parameters, we use the Gaussian kernel as $F(\cdot)$, with $h = \{0.10, 0.15, 0.20\}$.

By analyzing the CoVaR parameters, we checked that positive returns of the individual companies have a positive impact on the VaR of the financial system, as $\lambda_\theta^{(i)}$ takes, on average, positive values. Likewise, the average impact exerted by the companies to both $QL\text{-}CoVaR_\tau^{(i)}$ and $QL\text{-}CoVaR_{1/2}^{(i)}$ is positive, but greater with respect to the standard CoVaR (the medians of both $\widehat{\lambda}_{\theta,\tau}^{(i)}$ and $\widehat{\lambda}_{\theta,0.5}^{(i)}$ are greater than the median of $\widehat{\lambda}_\theta^{(i)}$).[3] Therefore, the relationships between the system and the companies become stronger by focusing on particular regions of the $x_{i,t}$ support, i.e. when $x_{i,t}$ is in a neighbourhood of a distress state. On average, we observe larger values for $\widehat{\lambda}_{\theta,\tau}^{(i)}$ at $\theta = 0.01$ than at $\theta = 0.05$, whereas the opposite holds for $\widehat{\lambda}_{\theta,0.5}^{(i)}$. $\widehat{\lambda}_{\theta,\tau}^{(i)}$ measures the relation between $x_{i,t}$ and y_t, when the companies and the system simultaneously lie in the left tail of their distributions. The fact that $\widehat{\lambda}_{\theta,\tau}^{(i)}$ increases as θ and τ simultaneously decrease means that the co-movements between the financial system and the companies become stronger when moving leftwards along the left tails of their distributions. Consequently, the risk of contagion increases by accentuating the distress degree in the connections between the financial system and the companies. For both CoVaR and QL-CoVaR, the coefficient measuring the impact of the individual companies—λ—is statistically significant at the 5% level for the majority of companies.

[2]The control variables listed in (i)—(v) are taken from Thomson Reuters Datastream, whereas EM and RE are recovered from the industry portfolios built by Kenneth R. French, available at http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html.

[3]Figures and tables reporting empirical results, that we omit here for the sake of space, are available upon request.

References

1. Adrian, T., Brunnermeier, K.: CoVaR. Am. Econ. Rev. **106**, 1705–1741 (2016)
2. Bondell, H.D., Reich, B.J., Wang, H.: Non-crossing quantile regression curve estimation. Biometrika **97**, 825–838 (2010)
3. Efron, B.: Bootstrap methods: another look at the jackknife. Ann. Stat. **7**, 1–26 (1979)
4. Koenker, R., Bassett, G.: Regression quantiles. Econometrica **46**, 33–50 (1978)
5. Sim, N., Zhou, H.: Oil prices, US stock return, and the dependence between their quantiles. J. Bank. Financ. **55**, 1–8 (2015)

Probability of Default Modeling: A Machine Learning Approach

Stefano Bonini and Giuliana Caivano

Abstract Default prediction through probability of default modeling has attracted lots of research interests in the past literature and recent studies have shown that Artificial Intelligence (AI) methods achieved better performance than traditional statistical methods. This paper empirically investigates the results of applying different machine learning techniques through the overall estimation process to reduce the running time, maximize—in the first stage—the predictive power and contribute of each variable to the estimation of PDs. In the second stage, we have identified the best multivariate combination of drivers by comparing the results of a set of supervised machine learning algorithm. In the last development stage, we have applied an unsupervised machine learning to calibrate parameters and ranked the customers within an ordinal n-class scale obtained through the application of an unsupervised learning classification technique. Finally, we have verified the calibration goodness through classical calibration test (e.g. binomial tests). The study has been done on big data sample with more than 800,000 Retail customers of a European Bank under ECB Supervision, with 10 years of historical information and more than 600 variables to be analyzed for each customer.

Keywords Quantitative finance · Risk management · Credit risk · Machine learning · Big data

1 Introduction

The last 12 months have been characterized by a technological evolution and a "digital" wave offering new opportunities for the improvement of operational practices and the adoption of more advanced methodological approaches in different fields of research. In a context of growing competition and falling of profit margins,

S. Bonini · G. Caivano (✉)
Accenture Management Consulting, Milan, Italy
e-mail: stefano.bonini@accenture.com; giuliana.caivano@accenture.com

M. Corazza et al. (eds.), *Mathematical and Statistical Methods for Actuarial Sciences and Finance*, https://doi.org/10.1007/978-3-319-89824-7_32

Machine Learning can play a vital role in both technology and business, by enabling financial institutions to maximize the value of their own data. In the field of Credit Risk modeling, there is an extensive literature finding out a massive use of traditional statistical techniques in Probability of Default modeling. There are also lots of studies focused on the adoption of Machine Learning techniques in modeling credit risk parameters, highlighting different methodologies for estimating probability of default: artificial neural networks (as in [1]), discriminant analysis in [2], cluster analysis in [3], logistic regression (as in in [4–6]), support vector machines in [4, 7], classification trees in [8], random forests (in [9, 10]). Some of these have shown the advantages of using machine learning systems in credit scoring problems, and how they can achieve superior performance to the traditional techniques, e.g. Logistic Regression (as in [11, 12]). The emergence of these methods in open source libraries (such as R or Weka) and in proprietary software solutions (e.g. SAS) has made them widely available to the general population and to the lenders themselves.

Other studies provide benchmarks of machine learning classifiers on credit scoring datasets, but they are generally focused on rank ability, through the Gini score or AUROC, rather than on calibration. Despite the enormous volume of the related literature, our understanding is that the existing models that estimate probabilities of default (PD) need further development and more applications. In this context, our study wants to contribute to the understanding of these techniques and to demonstrate their effective contribute (compared to the traditional statistical methods) not only in ranking population but on the overall PD estimation process (variable selection, multivariate analysis and rating scale definition). As results section will show, the application of machine learning techniques ensures an improvement of modeling power not only in terms of ranking power (as already highlighted by the existing literature) but also in terms of running time and calibration power.

2 Methodology

We have divided the development process in three different stages: variable selection, multivariate analysis and rating scale definition applying different machine learning techniques for each stage:

a) Variable selection: Logistic regression;
b) Multivariate model: simulation with Classification trees, Neural Networks, Random Forests;
c) Parameters calibration and rating class: k-means cluster analysis.

2.1 Data Used

In this study, we have data for around 800,000 individual customers from January 2006 to December 2016 and more than 600 variables coming from three data main data sources information:

- Credit Bureau: we have built around 270 information variables (e.g. number of reporting entities, amount of expired installment, External Score, number of existing contracts, Total Expired Amount, Overdrawn amount etc.);
- Sociodemographic: we have built around 160 variables (e.g. Annual Income, Housing situation, Marital status, Activity Sector, Profession, Negative Notes, Banking Seniority etc.);
- Product (transaction level): we have defined 190 indicators as, e.g., Loan duration, Mortgage degree, Interest rate type, Funding Purpose, Product Type, Installment/Income etc.

2.2 Main Results

For *variable selection*, in line with the literature analyzed, we have performed a traditional logistic regression combined with an automated algorithm able to automatically apply to each variable a set of predefined constraints (Accuracy ratio higher than 10%, percentage of missing variables lower than 15% and coherence of coefficient sign with the expected trend between the variable and the default probability). Starting from the long list of 600 variables, we have identified a short list of 60 variables in a very reduced running time.

In order to identify the best multivariate combination, a benchmark analysis has been performed on the short list of variables comparing different machine learning techniques in terms of Accuracy Ratio (AR) and Correct Classification Rate (CCR), as shown in the Table 1:

Table 1 Machine learning techniques benchmark results

Machine learning techniques	AR (%)	CCR (%)
Classification tree	62	83
Random forest	65	84
Neural networks	56	81
Logistic regression	44	78

Starting from the ranking produced by classification tree model (comparable to random forest in terms of predictive power but much more interpretable from an economic point of view), we have finally calibrated the scores and clustered the customers into a final 7-class rating scale as the result of the application of k-means clustering (used for dividing a population with more than 800,000 observations into 50 clusters) and an automated algorithm able to test around 2 mln of possible 7-class scales for identifying the more predictive one in terms of calibration power (measured through binomial test).

3 Conclusion

The goal of this paper was to provide the evidences deriving from the application of different machine learning techniques to the different steps of a Probability of Default model estimation. Through the definition of an automated algorithm for variable selection process combining logistic regression with some economic constraints, we have identified a short list of 60 variables (from the initial 600 indicators) with a very limited running time. We have then identified different machine learning techniques for defining the best multivariate model, choosing the classification tree because of its higher predictive power but also its higher economic interpretability if compared e.g. to random forest. Finally, starting from the ranking score defined by classification tree, we have identified the best 7-class rating scale maximizing the calibration power of the model (measured by the pass of binomial test on all the scales) in line with the current regulation framework.

References

1. Zhao, Z., Xu, S., Kang, B.H., Kabir, M.J., Liu, Y., Wasinger, R.: Investigation and improvement of multi-layer perceptron neural networks for credit scoring. Expert Syst. Appl. **48**, 3508–3516 (2015)
2. McLachlan, G.: Discriminant Analysis and Statistical Pattern Recognition. Wiley Interscience, London (2004)
3. Steinbach, M., Tan, P.N.: kNN: k-nearest neighbors. In: Wu, X., Kumar, V. (eds.) The Top Ten Algorithms in Data Mining, pp. 151–162. Chapman & Hall/CRC, Boca Raton (2009)
4. Antunes, F., Ribeiroa, B., Pereira, F.: Probabilistic modeling and visualization for bankruptcy prediction. Appl. Soft Comput. **60**, 831–843 (2017)
5. Dwyer, D.W., Stein, R.M.: Inferring the default rate in a population by comparing two incomplete default databases. J. Bank. Financ. **30**, 797–810 (2006)
6. Han, L., Fraser, S., Storey, D.J.: Are good or bad borrowers discouraged from applying for loans? J. Bank. Financ. **33**, 415–424 (2009)
7. Yu, L., Yue, W., Wang, S., Lai, K.K.: Support vector machine based multiagent ensemble learning for credit risk evaluation. Expert Syst. Appl. **37**, 1351–1360 (2010)
8. Steinberg, D.: CART: classification and regression trees. In: Wu, X., Kumar, V. (eds.) The Top Ten Algorithms in Data Mining, pp. 180–201. Chapman & Hall/CRC, Boca Raton (2009)

9. Fonseca, P., Lopes, H.: Calibration of Machine Learning Classifiers for Probability of Default Modelling. James Finance, Crowd Process Inc. (2017)
10. Khandani, A.E., Kim, J., Lo, A.W.: Consumer credit-risk models via machine-learning algorithms. J. Bank. Financ. **34**, 2767–2787 (2010)
11. Lessmann, S., Baesens, B., Seow, H.V., Thomas, L.C.: Benchmarking state-of-the-art classification algorithms for credit scoring: an update of research. Eur. J. Oper. Res. **247**(1), 124–136 (2015)
12. Nanni, L., Lumini, A.: An experimental comparison of ensemble of classifiers for bankruptcy prediction and credit scoring. Expert Syst. Appl. **36**, 3028–3033 (2009)

Risk/Return Analysis on Credit Exposure: Do Small Banks Really Apply a Pricing Risk-Based on Their Loans?

Stefano Bonini and Giuliana Caivano

Abstract In the current economic scenario, an efficient and effective credit underwriting based on a pricing adjusted to the internal credit risk policies is a pillar for the existence of the banks. Despite the recent Regulator's indications and the advanced methodologies available, small banks prefer to adopt less structured for assessing the risk profile of their portfolio. Our study empirically analyzes a Small Bank portfolio of 16,216 loans (Retail and Corporate) underwritten in the last 5 years (2013–2017) and try to investigate the nature of the link between borrower credit worthiness and the effective price applied to the loans. The study finds out that the link between borrower credit worthiness and the price applied is opposite to what expected (the higher credit quality, the lower interest rate applied to loans).

Keywords Risk management · Credit risk · Pricing · Strategic policies · Credit underwriting · Credit policies · Financial crisis

1 Introduction

The financial crisis of 2007, the following actions of Regulator on regulatory framework, the structural decrease of interest rates and the *European Central Bank* (ECB) expansive monetary policy caused an impact on financial institutions both in terms of a deep reduction of banks profitability and a change in their business model. Consequently, in 2016 the cost-to-income ratio of the leading European banks has been mainly leaded by direct and indirect remuneration payments. These data highlight how the banking system is trying to readapt its strategic choices to the changed market conditions, supporting a very onerous framework of fixed costs but, at the same time, continuing credit underwriting. In this context, an efficient and effective credit underwriting based on a pricing adjusted to the internal credit risk policies is a pillar for the existence of the banks.

S. Bonini (✉) · G. Caivano
Accenture Management Consulting, Milan, Italy
e-mail: stefano.bonini@accenture.com; giuliana.caivano@accenture.com

M. Corazza et al. (eds.), *Mathematical and Statistical Methods for Actuarial Sciences and Finance*, https://doi.org/10.1007/978-3-319-89824-7_33

Despite the recent Regulator's indications and the advanced methodologies available, small banks struggle to equip themselves with complex systems for risk measurement and choose to adopt less structured methods which are anyway capable of providing an adequate risk assessment. The relation between borrower creditworthiness and pricing of loans has been investigated by few studies (mainly referred to past years, 2004–2011). Most of them highlights the existence of a negative correlation between credit quality and pricing of loans (the higher is the credit quality of the borrower, the lower is the final interest rate applied by lender). These studies are almost concentrated on US Market (as in [1] and related in particular on mortgage (as in [2, 3]) and household's portfolios of commercial banks (as in [4–6]).

This study provides an added value to the existing literature enlarging the empirical evidences on European Market and focusing on different portfolios (Retail and Corporate) and different products (Mortgages, Personal Loans, Cash and Self-liquidating), taking also into account the presence of credit risk mitigatory instruments.

The outcomes of our analysis highlight the presence of some differences and incoherence between borrower credit worthiness and the price applied, mainly due to a misalignment between strategic choices and risk management, actually the bank in order to gain new clients is underpricing its loans and then not completely remunerate the risk taken.

2 Methodology

As a first research step, we have adopted an empirical approach based on a descriptive analysis with the aim of preliminarily verify the traditional idea that the riskier is the client, the higher is the loan price, performing also some data mining in order to have a consistent data base to be analysed.

2.1 Data Used

In this study, we have 5-years data from 01/01/2012 to 31/12/2016 for a total of 16,216 observations with general counterparties information and internal performance data: turnover and total budget, industry sector, industry branch, exposure practice number, number of the lending decision process practice, amount of the credit, expiration date, date of credit underwriting, description of the destination of the loan, type of interest rate (fixed or variable), presence and type of collateral (mortgage loan, personal guarantees, etc.) and Rating computed internally.

In order to carry out a risk-adjusted analysis, we decided to divide the portfolio in three classes by client type: Corporates (companies, small enterprises and professionals), Public Entities (Public institutions or associations) and Individuals (physical subjects).

Moreover, the clients were divide by type of product and collateral: Installment—Unsecured, Installment—Mortgage Secured, Installment—Personal Guarantees, Cash—Unsecured, Cash—Secured, Self-Liquidating—Unsecured, Self-Liquidating—Secured.

Each portfolio has been analyzed by the six rating classes internally used: from the lower to the higher risk level.

2.2 Results

Starting from the data, a quantitative analysis by counterparty and product has been performed. The results are shown in the following graphs.

For all the products Individuals and Corporate have been analyzed, on the other hand, due to its nature, public institutions have not been analyzed only for some products.

3 Conclusion

The paper analyses the credit portfolio of a medium European bank that is not using Validated *Advanced Internal Rating Based* (AIRB) models to evaluate and price the credit worthiness of its clients. According to the analysis it has been shown that the price is mainly uncorrelated with the risk rating classes and is generally common for each type of product and collateral.

Actually, analyzing Figs. 1, 2, 3, 4, 5 and 6 it is possible to observe that the relationships risk/price are often almost linear between class 1 and class 6.

The analysis has shown that there are distortions between the client risk and effective price applied, which can be motivated only by commercial policies that

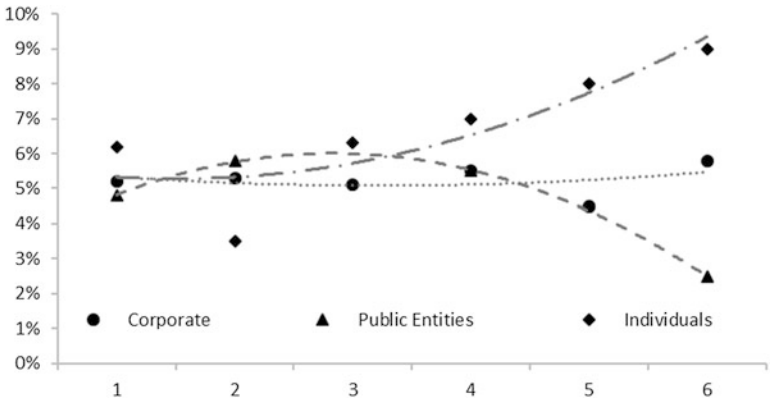


Fig. 1 Installment—Unsecured

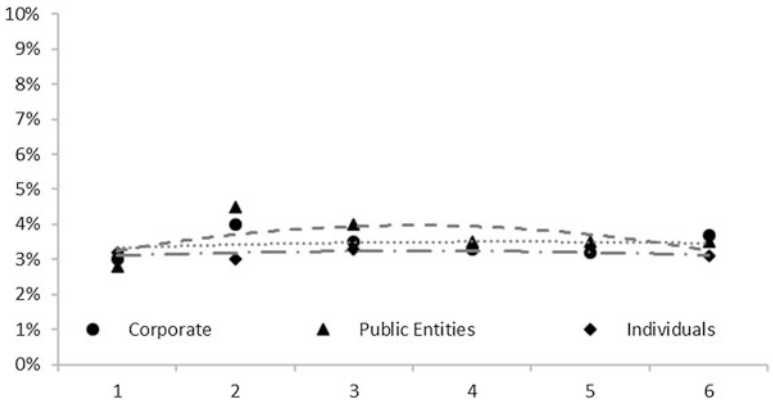


Fig. 2 Installment—Mortgage Secured

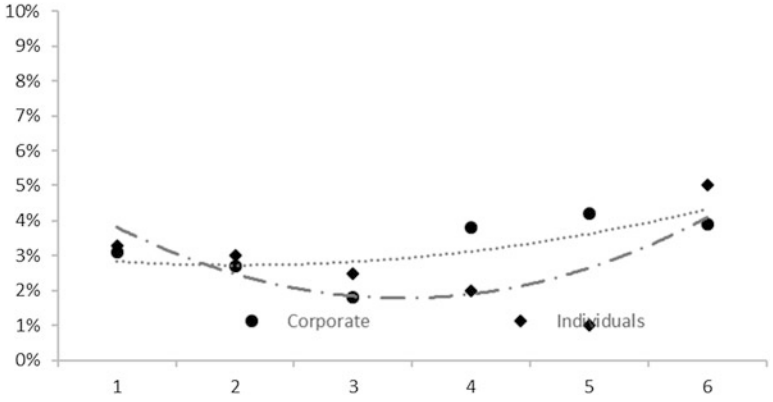


Fig. 3 Installment—Personal Guarantees

are not in line with the risk management best practices and are probably caused by many critical issues regarding the correct measurement of client credit worthiness, as well as by an inadequate remuneration policy during the pricing phase.

These empirical evidences suggest that if the banks do not have well developed risk management analytical instruments, they will continue to adopt a commercial strategy to acquire new clients from other banks with aggressive pricing policy without any risk profile adjustment.

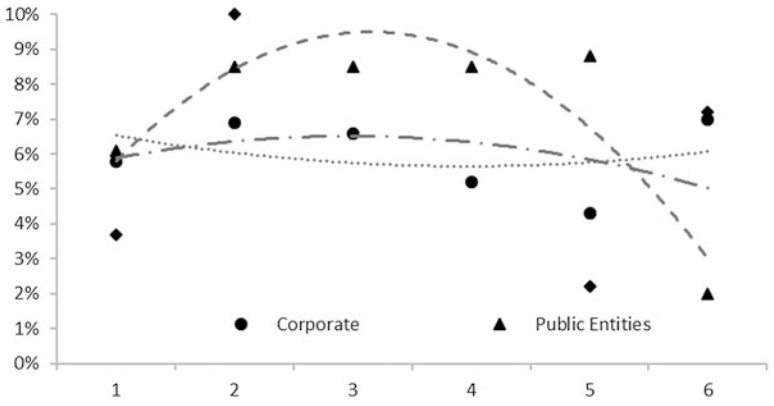


Fig. 4 Cash—Secured

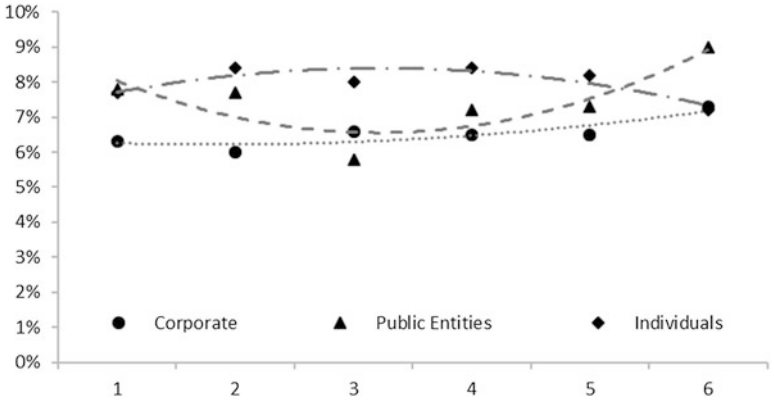


Fig. 5 Cash—Unsecured

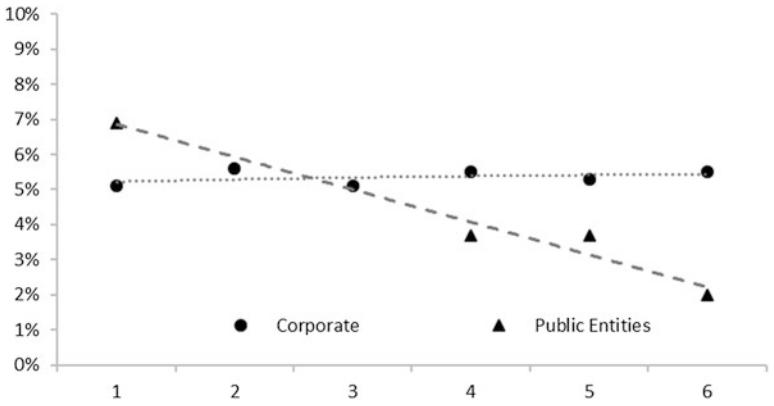


Fig. 6 Self-Liquidating—Secured

References

1. Getter, D.E.: Consumer credit risk and pricing. J. Consum. Aff. **40**, 41–63 (2006)
2. Al-Bahrani, A., Su, Q.: Determinants of mortgage pricing: a quantile regression analysis. J. Hous. Econ. **30**, 77–85 (2015)
3. Magri, S., Pico, R.: The rise of risk-based pricing of mortgage interest rates in Italy. J. Bank. Financ. **35**, 1277–1290 (2011)
4. Asquith, P., Beatty, A., Weber, J.: Performance pricing in bank debt contracts. J. Account. Econ. **40**, 101–128 (2005)
5. Deng, Y., Gabriel, S.: Risk-based pricing and the enhancement of mortgage credit availability among underserved and higher credit-risk populations. J. Money. Credit. Bank. **38**, 1431–1460 (2006)
6. Edelberg, W.: Risk-Based Pricing of Interest Rates in Household Loan Markets, vol. 62. FEDS Working Paper (2003)

Life Insurers' Asset-Liability Dependency and Low-Interest Rate Environment

Nicola Borri, Rosaria Cerrone, Rosa Cocozza, and Domenico Curcio

Abstract In this paper we study the relationships between life insurers' assets and liabilities and investigate how it evolved during the most recent years of unprecedented low interest rates. We use a canonical correlation analysis to measure the relationships among, and between, asset and liability accounts for the main EU life insurers in the years 2007, 2011 and 2015. We find strong and substantial evidence that assets and liabilities have become more independent over time. We argue that the declining trend of market interest rates has contributed to the generalized reduction in the linkage between the asset side and the liability side of EU life insurers, leaving them more exposed to ALM-related risks relative to the period before the financial crisis.

Keywords Insurance companies · Interest rates · Asset-liability management · Canonical correlation analysis

N. Borri
Department of Economics and Finance, LUISS University & CASMEF, Rome, Italy
e-mail: nborri@luiss.it

R. Cerrone
Department of Business Science–Management & Innovation Systems, University of Salerno, Fisciano, Italy
e-mail: rocerro@unisa.it

R. Cocozza
Department of Economics, Management, Institutions, University of Naples "Federico II", Naples, Italy
e-mail: rosa.cocozza@unina.it

D. Curcio (✉)
Department of Economics, Management, Institutions, University of Naples "Federico II" & CASMEF, Naples, Italy
e-mail: domenico.curcio@unina.it

M. Corazza et al. (eds.), *Mathematical and Statistical Methods for Actuarial Sciences and Finance*, https://doi.org/10.1007/978-3-319-89824-7_34

1 Introduction

This paper studies the effects of the declining and extremely low market interest rates on the asset and liability management (ALM) strategies of the main EU life insurers over the period 2007–2015. We examine the internal structure of the ALM–related decisions and use the canonical correlation analysis, introduced by Hotelling [4], to detect the relationships between life insurers' assets and liabilities and to investigate the evolution of these relationships during the recent past. We find evidence that insurers' assets and liabilities have indeed become more independent over time. We argue that the downward trend and the low level of market interest rates have contributed to the generalized reduction in the linkage between the asset and the liability sides of EU life insurers, leaving insurance companies more exposed to ALM-related risks.

The mean-variance approach provides a simple explanation of financial intermediation, where securities positively held can be considered as assets and those negatively held as liabilities. Pyle [5] and Francis [3] use a mean-variance model to study the optimal balance sheet structure of financial intermediaries. Stowe [7] applies a similar framework to life insurers. Consistently, we use canonical correlations to explicitly incorporate in our analysis dependencies between assets and liabilities. Simonson et al. [6] and DeYoung and Yom [2] use canonical correlation analysis to study US banks' asset-liability hedging decisions. Within the insurance literature, Stowe and Watson [8] argue that life insurers solve a portfolio optimization problem when structuring their assets and liabilities and highlight several significant cross-balance sheet relationships for a cross-section of 194 US large life insurers.

The remainder of this paper is organised as follows: Sect. 2 describes the data and methodology; Sect. 3 presents the empirical results; and Sect. 4 concludes.

2 Data and Methodology

We analyze a representative sample of 24 EU life insurers between 2007 and 2015, using year-end data from companies' balance sheets taken by the OSIRIS database [9]. We focus on three separate cross sections of data in 2007, 2011 and 2015, where each set of annual calculations is independent from the others. Examining the data in 4-year intervals gives us sufficient time between observations for asset-liability relationships to react (or to not react) to changes of market interest rates. In order to consider stable balance sheets, we include in our sample only companies that have been in the database for at least 10 years.

Canonical correlation is a multivariate analysis technique that we use to examine the relationships between the asset and liability/capital accounts of an insurers' balance sheet, respectively denoted by the matrices X and Y. The number of rows of each matrix represents the n insurers of our sample, while the number of columns

indicates the different categories of asset (q_1) and liability (q_2) taken into account and expressed as a proportion of total assets. The canonical correlation methodology attempts to find linear combinations of X and Y so that the correlation between them is as high as possible. The linear combinations of X and Y are denoted, respectively, by u_i and v_i:

$$u_i = Xa_i \quad v_i = Yb_i \quad i = 1, \ldots, p \quad \text{with} \quad p = \min(q_1, q_2)$$

where a_i and b_i are vectors to be estimated and have dimension, respectively, $(q_1 \times 1)$ and $(q_2 \times 1)$. We refer to the scalars that form the vectors as canonical coefficients, to the linear combinations of X and Y as canonical variables, and to the correlations between the canonical variables as canonical correlation coefficients. Since canonical correlation coefficients represent the variance shared by linear combinations of assets and liabilities, and not the variance shared by the original asset and liability accounts, a high correlation between only one asset variable and only one liability variable might lead to a very large canonical correlation coefficient. In order to address this issue and further investigate the links between asset and liability accounts, we calculate the redundancy coefficients that provide a measure of the average ability of asset (liability) variables, taken as a set, to explain variation in liability (asset) variables taken one at a time.

3 Empirical Results

We summarize our results in Table 1. Asset and liability variables exhibit a relatively high degree of collective dependence, even if we observe a declining trend in the strength of the asset-liability relationship from 2007 to 2015 (i.e., the number of statistically significant canonical correlations goes from 5 in 2007 to 2 in 2015).

Table 1 Canonical correlations

	2007		2011		2015	
Roots	Correlation	F-stat	Correlation	F-stat	Correlation	F-stat
1 through 5	0.99***	10.59	0.99***	5.95	0.99***	7.57
2 through 5	0.93***	4.26	0.89***	2.60	0.94***	3.16
3 through 5	0.83***	2.90	0.79*	1.75	0.77	1.42
4 through 5	0.65***	2.18	0.53	0.89	0.44	0.57
5 through 5	0.55*	0.06	0.00		0.00	

Notes: This table reports the values of the canonical correlation coefficients. The F-statistic tests whether there is any association between the p pairs of canonical variables. Note: ***, * = significance level of 1% and 10%, respectively, using Bartlett's Chi-square test [1]

Then, we calculate the proportion of variance coefficients and the redundancy coefficients.[1] In 2007, about 55.2% (60.5%) of the variation in the actual assets (liabilities) data is explained by the asset (liability) canonical variables in the first three loadings whereas, in 2015, 50.3% (66.7%) of the variation in the actual assets (liabilities) is explained by the asset (liability) canonical variables in the first loading. All else equal, the relationships among the various asset and liability accounts seem to become less complex over time, since we observe a reduction in the number of significant loadings and, accordingly, an increase in the share of the variance explained by the first loadings. As to the redundancy coefficients, in 2007, the liability canonical variables explain 51.45% of the variation in the asset variables, while the asset canonical variables explain 67.4% of the variation in the liability variables. Those figures drop to 34.2% and 57%, respectively, in 2015. Furthermore, also the number of significant canonical loadings decrease from 3 in 2007 to 2 in 2015. We interpret these results as follows. First, causation runs more strongly from assets to liabilities than from liabilities to assets (i.e., insurers are pools of deposits looking for investing opportunities). Second, despite the relatively large size of the redundancy coefficients, the importance of the first loadings suggests that a relatively small number of relationships among individual asset and liability accounts drives the strong canonical correlations.[2]

4 Conclusions

The declining trend and the low level of market interest rates might have contributed to the generalized reduction in the linkage between the asset and liability side of EU life insurers, and have made them more exposed to ALM-related risks than they were in the period before the financial crisis broke out. Further investigation on the relations between insurer assets and liabilities is crucial from both a regulatory and supervisory perspective, since it might help to define qualitative and quantitative measures of liquidity requirements more consistent with insurers' behaviour during both benign market conditions and stressed financial markets.

[1] Details on the calculations of the proportion of variance coefficients and the redundancy coefficients are available upon request.

[2] For robustness, and following [2], we also investigate whether asset-liability dependency changes with insurers' size. Specifically, we repeat the analysis on two subsamples containing life insurers from the first and bottom quartile of the average size distribution and find that results are qualitatively unchanged.

References

1. Bartlett, M.S.: A note on the multiplying factors for various χ 2 approximations. J. R. Stat. Soc. Ser. B Methodol. **16**, 296–298 (1954)
2. DeYoung, R., Yom, C.: On the independence of assets and liabilities: evidence from US commercial banks, 1990–2005. J. Financ. Stab. **4**, 275–303 (2008)
3. Francis, J.C.: Portfolio analysis of asset and liability management in small-, medium-and large-sized banks. J. Monet. Econ. **4**, 459–480 (1978)
4. Hotelling, H.: The most predictable criterion. J. Educ. Psychol. **26**, 139 (1935)
5. Pyle, D.H.: On the theory of financial intermediation. J. Financ. **26**, 737–747 (1971)
6. Simonson, D.G., Stowe, J.D., Watson, C.J.: A canonical correlation analysis of commercial bank asset/liability structures. J. Financ. Quant. Anal. **18**, 125–140 (1983)
7. Stowe, J.D.: Life insurance company portfolio behavior. J. Risk Insur. **XLV**(1), 431–447 (1978)
8. Stowe, J.D., Watson, C.J.: A multivariate analysis of the composition of life insurer balance sheets. J. Risk Insur. **52**(2), 222–240 (1985)
9. Van Dijk, B.: Osiris Database. Bureau Van Dijk, Amsterdam (2015)

Modelling the Australian Electricity Spot Prices: A VAR-BEKK Approach

Manuela Braione and Davide De Gaetano

Abstract This paper investigates the transmission of spot electricity prices and price volatility among the five Australian regional electricity markets. In particular, VAR(k)-BEKK(p, q) models with optimized lag lengths and different distributional assumptions are analysed. Empirical results suggest that a VAR(3)-BEKK(1,2) under Student-*t* assumption can better describe the complex dynamics between the markets.

Keywords Spot electricity prices · Mean and volatility spillovers · Multivariate volatility

1 Introduction and Motivation

This paper investigates the transmission of spot electricity prices and price volatility across the Australian regional electricity markets in New South Wales, Queensland, South Australia, Tasmania and Victoria. Being a highly volatile and significantly more spike-prone market than many comparable systems (see [4]), it has represented a topic of great interest in the past decade. Nevertheless, only a limited number of studies concentrated on the analysis of the market dynamics using a multivariate approach. In this respect, pioneer work has been carried out by [5] who proposed a joint approach consisting in a VAR(1) model for the conditional mean of the series

The views expressed in the article are those of the authors and do not involve the responsibility of SOSE Soluzioni per il Sistema Economico S.p.A.

M. Braione · D. De Gaetano (✉)
SOSE - Soluzioni per il Sistema Economico S.p.A., Roma, Italy
e-mail: mbraione@sose.it; ddegaetano@sose.it

M. Corazza et al. (eds.), *Mathematical and Statistical Methods for Actuarial Sciences and Finance*, https://doi.org/10.1007/978-3-319-89824-7_35

and a symmetric BEKK(1,1)[1] model for the dynamics of the conditional covariance matrix of the residuals. They found positive own-mean spillovers in only a small number of markets with the absence of mean spillovers among the others, along with significant ARCH and GARCH effects across markets. Recently, there have been significantly changes in the structure of the National Electricity Market (NEM) which is currently experiencing one of the most challenging periods in the energy sector history (see [1]). As such, the market has been dramatically volatile, calling for a better understanding of its new interactions. To this aim, we estimate a VAR(k) model coupled with a bunch of unrestricted BEKK(p, q) specifications under both normal and Student-t assumptions to find the preferred one to better capture the occurred changed market conditions.

2 Methods and Results

The data consists of daily spot electricity prices running over the period February 8, 2013 to November 30, 2017. The price series are obtained from the Australian Energy Market Operator and are calculated as daily arithmetic means over the 48 intra-day trading intervals. Similarly to [5], both the null hypotheses of normality and of non-stationarity can be rejected for all price series. In order to model the dynamics of the markets, we employ a two step procedure. First, the conditional mean process is modeled through a VAR(k) model accommodating a dummy variable γ_t to capture the day-of-the-week effect[2]. The augmented VAR(k) model is written as

$$\boldsymbol{P}_t = \boldsymbol{\alpha} + \sum_{i=1}^{k} \mathsf{A}_i \boldsymbol{P}_{t-i} + \gamma_t \boldsymbol{\Gamma} + \boldsymbol{\epsilon}_t, \tag{1}$$

with $\boldsymbol{\alpha}$ being the $(n \times 1)$ vector of long-term drift coefficients, A the $(n \times n)$ matrix whose elements measure the own and cross-mean spillovers, $\boldsymbol{\Gamma}$ the $(n \times 1)$ vector of dummy coefficients and $\boldsymbol{\epsilon}_t$ the $n-$dimensional white noise process. The optimal model according to the BIC is a VAR(3), reported in Table 1. It highlights a deep net of connections among markets with significant own- and cross-mean spillovers. Moreover, all the constant terms are significant and, similar to [3], weekend and public holidays effects are significant and negative in all five markets indicating that Saturday, Sunday and public holiday electricity prices are lower than weekday prices. In the second step, a set of full BEKK(p,q) models with lag order values ranging from (1,1) to (2,2) are fitted to the obtained standardized

[1] For a deep understanding of the BEKK-type specifications refer to [2].

[2] The dummy γ_t assumes value one if day t is a public holiday or a weekend and value zero otherwise.

Table 1 Estimated parameters and corresponding standard errors of the VAR(3) model

	NSW		QLD		SA		TAS		VIC	
	Coef.	Std.err	Coef.	Std.err	Coef.	Std.err	Coef.	Std.err	Coef.	Std.err
α	17.171**	(1.670)	15.791**	(2.348)	33.718**	(3.453)	8.065**	(1.109)	23.114**	(1.940)
A_1'	0.345**	(0.018)	0.180**	(0.026)	−0.052	(0.038)	0.006	(0.012)	0.047**	(0.021)
	0.036**	(0.013)	0.294**	(0.018)	−0.017	(0.027)	0.008	(0.009)	0.003	(0.015)
	0.186**	(0.010)	0.095**	(0.014)	0.312**	(0.020)	0.018**	(0.006)	0.134**	(0.011)
	−0.036	(0.025)	0.020	(0.036)	0.039	(0.052)	0.684**	(0.017)	0.026	(0.029)
	−0.038**	(0.018)	−0.025	(0.025)	−0.016	(0.036)	0.022*	(0.012)	0.133**	(0.020)
A_2'	0.001	(0.019)	0.005	(0.027)	0.110**	(0.039)	0.010	(0.013)	0.013	(0.022)
	−0.011	(0.013)	0.055**	(0.019)	0.041	(0.028)	0.004	(0.009)	0.006	(0.016)
	0.057**	(0.011)	0.033**	(0.015)	0.025	(0.022)	−0.004	(0.007)	−0.052**	(0.012)
	0.082**	(0.031)	0.029	(0.043)	0.039	(0.064)	−0.198**	(0.020)	0.025	(0.036)
	−0.027	(0.018)	−0.020	(0.025)	0.003	(0.037)	−0.017	(0.012)	0.065**	(0.021)
A_3'	0.038**	(0.017)	0.010	(0.024)	0.019	(0.036)	−0.008	(0.011)	0.016	(0.020)
	0.034**	(0.013)	0.071**	(0.018)	0.027	(0.027)	0.011	(0.009)	0.016	(0.015)
	−0.057**	(0.011)	−0.029**	(0.015)	0.017	(0.022)	−0.016**	(0.007)	0.005	(0.012)
	−0.008	(0.025)	0.029	(0.035)	0.008	(0.052)	0.352**	(0.017)	0.035	(0.029)
	0.059**	(0.017)	0.003	(0.025)	0.019	(0.036)	−0.002	(0.012)	0.038*	(0.020)
Γ	−8.432**	(1.857)	−9.019**	(2.610)	−21.103**	(3.839)	−3.683**	(1.233)	−10.250**	(2.156)

Notes: NSW is New South Wales, QLD is Queensland, SA is South Australia, TAS is Tasmania, VIC is Victoria. A_i' is the transpose of A_i. Asterisks ** and * denote significance at the 5% and 10% level, respectively

residuals, under both normal and Student-t distribution assumptions. The model is written as:

$$H_t = BB' + \sum_{i=1}^{p} C_i \epsilon_{t-i} \epsilon'_{t-i} C'_i + \sum_{j=1}^{q} G_j H_{t-j} G'_j, \tag{2}$$

where C_i $(i = 1, \ldots, p)$, G_j $(j = 1, \ldots, q)$ are unrestricted square matrices and B is a triangular matrix to ensure positive definiteness of H_t.

From Table 2 it emerges that for each lag order (p, q), the inclusion of the Student-t assumption significantly increases the value of the maximized log-likelihood function, indicative for a much better fit. Moreover, the largest eigenvalues of the matrices $\psi = |\sum_{j=1}^{max(p,q)} (C_j \otimes C_j) + (G_j \otimes G_j)|$, obtained for the Student-t class of models, are all less than one and smaller than under the normal assumption, thus suggesting that allowing for heavy-tails has significant effects on the dynamic structure of the BEKK. As emerges, the BIC-preferred specification is the Student-t BEKK(1,2). Empirical estimates and standard errors are summarized in Table 3. The estimates for the variance equation show significant ARCH coefficients along the main diagonal of the C_1 matrix, suggesting that volatility is persistent in all markets. In particular, VIC has the most persistent ARCH effect of 0.603^2. The off diagonal elements of C_1 evidence some significant spillover ARCH effects as well. In particular, the volatility of SA is positively affected by past shocks in the NSW and VIC markets, while is negatively affected by past shocks in the (NSW, SA) and (SA, VIC) markets. The main diagonal coefficients of the G_1 and G_2 matrices indicate that there are statistically significant own-GARCH effects only at lag one, except in the TAS market where the total effect is of $0.593^2 + 0.373^2$. Regarding spillover GARCH effects, these are mainly seen for the NSW market, which directly influences the volatility of QLD $(0.042^2 + 0.093^2)$, SA (0.221^2) and TAS $(0.093^2 + 0.110^2)$, thus being the main volatility transmitter within the national industry. On the other hand, it is found that while SA and TAS only take volatility without transmitting it, VIC only responds to its own past volatility changes (0.484^2).

Overall, the VAR(3)-BEKK(1,2) model uncovers a number of significant interactions among the five markets, including spillovers from surprise price changes in one market to the volatility of another market. Thanks to the long span data set and the powerful model structure, we are able to not only detect these effects, but also estimate their magnitude.

Table 2 Optimized Log-likelihood and diagnostics for the estimated BEKK(p,q) models

					p-Value for the Ljung-Box test for the residuals; 50 lags													
(p,q)	N.pars	Loglik	BIC	MaxEig	ϵ_{11}	ϵ_{12}	ϵ_{13}	ϵ_{14}	ϵ_{15}	ϵ_{22}	ϵ_{23}	ϵ_{24}	ϵ_{25}	ϵ_{33}	ϵ_{34}	ϵ_{35}	ϵ_{44}	ϵ_{45}	ϵ_{55}
BEKK(p,q) Student-t assumption																			
(1,1)	66	−62391	125316	0.910	**0.00**	**0.00**	1.00	1.00	1.00	**0.00**	1.00	1.00	**0.00**	**0.00**	**0.00**	1.00	**0.00**	1.00	0.04
(1,2)	91	−61605	123947	0.679	**0.00**	1.00	1.00	1.00	1.00	**0.00**	1.00	**0.00**	0.90	**0.00**	1.00	0.98	**0.00**	**0.00**	0.07
(2,1)	91	−61733	124202	0.643	**0.00**	**0.00**	**0.00**	1.00	**0.00**	**0.00**	1.00	1.00	1.00	**0.00**	1.00	1.00	**0.00**	**0.00**	0.75
(2,2)	116	−61558	124054	0.895	**0.00**	0.92	**0.00**	1.00	**0.00**	**0.00**	1.00	0.12	**0.00**	**0.00**	1.00	0.80	**0.00**	**0.00**	0.88
Full BEKK(p,q) normal assumption																			
(1,1)	65	−80719	161963	0.965	**0.00**	1.00	**0.00**	1.00	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	0.89	1.00	**0.00**	**0.00**	1.00
(1,2)	90	−80068	160864	0.946	**0.00**	**0.00**	0.94	0.29	0.89	**0.00**	0.13	1.00	**0.00**	**0.00**	**0.00**	1.00	**0.00**	1.00	1.00
(2,1)	90	−79325	159378	0.980	**0.00**	**0.00**	1.00	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	1.00	1.00
(2,2)	115	−77688	156305	0.929	**0.00**	**0.00**	0.63	1.00	1.00	**0.00**	1.00	**0.00**	**0.00**	**0.00**	1.00	1.00	**0.00**	**0.00**	1.00

Notes: the table reports, in order, the lag order (p,q), the number of model parameters (N.pars), the maximized log-likelihood value (Loglik), the BIC information criteria and the largest eigenvalues of the estimated matrices ψ; bold p-values indicate significance at the 1% level

Table 3 Estimated parameters and corresponding standard errors of the BEKK(1,2) under Student-t assumption

	NSW		QLD		SA		TAS		VIC	
	Coef.	Std.err	Coef.	Std.err	Coef.	Std.err	Coef.	Std.err	Coef.	Std.err
B	4.060**	(1.259)								
	5.002**	(1.349)	−0.812	(1.966)						
	15.541**	(0.924)	1.887	(5.952)	1.883	(2.623)				
	2.101**	(0.544)	−0.211	(0.281)	−0.294	(0.560)	2.815**	(0.351)		
	5.606**	(1.104)	0.304	(0.274)	−1.351	(0.968)	−0.043	(0.188)	−1.135	(2.257)
C_1	0.569**	(0.073)	0.037	(0.067)	−0.044	(0.095)	0.010	(0.022)	−0.061	(0.074)
	−0.013	(0.023)	0.546**	(0.093)	0.004	(0.019)	0.004	(0.028)	−0.004	(0.014)
	−0.093**	(0.045)	−0.057	(0.046)	0.321**	(0.059)	−0.004	(0.025)	−0.063**	(0.032)
	0.010	(0.007)	−0.003	(0.022)	−0.011	(0.009)	0.436**	(0.033)	−0.013	(0.015)
	0.017	(0.025)	−0.026	(0.031)	0.145	(0.142)	−0.011	(0.029)	0.603**	(0.058)
G_1	0.314*	(0.176)	0.063	(0.099)	−0.194	(0.141)	0.010	(0.023)	0.055*	(0.029)
	0.042**	(0.021)	0.484**	(0.146)	0.021	(0.016)	0.006	(0.014)	0.012	(0.008)
	0.052	(0.149)	0.104	(0.065)	0.345**	(0.101)	0.012	(0.030)	0.011	(0.036)
	−0.099**	(0.026)	−0.011	(0.032)	−0.018	(0.041)	0.593**	(0.051)	0.002	(0.020)
	0.127	(0.156)	−0.044	(0.093)	0.094	(0.165)	−0.048	(0.069)	0.484**	(0.096)
G_2	−0.014	(0.018)	−0.093	(0.106)	0.242	(0.293)	−0.003	(0.016)	0.060	(0.075)
	−0.093**	(0.033)	0.044	(0.130)	−0.046	(0.039)	−0.009	(0.030)	−0.027*	(0.014)
	0.221**	(0.026)	0.083**	(0.033)	0.121	(0.178)	0.013	(0.009)	0.043	(0.043)
	0.110**	(0.038)	0.058	(0.047)	0.046	(0.085)	0.373**	(0.068)	0.034	(0.040)
	−0.154	(0.137)	−0.118	(0.153)	0.096	(0.141)	0.039	(0.043)	0.172	(0.132)
ν	3.395**	(0.462)								

Notes: NSW is New South Wales, QLD is Queensland, SA is South Australia, TAS is Tasmania, VIC is Victoria. Asterisks ** and * denote significance at the 5% and 10% level, respectively

References

1. Australian Energy Regulator (AER): state of the energy market 2017 (2017)
2. Engle, R.F., Kroner, K.F.: Multivariate simultaneous generalized ARCH. Econometric. Theory **11**(1), 122–150 (1995)
3. Higgs, H.: Modelling price and volatility inter-relationships in the Australian wholesale spot electricity markets. Energy Econ. **31**(5), 748–756 (2009)
4. Worthington, A.C., Higgs, H.: Stochastic price modelling of high volatility, mean-reverting, spike-prone commodities: the Australian wholesale spot electricity market. Energy Econ. **30**(6), 3172–3185 (2008)
5. Worthington, A.C., Kay-Spratley, A., Higgs, H.: Transmission of prices and price volatility in Australian electricity spot markets: a multivariate GARCH analysis. Energy Econ. **27**(2), 337–350 (2005)

Cyber Risk Management: A New Challenge for Actuarial Mathematics

Maria Francesca Carfora, Fabio Martinelli, Francesco Mercaldo, Albina Orlando, and Artsiom Yautsiukhin

Abstract A specific kind of insurance that is emerging within the domain of cyber-systems is that of cyber-insurance. It allows for transferring the residual risk associated with network and computer incidents to a third party. Insurance companies are increasingly offering such policies, in particular in the USA, but also in Europe. The emerging trends in cyber insurance raise a number of unique challenges and force actuaries to reconsider how to think about underwriting, pricing and aggregation risk. Aim of this contribution is to offer a review of the recent literature on cyber risk management in the actuarial field. Moreover, basing on the most significant results in IT domain, we outline possible synergies between the two lines of research.

Keywords Risk management · Cyber risk · Cyber insurance

1 Introduction

The Internet evolution is one of the greatest innovations of the twentieth century and has changed lives of individuals and business organizations. As a consequence, cyber risk has emerged as one of the top challenges faced by companies worldwide. Executives and security professionals are accepting that it is not a matter of if but a matter of when their organization will be hit by a cyber-attack. Companies have to include cyber risk in their risk management framework, depicting their risk profile, assessing their risk appetite and looking for corresponding risk transfer solutions. Cyber-security insurance is designed to mitigate losses from a variety of cyber

M. F. Carfora · A. Orlando (✉)
Istituto per le Applicazioni del calcolo "M. Picone" - Consiglio Nazionale delle Ricerche, Napoli, Italy
e-mail: f.carfora@na.iac.cnr.it; a.orlando@na.iac.cnr.it

F. Martinelli · F. Mercaldo · A. Yautsiukhin
Istituto di Informatica e Telematica - Consiglio Nazionale delle Ricerche, Pisa, Italy
e-mail: fabio.martinelli@iit.cnr.it; francesco.mercaldo@iit.cnr.it; artsiom.yautsiukhin@iit.cnr.it

M. Corazza et al. (eds.), *Mathematical and Statistical Methods for Actuarial Sciences and Finance*, https://doi.org/10.1007/978-3-319-89824-7_36

incidents, including data breaches, data theft, business interruption and network damage. In general, immense difficulties emerge to insure cyber risk, especially due to the lack of data and modelling approaches, the risk of changes and the accumulation risks. Scientific interest on this topic is growing, but despite the increasing relevance for businesses, at present research on cyber risk is still limited. Many papers can be found in the IT domain, but relatively little research has been done in the actuarial, business and economic literature. Considering that cyber crime damage costs are expected to hit six trillion annually by 2021,[1] it is plausible that there is an increasing demand for methodologies with the aim to quantify the risk of cyber attacks exposure by industrial and public companies. This new scenario calls for bridging the gap between actuarial, economic and IT domain in order to address this increasing demand. Aim of the paper is to outline the peculiarities of cyber insurance and to show the main recent results in actuarial literature. Finally the interdisciplinarity of the topic is stressed together with the suggestion to look at the results in IT domain.

2 Peculiarities of Cyber Insurance

The main issues related to cyber insurance can be summarized as follows [6]:

- *Evolution of information system*: the information system of an organisation may easily change and new technologies appear, modifying the landscape of cyber risks;
- *Information asymmetry*: there are many obstacles for an insurer to get reliable information about the risk exposure of an insured and it is difficult to know if this exposure will be maintained during the whole period of policy operation;
- *Evolution of attacks*: it is very hard to determine the rate of occurrences and, as a consequence, the assessment of risk exposure;
- *Interdependence of security*: security level of an information system may depend on security of others;
- *Impact determination*: damage for cyber risks is very hard to estimate in advance because of the intangible nature of information assets. Moreover reputation cost, which accounts for a large portion of the whole damage, is very difficult to estimate;
- *Lack of statistical data*: data lie at the center of any actuarial project, but data are very limited in this field. Companies often do not want to reveal breaches, since they cause secondary damage , e.g. to reputation.
- *Premium estimation*: unlike traditional insurance policies, cybersecurity insurance has no standard scoring systems or actuarial tables for rate making. Moreover geographical similarities and monoculture make the task very hard.

[1] https://www.csoonline.com/article/3153707/security/top-5-cybersecurity-facts-figures-and-statistics-for-2017.html.

2.1 Some Recent Results in Economic and Actuarial Literature

A very interesting contribution is given in [2] where the authors aim to provide an overview of the main research topics in the emerging fields of cyber risk and cyber risk insurance. The results illustrate the immense difficulties to insure cyber risk and various ways to overcome the insurability limitations are discussed. The authors illustrate where research stands currently and outline directions for the future.

Regarding modelling and pricing cybersecurity risk, in [5] the authors propose a model consisting of three components: epidemic models, loss functions and premium strategies. A simulation approach is proposed to compute the premium for the cybersecurity risk for practical use. Herath and Herath [3] develop a cyber-insurance model using the emerging copula methodology. This approach is the first in the information security literature to integrate standard elements of insurance risk, with the robust copula methodology to determine cyber insurance premiums.

As regards the lack of data, Eling and Loperfido [1] link what has been done in information technology field with the current discussion on goodness of fit, pricing and risk measurement in actuarial domain. They analyze the data breach information taken from "chronology of Data Breaches" provided by the Privacy Rights Clearing house and use multidimensional scaling and goodness-of-fit tests, to analyze the distribution of data breach information. They illustrate the usefulness of their results in two applications on risk measurement and pricing.

3 Information Technology for Cyber Insurance

Considering the cyber nature of the attacks that are targeting cyber physical or network infrastructure, a normal consequence is that the computer science community started to develop methodologies in order to provide defense mechanism to IT systems.

As a matter of fact, while computer scientist in last years proposed techniques in order to mitigate cyber attacks, there is a lack knowledge about the quantification of these attacks in financial terms. Cyber-insurance is currently considered just an option for industrial and public companies but it represents the increasingly important way for businesses of all sizes to manage the threat of cybercrime. However, less than 10% of UK companies actually take out specific protection. Incredibly, cyber-insurance cover has been around 10 years but, it seems, that companies do not have confidence in the types of products or services currently being offered [4]. A recent paper in computer science literature [6] summarizes the basic knowledge about cyber insurance so far from both market and scientific perspectives. The survey discusses the issues which make this type of insurance unique and show how different technologies are affected by these issues. Among the peculiarities of cyber-insurance there is coverage specification. It is hard to specify what an insured wants to be covered from and what an insurer is willing to cover precisely.

For many insurers and brokers, the technicalities of information security and the details of how to deal with a data breach remain a mystery. We think that a good starting point is to determine the costs or expenses the company needs covering and the types of incidents that cyber-insurance wants cover for. Considering the peculiarity and the repercussions of different attacks, we think that each kind of threat should be managed by different insurance policies. Furthermore we observe that different companies can exhibit a different risk level connected to the same type of threat. In few words, we are proposing to insurance companies to create ad hoc policies in order to support the effective spread of cyber insurance. The insurer should gain the trust of the company discussing the possible threats to which the company is exposed to, being able to propose ad-hoc policies for different companies and, basing on a preliminary analysis of the current company infrastructure, highlighting the vulnerabilities.

We aim to create a virtuous circle between companies that benefit from cyber insurance ad-hoc policies, insurers that will stipulate policies against cyber attacks and computer scientist, that will be able to adopt their proposed methodologies in the real-world.

Acknowledgements This work has been partially supported by H2020 EU-funded projects NeCS and C3ISP and EIT-Digital Project HII and PRIN "Governing Adaptive and Unplanned Systems of Systems" and the EU project CyberSure 734815.

References

1. Eling, M., Loperfido, N.: Data breaches: goodness of fit, pricing, and risk measurement. Insur. Math. Econ. **75**, 126–136 (2017)
2. Eling, M., Schnell, W.: What we know about cyber risk and cyber risk insurance? J. Risk Financ. **17**(4), 474–491 (2016)
3. Herath, H.S.B., Herath, T.C.: Copula based actuarial model for pricing cyber insurance policies. Insur. Mark. Co. Anal. Actuar. Comput. **2**(1) (2011)
4. Low, P.: Insuring against cyber-attacks. Comput. Fraud Secur. **2017**(4), 18–20 (2017)
5. Maochao, X., Lei, H.: Cybersecurity insurance: modelling and pricing. Research paper. Society of Actuaries (SOA), pp. 1–38 (2017)
6. Marotta, A., Martinelli, F., Nanni, S., Orlando, A., Yautsiukhin, A.: Cyber-insurance survey. Comput. Sci. Rev. **24**, 35–61 (2017)

Predicting the Volatility of Cryptocurrency Time-Series

Leopoldo Catania, Stefano Grassi, and Francesco Ravazzolo

Abstract Cryptocurrencies have recently gained a lot of interest from investors, central banks and governments worldwide. The lack of any form of political regulation and their market far from being "efficient", require new forms of regulation in the near future. From an econometric viewpoint, the process underlying the evolution of the cryptocurrencies' volatility has been found to exhibit at the same time differences and similarities with other financial time-series, e.g. foreign exchanges returns. This short note focuses on predicting the conditional volatility of the four most traded cryptocurrencies: Bitcoin, Ethereum, Litecoin and Ripple. We investigate the effect of accounting for long memory in the volatility process as well as its asymmetric reaction to past values of the series to predict: 1 day, 1 and 2 weeks volatility levels.

Keywords Cryptocurrencies · Score-driven models · Volatility forecast

1 The Volatility of Cryptocurrencies

Many of the stylized facts that characterize usual financial time-series also apply to cryptocurrencies. For instance, similar to equity prices, cryptocurrencies exhibit: (1) time-varying volatility, (2) extreme observations, and (3) an asymmetric reaction of

L. Catania (✉)
Department of Economics and Business Economics, Aarhus BSS and CREATES, Aarhus N, Denmark
e-mail: leopoldo.catania@econ.au.dk

S. Grassi
Department of Economics and Finance, University of Rome Tor Vergata and CREATES, Rome, Italy
e-mail: stefanograssi@uniroma2.it

F. Ravazzolo
Faculty of Economics and Management, Free University of Bozen-Bolzano, Bolzano, Italy

CAMP, BI Norwegian Business School, Oslo, Norway
e-mail: francesco.ravazzolo@unibz.it

M. Corazza et al. (eds.), *Mathematical and Statistical Methods for Actuarial Sciences and Finance*, https://doi.org/10.1007/978-3-319-89824-7_37

the volatility process to the sign of past observations (i.e., leverage effect). However, standard dynamic volatility models like the Generalized Autoregressive Conditional Heteroscedasticity (GARCH) model of Bollerslev [3] do not perform accurately and [4] show that they are outperformed by more refined alternatives like the Score Driven model with conditional Generalized Hyperbolic Skew Student's t (GHSKT) innovations. The specification of the conditional distribution of the aforementioned Score Driven volatility model, GHSKT, is important since it characterises the filter for the conditional volatility, see [5] and [7]. For instance, [4] find that the robust volatility filter implied by the Score Driven–GHSKT model is of primary importance in describing the stochastic evolution of cryptocurrencies. Indeed, in their analysis involving 289 cryptocurrencies, GARCH is never preferred according to likelihood criteria.

The aim of this short note is to extend results of Catania and Grassi [4] to the important tasks of predicting future volatility levels of the four most representative cryptocurrencies: Bitcoin, Ethereum, Litecoin and Ripple. Those cryptocurrencies are the most important in terms of diffusion and market capitalization. At the time of writing market capitalization in USD dollars is 185.5 billion for Bitcoin, 44.3 billion dollars for Ethereum, 9.7 billion dollars for Ripple and 5.5 billion dollars for Litecoin. All together, these cryptocurrencies represent the 73% of the total cryptocurrency market value. See [4] for a detailed description of those cryptocurrencies.

Since volatility is unobserved and realized volatility measures are not available, in our forecasting analysis we proxy future volatility levels with the square of the realized log-returns. Squared returns are known to be a poor volatility proxy, and poor volatility proxies are known to affect forecast comparison, see [1]. To lower the influence of a volatility proxy on our results, model comparison is performed using the Quasi-Like (QLIKE) loss function which, as discussed by Patton et al. [8], is robust to this choice of volatility proxy. Specifically, let $\widehat{\sigma}_{j,t+h|t}$ be the h-step ahead volatility prediction made by model j at time t, and let r_{t+h} the log returns at time $t+h$, the QLIKE loss is defined as:

$$QLIKE(\widehat{\sigma}^2_{j,t+h|t}, \sigma^{2^*}_{j,t+h|t}) = \log\left(\widehat{\sigma}^2_{j,t+h|t}\right) + \frac{\widehat{\sigma}^2_{j,t+h|t}}{\sigma^{2^*}_{j,t+h|t}} \tag{1}$$

where $\sigma^{2^*}_{j,t+h|t} = r^2_{t+h}$ is the volatility proxy. QLIKE values associated to each model are computed recursively over a forecast horizon of length H. Values are then averaged and models with lower average values are preferred. In order to statistically assess the differences among alternative models, we employ the Model Confidence Set procedure of Hansen et al. [6] using the R package MCS detailed in [2].

2 Forecast Analysis and Model Comparison

The set of models we consider includes the GARCH model of Bollerslev [3] ($\mathscr{M}_1$), the Score Driven–GHSKT model ($\mathscr{M}_2$) along with three extensions with: (1) leverage ($\mathscr{M}_3$), (2) time-varying skewness ($\mathscr{M}_4$), and (3) fractional integration in the volatility process ($\mathscr{M}_5$), see [4] for a detailed specification of these models. It is worth noting that, the volatility filter of the Score Driven–GHSKT model also depends from the shape and skewness parameters of the GHSKT conditional distribution. This way, volatility predictions delivered by model $\mathscr{M}_4$ will be affected by the specification of time-varying skewness coefficients.

The data we consider are percentage log differences of the daily cryptocurrencies closing values. The Bitcoin and Litecoin series start the 29th of April, 2013, while Ethereum and Ripple series start the 8th and the 5th August, 2013, respectively. All series end the 1st of December, 2017.[1] Bitcoin and Litecoin have 1678 observations while Ethereum and Ripple have 847 and 1'580, respectively.[2] The full sample is equally divided in two parts: (1) the in-sample period where models' parameters are estimated the first time and, (2) the out-of-sample period where predictions are made. The length of the out-of-sample period is 839 for Bitcoin and Litecoin, and 424 and 790 for Ethereum and Ripple, respectively. Models' parameters are updated each time a new observation becomes available using an expanding window until the end of the sample. We select three forecast horizons: (1) 1 day ($h = 1$), (2) 1 week ($h = 7$) and, 2 weeks ($h = 14$).

Table 1 reports the average QLIKE values for all cryptocurrencies and forecast horizons. Results are reported relative to the GARCH model, $\mathscr{M}_1$, acting as a benchmark. That is, values lower than one indicate outperformance with respect to $\mathscr{M}_1$ and viceversa. Gray cells indicate those models that belong to the Superior Set of Models delivered by the Model Confidence Set procedure with confidence level 10%.

Results indicate that $\mathscr{M}_1$ is generally outperformed by the more refined Score Driven–GHSKT model, $\mathscr{M}_2$. Gains increase when the forecast horizon growths. We find that for Bitcoin, $\mathscr{M}_2$ reports better results than its extensions $\mathscr{M}_3$, $\mathscr{M}_4$ and $\mathscr{M}_5$. This result confirms the findings of Catania and Grassi [4] in their in-sample models comparison. Results for Ethereum show that many models belong to SSM indicating that all models perform similar in predicting future volatility levels. This result might be influenced by the low number of observations available for Ethereum. Results for Ripple and Litecoin are very clear: $\mathscr{M}_5$ is preferred for Ripple and $\mathscr{M}_3$ for Litecoin. That is, long memory is an important feature for the prediction of the Ripple's volatility, and the inclusion of an asymmetric reaction of the volatility process is of primary importance for Litecoin.

[1] Note that the cryptocurrency market trades 24 h a day, all days. Here with closing value we mean the price at (UTC) midnight.

[2] All series are available from https://coinmarketcap.com.

Table 1 Average QLIKE values for all cryptocurrencies and forecast horizons $h = 1, 7, 14$

	$\mathscr{M}_1$	$\mathscr{M}_2$	$\mathscr{M}_3$	$\mathscr{M}_4$	$\mathscr{M}_5$	$\mathscr{M}_1$	$\mathscr{M}_2$	$\mathscr{M}_3$	$\mathscr{M}_4$	$\mathscr{M}_5$
	Bitcoin					Ethereum				
$h = 1$	1.00	0.99	0.99	1.00	0.99	1.00	1.00	1.00	1.00	1.00
$h = 7$	1.00	0.97	0.97	0.99	0.98	1.00	0.97	0.98	0.98	1.00
$h = 14$	1.00	0.97	0.97	0.99	1.02	1.00	1.00	1.00	1.00	1.04
	Ripple					Litecoin				
$h = 1$	1.00	0.97	0.97	0.99	0.96	1.00	1.00	1.00	1.02	1.02
$h = 7$	1.00	0.98	0.98	1.01	0.97	1.00	0.98	0.97	0.99	0.98
$h = 14$	1.00	0.99	0.99	1.02	0.98	1.00	0.94	0.94	0.97	0.94

Results are reported for the five models the GARCH model of Bollerslev [3], $\mathscr{M}_1$, the Score Driven–GHSKT model detailed in [4], $\mathscr{M}_2$, and its three extensions including: (1) leverage ($\mathscr{M}_3$), (2) time-varying skewness ($\mathscr{M}_4$), and (3) fractional integration in the volatility process ($\mathscr{M}_5$), see [4]. Results are reported relative to $\mathscr{M}_1$. Values lower than one indicate outperformance with respect to $\mathscr{M}_1$ and viceversa. Gray cells indicate those models that belong to the Superior Set of Models delivered by the Model Confidence Set procedure with confidence level 10%

3 Conclusion

This short paper focuses on predicting the conditional volatility of the four most traded cryptocurrencies: Bitcoin, Ethereum, Litecoin and Ripple. We investigate the effect of accounting for long memory in the volatility process as well as its asymmetric reaction to past values of the series to predict volatility levels. Our findings indicate that more sophisticated volatility models that include leverage and time-varying skewness can improve volatility predictions at different forecast horizons from 1% to 6% compared to more standard alternatives. Applications in portfolio optimizations, hedging and pricing of derivative securities, where volatility modelling is of primary importance, can benefit from these findings.

References

1. Andersen, T.G., Bollerslev, T.: Answering the skeptics: yes, standard volatility models do provide accurate forecasts. Int. Econ. Rev. **39**(4), 885–905 (1998)
2. Bernardi, M., Catania, L.: The model confidence set package for R. Int. J. Comput. Econ. Econ. **8**(2), 144–158 (2018)
3. Bollerslev, T.: Generalized autoregressive conditional heteroskedasticity. J. Econ. **31**, 307–327 (1986)
4. Catania, L., Grassi, S.: Modelling crypto-currencies financial time-series. Available at SSRN: https://ssrn.com/abstract=3028486 or http://dx.doi.org/10.2139/ssrn.3028486 (16 Aug 2017)
5. Creal, D., Koopman, S.J., Lucas, A.: Generalized autoregressive score models with applications. J. Appl. Econ. **28**, 777–795 (2013)
6. Hansen, P.R., Lunde, A., Nason, J.M.: The model confidence set. Econometrica **79**(2), 453–497 (2011)

7. Harvey, A.C.: Dynamic Models for Volatility and Heavy Tails: With Applications to Financial and Economic Time Series, vol. 52. Cambridge University Press, Cambridge (2013)
8. Patton, A.J.: Volatility forecast comparison using imperfect volatility proxies. J. Econ. **160**(1), 246 – 256 (2011), Realized Volatility

A Generalized Error Distribution-Based Method for Conditional Value-at-Risk Evaluation

Roy Cerqueti, Massimiliano Giacalone, and Demetrio Panarello

Abstract One of the most important issues in finance is to correctly measure the risk profile of a portfolio, which is fundamental to take optimal decisions on the capital allocation. In this paper, we deal with the evaluation of portfolio's Conditional Value-at-Risk (CVaR) using a modified Gaussian Copula, where the correlation coefficient is replaced by a generalization of it, obtained as the correlation parameter of a bivariate Generalized Error Distribution (G.E.D.). We present an algorithm with the aim of verifying the performance of the G.E.D. method over the classical RiskMetrics one, resulting in higher performance of the G.E.D. method.

Keywords Portfolio theory · Gaussian Copula · Generalized Correlation Coefficient

1 Introduction

Value-at-Risk (VaR) has become a standard measurement tool in financial risk management due to its simplicity. However, it is an unstable and numerically difficult to use method when the losses do not follow a Gaussian distribution [7], which is usually the case in the analysis of financial data. Conditional VaR (CVaR, see Rockafellar and Uryasev [14]; Acerbi and Tasche [1, 10]) has been proposed by literature as an alternative to VaR [3]. For a better calculation of the risk, one of the proposals (e.g. [11]) is to model the interdependence of the returns by means

R. Cerqueti
Department of Economics and Law, University of Macerata, Macerata, Italy
e-mail: roy.cerqueti@unimc.it

M. Giacalone
Department of Economics and Statistics, University of Naples 'Federico II', Napoli, Italy
e-mail: massimiliano.giacalone@unina.it

D. Panarello (✉)
Department of Economic and Legal Studies, Parthenope University of Naples, Napoli, Italy
e-mail: demetrio.panarello@uniparthenope.it

M. Corazza et al. (eds.), *Mathematical and Statistical Methods for Actuarial Sciences and Finance*, https://doi.org/10.1007/978-3-319-89824-7_38

of Copula functions (see [12]). In this context, the problem can be split into two separate parts: first, to identify the marginal distributions of the returns of the single assets; second, to identify the specific copula which is more appropriate for representing the dependence structure of the returns (see [15]).

In Sect. 2, we introduce the quantitative ingredients of the study, with the main definitions. The proposed methodological setting for calculating the Conditional Value-at-Risk of a generic portfolio is presented in Sect. 3. Finally, in Sect. 4 some conclusive remarks are given.

2 The Generalized Error Distribution and the G.E.D. Copula

The Generalized Error Distribution (G.E.D.) family was introduced by Subbotin [16] and has been employed by various authors with different names and parameterizations (see e.g. [2, 4–6, 13]). A parameterization of the G.E.D. density function for a random variable X is:

$$f(x;\mu,\sigma_p,p)=\frac{1}{2\sigma_p p^{1/(p)}\Gamma(1+1/p)}exp(-\frac{1}{p}|\frac{x-\mu}{\sigma_p}|^p)\ \text{ for }\ -\infty<x<\infty \tag{1}$$

where $\mu = E(X)$ is the location parameter, $\sigma_p = [E|X-\mu|^p]^{1/p} > 0$ is the scale parameter, $p > 0$ is the shape parameter and Γ is the Euler Gamma function.

The density of a generic G.E.D. distribution is unimodal, symmetric and, for $p > 1$, bell-shaped. As particular cases we obtain the Laplace distribution ($p = 1$), the Normal ($p = 2$) and the Uniform ($p \to \infty$); for values of $1 < p < 2$ we obtain leptokurtic densities and for values of $p > 2$ we obtain platykurtic densities. Thus, the G.E.D. represents a generalization of a large set of distributions, allowing for a better description of financial data.

A bivariate copula is a function $C : [0, 1]^2 \to [0, 1]$ whose main interest in the field of probability is that it associates univariate marginal distributions to their joint ones [15]. We are here interested in the bivariate Gaussian Copula, which is defined as:

$$C(u,v|\rho)=\int_{-\infty}^{\Phi^{-1}(u)}\int_{-\infty}^{\Phi^{-1}(v)}\frac{1}{2\pi\sqrt{(1-\rho^2)}}exp\{\frac{-(r^2-2\rho rs+s^2)}{2(1-\rho^2)}\}drds, \tag{2}$$

where Φ^{-1} is the inverse of Gaussian distribution function, (u, v) uniform independent random variables generated from (X,Y) random variables, and $\rho \in [-1, 1]$ is a parameter representing the Pearson's correlation coefficient associated to the bivariate normal. The G.E.D. Copula for a generic random vector (X, Y) is obtained by replacing the parameter ρ by the Generalized Correlation Coefficient $\rho_p \in$

$[-1, 1]$, introduced by Taguchi [17] and estimated as follows:

$$\rho_p = \frac{codisp^{(p)}(X, Y)}{\sigma_p(X)\sigma_p(Y)}, \text{ with } -1 \leq \rho_p \leq 1, \tag{3}$$

where

$$|codisp^{(p)}(X, Y)|^p = |E[(Y-\mu_Y)|X-\mu_X|^{p-1}sign(X-\mu_X)]| \cdot |E[(X-\mu_X)|Y-\mu_Y|^{p-1}sign(Y-\mu_Y)]|,$$

$$\sigma_p(X) = [E|X-\mu_X|^p]^{1/p}, \qquad \sigma_p(Y) = [E|Y-\mu_Y|^p]^{1/p}$$

and μ_X and μ_Y are the expected values of X and Y, respectively. The parameters μ, p and σ_p could be estimated e.g. by using the Lp_{min} method [8, 9].

3 The Methodology

We here discuss theoretically how to compute the CVaR in a G.E.D. framework. Real world applications on financial data are available upon request.
We start from two sets of consecutive observations of the returns of two assets, assumed to be empirically distributed according to a G.E.D.:

$$X = \{x_1, \ldots, x_n\}, \qquad Y = \{y_1, \ldots, y_n\}.$$

For what concerns the stochastic dependence between X and Y, we propose a G.E.D. copula model.

The fundamental steps in the algorithm for the computation of the CVaR of a portfolio consisting of (α, β) are as follows:

0. we fix a portfolio (α, β), with $\alpha + \beta = 1$;
1. estimation of the parameters μ, p, σ_p in (1) for the two series of returns X and Y. In accord to the notation used above, we will denote the parameters as $\mu_X, p_X, \sigma_{p,X}$ and $\mu_Y, p_Y, \sigma_{p,Y}$;
2. estimation of the ρ_p parameter of the G.E.D. copula by using formula (3), with $p = \alpha p_X + \beta p_Y$;
3. generation of couples (x, y), which are the realization of the double stochastic variable (X, Y) having G.E.D. marginals of item 1. and stochastic dependence described by the G.E.D. copula with ρ_p of item 2.;
4. construction of the realizations of the returns of portfolio $P = \alpha X + \beta Y$ and of its empirical distribution;
5. computation of the Value-at-Risk of P at a confidence level $(1 - c)\%$;
6. computation of the Conditional Value-at-Risk of P at a confidence level $(1-c)\%$.

The algorithm above is repeated for all the considered portfolios, given by $\alpha = 0.01 : 0.01 : 0.99$ and $\beta = 1 - \alpha$.

4 Conclusions

The present paper is part of market risk calculation methods, whose purpose is to support risk managers' decision making processes. Among the different methods proposed in the literature for calculating Value-at-Risk, we took the well-known RiskMetrics into account. We introduced the new G.E.D. method and proposed the G.E.D. Copula as a generalization of the Gaussian Copula. Moreover, we introduced the Generalized Correlation Coefficient of norm p that, for the $p = 2$ case, equals the classic Bravais-Pearson correlation coefficient. We then presented an algorithm with the aim of verifying the performance of the new method over the classical RiskMetrics one. The problem of whether CVaR-G.E.D. can constitute a valid generalization of CVaR-R.M. is still debatable, and an empirical analysis would provide more insights on this relevant topic.

References

1. Acerbi, C., Tasche, D.: On the coherence of expected shortfall. J. Bank. Financ. **26**(7), 1487–1503 (2002)
2. Agrò, G.: Parameter orthogonality and conditional profile likelihood: the exponential power function case. Commun. Stat. Theory Methods **28**(8), 1759–1768 (1999)
3. Artzner, P., Delbaen, F., Eber, J.M., Heath, D.: Coherent measures of risk. Math. Financ. **9**(3), 203–228 (1999)
4. Bottazzi, G., Secchi, A.: A new class of asymmetric exponential power densities with applications to economics and finance. Ind. Corp. Chang. **20**, 991–1030 (2011)
5. Box, G.E., Tiao, G.C.: Bayesian Inference in Statistical Analysis. Addison-Wesley, Reading, MA (1973)
6. Chiodi, M.: Procedures for generating pseudo-random numbers from a normal distribution of order p. Riv. Stat. Appl. **19**, 7–26 (1986)
7. Ferraty, F., Quintela-Del-Río, A.: Conditional VaR and expected shortfall: a new functional approach. Econ. Rev. **35**(2), 263–292 (2016)
8. Giacalone, M.: Parameter evaluation of an exponential power function by simulation study. In: Shorts Communications and Posters, Compstat 96, Barcelona (1996)
9. Giacalone, M., Richiusa, R.: Lp-norm estimation: some simulation studies in presence of multicollinearity. Student **5**, 235–246 (2006)
10. Huang, D., Zhu, S., Fabozzi, F.J., Fukushima, M.: Portfolio selection under distributional uncertainty: a relative robust CVaR approach. Eur. J. Oper. Res. **203**(1), 185–194 (2010)
11. Malevergne, Y., Sornette, D.: Testing the Gaussian Copula hypothesis for financial assets dependences. Quant. Finan. **3**, 231–250 (2003)
12. McNeil, A.J., Frey, R., Embrechts, P.: Quantitative Risk Management: Concepts, Techniques and Tools. Princeton University Press, Princeton (2015)
13. Mineo, A.M.: On the estimation of the structure parameter of a normal distribution of order p. Statistica **63**(1), 109–122 (2007)
14. Rockafellar, R.T., Uryasev, S.: Conditional value-at-risk for general loss distributions. J. Bank. Financ. **26**(7), 1443–1471 (2002)
15. Sklar, M.: Fonctions de répartition à n dimensions et leurs marges. Université Paris **8** (1959)
16. Subbotin, M.: On the law of frequency of error. Mathematicheskii Sbornik **31**, 296–301 (1923)
17. Taguchi, T.: On Fechner's thesis and statistics with Norm-p. Ann. Inst. Stat. Math. **26**(1), 175–193 (1974)

Risk-Return Optimization for Life Insurance Portfolios

Riccardo Cesari and Vieri Mosco

Abstract Two Fong-Vasicek immunization results are discussed and applied in relation to life insurance fixed income portfolios. Firstly, we analyzed the contribution of Fong-Vasicek (J. Finance 39(5):1541–1546, 1984) providing a lower bound on the "shortfall" of an immunized asset portfolio in the face of an arbitrary shock to the term structure of interest rates. A "passive" strategy minimizing immunization (i.e., reinvestment) risk emerges, such that the exposure to an arbitrary variation of the shape of the term structure is minimized with respect to the "M-squared" risk measure representing the cash-flows dispersion around the duration matching target. Secondly, Fong and Vasicek (Financ. Anal. J. 39(5):73–78, 1983) risk-return approach is generalized in a model which seeks for only a partial risk minimization in exchange for more return potential. The empirical application hints at a perspective of "active" management, highlighting which segregated funds can be re-positioned along the efficient frontier, at a chosen level of the firm's risk appetite.

Keywords Life insurance · Immunization · Duration · Convexity · Fong-Vasicek theorem · Efficient frontier

R. Cesari (✉)
University of Bologna, Bologna, Italy

IVASS, Rome, Italy
e-mail: riccardo.cesari@ivass.it

Vieri Mosco
Research Department, IVASS, Rome, Italy
e-mail: vieri.mosco@ivass.it

M. Corazza et al. (eds.), *Mathematical and Statistical Methods for Actuarial Sciences and Finance*, https://doi.org/10.1007/978-3-319-89824-7_39

1 Introduction: The Life Insurance Matching Strategy and Movements of the Term Structure

Life insurance companies sell guarantees through a maturity matching between assets and liabilities. This matching and more generally the integrated management of assets and liabilities (ALM) has essentially the aim to cope with the interest rate risk, i.e. the risk of an asymmetric impact of interest rate movements to the asset and the liability side of the balance sheet. Depending on their asset and liability composition, insurance companies will be differently affected by future levels and shapes (slope and curvature) of interest rates. The duration mismatch gives an approximate measure of this exposure.

2 Passive Strategy: Minimizing the "Immunization Risk"

According to Fong and Vasicek [1], the investment horizon (holding period) might be considered a strategic horizon with respect to which to guarantee some target return, $\overline{R}_0$, fixed ex-ante at time 0. A lower bound for the ex post return is obtained by showing that the ex post portfolio value has a minimum percentage change (maximum shortfall) given by

$$\frac{\Delta A_0(H)}{A_0(H)} \geq -\frac{1}{2} M_0^2 \cdot \max_\tau \left\{ \Delta_0'(\tau) \right\} \tag{1}$$

where M_0^2 is a variance of times to payment, around asset duration D, and can be regarded as a risk measure for the imperfect immunization provided by the duration-matching strategy; $\max \left\{ \Delta_0'(\tau) \right\}$ is the maximum change of the slope of the current term structure across maturities. More precisely, given forward (instantaneous) rates $r_{FW}(0, \tau)$, for a given maturity τ their time change $\Delta_0(\tau)$ is calculated as their difference $r_{FW}(0+, \tau) - r_{FW}(0, \tau)$, providing a shift function, whose derivative with respect to maturity $\Delta_0'(\tau)$ expresses the "slope" of term structure of shifts[1]; while for M^2—representing the variance of times to payment around duration D, i.e.[2] $M^2 = \sum_{j=1}^{m} \left(s_j - H\right)^2 \cdot w_j$—a decomposition holds such that $M^2(H) = M^2(D) + (D - H)^2$.

According to the Fong and Vasicek result, M_0^2 is a risk measure in that it captures the exposure to any arbitrary movement of the discount curve. It suffices, therefore, to minimize M_0^2 in order to reduce such exposure and this prudential strategy might be regarded as a "passive" management.

Let us consider the case of two portfolios, the "*bullet*" and the "*barbell*". A "*bullet*" portfolio is composed by low-coupon securities with maturities close to duration

[1] Empirically approximated, over one year, as $\Delta_0(\tau) - \Delta_0(\tau - 1)$.

[2] Weighted in terms of present values of asset cash-flows.

D so that M^2 is close to zero (its minimum). Viceversa, a "*barbell*" portfolio is a set of very short and very long securities with large M^2 even if it has the same duration D. In fact, this twist affects both "reinvested income" and "capital gain" because short term rates have become lower, producing lower coupon from reinvested income whilst long term rates are now higher, producing higher capital losses realized at the horizon. This will produce a shortfall of the portfolio value with respect to the target value (a negative twist will produce the opposite). This means that both M^2 and the risk of the "*barbell*" portfolio are greater than those of the "*bullet*".

3 Active Strategy: The Risk-Return Tradeoff Optimization

Following Fong and Vasicek [2], a more general approach in terms of risk-return optimization problem (analogous to Markowitz's mean-variance analysis) could be set.

Let $\overline{\mathrm{R}}_0$ be the current annual ex-ante return over the horizon H. By definition:

$$(1 + R_0(H))^H \geq \frac{\Delta A_0(H)}{A_0(H)} \tag{2}$$

After an instantaneous non-constant shift of the term structure we have:

$$\Delta R_0(H) \cong \frac{1}{H} \frac{\Delta A_0(H)}{A_0(H)} \cong \frac{1}{H} M_0^2(H) \cdot \Delta S_0(H) \tag{3}$$

where

$$\Delta S_0(H) \equiv \frac{1}{2} \left[\Delta_0^2(H) - \Delta_0'(H) \right] \gtreqless 0 \tag{4}$$

is a special function of the term structure shift from time 0 to time 0+ and maturity H. Note that this function is the sum of a "shift in level" component ("*convexity effect*"), always positive, Δ_0^2, and a "*slope of shift*" component ("*risk effect*") of ambiguous sign. In case of adverse shift ($\Delta_0^2 < \Delta_0'$, $\Delta S_0 < 0$) the realized return will be under the target value.

The portfolio risk can be measured by the volatility (standard deviation) of $\Delta \overline{R}_0(H)$ and it is proportional to M_0^2, its dispersion measure[3].

[3] As explained by Fong and Vasicek [2]: "*A portfolio with half the value of M^2 than another portfolio can be expected to produce half the dispersion of realized returns around the target value when submitted to a variety of interest rate scenarios than the other portfolio*".

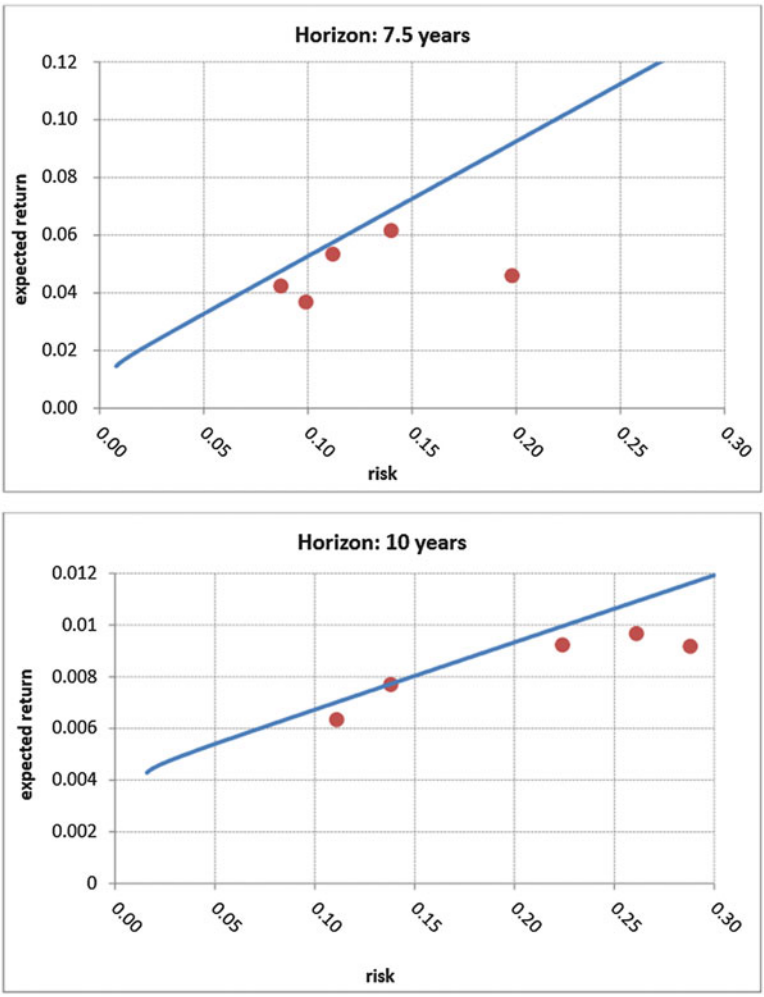


Fig. 1 Efficient frontiers for immunized portfolios at different horizons

Summing up (or integrating) all the shape changes between 0 and $H-1$ we have[4]:

$$R_H(H) - \overline{R}_0(H) = \frac{1}{H}\sum_{t=0}^{H-1} \Delta R_t(H) \cong \frac{1}{H}\sum_{t=0}^{H-1} M_t^2(H) \cdot \Delta S_t(H) \qquad (5)$$

[4]As noted by Fong and Vasicek [1]: "*the return differential can be thought of as the result of a large number of independent interest rate changes*" and "*assuming that the subsequent values of change in the slope of the yield curve" have "a common variance, the effects of the individual rate shocks can be integrated over the total horizon, subject to a function describing how* M^2 *changes with the remaining time to horizon*".

where $M_t^2(H)$ scales in time according to $M_t^2(H) = M_t^2(H) \cdot \left(\frac{H-t}{H}\right)^3$ while $M_t^2(H)$ is assumed to scale linearly in time as $\Delta S_t(H) \cong \Delta S_0(D) \cdot \left(\frac{H-t}{D}\right)$. Given a set of K different bonds the return is a linear combination of K random variable, $\sum_{i=1}^{K} w_i \left(R_{iH} - \overline{R}_{i0}\right)$ and, these formulas are applied in the computation of the mean vector and covariance matrix of the global return, $\mu_{i,H} = E\left[R_{i.H}(H) - \overline{R}_{i,0}(H)\right]$ and $\sigma_{ij} = E\left[R_{i.H}(H) - \overline{R}_{i,0}(H), R_{j.H}(H) - \overline{R}_{j,0}(H)\right]$, $i = 1, \ldots, K$.

Note that, in practice, in financial markets, cash flows can be typically bought and sold only in pre-defined "bundles" (the coupon bonds) so that optimal management must be set in terms of available bond portfolios.

As in the classical Markowitz approach, the active strategy is to find the portfolio $\{w_i, 1 \leq i \leq K\}$ minimizing the portfolio variance $w' \Sigma w$, for a given ex ante differential return m and the given horizon H:

$$\min_{w_i} \sqrt{w' \Sigma w} \tag{6}$$

$$\begin{cases} \sum_{1 \leq i \leq K} w_i \cdot \mu_{i,H} = m \\ \sum_{1 \leq i \leq K} w_i \cdot D_{i,0} = H \quad (duration\ consraint) \\ \sum_{1 \leq i \leq K} w_i = 1 \quad (budget\ constraint) \\ w_i \geq 0, \quad (no\ short\ selling) \end{cases}$$

By varying m, then, the efficient frontier $m(\sigma)$ can be traced, for a given H.

As an example, the following frontiers have been obtained by simplifying the Government bond market into four bond benchmarks with maturity 3, 5, 10 and 20 years respectively and by calculating the efficient frontiers with horizon (average duration) $H = 7.5$ and $H = 10$ (see Fig. 1).

4 Conclusions and Further Developments

Following the seminal work by Fong and Vasicek [1–3], it is possible to actively manage the "*immunization risk*" of a duration-matching bond portfolio of an insurance company. This risk is proportional to a measure (M^2) of the dispersion of the cash flow dates. The empirical application shows how distant the actual portfolio is from optimality on an "efficient frontier", according to the firm's risk appetite. Extensions are manifold. At the empirical level, a larger set of bond maturities could be taken into account. Theoretically, one can attempt to consider a multiple-liability framework. Moreover, an explicit stochastic dynamics can be assumed for the shocks to the forward rates, to be exploited in the calculation of moments. Finally, a three-dimensional frontier (risk, return and horizon) could be implemented as a more general approach.

References

1. Fong, H.G., Vasicek, O.A.: The tradeoff between return and risk in immunized portfolios. Financ. Anal. J. **39**(5), 73–78 (1983)
2. Fong, H.G., Vasicek, O.A.: Return maximization for immunized portfolios. In: Kaufman, G., Biervag, G.O., Toevs, A. (eds.) *Innovations in Bond Portfolios Management: Duration Analysis and Immunization*, pp. 227–238. JAI Press, London (1983)
3. Fong, H.G., Vasicek, O.A.: A risk minimizing strategy for portfolio immunization. J. Financ. **39**(5), 1541–1546 (1984)

When Is Utilitarian Welfare Higher Under Insurance Risk Pooling?

Indradeb Chatterjee, Angus S. Macdonald, Pradip Tapadar, and R. Guy Thomas

Abstract This paper focuses on the effects of bans on insurance risk classification on utilitarian social welfare. We consider two regimes: full risk classification, where insurers charge the actuarially fair premium for each risk, and pooling, where risk classification is banned and for institutional or regulatory reasons, insurers do not attempt to separate risk classes, but charge a common premium for all risks. For the case of iso-elastic insurance demand, we derive sufficient conditions on higher and lower risks' demand elasticities which ensure that utilitarian social welfare is higher under pooling than under full risk classification. Empirical evidence suggests that these conditions may be realistic for some insurance markets.

Keywords Social welfare · Elasticity of demand · Risk pooling

1 Outline of Our Approach

We consider two alternative regimes: full risk classification, where insurers charge the actuarially fair premium for each risk, and pooling, where risk classification is banned and insurers charge a common premium for all risks. Pooling implies a redistribution from lower risks towards higher risks. The outcome in terms of utilitarian social welfare depends on how we evaluate the trade-off between the utility gains and losses of the two types.

Such evaluations are typically made with models which assume that all individuals share a common utility function. Given an offered premium, individuals with the same probabilities of loss (i.e. individuals from the same *risk-group*) then all make the same purchasing decision. However, this does not correspond well to

I. Chatterjee (✉) · P. Tapadar · R. Guy Thomas
University of Kent, Canterbury, UK
e-mail: ic252@kent.ac.uk

A. S. Macdonald
Heriot-Watt University, Edinburgh, UK
e-mail: A.S. Macdonald@hw.ac.uk

M. Corazza et al. (eds.), *Mathematical and Statistical Methods for Actuarial Sciences and Finance*, https://doi.org/10.1007/978-3-319-89824-7_40

the observable reality of many insurance markets, where individuals with similar probabilities of loss often appear to make different decisions, and many individuals do not purchase insurance at all.

To reproduce observable reality, we instead introduce *heterogeneity* of utility functions (not necessarily all risk-averse) across individuals from any given risk-group. Individual utility functions then determine individual purchasing decisions, which (when aggregated) determine the insurance demand curve, and hence the equilibrium price of insurance when all risks are pooled.

Our measure of social welfare is expected utility *given* the distributions of loss probabilities and preferences in society, but evaluated *behind* a hypothetical 'veil of ignorance' which screens off knowledge of what position in society one occupies.

2 Model Set-Up

2.1 Insurance Demand from the Individual Viewpoint

Suppose that an individual has wealth W and risks losing an amount L with probability μ. The individual's utility of wealth is given by $u(\cdot)$, where $u'(\cdot) > 0$. The individual is offered insurance against the potential loss amount L at premium π (per unit of loss), i.e. for a payment of $\pi\, L$. He will purchase insurance if:

$$u(W - \pi\, L) > (1 - \mu)\, u(W) + \mu\, u(W - L). \tag{1}$$

Since certainty-equivalent decisions do not depend on the origin and scale of a utility function, standardising $u(W) = 1$ and $u(W - L) = 0$, simplifies the decision rule to:

$$u(W - \pi\, L) > (1 - \mu). \tag{2}$$

Assuming small premiums (such that the second and higher-order terms in the Taylor series of expansion of $u(W - \pi L)$ are negligible), we can then write:

$$u(W - \pi\, L) \approx u(W) - \pi\, L\, u'(W) = 1 - \pi\, L\, u'(W), \quad \text{as } u(W) = 1 \tag{3}$$

and hence the decision rule becomes:

$$\gamma < v. \tag{4}$$

where $\gamma = L\, u'(W)$ is the *risk preferences index* and $v = \mu/\pi$ is *risk-premium ratio*.

2.2 Insurance Demand from the Insurer's Viewpoint

From an insurer's perspective, it cannot observe individual utility functions; it observes only the *proportion* of each risk-group who choose to buy insurance. We call this a (proportional) demand function and define it as:

$$d(\pi) = \mathrm{P}[\,\gamma < v\,]\,. \tag{5}$$

It can be shown that if the underlying random variable Γ from which individual realisations of γ are generated has a particular distribution,[1] this implies the *iso-elastic* demand function:

$$d(\pi) = \tau \left(\frac{\mu}{\pi}\right)^{\lambda} \tag{6}$$

where λ is the constant *elasticity of demand* and τ is the *fair-premium demand*.

2.3 Market Equilibrium and Social Welfare

We assume a market with n risk-groups, where competition between insurers leads to zero expected profits in equilibrium. We define a *risk classification regime* as a vector of premiums $(\pi_1, \pi_2, \ldots, \pi_n)$ charged to the risk-groups. Social welfare, $S(\underline{\pi})$, under that regime is the expected utility of an individual selected at random from the population.[2] For the special case of iso-elastic demand, it can be shown that:

$$S(\underline{\pi}) = \sum_{i=1}^{n} p_i\, \tau_i\, \frac{1}{(\lambda_i + 1)} \left(\frac{\mu_i}{\pi_i}\right)^{\lambda_i + 1} \pi_i + K \tag{7}$$

where K is a constant, and the premium regime $\underline{\pi}$ satisfies the equilibrium condition:

$$\sum_{i=1}^{n} p_i\, \tau_i \left(\frac{\mu_i}{\pi_i}\right)^{\lambda} (\pi_i - \mu_i) = 0. \tag{8}$$

[1] Specifically, a generalised form of the Kumaraswamy [2] distribution. Similarly, any other distribution for γ will imply its own corresponding demand function.

[2] Social welfare at the 'end points' W and $W - L$ is standardised to $u(W) = 1$ and $u(W - L) = 0$ as in Eq. (2); this is necessary to avoid the possibility of a 'utility monster' dominating social welfare [1, 3].

3 Results for Iso-Elastic Demand

Result 3.1 *Suppose there are n risk-groups with risks* $\mu_1 < \mu_2 < \cdots < \mu_n$ *and the same iso-elastic demand elasticity* $\lambda > 0$. *Then* $\lambda \leq 1 \Rightarrow S(\pi_0) \geq S(\underline{\mu})$.

Result 3.1 says that if the common demand elasticity for all risk-groups is less than 1, pooling gives higher social welfare than full risk classification.

Result 3.1 assumes constant iso-elastic demand elasticity for all individuals. However, for most goods and services, we expect demand elasticity to rise with price, because of the income effect on demand: at higher prices, the good forms a larger part of the consumer's total budget constraint, and so the effect of a small percentage change in its price might be larger. For insurance this suggests that demand elasticity for higher risks might be higher. This motivates the following Result 3.2:

Result 3.2 *Suppose there are n risk-groups with risks* $\mu_1 < \mu_2 < \cdots < \mu_n$ *with iso-elastic demand elasticities* $\lambda_1, \lambda_2, \ldots, \lambda_n$ *respectively. Define* $\lambda_{lo} = \max\{\lambda_i : \mu_i \leq \pi_0\}$ *and* $\lambda_{hi} = \min\{\lambda_i : \mu_i > \pi_0\}$ *where* π_0 *is the pooled equilibrium premium. If* $\lambda_i < 1$ *for all* $i = 1, 2, \ldots, n$ *and* $\lambda_{lo} \leq \lambda_{hi}$, *then* $S(\pi_0) \geq S(\underline{\mu})$.

Roughly speaking, Result 3.2 says that if all higher risk-groups' (iso-elastic) demand elasticities are higher than all lower risk-groups' (iso-elastic) demand elasticities, *and* all demand elasticities are less than 1, then social welfare is higher under pooling than under full risk classification.

For the two risk-groups case, Result 3.2 references the green triangle in Fig. 1. The two axes represent demand elasticities for lower and higher risk-groups, λ_1

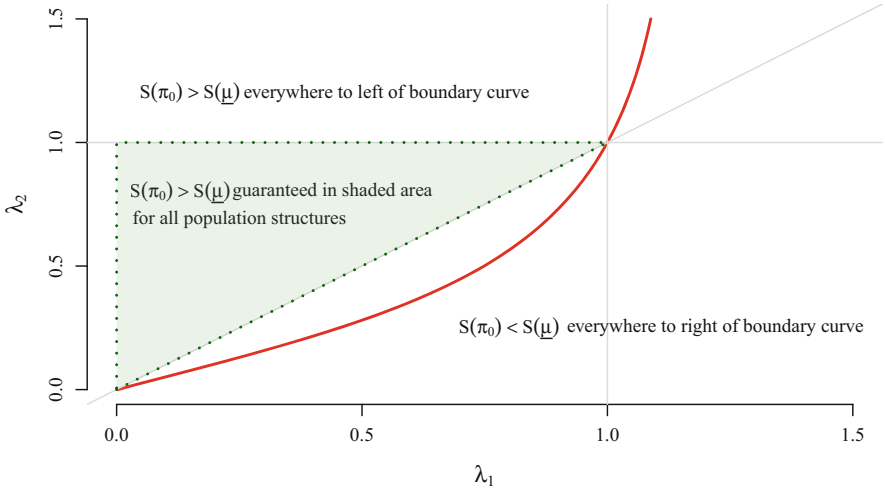


Fig. 1 Social welfare is higher under pooling to the left of the curve (guaranteed for *any* population structure in green triangle)

and λ_2. Social welfare under pooling is higher than under full risk classification everywhere on the left of the boundary curve, and lower everywhere on the right. The exact position of the boundary curve depends on the population structure and relative risks; the curve shown is for $\mu_2/\mu_1 = 4$ and 80% of the population are low risks. The sufficient conditions in Result 3.2 specify that in the shaded triangle where $\lambda_1 \leq \lambda_2 < 1$, social welfare under pooling is *always* higher than that under full risk classification, *irrespective* of the population structure and relative risks.

4 Discussion

The conditions in the above results encompass many plausible combinations of higher and lower risks' demand elasticities. The conditions are stringent because they are sufficient for *any* population structures and relative risks, but they are *not* necessary (as shown by the white areas to the left of the boundary in Fig. 1).

A condition common to both results is that all demand elasticities should be less than 1. Most relevant empirical estimates found in literature are of magnitude significantly less than 1. Whilst the various contexts in which these estimates were made may not correspond closely to the set-up in this paper, it is at least suggestive of the possibility that insurance demand elasticities may typically be less than 1.

References

1. Bailey, J.W.: Utilitarianism, Institutions and Justice. Oxford University Press, Oxford (1997)
2. Kumaraswamy, P.: A generalized probability density function for double-bounded random processes. J. Hydrol. **46**, 79–88 (1980)
3. Nozick, R.: Anarchy, State and Utopia. Basic Books, New York (1974)

The Value of Information for Optimal Portfolio Management

Katia Colaneri, Stefano Herzel, and Marco Nicolosi

Abstract We study the value of information for a manager who invests in a stock market to optimize the utility of her future wealth. We consider an incomplete financial market model with a mean reverting market price of risk that cannot be directly observed by the manager. The available information is represented by the filtration generated by the stock price process. We solve the classical Merton problem for an incomplete market under partial information by means of filtering techniques and the martingale approach.

Keywords Utility maximization · Merton model · Partial information · Martingale approach

1 Introduction

What is the value of information for a portfolio manager who invests in the stock market to optimize the utility of her future wealth? We study this problem in a market with a mean reverting market price of risk X and where returns of stocks are random but predictable. The process X cannot be directly observed by the manager, and it is driven by different risk factors from those that directly affect stock prices.

Research partially funded by the Swedish Research Council, Grant 2015-01713

K. Colaneri
University of Leeds, School of Mathematics, Leeds, UK
e-mail: k.colaneri@leeds.ac.uk

S. Herzel
University of Rome Tor Vergata, Department of Economics and Finance, Rome, Italy
e-mail: stefano.herzel@uniroma2.it

M. Nicolosi (✉)
University of Perugia, Department of Economics, Perugia, Italy
e-mail: marco.nicolosi@unipg.it

M. Corazza et al. (eds.), *Mathematical and Statistical Methods for Actuarial Sciences and Finance*, https://doi.org/10.1007/978-3-319-89824-7_41

The objective of the manager is to maximize the utility of her wealth at a given time T. In a nutshell, what we consider is the classical Merton problem for an incomplete market with partial information.

An important part of the current research is concerned with the fact that expected returns and risk premia vary over time but are somewhat predictable, see [1] for a nice review. Following this path we consider asset dynamics with a mean-reverting market price of risk, thus extending a model proposed by Wachter [5] to a setting with partial information and market incompleteness. By applying the martingale approach, proposed by Cox and Huang [2] and suitably adapted to the case of incomplete markets by He and Pearson [3], we solve the problem in case of perfect information. In case of imperfect information, we apply filtering techniques to transform our problem into an equivalent optimization problem under full information. Due to the particular problem structure we can derive solutions in closed form in both cases. In the present short communication we limit ourselves to provide a summary of the main results that will be published in a full paper.

2 The Model

We consider a market model with N risky assets and a money market account having constant, risk-free, interest rate r. Trading is continuous in time from $t = 0$ to a finite time horizon T. The N-dimensional risky assets' prices process follows

$$dS_t = \mathrm{diag}(S_t)\left(\mu_t dt + \sigma_t dZ_t^S\right), \quad S_0 = s \in \mathbb{R}^+, \tag{1}$$

where $\mathrm{diag}(S)$ is a $N \times N$ diagonal matrix whose diagonal is equal to the vector $S = \{S_t\}_{t\in[0,T]}$ and $Z^S = \{Z_t^S\}_{t\in[0,T]}$ is a N-dimensional Brownian motion. Throughout the paper we always consider column vectors and use the symbol $\cdot'$ to denote the transpose. Let $\mathbb{F}^S = \{\mathscr{F}_t^S, 0 \le t \le T\}$ be the natural filtration of S. The vector $\mu = \{\mu_t\}_{t\in[0,T]}$ and the matrix $\sigma = \{\sigma_t\}_{t\in[0,T]}$ are processes satisfying standard regularity conditions (see, e.g., [4]) in order to have existence and uniqueness of the solution to Eq. (1). We also assume that σ is invertible and adapted to $\mathbb{F}^S$. The process μ is defined by

$$\mu_t - r\mathbf{1} = \sigma_t X_t, \quad t \in [0, T],$$

where $\mathbf{1}$ is a N-vector of 1's and $X = \{X_t\}_{t\in[0,T]}$ is the N-vector of the market prices of risk whose dynamics is given by

$$dX_t = -\lambda_X(X_t - \bar{X})dt + \sigma_X dZ_t^X, \quad X_0 = x \in \mathbb{R}.$$

The strictly positive N-dimensional diagonal matrix λ_X provides the strength of attraction toward $\bar{X}$, that is the long term expected level for X, and σ_X is a N-dimensional square matrix that parameterizes the covariance of the market prices

of risk. The process $Z^X = \{Z_t^X\}_{t\in[0,T]}$ is a N dimensional Brownian motion, correlated with Z^S. The full information is given by filtration $\mathbb{F} = \{\mathscr{F}_t, 0 \le t \le T\}$ generated by Brownian motions Z^X and Z^S. We assume throughout the sequel that filtrations $\mathbb{F}$ and $\mathbb{F}^S$ satisfy the usual hypotheses of completeness and right continuity.

A portfolio manager dynamically allocates the fund's wealth through a self-financing strategy. Let θ_t be the N-dimensional vector representing the ratios of the wealth $W = \{W_t\}_{t\in[0,T]}$ invested in the risky assets at time t, the rate of return of the strategy follows

$$\frac{dW_t}{W_t} = (r + \theta_t'\sigma_t X_t)dt + \theta_t'\sigma_t dZ_t^S. \tag{2}$$

We consider a manager, with preferences described by a utility function u, who consumes all her wealth at time T and hence her optimization problem is

$$\max_{\theta} E[u(W_T)], \tag{3}$$

subject to the self-financing constraint, starting from a wealth W_0. In particular, in the sequel, we assume u to be a power utility function of the form

$$u(x) = \frac{1}{1-\gamma}x^{1-\gamma}, \qquad \gamma \neq 1.$$

3 The Full Information Case

We first consider the case where the manager can observe the market price of risk, that is the available information to the manager is given by filtration $\mathbb{F}$. Since the number of risk factors exceeds the number of securities, the market is incomplete.

According to [3], we apply a suitably adapted version of the martingale approach. Precisely, in the martingale approach the optimal final wealth W_T^* is obtained by solving the static problem

$$\max_{W_T} E[u(W_T)],$$

subject to the constraint

$$W_0 = E[\xi_T W_T | \mathscr{F}_0],$$

where ξ is the state price density that must be chosen in order to penalize unfeasible strategies, see [3]. The solution is given by

$$W_T^* = \frac{W_0}{E[\xi_T^{1-\frac{1}{\gamma}} | \mathscr{F}_0]}\xi_T^{-\frac{1}{\gamma}}.$$

The optimal wealth at any time t in $[0, T]$ is then given by

$$W_t^* = \frac{W_0}{E[\xi_T^{1-\frac{1}{\gamma}} | \mathscr{F}_0]} \frac{1}{\xi_t} E[\xi_T^{1-\frac{1}{\gamma}} | \mathscr{F}_t].$$

4 The Partial Information Case

Now we assume that the agent can only observe the prices process S and not the market prices of risk X. This implies that the available information is carried by filtration $\mathbb{F}^S$. The problem to be solved is to maximize expected utility from terminal wealth over a set of self-financing portfolio strategies that are $\mathbb{F}^S$-adapted. In this case we face an optimization problem under partial information. The standard procedure is to transform this problem into an equivalent optimization problem under full information by means of filtering. In other words, we replace unobservable quantities by their filtered estimates. By expressing the dynamics of stock price and *filtered* market price of risk in the observation filtration, we get that all state processes are driven by the Innovation process which is a N-dimensional Brownian motion with respect to filtration $\mathbb{F}^S$. Therefore the market model under partial information turns out to be complete and we can solve the optimization problem following the *classical martingale approach*.

5 Conclusion

After defining the dynamics of the assets price process and of the unobservable market price of risk process X, we study the corresponding filtering problem and solve a utility maximization problem under partial information. Then, we consider the same problem, this time assuming that the manager directly observes X. The difference of the certainty equivalents of the two optimal utilities is the maximum price that the agent with partial information would be willing to pay for the full information. Hence it represents the value of the information for the manager. In our setting, such a quantity can be expressed in closed form. The formula for the value of information can be used to provide direct answers to questions like: *A more risk-averse manager is willing to pay more or less to get full information? The value of information is higher or lower when the market gets closer to be a complete one? What is the impact on the value of the information of the uncertainty on the market price of risk?* We refer to the full paper for the explicit formula and some applications.

References

1. Cochrane, J.H.: Presidential address: discount rates. J. Financ. **66**, 1047–1108 (2011)
2. Cox, J.C., Huang, C.F.: Optimal consumptions and portfolio policies when asset prices follow a diffusion process. J. Econ. Theory **49**, 33–83 (1989)
3. He, H., Pearson, N.D.: Consumption and portfolio policies with incomplete markets and short-sale constraints: the infinite dimensional case. J. Econ. Theory **54**, 259–304 (1991)
4. Karatzas, I., Shreve, S.E.: Brownian Motion and Stochastic Calculus. Graduate Texts in Mathematics, vol. 113. Springer, New York (1991)
5. Wachter, J.A.: Portfolio and consumption decisions under mean-reverting returns: an exact solution for complete markets. J. Financ. Quant. Anal. **37**, 63–91 (2002)

Risk and Uncertainty for Flexible Retirement Schemes

Mariarosaria Coppola, Maria Russolillo, and Rosaria Simone

Abstract Nowadays, we are witnessing a wide and spread need to create flexible retirement schemes for facing global ageing and the prolonging working lives. Many countries have set up Social Security Systems, which link retirement age and/or pension benefits to life expectancy. In this context, we consider an indexing mechanism based on the expected residual life expectancy to adjust the retirement age and keep a constant Expected Pension Period Duration (EPPD). The analysis assesses the impact of different stochastic mortality models on the indexation by forecasting mortality paths based on extrapolative methods. Nevertheless, so far, in recent literature less attention has been given to the uncertainty issue related to model selection, although having appropriate estimates for the risk in mortality projections. With respect to the state of art, our proposal considers model-assembling techniques in order to balance fitting performances and uncertainty related to model selection, as well as uncertainty of parameters estimation.

The indexation mechanism obtained by joining the retirement age up with the expected life span is tested in actuarial terms by assessing the implied reduction of costs also when assuming worst and best scenarios. The analysis concerns the Italian population and outlines gender differences.

Keywords Longevity risk · Mortality projections · Uncertainty · Model averaging · Retirement scheme

M. Coppola · R. Simone
Department of Political Sciences, Federico II University, Naples, Italy
e-mail: m.coppola@unina.it; rosaria.simone@unina.it

M. Russolillo (✉)
Department of Economics and Statistics, University of Salerno, Salerno, Italy
e-mail: mrussolillo@unisa.it

M. Corazza et al. (eds.), *Mathematical and Statistical Methods for Actuarial Sciences and Finance*, https://doi.org/10.1007/978-3-319-89824-7_42

1 Motivation

The present work is tailored to study uncertainty in model fitting and selection with an application to the mainstream topic of longevity risk. In this setting, the problem of controlling for uncertainty of estimation and prediction is analyzed on a proposal for a flexible retirement scheme involving the indexation of the retirement age. The advanced lag mechanism is matched with life expectancy to keep a constant pension period duration and thus to control costs for National Social Security Systems. In an era where the ageing of population is a matter of serious concern for several countries, forecasts of mortality and related policy actions should be carefully managed. For this reason, a major importance has to be recognized to the model selection phase when assessing both the fitting performances and the predictive abilities of the candidate models. The discussion will rely on the Italian mortality experience and gender differences in risk and uncertainty of the proposed retirement scheme will be emphasized.

2 The Italian Mortality Experience

Data have been taken from the Human Mortality database and we shall focus on cohorts 1952–2012: indeed, those born in 1952 turned 65 in 2017, the historical benchmark for retirement age. For the forecasts, extrapolative methods have been applied by choosing the best fitting ARIMA process for the cohort effects when pertinent. We have considered the stochastic mortality models belonging to the Generalized Age Period Cohort Family. According to the AIC measure reported in Table 1, the M7 model is the best performing one for both male and female populations. However, we notice not negligible differences in the rank of the performances: for females, in particular with respect to the LC model.

3 Dealing with Model Uncertainty

We have restricted the analysis to the subset of models in the GAPC family with closer AIC value [1]. After this first step in the model selection, models have been assembled via the AIC weights [1] reported in Table 1. The LC model is maintained

Table 1 AIC and AIC weights for best fitting models

	Women		Men	
Model	AIC	AIC weights	AIC	AIC weights
LC	32894.02	0.225	43041.60	0.194
RH	26349.78	0.256	26610.66	0.267
M7	**25313.47**	**0.262**	**25793.43**	**0.271**
PLAT	26220.45	0.257	26440.79	0.268

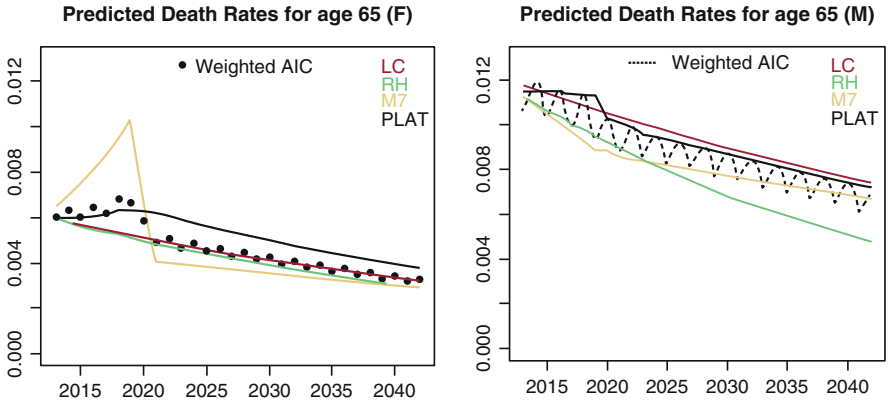


Fig. 1 Comparison between forecasted central death rates and model assembled death rates for aged 65 in 2012

as a benchmark. In this respect, for both women and men cohorts the RH, M7 and PLAT models are the competitors whose predictive abilities is worth to combine. This approach is particularly suitable and advisable for women since the best fitting model (M7) forecasts a steep increase in mortality trends around 2015–2020 as shown in Fig. 1. This appears to be justifiable on the basis on recent studies on the different impact that lung cancers, lower practice medical care because of the economic crises, and slower increase in longevity observed for women than for men. Models have been assembled via a weighted sum of fitted central death rates and projections: to that purpose, the AIC weights have been chosen to assess the relative convenience of each of the candidate models in order to perform a comparative analysis with classical model selection techniques.

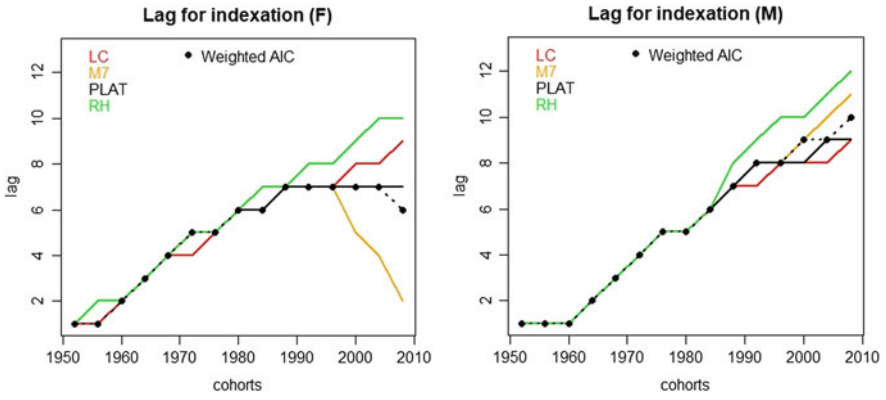
4 A Flexible Retirement Scheme

We propose an indexing mechanism for retirement age [4] based on the period life expectancy at age 65 for selected cohorts and for selected mortality model M belonging to the GAPC family [2]. We consider cohorts of males/females born from 1952 to 2012, setting the cohort 1952 as benchmark [3]. We follow an age-period approach in the sense that life expectancy is considered as a function of the age and the calendar year [5]. Specifically, let us consider an individual belonging to the cohort C, aged x_0 on the first of January of year t_0, when the expected lifetime according to a given stochastic mortality model M is equal to $e^{(M)}_{x_0,C}$. Let us suppose that the pension system we refer to foresees that x_0 is the fixed retirement age for all subsequent cohorts. The individual aged x_0 receives a constant monthly payment B as long as he/she survives. We can say that $e^{(M)}_{x_0,C}$ represents the Expected Pension

Period Duration according to model M (EPPD$^{(M)}$), that is the expected number of years during which pension payments are due. Then, for a fixed mortality model M and for each of the selected cohorts C, we determine the age at which life expectancy equals the EPPD$^{(M)}$. Specifically, we evaluate $e^{(M)}_{x_0+j,C}$ for increasing age span j = 1,2 . . ., and we index the retirement age x_0 by shifting it by the minimal amount $s^{(M)}_C$ to reach the EPPD$^{(M)}$, that is:

$$s^{(M)}_C = \min\left\{ j : e^{(M)}_{x_0+j,C} \leq EPPD^{(M)} \right\}.$$

Then, for each cohort (x-axis) the forward shift in retirement age derived from the application of the indexation mechanism is plotted for each of the best fitting models and for the assembled one (dotted line). We notice that this technique has a more prominent impact for the female population (right), because it allows to balance the increasing indexation foreseen by the LC, RH and PLAT model, with the decay implied by the M7 model. Indeed, the latter forecasts an increase in mortality rates for women cohorts 2000–2010 and thus, accordingly, the indexation procedure accounts for such reduced longevity risk by decreasing the forward shift for retirement age with respect to older cohorts.



5 Conclusion and Further Developments

The paper is devoted to assess the impact of different stochastic mortality models with respect to the proposed indexation of retirement age. A model-assembling technique based on the AIC weights is then considered in order to balance fitting performances and uncertainty related to model selection both for female and male population.

References

1. Benchimol, A.G., et al.: Model uncertainty approach in mortality projection with model assembling methodologies. Working papers. Statistics and econometrics. WS Universidad Carlos III de Madrid Departamento de Estadística DES (2016)
2. Villegas, A.R., Kaishev, V., Millossovich, P.: StMoMo: an R package for stochastic mortality modelling. Under review or revision 38 pages. http://openaccess.city.ac.uk/16834/1/StMoMoVignette.pdf (2016)
3. Coppola, M., Russolillo, M., Simone, R.: The impact of mortality projection models in case of flexible retirement schemes. Book of abstract 17th Applied Stochastic Models and Data Analysis International Conference with Demographics Workshop. ISBN: 978-618-5180-22-5. http://www.asmda.es/ (2017)
4. Bisetti, E., Favero, C.A.: Measuring the impact of longevity risk on pension systems: the case of Italy. N. Am. Actuar. J. **18**(1), 87–104 (2014)
5. Denuit, M., Haberman, S., Renshaw, A.E.: Longevity-contingent deferred life annuities. J. Pension Econ. Financ. **14**(3), 315–327 (2015)

Comparing Possibilistic Portfolios to Probabilistic Ones

Marco Corazza and Carla Nardelli

Abstract In this paper, we compare mean-variance portfolios based on the standard probabilistic representation of the stock returns to mean-variance portfolios built by using stock returns represented as possibilistic numbers. With reference to the latter, in this note we focus our attention on the definitions recently proposed in literature for modeling portfolio selection problems. In particular, first we investigate some theoretical properties of the possibilistic portfolios and compare them to the equivalent ones of the probabilistic portfolios, then, given the assets composing the Italian stock index FTSE MIB, we empirically compare the performances of the possibilistic portfolios to those of the probabilistic one. The results show that, generally, the probabilistic approach is more flexible than the possibilistic one in solving portfolio selection problems.

Keywords Trapezoidal fuzzy number · Possibilistic portfolio · Probabilistic portfolio · Variance-covariance matrix · FTSE MIB

1 Introduction

In general terms, portfolio selection consists in sharing a starting capital among various stocks whose future performances are unknown, this in order to optimize some risk-return profile of the portfolio itself. Given the uncertainty of the future stock returns, a crucial role is played by the measurement of the risk associated to such stock returns. The classical approach represents the future stock returns in terms of random variables. Alongside to this approach, other alternative ways have

M. Corazza (✉)
Department of Economics, Ca' Foscari University of Venice, Venice, Italy
e-mail: corazza@unive.it

C. Nardelli
Department of Management, Economics and Quantitative Methods, University of Bergamo, Bergamo, Italy
e-mail: carla.nardelli@unibg.it

M. Corazza et al. (eds.), *Mathematical and Statistical Methods for Actuarial Sciences and Finance*, https://doi.org/10.1007/978-3-319-89824-7_43

been proposed. One of the most successful is the representation of the stock returns in terms of trapezoidal fuzzy numbers, also called possibilistic numbers.

In this contribution, we compare mean-variance portfolios based on the standard probabilistic representation of the stock returns to mean-variance portfolios built by using stock returns represented as possibilistic numbers. There exists only a definition for the probabilistic mean and only one for the probabilistic variance,[1] whereas there exist several definitions of possibilistic mean and of possibilistic variance. With reference to the latter, in this note we focus our attention on the definitions recently proposed in literature for modeling portfolio selection problems: the lower possibilistic mean and variance [3], the upper possibilistic mean and variance [3], and the possibilistic mean and variance *à la* Zhang-Zhang-Xiao [4], hereinafter referred to as ZZX possibilistic mean and variance.

In order to present these different possibilistic means and variances, we before need to provide the definition of trapezoidal fuzzy number [1]. A fuzzy number $A = (a, b, \alpha, \beta)$ is called trapezoidal with tolerance interval $[a, b]$, in which $a, b \in \mathbb{R}$, with left width $\alpha > 0$ and with right width $\beta > 0$ if its membership function has the following form:

$$A(t) = \begin{cases} 1 - \dfrac{a - t}{\alpha} & \text{if } a - \alpha < t \leq a \\ 1 & \text{if } a < t \leq b \\ 1 - \dfrac{t - b}{\beta} & \text{if } b < t \leq b + \beta \\ 0 & \text{otherwise} \end{cases}.$$

Now, on the basis of this definition, it is possible to obtain the listed possibilistic means and variances by calculating simple integrals. In particular, given two possibilistic numbers $A_1 = (a_1, b_1, \alpha_1, \beta_1)$ and $A_2 = (a_2, b_2, \alpha_2, \beta_2)$, with regards to the lower possibilistic mean, variance and covariance, one has respectively:

$$M_*(A_1) = a_1 - \frac{\alpha_1}{3},\ Var_*(A_1) = \frac{1}{18}\alpha_1^2,\ Covar_*(A_1, A_2) = \frac{1}{18}\alpha_1\alpha_2.$$

Then, given the same possibilistic numbers, concerning the upper possibilistic mean, variance and covariance, one has respectively:

$$M^*(A_1) = b_1 + \frac{\beta_1}{3},\ Var^*(A_1) = \frac{1}{18}\beta_1^2,\ Covar^*(A_1, A_2) = \frac{1}{18}\beta_1\beta_2.$$

[1] From here on, with the term "variance" we imply also the term "covariance".

Finally, given again the possibilistic numbers A_1 and A_2, with reference to the ZZX possibilistic mean, variance and covariance, one has respectively:

$$M(A_1) = \frac{a_1 + b_1}{2} + \frac{\beta_1 - \alpha_1}{6},\ Var(A_1) = \frac{1}{4}(b_1 - a_1)^2 + \frac{1}{24}(\alpha_1 + \beta_1)^2 + \frac{1}{6}(b_1 - a_1) \cdot (\alpha_1 + \beta_1),\ Covar(A_1, A_2) = \frac{1}{4}(b_1 - a_1)(b_2 - a_2) + \frac{1}{24}(\alpha_1 + \beta_1)(\alpha_2 + \beta_2) + \frac{1}{12}(b_1 - a_1)(\alpha_2 + \beta_2) + \frac{1}{12}(b_2 - a_2)(\alpha_1 + \beta_1).$$

The remainder of this note is organized as follows. In the next section, we investigate some theoretical properties of the possibilistic portfolios and compare them to the equivalent ones of the probabilistic portfolios. In Sect. 3, given the real assets which are components of the Italian stock index FTSE MIB, we empirically compare the performances of the possibilistic portfolios to those of the probabilistic ones.

2 The Theoretical Comparison

As well known, in portfolio selection problems, an important role is played by the variance-covariance matrix. Therefore, the knowledge of its features is basic for dealing with such problems. In this regard, we present the statements of two new theoretical results concerning the possibilistic variance-covariance matrices.

Theorem 1 *Let $A_1, \ldots, A_N$ be $N \geq 2$ trapezoidal fuzzy numbers.*[2] *Both the lower possibilistic variance-covariance matrix and the upper possibilistic variance-covariance matrix have rank* 1.

Theorem 2 *Let $A_1, \ldots, A_N$ be $N \geq 2$ trapezoidal fuzzy numbers. The ZZX possibilistic variance-covariance matrix has rank* 1 *or* 2*, depending on the values of a_i, b_i, α_i and β_i, with $i = 1, \ldots, N$.*

The idea underlying both these theorems is based on the fact that all the considered variance-covariance matrices depend only on a_i, b_i, α_i and β_i, with $i = 1, \ldots, N$. By suitably exploiting this feature, it is possible to prove the theorems.

Given such results, it is trivial to prove that the lower and upper possibilistic variance-covariance matrices cannot be used in the basic models of portfolio selections (like, for instance, that considered in [2]). In fact, their solutions require the inverse of such matrices. Furthermore, for the same reasons, it is equally trivial to prove that the ZZX possibilistic variance-covariance matrix cannot be used in the basic models of portfolio selections for which $N > 2$.

[2] In our framework, we can imagine such possibilistic numbers as stock returns.

Under this point of view, the probabilistic approach is more flexible than the possibilistic one as it is able to solve portfolio selection problems that the latter can not deal with.

3 The Empirical Comparison

In this second step, we empirically compare a possibilistic-based portfolio selection model to the probabilistic-based version of the same model on the basis of their respective performances when applied to a real stock market.

In particular, as concerns the portfolio selection problem, we consider the following one:

$$\min_{\boldsymbol{x}} \ \boldsymbol{x}'\boldsymbol{V}\boldsymbol{x} \ \text{ s.t. } \left\{\boldsymbol{x}'\boldsymbol{r} = \pi, \ \ \boldsymbol{x}'\boldsymbol{e} = 1, \ \ x_i \geq 0 \ \forall \ i = 1, \ldots, N\right\},$$

where $\boldsymbol{x} = (x_i, \ldots, x_N)$ is the vector of the unknown decision variables, that is the vector of the percentages of capital to invest in the various assets, $\boldsymbol{V}$ is the variance-covariance matrix, $\boldsymbol{r}$ is the vector of the expected rates of return, π is the desired expected rate of the return of the portfolio, and $\boldsymbol{e}$ is a vector of ones.
Note that, given the above theoretical results, we are obliged to use a portfolio selection model whose solution does not require the inverse of the variance-covariance matrix.

Regarding the data, we consider the returns of the assets which are components of the Italian stock index FTSE MIB from January 1, 2006 to April 28, 2017. We use the first half of these returns to calculate the possibilistic and the probabilistic versions both of the vectors of the expected rates of return and of the variance-covariance matrices. Then, we use the second half of these returns to calculate the futures performances of both the approaches. Note that, with reference to the representation of the stock returns in terms of trapezoidal fuzzy numbers, we randomly generate the latter in order to taken into account different financial views about the future behaviors of the stock returns themselves (of course, the probabilistic representation of the stock returns is always the same). We repeat this procedure 100 times. In Fig. 1, we report as example both the probabilistic representation of the returns of the stock Buzzi Unicem S.p.A. and one of the randomly generated possibilistic representation of the same returns.

The performances of the two approaches are similar. Therefore, under this point of view, it is not possible to detect a "winner" approach, while, we recall, the probabilistic approach is more flexible from the theoretical standpoint.

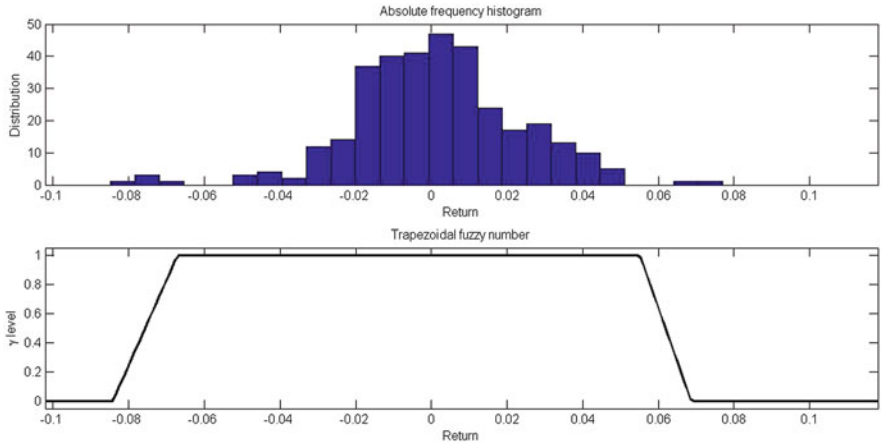


Fig. 1 In the upper figure we show the probabilistic representation of the stock Buzzi Unicem S.p.A. expressed in terms of absolute frequencies, and in the lower figure we show the possibilistic representation of the same returns

References

1. Carlsson, C., Fullér, R, Majlender, P.: A possibilistic approach to selecting portfolios with highest utility score. Fuzzy Sets Syst. **131**, 13–21 (2001)
2. Merton, R.C.: An analytic derivation of the efficient portfolio frontier. J. Financ. Quant. Anal. **7**, 1851–1872 (1972)
3. Zhang, W.-G.: Possibilistic mean-standard deviation models to portfolio selection for bounded assets. Appl. Math. Comput. **189**, 1614–1623 (2007)
4. Zhang, W.-G., Zhang, X.-L., Xiao, W.-L.: Portfolio selection under possibilistic mean-variance utility and a SMO algorithm. Eur. J. Oper. Res. **197**, 693–700 (2009)

Some Critical Insights on the Unbiased Efficient Frontier *à la* Bodnar&Bodnar

Marco Corazza and Claudio Pizzi

Abstract In a couple of recent papers, Bodnar and Bodnar have tackled the estimation problem of the efficient frontier of a risky asset portfolio. The authors prove that the sample estimator of such a frontier is biased and provide, under proper but questionable hypotheses, an analytical expression for its unbiased estimator. In this contribution, first, we study the behavior of the unbiased estimator of the efficient frontier when the length of the return time series tends to infinity, then, we investigate a "strange" behavior of the unbiased estimator in correspondence of particular combinations of the means of the returns of the assets and of their variances and covariances with respect to the number of the assets and the length of the associated time series of returns; finally, we analyze the operational effectiveness of the proposed unbiased estimator by a bootstrap-based approach.

Keywords Efficient frontier · Unbiased estimator · Asymptotic and "strange" behaviors · Bootstrap estimator

1 Introduction

In a couple of recent papers of Bodnar and Bodnar on the estimation of the efficient frontier of a portfolio of risky assets [1, 2], the Authors prove that the sample estimator of such a frontier is biased and provide, under proper but questionable hypotheses, an analytical expression for its unbiased estimator. Furthermore, they highlight that the sample estimator of the efficient frontier is overoptimistic, in the sense that the latter systematically underestimates the variance of each efficient portfolio.

M. Corazza (✉) · C. Pizzi
Department of Economics, Ca' Foscari University of Venice, Venice, Italy
e-mail: corazza@unive.it; pizzic@unvie.it

M. Corazza et al. (eds.), *Mathematical and Statistical Methods for Actuarial Sciences and Finance*, https://doi.org/10.1007/978-3-319-89824-7_44

In this contribution: first, we study the behavior of the unbiased estimator of the efficient frontier when the length of the return time series tends to infinity; then, we investigate a "strange" behavior of the same unbiased estimator in correspondence of particular combinations of the means of the returns of the assets and of their variances and covariances with respect to the number of the assets and the length of the associated time series of returns; finally, we analyze the operational effectiveness of the proposed unbiased estimator by a bootstrap-based approach.

The remainder of this note is organized as follows. In the next section, we synthetically present the model of O. Bodnar and T. Bodnar. In Sect. 3, we provide our main results.

2 Unbiased Estimator of the Efficient Frontier

The unbiased estimator of the efficient frontier proposed by Bodnar and Bodnar [2] is based on the following hypotheses:

1. All the considered $K > 2$ assets are risky;
2. The returns of these assets are serial independent;
3. The returns of these assets are normally distributed;
4. Indicated by N the length of the return time series, $N > K$.

Note that, without entering into detailed discussion, there exists an extensive literature following which the hypotheses (2) and (3) do not appear particularly realistic.

Given these hypotheses, the Authors prove that the following sample estimator of the efficient frontier is biased:

$$BEF := \left(R - \hat{R}_{GMV}\right)^2 - \hat{s}\left(V - \hat{V}_{GMV}\right)^2 = 0$$

where

R is the expected return of the portfolio,
$\hat{R}_{GMV}$ is the sample estimate of the expected return of the global minimum variance portfolio,
$\hat{s} = \hat{\boldsymbol{\mu}}' \hat{\boldsymbol{R}} \hat{\boldsymbol{\mu}}$, in which: $\hat{\boldsymbol{\mu}}$ is the vector of the sample estimates of the expected rates of return of the assets; $\hat{\boldsymbol{R}} = \hat{\boldsymbol{\Sigma}}^{-1} - \left(\hat{\boldsymbol{\Sigma}}^{-1}\mathbf{1}'\mathbf{1}\hat{\boldsymbol{\Sigma}}^{-1}\right) \Big/ \left(\mathbf{1}'\hat{\boldsymbol{\Sigma}}^{-1}\mathbf{1}\right)$, where $\hat{\boldsymbol{\Sigma}}$ is the matrix of the sample estimates of the variances and of the covariances of the rates of return, and $\mathbf{1}$ is a vector of ones,
V is the variance of the return of the portfolio,
$\hat{V}_{GMV}$ is the sample estimate of the variance of the return of the global minimum variance portfolio.

Then, under the same conditions, the Authors provide the following expression for the unbiased estimator of the efficient frontier:

$$UEF := \left(R - \hat{R}_{GMV}\right)^2 - A \cdot \hat{V}_{GMV} - \left(B \cdot \hat{s} - C\right)\left(V - D \cdot \hat{V}_{GMV}\right)^2 = 0$$

where

$A = [(n-2)(n-1)] \big/ [(n-k)(n-k+1)n]$,
$B = (n-k-1) \big/ [(n-k)(n-1)]$,
$C = (k-1)/n$,
$D = [(n-k-2)(n-1)] \big/ [(n-k)(n-k+1)]$.

Finally, the Authors present an empirical application in which they consider time series of monthly returns from the equity markets of ten developed countries. Note that for these monthly returns, the above hypotheses (2) and (3) seem acceptable.

3 Our Main Results

With respect to the framework described above, in this section we present our main results.

3.1 The Asymptotic Behavior of the Biased Estimator

First, we study the behavior of the unbiased estimator of the efficient frontier when the length of the return time series tends to infinity. In this regard, we present the statement of our following theoretical result.

Theorem 1 $\lim_{N \to +\infty} BEF = UEF$.

From practical purpose, this result ensures the convergence of the biased efficient frontier to the unbiased one when the length of the return time series is large enough.

3.2 The "Strange" Behavior of the Unbiased Estimator

The unbiased estimator of the efficient frontier is characterized by a "strange" behavior in correspondence of particular combinations of the means of the returns of the assets and of their variances and covariances with respect to the number of the assets and the length of the associated time series of returns. In fact, in these cases, the estimator of the efficient frontier develops toward North-West instead of towards North-East, as usual (and theoretically correct). In particular, this "strange"

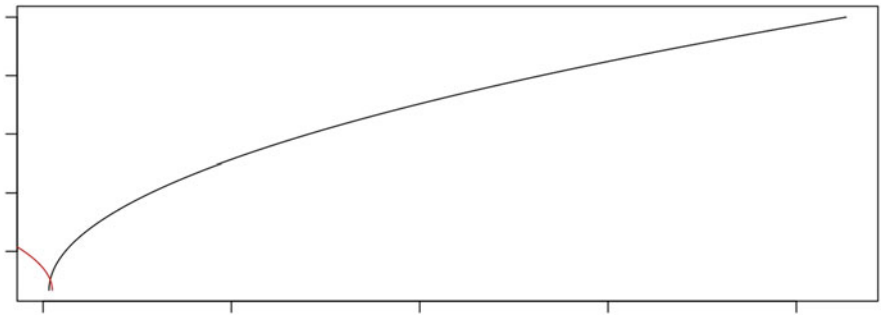


Fig. 1 Example of the "strange" behavior of the unbiased estimator of the efficient frontier. As usual, the x-axis stands for the variance of the return of the portfolio and the y-axis stands for the expected return of the portfolio

behavior appears when $N \gg 0$ and the returns are daily. In Fig. 1 we give a graphical exemplification of such an occurrence: the North-East efficient frontier and the North-West one are both generated by the same stocks but using different subsets of the original return time series.

Also in this regard, we present the statement of our following theoretical result.

Theorem 2 *The unbiased estimator of the efficient frontier develops towards North-West when*

$$\hat{s} < \frac{(n-1)(k-1)}{n(n-k-1)} \qquad (*)$$

3.3 The Operational Effectiveness of the Unbiased Estimator

In conclusion, an important question rises: is the proposed unbiased estimator of efficient frontier operationally effective?

In order to give an answer to this question, given a set of real assets which are components of the Italian stock index FTSE MIB, we perform two empirical analysis.

– The first one concerns the determination of the confidence interval of the sample estimate of the efficient frontier through a bootstrap approach (see [3]) by using 100 resampling. In particular, for preserving the autocorrelation structure of the return time series we adopt the block bootstrap (see [4]). We obtain that the unbiased estimate of the efficient frontier is not statistically different with respect to the sample estimate of the same efficient.

– The second empirical analysis, similar to the previous one, is based on 100 simulated portfolio with different distributional properties and different combination of means, variances and covariances of the returns of the assets. By doing so, we obtain interesting results. In fact, different distributional properties produce different results. In case of the normal probability distribution, in the 69% of the simulations the unbiased estimate of the efficient frontier and the sample estimate of the same frontier are statistically equal, while in case of the Student's t probability distribution only the 28% are. Furthermore, repeating this analysis by removing the serial independence in the return time series, the percentage in case of normal probability distribution and of Student's t one are similar (54% and 58%, respectively).

References

1. Bodnar, O., Bodnar, T.: Statistical inference procedure for the mean–variance efficient frontier with estimated parameters. Adv. Stat. Anal. **93**, 295–306 (2009)
2. Bodnar, O., Bodnar, T.: On the unbiased estimator of the efficient frontier. Int. J. Theor. Appl. Financ. **7**, 1065–1073 (2010)
3. Efron, B., Tibshirani, R.: Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy. Stat. Sci. **1**, 54–75 (1986)
4. Politis, D.N., Romano, J.P.: The stationary bootstrap. J. Am. Stat. Assoc. **89**, 1303–1313 (1994)

Numerical Solution of the Regularized Portfolio Selection Problem

Stefania Corsaro, Valentina De Simone, Zelda Marino, and Francesca Perla

Abstract We investigate the use of Bregman iteration method for the solution of the portfolio selection problem, both in the single and in the multi-period case. Our starting point is the classical Markowitz mean-variance model, properly extended to deal with the multi-period case. The constrained optimization problem at the core of the model is typically ill-conditioned, due to correlation between assets. We consider l_1-regularization techniques to stabilize the solution process, since this has also relevant financial interpretations.

Keywords Portfolio optimization · l_1 Regularization · Bregman iteration

1 Introduction

In this work we discuss the numerical solution of the portfolio selection problem. Our starting point is the classical Markowitz mean-variance framework, in which one aims at the construction of an investment portfolio that exposes investor to minimum risk providing him a fixed expected return. A common strategy to estimate Markowitz model parameters is to use historical data as predictive of the future behaviour of asset returns. This typically leads to ill-conditioned numerical problems. We then consider l_1 regularization techniques; the single-period regularized model was introduced in [3], where a l_1-penalty term is added to the objective function of the optimization problem at the core of the model. This has also nice

S. Corsaro (✉) · Z. Marino · F. Perla
Dipartimento di Studi aziendali e quantitativi, Università degli Studi di Napoli "Parthenope", Napoli, Italy
e-mail: stefania.corsaro@uniparthenope.it; zelda.marino@uniparthenope.it; francesca.perla@uniparthenope.it

V. De Simone
Dipartimento di Matematica e Fisica, Università degli Studi della Campania "Luigi Vanvitelli", Caserta, Italy
e-mail: valentina.desimone@unicampania.it

M. Corazza et al. (eds.), *Mathematical and Statistical Methods for Actuarial Sciences and Finance*, https://doi.org/10.1007/978-3-319-89824-7_45

financial interpretations, both in terms of transaction costs and control of short-positions. We then extend the regularized model to the multi-period case. Our model satisfies time consistency, a fundamental requirement in this framework. Different definitions of time consistency can be found in literature, either related to dynamic risk measures or investment policies [5]; this concept deals with the consistency over time of decisions taken with the support of evolving available information. We discuss the numerical solution in both cases. We develop iterative algorithms based on Bregman iteration method, that converts the constrained problem into a short sequence of unconstrained ones. The presence of the l_1-term makes the solution of the involved optimization sub-problem not trivial, thus we apply ad hoc methods to deal with non-smoothness [1].

In Sect. 2 we describe the regularized portfolio selection model; in Sect. 3 we describe Bregman iteration method.

2 Regularized Portfolio Selection Model

Let n be the number of traded assets. We assume self-financing investment strategies, both in the single and in the multi-period case. We start by describing the static mean-variance problem. We suppose that one unit of capital is available and define

$$\mathbf{w} = (w_1, w_2, \ldots, w_n)^T$$

the portfolio weight vector, where w_i is the amount invested in the i-th security. We furthermore denote with

$$\mathbf{r} = (r_1, r_2, \ldots, r_n)^T$$

the vector of expected asset returns. Regularized portfolio selection is formulated as the following quadratic constrained optimization problem:

$$\begin{aligned} &\min_{\mathbf{w}} \mathbf{w}^T \Sigma \mathbf{w} + \tau \|\mathbf{w}\|_1 \\ &\text{s.t.} \\ &\mathbf{w}^T \mathbf{1}_n = 1, \\ &\mathbf{w}^T \mathbf{r} = \rho \end{aligned}$$

where $\mathbf{1}_n$ is the column vector of ones of dimension n, ρ is the fixed expected portfolio return, and Σ is the covariance matrix of returns. The first constraint is a budget constraint which establishes that all the available capital is invested. The second one fixes the expected return.

Let us now turn to the dynamic case and consider m dates, which define $m-1$ periods of investment. Decisions are assumed at time t_i, $i = 1, \ldots, m-1$; decision

taken at time t_i is kept in the period $[t_i, t_{i+1})$. Portfolio weights and asset returns are now stored in matrices $W \in \Re^{n\times m}$, $R \in \Re^{n\times(m-1)}$, in which the i-th columns $\mathbf{w}_i, \mathbf{r}_i$ contain, respectively, the weight vector at time t_i and the expected return vector in the period $[t_i, t_{i+1})$. Regularized portfolio selection is formulated as the following constrained optimization problem:

$$\min_W \sum_{i=1}^{m} \left[\mathbf{w}_i^T \Sigma_i \mathbf{w}_i + \tau \|\mathbf{w}_i\|_1 \right]$$

s.t.

$$\mathbf{w}_1^T \mathbf{1}_n = 1 \tag{1}$$

$$\mathbf{w}_i^T \mathbf{1}_n = (1 + r_{i-1})^T \mathbf{w}_{i-1}, \quad i = 2, \ldots, m \tag{2}$$

$$\mathbf{w}_n^T \mathbf{1}_n = x_{\text{term}} \tag{3}$$

where x_{term} is the expected wealth provided by the overall investment and Σ_i is the covariance matrix estimated in the i-th period. As in the one-period case, constraint (1) is the budget constraint, constraint (2) means that the investment strategy is self-financing, thus, at the end of each period the wealth is given by the revaluation of the previous one. Finally, constraint (3) fixes the investment target. We adopt a separable formulation for the risk measure, so, following [4], we show that our approach is time consistent.

3 Bregman Iteration for Portfolio Selection

Regularized portfolio selection can be in general formulated as the constrained nonlinear optimization problem:

$$\begin{aligned} &\min_{\mathbf{w}} E(\mathbf{w}) \\ &\text{s.t.} \\ &A\mathbf{w} = \mathbf{b}, \end{aligned} \tag{4}$$

where, defined $M = m + 1$, $N = m \cdot n$, the functional $E(\mathbf{w}) : \Re^N \longrightarrow \Re$ is strictly convex and non-smooth due to the presence of the l_1-penalty term and $A \in \Re^{M\times N}$ is the matrix form of the constraints and $\mathbf{b} \in \Re^M$. Bregman iteration can be used to reduce (4) in a short sequence of unconstrained problems by using the *Bregman distance* associated with E [2].

The Bregman distance associated with a proper convex functional $E(\mathbf{w})$ at point $\mathbf{v}$ is defined as:

$$D_E^{\mathbf{p}}(\mathbf{w}, \mathbf{v}) = E(\mathbf{w}) - E(\mathbf{v}) - < \mathbf{p}, \mathbf{w} - \mathbf{v} >,$$

where $\mathbf{p} \in \partial E(\mathbf{v})$ is a subgradient in the subdifferential of E at point $\mathbf{v}$ and $< ., . >$ denotes the canonical inner product in R^N. First of all, the constrained problem (4) is converted into the unconstrained one:

$$\min_{\mathbf{w}} E(\mathbf{w}) + \tfrac{\lambda}{2}\|A\mathbf{w} - \mathbf{b}\|_2^2 \tag{5}$$

for a fixed $\lambda > 0$. Then, at each Bregman iteration $E(\mathbf{w})$ is replaced by the Bregman distance so a sub-problem in the form of (5) is solved according to the following iterative scheme:

$$\begin{cases} \mathbf{w}_{k+1} = \operatorname{argmin}_{\mathbf{w}} D_E^{\mathbf{p_k}}(\mathbf{w}, \mathbf{w}_k) + \frac{\lambda}{2}\|A\mathbf{w} - \mathbf{b}\|_2^2, \\ \mathbf{p}_{k+1} = \mathbf{p}_k - \lambda A^T(A\mathbf{w}_{k+1} - b) \in \partial E(\mathbf{w}_{k+1}). \end{cases} \tag{6}$$

Under suitable hypotheses the convergence of the sequence $\{\mathbf{w}_k\}$ to the solution of the constrained problem (4) is guaranteed in a finite number of steps [6]. Since there is generally no explicit expression for the solution of the sub-minimization problem involved in (6), at each iteration the solution is computed inexactly using an iterative solver. At this purpose, we focus on first order methods, which are gradient-based that converge rather slowly; however, for large problem dimensions, usually a fast lower-precision solution is favoured. In particular, we use the Fast Proximal Gradient method with backtracking stepsize rule (FISTA) [1], an accelerated variant of Forward Backward algorithm, suitable for minimizing convex objective functions given by summation of smooth and non-smooth terms. We test our algorithms on real market data, and validate our approach observing the out-of-sample performances of optimal portfolios.

Acknowledgements This work was partially supported by the Research grant of Università Parthenope, DR no. 953, November 28th, 2016, and by INdAM-GNCS, under projects 2017 and 2018.

References

1. Beck, A., Teboulle, M.: A fast iterative shrinkage-thresholding algorithm for linear inverse problems. SIAM J. Imag. Sci. **2**, 183–202 (2009)
2. Bregman, L.: The relaxation method of finding the common point of convex sets and its application to the solution of problems in convex programming. USSR Comput. Math. Math. Phys. **7**, 200–217 (1967)
3. Brodie J., Daubechies I., De Mol, C., Giannone, D., Loris, I.: Sparse and stable Markowitz portfolios. PNAS **106**, 12267–12272 (2009). https://doi.org/10.1073/pnas.0904287106
4. Chen, Z., Guo, J.E., Li, G.: Optimal investment policy in the time consistent mean–variance formulation. Insur. Math. Econ. **52**, 145–256 (2013)
5. Chen, Z., Consigli, G., Liu, J., Li, G., Hu, Q.: Multi-period risk measures and optimal investment policies. In: Consigli, G., Kuhn, D., Brandimarte, P. (eds.) Optimal Financial Decision Making under Uncertainty, pp. 1–34. Springer International Publishing, Cham (2017)
6. Osher, S., Burger, M., Goldfarb, D., Xu, J., Yin, W.: An iterative regularization method for total variation-based image restoration. SIAM Multiscale Model. Simul. **4**(2), 460–489 (2005)

Forecasting the Equity Risk Premium in the European Monetary Union

David Cortés-Sánchez and Pilar Soriano-Felipe

Abstract This article examines the performance of several variables that could be good predictors of the equity risk premium in the European Monetary Union for a period that spans from 2000 to 2015. In-sample, technical indicators display predictive power, matching or exceeding that of traditional economic forecasting variables. We also find consistent results in the fact that combining information from technical and economic variables improves equity risk premium forecasts, compared to using these variables alone. Nevertheless, out-of-sample exercises do not confirm in-sample results. Economic predictors show stronger out-of-sample forecasting ability than technical indicators, and apart from volume rules, technical indicators do not show forecasting power. Overall, only a few economic and technical predictors display forecasting power in-sample and out-of-sample, and provide economic value for a risk-averse investor.

Keywords Equity risk premium · Economic variables · Technical indicators · Principal components analysis · In-sample forecasts · Out-of-sample forecasts · Asset allocation

1 Introduction

Forecasting stock returns is one of the most popular themes to both academics and practitioners in finance. The existing literature has studied many types of variables and proposed multiple econometric models trying to see if there is significant evidence of returns predictability. It is generally accepted among financial economist, that stock returns contain a significant predictable component in-sample. For example Rozeff [1], Campbell and Shiller [2] or Cochrane [3] find evidence in favor of return predictability using the dividend yield. Nonetheless, when we

D. Cortés-Sánchez (✉) · P. Soriano-Felipe
Universitat de Valencia, Valencia, Spain
e-mail: david.cortes@bbva.com

M. Corazza et al. (eds.), *Mathematical and Statistical Methods for Actuarial Sciences and Finance*, https://doi.org/10.1007/978-3-319-89824-7_46

review economic literature that focus on out-of-sample forecasting ability, there is not such consensus among economists. Bossaerts and Hillion [4], and Welch and Goyal [5] show that a long list of predictors from the literature are unable to perform consistently better out-of-sample than a simple forecast based on the historical average.

In this paper, as suggested by Neely et al. [6], we investigate the capacity of multiple economic variables and technical indicators to forecast the equity risk premium. We evaluate the same variables and econometric models, but for a European data set.

2 In-Sample Analysis

The conventional framework for analyzing the equity risk premium predictability is based on simple linear regressions. The idea is to run a predictive linear regression of realized excess of returns on lagged explanatory variables.

$$r_{t+1} = \alpha_i + \beta_i X_t + \varepsilon_{t+1} \tag{1}$$

Where the equity risk premium, r_{t+1}, is the return on a broad stock market index in excess of a risk free asset from period t to t + 1, x_t is the predictor, and ε_{t+1} is a zero mean disturbance.

We replicate Neely et al. [6] paper but for the European Monetary Union (EMU), to see if we can get similar results as they obtained for the United States. In this regard, we use the same variables,[1] and we work with 12 economic variables, and 14 technical indicators. Data spans from January 2000 to September 2015, frequency is monthly, and to calculate the equity return, we selected a popular stock market index, the MSCI EMU index.

Adding more predictors could incorporate relevant information and improve regression forecasting accuracy. Hence we also estimate multivariate predictive regressions, but instead of using all the predictor variables, we use principal components analysis to summarize all relevant information of our predictors into few principal components. To calculate these components we work with three sets of variables: F^{ECO}, calculates principal components for all economic predictors, F^{TECH} for all the technical indicators, and F^{ALL} takes into account all predictive variables.

[1] Economic Variables include: Dividend-Price ratio, Dividend-Yield ratio, Earnings-Price ratio, Payout ratio, Book-to-Market ratio, Euribor 3M rate, 10Y Euro Swap rate, Term Spread, Default Yield Spread, Default Return Spread, Inflation rate and Equity Risk Premium Volatility. Technical Indicators include: Moving Averages rules, Momentum rules and Volume rules.

3 Out-of-Sample Analysis

As a robustness check, we run out-of-sample forecasting tests to check if in-sample results are also valid out-of-sample. To calculate out-of-sample forecasts, we divide the whole period of data into two subperiods.

To analyze out-of-sample stock return predictability, the accuracy of the proposed predictive regressions forecasts is compared to the historical mean average. Moreover, the Mean Squared Forecast Error (MSFE) is calculated to evaluate forecast accuracy, and we analyze forecasts in terms of the Campbell and Thomson [7] out-of-sample R-squared (R^2_{OS}), which compares the MSFE of regressions constructed with selected predictors to the MSFE of the benchmark.

$$R^2_{os} = 1 - \frac{MSFE_r}{MSFE_b} \tag{2}$$

Where $MSFE_r$ is the mean squared forecast error of the predictive regression, and $MSFE_b$ is the mean squared forecast error of the benchmark. If $R^2_{OS} > 0$, then the predictive regression forecast relative to historical average is more accurate. But, if $R^2_{OS} < 0$, then the opposite happens.

4 Asset Allocation

As a final exercise, we measure if there is any economic value in the out-of-sample predictions for a risk-averse investor that has a quadratic utility function. We maximize investors' utility function using a simple asset allocation model that, either invests in equity assets, or in the risk-free asset. If predictive regressions improve investors' utility relative to the benchmark predictions, then those predictors create economic value for this investor.

5 Summary of Most Important Results

Table 1 summarizes all the results obtained in previous sections.

First, there are six variables that show ability to forecast monthly equity risk premiums in-sample and out-sample, and moreover produce utility gains for a risk-averse invertor. Three of these predictors are economic, the long and short interest rates (EUR3M, SWAP10), and the book market value (BM). The other three predictors are technical, and are on-balance volume indicators OBV212, OBV39 and OBV312. These results are similar to those found by Neely et al. [6].

Second, this exercise also finds evidence that multivariate analysis using PCA gathers relevant information of economic and technical predictors, shows predictive

Table 1 Summary of in-sample, out-of-sample and asset allocation results

Predictor	In-sample		Out-of-sample		A.A. performance	
	t-stat	R2 (%)	MSFE-Adj	R2OS (%)	Relative gain (%)	Sharpe
DP	0.17	0.02	0.96	0.36	0.00	0.00
DY	0.31	0.07	1.52*	0.89	0.00	0.00
EP	0.17	0.02	1.95**	0.48	0.00	0.00
DE	0.48	0.12	0.76	1.08	−0.34	−0.26
RVOL	−0.21	0.03	−0.85	−0.67	0.00	0.00
BM	2.27**	2.80	2.54***	8.04	0.85	0.25
EUR3M	−3.58***	6.49	1.6*	−2.88	0.84	0.15
SWAP10Y	−2.97***	4.04	1.55**	−9.96	0.64	0.14
TMS	2.32***	3.26	0.19	−0.90	0.62	0.21
DFY	−1.55	1.78	1.37*	4.01	−0.40	−0.01
DFR	−1.26	1.39	1.57*	4.78	−0.19	−0.01
INF	−2.97***	5.44	1.03	−22.65	−1.75	0.09
MA1MA9	2.52***	4.25	0.46	−7.00	−0.49	0.03
MA1MA12	2.52***	4.44	1.45	−0.15	1.09	0.10
MA2MA9	2.31***	3.65	1.07	−1.34	0.23	0.06
MA2MA12	2.04***	2.90	0.99	−1.36	−0.07	0.04
MA3MA9	2.23***	3.44	0.59	−9.56	−0.19	0.03
MA3MA12	1.83***	2.34	1.55	1.69	0.24	0.05
MOM9	2.25***	3.58	0.7	−1.60	−0.33	0.03
MOM12	1.46	1.46	1.51	2.07	−0.23	0.00
OBV19	1.85***	2.24	0.22	−2.02	−0.72	−0.01
OBV112	2.78***	4.80	0.75	−1.55	−0.29	0.05
OBV29	2.93***	5.24	1.68	2.11	3.52	0.21
OBV212	3.07***	5.97	2.32*	5.85	4.58	0.23
OBV39	3.05***	5.66	2.03*	4.56	4.19	0.22
OBV312	3.34***	7.09	2.99**	9.55	6.26	0.28
FECO	5.15***	7.75	1.46*	−4.39	0.51	0.14
FTECH	5.56***	5.77	1.33	−1.09	2.62	0.16
FALL	6.22***	9.19	1.74*	−0.25	2.22	0.14

R2 is the simple R2 obtained from ols estimation and R2OS is the Campbell and Thompson [7] R2 Relative gain id the difference between the CER generated by a predictor and the CER of the benchmark

*, ** and *** indicate significance at the 10%, 5% and 1% levels, respectively. Based on two-tail wild bootstrapped p-values for in-sample estimations and on Clark and West [8] MSFE-adj for out-of-sample predictions

power, and creates economic value for risk-averse investors. In most of the cases, multivariate predictive regressions using principal components outperform forecast based on individual predictors. This is also line with results obtained by Neely et al. [6].

Third, technical indicators display stronger forecasting ability in-sample, but this result is not confirmed out-of-sample, where economic variables show a better

performance than technical indicators. This is also confirmed by the asset allocation exercise, where most of the trend-following rules produce more trading and higher transaction costs, but underperform the historical average benchmark. This third point clearly contradicts Neely et al. [6] findings that monthly equity risk premium forecast based on technical indicators outperform forecasts based on economic variables.

Fourth, within the economic variables, macroeconomic variables perform better than valuation ratios for short term forecasting and portfolio rebalancing. Further studies using longer forecasting periods might turn around this result and improve valuation ratios forecasting power.

References

1. Rozeff, M.S.: Dividend yields are equity risk premiums. J. Portf. Manag. **11**, 68–75 (1984)
2. Campbell, J.Y., Shiller, R.J.: The dividend-price ratio and expectations of future dividends and discount factors. Rev. Financ. Stud. **1**, 195–228 (1988)
3. Cochrane, J.H.: The dog that did not bark: a defense of return predictability. Rev. Financ. Stud. **21**, 1533–1575 (2008)
4. Bossaerts, P., Hillion, P.: Implementing statistical criteria to select return forecasting models: what do we learn. Rev. Financ. Stud. **12**, 405–428 (1999)
5. Welch, I., Goyal, A.: A comprehensive look at the empirical performance of equity premium prediction. Rev. Financ. Stud. **21**, 1455–1508 (2008)
6. Neely, J., Rapach, D.E., Tu, J., Zhou, G.: Forecasting the equity risk premium: the role of technical indicators. Manag. Sci. **60**, 1772–1791 (2014)
7. Campbell, J.Y., Thompson, S.B.: Predicting excess stock returns out of sample: can anything beat the historical average. Rev. Financ. Stud. **21**(4), 1509–1531 (2008)
8. Clark, T.E., West, K.D.: Approximately normal tests for equal predictive accuracy in nested models. J. Econ. **138**, 291–311 (2007)

Statistical Learning Algorithms to Forecast the Equity Risk Premium in the European Union

David Cortés-Sánchez and Pilar Soriano-Felipe

Abstract With the explosion of "Big Data", the application of statistical learning models has become popular in multiple scientific areas as well as in marketing, finance or other business disciplines. Nonetheless, there is not yet an abundant literature that covers the application of these learning algorithms to forecast the equity risk premium. In this paper we investigate whether Classification and Regression Trees (CART) algorithms and several ensemble methods, such as bagging, random forests and boosting, improve traditional parametric models to forecast the equity risk premium. In particular, we work with European Monetary Union data for a period that spans from the EMU foundation at the beginning of 2000 to half of 2017.

The paper first compares monthly out-of-sample forecasting ability of multiple economic and technical variables using linear regression models and regression trees techniques. To check the out-of-sample accuracy, predictive regressions are compared to a popular benchmark in the literature: the historical mean average. Forecasts performance is analyzed in terms of the Campbell and Thompson R-squared (R^2_{OS}), which compares the MSFE of regressions constructed with selected predictors, against the MSFE of the benchmark.

Keywords CART models · Ensemble techniques · Equity risk premium forecast

1 Introduction

Forecasting stock returns is one of the most popular themes to both academics and practitioners in finance. The existing literature has studied many types of variables and proposed multiple econometric models trying to see if there is significant evidence of returns predictability. A big part of the literature that tries to find out-of-

D. Cortés-Sánchez (✉) · P. Soriano-Felipe
Universitat de Valencia, Valencia, Spain
e-mail: david.cortes@bbva.com

M. Corazza et al. (eds.), *Mathematical and Statistical Methods for Actuarial Sciences and Finance*, https://doi.org/10.1007/978-3-319-89824-7_47

sample evidence of the equity risk premium predictability has focused on parametric models, mainly simple linear regressions models, and little attention has been given to statistical learning algorithms.

In this paper we focus our attention on regression trees that were introduced by Breiman et al. [1]. We selected these algorithms because they are simple to implement and to interpret, they are not subject to specific model assumptions and they can deal with multiple explicative variables giving consistent results. Nonetheless, prediction results given by regression trees are quite unstable. A solution to this problem of instability is to apply ensemble methods that create multiple trees through bootstrapping techniques, or any other iteration process, and ensemble all results obtained.

2 Methodology

2.1 CART Models

CART is a non-parametric modeling technique that fixes a set of rules upon the explanatory variables to classify the explained variable into categories. It looks at the variables in the data set, determines which are most important, and results in a tree of decisions that best partition the data. The principal idea behind a decision tree is to recursively partition the space into smaller sub-spaces where similar response values are grouped. After the partition is completed, a constant value of the response variable is predicted within each area.

The basic steps to construct a classification or regression tree are:

1. **Start with and empty tree**
2. **Select a feature to split the data.** Tree-structured classifiers are constructed by making repetitive splits of space X, so that a hierarchical structure is formed. The process follows a top-down approach and it is applied at every node and for every k explicative variables. It is known as "top-down greedy recursive partitioning". It begins at the top of the tree, includes all the data, and then successively splits the predictor space. Each split is indicated via two new branches further down on the tree. It is considered "greedy" because at each step the best split is made at that particular step, without any consideration of whether those choices remain optimal in future stages.
3. **Stop splitting and decide which nodes are terminal.** Node splitting would continue recursively until some stop condition is reached and the grown tree has the right complexity. There are several stop criteria, but one of the best approaches was proposed by Breiman et al. [1] and consists to let the tree grow to saturation and then prune it back.

2.2 Ensemble Models

CART models show a high degree of instability to small changes in the learning data. A solution to this problem is given by ensemble methods that generate multiple training samples, get separate predictions for each sample, and somehow aggregate those multiple predictions. The most common types of ensembles used with trees are Bagging, Random Forest and Boosting. All these techniques are applied in the paper.

3 The Model

We used a regression tree model with the following form:

$$ERP_t = \sum_{m=1}^{M} c_m x \prod_{i=1}^{n} I\ (x_i \in R_m) \tag{1}$$

Where ERP_t is the Equity Risk Premium in moment t; C_m are constants; I(.) is an indicator function returning 1 if its argument is true and 0 otherwise; Xi is the lagged predictor variable I; and R1 . . . Rm represent a partition of the feature space.

3.1 The Data

Our empirical analysis is conducted for the European Monetary Union with a monthly dataset, and covering a period that goes from June 2000 to July 2017. Data is also divided into two groups: the training set that is used to build initial trees, and the test set, which checks if forecasted results perform well out of sample. Three training sets were tested: June 2000–December 2012, June 2000–December 2013 and June 2000–December 2014.

Table 1 show all predictors used in the algorithms. These variables were measured in levels, 1 month changes, and 3 months and 6 months z-scores. Z-scores are normalized variables calculated as the value in a particular date minus its mean divided by its standard deviation. This gives a total of 104 explicative variables that are loaded into the algorithm.

Table 1 Predictive variables

Economic and technical variables				
Valuation/fundamentals	Rates and inflation	Risk indicators	Confidence surveys	Technical indicators
Dividend yield	Euribor 3M	VIX	Consumer confidence	Moving averages
Price to book value	Swap rate 2Y	ASW IG corporates	EU business climate indicator	OBV
Price to earnings (trailing)	Swap rate 10Y	ASW HY corporates	Industrial confidence	Earnings momentum
Price to earnings (forward)	Yield curve 3m-10Y	Spread IG-HY corporates	Retail confidence	
Price to earnings (trailing, 5Y averages)	Yield curve 2Y-10Y	TED spread	OECD EU leading indicator	
Price to earnings (forward, 5Y averages)	NYMEX level			
	Consumer prices			

3.2 Out-of-Sample Predictions

To analyze out-of-sample stock return predictability, the accuracy of the proposed predictive regressions forecasts is compared to the historical mean average. Moreover, the Mean Squared Forecast Error (MSFE) is calculated to evaluate forecast accuracy, and we analyze forecasts in terms of the Campbell and Thomson [2] out-of-sample R-squared (R^2_{OS}), which compares the MSFE of regressions constructed with selected predictors to the MSFE of the benchmark. If $R^2_{OS} > 0$, then the predictive regression forecast relative to historical average is more accurate. But, if $R^2_{OS} < 0$, then the opposite happens.

4 Results

Table 2 illustrates the most interesting results. The first one is that regression trees techniques do not always outperform simple linear regression models. In most of the cases, ensemble regression trees show a better performance than simple linear regressions, but in the case of the OECD leading indicator in 1 month changes (OECD_1MCHG), for all the out-of-sample periods tested, linear regression showed higher forecasting ability. For instance, for the out-of-sample period 2015–2017 the R^2_{OS} was 19.1% for the OECD_1MCHG and 17.1% for the boosting algorithm. Nonetheless, ensemble regression trees outperformed the other 103 regressions built with the rest of predictor variables.

Another interesting result is that, among regression trees, boosting was the best performer algorithm. In all periods, the rank between regression trees models was the same, and all the models except the not-pruned regression tree beat the benchmark. Best forecasting ability was achieved by boosting methods, followed by random forest, bagging and pruned trees. Not pruned regression trees displayed lower forecasting ability that its benchmark, having negative R^2_{OS} in all periods. This confirms that letting trees to grow tends to overfit the learning sample, fitting non-systematic changes of the data, and leading to poor out-of-sample forecast.

Table 2 Predictive variables

		Testing periods					
		2013–2017		2014–2017		2015–2017	
		MSFE	R^2_{OS} (%)	MSFE	R^2_{OS} (%)	MSFE	R^2_{OS} (%)
	Benchmark	0.156		0.161		0.195	
Top-5 simple linear regression forecasts	OECDLEAD_1MCHG	0.125	19.6	0.130	19.0	0.158	19.1
	SWAP2A_1MCHG	0.119	23.8	0.143	10.9	0.169	13.6
	MOVAVGS_1MCHG	0.141	9.5	0.146	9.5	0.182	7.0
	BUSCLIM_1MCHG	0.145	7.0	0.150	6.7	0.182	6.7
	EMOM	0.146	6.8	0.152	5.5	0.187	4.4
Regression trees forecasts	Tree	0.229	−46.8	0.231	−43.5	0.218	−11.8
	Pruned tree	0.148	4.9	0.157	2.6	0.190	2.5
	Bagging	0.141	9.9	0.149	7.0	0.178	8.8
	Random forest	0.137	12.1	0.148	7.7	0.176	10.1
	Boosting	0.133	14.8	0.142	11.9	0.162	17.1

References

1. Breiman, L., Friedman, J.H., Olshen, J.H., Stone, C.I.: Classification and Regression Trees. Wadsworth, Belmont (1984)
2. Campbell, J.Y., Thompson, S.B.: Predicting excess stock returns out of sample: can anything beat the historical average? Rev. Financ. Stud. **21**(4), 1509–1531 (2008)

Evaluating Variable Annuities with GMWB When Exogenous Factors Influence the Policy-Holder Withdrawals

Massimo Costabile, Ivar Massabó, and Emilio Russo

Abstract We propose a model for evaluating variable annuities with guaranteed minimum withdrawal benefits in which a rational policy-holder, who would withdraw the optimal amounts maximizing the current policy value only with respect to the endogenous variables of the evaluation problem, acts in a more realistic context where her/his choices may be influenced by exogenous variables that may lead to withdraw sub-optimal amounts. The model is based on a trinomial approximation of the personal sub-account dynamics that, despite the presence of a downward jump due to the payed withdrawal at each anniversary of the contract, guarantees the reconnecting property. A backward induction scheme is used to compute the insurance fair fee paid for the guarantee.

Keywords Variable annuity · Guaranteed minimum withdrawal benefits · Trinomial tree

1 The Framework

We consider a single premium variable annuity (VA) with GMWB issued at time 0 and maturing at time T with withdrawals allowed at each contract anniversary $t_h = h$, with $h = 1, \ldots, T$, versus an initial investment W_0 in a fund. As a consequence, such a contract may be described introducing two accounts for the policy-holder: a personal sub-account invested in a well diversified reference fund made up of risky assets, whose value at a generic time $t \in [0, T]$ is denoted by $W(t)$, and a guarantee account, having value A_t at the same time t. The initial investment W_0 forms the initial balance of both the accounts, i.e., $W(0) = W_0$, and $A_0 = W_0$, as each withdrawal reduces both the accounts. The personal sub-account presents the

M. Costabile · I. Massabó · E. Russo (✉)
Department of Economics, Statistics, and Finance, University of Calabria, Arcavacata di Rende (CS), Italy
e-mail: emilio.russo@unical.it

M. Corazza et al. (eds.), *Mathematical and Statistical Methods for Actuarial Sciences and Finance*, https://doi.org/10.1007/978-3-319-89824-7_48

following dynamics under the risk-neutral probability measure,

$$dW(t) = (r - \alpha)W(t)dt + \sigma W(t)dB(t) \quad \text{with} \quad t_h \leq t < t_{h+1}, \tag{1}$$

where r is the constant instantaneous risk-free rate of return, α captures the insurance fee for the GMWB guarantee, σ is the return volatility, and $B(t)$ is a standard Brownian motion. At each withdrawal epoch t_h, the policy holder may withdraw an arbitrary amount $\omega_h \leq W(t_h^-)$, so that the sub-account value decreases as $W(t_h) = \max(W(t_h^-) - \omega_h, 0)$, and the guarantee account balance, that remains constant when $t_{h-1} \leq t < t_h$ at level $A_t = A_{t_{h-1}}$ since no withdrawal is allowed between two consecutive anniversaries, assumes value $A_{t_h} = \max(A_{t_{h-1}} - \omega_h, 0)$. For instance, at the first anniversary $t_1 = 1$, the sub-account has value $W(t_1^-)$ generated by (1), while the guarantee account has value $A_{t_1^-} = W_0$. The policy-holder may decide to withdraw a generic amount ω_1, with $G \leq \omega_1 \leq W(t_1^-)$, where $G = W_0/T$, which will be penalized only for the part exceeding G with a contractual fixed penalty φ. Hence, she/he receives

$$f(\omega_1) = G + (1 - \varphi)(\omega_1 - G), \tag{2}$$

while both the sub-account and the guarantee account decrease by ω_1 assuming value $W(t_1) = \max(W(t_1^-) - \omega_1, 0)$, and $A_{t_1} = \max(A_{t_1^-} - \omega_1, 0)$, respectively.

2 The Evaluation Model

The novelty of our approach consists on the fact that the choice of the withdrawn amount ω_h may be driven not only by an endogenous financial valuation, but also by an exogenous variable that influences the policy-holder's decisions. The role of such variable is to model the possible exogenous factors that may lead the policy-holder to make sub-optimal choices of the withdrawn amounts in place of the optimal ones. To capture this effect on the holder decisions, we denote by ω_h^F the financially driven withdrawn amount (i.e., endogenous within the evaluation problem), and by ω_h^I the amount needed by the policy-holder whenever a certain exogenous event happens. We suppose that ω_h^I is a realization of a random variable with a given distribution function (then specified in Sect. 3 to make numerical experiments), so that the policy-holder withdraws ω_h^I with probability p^I, while she/he withdraws ω_h^F with probability $1 - p^I$. Here, $p^I = 1 - e^{-\iota(\cdot)}$ represents the probability characterizing the external factor and $\iota(\cdot)$ is the hazard rate function.

The model is based on a trinomial approximation of the sub-account dynamics (1) to evaluate VA with GMWB and, consequently, the fair fee α paid for the GMWB guarantee, even in presence of some exogenous factors affecting the policy-holder choices of the withdrawn amounts at each contract anniversary. At first, a trinomial approximation of the stochastic differential equation (1) is established when a fixed amount is withdrawn at each contract anniversary, as in the

static approach. The previous approximation scheme is then adapted to handle the possibility for the policy-holder to withdraw an arbitrary amount, as in the dynamic approach. In this case, we assign to each sub-account value, in correspondence of the policy anniversaries, the additional role of keeping information concerning the total withdrawn amount up to that epoch. Finally, an exogenous variable is introduced to show its influence on the choice of the withdrawal amounts when we compute the policy value at inception proceeding backward on the tree and, in particular, the fee α in (1) charged for the GMWB guarantee.

3 Numerical Results

We test the pricing model by showing some numerical experiments where we compute the fee α for the GMWB guarantee. At first, to assess the goodness of the proposed model, we compare the values of α computed by the proposed model with the ones reported in the paper of Yang and Dai [1].

$W_0 = 100, r = 0.0325, n = 1000$						
	$T = 25$			$T = 20$		
Static, $\varphi = 1$	$\sigma = 0.2$	$\sigma = 0.3$	$\sigma = 0.4$	$\sigma = 0.2$	$\sigma = 0.3$	$\sigma = 0.4$
Trin.	0.004599	0.010182	0.015675	0.006607	0.014147	0.021599
YD	0.0046	0.0102	0.0157	0.0066	0.0142	0.0216
	$T = 25$			$T = 20$		
Static with surrender, $\varphi = 0.1$	$\sigma = 0.2$	$\sigma = 0.3$	$\sigma = 0.4$	$\sigma = 0.2$	$\sigma = 0.3$	$\sigma = 0.4$
Trin.	0.004599	0.015796	0.039517	0.006607	0.022359	0.052297
YD	0.0046	0.0158	0.0395	0.0066	0.0224	0.0523

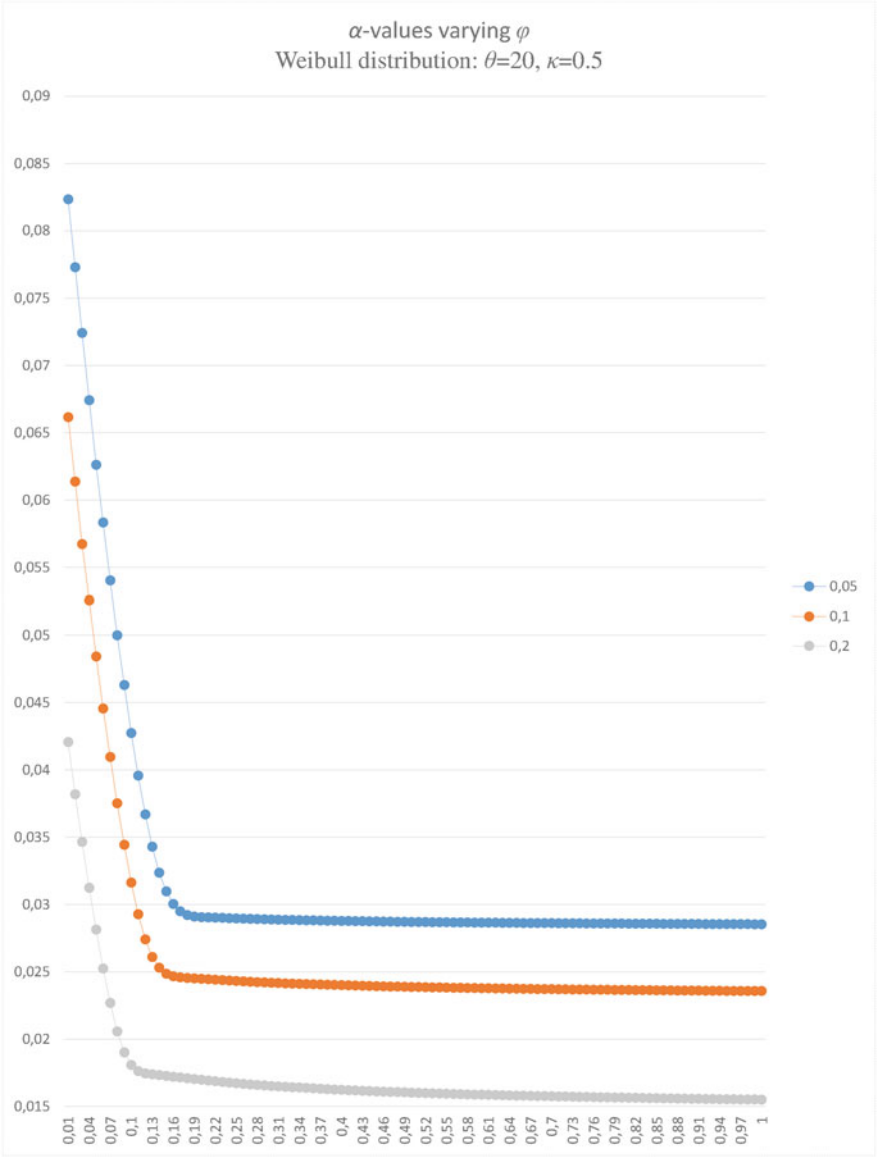
To provide an analysis of the performance of the proposed model embedding the effect of the exogenous variable, we consider a policy having time to maturity 10 years with parameters $W_0 = 100$, $r = 0.05$, and $\sigma = 0.4$. In this case, the static model obtained by imposing $\varphi = 1$ and $\iota(\cdot) = 0$ provides a fee for the GMWB guarantee $\alpha = 0.0344$. Imposing the penalty $\varphi = 0.01$, the static model with surrender option provides a value for the fee equal to $\alpha = 0.1043$. To show the effect of the exogenous variable on the fee for the GMWB characterizing the considered policy, we consider a Weibull distribution having cumulative distribution function

$$F(x) = 1 - e^{-\left(\frac{x}{\theta}\right)^{\kappa}} \tag{3}$$

where $x > 0$, $\theta > 0$, and $\kappa > 0$. To generate numerical results, we choose the hazard rate as a decreasing function of the sub-account value, that is

$$\iota(W(i, j)) = \rho \frac{W_0}{W(i, j)}, \tag{4}$$

where ρ is a positive constant and W_0 is the policy minimum guaranteed global amount. In the figure, we provide a comparison of the α values when the penalty $\varphi \in [0.01, 1]$. To do this, we consider for ρ in (4) three different levels 0.05, 0.1, and 0.2, while the Weibull parameters are set to $\theta = 20$ and $\kappa = 0.5$. The effect of the exogenous variable on the values for α is more evident when increasing the ρ-level, that is sub-optimal amounts are more probable to be withdrawn. Indeed, the greater



is the probability that sub-optimal amounts are withdrawn, the smaller is the present policy value and, consequently, smaller is the fee α for the GMWB guarantee.

Reference

1. Yang, S.S., Dai, T.S.: A flexible tree for evaluating guaranteed minimum withdrawal benefits under deferred life annuity contracts with various provisions. Insur. Math. Econ. **52**(2), 231–242 (2013)

A Continuous Time Model for Bitcoin Price Dynamics

Alessandra Cretarola, Gianna Figà-Talamanca, and Marco Patacca

Abstract This chapter illustrates a continuous time model for the dynamics of Bitcoin price, which depends on an attention or sentiment factor. The model is proven arbitrage-free under mild conditions and a quasi-closed pricing formula for European style derivatives is provided.

Keywords Bitcoin · Sentiment · Option pricing

1 Introduction

Bitcoin has been introduced by Satoshi Nakamoto (a pseudonym) in 2009 as a peer-to-peer payment system based on the blockchain technology, a decentralized ledger which does not rely on any central authority. Indeed, Bitcoin has shown a high volatile behavior and it is claimed to be a speculative asset rather than a currency, see [5]. Besides, Bitcoin price is known to be affected by market sentiment and it is believed that over-confidence about the underlying technology might boost in a pricing bubble, see, among others [1–3]. It is also worth noticing that a standardized market has started within the Chicago Option Exchange (CBOE) for Bitcoin futures, opening the path to other derivatives. In this paper we build on the above literature and on the new interest in Bitcoin derivatives to develop a bivariate model in continuous time to describe the Bitcoin price dynamics depending on the behavior of a second stochastic factor, which represents confidence/sentiment in the Bitcoin system. We also take into account a delay between the confidence factor and its effect on the Bitcoin price. The model is proven to be arbitrage-free under mild conditions and, based on risk-neutral evaluation, a quasi-closed formula is derived

A. Cretarola
Department of Mathematics and Computer Science, University of Perugia, Perugia, Italy
e-mail: alessandra.cretarola@unipg.it

G. Figà-Talamanca (✉) · M. Patacca
Department of Economics, University of Perugia, Perugia, Italy
e-mail: gianna.figatalamanca@unipg.it; marco.patacca@studenti.unipg.it

M. Corazza et al. (eds.), *Mathematical and Statistical Methods for Actuarial Sciences and Finance*, https://doi.org/10.1007/978-3-319-89824-7_49

for European style derivatives on the Bitcoin. A short numerical exercise is finally provided to assess the pricing formula sensitivity with respect to model parameters.

2 The Model

Consider the probability space $(\Omega, \mathscr{F}, \mathbf{P})$ with a filtration $\mathbb{F} = \{\mathscr{F}_t, 0 \leq t \leq T\}$ satisfying classical assumptions of right-continuity and completeness, where $T > 0$ is a fixed time horizon; in addition, we assume that $\mathscr{F} = \mathscr{F}_T$. We denote by $S = \{S_t,\ 0 \leq t \leq T\}$ the Bitcoin price process and we assume that the Bitcoin price dynamics is described by the following equation:

$$\mathrm{d}S_t = \mu_S P_{t-\tau} S_t \mathrm{d}t + \sigma_S \sqrt{P_{t-\tau}} S_t \mathrm{d}W_t, \quad S_0 = s_0 \in \mathbb{R}_+, \tag{1}$$

where $\mu_S \in \mathbb{R} \setminus \{0\}, \sigma_S \in \mathbb{R}_+,\ \tau \in \mathbb{R}_+$ represent model parameters; $W = \{W_t,\ 0 \leq t \leq T\}$ is a standard $\mathbb{F}$-Brownian motion on $(\Omega, \mathscr{F}, \mathbf{P})$ and $P = \{P_t,\ 0 \leq t \leq T\}$ is a stochastic factor, representing confidence/sentiment about the Bitcoin system, whose dynamics is given by:

$$\mathrm{d}P_t = \mu_P P_t \mathrm{d}t + \sigma_P P_t \mathrm{d}Z_t, \quad P_t = \phi(t),\ t \in [-L, 0]. \tag{2}$$

Here, $\mu_P \in \mathbb{R} \setminus \{0\}, \sigma_P \in \mathbb{R}_+$, $Z = \{Z_t,\ 0 \leq t \leq T\}$ is a standard $\mathbb{F}$-Brownian motion on $(\Omega, \mathscr{F}, \mathbf{P})$, independent of W; $\phi : [-L, 0] \to [0, +\infty)$ is a continuous and deterministic initial function. In (2) we assume that the confidence factor P affects explicitly the Bitcoin price S_t up to a certain preceding time $t - \tau$ but requiring P to be observed in $[-L, 0]$ makes the bivariate model jointly feasible. Let us define the *integrated confidence process* $X^\tau = \{X^\tau_t,\ 0 \leq t \leq T\}$ as:

$$X^\tau_t := \int_0^t P_{u-\tau} \mathrm{d}u = \int_{-\tau}^0 \phi(u) \mathrm{d}u + \int_0^{t-\tau} P_u \mathrm{d}u = X^\tau_\tau + \int_0^{t-\tau} P_u \mathrm{d}u, \quad 0 \leq t \leq T.$$

Note that, X^τ_t turns out to be deterministic for $t \in [0, \tau]$.

3 Risk Neutral Measure and Option Pricing

Let us assume the existence of a riskless asset, with value process $B = \{B_t,\ t \in [0, T]\}$ given by

$$B_t = \exp\left(\int_0^t r(s) \mathrm{d}s\right), \quad t \in [0, T],$$

where $r : [0, T] \to \mathbb{R}$ is a bounded, deterministic function representing the instantaneous risk-free interest rate.

Theorem 1 *In the market model outlined in Sect. 2, the followings hold:*

(i) the bivariate stochastic delayed differential equation

$$\begin{cases} \mathrm{d}S_t = \mu_S P_{t-\tau} S_t \mathrm{d}t + \sigma_S \sqrt{P_{t-\tau}} S_t \mathrm{d}W_t, \quad S_0 = s_0 \in \mathbb{R}_+, \\ \mathrm{d}P_t = \mu_P P_t \mathrm{d}t + \sigma_P P_t \mathrm{d}Z_t, \quad P_t = \phi(t),\ t \in [-L, 0], \end{cases}$$

has a continuous, $\mathbb{F}$-adapted, unique solution $(S, P) = \{(S_t, P_t), 0 \le t \le T\}$ given by

$$S_t = s_0 \exp\left(\left(\mu_S - \frac{\sigma_S^2}{2}\right)\int_0^t P_{u-\tau}\mathrm{d}u + \sigma_S \int_0^t \sqrt{P_{u-\tau}}\mathrm{d}W_u\right), \quad t \in [0, T],$$

$$P_t = \phi(0) \exp\left(\left(\mu_P - \frac{\sigma_P^2}{2}\right)t + \sigma_P Z_t\right), \quad t \in [0, T].$$

(ii) Let $\phi(t) > 0$, for every $t \in [-L, 0]$. Then, there exists an equivalent martingale measure $\mathbf{Q}$ for S defined on $(\Omega, \mathscr{F}_T)$, with the following density

$$\left.\frac{\mathrm{d}\mathbf{Q}}{\mathrm{d}\mathbf{P}}\right|_{\mathscr{F}_T} =: L_T^{\mathbf{Q}}, \quad \mathbf{P} - a.s.,$$

where $L_T^{\mathbf{Q}}$ is the terminal value of the $(\mathbb{F}, \mathbf{P})$-martingale $L^{\mathbf{Q}} = \{L_t^{\mathbf{Q}},\ 0 \le t \le T\}$ given by

$$L_t^{\mathbf{Q}} := \mathscr{E}\left(-\int_0^{\cdot} \frac{\mu_S P_{s-\tau} - r(s)}{\sigma_S \sqrt{P_{s-\tau}}}\mathrm{d}W_s - \int_0^{\cdot} \gamma_s \mathrm{d}Z_s\right)_t, \quad t \in [0, T],$$

for a suitable $\mathbb{F}$-progressively measurable process $\gamma = \{\gamma_t,\ 0 \le t \le T\}$, representing the market price of risk for the confidence factor. Here, $\mathscr{E}$ refers to the Doléans-Dade exponential.

(iii) If we further assume $\gamma = 0$ (minimal martingale measure) then the risk-neutral price C_t at time t of a European Call option written on the Bitcoin, expiring in T and with strike price K, is given by the formula

$$C_t = \int_0^{+\infty} C^{BS}(t, S_t, x) f_{X_{t,T}^{\tau}}(x)\mathrm{d}x, \tag{3}$$

where $f_{X^{\tau}_{t,T}}(x)$ *denotes the density function of* $X^{\tau}_{t,T}$*, for each* $t \in [0, T)$ *and*

$$C^{BS}(t, s, x) := s\mathcal{N}(d_1(t, s, x)) - K \exp\left(-\int_0^t r(u)\mathrm{d}u\right)\mathcal{N}(d_2(t, s, x)),$$

with

$$d_1(t, s, x) = \frac{\log\left(\frac{s}{K}\right) + \int_0^t r(u)\mathrm{d}u + \frac{\sigma_S^2}{2}x}{\sigma_S\sqrt{x}}$$

and $d_2(t, s, x) = d_1(t, s, x) - \sigma_S\sqrt{x}$*;* $\mathcal{N}$ *denotes the standard Gaussian cumulative distribution function.*

Due to space constraints the proof is available upon request.

4 A Numerical Example

By applying the approximation in [4] in order to derive $f_{X^{\tau}_{t,T}}(x)$, we are able to compute Call prices in (3) in closed form. In Table 1, Call option prices are reported in the case where $S_0 = 450$, $r = 0.01$, $\mu_P = 0.03$, $\sigma_P = 0.35$, $\sigma_S = 0.04$, $T = 3$ months, τ= 1 week (5 working days) for different values of the confidence P_0 (rows) and strike prices (columns). As expected, Call option prices are increasing with respect to confidence in the market and decreasing with respect to strike prices.

In Table 2, Call option prices are summed up, for initial confidence value $P_0 = 100$, by letting the expiration date T and the information lag τ vary. Again as expected, for Plain Vanilla Calls the price increases with time to maturity.

Increasing the delay reduces option prices; of course, the spread is inversely related to the time to maturity of the option.

Table 1 Call option prices against different strikes K and for different values of the confidence P_0

K	400	425	450	475	500
$P_0 = 10$	51.24	28.35	11.46	3.09	0.54
$P_0 = 100$	64.12	48.05	34.94	24.69	16.97
$P_0 = 1000$	128.68	117.75	107.77	98.66	90.35

Table 2 Call option prices against different Strikes K and for different values of T and τ

K	400	425	450	475	500
T=1 month, τ=1 week	52.85	33.09	18.27	8.81	3.71
T=1 month, τ=2 weeks	51.58	30.62	15.18	6.13	2.00
T=3 months, τ=1 week	64.12	48.05	34.94	24.69	16.97
T=3 months, τ=2 weeks	62.95	46.65	33.42	23.18	15.60

References

1. Bukovina, J., Marticek, M.: Sentiment and bitcoin volatility. Technical report, Mendel University in Brno, Faculty of Business and Economics (2016)
2. Kristoufek, L.: Bitcoin meets Google Trends and Wikipedia: quantifying the relationship between phenomena of the Internet era. Sci. Rep. **3** (2013)
3. Kristoufek, L.: What are the main drivers of the bitcoin price? Evidence from wavelet coherence analysis. PLoS One **10**(4) (2015)
4. Lévy, E.: Pricing European average rate currency options. J. Int. Money Financ. **11**(5), 474–491 (1992)
5. Yermack, D.: Is Bitcoin a real currency? An Economic appraisal. In: Handbook of Digital Currency, pp. 31–43. Elsevier, New York (2015)

Forecasting the Volatility of Electricity Prices by Robust Estimation: An Application to the Italian Market

Lisa Crosato, Luigi Grossi, and Fany Nan

Abstract Volatility of electricity prices has been often estimated through GARCH-type models which can be strongly affected by the presence of extreme observations. Although the presence of spikes is a well-known stylized effect observed on electricity markets, robust volatility estimators have not been applied so far. In this paper we try to fill this gap by suggesting a robust procedure to the study of the dynamics of electricity prices. The conditional mean of de-trended and seasonally adjusted prices is modeled through a robust estimator of SETAR processes based on a polynomial weighting function while a robust GARCH is used for the conditional variance. The robust GARCH estimator relies on the extension of the forward search by Crosato and Grossi. The robust SETAR-GARCH model is applied to the Italian day-ahead electricity market using data in the period spanning from 2013 to 2015.

Keywords Italian electricity market · Robust forecasting · Forward search · SETAR models

The views expressed are purely those of the authors and may not in any circumstances be regarded as stating an official position of the European Commission.

L. Crosato (✉)
University of Milano-Bicocca, Milano, Italy
e-mail: lisa.crosato@unimib.it

L. Grossi
University of Verona, Verona, Italy
e-mail: luigi.grossi@univr.it

F. Nan
European Commission, Joint Research Center (JRC), Ispra (VA), Italy
e-mail: fany.nan@ec.europa.eu

M. Corazza et al. (eds.), *Mathematical and Statistical Methods for Actuarial Sciences and Finance*, https://doi.org/10.1007/978-3-319-89824-7_50

1 Introduction

In this paper we introduce a doubly robust approach to modelling the volatility of electricity spot prices, minimizing the misleading effects of the extreme jumps that characterize this particular kind of data on the predictions. The purpose of this paper is twice: first, it is possible to enhance the prediction from point to intervals with associated probability levels, second, we can detect possible extreme prices which are commonly observed in electricity markets. Although many papers have applied quite sophisticated time series models to prices and demand time series of electricity and gas, very few have considered the strong influence of jumps on estimates and the need to move to robust estimators [4]. Thus, the big differences between our paper and the previous literature related to jumps in electricity prices are twofold: (1) we don't focus on the prediction of jumps; (2) we apply robust estimators which are not strongly influenced by the jumps. In this way we improve the prediction of normal prices which represent the majority of data. Our approach could then be integrated with other methods dealing with jumps forecasting.

Our methodology is applied separately to the hourly time series of Italian day-ahead electricity price data from January 1st, 2013 to December 31th, 2015. In the first step, we apply a threshold autoregressive model (SETAR) to the time series of logarithmic prices. In the second step the weighted forward search estimator (WFSE) for GARCH(1,1) models is applied to the residuals extracted from the first step in order to estimate and forecast volatility. The weighted forward search is a modification of the Forward Search intending to correct the effects of extreme observations on the estimates of GARCH(1,1) models. Differently from the original Forward Search, at each step of the search estimation involves all observations which are weighted according to their degree of outlyingness.

2 A Robust SETAR Model for Electricity Prices

It is well-known that the series of electricity prices have long-run behaviour and annual dynamics, which change according with the load period. A common characteristic of price time series is the weekly periodic component (of period 7), suggested by the spectra that show three peaks at the frequencies $1/7$, $2/7$ and $3/7$, and a very persistent autocorrelation function.

We assume that the dynamics of log prices can be represented by a nonstationary level component L_{tj}, accounting for level changes and/or long-term behaviour, and a residual stationary component p_{tj}, formally:

$$\log P_{tj} = L_{tj} + p_{tj}. \tag{1}$$

To estimate L_{tj} we use the wavelets approach [3]. We consider the Daubechies least asymmetric wavelet family, LA(8), and the coefficients were estimated *via*

the maximal overlap discrete wavelet transform (MODWT) method (for details, see Percival and Walden [5]).

The detrended prices p_t will then be modeled by a two-regime Self-Exciting Threshold AutoRegressive model SETAR(7,1) which is specified as

$$p_t = \begin{cases} \mathbf{x}_t \boldsymbol{\beta}_1 + \varepsilon_t, & \text{if} \quad p_{t-1} \leq \gamma \quad \text{and} \\ \mathbf{x}_t \boldsymbol{\beta}_2 + \varepsilon_t & \text{if} \quad p_{t-1} > \gamma \end{cases} \tag{2}$$

for $t = 1, \ldots, N$, where p_{t-1} is the threshold variable and γ is the threshold value. The relation between p_{t-1} and γ states if p_t is observed in regime 1 or 2. $\boldsymbol{\beta}_j$ is the parameter vector for regime $j = 1, 2$ containing 7 coefficients to account for weekly periodicity. $\mathbf{x}_t$ is the t-th row of the $(N \times 7)$ matrix $\mathbf{X}$ comprising 7 lagged variables of p_t (and eventually a constant). Errors ε_t are assumed to follow an iid$(0, \sigma_\varepsilon)$ distribution.

Parameters can be estimated by sequential conditional least squares. In the case of robust two-regime SETAR model, for a fixed threshold γ the GM estimate of the autoregressive parameters can be obtained by applying the iterative weighted least squares:

$$\hat{\boldsymbol{\beta}}_j^{(n+1)} = \left(\mathbf{X}_j' \mathbf{W}^{(n)} \mathbf{X}_j \right)^{-1} \mathbf{X}_j' \mathbf{W}^{(n)} \mathbf{p}_j \tag{3}$$

where $\hat{\boldsymbol{\beta}}_j^{(n+1)}$ is the GM estimate for the parameter vector in regime $j = 1, 2$ after the n-th iteration from an initial estimate $\hat{\boldsymbol{\beta}}_j^{(0)}$, and $\mathbf{W}^{(n)}$ is a weight diagonal matrix, those elements depend on a weighting function $w(\hat{\boldsymbol{\beta}}_j^{(n)}, \hat{\sigma}_{\varepsilon,j}^{(n)})$ bounded between 0 and 1. The GM weights are presented in Schweppe's form $w(\hat{\boldsymbol{\beta}}_j, \hat{\sigma}_{\varepsilon,j}) = \psi(r_t)/r_t$ with standardized residuals $r_t = (p_t - \mathbf{x}_t \hat{\boldsymbol{\beta}}_j)/(\hat{\sigma}_{\varepsilon,j} w(\mathbf{x}_t))$ and $w(\mathbf{x}_t) = \psi(\mathrm{d}(\mathbf{x}_t)^\alpha)/\mathrm{d}(\mathbf{x}_t)^\alpha$. $\mathrm{d}(\mathbf{x}_t) = |\mathbf{x}_t - m_{p,j}|/\hat{\sigma}_{p,j}$ is the Mahalanobis distance and α is a constant usually set equal to 2 to obtain robustness of standard errors. The weight function is the Polynomial ψ function as in Grossi and Nan [2]. The threshold γ is estimated by minimizing the objective function $\sum_{t=1}^{N} w(\hat{\boldsymbol{\beta}}, \hat{\sigma}_\varepsilon)(p_t - \mathbf{x}_t \hat{\boldsymbol{\beta}})^2$ over the set Γ of allowable threshold values.

3 Robust Volatility Estimation Through the Forward Search

Now we apply the weighted forward search (WFS) estimator to derive robust prediction intervals for the volatility of electricity prices, starting from the residuals of the SETAR model estimated in Sect. 1. Let ϵ_t denote an observed time series of heteroscedastic residuals. For electricity prices $\epsilon_t = p_t - \hat{p}_t$ where $\hat{p}_t$ is the price fitted by the robust SETAR model. We proceed now by estimating a standard

GARCH(1,1) model on residuals ϵ_t, so that $\epsilon_t | F_{t-1} \sim N\left(0, \sigma_t^2\right)$ and

$$\sigma_t^2 = \alpha_0 + \alpha_1 \epsilon_{t-1}^2 + \beta \sigma_{t-1}^2 \tag{4}$$

with $\alpha_0 > 0, \alpha_1 \geq 0, \beta \geq 0, \alpha_1 + \beta < 1$.

The WFS (see Crosato and Grossi [1], for details) starts with the selection of a small subset of observations showing the minimum median of squared residuals. Then, at each step of the search, MLE estimation is carried out on all observations weighted to account for their degree of outlyingness so that the estimation pattern along the search is not influenced by the presence of spikes. All units, a part from the first b, are ranked according to their squared standardized residuals with respect to the model estimated at the previous step of the search. Proceeding along the search, an increasing number of units are given weight 1 while the remaining are downweighted by the corresponding value of the complementary cumulative distribution function of the squared standardized residuals defined on the whole sample. The weights range from approximately 0 to 1 so that outliers will be heavily downweighted until the end of the search while the closer the weight to 1 the stronger is the degree of agreement of the observation with the estimated GARCH. This way the temporal structure of the time series is respected, filling the time gaps created by the forward search ordering but on the same time observations are ordered consistently with the GARCH model estimated until the last steps when all observations enter with their original value.

A plus of the WFS approach is that outliers are identified automatically through a test reducing the arbitrariness in declaring a given observation as outlier. Once outliers have been identified the forward plots of coefficient estimates are cut automatically and WFS GARCH estimates are obtained.

The outliers identified through the WFS test vary from a minimum of 29 (hour 21) to a maximum of 59 (hours 8 and 9). A few outliers characterize many hours of the same day, as is the case of the 21st and 27th of April, 2013 for 14 h as well as mot days of June 2013. Spikes in 2014 characterize mainly the months of February, March and April. Outliers for 2015, hours 14 and 16, are highlighted by red circles in Fig. 1.

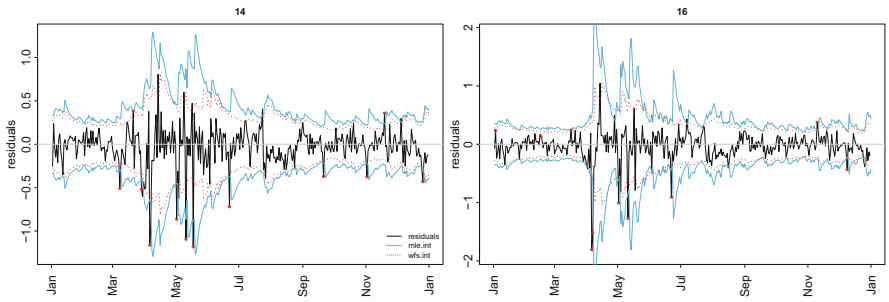


Fig. 1 Forecasted conditional confidence intervals of volatility for hourly electricity prices in Italy (from 13 to 18, January to December 2015) obtained by MLE (light blue) and WFS (dotted red)

Using the MLE and WFS estimated coefficients and applying Eq. (4) for the conditional variance recursively, we then obtain the forecasted price volatility and the corresponding 95% confidence intervals (see Fig. 1). As can be seen, the WFS intervals (dotted red) are tighter around the realized volatility than the MLE ones (light blue). The gain in prediction is evident in the aftermath of a spike or a drop, for instance in April and May. Note that the correction provided by the WFS works also for smallest jumps, as in July and December.

References

1. Crosato, L., Grossi, L.: Correcting outliers in GARCH models: a weighted forward approach. Stat. Pap. (2017). https://doi.org/10.1007/s00362-017-0903-y
2. Grossi, L., Nan, F.: Robust estimation of regime switching models. In: Morlini, I., Minerva, T., Palumbo, F. (eds.) Advances in Statistical Models for Data Analysis, pp. 125–135. Springer International Publishing, Cham (2015)
3. Lisi, F., Nan, F.: Component estimation for electricity prices: procedures and comparisons. Energy Econ. **44**, 143–159 (2014)
4. Nowotarski, J., Tomczyk, J., Weron, R.: Robust estimation and forecasting of the long-term seasonal component of electricity spot prices. Energy Econ. **39**, 13–27 (2013)
5. Percival, D., Walden, A.: Wavelet Methods for Time Series Analysis. Cambridge University Press, Cambridge (2000)

"Money Purchase" Pensions: Contract Proposals and Risk Analysis

Valeria D'Amato, Emilia Di Lorenzo, Marilena Sibillo, and Roberto Tizzano

Abstract The Authors propose a personal pension product, consisting of a non-traditional profit sharing life insurance contract where the insured is allowed to share the profit of the pension's invested funds all along the contract duration, that is from the issue time till the insured's death. In its concrete realization, the idea comes true as a sequence of premiums characterized by a level cap, followed by the sequence of benefits characterized by a level floor. The two embedded options are inserted in the basic structure of a pension annuity. Due to the negligibility of the pooling effect in such kind of portfolios, the impact of the accidental demographic risk source is investigated.

Keywords Personal pension product · Participating benefits · Demographic risk

1 Personal Pension Products

The growth and the consolidation of the third pillar of retirement, posit a significant challenge to the combination of long-term savings and long-term investments. The varied world of the Personal Pensions basically consists of open funds or Private Personal Pensions (PPP), the latter being individual plans built up by insurance contracts, people may join as a consequence of discretionary choices.

Those who enter this kind of contracts contribute to the accumulation of an individually accounted private pension fund to complement the pension income provided by a basic public and/or occupational pension scheme, in order to meet the need of future adequacy of income at old age.

V. D'Amato · M. Sibillo
Department of Economics and Statistics, University of Salerno, Fisciano (SA), Italy
e-mail: vdamato@unisa.it; msibillo@unisa.it

E. Di Lorenzo (✉) · R. Tizzano
Department of Economic and Statistical Sciences, University of Naples Federico II, Naples, Italy
e-mail: diloremi@unina.it; tizzano@unina.it

M. Corazza et al. (eds.), *Mathematical and Statistical Methods for Actuarial Sciences and Finance*, https://doi.org/10.1007/978-3-319-89824-7_51

Currently, a striking variety of life insurance contracts, investment funds, pension insurances confuse the prospective users, often discouraged by high costs and contract complexity. On the other hand, providers have to cope with the impact of long-term financial risks, as well as continuous and systematic demographic changes. It is no accident that in 2016 the Organisation for Economic Cooperation and Development (OECD), the International Organisation of Pension Supervisors (IOPS) and the World Bank [6] decided to collect and disclose a global database of private pensions. It is precisely the PPP socio-economic centrality (cf. [5]) that brought them to the forefront of scientific debate, as reflected by the variety of contributions in recent actuarial literature (see for instance [2]). Personal pension schemes with the investment risk and the longevity risk sharing were proposed by Di Lorenzo et al. [4] and D'Amato et al. [3]. In this paper we develop the contract we proposed in [3], placing it in the portfolio logic: due to the number of homogeneous policies most likely not sufficiently high, the insurer could not appeal to the pooling effect for managing the accidental deviations of the number of deaths from the forecasted ones. Due to this reason, we investigate the impact of the accidental demographic risk source on a portfolio of such policies.

2 Variable Annuities with Participating Benefits: Notations and Recalls

In the following we will develop the proposal of a life annuities with participating benefits, under the hypotheses that the installments are increased by a percentage of the period financial result of a pension fund, when it reaches a predefined value at least, that can be withdrawn by the insureds even before reaching the retirement age. Aim of the study is the quantification of the impact of the accidental demographic risk source on a portfolio of such contracts, that we will develop in the next section.

As we explained in [3], the implementation of the model proposed therein requires the insurer to disclose an appropriate set of audited financial statements for each and every managed pension fund, in order to determine the insureds' participation in the income of the fund, that may be linked to the Cash Net Profit (CNP) of the fund, that is the cash part of its accrual income, resulting from the (segregated) income statement of the fund for each financial year. Since the CNP is equivalent to an increase in the net financial assets of the fund, its distribution to the insureds will not impair the investment base upon which the insureds future income (annuities) are determined. By definition, CNP will never be larger than the accrual income from which it is derived even though the net cash flow of operation is larger. It goes without saying that if the accrual income is negative, that is if the pension fund income statement discloses a loss, no distribution is possible at all.

Each contract consists of a variable annuity with deferment period T, with annual installments due at the end of each year s ($s \geq T$), each one given by the sum of a predefined installment b_t and the participation in the fund CNP provided this being positive. The annual premiums P, due at the beginning of each year until the time τ, $\tau < T$, are allocated into the pension fund, which yields a variable annual interest rate. Since the contract provides for the insureds' participation in the fund annual financial profit also during the deferment period, the participation amount, if due, is subtracted from the constant premium. In summary, the contract incorporates an embedded option involving the insureds' participation in the fund annual financial profit CNP, where such profit is positive. So, denoting by r_t the CNP rate at time t, ρ the participation rate and i the first order financial technical base, by virtue of the above exposed contract guidelines, the benefit stream $\tilde{b_t}$ is the following, as explained in [3]:

$$\tilde{b_t} = \begin{cases} [\rho(r_t - i)^+ \sum_{j=0}^{t-1} P_j (1+i)^{h-j-1}]\mathbf{1}_{(1 \leq t < T | K(x) \geq t)} \\ [b_t + \rho(r_t - i)^+ \sum_{j=0}^{t-1} P_j (1+i)^{h-j-1}]\mathbf{1}_{(T \leq t | K(x) \geq t)} \end{cases} \tag{1}$$

in which $b_0 = 0$ and the indicator functions take the values 1 or 0 if the event at subscript happens or not, respectively; $K(x)$ is the random curtate future lifetime of an annuitant aged x at the issue time; $(r_t - i)^+$ is the maximum between $(r_t - i)$ and 0.

On the other hand, the premiums' level cap P due at time t ($t = 0, 1, \ldots, \tau - 1$) is given by:

$$\sum_{t \geq \tau} \mathbb{E}[(\tilde{b}_{tt} \, {}_t p_x v(0,t)] = \sum_{t=0}^{\tau-1} \mathbb{E}[(P - \tilde{b}_t)_t \, p_x v(0,t)]. \tag{2}$$

where $v(0, t)$ is the stochastic value in $t = 0$ of a monetary unit in t.

3 The Demographic Risk Filters

The demographic variability may cause significant consequences on the contract, because of the long run time perspective There are two main aspects to be considered: the unexpected changes as well the systematic movements (that is the systematic improvements of life expectancy), which is particularly significant in light of the Solvency II directives, in order to guarantee the sustainability performance, mostly when the longevity risk cannot be hedged or marked to market (see [1]).

At any rate, the proposed contract's innovative character entails, in a more realistic sense, providers' consideration of small-sized portfolios (in terms of policy ranges). The definition of mortality unexpected changes' impact, in this context, is particularly meaningful. In the following we consider a homogeneous portfolio of

Table 1 ADR (100 contracts, age 45 at issue)

Time	2017	2022	2027	2032	2037	2042	2047	2052	2057
ADR	0	0.01417	0.025201	0.03854	2.7540	3.6107	4.6595	5.1009	4.9914

contracts issued to $N(x)$ lives aged x at issue, denoting by Sf the survival function used to infer the survival probabilities; we define ADR $= \left[\text{Var}\left[N(x+t)\tilde{b}_t|Sf\right)\right]\right]^{0.5}$ the measure of the risk of the accidental deviations in the life durations.

In order to give a numerical exemplification, we consider a portfolio of 100 single pension annuity policies, issued on policyholders aged $x = 45$ in 2017, with the minimum benefits of 1 paid at the end of each year in the case of life, after the deferment period of 20 years. The premiums are paid during the whole deferment period. The participation rate is $\rho = 0.8$ and the guaranteed interest rate is 0.01. The profit of the invested fund is modeled as in [3] by means of a Vasicek process, whose parameters are estimated on Federal Reserve Economic Data by means of a 30-year swap rate percent (sampled daily), from 2011-10-31 to 2016-10-28. The demographic technical base is modeled by a Lee Carter process on the American Male data from 1933:2007, age 0:100 (Human Mortality Database). The first order financial technical base is $i = 0.02$.

Table 1 shows how the risk measure expressed by the ADR varies with t.

References

1. Blake, D., Turner, J. A.: Annuities: lessons from the United Kingdom. First Quarter 2014 issue of BENEFITS QUARTERLY. http://www.ifebp.org/inforequest/0165164.pdf
2. Denuit, M., Haberman, S., Renshaw, A.: Longevity-contingent deferred life annuities. J. Pension Econ. Financ. **14**(3), 315–327 (2015)
3. D'Amato, V., Di Lorenzo, E., Sibillo, M., Tizzano, R.: Profit-sharing and personal pension products: a proposal. In: Pensions: Global Issues, Perspectives and Challenges, pp. 97–111. Nova Science Publisher, New York (2017)
4. Di Lorenzo, E., Orlando, A., Sibillo, M.: Measuring risk-adjusted performance within the actuarial framework, and product attractiveness of a life annuity portfolio. J. Math. Financ. **7**, 83–101 (2017). https://doi.org/10.4236/jmf.2017.71005
5. EIOPACP 16/001: Consultation paper on EIOPAs advice on the development of an EU Single Market for personal pension products (PPP). https://eiopa.europa.eu/Pages/Consultations/EIOPA-CP-16-001-Consultation-Paper-on-EIOPA%E2%80%99s-advice-on-the-development-of-an-EU-Single-Market-for-personal-pension-product.aspx (1 February 2016)
6. OECD: OECD, IOPS and World Bank join forces to improve private pension statistics. http://www.oecd.org/daf/fin/private-pensions/oecd-iops-world-bank-join-forces-to-improve-private-pension-statistics.htm (2016)

What If Two Different Interest Rates Datasets Allow for Describing the Same Financial Product?

Valeria D'Amato, Antonio Díaz, Emilia Di Lorenzo, Eliseo Navarro, and Marilena Sibillo

Abstract The chance to choose among more than one dataset for representing and describing the movements in the financial market of the same financial entity has noteworthy effects on the practical quantifications. The case we consider in the paper concerns two datasets, different and deemed to be equivalent between them, referred to risk free interest rates. In light of the volatility term structure discrepancies between the two databases and of some closed formulas for stochastically describing the behavior of the financial valuation discrepancies by means of the Vasicek interest rate process, we show two relevant practical evidences. The application concerns the pricing of two derivative cases. The aim is to quantify how much the use of one dataset rather than the other impacts on the final result.

Keywords Yield curve estimation · Volatility term structure · Vasicek process · Derivative pricing

V. D'Amato · M. Sibillo
Department of Economics and Statistics, Campus Universitario, University of Salerno, Fisciano, Italy
e-mail: vdamato@unisa.it; msibillo@unisa.it

A. Díaz (✉)
Department of Economics and Finance, Universidad de Castilla-La Mancha, Albacete, Spain
e-mail: antonio.diaz@uclm.es

E. Di Lorenzo
Department of Economic and Statistical Sciences, via Cintia, University of Naples Federico II, Naples, Italy
e-mail: diloremi@unina.it

E. Navarro
Departamento de Economía y Dirección de Empresa, Universidad de Alcalá, Alcalá de Henares, Spain
e-mail: eliseo.navarro@uah.es

M. Corazza et al. (eds.), *Mathematical and Statistical Methods for Actuarial Sciences and Finance*, https://doi.org/10.1007/978-3-319-89824-7_52

1 Introduction

Our research starts from the observation that the relationship between specified financial instruments and datasets describing their behavior could not be bijective. In particular we will deepen the case of the government debt market of United States Treasury: two different interest rate structures can be downloaded from the websites of the US Department of the Treasury (DoT from here onwards) and of the Federal Reserve Board (FRB from here onwards). The two datasets refer to the same yield of the same financial instrument and are considered to be equivalent. The discrepancies between the two datasets are due to the model and numerical techniques used to estimate the zero coupon rates (e.g. [1, 2]) and to the fact that each provider use as input different prices or yields (cf. [3]).

What we want to investigate in this paper is if and how much the choice of one of the two datasets impact on the pricing valuation of derivatives characterized for being strongly affected by the evolution in time of the interest rates: the Interest Rate Swap and the Swaption. The analysis will be developed in a stochastic framework.

Basing on D'Amato et al. [3], in Sect. 2 we concisely report some results concerning the Volatility Term Structure analysis developed on the FRB and DoT datasets and the main characteristics of the stochastic process representing the difference between the risk free bond prices got by both datasets, in the Vasicek process environment. In Sect. 3 we price an Interest Rate Swap and a Swaption basing on each of the two datasets.

2 Discrepancies in the Datasets: Impact on the Volatility Term Structure and on the Interest Rate Projections

The FRB dataset is estimated using a weighted version of the Svensson's method from end-of-day prices of all of the outstanding off-the-run bonds. Among other securities, it excludes from the estimation all Treasury bills and the on-the-run and "first-off-the-run" issues of bonds and notes. The DoT dataset is computed from a quasi-cubic Hermite spline function that passes exactly through the yields for the chosen securities. It uses the market yields to maturity for the most recently auctioned bills (4-, 13-, 26-, and 52-weeks), six maturities of just-issued bonds and notes (2-, 3-, 5-, 7-, 10-, and 30-years), and the composite rate in the 20-year maturity range. We show statistically and economically significant differences between estimates of the volatility term structure depending on the interest rate data set used in our own yield curve estimates. These differences are observed mainly in the short-term, but also in the long-term volatility.

The projection of the interest rate term structure is the essential tool when one has to price interest rate sensitive financial products. As the two datasets DoT and FRB differ each other from the point of view of their stochastic behavior can be well highlighted observing the characteristics of the stochastic process chosen for

Table 1 3-Month time-series evaluation—Vasicek parameters %

	α_i	β_i	σ_i	r_0
Vasicek (dataset DoT)	4.0065	2.5605	43.2157	0.45
Vasicek (dataset FRB)	1.8286	2.4805	41.7105	0.45
Vasicek (difference D)	2.1779	2.6277	1.5052	0.00

describing the future evolution in time. To this aim we will choose one of the most popular interest rate stochastic model, suitable to design the trend of the risk free interest rate able to assume also negative values, the Vasicek process [4]. Its well known SDE is:

$$\mathrm{dr}_i\,(\mathrm{s}) = \alpha_i\,(\beta_i - \mathrm{r}_i)\,\mathrm{ds} + \sigma_i \mathrm{dW}_\mathrm{s} \qquad i = A, B \qquad (1)$$

in which β_i is the long term mean, towards which the process is continuously pulled with the force α_i, σ_i is the diffusion coefficient, A is DoT and B is FRB. Recalling D'Amato et al. [3], it is possible to prove that the stochastic process $r_D(s)$ arising from the difference between the two Vasicek processes $r_A(s)$ and $r_B(s)$ is Vasicek with parameters:

$$\beta_\mathrm{D} = \frac{\alpha_\mathrm{A}\beta_\mathrm{A} - \alpha_\mathrm{B}\beta_\mathrm{B}}{\alpha_\mathrm{A} - \alpha_\mathrm{B}} \qquad \alpha_D = \alpha_\mathrm{A} - \alpha_\mathrm{B} \qquad \sigma_\mathrm{D} = \sigma_\mathrm{A} - \sigma_\mathrm{B} \qquad (2)$$

We performed the model calibration by maximum likelihood estimation on 3-months interest rates of both datasets. In this way we found the set values for which the observed sample was most likely, according to Remillard [5], as implemented in SMFI5 R package. The parameter estimations is reported in Table 1 (cf. [3]). The main influence of the change in the dataset concerns the α parameter, pointing out the stronger attraction power of the dataset A towards the long term mean.

3 Pricing Derivatives: Two Applicative Cases

In this section we consider the impact of the different projections of the interest rates on a financial interest rate swap (IRS) and a swaption. Firstly we take into account an IRS with remaining time to maturity of 10-years, where the periodic payments have been forecasted on the basis of a Vasicek model. We price the derivative, i.e. we calculate the fair fixed rate on the basis of the equilibrium principle, by setting the value of the IRS equal to zero at inception. This is accomplished when the present value of the floating and fixed expected cash flow streams equal each other. In particular the fair fixed rate k can be expressed as the following formula:

$$\$1 = [\mathrm{k}\alpha_0 \mathrm{B}\,(\mathrm{t}_0, \mathrm{t}_1) + \mathrm{k}\alpha_1 \mathrm{B}\,(\mathrm{t}_0, \mathrm{t}_2) + \cdots + \mathrm{k}\alpha_{\mathrm{n}-2} \mathrm{B}\,(\mathrm{t}_0, \mathrm{t}_{n-1})] + \mathrm{B}\,(\mathrm{t}_0, \mathrm{t}_\mathrm{n}) \qquad (3)$$

where the floating leg at inception is equal to \$1, α_i, i = 0, 1, . . , n − 1 is the accrual factor for each period based on a specified accrual method and B(t_0, t_i), $i = 1, 2, \ldots, n$ is the present value of \$1 at i-time.

Solving by k the Eq. (3), the IRS fixed rate can be obtained as

$$k = \frac{1 - B(t_0, t_n)}{\alpha_0 \beta(t_0, t_1) + \alpha_1 \beta(t_0, t_2) + \cdots + \alpha_{n-2} B(t_0, t_{n-1})} \tag{4}$$

so that it is indifferent between investing in the floating versus investing in a fixed rate bond that pays k.

Results show that the issuer (fixed-rate payer) will be willing to pay k = 0.020 per dollar for a 10-year IRS in order to obtain a floating stream, in DoT environment. In the FRB context the issuer will not willing to issue the derivative due to the different projections (k = −0.056). Both of them are calculated on the basis of the Vasicek model. The glaring evidence of the numerical results points out the impact of the discrepancies of the two datasets.

Furthermore we investigate the effect of interest rate projections in the swaption pricing. The swaption is an option contract on a swap, typically an IRS. The buyer of a payer swaption can be enter into a swap contract where he becomes the fixed-rate payer and the floating-rate receiver. The buyer of a receiver swaption enters into a swap contract where he will receive the fixed rate and pay the floating rate.

We consider a European payer swaption which gives the holder the right to enter into an IRS at the expiration date. We find that this derivative price meaningfully changes depending on the provider of the data used for representing the future evolution of interest rates. We quote the contract by the well-known Black's formula [6], a variant of the Black and Scholes option pricing model.

Let us assume the following notations: T, expiration date of swaption, δ,risk-free interest rate, Φ Gaussian distribution function, σ_F, volatility of forward swap rate, i_S, strike rate, i_F, market swap rate (at par) at time T, N, notional capital.

Let $t_1 < t_2 < \ldots < t_n$ represent the coupon dates for the swap and $t_0 = T$. The swap begins on the expiration date T of the swaption and ends at time t_n.

The cash flow made to the buyer of a payer swaption at time T amounts to

$$\sum_{i=1}^{n} N e^{-\delta(t_i - T)} (i_F - i_S)(t_i - t_{i-1}) \tag{5}$$

if $i_F > i_S$ and 0 otherwise. According to the Black model, the price of this option at time 0 is expressed by the following formula:

$$N e^{-\delta t_i} (t_i - t_{i-1}) [i_F \Phi(d_1) - i_S \Phi(d_2)] \tag{6}$$

where $d_1 = \frac{ln\frac{i_F}{i_S} + \frac{1}{2}\sigma_F^2 T}{\sigma_F \sqrt{T}}$ and $d_2 = d_1 - \sigma_F \sqrt{T}$.

In this numerical application, we set the input parameters as follows: Strike Rate is 0.0620, Forward Swap Rate Continuous Compounding is 0.0600, Volatility of Forward Swap Rate is 0.20, Notional Principal is 100, Number of Payments per year is 2, and Option Maturity is 5.

We compute the price of the European Swaption based on the different datasets. It can be noticed the mispricing between the different datasets where we quoted a swaption by the above formula on a notional of 100 million. In particular, ceteris paribus, the FRB Black valuation (3.00) shows a higher swaption price than that one based on the DoT data (2.66).

As we saw earlier, the numerical results presented in this section emphasized that discrepancies of the datasets used for calibrating the interest rate structure lead to discrepant financial valuations and measures.

References

1. Bliss, R.R.: Testing term structure estimation methods. FRB Atlanta WP 96-12a (1996)
2. Bolder, D.J., Gusba, S.: Exponentials, polynomials, and fourier series: more yield curve modelling at the Bank of Canada. Bank of Canada working paper, no. 02–29 (2002)
3. D'Amato, V., Díaz, A., Di Lorenzo, E., Navarro, E., Sibillo, M.: The impact of the discrepancies in the yield curve on actuarial forecasting. Proceedings di XVI Iberian Italian Conference on Financial and Actuarial Mathematics, Paestum (Salerno), 26–27 Maggio 2016, pp. 49–55. ISBN 9788861970601 (2016)
4. Vasicek, O.: An equilibrium characterization of the term structure. J. Financ. Econ. **5**, 177–188 (1977)
5. Remillard, B.: Statistical Methods for Financial Engineering. CRC Press, Boca Raton (2013)
6. Black, F.: The pricing of commodity contracts. J. Financ. Econ. **3**, 167–179 (1976)

An Integrated Approach to Explore the Complexity of Interest Rates Network Structure

Maria Elena De Giuli, Marco Neffelli, and Marina Resta

Abstract We represent the relationships among interest rates of the same term structure using an integrated approach, which combines quantile regression and graphs. First, the correlation matrix estimated via the quantile regression (QR) is used to explore the inner links among interest rates with different maturity. This lets us possible to check for quantile cointegration among short and long-term interest rates and to assess the Expectations Hypothesis of the term structure. Second, we use these inner links to build the Minimum Spanning Tree (MST) and we investigate the topological role of maturities as centres of a network, in an application focusing on the European interest rates term structure in the period 2006–2017. To validate our choice, we compare the MST built upon the quantile regression to the one based on the sample correlation matrix. The results highlight that the QR exalts the prominent role of short-term interest rates; moreover, the connections among interest rates of the same term structure seem being better captured and described by our procedure rather than by the methodology relying on the estimation of the sample correlation matrix.

Keywords Quantile regression · Graph theory · Minimum spanning tree · European term structure

1 Introduction

The European interest rate term structure is a tool of paramount relevance, used by the European Central Bank to take important policy decisions and by European governments as a measure of market participants' expectations about current and

M. E. De Giuli
University of Pavia, Pavia, Italy

M. Neffelli (✉) · M. Resta
University of Genova, Genova, Italy
e-mail: marco.neffelli@edu.unige.it

M. Corazza et al. (eds.), *Mathematical and Statistical Methods for Actuarial Sciences and Finance*, https://doi.org/10.1007/978-3-319-89824-7_53

future state of the economy. In an econometric perspective, modelling the European interest rate term structure is a bit tricky: reconstructing its dynamics, in fact, relies on a set of non-stationary and leptokurtic time series, each of those corresponding to a different maturity. We address the problem following the stream proposed in [1], with interest rates as the nodes of a network. Nodes connections have been variously modeled in literature: [2] offers a great review for stochastics models; [3] does the same but for parametric models. However, two issues generally arise related to: (i) the trade-off between sampling and model error; (ii) unanswered questions regarding the relationships among short, medium and long-term interest rates. With respect to the first point, we propose an integrated approach, which combines quantile regression (QR) and graphs. QR can cope with stylized facts of interest rates, e.g. leptokurtic and heavy tailed distribution, hence capturing the relationships at different quantiles of the error distribution better than conventional models. Moreover, this enables us to fully answer to (ii), as following [4, 5] we jointly run the CUSUM test for quantile cointegration, thus assessing the Expectations Hypothesis of the term structure [6]. We then use the correlation matrix extracted from the quantile regression to calculate the Minimum Spanning Tree (MST), which offers a visual and immediate tool for representing complex systems [7].

The paper is structured as follows: Sect. 2 describes the mathematical background, Sect. 3 contains the empirical application on the European interest rates term structure in the period 2006–2017; Sect. 4 concludes.

2 Mathematical Background

We are going to illustrate an integrated approach combining the visual power of networks to the explicative power of the quantile regression. In detail, our procedure consists of two steps:

1. we derive the correlation matrix among variables in the input space through the quantile regression;
2. we employ the matrix obtained in Step 1 for building the MST.

In the following rows, we give some insights about the quantile regression, only. The use of graphs to represent financial systems is in fact widely documented in literature: for a deeper insight, the interested reader can refer to [8].

2.1 Quantile Regression

Following [9], we build the quantile cointegration model from the long-run equilibrium equation:

$$y_t = \alpha_t + \boldsymbol{\beta}_t{}' \boldsymbol{x}_t + u_t \tag{1}$$

where the scalar y_t and the m-dimensions vector $\mathbf{x}_t$ are I(1) variables, u_t is the zero mean stationary error term and $\alpha_t \in \mathbb{R}$, $\boldsymbol{\beta}_t \in \mathbb{R}^m$ are time-varying parameters. In our application y_t represents the shortest term interest rate, while $\mathbf{x}_t$ is the vector containing the values of remaining short/medium/long term interest rates. In contrast to the Least Squares technique, that estimates the conditional expected value of y_t, quantile regression expresses the τ-th quantile of y_t conditional on the information set $\mathcal{F}_t$ in period t,

$$\widehat{Q}_{y_t}(\tau|\mathcal{F}_t) = \widehat{\alpha}(\tau) + \widehat{\boldsymbol{\beta}}_t(\tau)'\boldsymbol{x}_t + F_u^{-1}(\tau) \tag{2}$$

where $F_u(\cdot)$ is the cumulative distribution function for the error term u_t , the residual weights are computed as: $\psi_\tau(u) = \tau - 1\{u < 0\}$, and the residual series is given by: $u_{\tau,t} = y_t - \widehat{\alpha}(\tau) - \widehat{\boldsymbol{\beta}}_t(\tau)'\boldsymbol{x}_t$.

The stability of the parameters is verified by the CUSUM test, as discussed in [4] to which we refer for further information and a more complete review. This lets us possible to investigate the stability of the equilibrium relationships across different quantiles of the distribution of the response variable.

3 A Practical Application

Our dataset is built upon 24 interest rates time series making the EU term structure, namely: EONIA, EU futures contracts with 1 week (W1), 1–12-month maturities (M1–M12), and EU Swap with 2–15, 20, 25 and 30-year maturities (Y2–Y30), respectively. We used daily data, for an overall number of 3059 observations from January 2006 to December 2017. According to their maturity, the selected interest rates can be grouped into short, medium and long-term rates, as described in Table 1 in the online appendix.[1] In order to analyse the inner relations among groups, we preliminary run the quantile cointegration CUSUM test. Here we are mainly interested in checking whether cointegration exists among interest rates at short and medium/long terms. To such aim, we selected the EONIA to represent short-term interest rates, and the 3-year and the 11-year swap contracts raw time series to characterize medium and long terms rates, respectively. These three time series are non-stationary, while their first-difference is; thus, they are drawn from I(1) processes. We then run the CUSUM test with a cross-estimation setting, checking for nine quantiles (from $\tau = 0.1$ to $\tau = 0.9$), assuming the presence of cointegration as null hypothesis H0. The results (given in Table 2 of the online appendix) highlight a larger consensus for the presence of cointegration: the test, in fact, generally does not reject the H0 hypothesis, thus emphasizing the presence of common trends

[1] Due to the limited number of pages, results are not reported here, but they are freely available in the authors' Researchgate page, together with other additional material. Please refer to the following URL: https://www.researchgate.net/profile/Marina_Resta

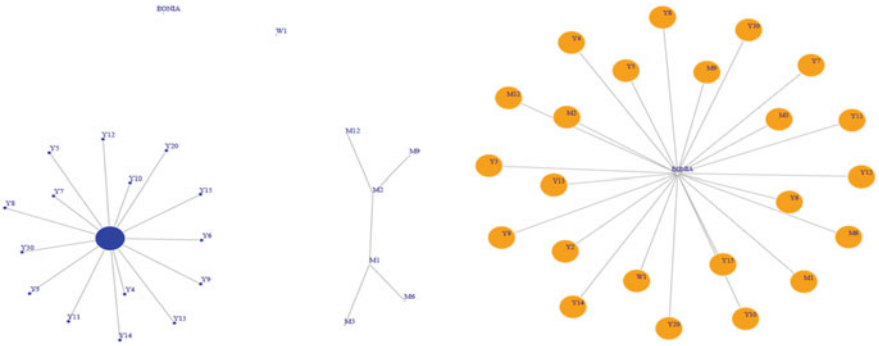


Fig. 1 MSTs built from our dataset; CV-MST on the left; QR-MST on the right

among short, medium and long interest rates. This, in turn, allows us to confirm the Expectation Hypothesis, which links short, medium and long-term interest rates with the expectations of market participants. Going back to the original dataset, we then run a comprehensive estimation of the quantile regression on interest rates changes: log-changes cannot be implemented due to the presence of negative interest rates. Again, the estimation was performed for nine quantiles. Results are listed in Table 3, in the online appendix, where we show the values of β for interest rates at various maturity. We adopted the notational convention of indicating the rates in terms of their maturity. Values in Table 3 are the starting point for building the adjacency matrix for the MST. Furthermore, we compared this MST (since now denoted by QR-MST) to its counterpart built on the sample correlation matrix (since now on: CV-MST), as given in Fig. 1. The structure of the two networks is sensitively different from the topological viewpoint; the CV-MST seems offering less information than the QR-MST, which, on the other hand, highlights strongest connections among various interest rates, as well as the central role of EONIA interest rates.

4 Conclusion

We discuss an integrated approach for evaluating the inner links among maturities, which lets also possible answering to crucial questions regarding the relationships among short, medium and long-term interest rates. In our example on the European term structure, we assess the presence of cointegration, confirming the Expectation Hypothesis, through the quantile cointegration CUSUM test. Moreover, combining this information in a graph, we highlighted the prominent role of short interest rates in the interest rates network, and we provided a network representation based on the quantile regression, which seems being more robust than the one relying on the estimation of the sample correlation matrix.

References

1. Tabak, B.M., Serra, T.R., Cajueiro, D.O.: The expectation hypothesis of interest rates and network theory: the case of Brazil. Physica A: Stat. Mech. Appl. **388**(7), 1137–1149 (2009)
2. Brigo, D., Mercurio, F.: Interest Rate Models-Theory and Practice: With Smile, Inflation and Credit. Springer Science & Business Media, Berlin (2007)
3. Diebold, F.X., Li, C.: Forecasting the term structure of government bond yields. J. Econ. **130**(2), 337–364 (2006)
4. Kuriyama, N.: Testing cointegration in quantile regressions with an application to the term structure of interest rates. Stud. Nonlinear Dyn. Econ. **20**(2), 107–121 (2016)
5. Xiao, Z., Phillips, P.C.: A CUSUM test for cointegration using regression residuals. J. Econ. **108**(1), 43–61 (2002)
6. Gerlach, S., Smets, F.: The term structure of Euro-rates: evidence in support of the expectations hypothesis. J. Int. Money Financ. **16**(2), 305–321 (1997)
7. Mantegna, R.N.: Hierarchical structure in financial markets. Eur. Phys. J. B: Condens. Matter Complex Syst. **11**(1), 193–197 (1999)
8. Resta, M.: Computational Intelligence Paradigms in Economic and Financial Decision Making. Springer, Basel (2016)
9. Xiao, Z.: Quantile cointegrating regression. J. Econ. **150**(2), 248–260 (2009)

Estimating Regulatory Capital Requirements for Reverse Mortgages: An International Comparison

Iván de la Fuente, Eliseo Navarro, and Gregorio Serna

Abstract In this paper, we estimate the value of the no-negative-equity guarantee (NNEG) embedded in reverse mortgage contracts and develop a method for calculating regulatory capital requirements according to Basel II and III. We employ a Monte Carlo simulation method that assumes an ARMA-EGARCH process for house prices in four European countries: France, Germany, Spain and the United Kingdom. The results show different estimated values for the NNEG among countries. Specifically, the value of the NNEG tends to be related to the level of the interest rates, the rental yield and house price volatility in each country, as well as the age of the borrower. Different values for value-at-risk and the expected shortfall among countries are also found, which depend on the volatility of each country's house price series.

Keywords Reverse mortgages · Option pricing · No-negative-equity guarantee · Mortality modeling · House price modeling · Regulatory capital requirements

1 Paper Summary

During recent years, there have been many papers addressing the problems of pricing and hedging the no-negative-equity guarantee (NNEG) embedded in reverse mortgages. However, despite the extensive literature on reverse mortgage pricing, there are few papers focused on regulatory capital requirements calculation for reverse mortgage providers. Therefore, this paper addresses this gap in the literature by developing a method for measuring risk and calculating the regulatory capital requirements for lump-sum reverse mortgage providers, based on value-at-risk and the expected shortfall according to Basel II and Basel III, respectively, using data from four European countries: France, Germany, Spain and the United Kingdom.

I. de la Fuente · E. Navarro · G. Serna (✉)
University of Alcala, Madrid, Spain
e-mail: ivan.fuente@uah.es; eliseo.navarro@uah.es; gregorio.serna@uah.es

M. Corazza et al. (eds.), *Mathematical and Statistical Methods for Actuarial Sciences and Finance*, https://doi.org/10.1007/978-3-319-89824-7_54

To do so, several data sets must be combined. Specifically, the proxy for the risk-free rate will be the 10-year zero-coupon government bond rate (average rate from the year 2014, obtained from Eurostat). The proxy for the rental yield rate has been obtained from the Global Property Guide 2014, which estimates the annual rental yield rate for each country as the average annual rental yield rate in the main cities of the country. In the case of mortality rates, the information used in this paper comes from the Human Mortality Database, which includes detailed information about demographic variables from 38 countries. Due to data availability at the time of data set creation, mortality data from 1990 to 2012 have been employed for all the countries considered in this study. Finally, the house price data used in this study comes from the International House Price Database, which is developed by the Federal Reserve Bank of Dallas. This database contains quarterly information since 1975. Specifically, we have employed the House Price Index from the first quarter of 1975 until the fourth quarter of 2014 (160 quarterly observations). This House Price Index is an appraisal-based property index. To solve this problem, the classical unsmoothing method proposed by Geltner [2] and Fisher et al.[1] has been applied.

Before estimating the value-at-risk and expected shortfall, we have estimated the value of the NNEG embedded in reverse mortgage contracts for the European countries considered. In this paper, the value of the NNEG has been estimated by means of the method proposed by Li et al. [3]. These authors propose an ARMA-EGARCH model for house price index returns. The results indicate that the value of the NNEG shows great differences among countries. These differences are explained by differences in interest rate levels, house price volatility and the rental yield level. These results also show important implications from the reverse mortgage provider's point of view. Specifically, we obtain large differences in the value of the NNEG if we consider the German government bond rate as the appropriate risk-free rate for all countries instead of each country's respective government bond rate. Using the German risk-free rate for all Eurozone countries, the reverse mortgage provider faces a higher NNEG value in the case of Spain, one of the countries that suffered much in the crisis. Moreover, in all cases it is found that reverse mortgage providers face a higher NNEG value for the female population, for higher roll-up rates and for relatively young borrowers. Specifically, the value of the NNEG is extremely high for borrowers under 80 years in all countries considered, making the reverse mortgage a non-viable product for those ages.

Consider a financial institution with an initial portfolio of reverse mortgages granted to a population of $N_0 = 1000$ men aged 70 years. This financial institution faces two different types of risk, mortality risk and house price risk. Therefore, to calculate one-year VaR and ES, it must simulate both the number of survivors and the house price in one year. The number of survivors is simulated assuming that the number of deaths follows a binomial distribution $B(N_0, q_x)$, where q_x is the probability of death for this male population, whereas the house price is simulated using the ARMA-EGARCH process with the parameters estimated

previously. We assume that all deaths occur at mid-year and that there is a delay of 6 months from the home exit until the sale of the property. Consequently, the reverse mortgage provider receives an intermediate cash flow after one year due the sale of the properties owned by borrowers that died during the year. In the estimations presented below, we assume that the amount advanced by the reverse mortgage provider is $Y_0 = 30,000$ euros (pounds in the case of the U.K.) in all cases and that the roll-up rate is 4%.

The initial value of this portfolio of reverse mortgages, from the point of view of the reverse mortgage provider, is $N_0 \cdot (Y_0 - \text{NNEG}_0)$, where NNEG_0 is the put option price estimated based on an initial house price of $S_0 = 111{,}000$ euros (pounds). The next step is to simulate a pair (N_1, S_1) for the number of survivors and the house price in one year's time. Then, based on S_1, we estimate the option price after one year (NNEG_1). Therefore, the value of this reverse mortgage portfolio after one year for this specific pair (N_1, S_1) is as follows:

$$N_1 \cdot (Y_1 - \text{NNEG}_1) + (N_0 - N_1) \cdot [Y_1 - max\,(Y_1 - S_1; 0)] \tag{1}$$

In the above expression, the second term accounts for the intermediate cash flows.

The process is repeated 10,000 times, so we obtain a distribution for the value of the reverse mortgage portfolio after one year. In the case of VaR, from this distribution we take the 99.9th percentile, whereas in the case of the ES, we compute the average of all the values below the 97.5th percentile. Finally, we calculate the loss with respect to the initial portfolio value as the present value of the portfolio after one year (the 99.9th percentile or the average of all values below the 97.5th percentile for VaR of ES respectively) minus its initial value.

The results are presented in the table below (Table 1). It is quite striking to observe the relatively high values shown by France and Germany (with losses between 14% and 18%). The U.K. exhibits much more moderate losses, approximately 6%. Finally, the low values shown by Spain (around −3%) are due to the low volatility of house price returns observed during the last part of the sample. We must keep in mind that in estimating the VaR or the ES, we are dealing with the probability of sharp decreases in house prices. Specifically, a very high volatility increases the probability of house prices reaching very low values, increasing the VaR and the ES and vice versa.

Table 1 VaR and ES estimations

	France	Germany	Spain	United Kingdom
99.9% VaR	−16.15%	−13.98%	−0.63%	−5.99%
97.5% ES	−16.47%	−17.57%	−2.70%	−6.09%

The table shows the value of the one year 99.9% VaR and 97.5% ES for a standard reverse mortgage portfolio, for all countries under study

In this sense, it is important to note that these results are in accordance with the values estimated previously. Specifically, the highest value of the volatility is found in the case of Germany, followed by France, the U.K. and Spain. These results explain the relatively high values obtained for the VaR and the ES in the cases of Germany and France and the relatively low values obtained for the U.K. and Spain.

Finally, it is important to note that another application of the method shown above for calculating regulatory capital requirements is that the value of the roll-up mortgage rate (u) can be determined so that the VaR (or ES) reaches a certain threshold.

References

1. Fisher, J.D., Geltner, D.M., Webb, R.B.: Value indices of commercial real estate: a comparison of index construction methods. J. Real Estate Financ. Econ. **9**, 137–164 (1994)
2. Geltner, D.:. Estimating market values from appriased values without assuming and efficient market. J. Real Estate Res. **8**, 325–345 (1993)
3. Li, J.S.H., Hardy, M., Tan, K.S.:. On pricing and hedging the no-negative-equity guarantee in equity release mechanisms. J. Risk Insur. **77**, 499–522 (2010)

A Basic Social Pension for Everyone?

Joseba Iñaki De La Peña and Noemí Peña-Miguel

Abstract The paper proposes a general basic pension system backed by a mixed financing model. A "basic social pension" (BSP) brings together all different aids given by different administration bodies, in a single scheme would do away with the inconsistencies and shortfalls observed in many current schemes, which result in disparities in the degree of protection received by different segments of the population. Such minimum basic social coverage would need to be backed up by a financing structure capable of guaranteeing its viability and sustainability over time in financial and social terms. It needs to reach most of the population and cover their basic necessities. Setting up a level of social protection sufficient to cover basic necessities would of course entail increasing the level of social assistance provided by the social security system, and this in turn would mean redefining the amounts payable through contributions. This redefinition has implications for sources of financing: public funding from taxation to cover the social assistance part and contributions from employers and workers to fund the contributory part.

Keywords Sustainable benefits · Minimum pension benefit · Pensions · Social security

1 Introduction

The idea of basic pensions is not new. Some researchers have put forward the idea of single universal benefit payments with amounts that vary in line with factors such as place of residence, age and family circumstances [1–3].

J. I. De La Peña (✉) · N. Peña-Miguel
University of the Basque Country (UPV/EHU), Vizcaya, Spain
e-mail: jinaki.delapena@ehu.es

M. Corazza et al. (eds.), *Mathematical and Statistical Methods for Actuarial Sciences and Finance*, https://doi.org/10.1007/978-3-319-89824-7_55

2 One Basic Social Pension and One Funding Model

BSP as a single financial allowance paid each year to each individual that is sufficient to meet the minimum requirements for survival in line with basic needs for subsistence, family situation and place of residence.

The basic needs for subsistence—food, clothing and footwear, accommodation and public transport [4]—are hard for people who live below the poverty line to obtain [5]. [6] detail the regression factors in the sample used in the Spanish EPF 2010 household budget survey, which quantifies the aggregate and individual benefits payable to 22,203 respondents (Table 1):

Technically, the funding model is based on financial actuarial equivalence [7–10]. If it entails financing the BSP not just from contributions but also from the tax revenue currently earmarked for the aids that it is to replace. Being,

$S(\tau)$: Sum of pensionable earnings of contributing workers for period τ
K_n: Percentage of contribution.
$\delta(\tau)$: Strength of interest.
$B(\tau)$: Pension spending function at time τ

$$K_n \cdot \int_{t_n}^{t_{n+1}} S(\tau) \cdot e^{-\int_{t_n}^{t_{n+1}} \delta(\tau) \cdot d\tau} \cdot d\tau = \int_{t_n}^{t_{n+1}} B(\tau) \cdot e^{-\int_{t_n}^{t_{n+1}} \delta(\tau) \cdot d\tau} \cdot d\tau$$

If,

$\int_{t_n}^{t_{n+1}} S(\tau) \cdot e^{-\int_{t_n}^{t_{n+1}} \delta(\tau) \cdot d\tau} \cdot d\tau$: This indicates the total wages of contributing workers updated to the start of the interval $[t_n, t_{n+1}]$.

Table 1 First quantile regression by category

Age	49.71∗	(128.078)
Age squared	−0.27∗	(−72.2)
Number of descendents	−648.87∗	(−681.21)
Town		
(10,000; 50,000)	508.35∗	(256.7)
More than 50,000	214.60∗	(96.58)
Head of household (HH)		
Unemployed	−273.09∗	(−89.94)
Retired	216.11∗	(73.17)
Other	86.37∗	(21.71)
Constant	1107.64∗	(105.36)
Income	0.12∗	(2662.4)
Gender	459.26∗	(248.29)

T-statistics are in parentheses. Regional dummies are also included
Source: [6]
* represents significance at 1% level

The initial population to be covered at the start of the period $l(t_n)$, breaking down the wage function $S(\tau)$ into two factors ($s(\tau)$ or the average wage at time τ and the number of people in work $l^a(t_n)$)

$\int_{t_n}^{t_{n+1}} B(\tau) \cdot e^{-\int_{t_n}^{t_{n+1}} \delta(\tau) \cdot d\tau} \cdot d\tau$: This indicates the total pension spending updated to the start of the interval $[t_n, t_{n+1}]$. The spending function $B(\tau)$ is also broken down into two factors ($PM(\tau)$ or the average pension received at time τ and the group of beneficiaries $l^b(t_n)$)

$BSP(\tau)$: Sum of social pension benefits in period τ

$AE(\tau)$: Sum of state contributions for period τ.

We have,

$$K_n \cdot \int_{t_n}^{t_{n+1}} \overbrace{\underbrace{s(\tau)}_{Productivity} \cdot \underbrace{\frac{l^a(t_n)}{l(t_n)}}_{Employment}}^{Labour\ market} \cdot e^{-\int_{t_n}^{t_{n+1}} \delta(\tau) \cdot d\tau} \cdot d\tau$$

$$+ \overbrace{\int_{t_n}^{t_{n+1}} AE(\tau) \cdot e^{-\int_{t_n}^{t_{n+1}} \delta(\tau) \cdot d\tau} \cdot d\tau}^{Tax\ factor}$$

$$= \int_{t_n}^{t_{n+1}} \overbrace{\underbrace{BSP(\tau)}_{\substack{Generosity \\ of\ the\ system}} \cdot \underbrace{\frac{l^b(t_n)}{l(t_n)}}_{\substack{Coverage \\ of\ the\ system}}}^{Institutional\ factor} \cdot e^{-\int_{t_n}^{t_{n+1}} \delta(\tau) \cdot d\tau} \cdot d\tau$$

3 Application

The cost of providing a basic social pension to part of the population (those without wage earnings)—BSP_{2010}—is (Table 2):

$$BSP_{2010} = \int_{t_n}^{t_{n+1}} B(\tau) \cdot e^{-\int_{t_n}^{t_{n+1}} \delta(\tau) \cdot d\tau} \cdot d\tau = €\,74{,}862\ million$$

Table 2 2010 state contributions to be replaced by the BSP, in millions of Euros

Item	Contribution
Regional guaranteed minimum income (RMI) schemes (100%)	766.73
Non contributory pension schemes, LISMI (social integration and job placement for the disabled), SOVI (statutory old-age and invalidity cover) and other subsidies (100%)	13,828.12
Contribution-based pensions (35%)	37,033.06
Unemployment: contributory level (85%)	20,931.25
Non contributory pension quota (100%)	142.57
State contributions$_{2010}$	72,701.73

Source: [12]

The labour market factor is obtained from the Spanish Wage Structure Survey [11]:

$$S_{2010} = \int_{t_n}^{t_{n+1}} S(\tau) \cdot e^{-\int_{t_n}^{t_{n+1}} \delta(\tau) \cdot d\tau} \cdot d\tau = € \, 517{,}909 \; million$$

By applying the above factors, the proportion of the total wage bill accounted for by this benefit in Spain in 2010 could be obtained. The resulting under a mixed financing model only 0.44% of the total wage bill is needed.

4 Conclusions

Although figures are based on a year that was in the middle of a huge recession, the BSP is based on basic needs for subsistence for people with low salaries. Thus, the changes by the crisis should be insignificant in basic consumption patterns. So, one immediate consequence of this proposal may be the restructuring of the first level, i.e. the contributory pillar. On becoming pension beneficiaries all citizens would receive BSP adjusted to their own personal characteristics (place of residence, age, dependents, etc.). If their welfare contributions over their years at work add up to a contribution-based benefit lower than the BSP then they receive the BSP. However if the sum is higher than the BSP they also receive an additional amount in line with the contributions paid (Fig. 1):

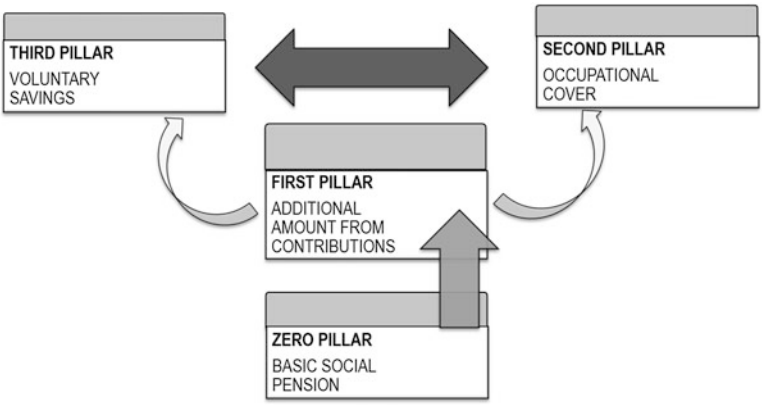


Fig. 1 Layout of the social welfare pillars and their levels of coverage. Source: Own work

Acknowledgements This study was supported by Polibienestar Research Institute and Consolidated Research Group EJ/GV: IT 897-16.

References

1. Pinilla, R.: The Basic Income of Citizenship. A Key Offer for the Renovation of the Welfare State. Icaria Ed, Barcelona (2004)
2. Pinilla, R.: Beyond the Well-Being. The Basic Income of the Citizenship Like Social Innovation Based on the Evidence. Icaria Ed, Barcelona (2006)
3. Sanzo, L.: Basic income and social protection in Spain. In: La Renta Básica en la era de las grandes desigualdades. Editorial Montesinos, Barcelona (2011)
4. Clark, D.: Sen's capability approach and the many spaces of human well-being. The Journal of Develoment Studies. **41**(8), 1139–1368 (2005)
5. Guio, A.: Material Deprivation in the EU. En EUROSTAT Statistics in Focus. Office of Official Publications of the European Communities, Luxembourg (2005)
6. Peña-Miguel, N., De La Peña, J.I., Fernández-Sainz, A.: Main factors for a proposal for a social protection floor. Social Indicators Research. **123**(1), 59–76 (2015)
7. Kaiser, E.: Functional Equations for the Social Mathematics. A.I.S.S. III. International Congress, Madrid (1962)
8. Nieto, U., Vegas, J.: Actuarial Mathematics. Ed. Mapfre, Madrid (1993)
9. Thullen, P.: Mathematische Methoden der Socialen Sicherheit: Systeme der Invaliden Alters- und. Ed. Buch. Karlsruhe Werl, Versicherungswirtschaf (1977)
10. Zelenka, A.: Quelques remarques sur le régime financier in ISSA: actuarial and statistical problems of social security, vol. III (Geneva and Rome) (1958)
11. EPA – Labour Force Survey: INE – Spanish National Statistics Office (2010)
12. IGAE – Government Administration's General intervention body. Ministerio de Hacienda y Administraciones Públicas (2010)

A Copula-Based Quantile Model

Giovanni De Luca, Giorgia Rivieccio, and Stefania Corsaro

Abstract A copula-based quantile model is built. The estimates are compared to the estimates obtained using the multivariate CAViaR model, which extends the univariate version of the model. The comparison is firstly made in terms of Kupiec and Christoffersen test. Moreover, a further comparison is made using two loss functions that evaluate the distances between the losses and the VaR measures in presence of a violation. The results show that the copula approach is highly competitive providing, in particular, estimated quantiles which generally imply a lower value for the two loss functions.

Keywords Value-at-Risk · Copula function · Loss function

1 Introduction

Risk management has typically focused on the Value-at-Risk (VaR) as the main risk measure. This value is financially interpreted as the worst loss expected over a given period of time with a given probability. From a statistical point view, VaR is a quantile of the losses distribution, that is an unobservable quantity that can be estimated once the distribution of the losses is known. The most traditional technique is the estimation of a dynamic VaR as a byproduct of a heteroskedastic model, e.g. a GARCH model. A second approach stresses the possibility of estimating directly the dynamics of the quantile: instead of considering a time-varying variance which leads to a time-varying quantile, the dynamics is estimated directly on the quantiles. Engle and Manganelli [3] proposed the CAViaR model, with a specified quantile a time t depending mainly on its own lagged values and on a function of past returns. Recently, the univariate approach has been extended to take into account possible spillovers over VaRs [6] in a bivariate formulation.

G. De Luca (✉) · G. Rivieccio · S. Corsaro
University of Naples Parthenope, Naples, Italy
e-mail: giovanni.deluca@uniparthenope.it; giorgia.rivieccio@uniparthenope.it; stefania.corsaro@uniparthenope.it

M. Corazza et al. (eds.), *Mathematical and Statistical Methods for Actuarial Sciences and Finance*, https://doi.org/10.1007/978-3-319-89824-7_56

In this paper the quantiles obtained by the bivariate CAViaR model are compared to the quantiles estimated exploiting a copula as joint function of some univariate quantiles.

2 Modelling and Data

Let Y_t be the random variable denoting the daily return at time t, and $q_t(\theta)$ the corresponding quantile at θ (with $0 < \theta < 1$). In the CAViaR model by Engle and Manganelli [3] the dynamics of the quantiles is the following:

$$q_t(\theta) = \beta_0 + \beta_1 q_{t-1}(\theta) + \beta_2 f(y_{t-1}).$$

In a multivariate context, [6] have proposed a framework which can be conveniently thought of as a Vector Autoregressive (VAR) extension to quantile models. The aim is to analyze possible spillovers on the VaRs. A simple version of the proposed structure in a bivariate case relates the conditional quantiles of the random variables Y_{1t} and Y_{2t} according to a VAR structure:

$$q_{1t}(\theta) = b_{10} + b_{11} q_{1t-1}(\theta) + b_{12} q_{2t-1}(\theta)$$
$$q_{2t}(\theta) = b_{20} + b_{21} q_{1t-1}(\theta) + b_{22} q_{2t-1}(\theta)$$

The off-diagonal coefficients b_{12} and b_{21} represent the measure of tail codependence between the two random variables. If $b_{12} = b_{21} = 0$ the model reduces to two independent univariate CAViaR models.

An alternative approach is based on the copula functions. The idea is that of finding a copula function that could provide a flexible representation of the relationship among q_{1t} (or q_{2t}), q_{1t-1} and q_{2t-1}. For asset 1, the conditional estimates are given by $\hat{q}_{1t}^{cop}(\theta) = E(q_{1t}|q_{1t-1}, q_{2t-1})$ evaluated using the conditional copula function

$$C(F_{1t}|F_{1t-1}, F_{2t-1}) \tag{1}$$

where F_{it} is the estimated distribution function of the i-th asset θ-quantile at time t.

Because of the unobservable nature of quantiles, initial estimates are required to let the procedure start. Starting from the quantile estimates provided by the univariate CAViaR models, $\hat{q}_{it}^{U}(\theta)$, $i = 1, 2$, the idea is to update these initial estimates considering the conditional dynamics of the remaining time series. Crucial points are the identification of satisfactory marginal distribution to obtain F_{it} and the choice of the copula function in (1).

Data sample includes daily returns (from January 2008 to February 2014, for a total of 1584 observations) of 20 assets included in Eurostoxx50.[1] Each quantile ($\theta = 0.05$) dynamics is analyzed in relation to the quantile of an equally weighted average of all assets in portfolio, denoted as q_{Pt}. In particular, we have estimated the time-varying quantiles q_{it}^{COP} ($i = 1, \ldots, 20$) specified in Eq. (1), and compared them with the time-varying quantiles computed according to the multivariate CAViaR model (q_{it}^{MCAV}). For the estimation of q_{it}^{COP}, the initial quantile estimates are provided by the univariate CAViaR model with $f(y_{t-1}) = |y_{t-1}|$. To identify the marginal distributions of $\hat{q}_{it}^{U}$ we have chosen a mixture of three log-normal distributions, that is a fairly general assumption.

Then, a mixture of copula functions has been used for applying (1). The reason of such a choice is given by the need to ensure asymmetric tail dependence. In fact, the popular Student's t copula has the limit of symmetric tail dependence coefficients, while it is easy to check that the association between estimated quantiles of each asset and estimated quantile of the portfolio is greater in the lower tail. So, the Student's t copula has been mixed with a copula function which allows for a different degree of association in the two tails. The selected mixture is then composed by a Clayton copula, which presents only lower tail dependence, and a Student's t copula with weights, respectively, p e $1 - p$, that is

$$C_M = pC_C + (1 - p)C_T.$$

As a result, the lower tail dependence coefficient is given by

$$\lambda_L^M = p\lambda_L^{C_C} + (1 - p)\lambda_L^{C_T}$$

and the upper tail dependence coefficient is

$$\lambda_U^M = (1 - p)\lambda_U^{C_T}.$$

In order to evaluate the estimated quantiles, we have used the unconditional coverage test (see [4]) and the conditional coverage test (see[2]). Figure 1 (left) reports the empirical coverage of $\hat{q}_{it}^{MCAV}$ and $\hat{q}_{it}^{cop}$ together with the critical values denoting the acceptance region. The former estimates present an empirical coverage very close to the nominal one, the latter being generally a bit more far. However, for all the assets, both the estimates pass the conditional and unconditional coverage tests.

[1] AIRBUS, ALLIANZ, AXA, BANCO SANTANDER, BASF, BAYER, BNP PARIBAS, DAIMLER, DEUTSCHE POST, DEUTSCHE TELEKOM, ENI, L'OREAL, LVMH, PHILIPS, SIEMENS, SOCIETE GEN, TELEFONICA, TOTAL, UNILEVER, VOLKSWAGEN.

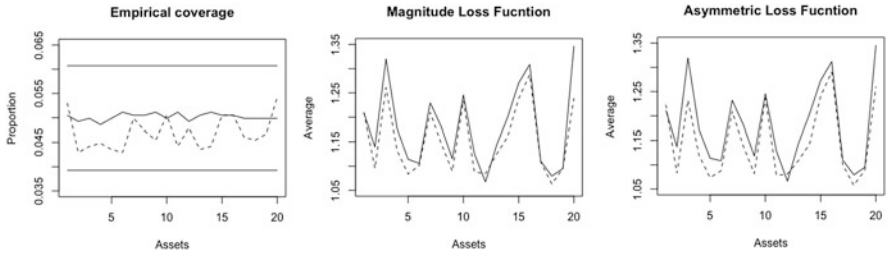


Fig. 1 Proportion of violations (left), average values of $L_{M,t}$ (middle) and $L_{A,t}$ (right) for the 20 assets. Solid and dashed lines refer, respectively, to $\hat{q}_{it}^{MCAV}$ and $\hat{q}_{it}^{cop}$

Moreover, we have considered two loss functions:

1. the Magnitude Loss Function [5] that takes into account the distance between the loss and a VaR measures (i.e. a quantile) in presence of a violation,

$$L_{M,t} = \begin{cases} 1 + (y_t - VaR_t)^2 & \text{if } y_t < VaR_t \\ 0 & \text{if } y_t \geq VaR_t \end{cases}$$

2. an asymmetric version due to [1] depending on $P = \exp(R)(R > 0) + 1 \cdot (R \leq 0)$ where $R = (\hat{\theta} - 0.05)/0.05$, and $\hat{\theta}$ is the empirical coverage,

$$L_{A,t} = \begin{cases} 1 + P(y_t - VaR_t)^2 & \text{if } y_t < VaR_t \\ 0 & \text{if } y_t \geq VaR_t \end{cases}$$

The results show that the averages of L_{Mt} and $L_{A,t}$ are smaller for the copula-based quantiles, respectively, in 17 and 18 cases out 20 (see Fig. 1).

3 Conclusions

We have explored a flexible copula approach to obtain quantile estimates to be compared with the quantiles provided by the bivariate CAViaR model. The application of the copula-based model to 20 asset returns of Eurostoxx50 emphasizes in most cases the better performance of the novel quantiles in terms of the distance assessment between a loss and a VaR measure.

References

1. Amendola, A., Candila, V.: Evaluation of volatility predictions in a VaR framework. Quant. Finance **16**, 695–709 (2016)
2. Christofferssen, P.: Evaluating interval forecasts. Int. Econ. Rev. **39**, 841–862 (1998)
3. Engle, R.F., Manganelli, S.: CAViaR: conditional autoregressive value at risk by regression quantiles. J. Bus. Econ. Stat. **22**, 367–381 (2004)
4. Kupiec, P.: Techniques for verifying the accuracy of risk management models. J. Deriv. **3**, 73–84 (1995)
5. Lopez, J.A.: Methods for evaluating value-at-risk estimates. Econ. Rev. **2**, 3–17 (1999)
6. White, H, Kim T.H., Manganelli, S.: VAR for VaR: measuring tail dependence using multivariate regression quantiles. J. Economet. **187**, 169–188 (2015)

International Longevity Risk Pooling

Clemente De Rosa, Elisa Luciano, and Luca Regis

Abstract This paper studies the problem of an insurance company that has to decide whether to expand her portfolio of policies selling contracts written on a foreign population. We quantify diversification across populations and cohorts using a parsimonious continuous-time model for longevity risk. We present a calibrated example, based on annuity portfolios of UK and Italian males aged 65–75. We show that diversification gains, evaluated as the reduction in the portfolio risk margin following the international expansion, can be non-negligible.

Keywords Longevity risk · International expansion of insurance companies · Longevity risk pooling

1 Motivation and Related Literature

In the last 20 years, insurance companies have been expanding internationally, via subsidiaries operating in different countries or via cross-border mergers and acquisitions.[1] Reinsurance companies have always been more geographically diversified than insurance companies [3], because their portfolios are more easily disconnected

[1] In Europe, for instance, the creation of a common regulation framework in the middle of the Nineties gave rise to a wave of international expansions and M&A operations.

C. De Rosa
Scuola Normale Superiore Pisa, Pisa, Italy
e-mail: clemente.derosa@sns.it

E. Luciano
Collegio Carlo Alberto and ESOMAS Department, University of Torino, Torino, Italy
e-mail: elisa.luciano@unito.it

L. Regis (✉)
Department of Economics and Statistics, University of Siena, Siena, Italy
e-mail: luca.regis@unisi.it

M. Corazza et al. (eds.), *Mathematical and Statistical Methods for Actuarial Sciences and Finance*, https://doi.org/10.1007/978-3-319-89824-7_57

from the geographical localization of their branches. As a result, the largest insurers and reinsurers in most cases concentrate the bulk of their activities in foreign countries. Evidence is provided for instance in [5].

The connection between internationalization and profitability has been the subject of some recent studies, while the link between internationalization and risk or capital requirement reduction has been very little explored. This paper addresses a typical diversification gain from internationalization of life insurance or reinsurance companies, namely longevity-risk pooling across populations. We consider the situation of an insurer who is facing the choice of expanding her portfolio of liabilities (for simplicity, an annuity portfolio held by policyholders of different ages) either in her domestic market or in a foreign one. We want to determine the impact of the expansion in terms of regulatory risk margin for Solvency II purposes.

Why is international diversification of life insurance portfolios beneficial? Because, even if—in expectation—longevity has been steadily increasing on a worldwide scale, longevity risk of different populations may differ. By longevity risk we mean the amount of unexpected decrease in mortality rates, that may be non-perfectly correlated across countries. Pooling portfolios of policies written on the lives of different populations allows to diversify longevity risk. To model and calibrate longevity risk of different cohorts and nations, we make use of a novel model developed in [4]. We consider a domestic and a foreign market, and heterogeneous cohorts in each market. We apply the model to estimate the joint dynamics of the mortality rates of UK and Italian males aged between 65 and 75, estimating the whole correlation matrix across the eleven cohorts of the two different populations.

Based on the model, we measure the risk effects of an international expansion through the risk margin reduction. Reminiscent of the Solvency II regulation framework, we compute the risk margin of a life insurance portfolio as its value-at-risk (VaR). The difference between the risk margin of a life insurer who expands its portfolio domestically and, all others equal, diversifies it internationally, provides us with a (dollar-based) measure of the diversification benefits. Our application, that considers a UK-based annuity provider, shows that the risk margin can reduce up to 3% as a proportion of the actuarial value, in the case of a foreign expansion.

2 Pooling Metrics and Impact

We consider an insurer, based in the UK, who, at time t, has a portfolio Π^0 with actuarial (best estimate) value $AV_{\Pi^0}(t)$. She wants to increase the size of her annuity portfolio and needs to choose between two possible strategies. The first one is just to sell new contracts to her Domestic population. In this case, we denote with n_i' the number of new contracts sold to people aged x_i, with $N_i(t)$ the actuarial value of each annuity, with Π^D the portfolio composed of just these new annuities, and with Π^1 the total portfolio of old and new contracts. The actuarial value of the new

portfolio is

$$AV_{\Pi^D}(t) = \sum_{i=1}^{m} n_i' N_i(t), \tag{1}$$

and

$$AV_{\Pi^1}(t) = AV_{\Pi^0}(t) + AV_{\Pi^D}(t). \tag{2}$$

The total value of the portfolio Π^1 is the sum of this actuarial value and the risk margin $RM_{\Pi^1}(t)$, computed as the discounted VaR of the total portfolio at a given confidence level $\alpha \in (0, 1)$:

$$\Pi^1(t) = AV_{\Pi^1}(t) + RM_{\Pi^1}(t) = AV_{\Pi^0}(t) + AV_{\Pi^D}(t) + RM_{\Pi^1}(t), \tag{3}$$

where

$$\begin{aligned} RM_{\Pi^1}(t) &= D(t, t+T) \cdot VaR_\alpha\big(AV_{\Pi^1}(t+T) - \mathbb{E}_t[AV_{\Pi^1}(t+T)]\big) = \\ &= D(t, t+T) \cdot inf\{l \in \mathbb{R}^+ : P(AV_{\Pi^1}(t+T) \\ &\quad - \mathbb{E}_t[AV_{\Pi^1}(t+T)] > l) < 1-\alpha\}, \end{aligned}$$

where $P(\cdot)$ is the probability of the event in brackets, E_t the time-t expected value operator and $D(t, t+T)$ is the discount factor from $t+T$ to t.

The second possible strategy is to acquire a new portfolio of annuities Π^F, written on a Foreign population, namely the Italian one. To compare the two strategies, we simply assume that for each age x_i the number of annuities written on people aged x_i in the foreign population is still n_i'. The actuarial value of this portfolio is

$$AV_{\Pi^F}(t) = \sum_{i=1}^{m} n_i' N_i^F(t), \tag{4}$$

and

$$AV_{\Pi^2}(t) = AV_{\Pi^0}(t) + AV_{\Pi^F}(t). \tag{5}$$

Moreover,

$$\Pi^2(t) = AV_{\Pi^2}(t) + RM_{\Pi^2}(t) = AV_{\Pi^0}(t) + AV_{\Pi^F}(t) + RM_{\Pi^2}(t). \tag{6}$$

Table 1 presents the actuarial value, risk margin with $\alpha = 99.5\%$ and $T = 15$ years, total value and risk margin as a percentage of the actuarial value for different portfolios ($\%RM$), which obtain from an initial portfolio Π^0 of 1100

Table 1 Effects of geographical diversification ($r = 0\%$)

Portfolio	AV	RM	Π	$\%RM$	Δ
Π^0	1.9097×10^4	2.1318×10^3	2.1228×10^4	11.16%	–
Π^F	2.0093×10^4	1.9060×10^3	2.1999×10^4	9.49%	1.67%
Π^1	3.8193×10^4	4.2636×10^3	4.2457×10^4	11.16%	0
Π^2	3.9189×10^4	4.0378×10^3	4.3227×10^4	10.30%	0.86%
Π^3	5.9282×10^4	5.9437×10^3	6.5226×10^4	10.03%	1.13%
Π^1_{opt}	4.1675×10^4	3.6480×10^3	4.5323×10^4	8.75%	2.41%
Π^2_{opt}	4.2400×10^4	3.4234×10^3	4.5824×10^4	8.07%	3.09%

Π^0 is a portfolio of 1100 policies sold to UK insureds; Π^F is composed of 1100 policies sold to Italian insureds; $\Pi^1 = 2\,\Pi^0$; $\Pi^2 = \Pi^0 + \Pi^F$; $\Pi^3 = \Pi^0 + 2\Pi^F$; Π^1_{opt} couples Π^0 with 1100 policies sold to 66 year-old UK males; Π^2_{opt} couples Π^0 with 1100 policies sold to 66 year-old Italian males. $\Delta = -(\%RM(\Pi^0) - \%RM(\Pi^i))$ is the percentage risk margin reduction of portfolio Π^i relative to Π^0

UK insureds (100 for each age from age 65 to age 75). Diversification benefits are evaluated as the percentage risk margin reduction, Δ, of portfolio i relative to Π^0, $\Delta = -(\%RM(\Pi^i) - \%RM(\Pi^0))$. The different alternatives imply selling additional policies, and must be compared using their percentage (not absolute) risk margin as a pooling or riskiness metrics. Results are taken from the paper [4], which details also the estimation of the underlying mortality model and calibrates the correlation structure across ages and populations. The portfolios Π^F, Π^1 and Π^2 are as described above. $\Pi^3 = \Pi^0 + 2\Pi^F$ represents a more aggressive foreign expansion obtained by doubling the size of the Italian expansion. We also consider the cases Π^1_{opt} and Π^2_{opt}, in which the domestic and foreign expansions are performed optimizing the mix of the newly added contracts, targeting the cohorts that lead to the maximally diversified portfolios, without and with geographical diversification respectively. Based on the correlation matrices estimated in [4], they are obtained by selling 1100 contracts to the UK and Italian 66 years old males, respectively. Π^2_{opt} is the most diversified portfolio, as its percentage risk margin is minimal (8.07%). From the table, we can appreciate the gains deriving from international diversification according to our metrics Δ. It amounts to 0.86% in the case of a sub-optimal expansion, and reaches 3.09% for an optimized foreign expansion.

3 Conclusions

In this short paper we established a metrics to compute the diversification benefits of an international expansion by a life insurer and provided a calibrated example. In practice, such benefits may happen to be counterbalanced by the costs of the foreign portfolio acquisition process. These costs, that are—say—the costs of opening a foreign affiliate, or the fees for an M&A operation, etc., may be

substantial. As an alternative to a physical expansion, the insurer may obtain the same diversification benefit operating on the longevity derivatives market (see [2] for instance). In particular, the insurer can expand internationally by receiving a fixed periodical fee and paying the survivorship of the foreign cohorts. Thus, the risk margin reduction benefits of a foreign expansion can be replicated by selling protection through a longevity swap. Even in this case, however, the costs of structuring the agreement and coping with informational asymmetries [1], can reduce the diversification gains. The final balance can be decided only case by case.

Acknowledgements The Authors would like to thank the Global Risk Institute, Toronto, for financial support, as well as its seminar participants in October 2015 and January 2017.

References

1. Biffis, E., Blake, D., Pitotti, L., Sun, A.: The cost of counterparty risk and collateralization in longevity swaps. J. Risk Insur. **83**(2), 387–419 (2016)
2. Blake, D., Cairns, A., Dowd, K., MacMinn, R.: Longevity bonds: financial engineering, valuation, and hedging. J. Risk Insur. **73**(4), 647–672 (2006)
3. Cummins, J.D., Xie, X.: Mergers and acquisitions in the US property-liability insurance industry: productivity and efficiency effects. J. Bank. Financ. **32**(1), 30–55 (2008)
4. De Rosa, C., Luciano, E., Regis, L.: Geographical Diversification in Annuity Portfolios. Collegio Carlo Alberto Notebook 546 (2017)
5. Outreville, J.F.: Foreign affiliates of the largest insurance groups: location-specific advantages. J. Risk Insur. **75**(2), 463–491 (2008)

A Two-Steps Mixed Pension System: An Aggregate Analysis

Pierre Devolder, Inmaculada Domínguez-Fabián, Francisco del Olmo-García, and José A. Herce

Abstract The change in the economic and sociodemographic context, framed by a continuous increase in longevity, the consequences of the economic crisis as well as the lack of adequate adjustments of the Social Security retirement pension systems everywhere, entail risks for workers and for the Social Security itself. Against the background of the change in agents' behaviors throughout the life cycle and the presence of an adverse selection problem in the annuities market, we describe in this paper a 'two-steps mixed pension system' that tries to solve the pressure that increasing longevity is putting on conventional pension schemes to provide adequate and sustainable pensions for all. In our proposal, Social Security, preceded by a term-annuities scheme, is 'reinvented' to continue to ensure retirees' incomes from their 'grand age' onwards.

Keywords Two-steps mixed pension system · Term annuity · Aggregate analysis · Grand age

P. Devolder
Institute of Statistics, Biostatistics and Actuarial Sciences, Catholic University of Louvain, Louvain-la-Neuve, Belgium
e-mail: pierre.devolder@uclouvain.be

I. Domínguez-Fabián (✉)
University of Extremadura, Cáceres, Spain
e-mail: idomingu@unex.es

F. del Olmo-García
Universidad de Alcalá, Alcalá de Henares, Spain
e-mail: francisco.olmo@uah.es

J. A. Herce
Afi Finance School, Madrid, Spain
e-mail: jherce@afi.es

M. Corazza et al. (eds.), *Mathematical and Statistical Methods for Actuarial Sciences and Finance*, https://doi.org/10.1007/978-3-319-89824-7_58

1 Introduction

Economic and demographic change, characterized by a continuous increase in longevity, the impact of the economic crisis, the digital transformation as well as the lack of adequate adjustments in the Social Security (SS, in what follows) retirement pension systems everywhere, entail risks for workers and for SS itself. Many reforms of public pension systems have been carried out in recent years, either changing system's parameters or adopting structural changes. Some reforms have aimed at increasing capitalization components in the determination of the final pension through a life annuity to complement the public retirement pension as a significant second retirement income.

The conventional mixed pension systems are based on a two-pillar structure with a first pillar being the conventional SS, pay-as-you-go scheme that provides a public retirement pension which is complemented with a life annuity, generated by a fully funded, employers' sponsored scheme. Both benefits are received during the entire retirement period in a simultaneous and complementary way. This type of conventional complementary system has several problems. The first is that annuities entail a severe problem of adverse selection (see [1–4]) and thus they become unduly expensive. On the other hand, the kind of longevity insurance, offered by SS, has barely changed the retirement age since it was created, when life expectancy at birth was around 40 years, and around 10 years at age 65.

2 The Two-Steps Mixed System

The many pensions reforms proposed and implemented till now have not solved the problems of long-term sustainability and adequacy of retirement pension systems and some alternatives are truly needed. Within the line of research we describe here, several papers have been presented both by the authors [2, 3] and by other researchers [5, 7]. This literature is proposing the 'reinvention' of SS through a mixed system implemented in two steps. In the first step a term annuity offered by a fully funded (compulsory but private) scheme would cover retirees between the established legal retirement age and what we call the 'grand age'. This grand age could be equivalent today to what age 65 meant one century ago (around age 80 today, see [3]), but could also be lower. After this grand age, the retiree would obtain a SS annuity (pay-as-you-go based if needed) sustainable and adequate [2, 3, 6, 7]. Term annuities from insurance companies are much more efficient than life annuities while SS life annuities are the best longevity insurance ever invented. Insuring retired workers after grand age is what SS precisely did more than one hundred years ago.

Figure 1 shows the contributions that a typical individual would make to and the retirement income that he or she would receive both in the standard mixed model and in the two-steps mixed model. For this illustration a contribution rate of 15%

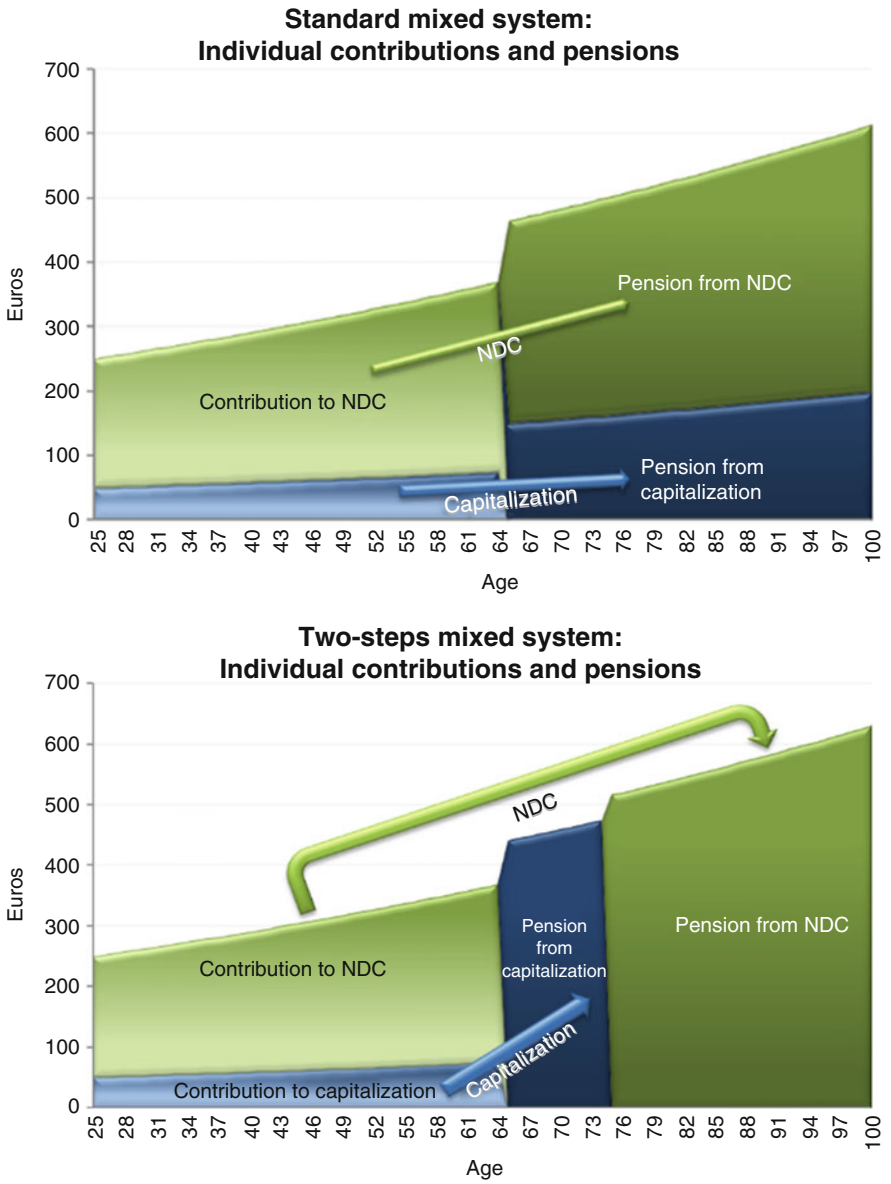


Fig. 1 (Illustrative figures). Source: Own computations

of wages to the SS scheme (Notional Defined Contribution variety or NDC, in what follows) and a 5% rate to the funded scheme have been assumed. Initial working age is 25 years, the contributions career is not interrupted, the retirement age is 65 years and the grand age has been set at 75 years, with a maximum survival age

of 100 years. PEM-2000 mortality tables have been used. A nominal interest rate of 3%, a notional rate of 1% and a wage revaluation of 1.5% per year have also been assumed for the computations.

This way of presenting the sequence of contributions and benefits combined in time, is what allows us to talk about the 'reinvention' of SS [3]. The two-steps mixed system solves the vital algebra's inconsistency problem mentioned above since the pension funded by PAYG is paid from grand age until death; which means that the number of years the scheme must pay benefits to any retiree is far lower than under the current system. On the other hand, the fully funded scheme term annuities, payable since retirement until grand age, are marketed more efficiently than a life annuity, since the former are significantly cheaper on unit terms as they must not be insured against significant longevity but in a small part of this risk. This product is thus more interesting both for workers and for insurance companies, and could eventually lead to a significant expansion of annuities markets everywhere in a more natural way.

3 Aggregate Perspective

Figure 2 shows the system's flows of income and expenses for the base year 2017 and for 2050. For each of the panels, on the left side, the standard mixed system's income and expenses are shown where those for the two-steps mixed system are shown in the right side of the panels. Income flows are the same for both systems as contribution rates are the same but, concerning expenditure flows, the two-steps system is cheaper to run, since retirees do not receive SS income until they are 75 years old (the grand age), although system's expenses from 75 years on are higher since the retirement income generated in the NDC scheme are higher than in the standard PAYG one.

The graph also shows how the financial problems of the standard PAYG scheme are exacerbated in 2050, as pensions expenses are higher and income is lower, due respectively, to increased longevity and a reduced workforce. The NDC scheme in the two-steps mixed system generates savings since inception year and increasingly so until 2050, that are accumulated over time.

4 Concluding Comments on Transition Possibilities

The two-steps system we propose means a radical reform of the standard mixed system that will bring advantages both for the individuals and for the two schemes that make up the whole system. In both cases a fully funded, quasi-mandatory private scheme must be part of the pension system to provide, one way or another, adequate total retirement income since ordinary retirement age (that can be adapted or made optional to a degree) until death. For a smooth transition, it would be

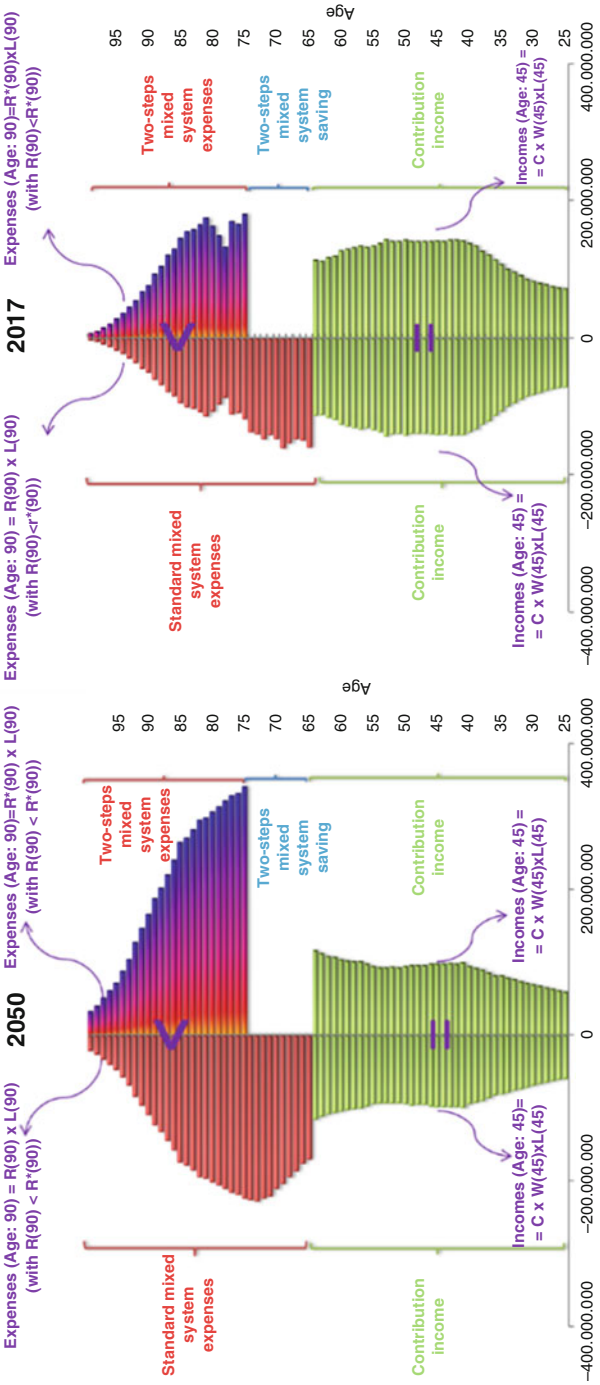


Fig. 2 Note: The fully funded scheme is not shown in this graph as for both systems it is sustainable. Source: Own computations

advisable to establish an age line between those who remain in the standard system and those who can pass to the two-steps system. Precise computations of the system's implicit debt would be necessary to determine that dividing age line. If the transition is made from a standard complementary system, in which individuals have accumulated retirement savings during their labor lives, the new NDC SS branch could start generating savings since inception until retirees reach their corresponding grand ages. This could allow the scheme to cope with the large implicit SS debt accumulated by historical PAYG systems everywhere. These savings could also be used to encourage, through tax deductions and/or allowances, voluntary transitions to the new two-steps mixed system. Other tax incentives can stimulate the accumulation of retirement savings during labor life, either in final savings products or in other assets such as housing, to be converted into a term annuities to be received between the legal retirement age and the grand age, also helping the transition towards a two-steps mixed system of the kind we are proposing here.

References

1. Blake, D., Dowd, K., Cairns, A.: Longevity risk and the Grim Reaper's toxic tail: the survivor fan charts. Insur. Math. Econ. **42**, 1062–1066 (2008).
2. Domínguez, I., Herce, J.A., del Olmo, F.: Reinventing social security: towards a mixed two-steps pension system. In: Álvarez, J. (ed.) Public Pension Systems: The Greatest Economic Challenge of the 21st Century, Chapter 16. Springer, Forthcoming (2018)
3. Herce, J.A., del Olmo, F.: Reinventar la Seguridad Social asegurando la gran edad. In: Herce, J.A. (ed.) Pensiones: Una reforma medular, pp. 191–209. Fundación de Estudios Financieros y Círculo de Empresarios, Madrid (2013)
4. Whitehouse, E.R., Zaidi, A.: Socio-economic differences in mortality: implications for pensions policy. OECD social, employment and migration working papers, no. 71. OECD Publishing (2008)
5. Tomoyuki, K.: Necessary Adjustment of Pensions Under Aging Societies. PBSS Coloquium, Cancún (2017)
6. Bertein, S., Larraín, G., Puenta, A.: Rol de un seguro de longevidad en América Latina: casos de Chile, Colombia, México y Perú. Seminario Internacional FIAP 2015, Montevideo (2015)
7. Larraín, G., Ballesteros, S., García, S.: Longevidad y pensiones: una propuesta de seguro para la cuarta edad. Working paper. Facultad de Economía y Negocios de la Universidad de Chile, No 441, Santiago de Chile (2017)

The Influence of Dynamic Risk Aversion in the Optimal Portfolio Context

Antonio Díaz and Carlos Esparcia

Abstract Despite the influence of risk aversion in the optimal portfolio context, there are not many studies which have explicitly estimated the risk aversion parameter. Instead of that, researchers almost always choose random fixed values to reflect the common levels of risk aversion. However, the above could generate optimal portfolios, which do not reflect the actual investor's attitude towards risk. Otherwise, as it is well known, an individual is more or less risk averse according to the economic and political circumstances. Given the above, we model the risk aversion attitude so that it changes over time, in order to take into account the variability in agents' expectations. Therefore, the aim of this paper is to shed light on the choice of the risk aversion parameter that correctly represents the investors' behaviour. For that purpose, we build optimal portfolios for different types of investment profiles in order to compare whether it is better to use a constant risk aversion parameter or a dynamic one. In particular, our proposal is based on estimating the time-varying risk aversion parameter as a derivation of the market risk premium. For that purpose, we implement several statistical univariate and multivariate models. Specifically, we use conditional variance and correlation models, such as GARCH (1, 1), GARCH-M (1, 1) and DCC-GARCH.

Keywords Optimal portfolio · Time-varying risk aversion · Market risk premium · GARCH models

1 Introduction

In accordance with the mean-variance approach, we can partially order the set of investment opportunities, reducing the choice of investors to those portfolios located on the efficient frontier. However, with this approach, the investors cannot compare

A. Díaz · C. Esparcia (✉)
Facultad de cc. Económicas y Empresariales, Universidad de Castilla-La Mancha, Albacete, Spain
e-mail: Antonio.Diaz@uclm.es; Carlos.Esparcia@uclm.es

M. Corazza et al. (eds.), *Mathematical and Statistical Methods for Actuarial Sciences and Finance*, https://doi.org/10.1007/978-3-319-89824-7_59

which alternatives are dominant among themselves; therefore, they are not allowed to select the investment portfolio that best meets their economic objectives. To find this portfolio, we must use a different criterion, incorporating the individual risk attitude. Although these preferences are very complex (they depend on, for instance, the age, gender, education level, and income of the individual), to make their implementation easier, they are represented by a single parameter that summarizes the personal level of risk aversion, the risk aversion parameter.

In uncertainty contexts, it is possible to represent the preferences of economic agents from their expected utility. In short, it is suggested that financial theory has developed utility functions to assess how good an investment is. In this context, the second derivative of the utility function (concavity) reflects the risk aversion level of an individual, but it is necessary to adjust this measure by the first derivative to ensure that it does not change under linear transformations. In general, we assume that investors are risk-averse. For that reason, in this work, we only focus on the analysis of two increasing and concave utility functions, such as the quadratic and CARA specifications.

In spite of playing a key role in the optimal portfolio construction, there are few studies that have explicitly estimated the risk aversion of an investor. Instead, they choose random values to reflect the common levels of risk aversion. The equity literature on risk aversion developed based on the review proposed by Arrow [1], who affirmed that the risk aversion parameter should be approximately 1. Otherwise, in the equity context, several studies have been published that differ in their estimations of risk aversion. For instance, Mehra and Prescott [2] argued that this parameter should be greater than 10. Moreover, Ghysels et al. [3] affirmed the risk attitude should be between 1.5 and 2, on average, while Guo and Whitelaw [4] established the mentioned parameter of 4.93.

However, common sense tells us that the use of fixed arbitrary values for this parameter could yield optimal portfolios that do not reflect the actual investor's attitude towards risk. An individual is more or less risk averse according to the economic and political circumstances. For instance, we are currently in a period in which even the most adventurous investor has had to reduce his optimistic expectations. Given that, it seems reasonable to model the risk aversion parameter so that it changes over time, to consider the variability in the agents' expectations. In this context, there are some studies in the financial literature that refer to time-varying risk aversion. For instance, Kim [5] proposes a consistent indicator of conditional risk aversion in consumption-based CAPM. Other studies have differed widely in their estimates of time-varying risk aversion, such as Dionne [6], who aims to extend the concept of orders of conditional risk aversion to orders of conditional dependent risk aversion. However, our motivation follows the framework proposed by Frankel [7] and revised by Giovannini and Jorion [8] and Cotter and Hanly [9],

which is based on estimating the risk aversion parameter as a derivation of the CRRA.[1]

According to the last paragraphs, the aim of this paper is to reveal the optimal parameter choice that provides a better representation of the investors' attitude towards risk. In particular, our proposal is based on estimating the time-varying risk aversion parameter as a derivation of the mentioned CRRA and strongly related to the market risk premium context. For that purpose, we build a well-diversified portfolio through the selection of ten risky equities traded on the Eurostoxx-50 index. From here, we introduce the time-varying modeling of probability distribution moments, to consider the optimal portfolios changing over time. To reach the above aim, and focusing on the optimal portfolio problem, we propose the application of conditional variance and correlation schemes such as GARCH (1, 1) and DCC-GARCH.

Otherwise, we estimate the CRRA from the market risk premium, which depends on the mean and variance of the market.[2] This estimation allows us to obtain the risk aversion attitude of an investor in a single number. However, in this research we are more interested in the time-varying risk aversion, not in a constant parameter. Thus, we model the market mean and variance through conditional models such as the GARCH (1, 1)-M specification. Further, we aim to assess whether it is better to work with a constant or changing risk aversion parameter. For that purpose, we build optimal portfolios for different types of investment profiles, a conditional one associated with the CRRA, Model A, and one based on constant risk aversion, Model B. Note that we assess the portfolios for different time frames, ones related to calm periods and others related to economic recession.

2 Main Results and Findings: The Certainty-Equivalent

In this section, we show a well-known performance measure to assess our portfolio management. Figure 1 shows the Certainty-Equivalent. Then, analysing the performance results from Fig. 1, we have tested that the highest risk premium offered to exchange our portfolio, is the one showed by Model A (dynamic risk aversion), since it is the one that usually offers us the greatest relationship over time. We obtain evidence that portfolios with better performances based on Certainty-Equivalent ratio are those ones associated to the time-varying risk aversion attitude, while those with a constant risk aversion parameter (Model B), have a negative risk-return relationship and a very unstable trend throughout the whole studied period (calm and stress).

[1]This term refers to the changes in relative risk aversion, which is a way to express the risk aversion attitude of an investor through his utility function.

[2]Note that we use the daily closing prices of EuroStoxx-50 Index as market portfolio and the 3-month German Treasury Bills as risk free rate.

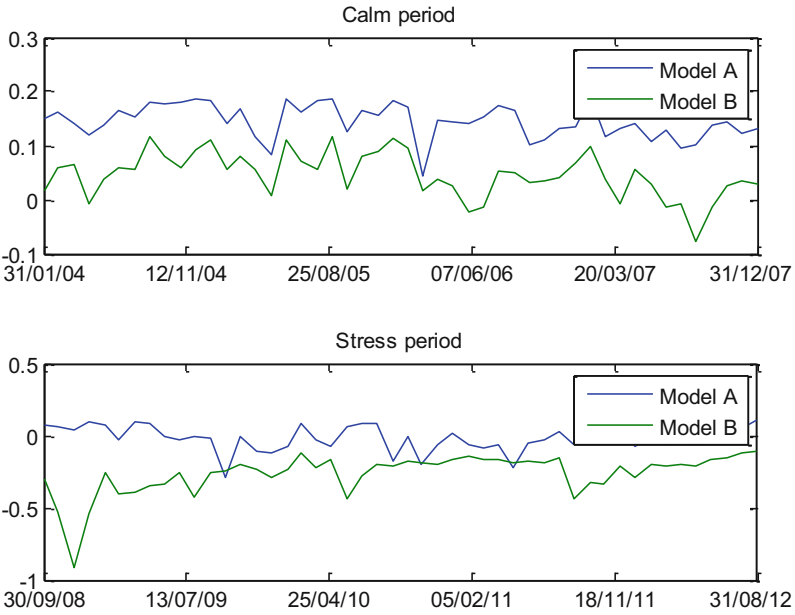


Fig. 1 We plot the monthly evolution of the Certainty-Equivalent for both studied models (dynamic and constant risk aversion). This performance measure is expressed in annual terms. Note that we generate different kind of time-varying portfolios for two sample periods. Top graph represents a calm period, which runs from 01/01/2004 to 31/12/2007. Moreover, the second timeframe is a more stressed one (located at the bottom) and comes from 01/09/2008 to 31/08/2012

In addition, we implement a mean-difference test (parametric test) for two independent samples, that is to say, we compare whether the differences between the averaged ratios of the dynamic and constant models are significant or not. Then, if we had to invest some money in a risky portfolio, we would choose the one associated with Model A. However, according to the results obtained in the mentioned test, we find that there are few differences between the different selected models. In fact, if we prefer a less complex method (by calculating the conditional risk aversion attitude through different mathematical equations), we can select the constant risk aversion scheme. As we have mentioned before, this is because the differences between the best model (Model A, based on time-varying risk aversion) and the worst one (Model B, based on a constant risk attitude parameter) are not significant.

References

1. Arrow, K.: Essays in the Theory of Risk Bearing. Markham, New York (1971)
2. Mehra, R., Prescott, E.: The equity premium: a puzzle. J. Monet. Econ. **15**, 145–161 (1985)

3. Ghysels, E., Santa-Clara, P., Valkanov, R.: There is a risk-return trade off after all. J. Financ. Econ. **76**, 509–548 (2005)
4. Guo, H., Whitelaw, R.: Uncovering the risk–return relation in the stock market. J. Financ. **61**(3), 1433–1463 (2006)
5. Kim, K.H.: Counter-cyclical risk aversion. J. Empir. Financ. **29**, 384–401 (2014)
6. Dionne, G.: When can expected utility handle first-order risk aversion? J. Econ. Theory. **154**, 403–422 (2014)
7. Frankel, J.: In search of the exchange risk premium: a six currency test assuming mean-variance optimization. J. Int. Money Financ. **1**, 255–274 (1982)
8. Giovannini, A., Jorion, P.: The time-variation of risk and return in the foreign exchange and stock markets. J. Financ. **44**(2), 307–325 (1989)
9. Cotter, J., Hanly, J.: Time-varying risk aversion: an application to energy hedging. Energy Econ. **32**, 432–441 (2010)

Socially Responsible Investment, Should You Bother?

Antonio Díaz and Gloria Garrido

Abstract Companies included on a sustainability index meet several criteria based on an assessment of their economic, environmental and social practices. Each of these companies satisfies a different number of criteria and these standards can be quite different in quality and rigor. In this sense, RobecoSAM provides a Corporate Sustainability Assessment of the companies included in the Dow Jones Sustainable Index. The three proposed classes (Gold, Silver and Bronze) can be considered as social responsible (SR) ratings. Therefore, we examine the financial performance of portfolios composed of stocks according to these ratings. We assume that highly conscious SR investors could base their portfolio decision-making process on these SR ratings. From an extensive dataset, our results show that SR investments not only have no cost for investors but also outperform the market. Additionally, there are no significant differences among SR portfolios depending on the SR rating.

Keywords Socially responsible investment · Investment screening · Dow Jones sustainable index

1 Introduction

Socially Responsible Investment (henceforth SRI) is an increasing finance trend, covering a great part of the upcoming literature, but, indeed, it has shaped how finance has developed over the past years, and probably how it will continue to do so.

SRI consists of selecting or excluding through investment screens based on environmental, social, or corporate governance criteria. Furthermore, engagement with local communities and shareholder activism is considered as out-standing criteria [1]. There are different ways of selecting SRI from the whole bunch of

A. Díaz (✉) · G. Garrido
Department of Economics and Finance, Universidad de Castilla-La Mancha, Albacete, Spain
e-mail: Antonio.Diaz@uclm.es; GloriaPilar.Garrido@alu.uclm.es

M. Corazza et al. (eds.), *Mathematical and Statistical Methods for Actuarial Sciences and Finance*, https://doi.org/10.1007/978-3-319-89824-7_60

assets available in the market, investors can use negative, positive sustainability, or a combination between "sustainability" and shareholder activism [2].

The literature on the subject is focused on examining the performance of global SRI indices compared to conventional investments, the performance of companies grouped as homogenous because they are included in an ethical index or other SR benchmarks compared to conventional investments, and the performance of mutual funds or Ethical Investment Funds. Consensus has not been achieved.

The aim of this paper is to examine the role of corporate sustainable ratings as portfolio decision-making criterion for highly conscious SR investors. Some international investment companies have carried out sustainability assessments to build up SRI indices. This is the case of RobecoSAM, company which builds up the Dow Jones Sustainable Index (henceforth DJSI) and classifies SR companies in different levels or ratings based on the Corporate Sustainability Assessment (henceforth CSA). We assume that highly conscious SR investors could base their investment strategies on the three categories or CSA levels (Gold, Silver, and Bronze). In addition, we study is that SRI can be as good as conventional investments in terms of the financial performance, for which, an empirical analysis that measures financial performance of a different degree of sustainable companies versus non-sustainable companies has been carried out. This study contributes to the existing literature using a large sample period, reliable data to rank companies' socially responsible performance and a wider range of performance measures. From the empirical analysis, we can conclude that SRI not only are not a cost for investors but also they outperform the market, having better performance than traditional diversified portfolios.

2 Relationship Between Socially Responsible Performance and Financial Performance

A branch of the literature on SRI is focused on examining the performance of global SRI indices compared to conventional investments, the performance of companies grouped as homogenous because they are included in an ethical index or other SR benchmarks compared to conventional investments, and the performance of mutual funds compared to EIFs. Consensus has not been reached either in the data base used neither the methodology. While some use liquidity ratios, other use risk-adjusted performance measures.

This study goes beyond this literature by examining portfolios that contain individual companies, but while distinguishing among different degrees of corporate sustainability involvement. Some authors [3–6] examine the performance of the own DJSI or the alternative FTSE4Good Indices or use being included in those indices as homogeneous Corporate Sustainability Performance (CSP). Other recent papers analyse companies that follows or publicly maintain that they use SBP [7–9]. We try to improve the methodology previously used by considering different degrees of

CSR. We use the CSA, interpreted like a sort of CSR ratings, to examine separately the performance of portfolios of companies included in each category: Gold, Silver, and Bronze.

3 Portfolios Composition and Data Sample

Based on the yearly classification of the CSA provided by RobecoSAM, we look for companies also included in the Euro Stoxx 50, i.e., we match both datasets. We select manually companies that have been classified in one of the three main categories (Gold, Silver and Bronze) from The Sustainability Yearbook 2016, then the companies from the Eurozone need to be extracted from the list.[1]

The study has been carried out so that the investor can build up different value-weighted portfolios depending on its SRI demand, i.e., depending on the "socially responsible ratings". The portfolios would be Gold (3 stocks), Silver (9 stocks), Bronze (19 stocks), and Market portfolios. The last portfolio is the stock market portfolio Euro Stoxx 50, and is the benchmark or market portfolio, since it includes all the companies that are in each portfolio and the 50 leader companies of the euro zone.

The time horizon of this analysis is 15 years (January 2002–December 2016), calculating the weekly rates of return of each of the different portfolios. According to the volatility given by the VIX Index, the sample period has been divided into three different subperiods: pre-crisis and calm period (January 2002–June 2007), the financial crisis or market stress (includes the sovereign debt crisis in PIGS countries) (July 2007–June 2012), and bailout and economical overcome (July 2012–December 2016).

4 Empirical Analysis

In this section, we compute different risk-adjusted performance measures to our SRI portfolios constructed from three different levels of exigency of social responsibility: only Gold stocks, Gold and Silver stocks, and Gold, Silver, and Bronze stocks. From weekly discrete rates of returns, we compute basic statistical metrics and the performance measures to test the hypotheses previously exposed.

Both sets of performance measures show that the three sustainable portfolios outperform the market for the whole period (January 2002–December 2016). Summarizing results obtained by our analysis of the risk-adjusted performance

[1]The stock prices have been obtained from Yahoo Finance, the market capitalization of each company has been obtained from ElEconomista website. As a risk-free asset, we use the daily series of the 1-week Euribor obtained from the European Money Markets Institute (EMMI).

measures, the three portfolios could be sorted into a clear order according to the values of all the performance indices: first the Silver portfolio, second the Bronze, and third the Gold.

We now try to check if this apparent result is statistically significant for the full sample period and for each particular subsample. This analysis makes sense from a statistical point of view and thus portfolios can be compared between each other to state which one is superior to the other. Using 6-month non-overlapping windows during the period 2002–2016, we test the equality of mean, and two non-parametric tests, the sign test (binomial test) and the Wilcoxon signed-ranked test to test the null hypothesis of equality of median, and the Levene statistic.

This analysis shows that the only statistically significant difference between portfolios appear in the case of the Sharpe ratio. For the full sample period, the Gold portfolio systematically performs below the Silver and above the Bronze portfolios. For the rest of the analysed performance measures, their mean, median, and variance are statistically similar. No differences are observed from the performance analysis based on these measures of the three portfolios constructed with different CSA categories.

5 Conclusion

This study delivers a new point of view to the existing literature. We examine whether investing in SRI has a cost in terms of lower rates of return or underperformance. Literature has not come to an agreement about the relationship between being socially responsible and financial performance, this study aims to fill the gap. The empirical background of this study, as far as we know, is the first analysis within which different portfolios are constructed depending on a ranking that classify SRI in different levels. These "socially responsible ratings" consider the number and quality of the SBP that the companies meet. This fact added to the long sample period analysed, which covers large periods of calm and turmoil, and the reliable data used to measure the companies' SR performance, obtained from one of the most reliable databases in the world, gives a new outlook to the literature.

This study has been mainly developed with the aim of checking a relevant hypothesis: the performance of the portfolio is different depending on the "socially responsible rating" achieved. For which, it can be concluded that there are not significant differences between portfolios depending on their "socially responsible ratings", then SRI is not a cost for investors, even, they can outperform the market.

References

1. Renneboog, L., Ter Horst, J., Zhang, C.: Socially responsible investments: institutional aspects, performance, and investor behavior. J. Bank. Financ. **32**(9), 1723–1742 (2008)
2. Ward, S.: Socially Responsible Investment. Directory of Social Change Publication, London (1991)
3. Collison, D.J., Cobb, G., Power, D.M., Stevenson, L.A.: The financial performance of the FTSE4Good indices. Corp. Soc. Responsib. Environ. Manag. **15**(1), 14–28 (2008)
4. Oberndorfer, U., Schmidt, P., Wagner, M., Ziegler, A.: Does the stock market value the inclusion in a sustainability stock index? An event study analysis for German firms. J. Environ. Econ. Manag. **66**(3), 497–509 (2013)
5. Perez-Calderon, E., Milanes-Montero, P., Ortega-Rossell, F.J.: Environmental performance and firm value: evidence from Dow Jones Sustainability Index Europe. Int. J. Environ. Res. **6**(4), 1007–1014 (2012)
6. Ziegler, A.: Is it beneficial to be included in a sustainability stock index? A panel data study for European firms. Environ. Resour. Econ. **52**(3), 301–325 (2012)
7. Brammer, S., Brooks, C., Pavelin, S.: Corporate social performance and stock returns: UK evidence from disaggregate measures. Financ. Manag. **35**(3), 97–116 (2006)
8. Callan, S.J., Thomas, J.M.: Corporate financial performance and corporate social performance: an update and reinvestigation. Corp. Soc. Responsib. Environ. Manag. **16**(2), 61–78 (2009)
9. Godfrey, P.C., Merrill, C.B., Hansen, J.M.: The relationship between corporate social responsibility and shareholder value: an empirical test of the risk management hypothesis. Strateg. Manag. J. **30**(49), 425–445 (2009)

Measuring Financial Risk Co-movement in Commodity Markets

Gema Fernández-Avilés, Jose-María Montero, and Lidia Sanchis-Marco

Abstract Commodities play a more and more central role in financial markets. There are currently around fifty major commodity markets where more than a hundred hard and soft primary commodities are traded. Financialization has made purchasing index funds one of the most popular ways to invest on commodities. Consequently, understanding the dynamics of commodity indexes and whether or not they co-move becomes crucial for investors, especially in distress periods, where risk sharply increases. In this short paper, we analyze the downside risk co-movement of a number of primary commodity indexes in a distress period known as the 2007–2008 oil and food crisis. For this purpose, we use the expected shortfall and multidimensional scaling as a technique to produce low-dimensional financial risk maps.

Keywords Commodity market returns · Financial propagation · Financial maps · Multidimensional scaling

1 Introduction

There are currently about 50 major commodity markets where around one hundred of hard and soft commodities are traded. Today's commodity markets can be considered as mature and highly developed institutions, playing a very important role in the modern economy [11]. In this article we focus on one of the open questions in these markets: the co-movement in commodity prices, and more specifically the co-movement in commodity indexes. However, we do not focus on co-movement in mean and/or volatility, as usual in the literature on the topic. We go beyond this point and deal with co-movement in tail or downside risk, which is more interesting for agents participating in commodity markets. Co-movement (also spill-

G. Fernández-Avilés · J.-M. Montero · L. Sanchis-Marco (✉)
University of Castilla la Mancha, Toledo, Spain
e-mail: gema.FAviles@uclm.es; Jose.MLorenzo@uclm.es; lidia.sanchis@uclm.es

M. Corazza et al. (eds.), *Mathematical and Statistical Methods for Actuarial Sciences and Finance*, https://doi.org/10.1007/978-3-319-89824-7_61

over processes and contagion) has been usually studied for returns and, less usually, for volatility. However, there exist only a few number of research studies on co-movement in the tail of the returns distribution [1, 9]. We contribute to this stream of literature by analyzing the tail risk co-movement in commodity markets. To this purpose, we first estimate the expected shortfall (ES) as a downside risk measure. Then, ES is used as the input of a multidimensional scaling (MDS) procedure to constructing financial downside risk maps. One of the major advantages of this ES-MDS combined procedure is that it provides the representation of the objects under study (in our case ES commodity time series) as points in a map, so that highly correlated objects will be close each other in such a map and objects with a low correlation will be represented with very distant points in the map. To the best of our knowledge, this is the first study attempting to create financial risk maps using MDS. It is of note that this is a very easy-to-interpret technique whose results can serve as an input for further spatial (or spatio-temporal) statistical analysis of tail risk (geostatistics, spatial econometrics, local modelling, point patterns analysis, etc.). In this short article, we deal with nine major world commodity indexes from the sectors: energy, metals minerals, Beverages, Fats and Oils, Fertilizers, Grains, Food, Row materials, and Timber.[1]

The remainder of the paper is organized as follows. Section 2 addresses methodological questions. Specifically, Sect. 2.1 states the downside risk measure considered and Sect. 2.2 describes our approach to creating financial downside risk maps using MDS. Section 3 presents the main results obtained. Finally, Sect. 4 concludes.

2 Methodology

2.1 Selecting a Downside Risk Measure

In order to create financial downside risk maps for commodity markets, we first have to choose an optimal measure of downside risk. We use ES because it is a coherent risk measure and takes into account extreme negative returns. Moreover, ES is comonotonic additive, robust and elicitable (see [4]). In addition, the Basel Committee [3] has also confirmed that ES will replace Value at Risk (VaR) for regulatory capital purposes in the trading book. ES was introduced by Artzner et al. [2]. We use the parametric GJR-GARCH(p,r,q) with t-student distribution in the updating process of ES volatility estimates. Technical details can be seen in [10].

[1]The indices composition is detailed in http://pubdocs.worldbank.org/en/752911517610537957/CMO-Pink-Sheet-February-2018.pdf.

2.2 *Constructing Financial Downside Risk Maps with MDS*

Financial downside risk maps are constructed applying MDS to the ES series of the commodity indexes. The MDS approach has the advantage of reproducing the main features of the data in the form of maps that not only lead themselves to intuitive interpretation but also provide the set of financial risk distances. The main aim of MDS is to discover structures in multidimensional data. Based on a proximity matrix, typically derived from variables measured on objects as input entity, these dissimilarities are mapped on a low-dimensional spatial representation [7]. More in detail, given a matrix of measured or perceived similarities among various items (in our study, ES commodity indexes time series), MDS plots the items on a map (a financial risk map) such that those which are perceived to be similar are placed near one another. Technical details can be seen in [5, 6] and [8].

3 Main Results

Figure 1 shows, in Panel (a), the financial downside risk map of the nine commodity indexes considered in the analysis during the global oil and food crisis in 2007–2008. As can be seen, Energy is the most risky commodity index and do not co-move with any other index; it exhibits a specific dynamics. The second most risky index is Fat and oils, related with oil crisis as soybean was used as biofuel. It is also far from co-moving with any other index. Beverage and Timber exhibit a certain co-movement. Metals and minerals, Raw materials, Grains, Food and Fertilizers, the least risky indexes, show the highest degree of co-movement. Panel (b) shows the Shepard plot, which depicts a good distribution of points around the 45°-line. Moreover, the Kruskals stress-1 indicates a god fit (0.0092).

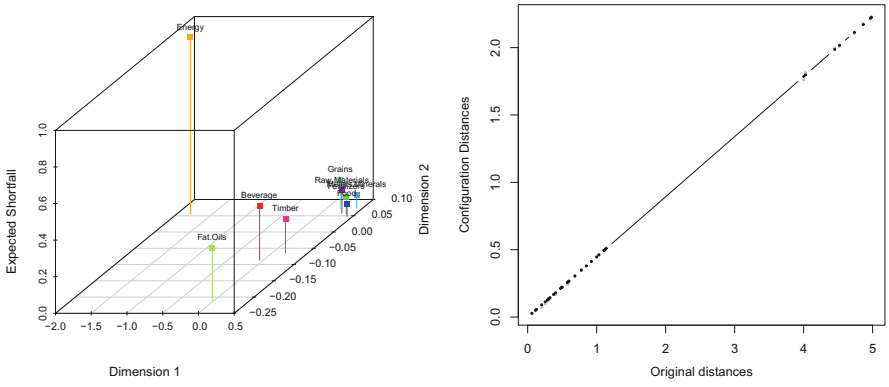


Fig. 1 Financial downside risk map (2007–2008 oil and food crisis)

4 Conclusions

In this short paper we study whether or not the downside risk of a number of major commodity indexes co-move. We use a combined expected shortfall/multidimensional scaling approach which provides intuitive and easy-to-interpret low dimensional downside risk maps. During the 2007–2008 global oil and food crisis, the distress period under study, Metals and minerals, Raw materials, Grains, Food, and Fertilizers, the least risky indexes, exhibit a high degree of downside risk co-movement, whereas Energy, the riskiest index, shows a particular dynamics that has nothing to do with the dynamics exhibited by the rest of indexes. Fat and oils, the second most risky index, is also far from co-moving with the other indexes, and Beverage and Timber exhibit a certain co-movement, but not of the intensity of the co-movement in Metals and minerals, Raw materials, Grains, Food, and Fertilizers.

References

1. Algieri, B., Leccadito, A.: Assessing contagion risk from energy and non-energy commodity markets. Energy Econ. **62**, 312–322 (2017)
2. Artzner, P., Delbaen, F., Eber, J.M., Heath, D.: Coherent measures of risk. Math. Financ. **9**(3), 203–228 (1999)
3. Basel Committee on Banking Supervision (BCBS): Consultative Document, Fundamental review of the trading book: a revised market risk framework, BIS, Basel. http://www.bis.org/publ/bcbs265.pdf (2013)
4. Emmer, S., Kratz, M.,Tasche, D.: What is the best risk measure in practice? A comparison of standard measures. J. Risk. **18**(2), 31–60 (2015)
5. Jobson, J.D.: Applied Multivariate Analysis. Volume II: Categorical and Multivariate Methods. Springer, New York (1992)
6. Kruskal, J., Wish, M.: Multidimensional Scaling. Sage Publications, Newbury Park, CA (1978)
7. Mair, P., Leeuw, J., Groenen, P.J.F.: Multidimensional scaling in R: SMACOF. Working Paper, UCLA (2015)
8. Peña, D.: Análisis de Datos Multivariantes. McGraw-Hill, Madrid (2002)
9. Pierret, D.: The systemic risk of energy markets. https://ssrn.com/abstract=2245811 or http://dx.doi.org/10.2139/ssrn.2245811 (2013)
10. Righi, M.B., Ceretta, P.S.: A comparison of expected shortfall estimation methods. J. Econ. Bus. **78**, 14–47 (2015)
11. Sieczka, P., Holyst, J.A.: Correlations in commodity markets. Phys. A **388**(8), 1621–1630 (2009)

Helping Long Term Care Coverage via Differential on Mortality?

María Cristina Fernández-Ramos, Joseba Iñaki De La Peña, Ana Teresa Herrera, Iván Iturricastillo, and Noemí Peña-Miguel

Abstract This paper seeks to help draw up a flexible design for pensions for dependents that can help reduce the costs of their situation while precisely increasing the amounts that they receive. The way is a system for the automatic adjustment of pension benefits taking into account the dependency level of the beneficiary. Thus, pension benefits increase in the new state as the cost of care increases. To that end we propose a model with a benefit correction factor that includes a specific mortality rate for dependents, thus enabling us to adapt benefits to the profile of each beneficiary. Special attention is paid to mortality rates among dependents as the determinant for the correction factor. This new model has many practical implications, as it can be implemented without much difficulty and indeed at no additional cost. This enables coverage to be universal in private capitalization-type pension plans. However, it does increase the cost of social security systems funded on a pay-as-you-go basis.

Keywords Elderly · Pension evaluation · Sustainability factor · Pension schemes

1 Long Term Care as a Coverage

In the field of private insurance, a distinction is drawn between natural coverage and long-term care [1, 2], with problems of dependency being alleviated with products suited to demand. The combination of different benefits [3] simplifies matters and

M. C. Fernández-Ramos
School of Education, Junta de Castilla y León, Valladolid, Spain

J. I. De La Peña (✉) · A. T. Herrera · I. Iturricastillo · N. Peña-Miguel
University of the Basque Country (UPV/EHU), Vizcaya, Spain
e-mail: jinaki.delapena@ehu.es

M. Corazza et al. (eds.), *Mathematical and Statistical Methods for Actuarial Sciences and Finance*, https://doi.org/10.1007/978-3-319-89824-7_62

includes an important aspect of retirement pensions, which is usually dealt, with separately: an acknowledgement of the potential need for dependency care, resulting in higher benefits being paid when the beneficiary is dependent. This approach is proposed by [4] and [5] as a combination of retirement income and higher income on becoming dependent.

2 Differential on Mortality

We assume that when the beneficiary becomes a high-level dependent at age x the amount of the benefit is automatically increased by a percentage λ_x^d, which helps to pay for dependency care services. This factor λ_x^d is applied when a beneficiary becomes a dependent then only the probability of death while classed as dependent remains to be determined. The probabilities of suffering from high-level dependency have been determined in various studies [6], on the basis of which life expectancy figures for individuals in the severest states of dependency have been calculated.

On that basis, [7] determine the probabilities of death among high-level in Spain. They find that the gap between excess mortality and general mortality rates decreases from age 96 onwards. To reflect this effect they include a mixed correction factor: an additive modification, and a multiplicative correction:

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

δ: Maximum value to be incorporated in line with the age at which figures converge asymptotically.
γ: Slope factor.
x_i: Age at the point of inflection where the curve changes from convex to concave.
β: Multiplicative correction factor applied to general mortality.

Once the probability of death of severe and high-level dependents is known, the correction factor to be applied is the following:

$$ \lambda_x^d = \frac{\sum_{h=x}^{w} {}_{h-x}p_x^m}{\sum_{h=x}^{w} {}_{h-x}{}^{d}p_x^m} = \frac{e_x^m}{{}^{d}e_x^m} $$

$\sum_{h=x}^{w} {}_{h-x}p_x^m$: Sum of probabilities of being alive from age x to h years more.
$\sum_{h=x}^{w} {}_{h-x}{}^{d}p_x^m$: Sum of probabilities of a dependent to be alive from age x to $x+h$.

3 An Application to Spain

Based on PERM/F-2000P dynamic tables for Spain fitted to HID 98-01 statistics for France with the values obtained by [7] for δ, γ, β & x_i with an ordinary least squares procedure for the gross values for high-level dependency estimated for Spain (Table 1):

The mortality rates obtained for high-level dependents are markedly higher than general mortality rates from age 35 onwards (Figs. 1, 2 and 3).

The application of these calculations to high-level dependents in line with their year of birth shows pension increases of practically threefold in all cases. At younger ages the correction factor has values of just over one, in sharp contrast with the values found from retirement age onwards.

Table 1 Excess mortality factors for dependents

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

Source: [7]

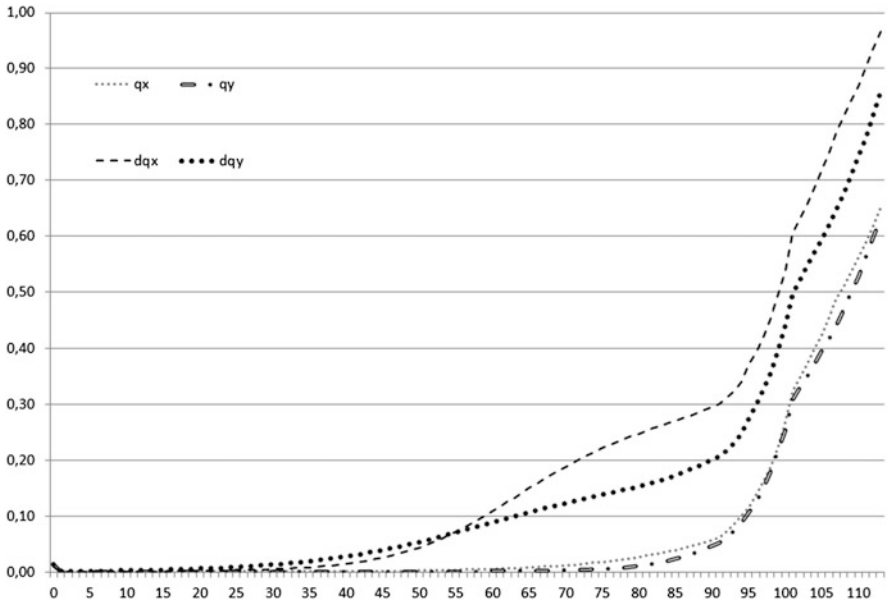


Fig. 1 Mortality among dependents and general mortality rates per age and gender

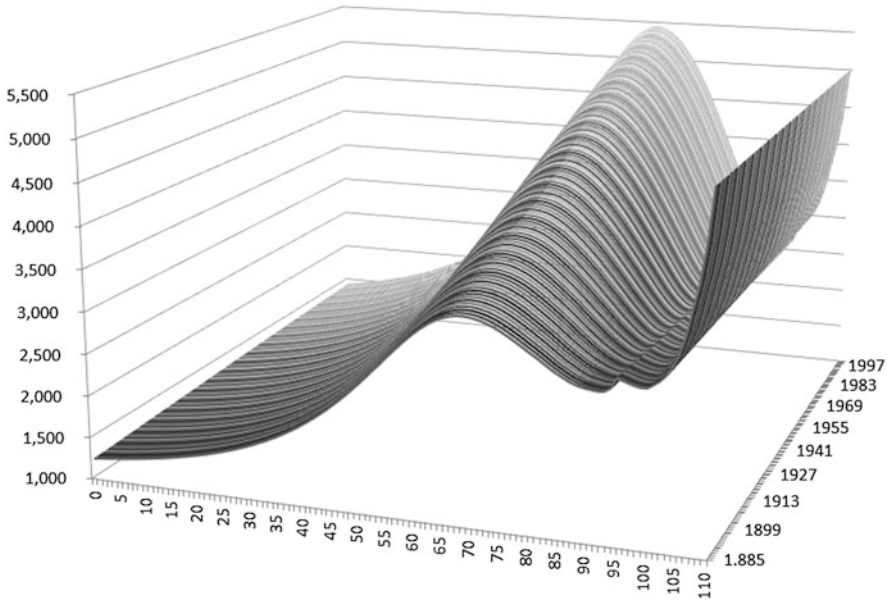


Fig. 2 Correction factor with actuarial income for dependency broken down by age and generation (men). Source: Own work

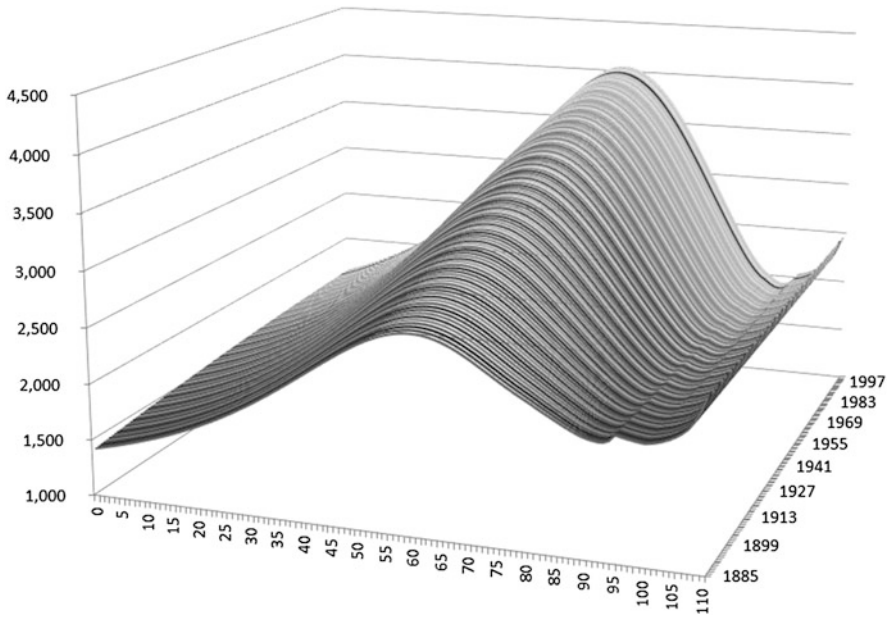


Fig. 3 Correction factor with actuarial income for dependency broken down by age and generation (women). Source: Own work

4 Conclusions

The model has many practical implications, and can be implemented with little difficulty and no additional cost, in capitalization-based private pension schemes. However, if this factor were to be included in a public, defined-benefit system such as the pay-as-you-go social security system it would lead to a direct increase in cost equivalent to the amount of the increase in benefits. Contributions would not increase, so initially a deficit would result.

Finally, public and private dependency coverage schemes alike seek to help meet the costs that dependency entails for individuals, but without necessarily providing all the resources needed to meet demands for coverage. Individuals are provided with a set of measures that can meet their needs as dependents in full: services, use of residence and financial benefits, thus providing higher levels of satisfaction and better monitoring of dependents.

Acknowledgements This study was supported by Polibienestar Research Institute and Consolidated Research Group EJ/GV: IT 897-16.

References

1. Brown, J., Warshawsky, M.: The life care annuity: a new empirical examination of an insurance innovation that addresses problems in the markets for life annuities & long-term care insurance. J. Risk Insur. **80**(3), 677–703 (2013)
2. Davidoff, T.: Housing, health, & annuities. J. Risk Insur. **76**(1), 31–52 (2009)
3. Spillman, B.C., Murtaugh, C., Warshawsky, M.J.: Policy implications of an annuity approach to integrating long-term care financing & retirement income. J. Aging Health. **15**(1), 45–73 (2003)
4. Haberman, S., Pitacco, E.: Actuarial Models for Disability Insurance. Chapman & Hall, London (1999)
5. Pitacco, E.: Biometric risk transfers in life annuities & pension products: a survey. CEPAR working paper 2013/25 (2013)
6. Fernández-Ramos, M.C.: Soluciones pragmáticas en el campo privado para la cobertura de la Dependencia en España. Doctoral Thesis, Universidad del País Vasco-Euskal Herriko Unibertsitatea, Bilbao (2015)
7. Sánchez, E., López, J.M., de Paz, S.: La corrección de los tantos de mortalidad de los dependientes: una aplicación al caso español. An. Inst. Actuar. Esp. **13**, 135–151 (2008)

Tuning a Deep Learning Network for Solvency II: Preliminary Results

Ugo Fiore, Zelda Marino, Luca Passalacqua, Francesca Perla, Salvatore Scognamiglio, and Paolo Zanetti

Abstract Under the Solvency II Directive, insurance and reinsurance undertakings are required to perform continuous monitoring of risks and market consistent valuation of assets and liabilities. Solvency II application is particularly demanding, both theoretically and under the computational point of view. At present, any technique able to improve on accuracy or to reduce computing time is highly desirable. This works reports initial results on the design of a Deep Learning Network, aimed to reduce computing time by avoiding the *standard* full nested Monte Carlo approach.

Keywords Solvency II · Deep learning · Monte Carlo · Profit insurance policies

1 Introduction

The European Union 2009/138 Directive, better known as Solvency II [1], requires insurance and reinsurance undertakings to hold eligible own funds covering the Solvency Capital Requirement (SCR), defined as the Value-at-Risk of the basic own funds subject to a confidence level of 99.5% over a one-year period. The SCR shall be calculated, either in accordance with the so-called standard formula—as detailed in the Directive—or using an internal model (IM), subject to approval by supervisory authorities. Within the IM approach, the SCR shall be derived from the probability distribution forecast (PDF) of the net asset value (NAV), i.e. of the difference between the value of the assets and that of the liabilities, generated by the internal model itself. In practice, internal models are multi-dimensional stochastic models that allow the evaluation of asset and liabilities at the end of the one-

U. Fiore · Z. Marino · F. Perla · S. Scognamiglio · P. Zanetti (✉)
Department of Management and Quantitative Studies, Parthenope University, Napoli, Italy
e-mail: salvatore.scognamiglio@uniparthenope.it; paolo.zanetti@uniparthenope.it

L. Passalacqua
Department of Statistical Sciences, Sapienza University, Rome, Italy

M. Corazza et al. (eds.), *Mathematical and Statistical Methods for Actuarial Sciences and Finance*, https://doi.org/10.1007/978-3-319-89824-7_63

year horizon. From the algorithmic point of view, they are based on a two-step procedure. The first step is the generation of the value of the relevant risk drivers (e.g. interest rates and mortality rates) over the one-year time horizon; this is usually performed by Monte Carlo simulation over a fixed time grid, using a discretised version of the stochastic differential equations describing the dynamics of the risk factors. The second step is the evaluation of asset and liabilities, conditional of the trajectory followed by the risk drivers. Depending on the complexity of the contracts involved, the evaluation may be performed by mean of theoretically derived closed form expressions or using a numerical procedure (unavoidable for contracts with embedded American style options, whose evaluation would require the solution of free-boundary problems). Thus, when the chosen numerical procedure is Monte Carlo, the evaluation of SCR involves Nested Monte Carlo (NMC) simulations. Since, to achieve the desired accuracy, the number of simulations of both steps must be large, the computational cost is very high. To avoid simulation in the second step, an *effective* closed form expression must be found, using a reduced size NMC for calibration.

At present, different techniques have been proposed in the literature [3, 6, 7], out of which *Least-squares Monte Carlo* based on orthogonal polynomial expansion seems the most promising; however the question is still debated [2]. *Machine learning* (ML) techniques appear to be viable alternatives, since the dependence (*target function*) of the value of asset and liabilities on the risk drivers is unknown, and *supervised learning* can be carried out on *training sets* produced by NMC.

Among ML techniques, Deep Learning Networks (DLNs) are a powerful and flexible technique which has proven its effectiveness in several research areas. A DLN is a multilayer Artificial Neural Network. The units of a neural network are associated to a nonlinear *activation function* applied on a linear combination of the inputs to produce the output. Learning refers to updating the parameters of such linear combinations so that a given loss function is minimized. A particularly relevant property of neural networks is that they are known to be universal approximators. A single-layer feed-forward neural network can approximate any continuous function. In DLNs, multiple layers are composed together, so that they can efficiently approximate the value of assets and liabilities as function of risk drivers. The first layer is associated with input, while the units in each subsequent layer act on a combination of the outputs of all units in the preceding layer. The resulting (deep) chain of transformations can model complex structure in its inputs more compactly than a shallow network can do. The expressive power of DLNs is counterbalanced by the extensive tuning needed to achieve an acceptable performance. Hyperparameters to be tuned include the number of layers, number of units in each layer, their activation function, in addition to other characteristics that influence the behavior of the learning algorithm.

In this work, initial results are presented on the tuning of the hyperparameters of a DLN to compute the value of the liabilities for *profit sharing life insurance policies* with minimum guarantees. In these contracts—widely diffused in Italy—benefits credited to the policyholder are indexed to the annual return of a specified

investment portfolio, called the *segregated fund*. A profit sharing policy is a "complex" structured contract, with the segregated fund return as underlying.[1]

2 Numerical Set-Up

As in the literature there is no unanimous point of view on the best combination of hyperparameters, fine tuning of a DLN is still a demanding task. As expected, the DLN performance was found to depend significantly on the hyperparameters. The analysis considered the segregated fund of an Italian insurance company with seven risk drivers, including interest rate, equity, inflation, and exchange rate risk. The NMC sample produced by the IM is composed of 10,000 simulations, out of which 6667 used for training, while the remaining 3333 were reserved for validation. The effect of the following hyperparameters was analyzed: number of hidden layers, activation function, learning rate (how quickly a network updates its weights to adapt to training data), and number of nodes in each layer. To compare different configurations, the performance metric used is the Normalized RMSE (NRMSE), i.e., the RMSE divided by the range of the response variable. To understand the role of each hyperparameter in the DLN performance, one hyperparameter at a time was changed. Better performances were observed when input data were normalised. Standardization and rescaling were also tested, with rescaling achieving the best performance. All tests were performed with the MXNet package[2] of R.

3 Results

DLNs with one, two, and three hidden layers were tested, concluding that a DLN with two hidden layers offers the best performance. As per the number of nodes in the hidden layers, different number of units in the two hidden layers and different learning rates were tested. In most cases, performance improvement was observed with learning rates around 0.2. Determining an adequate number of hidden layer units has always been a complex question in neural networks. Empirical rules that have been proposed in the literature are: 2/3 of the nodes in the input layer [5]; twice the number of nodes in the input layer [9]; a number between the input and the output layer nodes [4]; $\log_2 S$, where S is the number of training samples [10]. The performance of the DLN was tested following the previous rules and with all the numbers from 2 to twice the number of input layer units. In Fig. 1 (left), the NRMSE is reported for the analyzed combinations of the number of units in the

[1] For an exhaustive analysis of the basic principles and methodological approach for a valuation system of profit sharing policies with minimum guarantees the reader may refer to [8].

[2] https://mxnet.apache.org/.

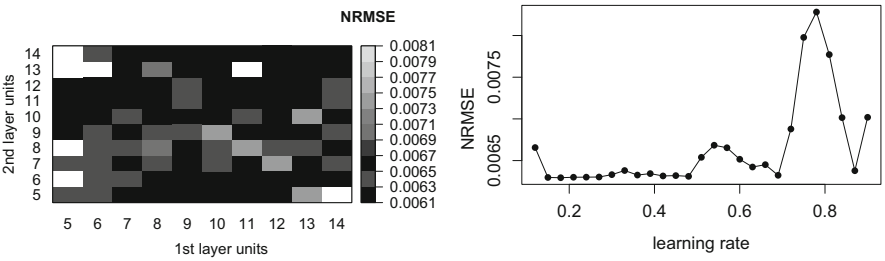


Fig. 1 Performance with (left) various numbers of units per layer and (right) for different learning rates

two hidden layers, with the learning rate fixed at 0.2 (white boxes indicate outliers). It can be seen that the performance is almost stable on good values for numerous combinations (with the top-right quadrant seemingly offering the best scenario), but strong and sudden variations are visible. The absolute minimum (though by a very slight margin) was observed, however, for nine nodes in the first layer and eight in the second. This evidence shows that none of the previous rules-of-thumb is always true. Figure 1 (right) shows in detail the DLN performance as the learning rate changes, when the number of hidden units is kept fixed at the best configuration as specified above. A learning rate equal to 0.18 is the best value. The last hyperparameter considered is the activation function of nodes. For all the configurations that were previously evaluated, the best performance was obtained with the Rectified Linear Unit (ReLU) activation function. Plans for further study include the refinement of the analysis and the realization of large-scale tests on an High Performance Computing system.

Acknowledgements This work was partially supported by the Research grant of Università di Napoli Parthenope, D.R. no. 953, November 28th, 2016.

References

1. Directive 2009/138/EC of the European Parliament and of the Council of 25 November 2009 on the taking-up and pursuit of the business of insurance and reinsurance (Solvency II). Off. J. Eur. Union (2009)
2. Bauer, D., Bergmann, D., Kiesel, R.: On the risk-neutral valuation of life insurance contracts with numerical methods in view. ASTIN Bull. J. IAA **40**(1), 65–95 (2010)
3. Bauer, D., Reuss, A., Singer, D.: On the calculation of the solvency capital requirement based on nested simulations. ASTIN Bull. J. IAA **42**(2), 453–499 (2012)
4. Blum, A.: Neural Networks in C++: An Object-Oriented Framework for Building Connectionist Systems. Wiley, New York (1992)
5. Boger, Z., Guterman, H.: Knowledge extraction from artificial neural network models. In: Systems, Man, and Cybernetics, vol. 4. IEEE International Conference on Computational Cybernetics and Simulation, 1997, pp. 3030–3035. IEEE, Piscataway (1997)

6. Casarano, G., Castellani, G., Passalacqua, L., Perla, F., Zanetti, P.: Relevant applications of Monte Carlo simulation in Solvency II. Soft. Comput. **21**(5), 1181–1192 (2017)
7. Castellani, G., Passalacqua, L., et al.: Applications of distributed and parallel computing in the Solvency II framework: the DISAR system. In: Euro-Par Workshops, pp. 413–421. Springer, Berlin (2010)
8. De Felice, M., Moriconi, F.: Market based tools for managing the life insurance company. ASTIN Bull. J. IAA **35**(1), 79–111 (2005)
9. Linoff, G.S., Berry, M.J.: Data Mining Techniques: For Marketing, Sales, and Customer Relationship Management. Wiley, New York (2011)
10. Wanas, N., Auda, G., Kamel, M.S., Karray, F.: On the optimal number of hidden nodes in a neural network. In: IEEE Canadian Conference on Electrical and Computer Engineering, 1998, vol. 2, pp. 918–921. IEEE, Piscataway (1998)

Exploratory Projection Pursuit for Multivariate Financial Data

Cinzia Franceschini

Abstract Projection pursuit is a multivariate statistical technique aimed at finding interesting low-dimensional data projections. It deals with three major challenges of multivariate analysis: the curse of dimensionality, the presence of irrelevant features and the limitations of visual perception. In particular, kurtosis-based projection pursuit looks for interesting data features by means of data projections with either minimal or maximal kurtosis. Its applications include independent component analysis, cluster analysis, discriminant analysis, multivariate normality testing and outliers detection. To the best of the author's knowledge, this paper constitutes the first application of kurtosis-based projection pursuit to the exploratory analysis of multivariate financial time series.

Keywords Projection pursuit · Kurtosis · Financial data

1 Introduction

According to [6], "Projection pursuit is a technique for locating projections from high to low-dimensional space that reveal interesting non-linear features of a data set, such as clustering and outliers. The two key components of projection pursuit are the chosen measure of interesting features (the projection index) and its algorithm."

Projection pursuit deals with three major challenges of multivariate analysis: the curse of dimensionality, the presence of irrelevant features, and the limitations of visual perception. Projection pursuit is particularly useful when data are high-dimensional, data features are unclear, and the approach is exploratory.

The absolute value of the fourth standardized cumulant is a valid projection pursuit index, according to the criteria stated in [5], and leads to kurtosis-based

C. Franceschini (✉)
Università degli Studi di Bologna, Vicepresidenza della Scuola di Economia, Management e Statistica - Bologna, Bologna, BO, Italy
e-mail: cinzia.franceschini2@unibo.it

M. Corazza et al. (eds.), *Mathematical and Statistical Methods for Actuarial Sciences and Finance*, https://doi.org/10.1007/978-3-319-89824-7_64

projection pursuit. Its statistical applications include multivariate normality testing [8, 12, 14], cluster and outlier detection [15–17], independent component analysis [7, 13]. The prominent role of kurtosis as a projection pursuit index is emphasized by several authors, as for example [1, 3, 4, 19, 20]. To the best of the author's knowledge, this paper constitutes the first application of kurtosis-based projection pursuit to multivariate financial time series.

2 Results

This section examines a large dataset in order to illustrate the potential of the proposed method for visualizing clusters. Each observation is the closing price of an European financial market, as recorded by MSCI Barra. The included countries are Austria, Belgium, Denmark, Finland, France, Germany, Greece, England, Ireland, Italy, Norway, Holland, Portugal, Spain, Sweden, Switzerland. The first and last closing prices were recorded during 24/06/2003 and 23/06/2008, respectively. Data are arranged in a matrix where each row represents a day and each column a country. Hence the size of the data matrix is 1305×16.

All variables are platykurtic: their kurtosis, as measured by the fourth standardized moment, is never greater than 2.3 and it exceeds 2 for Finland and England only. As remarked by several authors [17, 18] kurtosis is likely to be small in the presence of two well-separated clusters with similar size and small variances. The same happens when three clusters are presents and the clusters in the extremes are of similar size. However, visual inspection of the histograms is not conclusive with respect to the cluster structure. Some of them (as Finland, England, and Spain) suggest the presence of two clusters, while three or more clusters appear to be present in others (like Austria, Belgium, and Denmark).

More insight into the cluster structure of the data might be gained by the proposed method. We shall denote by MIN and MAX the projections of the data onto the directions approximating those with minimal and maximal kurtosis, respectively. Both projections were computed with the algorithm whose starting value and iteration step are described in [2] and [10], respectively. The histogram of MIN (Fig. 1) is markedly bimodal, thus suggesting the presence of two clusters. On the other hand, the histogram of MAX (Fig. 2) has a very long left tail, often associated with the presence of outliers. The apparent contradiction disappears when plotting MIN against MAX. The resulting scatterplot (Fig. 3) clearly reveals the presence of three clusters forming a tree-like structure. More precisely, the upper end of the cluster which is more dispersed along the vertical axis overlaps with the lower end of the cluster of a more spherical shape, which in turn is very close to the left side of the remaining cluster. The example encourages the joint use of data projections which either minimize or maximize kurtosis, as proposed by Peña and Prieto [16, 17]. Further research in this area might consider projections maximizing skewness, as in [9, 11].

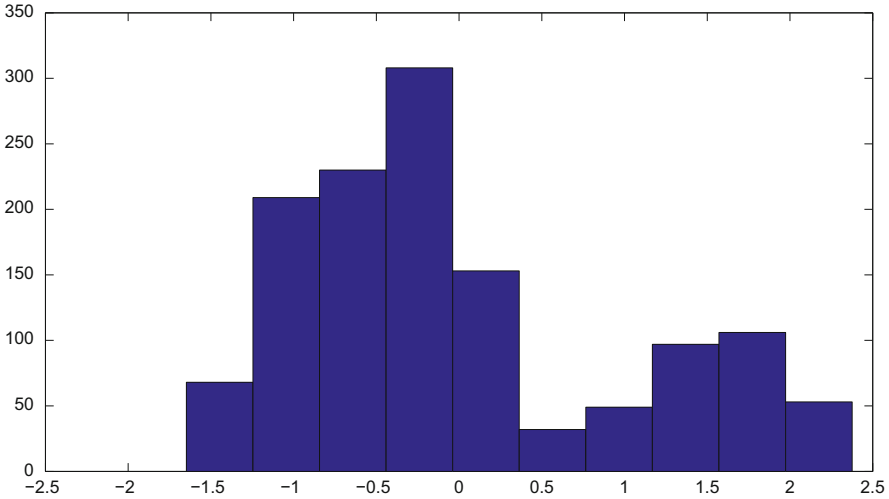


Fig. 1 Histogram of the most platykurtic projection

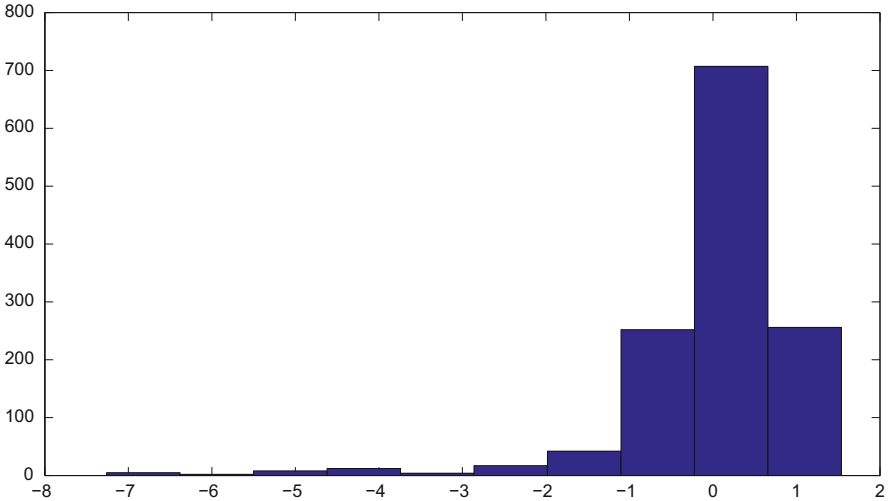


Fig. 2 Histogram of the most leptokurtic projection

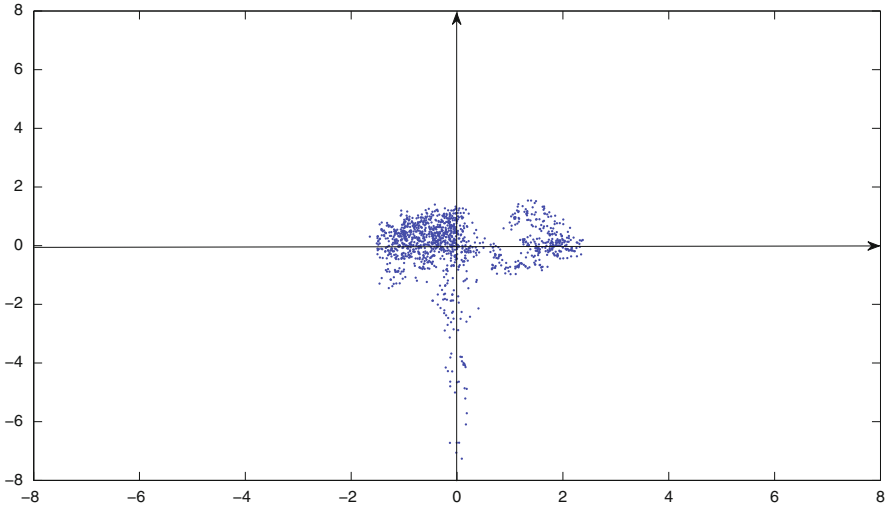


Fig. 3 Scatterplot of the most platykurtic (horizontal axis) and leptokurtic (vertical axis) projection

References

1. Daszykowski, M.: From projection pursuit to other unsupervised chemometric techniques. J. Chemom. **21**, 270–279 (2007)
2. Franceschini, C., Loperfido, N.: An algorithm for finding projections with extreme kurtosis. In: Perna, C., Pratesi, M., Ruiz-Gazen, A. (eds.) Studies in Theoretical and Applied Statistics: SIS2016-48th Meeting of the Italian Statistical Society, Salerno 8-10 June 2016. Springer, Cham (2018)
3. Hou, S., Wentzell, P.D.: Fast and simple methods for the optimization of kurtosis used as a projection pursuit index. Anal. Chim. Acta **704**, 1–15 (2011)
4. Huang, Y., Cheng, C.R., Wang, T.H.: Influence analysis of nongaussianity by applying projection pursuit. Stat. Probab. Lett. **77**, 1515–1521 (2007)
5. Huber, P.: Projection pursuit (with discussion). Ann. Stat. **13**, 435–475 (1985)
6. Hui, G., Lindsay, B.G.: Projection pursuit via white noise matrices. Sankhya B **72**, 123–153 (2010)
7. Hyvarinen, A., Karhunen, J., Oja, E.: Independent Component Analysis. Wiley, New York (2001)
8. Kuriki, S., Takemura, A.: The tube method for the moment index in projection pursuit. J. Stat. Plann. Inference **138**, 2749–2762 (2010)
9. Loperfido, N.: Skewness and the linear discriminant function. Stat. Probab. Lett. **83**, 93–99 (2013)
10. Loperfido, N.: A new kurtosis matrix, with statistical applications. Linear Algebra Appl. **512**, 1–17 (2017)
11. Loperfido, N.: Skewness-based projection pursuit: a computational approach. Comput. Stat. Data Anal. **120**, 42–57 (2018)
12. Malkovich, J.F., Afifi, A.A.: On tests for multivariate normality. J. Am. Stat. Assoc. **68**, 176–179 (1973)
13. Miettinen, J., Taskinen, S., Nordhausen, K., Oja, H.: Fourth moments and independent component analysis. Stat. Sci. **30**, 372–390 (2015)

14. Naito, K.: A generalized projection pursuit procedure and its significance level. Hiroshima Math. J. **27**, 513–554 (1997)
15. Peña, D., Prieto, F.J.: The kurtosis coefficient and the linear discriminant function. Stat. Probab. Lett. **49**, 257–261 (2000)
16. Peña, D., Prieto, F.J.: Cluster identification using projections. J. Am. Stat. Assoc. **96**, 1433–1445 (2001)
17. Peña, D., Prieto, F.J.: Multivariate outlier detection and robust covariance estimation (with discussion). Technometrics **43**, 286–310 (2001)
18. Peña, D., Prieto, F.J.: Combining random and specific directions for outlier detection and robust estimation of high-dimensional multivariate data. J. Comput. Graph. Stat. **16**, 228–254 (2007)
19. Rubinshtein, E., Anuj Srivastava, A.: Optimal linear projections for enhancing desired data statistics. Stat. Comput. **20**, 267–282 (2010)
20. Ruiz-Gazen, A., Marie-Sainte, S.L., Berro, A.: Detecting multivariate outliers using projection pursuit with particle swarm optimization. In: Proceedings of COMPSTAT 2010, pp. 89–98 (2010)

The Rearrangement Algorithm of Puccetti and Rüschendorf: Proving the Convergence

Marcello Galeotti, Giovanni Rabitti, and Emanuele Vannucci

Abstract In 2012 Puccetti and Rüschendorf [J. Comp. Appl. Math., 236 (2012)] proposed a new algorithm to compute the *upper* Value-at-Risk (VaR), at a given level of confidence, of a portfolio of risky positions, whose mutual dependence is unknown. The algorithm was called Rearrangement, as it consists precisely in rearranging the columns of a matrix, whose entries are quantiles of the marginal distributions. In the following years the algorithm has performed quite well in several practical situations, but the convergence has remained an open problem. In the present paper we show that the rearrangement algorithm converges, once the deterministic procedure has been precisely defined and an initial *optimality* condition is satisfied.

Keywords Value-at-Risk · Rearrangement · Ordered matrices

1 Introduction

The Rearrangement Algorithm comes from two apparently distant problems:

1. a numerical problem (see, e.g., [2]): given an $n \times d$ matrix with positive entries, rearrange its columns as to minimize the variance of the row-sum vector;

M. Galeotti (✉)
Department of Statistics, Informatics, Applications, University of Florence, Florence, Italy
e-mail: marcello.galeotti@unifi.it

G. Rabitti
Department of Decision Sciences, Bocconi University, Milan, Italy
e-mail: giovanni.rabitti@unibocconi.it

E. Vannucci
Department of Economics and Management, University of Pisa, Pisa, Italy
e-mail: emanuele.vannucci@unipi.it

M. Corazza et al. (eds.), *Mathematical and Statistical Methods for Actuarial Sciences and Finance*, https://doi.org/10.1007/978-3-319-89824-7_65

2. a financial and probabilistic problem: let $\mathscr{L} = (L_1, \ldots, L_d)$ be a portfolio of risky positions ($L_1, \ldots, L_d$ are losses), whose marginal distributions are known, but their dependence is not. Given a confidence level α, $0 < \alpha < 1$, find the highest $VaR_\alpha(L)$ of the total loss $L = L_1 + \ldots + L_d$.

In order to see the connection between the two problems, consider the partition of the interval $[\alpha, 1]$ into n equal subintervals and, calling $F_j(x)$ the marginal of the loss L_j, pose, for $i = 1, \ldots, n$,

$$a_{ij} = F_j^{-1}\left(\alpha + \frac{(i-1)(1-\alpha)}{n}\right)$$

and

$$b_{ij} = F_j^{-1}\left(\alpha + \frac{i(1-\alpha)}{n}\right).$$

Hence, we construct two $n \times d$ matrices A and B. Since the VaR is not sub-additive, the highest $VaR_\alpha(L)$, say $\overline{VaR_\alpha(L)}$, is generally higher than $\sum_{j=1}^{d} a_{1j}$. Thus, rearrange A and B into new matrices A^* and B^* so as to minimize the variances of the row-sum vectors, say X and Y. Then one can prove that

$$S_X \leq \overline{VaR_\alpha(L)} \leq S_Y$$

where S_X and S_Y denote, respectively, the minimal entries of X and Y.

2 The Rearrangement Algorithm

The key-step in the algorithm is the following. For the minimal variance of X (and Y) to be attained, the rearranged matrix A^* (B^*) must be "ordered", i.e. each column must be oppositely ordered to the sum of the others (see [2]).

Hence an algorithm procedure can be designed:

1. from a partition of $[\alpha, 1]$ into n subintervals, proceed to a partition of $2n$ subintervals, e.g. by dividing each subinterval into two halves;
2. write the new corresponding matrices A and B;
3. rearrange them into ordered matrices;
4. calculate the corresponding S_X and S_Y, and so on.

Therefore, one expects that eventually

$$S_X, S_Y \longrightarrow \overline{VaR_\alpha(L)}.$$

Given a matrix A or B as above, the ordered matrices which can be drawn from it may be, when $d > 2$, several, so that we are not sure whether the one we find first is "optimal" (i.e., it is the one whose row-sum vector attains the minimal variance). Consider a very simple example:

$$B_1 = \begin{pmatrix} 2 & 2 & 2 \\ 4 & 4 & 4 \end{pmatrix} \qquad \text{so as} \qquad B_1^* = \begin{pmatrix} 2 & 2 & 4 \\ 4 & 4 & 2 \end{pmatrix}$$

Now add to the ordered B_1^* two new rows corresponding to intermediate values 1, 3. Then

$$B_2 = \begin{pmatrix} 1 & 1 & 3 \\ 2 & 2 & 4 \\ 3 & 3 & 1 \\ 4 & 4 & 2 \end{pmatrix}$$

Rearrange B_2 starting from the first column

$$B_2^{'} = \begin{pmatrix} 4 & 1 & 3 \\ 3 & 3 & 1 \\ 2 & 2 & 4 \\ 1 & 4 & 2 \end{pmatrix} \longrightarrow Y^{'} = \begin{pmatrix} 8 \\ 7 \\ 8 \\ 7 \end{pmatrix}$$

Rearrange B_2, instead, from the third column

$$B_2^{''} = \begin{pmatrix} 1 & 1 & 4 \\ 2 & 2 & 3 \\ 3 & 3 & 1 \\ 4 & 4 & 1 \end{pmatrix} \longrightarrow Y^{''} = \begin{pmatrix} 6 \\ 7 \\ 8 \\ 9 \end{pmatrix}$$

Both $B_2^{'}$ and $B_2^{''}$ are ordered, but

$$\text{Variance } (Y^{'}) < \text{Variance } (Y^{''})$$

while

$$S_{Y'} = 7 > S_{Y''} = 6.$$

In [1], Embrechts, Puccetti and Rüschendorf proposed, then, to rearrange randomly the columns of A and B at each step before looking for ordered matrices. This way, though, the algorithm is no more deterministic.

3 Proof of Convergence

The road we explored was a different one. Let us start from an $n \times d$ matrix with positive entries, say B, where $d > 2$ (the case $d = 2$ is trivial). Let $\mathscr{O}(B)$ be the set of the ordered rearrangements of B. Then we introduce a topological structure in $\mathscr{O}(B)$. We can call it a "norm", h. If $B^* \in \mathscr{O}(B)$ and Y^* is the corresponding row-sum vector, define

$$h(B^*) := \text{Variance } (Y^*).$$

Then we prove the following Lemma.

Lemma 1 *We call B^* critical if one of the vectors Y_j^*, Y_{-j}^*, Y^* has two equal entries (where Y_j^* is the j-th column, Y^* is the sum of the columns, Y_{-j}^* is the sum of all the columns except the j-th one). If (up to an exchange of rows and columns) $B_1^* \neq B_2^* \in \mathscr{O}(B)$ are not critical, then*

$$h(B_1^*) \neq h(B_2^*).$$

The next step has to do with the construction of paths, that is of a flow, when B depends, in a suitable way, on a continuous parameter $\epsilon \in [0, 1]$. Namely, we consider a map

$$F : \big[\epsilon, B^*(0)\big] \longrightarrow B^*(\epsilon)$$

where $B^*(0) \in \mathscr{O}\big(B(0)\big)$ and $B^*(\epsilon) \in \mathscr{O}\big(B(\epsilon)\big)$. Such map is proved to have the following property.

Lemma 2 *Assume $B_1^* = F(0)$ is not critical and consider $F(1) = B_1^*(1)$. Then the curve joining $h\big(B_1^*(0)\big)$ and $h\big(B_1^*(1)\big)$ is the graph of a continuous function $\phi(\epsilon)$. Moreover, if $h\big(B_1^*(0)\big)$ is minimal in $\mathscr{O}\big(B(0)\big)$, then $h\big(B_1^*(1)\big)$ is minimal in $\mathscr{O}\big(B(1)\big)$.*

The main theorem follows.

Theorem 1 *Suppose, at the m-step of the rearrangement algorithm, the $n \times d$ ordered matrix B_m^* is not critical. Assume B_m^* is optimal (i.e. $h(B_m^*)$ is minimal in $\mathscr{O}\big(B_m\big)$). Then add n new rows, in the way we suggested above, and rearrange column-by-column the new $2n \times d$ matrix so as to find an ordered matrix B_{m+1}^*. Then, independently from what column you started first in the rearrangement, B_{m+1}^* is optimal.*

A critical matrix B_m^* can be modified, as slightly as we want, into a not critical one, and in the end the convergence can still be granted.

Example 1 Modify

$$B_1^* = \begin{pmatrix} 2 & 2 & 4 \\ 4 & 4 & 2 \end{pmatrix} \qquad \text{into} \qquad \bar{B}_1^* = \begin{pmatrix} 2+\sigma & 2+\sigma & 4 \\ 4 & 4 & 2 \end{pmatrix}$$

where $\sigma > 0$ is as small as we want. Adding the above new values 1, 3, whatever column we start from in the rearrangement, we obtain

$$B_2^* = \begin{pmatrix} 1 & 2+\sigma & 4 \\ 2+\sigma & 4 & 1 \\ 3 & 3 & 2 \\ 4 & 1 & 3 \end{pmatrix} \longrightarrow Y_2 = \begin{pmatrix} 7+\sigma \\ 7+\sigma \\ 8 \\ 8 \end{pmatrix}$$

References

1. Embrechts, P., Puccetti, G., Rüschendorf, L.: Model uncertainty and VaR aggregation. J. Bank. Financ. **37**, 2750–2764 (2013)
2. Rüschendorf, L.: Solution of a statistical optimization problem by rearrangement methods. Metrika **30**, 55–61 (1983)

Automatic Balancing Mechanisms in Practice: What Lessons for Pension Policy Makers?

Frederic Gannon, Florence Legros, and Vincent Touzé

Abstract Despite numerous reforms and the introduction of automatic or semi-automatic adjustment mechanisms, pension system future solvency is not guaranteed. Then, setting up an automatic balancing mechanism can offer several advantages. This article proposes to detail the specific properties of various adjustment rules prevailing in different countries to see to what extent their understanding may be helpful to determine public choices ensuring sustainable retirement systems. Several of such adjustment rules are possible. Three rules will get our attention. The American case is radical. The prohibition to resort to public debt, the so-called "fiscal cliff", forces the balance by a drastic reduction of pensions whenever the reserve fund is exhausted. The underlying idea is that this socially unacceptable perspective will force the parliament to take measures to restore solvency. The Swedish approach relies on an adjustment through the general level of pensions to guarantee a notional asset/liability ratio. A huge reserve fund smooths the shock associated with the aging of the population. The Canadian approach is based on an "inadequate rate provision" which increases the contribution rate and a pension freeze as long as the federal and provincial finance ministers do not reach an agreement.

Keywords Unfunded pension scheme · Automatic balancing mechanism

F. Gannon (✉)
University of Le Havre, Le Havre, France

OFCE - Sciences Po, Paris, France
e-mail: frederic.gannon@univ-lehavre.fr

F. Legros
ICN – Business School, Nancy, France
e-mail: florence.legros@icn-artem.com

V. Touzé
OFCE – Sciences Po, Paris, France
e-mail: vincent.touze@sciencespo.fr

M. Corazza et al. (eds.), *Mathematical and Statistical Methods for Actuarial Sciences and Finance*, https://doi.org/10.1007/978-3-319-89824-7_66

1 Introduction

Pension system solvency has become one of policymakers' core issues in aging developed countries. Many of these have opted for parametric reforms either by adopting precociously announced mechanisms of automatic adjustment or by discretionarily changing, in crisis mode, pension rules and/or contribution rates. But, in practice, these measures do not always guarantee solvency over the long term. This is the reason why some countries have explicitly or implicitly adopted an additional automatic balancing mechanism (ABM). Sweden is unanimously considered as a pioneer and a leader in this matter, with its ABM launched at the dawn of the twenty-first century. Other countries use alternative solvency rules. We focus on only two of these, namely the United States and Canada to evidence their differences and complementarities.

2 The US Social Security Fiscal Cliff: Automatic and Rough Adjustment Through Pension Benefits

In 1983, the U.S. government launched a radical long-run reform mainly by increasing the payroll taxes and raising the full pension age. This reform prevented a pending Social Security crisis. Moreover, it has guaranteed an intertemporal balanced budget for about half a century. Nevertheless, as stressed by Aaron [1], the weakness of this reform is that it "virtually guaranteed the return of deficits and a funding gap, and the need for further legislation to close it". In other words, solvency is not guaranteed permanently.

It has to be reminded that the U.S. Social Security trust funds are not allowed to borrow. This financial and legal constraint is a strong incentive to plan surpluses to compensate anticipated deficits, acting as a credible restoring force.

The case of U.S. Social Security budgetary rule is interesting since it must comply with a rule that prohibits debt. Should the reserve fund be exhausted, then the adjustment would be immediate and rough due to bankruptcy of the pension scheme. Social Security can only pay out pensions at the height of its revenue, which, de facto, implies a sharp decline in pensions. The underlying idea is that this socially unacceptable perspective will force the parliament to take measures to restore solvency. The prudential objective of the Social Security Trustees is straightforward, warranting a minimal reserve fund to smooth out the adjustments and to let lawmakers enough time to change the program.

From a prospective point of view, the Social Security Act requires that the Board of Trustees of the Federal Old-Age and Survivors Insurance (OASI) and Disability Insurance (DI) Trust Funds submits to the Congress an annual report on the actuarial and financial state of the PAYG. This report allows a thorough analysis of solvency over a 75-year time horizon and, notably, provides an estimation of the critical year when the system reaches bankruptcy. The last report published in July 2017

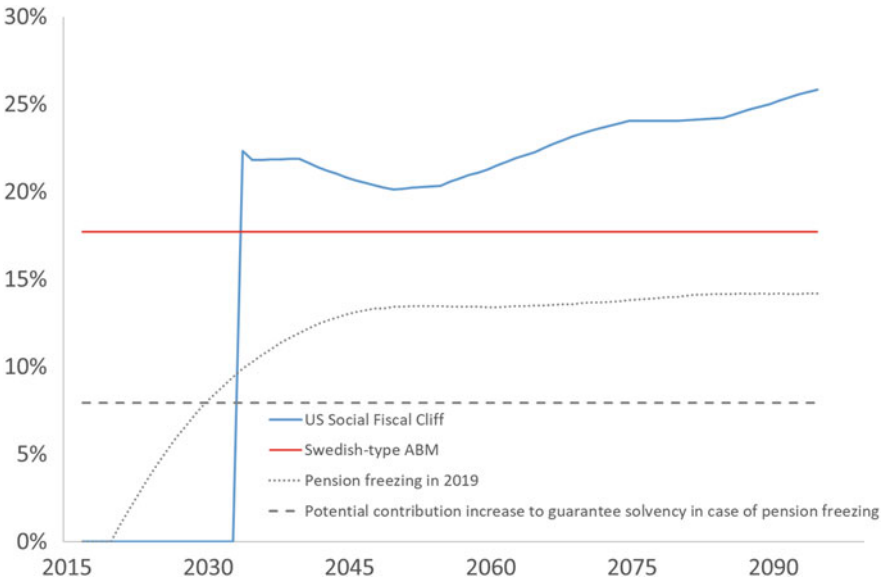


Fig. 1 Application of three types of ABM to US Social Security with 75-year solvency horizon. Source: All computations are based on data from US Social Security administration and are available upon request. Each ABM satisfies the balance between discounted sum of receipts plus the initial reserve fund and discounted sum of expenditures

estimates that in 2033, if no corrective governmental measures have been taken, an automatic and brutal adjustment would induce an immediate pension slash which would increase from 22% in 2033 to 26% in 2091 (see Fig 1).

3 The Swedish NDC Experiment: Reinforcing Automatic Adjustment Mechanisms by Introducing an Explicit ABM

Sweden is considered as a major pioneer, since it adopted Notional Defined Contributions (NDC) plans in 1994. This led the Swedish government to introduce actuarial rules to compute individual pensions, which rely on regular and fair revisions of the conversion coefficients for annuities comparable to an automatic adjustment mechanism. The challenge of the increase in life expectancy is clearly taken into account by inserting an explicit conversion rate in the actuarial formula used to calculate pensions. They are actuarially neutral, varying by cohort and age at which the individual retires.

To reinforce the robustness of the system, an Automatic Balance Mechanism (ABM) was launched in 2001 [2].

Sweden is only interested in working generations alive now, their current accumulated pension rights and future contributions they will pay in. Sweden has adopted a full adjustment mechanism where a global index on pension benefits is used to guarantee each year an intertemporal budget balance. The latter is computed as the equality between the discounted sum of current and future payroll taxes paid by the current workers and the implicit liabilities net of the reserve fund.

However, the Swedish model suffers a major flaw. When the 2008 economic and financial crisis resulted in both a capital loss—due to more than 23% decrease in its equity assets—of the reserve fund and a reduction in the growth of revenue, the Swedish capital ratio fell below its critical value. Hence, in order to return to balance, the amount of pensions was reduced.

To partially compensate the fall of pensioners' income, the Swedish government reduced taxes levied on pensions. In addition, the pension plan decided to spread the adjustment over 3 years, so as to smooth the necessary decrease in pension benefits. The adjustment rules have been modified in a discretionary manner. Another interpretation is that the pension system is still in control and that the Great Recession has raised a financial emergency which had not—or rather could not have—been anticipated.

An application of this Swedish-type ABM to the US Social Security would lead to an immediate and permanent 18% pension decrease (see Fig. 1).

4 Canada's Second Pillar: An Automatic Adjustment Through Contribution Scattered by the Absence of Political Choice

In Canada, the second pillar is made up of two mandatory partially funded plans: Canadian Pension Plan (CPP) and Quebec Pension Plan (QPP). Statutory periodic reviews of the CPP are scheduled from once every 5 years to once every 3 years, when financial status of the CPP is analyzed. Recommendations are made as to whether benefit or contribution rates, or both, should be changed. One of the main sources of information for the reviews is the actuarial report on the CPP by the chief actuary. Best-estimate assumptions are made without any provisions for adverse deviations, to avoid bias with respect to either current or future generations. As to the CPP reports, they are reviewed by an independent panel of Canadian actuaries.

The financial sustainability and intergenerational equity of the pension plan are closely monitored. Recent changes to the CPP aim at a better intergenerational fairness. One of these changes consisted in restoring CPP pension adjustment factors to their actuarially fair value, which implied both subsidiarizing early benefit uptake and penalizing late benefit uptake (after age 65). Canada's "self-adjustment mechanism" stipulates the simultaneous increase in contribution rate and the temporary freezing of the current pension benefit indexation.

Whenever, in the current actuarial valuation, legal contribution rate is lower than the minimum contribution rate to guarantee solvency and no agreement can be reached between federal and provincial finance ministers to increase or maintain the legislated rate, the ABM applying to the CPP lays down the following. For a 3-year period, contribution rate is increased by half the difference between the two rates, and pension benefits are frozen until the next actuarial review. This procedure is called "insufficient rates provisions" (Ménard and Billig [3]) which, in the case of CPP, plays as an ABM. Sakamoto [4] stresses that one of the advantages of this ABM specific to CPP is to "make policymakers conscious of intergenerational fairness". But, on the other hand, since it is activated only when the federal and province finance ministers do not reach an agreement, "it is unlikely that (it) will be activated" in practice.

As for the US Social Security, a permanent and partial pension freezing (inflation minus 1% indexing) would induce a slash ranging from 0.9% in 2019 to 14.2% slash in 2095 and would necessitate a permanent 8% contribution rate increase (see Fig. 1).

References

1. Aaron, F.J.: Social security reconsidered. Natl Tax J. **64**(2, Part 1), 385–414 (2011)
2. Settergren, O.: The Automatic Balance Mechanism of the Swedish Pension System. Working paper, National Social Insurance Board (2001)
3. Ménard, J.-C., Billig, A.: Intergenerational Balance of the Canadian Retirement Income System. Office of the Superintendent of Financial Institutions, Canada (2013)
4. Sakamoto, J.: Automatic Balancing Mechanisms. PBSS colloquium in Lyon. June (2013)

Empirical Evidence from the Three-Way LC Model

Giuseppe Giordano, Steven Haberman, and Maria Russolillo

Abstract The three-way Lee-Carter (LC) model was proposed as an extension of the original LC model when a three-mode data structure is available. It provides an alternative for modelling mortality differentials. This variant of the LC model adds a subpopulation parameter that deals with different drifts in mortality. Making use of several tools of exploratory data analysis, it allows giving a new perspective to the demographic analysis supporting the analytical results with a geometrical interpretation and a graphical representation. When facing with a three-way data structure, several choices on data pre-treatment will affect the whole data modelling. The first step of three-way mortality data investigation should be addressed by exploring the different source of variations and highlighting the significant ones. In this contribution, we consider the three-way LC model investigated by means of a three-way analysis of variance with fixed effects, where the three main effects, the three two-way interactions and one three-way interaction are analyzed. Aim of the paper is to highlight the technical-applicative infrastructure behind the methodology.

Keywords Anova · Three-way Lee-Carter model · Cross validation

1 Introduction

In the last few years, in the actuarial literature has been shown an increasing interest in population dynamics categorized by similar socio-economic conditions and by geographical proximity. Over time, it has been noticed that mortality rates and indices such as life expectancy frequently differ across sub-populations, given

G. Giordano · M. Russolillo (✉)
Department of Economics and Statistics, University of Salerno, Salerno, Italy
e-mail: ggiordano@unisa.it; mrussolillo@unisa.it

S. Haberman
Cass Business School, City University London, London, UK
e-mail: S.Haberman@city.ac.uk

M. Corazza et al. (eds.), *Mathematical and Statistical Methods for Actuarial Sciences and Finance*, https://doi.org/10.1007/978-3-319-89824-7_67

by gender, geographic area or socio-economic variables (e.g. occupation, level of education, income). These discrepancies, in particular those related to socio-economic conditions, pose significant challenges for planning pension systems and managing longevity risk in pension funds and annuity portfolios as well as for planning public policies for facing social disparities. In this framework, we consider a multiple population extension of the Lee-Carter (LC) model [1], the three-way LC model [2], proposed as an extension of the original LC model when a three-mode data structure is available. In particular, it provides an alternative for modelling mortality differentials, when socio-economic differences are taken into account. This alternative of the LC model contemplates a subpopulation parameter that deals with different drifts in mortality. Aim of the paper is to highlight the technical-applicative infrastructure behind the methodology. With this aim in mind, we consider England mortality data for socio-economic subpopulations defined using a deprivation index. We consider a first dataset (splitted into training set and test set) based on the idea of a validation process and a further data set to run an extensive empirical test. The results will be presented also from a standpoint of basic methodological issues and choices dealing with the proposed method.

2 The Three-Way Lee-Carter Model

In this section, we describe a multipopulation extension of the LC model which provides an alternative for modelling mortality differentials by adding a subpopulation parameter that deals with trend differences in mortality [4, 5]. This model proposes a parametric representation of the central death rate m_{xtp} in year t for people aged x in subpopulation p, as follows:

$$\ln\left(m_{xtp}\right) = \alpha_{x.p} + \beta_x \kappa_t \gamma_p + \varepsilon_{xtp}$$

where $\alpha_{x.p}$ measures the age-subpopulation-specific pattern of mortality, β_x and κ_t have the same interpretation as in the LC model, and γ_p is a new term capturing the variability in the improvement rates of the subpopulations. To make easier the interpretation of mortality differentials we can consider the following re-parametrisation [3]:

$$\ln\left(m_{xtp}\right) = \alpha_x + \alpha_{x.p} + \beta_x \kappa_t \gamma_p + \varepsilon_{xtp}$$

The three-way Lee-Carter offers a parsimonious and clear way of assessing mortality discrepancies and allows to simultaneous modelling mortality in a group of subpopulations.

3 The Analysis and Interpretation of the 3-Way LC Model

We discuss the three-way LC model in view of a fictitious case study.

First we consider the variability attached to three ways, let them be the traditional Age and Years and let us supposed to consider different Countries as the third way. Furthermore, we may consider the variability induced by the paired interactions between the three ways. Finally, the three way interaction could give information on which countries have specific trends (along years) in each age-group.

This kind of analysis is recommended to assess the source of variation in the mortality data before to extract rank-one components. In case of a three way data structure, this is much more important because of several choices on data pre-treatment that will affect the whole data modelling.

The three-way LC model should be investigated through a three-way analysis of variance with fixed effects, where each cell is given by the mortality rate in a given year of a specific age-group for a country.

We can thus analyse the three main effects, the 3 two-way interactions and the three-way interaction.

4 The Advantage of Introducing a Third Way in the Analysis

Roughly speaking, the advantage to introduce a third way in the analysis of a traditional years per age-class mortality data is much more clear if we can hypothesize that the decomposition along the third way bring information.

We can distinguish three cases: i) the different countries show homogeneous mortality patterns; ii) the different countries have large differences in their trend patterns and, iii) there exist groups of countries with homogeneous trends within the same group and showing different trend among groups.

In the first case, since the mortality data are similar, it is possible to aggregate the different mortality experiences. On the other hand, a factorial decomposition will furnish a single component that will explain much of the inherent variability.

In the second case, any data aggregation is awkward, and any solution could lead to unreliable results. In this case a factorial decomposition will give a poor synthesis on the first component.

The third case is of more interest from our point of view. In this case, we may argue that a unique synthesis is not reliable, but several synthesis are possible and they can be explored by the analytic tools of our exploratory approach.

Furthermore, the presence of different patterns should be indicated by a significant source of variation along the country way. Indeed, the first step of a mortality data investigation should be addressed by exploring the different source of variations and highlighting the significant ones.

The basic condition to analyse a mortality data structure is that there exists evidence of both a Years and Age-group effect. Moreover, we will explore the main effect of Country together with any two-way interactions.

The Age-group and Countries interaction could allow to explore the relative patterns of mortality rates for each specific Country in any Age-group and, conversely, which age-group is affected by a relative high (low) mortality rate for each Country.

The Years and Age-group two-way interaction represents the trends of the various age-group across the Years. Plotting Years versus Age-groups gives the classical representation before extrapolating the pooled time series.

The Years by Countries interaction shows how the trends vary with respect to each specific country, so exploring which years have relative high (low) mortality rates for any country and which country shows a relative high (low) impact in each specific year.

5 The Factorial Representations

In the framework of exploratory factorial methods, all this information will be displayed through the factorial representations. These plots gives an immediate overview by comparing and interpreting distance among points and their relative position on the factorial plan. The centre of the plot represents the average mortality rate for each variables (either Age-group or Country).

A year-point located at the origin has an average mortality rate for all age-group (respectively for all countries). Regarding the data structure, we consider the Countries and Age-groups as variables represented by vectors, while Years play the role of units and are represented by dots. According to the classical interpretation, angles between two vectors reflect the underlying correlation between the variables. For instance, countries whose vectors are collinear show similar trends along the years. Years-points relatively close have comparable mortality rates across countries. Similar interpretation for Age-groups can be derived.

Three-mode principal component analysis is an extension of ordinary principal component analysis to three-way data. Instead of two sets of coordinates for rows (Years) and columns (age-groups), there are now three sets of coordinates for components), one for each way, each with its own number of components. In addition, a set of parameters, the core array, indicates the importance of a particular combination of the components of each way. The interpretation of a three-mode component analysis of the three-way interaction requires the joint examination i) of the factorial plot (in this context generally called a *joint plot*) containing the relationships between the age-group and the Country and ii) the coefficients of the time components, which contain information on the changes over time in the relationships between age and Countries. Supposing that a time component show an increasing linear trend, a positive value (inner product) in the joint plot corresponds

to a linear increase in the interaction over the ages and a negative value a linear decrease over the ages.

This component will be the starting point for further analysis and for forecasts.

References

1. Lee, R.D., Carter, L.R.: Modelling and forecasting U.S. mortality. J. Am. Stat. Assoc. **87**, 659–671 (1992)
2. Russolillo, M., Giordano, G., Haberman, S.: Extending the Lee-Carter model: a three-way decomposition. Scand. Actuar. J. **2011**(2), 96–117 (2011)
3. Villegas, A., Haberman, S.: On the modeling and forecasting of socioeconomic mortality differentials: an application to deprivation and mortality in England. N. Am. Actuar. J. **18**(1), 168–193 (2014). https://doi.org/10.1080/10920277.2013.866034
4. Kroonenberg, P.M.: Three-Mode Principal Component Analysis. DSWO Press, Leiden (1983)
5. Kroonenberg, P.M., Murakami, T., Coebergh, J.W.W.: Added value of the three-way methods for the analysis of mortality trends illustrated with worldwide female cancer mortality (1968–1985). Stat. Methods Med. Res. **11**, 275–292 (2002)

Variable Selection in Estimating Bank Default

Francesco Giordano, Marcella Niglio, and Marialuisa Restaino

Abstract The crisis of the first decade of the 21st century has definitely changed the approaches used to analyze data originated from financial markets. This break and the growing availability of information have lead to revise the methodologies traditionally used to model and evaluate phenomena related to financial institutions. In this context we focus the attention on the estimation of bank defaults: a large literature has been proposed to model the binary dependent variable that characterizes this empirical domain and promising results have been obtained from the application of regression methods based on the extreme value theory. In this context we consider, as dependent variable, a strongly asymmetric binary variable whose probabilistic structure can be related to the Generalized Extreme Value (GEV) distribution. Further we propose to select the independent variables through proper penalty procedures and appropriate data screenings that could be of great interest in presence of large datasets.

Keywords Rare events · Variable selection · Banks failure prediction · GEV models

1 Introduction

Due to the last financial crisis, the failure risk prediction has attracted new attention, and the importance of having reliable statistical models able to provide early warning indicators of crisis and to identify the most likely predictors of banking failure is increased. In this context since there are some important challenges to be addressed, we propose a novel predictive method able to overcome two relevant drawbacks.

First, the number of banks' failures is usually very small, and the use of a logit model with symmetric link function may not be appropriate because of its

F. Giordano · M. Niglio (✉) · M. Restaino
Department of Economics and Statistics, University of Salerno, Salerno, Fisciano, Italy
e-mail: giordano@unisa.it; mniglio@unisa.it; mlrestaino@unisa.it

M. Corazza et al. (eds.), *Mathematical and Statistical Methods for Actuarial Sciences and Finance*, https://doi.org/10.1007/978-3-319-89824-7_68

symmetry around 0.5, which implies that the probability of default approaches zero at the same rate as it approaches one. Moreover the default probability will be underestimated. In order to overcome these limits, in line with other contributions [2, 9, among the others], we suggest to concentrate the estimation efforts on the tail of the distribution, adopting an asymmetric link function that lets the predicted default probability to approach one slower than it approaches zero. In more detail, we use the inverse of the distribution function of a generalized extreme value (GEV) random variable as link function in a generalized linear model with Bernoulli response variable.

Second, the number of predictors of banking failures could be very high and therefore it is crucial to identify a optimal subset of variables. The most used methods for selecting the best set of possible predictors do not take into account that some of the variables have a strong linear correlation among themselves (*multicollinearity*), due to the fact that numerator and denominator of some ratios are based on the same variables [1]. To overcome this limit, in our contribution we propose the use of an appropriate screening procedure for GEV regression models that is applicable in presence of large datasets (when the number of covariates is larger than the number of units) with independent and dependent predictors.

2 Statistical Model

Let Y be a response variable that describes a binary rare events, such that a random sample from Y has a small number of ones and let X be the $(n \times p)$ matrix of predictors. Since a symmetric link function (that traditionally characterizes logit models) can lead to biased estimates of defaults probability, an asymmetric link function is preferable. In [2] the quantile function of the GEV distribution has been suggested as link function, such that its distribution function is given by

$$F(X) = \exp\left\{-\left[1 + \tau\left(\frac{x-\mu}{\sigma}\right)\right]^{-\frac{1}{\tau}}\right\} \tag{1}$$

which is defined on $S_x = \{x : (1 + \tau(x - \mu)/\sigma) > 0\}$, with $-\infty < \mu < \infty$ and $\sigma > 0$. The parameter τ is the shape parameter, μ and σ are the location and scale parameters, respectively. For different values of τ different distributions can be obtained: if $\tau > 0$, we have Fréchet-type distribution; if $\tau < 0$, we have Weibull-type distribution and if $\tau \to 0$, the Gumbel-type distribution is obtained.

Then the GEV distribution function is used as response curve, so that the probability of default for each unit i $(i = 1, \ldots, n)$ is given by

$$\pi(\mathbf{x}_i) = P(Y_i = 1|\mathbf{x}_i) = \exp\left\{-[1 + \tau(\beta^{'}\mathbf{x}_i)]^{-\frac{1}{\tau}}\right\} \tag{2}$$

provided that $1 + \tau(\beta'\mathbf{x}_i) > 0$.

More details on the GEV distributions can be found in [8] and [3].

3 Screening for High Dimensional GEV Models

A feature of data related to bank default is the high dimension: the dimensionality of the p covariates can be definitely greater than the number of units n. In this context an essential challenge is the selection of the more relevant covariates that have to be properly chosen among hundreds or thousands of predictors.

This problem, not contemplated in the previous cited literature, has been differently faced (among the others see the review article [5]). In this domain we focus the attention on the *Sure Independence Screening* (SIS) introduced in [4], which is based on a quite sample idea that appear to be very powerful in variable selection.

The use of SIS in presence of generalized linear models has been largely investigated in [6] and [7] but its extension to GEV regressions has not yet examined.

The idea behind the SIS is here shortly summarized, referring the technical details to the cited literature. Following [6] and generalizing their approach to the GEV regression with response curve (2), we fit p marginal models (one for each predictor) such that the marginal log-likelihood $\ell(\beta_0 + \beta_j X_j, Y)$ is maximized with respect to β_0 and β_j, for $j = 1, 2, \ldots, p$, and we select the set of variables X_j such that $|\hat{\beta}_j| \geq \gamma$, with γ a predefined threshold value. In practice, the relevance for Y of the X_j variable is related to the magnitude of its marginal coefficient.

To evaluate the SIS performance in presence of GEV regressions (SIS-GEV) and to compare the performances with the logistic case, we have made a simulation study where we have considered $n = 200$ units, two different values for the number of predictors, $p = \{400, 600\}$ and shape parameter $\tau = -0.15$. We have further considered two scenarios:

S1. the p covariates are independent standard Gaussian random variables whereas the vector β is $\beta_0 = 0$, $\beta_j = 0.35$, for $j = 1, 2, 3$ and $\beta_j = 0$, for $j > 3$;

S2. the p variables are generated considering: X_1 and X_2 are two independent Gaussian random variables, $X_3, \ldots, X_{12}$ are marginally standard Gaussian with $corr(X_i, X_j) = 0.3$ in the first case (S2-C1) and $corr(X_i, X_j) = 0.5$ in the second case (S2-C2), with $i \neq j$, whereas the remaining X_j variables, for $j > 12$, are independent standard Gaussian random variables. The vector β is $\beta_0 = 0$, $\beta_1 = 1.3$, $\beta_2 = 1$, $\beta_3 = 1.3$ and $\beta_j = 0$, for $j > 3$.

In S1 and S2 the Y data are finally generated using the GEV model with two fixed percentage of ones: $\alpha = 5\%$ and $\alpha = 2.5\%$.

For all scenarios we have made 200 Monte Carlo replications and at each iteration we have evaluated if the relevant variables (X_1, X_2, X_3 in our case) are selected from the SIS-GEV. Following [7], at each iteration we have selected $d \approx n/(4 \log n)$ variables (that in practice allows to define the threshold γ) and we have evaluated the proportion of times in which one, two or all three relevant variables are selected from the SIS-GEV procedure. The results are presented in Table 1. It can be noted the better performance of the GEV model with respect to the logistic case in scenario S1, where the independence among variables and the small values of

Table 1 Proportions of inclusion of one (*1 var.*), two (*2 var.*) or three (*3 var.*) relevant covariates, among the $d = 10$ variables selected from the SIS procedure, when the estimated model is a GEV regression or a logit

	$\alpha = 5\%$						$\alpha = 2.5\%$					
S1	$d = 10$, $\beta = (0.35, 0.35, 0.35)$											
p	GEV			LOGIT			GEV			LOGIT		
	1 var.	2 var.	3 var.	1 var.	2 var.	3 var.	1 var.	2 var.	3 var.	1 var.	2 var.	3 var.
400	0.000	0.095	0.905	0.000	0.130	0.870	0.035	0.455	0.510	0.010	0.605	0.355
600	0.005	0.095	0.900	0.005	0.135	0.860	0.120	0.505	0.375	0.040	0.610	0.335
S2-C1	$d = 10$, $\beta = (1.3, 1, 1.3)$, $\rho = 0.3$											
p	GEV			LOGIT			GEV			LOGIT		
	1 var.	2 var.	3 var.	1 var.	2 var.	3 var.	1 var.	2 var.	3 var.	1 var.	2 var.	3 var.
400	0.000	0.095	0.905	0.000	0.075	0.925	0.025	0.405	0.570	0.015	0.440	0.545
600	0.000	0.105	0.895	0.000	0.140	0.860	0.050	0.460	0.490	0.025	0.500	0.475
S2-C2	$d = 10$, $\beta = (1.3, 1, 1.3)$, $\rho = 0.5$											
p	GEV			LOGIT			GEV			LOGIT		
	1 var.	2 var.	3 var.	1 var.	2 var.	3 var.	1 var.	2 var.	3 var.	1 var.	2 var.	3 var.
400	0.000	0.200	0.800	0.000	0.170	0.830	0.010	0.395	0.595	0.015	0.495	0.490
600	0.000	0.210	0.790	0.000	0.295	0.705	0.020	0.500	0.480	0.010	0.490	0.500

$\alpha = 5\%$ and $\alpha = 2.5\%$ are the percentage of ones in Y

the β_j, $j = 1, 2, 3$ makes the screening in the logit case more difficult. From the two cases of scenario S2, called S2-C1 and S2-C2, it can be noted that as the correlation among variables grows, and even correspondingly the β_j coefficients, the SIS variable selection based on the GEV regression performs quite always better than the corresponding logit model. This result is more pronounced in the S2-C2 case where the dependence from X_3 and other non relevant (but all correlated) variables makes more difficult the selection.

References

1. Amendola, A., Giordano, F., Parrella, M.L., Restaino, M.: Variable selection in high-dimensional regression: a nonparametric procedure for business failure prediction. Appl. Stoch. Mod. Bus. Ind. **33**, 355–368 (2017)
2. Calabrese, R., Osmetti, S.A.: Modelling sme loan defaults as rare events: the generalized extreme value regression model. J. Appl. Stat. **40**(6), 1172–1188 (2013)
3. Coles, S.: An Introduction to Statistical Modeling of Extreme Values. Springer, London (2001)
4. Fan, J., Lv, J.: Sure independence screening for ultrahigh dimensional feature space. J. Roy. Statist. Soc. B **70**(5), 849–911 (2008)
5. Fan, J., Lv, J.: A selective overview of variable selection in high dimensional feature space. Stat. Sin. **20**, 101–148 (2010)
6. Fan, J., Song, R.: Sure independence screening in generalized linear models with NP-dimensionality. Ann. Stat. **38**(6), 3567–3601 (2010)

7. Fan, J., Samworth, R., Wu, Y.: Ultrahigh dimensional feature selection: beyond the linear model. J. Mach. Learn. Res. **10**, 2013–2038 (2009)
8. Kotz, S., Nadarajah, S.: Extreme Value Distributions. Theory and Applications. Imperial College Press, London (2000)
9. 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**(4), 2000–2023 (2010)

Multiple Testing for Different Structures of Spatial Dynamic Panel Data Models

Francesco Giordano, Massimo Pacella, and Maria Lucia Parrella

Abstract In the econometric field, spatio-temporal data is often modeled by spatial dynamic panel data models (SDPD). In the last decade, several versions of the SDPD model have been proposed, based on different assumptions on the spatial parameters and different properties of the estimators. In particular, the classic version of the model assumes that the spatial parameters are homogeneous over location. Another version, proposed recently and called *generalized SDPD*, assumes that the spatial parameters are adaptive over location. In this work we propose a strategy for testing the particular structure of the spatial dynamic panel data model, by means of a multiple testing procedure that allows to choose between the generalized version of the model and some specific versions derived from the general one by imposing particular constraints on the parameters. The multiple test is made using the Bonferroni technique and the distribution of the multiple test statistic is derived by a residual bootstrap resampling procedure.

Keywords Spatio-temporal models · Model testing · Bootstrap

1 The SDPD Models

Consider a multivariate stationary process $\{\mathbf{y}_t\}$ of order p generating the observations at time t from p different locations. The following model

$$\mathbf{y}_t = D(\mathbf{l}_0)\mathbf{W}\mathbf{y}_t + D(\mathbf{l}_1)\mathbf{y}_{t-1} + D(\mathbf{l}_2)\mathbf{W}\mathbf{y}_{t-1} + \boldsymbol{\varepsilon}_t, \tag{1}$$

has been proposed by [1] as a generalized version of the spatial dynamic panel data model of [2]. The errors $\boldsymbol{\varepsilon}_t$ are serially uncorrelated, they have zero mean value and may show cross-sectional correlation and heteroscedasticity, which means that $\boldsymbol{\varepsilon}_t$

F. Giordano · M. Pacella (✉) · M. L. Parrella
University of Salerno, Fisciano, Italy
e-mail: giordano@unisa.it; mpacella@unisa.it; mparrella@unisa.it

M. Corazza et al. (eds.), *Mathematical and Statistical Methods for Actuarial Sciences and Finance*, https://doi.org/10.1007/978-3-319-89824-7_69

have a full variance/covariance matrix $\mathbf{\Sigma}_\varepsilon$; the *spatial matrix* $\mathbf{W}$ is a weight matrix with zero main diagonal; the matrices $D(\mathbf{l}_j)$ are diagonal, for $j = 0, 1, 2$, with main diagonal equal to vectors $\mathbf{l}_j = (\lambda_{j1}, \ldots, \lambda_{jp})$, respectively. Model (1) guarantees adaptivity by means of its $3p$ parameters λ_{ji}, $i = 1, \ldots, p$ and $j = 0, 1, 2$, and it is characterized by the sum of three terms: the *spatial component*, driven by matrix $\mathbf{W}$ and the vector parameter $\mathbf{l}_0$; the *dynamic component*, driven by $\mathbf{l}_1$; and the *spatial–dynamic component*, driven by $\mathbf{W}$ and $\mathbf{l}_2$.

Starting from the general model in (1), denoted as *generalized SDPD* model, we derive different models as special cases by considering some constraints on the parameters. The most used among these is the classic *SDPD* of [2], with only three parameters, where the spatial coefficients are constant among locations

$$\mathbf{y}_t = \lambda_0 \mathbf{W}\mathbf{y}_t + \lambda_1 \mathbf{y}_{t-1} + \lambda_2 \mathbf{W}\mathbf{y}_{t-1} + \boldsymbol{\varepsilon}'_t, \tag{2}$$

and the errors are homoscedastic and uncorrelated. We call this model *constant SDPD*. Other special cases of the model can be derived from the *generalized SDPD* by testing the significance of specific λ_{ji} coefficients.

2 A Strategy for Testing the Particular Structure of SDPD Models

In the sequel, we assume that $\mathbf{y}_1, \cdots, \mathbf{y}_T$ are T observations from a stationary process defined by (1) or (2). We assume that the process has mean zero and denote with $\mathbf{\Sigma}_j = Cov(\mathbf{y}_t, \mathbf{y}_{t-j}) = E(\mathbf{y}_t \mathbf{y}'_{t-j})$ the autocovariance matrix of the process at lag j, where the prime subscript denotes the transpose operator.

The estimators of the parameters for the *generalized SDPD* model (1) have been proposed and analyzed by Dou et al. [1]. Denote such estimators with $(\hat{\lambda}_{0i}, \hat{\lambda}_{1i}, \hat{\lambda}_{2i})'$, where the index $i = 1, \ldots, p$ indicates the specific location. For the sake of brevity, we do not report the details of such estimators here.

In order to test the structure of the *SDPD* model, we define the test statistics

$$\hat{D}_{ji} = \sqrt{n}\left(\hat{\lambda}_{ji} - \bar{\lambda}_j\right), \quad i = 1, \ldots, p, \text{ and } j = 0, 1, 2. \tag{3}$$

In the (3), we are comparing the estimator under the generalized model, $\hat{\lambda}_{ji}$, with the estimator under the standard model with constant coefficients, which is evaluated by $\bar{\lambda}_j = \frac{1}{p}\sum_{k=1}^{p} \hat{\lambda}_{jk}$, the mean value of the estimates over different locations, for $j = 0, 1, 2$. Note that large values of the statistics in the (3) denote a preference for the *generalized SDPD* model. Instead, when the true model has constant parameters, as in the *SDPD* model of [2], the statistics in (3) are expected to be around zero. In order to give an empirical evidence of this, Fig. 1 shows the estimated density (based on $N = 250$ replications of the model) of the statistic $\hat{D}_{ji} = \sqrt{n}\left(\hat{\lambda}_{ji} - \bar{\lambda}_j\right)$, for

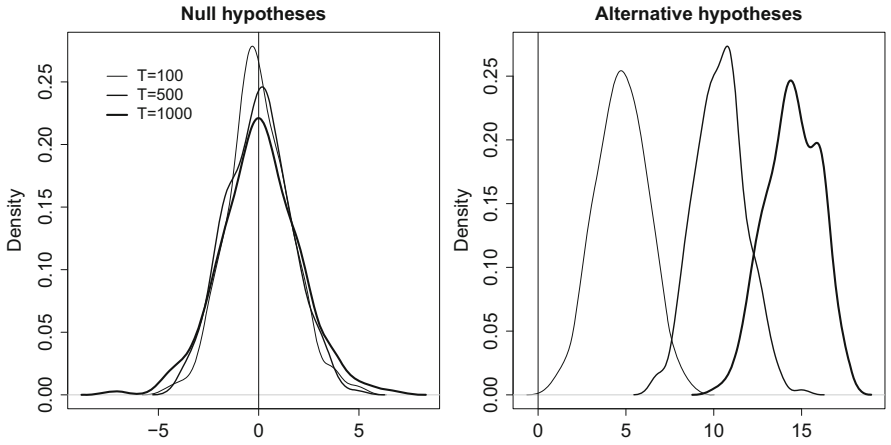


Fig. 1 Estimated densities (based on $N = 250$ replications of the model) of the statistic in (3), for $j = 2$, $i = 1$ and dimension $p = 50$, with different time series lengths (denoted by the line width, as indicated in the legend). The left side refers to the case generated under the Null hypothesis of true *constant SDPD* model. The right side refers to the case generated under the alternative hypothesis of true *generalized SDPD* model

$j = 2, i = 1$ and dimension $p = 50$, with different time series lengths (going from $T = 100$ to $T = 1000$ and denoted by the line width, as indicated in the legend). The left side of the figure refers to a case where the true model is the *constant SDPD* model, with constant parameters, therefore this is a case generated under the Null hypothesis. In such a case, as expected, the distribution of the statistic is centered around zero. The right side of the figure refers to a case where the true model is a *generalized SDPD*, with non-constant parameters, therefore this is a case generated under the alternative hypothesis. In the last case, as expected, the statistic $\hat{D}_{ji}$ is far away from zero. Moreover, as required for consistency, the value of the statistic increases for increasing time series length. Similar results can be shown for other values of i, j and p.

3 Bootstrap Scheme for the Multiple Testing Procedure

Figure 1 shows that the statistics in (3) can be used as building blocks of a testing procedure in order to identify the specific structure of the spatial dynamic model and to classify it between the two categories of *constant SDPD* and *generalized SDPD*. The hypotheses we need to test are

$$H_i : D_{ji} = 0, \quad \text{vs} \quad H_i' : D_{ji} \neq 0 \quad \text{for } i = 1, \ldots, p, \tag{4}$$

where j denotes the specific spatial parameter, $j = 0, 1, 2$. Test (4) has a multiple testing structure and the problem then becomes how to decide which hypotheses to reject, taking into account the multitude of tests. If many hypotheses are tested

jointly, some are bound to appear as significant by chance alone, even if in reality they are not relevant. To prevent us from declaring true null hypotheses to be false, we seek control (at least asymptotically) of the familywise error rate (FWE), which is the probability of making at least one false rejection. The most familiar scheme for controlling the FWE is the well known Bonferroni method: for each null hypothesis H_i , individual p-values p_is are computed and the hypothesis H_i is rejected at global level α if $p_i \leq \alpha/m$.

In order to derive the individual p-values p_is, we use a resampling procedure based on the residual bootstrap approach, to approximate the distribution of the test statistics $\hat{D}_{ji}$. This procedure runs as follows.

1. First obtain the bootstrap errors $\{\boldsymbol{\varepsilon}_t^*\}$ by drawing $B = 999$ replicates independently from the residuals $\hat{\boldsymbol{\varepsilon}}_t = \mathbf{y}_t - \hat{\mathbf{y}}_t$, where $\hat{\mathbf{y}}_t = \bar{\lambda}_0 \mathbf{W}\mathbf{y}_t + \bar{\lambda}_1 \mathbf{y}_{t-1} + \bar{\lambda}_2 \mathbf{W}\mathbf{y}_{t-1}$.
2. Generate the bootstrap series, under the Null hypothesis, as

$$\hat{\mathbf{y}}_t^* = (\mathbf{I}_p - \bar{\lambda}_0 \mathbf{W})^{-1}(\bar{\lambda}_1 \mathbf{I}_p + \bar{\lambda}_2 \mathbf{W})\mathbf{y}_{t-1}^* + \epsilon_t^*.$$

3. Compute the bootstrap statistics $\hat{D}_{ji}^* = \sqrt{n}\left(\hat{\lambda}_{ji}^* - \bar{\lambda}_j^*\right)$, as in (3), with $\hat{\lambda}_{ji}^*$ and $\bar{\lambda}_j^*$ estimated from the bootstrap data $\hat{\mathbf{y}}_t^*$.
4. For a given $i = 1, \ldots, p$, the individual p-value p_i for testing H_i is defined as the probability $P(|D_{ji}^*| > |\hat{D}_{ji}| \Big| \mathbf{y}_1, ..., \mathbf{y}_T)$, which is approximated by the relative frequency of the event $(|D_{ji}^*| > |\hat{D}_{ji}|)$ over the 999 bootstrap replications.

The size of the test (with nominal size $\alpha = 0.1$) and the power have been evaluated empirically for different values of p and T and reported in the following table.

Under the Null	for $j = 0$			for $j = 1$			for $j = 2$		
(=size)	$T = 100$	500	1000	100	500	1000	100	500	1000
$p = 10$	0.124	0.1	0.072	0.184	0.144	0.148	0.136	0.112	0.108
$p = 50$	0.024	0.12	0.092	0.156	0.144	0.172	0.144	0.164	0.24
$p = 100$	0.888	0.2	0.204	0.82	0.216	0.244	0.884	0.128	0.136
Under the Alternative	for $j = 0$			for $j = 1$			for $j = 2$		
(=power)	$T = 100$	500	1000	100	500	1000	100	500	1000
$p = 10$	0.204	1	1	1	1	1	0.988	1	1
$p = 50$	0.056	0.108	0.148	1	1	1	1	1	1
$p = 100$	0.44	0.968	1	1	1	1	1	1	1

References

1. Dou, B., Parrella, M.L., Yao, Q.: Generalized Yule-Walker estimation for spatio-temporal models with unknown diagonal coefficients. J. Econ. **194**(2), 369–382 (2016)
2. Yu, J., de Jong, R., Lee, L.F.: Quasi-maximum likelihood estimators for spatial dynamic panel data with fixed effects when both n and T are large. J. Econ. **146**, 118–134 (2008)

Loss Data Analysis with Maximum Entropy

Erika Gomes-Gonçalves, Henryk Gzyl, and Silvia Mayoral

Abstract We present some results of the application of maximum entropy methods to determine the probability density of compound random variables. This problem is very important in the banking and insurance business, but also appears in system reliability and in operations research.

The mathematical tool consists of inverting Laplace transforms of positive compound random variables using the maximum entropy method. This method needs a very small number of (real) values of the Laplace transform, is robust, works with small data sets, and it can be extended to include errors in the data as well as data specified up to intervals.

In symbols, the basic typical problem consist in determining the density f_S of a compound random variable like $S = \sum_{n=1}^{N} X_n$, or that of a sum of such random variables. There, N is an integer random variable and X_n is a sequence of positive, continuous random variables, independent among themselves and of N. Our methodology can be applied to determine the probability density of the total loss S and that of the individual losses.

Keywords Loss distributions · Sample dependence · Maximum entropy

E. Gomes-Gonçalves (✉)
Independent Consultant, Madrid, Spain

H. Gzyl
Instituto de Estudios Superiores de Administración - IESA, Caracas, Venezuela
e-mail: henryk.gzyl@iesa.edu.ve

S. Mayoral
Univ. Carlos III, Madrid, Department of Business Administration, Madrid, Spain
e-mail: smayoral@emp.uc3m.es

M. Corazza et al. (eds.), *Mathematical and Statistical Methods for Actuarial Sciences and Finance*, https://doi.org/10.1007/978-3-319-89824-7_70

1 Introduction

Estimate the correct density distribution f_S of a positive compound random variable S, is quite important in several fields. For example, in insurance, risk management operations research, reliability engineering, etc. There are situations in which models may not be trustworthy because we have little data, like in the case of very large disasters or large operational or credit risk losses. To begin with a motivational example, consider the need of calculating a (re)insurance premium like

$$\pi(S) = E\big[min\big(M, (S-d)^+\big)\big]. \tag{1}$$

It should be clear that when d is large (a high percentile) and M is greater than d by a couple of variances, compute $\pi(S)$ using empirical data may be quite unreliable.

To go around this difficulty we need a method to estimate probability densities from empirical data (or from models in the best of cases), that:

1. Produces densities from small data sets.
2. Is robust, that is, small variations in the data set produce small variations on the reconstructed density.
3. Needs a small "statistics" obtained from the data set.
4. Is easy to implement and fast to execute.
5. It can accommodate for measurement errors or data given up to intervals.

It is the object of this note to present some examples that illustrate these features in the maximum entropy procedure.

This note is organized as follows. In Sect. 2 we very briefly describe the mathematical problem that we solve and how the input to the maximum entropy procedure is computed. Then, in Sect. 3 we present two examples. One in which the power of the maxent is implicit. There we display the densities obtained from data of different sample sizes and we see how the densities tend to the "true" density as the sample becomes large. After that, we compute risk premiums according to (1) and plot the results in a box plot to display their sample variability.

We direct the reader to the volume [1], in which full detail about theoretical and numerical matters supporting the examples is presented in full detail, as well as a much larger bibliography.

2 The Standard Method of Maximum Entropy (SME)

The SME method is a variational procedure to obtain a numerical solution to the problem of finding a probability density f when we know the values of $\mu(\alpha)$ related to f by:

$$\int_0^\infty e^{-\alpha_k s} f(s)ds = \int_0^1 y^{\alpha_k} g(y)dy = \mu(\alpha_k), \quad k = 1, \ldots, K; \tag{2}$$

Actually, a variation on the procedure allows us to tackle the following problem: Determine a density f and estimate possible measurement errors ϵ_i such that

$$\int_0^\infty e^{-\alpha_k s} f(s)ds + \epsilon_k = \int_0^1 y^{\alpha_k} g(y)dy + \epsilon_k \in [a_k, b_k], \quad k = 1, \ldots, K; \quad (3)$$

where the intervals $[a_k, b_k]$ are part of the available information. Clearly, $g(y)$ is obtained from $f(s)$ after the change of variables $y = e^{-s}$, which can be reversed after $g(y)$ is obtained numerically.

In most cases we do not have models to obtain the $\mu(\alpha)$ from, and we have to content ourselves with estimating the moments from the data by

$$\int_0^\infty e^{-\alpha_k s} f(s)ds = \mu(\alpha_k) = \frac{1}{M}\sum_{t=1}^{M} e^{-\alpha_k S_t} \quad (4)$$

where $k = 1, \ldots K$, $t = 1, 2, \ldots, M$, and M denotes the number of non-zero data points in the data collected over some given period of time. These moments are the "statistics" mentioned in item (3) of the wish list mentioned above.

3 Examples

To begin with consider an example in which the reconstructed densities reflect the sample variance. Consider the following figure.

To obtain Fig. 1 we generated a large sample (5000 data points) corresponding to aggregating various sources of risk with different frequencies and different types of individual losses. Using this large sample we obtain the true density (the dark line in the various panels). And then we generated smaller samples. Here we present the densities obtained from samples of sizes $M = 10, 20, 50, 100, 500, 1000$, obtained by re-sampling (200 times) from the original sample. The gray clouds are the plotted curves of the densities. As it is to be expected, the clouds shrink as the size of the sample increases.

Once we have these densities, we can examine the sample variability of quantities of interest, like risk measures, risk premia, and so on. In particular, consider the box plot of the (re)insurance premia given by (1) computed with each of these densities. Clearly the variability of the premia decreases as the quality of the estimate (Fig. 2).

3.1 Notes and Comments

To conclude we mention that the maxent approach provides good reconstructions using a very small amount of information, which consists of:

1. Possibly a small sample of data points. Consider for example the top panel in Fig. 1.

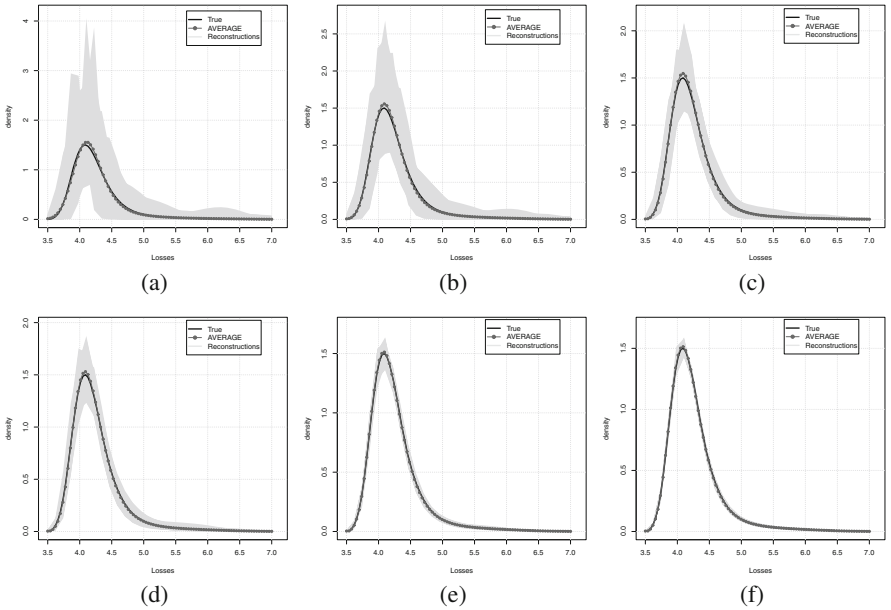


Fig. 1 SME reconstructions with different sample sizes. (**a**) M = 10, (**b**) M = 20, (**c**) M = 50, (**d**) M = 100, (**e**) M = 500, (**f**) M = 1000

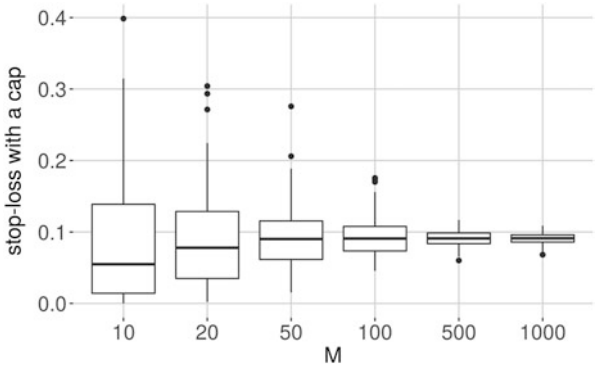


Fig. 2 Sample variability of stop-loss premia

2. Small number of generalized moments ("statistics"). In our case this consists of the knowledge of the Laplace transform at $K = 8$ real values of its transform's parameter, or according to (2), 8 fractional moments. Why this is so is explained in chapter 10 of [1].

Not only that, when all that we have is total loss data, and we know that it is data of a compound random variable, in a two step procedure, we can first determine the

probability density of the total losses, and invoking the maxent procedure, we can also determine the density of the individual losses.

Reference

1. Gomes-Gonçalves, E., Gzyl, H., Mayoral, S.: Loss Data Analysis: The Maximum Entropy Approach. De Gruyter, Berlin (2018)

Real-World Versus Risk-Neutral Measures in the Estimation of an Interest Rate Model with Stochastic Volatility

Lourdes Gómez-Valle and Julia Martínez-Rodríguez

Abstract In this paper, we consider a jump-diffusion two-factor model which stochastic volatility to obtain the yield curves efficiently. As this is a jump-diffusion model, the estimation of the market prices of risk is not possible unless a closed form solution is known for the model. Then, we obtain some results that allow us to estimate all the risk-neutral functions, which are necessary to obtain the yield curves, directly from data in the markets. As the market prices of risk are included in the risk-neutral functions, they can also be obtained. Finally, we use US Treasury Bill data, a nonparametric approach, numerical differentiation and Monte Carlo simulation approach to obtain the yield curves. Then, we show the advantages of considering the volatility as second stochastic factor and our approach in an interest rate model.

Keywords Interest rates · Stochastic volatility · Jump-diffusion · Stochastic processes · Nonparametric · Estimation · Numerical differentiation

1 Introduction

Traditionally, the financial literature assumes that interest rates move continuously and they are modelled as diffusion processes, as in [3, 11] and so on. However, more recent studies have showed that interest rates contained unexpected discontinuous changes, see for example [4, 9]. Jumps in interest rates are, probably, due to different market phenomena such as surprises or shocks in foreign exchange markets. Moreover, when pricing and hedging financial derivatives jump-diffusion models are very important, since ignoring jumps can produce hedging and pricing risks, see [10].

L. Gómez-Valle (✉) · J. Martínez-Rodríguez
Departamento de Economa Aplicada and IMUVA, Valladolid, Spain
e-mail: lourdes@eco.uva.es; julia@eco.uva.es

M. Corazza et al. (eds.), *Mathematical and Statistical Methods for Actuarial Sciences and Finance*, https://doi.org/10.1007/978-3-319-89824-7_71

It is widely known that one-factor interest rate models are very attractive for practitioners because its simplicity and computational convenience. However, these models have also unrealistic properties. First, they cannot generate all the yield curve shapes and changes that we can find in the markets. Second, the changes over infinitesimal periods of any two interest-rate dependent variables will be perfectly correlated. Finally, as [8] shows, none of their analyzed one-factor models captures the interest rate dynamics adequately. Therefore, we consider that at least two factors are necessary to model the term structure of interest rates. In fact, the number of factor must be a compromise between numerical efficient implementation and the capability of the model to fit data.

The main goal of this paper is twofold. We consider that the volatility is an important stochastic factor which can help to obtain the yield curves accurately. Then we show some results to estimate all the functions of this model even when a closed form solution is not known.

The rest of the paper is organized as follows. Section 2 shows a two-factor interest rate jump-diffusion model. In Sect. 3, we propose some results in order to estimate all the functions of the model efficiently. Finally, in Sect. 4 we show the supremacy of considering the volatility as stochastic and the estimation of the whole functions of the model directly from data in the market using interest rate data from US Federal Reserve.

2 The Valuation Model

Define $(\Omega, \mathscr{F}, \{\mathscr{F}_t\}_{t\geq 0}, \mathscr{P})$ as a complete filtered probability space which satisfies the usual conditions where $\{\mathscr{F}_t\}_{t\geq 0}$ is a filtration, see [1, 2].

In this paper, we consider an interest rate model with two factors: the instantaneous interest rate (r) and the volatility (V). We assume that under the risk-neutral measure the interest rate follows a jump-diffusion process and the volatility a diffusion process as follows,

$$r(t) = r(0) + \int_0^t \mu_r^{\mathscr{Q}}(r(z), V(z))\, dz + \int_0^t V(z) dW_r^{\mathscr{Q}}(z) + \int_0^t d\tilde{J}^{\mathscr{Q}}(z), \quad (1)$$

$$V(t) = V(0) + \int_0^t \mu_V^{\mathscr{Q}}(r(z), V(z))\, dz + \int_0^t \sigma_V(r(z), V(z)) dW_V^{\mathscr{Q}}(z), \quad (2)$$

where $\mu_r^{\mathscr{Q}} = \mu_r - V\theta^{W_r}$, $\mu_V^{\mathscr{Q}} = \mu_V - \sigma_V\theta^{W_V}$, with μ_r and μ_V the drifts of the interest rate and volatility processes under the Physical measure respectively, $W_r^{\mathscr{Q}}$ and $W_V^{\mathscr{Q}}$ are the Wiener processes under $\mathscr{Q}$-measure and $[W_r^{\mathscr{Q}}, W_V^{\mathscr{Q}}](t) = \rho t$. The market prices of risk associated to W_r and W_V Wiener processes are $\theta^{W_r}(r, V)$ and $\theta^{W_V}(r, V)$, respectively. Finally, $\tilde{J}^{\mathscr{Q}}(t) = \sum_{i=1}^{N^{\mathscr{Q}}(t)} Y_i - \lambda^{\mathscr{Q}} t E_Y^{\mathscr{Q}}[Y_1]$ is the compensated compound Poisson process under $\mathscr{Q}$-measure and the intensity of

the Poisson process $N^{\mathcal{Q}}(t)$ is $\lambda^{\mathcal{Q}}(r, V)$. Furthermore, $Y_1, Y_2, \ldots$ is a sequence of identically distributed random variables with a normal distribution $N(0, \sigma_Y^2)$.

A zero-coupon bond price at time t with maturity at time T, $t \leq T$, under the above assumptions, can be expressed as $P(t, r, V; T)$. It verifies that $P(T, r, V; T) = 1$ at maturity and it can be expressed by

$$P(t, r, V; T) = E^{\mathcal{Q}}[\mathrm{e}^{-\int_t^T r(u)\,du} | r(t) = r, V(t) = V]. \tag{3}$$

Moreover, the yield curve can be obtained as

$$R(t, r, V; T) = \frac{-\ln(P(t, r, V; T))}{T - t}. \tag{4}$$

3 Results and Approximations

In the spirit of studies such as [5–7] in one-factor interest rate models, which provide some relations to estimate the risk-neutral functions directly from data in the markets, we propose a novel approach to estimate the whole functions of a two factor jump-diffusion term structure model directly from market data, although a closed-form solution for the zero-coupon bond prices is not known. As we estimate directly the risk-neutral functions from interest rate data, the market prices of risk has not to be arbitrary specified. Moreover, this approach can be applied to parametric as well as nonparametric models.

We prove that if $P(t, r, V; T)$ is the price of a zero-coupon bond and r and V follow the joint stochastic processes given by (1)–(2), then:

$$\begin{aligned}
\frac{\partial R}{\partial T}|_{T=t} &= \frac{1}{2}\left(\mu_r^{\mathcal{Q}} + \lambda^{\mathcal{Q}} E^{\mathcal{Q}}[Y_1]\right)(t),\\
\frac{\partial (r^2 P)}{\partial T}|_{T=t} &= \left(-r^3 + 4r\frac{\partial R}{\partial T}|_{T=t} + V^2 + \lambda^{\mathcal{Q}} E^{\mathcal{Q}}[Y_1^2]\right)(t),\\
\frac{\partial (r^3 P)}{\partial T}|_{T=t} &= \left(-r^4 + 2r^2\frac{\partial R}{\partial T}|_{T=t} + 3r(V^2 + \lambda^{\mathcal{Q}} E^{\mathcal{Q}}[Y_1^2])\lambda^{\mathcal{Q}} E^{\mathcal{Q}}[Y_1^3]\right)(t),\\
\frac{\partial (r^4 P)}{\partial T}|_{T=t} &= \Big(-r^5 + 8r^3\frac{\partial R}{\partial T}|_{T=t} + 6r^2(V^2 + \lambda^{\mathcal{Q}} E^{\mathcal{Q}}[Y_1^2]) + 4r\lambda^{\mathcal{Q}} E^{\mathcal{Q}}[Y_1^3]\\
&\qquad + \lambda^{\mathcal{Q}} E^{\mathcal{Q}}[Y_1^4]\Big)(t),\\
\frac{\partial (V P)}{\partial T}|_{T=t} &= \left(\mu_V^{\mathcal{Q}} - rV\right)(t),
\end{aligned}$$

The derivatives above should be assumed as right derivatives when $T \in (\tau_i)_{i\geq 1}$, that is, when T is a jump time.

4 Empirical Application

In order to implement our approach, we take interest rate data from the Federal Reserve in USA from January 1997 to February 2017. We use data till December 2016, a nonparametric method (the Nadaraya-Watson estimator) and numerical differentiation to estimate all the functions of the two-factor model with the results in Sect. 3. Then, we obtain zero-coupon bond prices with maturities: 6 months, 1, 2, 5, and 7 years; and the yield curves from January to February 2017.

Our empirical results show that the jump-diffusion two-factor model in Sect. 2 estimated with our approach is more accurate than a one-factor jump-diffusion model. The Root Mean Square Error with the two-factor model is $1.0742e^{-3}$ but it is $3.3675e^{-3}$ with the one-factor model. Finally, we also find that meanwhile the one-factor model, in general, underprices the yield curves our two-factor model slightly overprices. This fact should be taken into account by practitioners in the markets for pricing and hedging interest rate derivatives.

Acknowledgements This work is partly supported by the GIR Optimización Dinámica, Finanzas Matemáticas y Utilidad Recursiva of the University of Valladolid and the projects: VA041P17 of the Consejería de Educación, MTM2014-56022-C2-2-P and MTM2017-85476-C2-P of the Spanish Ministerio de Economía y Competitividad and European FEDER Funds.

References

1. Applebaum, D.: Lévy Processes and Stochastic Calculus. Cambridge University Press, Cambridge (2009)
2. Cont, R., Tankov, P.: Financial Modeling with Jump Processes. Chapman and Hall/CRC, Boca Raton (2004)
3. Cox, J.C., Ingersoll J.E., Ross S.A.: A theory of the term structure of interest rates. Econometrica **53**, 385–407 (1985)
4. Das, S.R., Foresi, S.: Exact solutions for bond and option prices with systematic jump risk. Rev. Deriv. Res. **1**, 7–24 (1996)
5. Gómez-Valle, L., Martínez-Rodríguez, J.: Modelling the term structure of interest rates: an efficient nonparametric approach. J. Bank. Financ. **32**, 614–623 (2008)
6. Gómez-Valle, L., Martínez-Rodríguez, J.: The role of risk-neutral jump size distribution in single-factor interest rate models. Abstr. Appl. Anal. **2015**, Article ID 805695, 8 pages (2015)
7. Gómez-Valle, L., Martínez-Rodríguez, J: Estimation of risk-neutral processes in single-factor jump-diffusion interest rate models. J. Comput. Appl. Math. **291**, 48–57 (2016)
8. Hong, Y., Li, H.: Nonparametric specification testing for continuous-time models with applications to term structure of interest rates. Rev. Financ. Stud. **18**, 37–84 (2005)
9. Johannes, M.: The statistical and economic role of jumps in continuous-time interest rate models. J. Financ. **59**, 227–259 (2004)

10. Lin, B.H., Yeh, S.K.: Jump diffusion interest rate process: an empirical examination. J. Bus. Financ. Account. **26**, 967–995 (1999)
11. Vasiceck, O.: An equilibrium characterization of the term structure. J. Financ. Econ. **5**, 177–188 (1977)

Extensions of Fama and French Models

María de la O González and Francisco Jareño

Abstract This short paper proposes extensions of Fama and French models and compares their explanatory power. In concrete, it tests fluctuations in US sector returns between November 1989 and February 2014. In addition, this paper estimates the models using the quantile regression approach. In short, the most complete model shows the highest explanatory power and the extreme quantiles ($tau = 0.1$) show the best results.

Keywords Risk factors · Interest rates · Stock returns

1 Introduction and Methodology

1.1 Aim

This research compares different factor models in explaining variations in US sector returns between November 1989 and February 2014 using the quantile regression approach.

The models proposed in this study are based on the Fama and French [1] three-factor model and the Fama and French [2] five-factor model.

1.2 Methodology

As commented previously, this short paper extends the Fama and French [1] models, adding explanatory variables, such as nominal interest rates and its components, real interest and expected inflation rates [3–7]. Furthermore, it includes the Carhart [8]

M. de la O González · F. Jareño (✉)
University of Castilla-La Mancha, Albacete, Spain
e-mail: Mariao.Gonzalez@uclm.es; Francisco.Jareno@uclm.es

M. Corazza et al. (eds.), *Mathematical and Statistical Methods for Actuarial Sciences and Finance*, https://doi.org/10.1007/978-3-319-89824-7_72

risk factor for momentum (MOM) and for momentum reversal (LTREV) and the Pastor and Stambaugh [9] traded liquidity factor (LIQV).

The most complete models are the following:

Extension of the Fama and French [1] three-factor model

$$\begin{aligned} R_{jt} - R_{Ft} = a_j + b_j \cdot (R_{Mt} - R_{Ft}) + d_{jr} \cdot \Delta r_t + o_{j\pi} \cdot \Delta E_t^{ORT} \left(\pi_{t,t+1}\right) \\ + s_j \cdot SMB_t + h_j \cdot HML_t + m_j \cdot MOM_t + v_j \cdot LTREV_t \\ + l_j \cdot LIQV_t + e_{jt} \end{aligned} \quad (1)$$

where R_{jt} is the stock (sector) j return in month t, R_{Ft} is the riskfree return, R_{Mt} is the return on the value-weight market portfolio, Δr_t represents unexpected changes in real interest rates, $\Delta \mathrm{E}_t^{ORT}$ $(\pi_{t,t+1})$ shows shocks in the expected inflation rate (orthogonalized), SMB_t is the return on a diversified portfolio of small stocks minus the return on a diversified portfolio of big stocks, HML_t is the difference between the returns on diversified portfolios of high and low B/M stocks, MOM_t is the average return on the two high prior return portfolios minus the average return on the two low prior return portfolios, $LTREV_t$ is the average return on the two low prior return portfolios minus the average return on the two high prior return portfolios, $LIQV_t$ is the value-weighted return on the 10-1 portfolio from a sort on historical liquidity betas and, finally, e_{jt} is the error term.

Extension of the Fama and French [2] five-factor model

$$\begin{aligned} R_{jt} - R_{Ft} = a_j + b_j \cdot (R_{Mt} - R_{Ft}) + d_{jr} \cdot \Delta r_t + o_{j\pi} \cdot \Delta E_t^{ORT} \left(\pi_{t,t+1}\right) \\ + s_j \cdot SMB_t + h_j \cdot HML_t + r_j \cdot RMW_t + c_j \cdot CMA_t \\ + m_j \cdot MOM_t + v_j \cdot LTREV_t + l_j \cdot LIQV_t + e_{jt} \end{aligned} \quad (2)$$

where RMW_t is the difference between the returns on diversified portfolios of stocks with robust and weak profitability, and CMA_t is the difference between the returns on diversified portfolios of the stocks of low and high investment firms.

2 Main Results and Concluding Remarks

This paper shows that the most complete model, based on the Fama and French [2] five-factor model, that breaks down nominal interest rates into real interest and expected inflation rates and also aggregates the three risk factors is the model with the highest explanatory power (with a mean Adj. R^2 about 50%).

Moreover, this research points out that the extreme quantile 0.1 of the return distribution (Table 1) shows better results in all the factor models in comparison with the rest of quantiles (for example, tau = 0.9, in Table 2).

Table 1 Explanatory power of models at quantile 0.1

Models	Min. Adj. R^2	Max. Adj. R^2	Mean Adj. R^2
Ext. F-F 3 factor	0.2472	0.6577	0.4631
Ext. F-F 5 factor	0.2745	0.6611	0.4881

Table 2 Explanatory power of models at quantile 0.9

Models	Min. Adj. R^2	Max. Adj. R^2	Mean Adj. R^2
Ext. F-F 3 factor	0.1131	0.5953	0.3660
Ext. F-F 5 factor	0.1440	0.6066	0.3997

References

1. Fama, E.F., French, K.R.: Common risk factors in the returns on stocks and bonds. J. Financ. Econ. **33**(1), 3–56 (1993)
2. Fama, E., French, K.: A five-factor asset pricing model. J. Financ. Econ. **116**(1), 1–22 (2015)
3. González, M., Jareño, F., Skinner, F.: Interest and inflation risk: Investor behavior. Front. Psychol. **7**(390), 1–18 (2016)
4. González, M., Jareño, F., Skinner, F.S.: The financial crisis impact: An industry level analysis of the US stock market. Appl. Econ. Int. Dev. **17-2**, 61–74 (2017)
5. Jareño, F.: Spanish stock market sensitivity to real interest and inflation rates: An extension of the Stone two-factor model with factors of Fama and French three-factor model. Appl. Econ. **40**(24), 3159–3171 (2008)
6. Jareño, F., Ferrer, R., Miroslavova, S.: US stock market sensitivity to interest and inflation rates. Appl. Econ. **48**, 2469–2481 (2016)
7. Ferrando, L., Ferrer, R., Jareño, F.: Interest rate sensitivity of Spanish industries: A quantile regression approach. Manch. Sch. **85**(2), 212–242 (2017)
8. Carhart, M.M.: On persistence in mutual fund performance. J. Financ. **52**(1), 57–82 (1997)
9. Pastor, L., Stambaugh, R.F.: Liquidity risk and expected stock returns. J. Polit. Econ. **111**(3), 642–685 (2003)

The Islamic Financial Industry: Performance of Islamic vs. Conventional Sector Portfolios

María de la O González, Francisco Jareño, and Camalea El Haddouti

Abstract This paper studies the basic principles of the Islamic financial system to know the positive aspects that make it more solid and stable than the conventional financial system during financial crises. On the other hand, this research carries out a comparison between conventional and Islamic sectoral portfolios for the period from January 1996 to December 2015, through the use of different performance measures. Specifically, the performance measures used in this paper include Jensen's, Treynor's, Sharpe's and Sortino's classical performance ratios and two of the most recent and accurate performance measures that take into account the four statistical moments of the probability distribution function, the Omega's ratio and the MPPM statistic. In addition, for robustness, this paper analyses whether the performance results of conventional and Islamic sector portfolios depend on the state of the economy, by splitting the whole sample period into three sub-periods: pre-crisis, crisis and post-crisis. So, this paper would determine which types of portfolios offer better performance depending on the economic cycle. The main results confirm that, in general, the best performing sector is Health Care, while the worst performing sector is Financials. Furthermore, Islamic portfolios provide higher returns than conventional portfolios during the full period as well as the three sub-periods.

Keywords Islamic financial system · Islamic principles · Performance measurement

1 Introduction

Islamic banking emerges to help Muslim clients to obtain funding for their economic activities while they are faithful to their religious beliefs due to the fact that it is based on Sharia or Islamic law that regulates human behavior. Moreover, the

M. de la O González (✉) · F. Jareño · C. El Haddouti
University of Castilla-La Mancha, Albacete, Spain
e-mail: Mariao.Gonzalez@uclm.es; Francisco.Jareno@uclm.es; Camalea.ElHaddouti@alu.uclm.es

M. Corazza et al. (eds.), *Mathematical and Statistical Methods for Actuarial Sciences and Finance*, https://doi.org/10.1007/978-3-319-89824-7_73

adaptation of contracts to Islamic principles allows Islamic portfolios to be less affected by periods of economic recession, since the risks assumed by them are lower than those of conventional portfolios.

On the other hand, some recent studies about Islamic stocks use classical performance measures, remarking the great utility of these ratios for the analysis and comparison of portfolios. Umar [1] uses the Sharpe ratio, concluding that Islamic stocks are preferred for a short period of time, whereas conventional stocks are preferred for a long period of time. Jawadi et al. [2] use the Sharpe, Treynor and Jensen ratios and they conclude that the conventional indexes perform better than the Islamic ones in the period before the financial crisis, but they perform worse during the financial crisis. Thus, the state of the economy may be a key factor to analyze the portfolio performance.

Therefore, one of the main purposes of this paper is to compare conventional and Islamic sector portfolios by using not only classical but also recent performance measures such as the Omega ratio and the MPPM statistic [3, 4].

2 Performance

Specifically, the classical performance measures studied in this paper are the well-known Jensen's [5] alpha ratio, Treynor's [6] reward to volatility ratio and Sharpe's [7] reward to variability ratio.

The fourth performance measure proposed in this paper is the Sortino ratio, developed by Sortino and van der Meer [8]. This ratio is a modification to the Sharpe ratio that just penalizes those returns falling below a threshold return.

The fifth performance measure is the Omega ratio, developed by Keating and Shadwick [9] as an alternative to classical performance measures. This reliable risk indicator takes into consideration the four statistical moments of the probability distribution function (mean, variance, skewness and kurtosis), instead of just the first and second moments of the distribution (mean and variance).

$$O_p = \frac{E\left[\text{Max}\ \left(E\left(r_p\right) - \tau, 0\right)\right]}{E\left[\text{Max}\ \left(\tau - E\left(r_p\right), 0\right)\right]} \tag{1}$$

where $E(r_p)$ is the return of the portfolio and τ is the threshold of the expected return.

The sixth performance measure is the Manipulation Proof Performance Measure (MPPM) developed by Ingersoll et al. [10] to avoid that portfolio managers could manipulate their portfolios, either deliberately or otherwise, in order to score well on the previous static performance measures even though the manager has no private information [3, 4]. Moreover, the MPPM does not rely on any distributional assumption and so, this measure takes into account the four statistical moments of

the probability distribution function (mean, variance, skewness and kurtosis).

$$MPPM(A) = \left[\frac{1}{(1-A)\,\Delta t} \ln \left(\frac{1}{T} \sum_{t=1}^{T} \left[\frac{(1+r_t)}{(1+r_{ft})} \right]^{(1-A)} \right) \right] \tag{2}$$

where the parameter A is the measure of relative risk aversion that, like Ingersoll et al. [10], we use a A value of 3. Meanwhile, Δt is the time length between observations, T is the number of observations, r_t is the return of the selected portfolio at time t, r_{ft} is the risk-free interest rate at time t.

In general, the higher the ratios the better the risk-adjusted returns and so, the better the portfolio.

3 Main Results and Concluding Remarks

The performance measures of conventional and Islamic sector portfolios have been calculated in the whole sample period (January 1996–December 2015) and in three sub-periods: pre-crisis sub-period (January 1996–June 2007), crisis sub-period (July 2007–December 2010) and post-financial crisis sub-period (January 2011–December 2015).

Table 1 summarizes the values of the MPPM statistic for conventional and Islamic sector portfolios.

In detail, the values of the MPPM statistic, for the best performing sector, Health Care, is −1.148 for the conventional portfolio and −1.150 for the Islamic portfolio in the full period, while the MPPM value of this sector for conventional and Islamic portfolios, respectively, is −1.801 and −1.802 in the pre-crisis sub-period, −0.494 and −0.490 in the crisis sub-period and 0.116 and 0.110 in the post-crisis sub-period.

On the other hand, the values of MPPM for the worst performing sector, Financials, is −1.240 for the conventional portfolio and −1.265 for the Islamic portfolio in the total period, while this performance measure for Financials in the conventional and Islamic market is, respectively, −1.814 and −1.918 in the pre-crisis sub-period, −0.844 and −0.630 in the crisis sub-period and, finally, −0.003 and 0.036 in the post-crisis sub-period.

Therefore, the best performance according to the MPPM statistic both for the best performing sector, Health Care, and for the worst performing sector, Financials, corresponds to the post-crisis sub-period both in conventional and in Islamic markets.

Furthermore, Islamic sector portfolios perform better than conventional sector portfolios not only for the whole sample period but also for the three sub-periods. This could be due to the fact that the low level of uncertainty and speculation in Islamic finance and the prohibition of interest rates that negatively affect the

Table 1 MPPM of conventional and Islamic sector portfolios

	Whole sample period	Pre-crisis sub-period	Crisis sub-period	Post-crisis sub-period
Conventional				
Basic materials	−1.243	−1.830	−0.714	−0.167
Consumer goods	−1.163	−1.811	−0.489	0.040
Consumer services	−1.174	−1.827	−0.519	0.079
Oil and gas	−1.226	−1.801	−0.735	−0.146
Financials	−1.240	−1.814	−0.844	−0.003
Health care	−1.148	−1.801	−0.494	0.116
Industrials	−1.203	−1.837	−0.615	0.006
Technology	−1.233	−1.922	−0.525	0.042
Telecommunications	−1.211	−1.851	−0.599	−0.019
Utilities	−1.195	−1.803	−0.609	−0.037
Islamic				
Basic materials	−1.235	−1.815	−0.722	−0.167
Consumer goods	−1.163	−1.811	−0.493	0.046
Consumer services	−1.156	−1.804	−0.486	0.094
Oil and gas	−1.227	−1.807	−0.724	−0.143
Financials	−1.265	−1.918	−0.630	0.036
Health care	−1.150	−1.802	−0.490	0.110
Industrials	−1.196	−1.835	−0.584	0.010
Technology	−1.238	−1.928	−0.535	0.042
Telecommunications	−1.201	−1.836	−0.569	−0.030
Utilities	−1.226	−1.809	−0.750	−0.105

economic evolution would justify the greater profitability obtained by the Islamic sectors, even along the crisis sub-period.

References

1. Umar, Z.: Islamic vs conventional equities in a strategic asset allocation framework. Pac. Basin Financ. J. **42**, 1–10 (2015)
2. Jawadi, F., Jawadi, N., Louhichi, W.: Does Islamic finances outperform conventional finance? Further evidence from the recent financial crisis. Ipag business school research working paper, 279 (2014)
3. González, M.O., Papageorgiou, N.A., Skinner, F.S.: Persistent doubt: an examination of hedge fund performance. Eur. Financ. Manag. **22**(4), 613–639 (2016)
4. González, M.O., Skinner, F.S., Agyei-Ampomah, S.: Term structure information and bond investment strategies. Rev. Quant. Finan. Acc. **41**(1), 53–74 (2013)
5. Jensen, M.C.: The performance of mutual funds in the period 1945–1964. J. Financ. **23**(2), 389–416 (1968)
6. Treynor, J.L.: How to rate management of investment funds. Harv. Bus. Rev. **43**, 63–75 (1965)
7. Sharpe, W.F.: Mutual frund performance. J. Bus. **39**(1), 119–138 (1966)
8. Sortino, F., van der Meer, R.: Downside risk. J. Portf. Manag. **17**(4), 27–31 (1991)
9. Keating, C., Shadwick, W.F.: A Universal Performance Measure. The Finance Development Centre, London (2002)
10. Ingersoll, J., Spiegel, M., Goetzmann, W., Welch, I.: Portfolio performance manipulation and manipulation-proof performance measures. Rev. Financ. Stud. **20**(5), 1503–1546 (2007)

Do Google Trends Help to Forecast Sovereign Risk in Europe?

Marcos González-Fernández and Carmen González-Velasco

Abstract The aim of this paper is to analyze whether internet activity, as measured through Google data, influences the evolution of sovereign bond yields. For this purpose, we focus on ten European countries. We run VAR models and Granger causality tests between the Google Search Volume Index (GSVI) and sovereign risk. The VAR models and the causality tests for five core and five peripheral countries suggest that in the latter, especially Greece, Google data have had the highest positive impact on sovereign yields.

Keywords Google trends · Sovereign risk · Europe

1 Introduction

It is well known that over the past few years, Europe has faced a critical sovereign debt crisis that has had a great impact on sovereign bond yields and risk premiums, especially for peripheral countries, such as Greece, Ireland, Italy, Portugal and Spain. This sovereign debt crisis has been widely analyzed in the literature due to its great impact in financial markets and in the economic situation of most of European countries. With this paper which, to the best of our knowledge, is pioneering work in this line of research, along with the paper by Dergiades et al. [1] we attempt to determine whether internet activity is a useful tool to forecast changes in sovereign yields. If it is, it will represent a valuable tool that acts as a signal for ups and downs in yields and is helpful for financial market participants. To achieve this objective, we use Google data to analyze the impact of internet activity on sovereign bond yields.

For the purpose of our paper, we analyze 10 European countries using monthly data for the period between January 2004 and October 2017. We run vector

M. González-Fernández (✉) · C. González-Velasco
Universidad de León, León, Spain
e-mail: mgonf@unileon.es

M. Corazza et al. (eds.), *Mathematical and Statistical Methods for Actuarial Sciences and Finance*, https://doi.org/10.1007/978-3-319-89824-7_74

autoregressive models (VAR) and Granger causality, distinguishing between core and peripheral countries, to test the ability of Google data to predict yields. As a summary of the panel data results, we can state that VAR models indicate that this positive relationship is more intense for peripheral countries, especially Greece.

2 Background Literature

The use of internet search activity data, and especially Google data, is relatively recent in the literature, and the interest in this topic among both academics and professionals has been growing over the past decade. In economics and finance, there also multiple cases in which Google data have been used as a proxy in recent years [2–5], inter alia.

In relation to the aim of this paper, only a few papers analyze the impact of Google data on sovereign risk. To the best of our knowledge, only two papers in the literature address this subject. The first of them, Rose and Spiegel [6], analyzes dollar illiquidity during the global financial crisis. They use Google data for a selection of keywords related to the financial crisis, such as *crisis*, *financial* or *recession* and check whether it can be used as a proxy for default risk to include in their model. The second paper that addresses this issue is the one by Dergiades et al. [1]. They analyze the impact of Google searches on financial markets focusing on Europe's peripheral countries.

Following this line of research, our paper aims to contribute to the lack of literature in this area. We use a larger sample and a longer period: between 2004 and 2017. Our findings aim to provide more information about the importance of internet information in predicting sovereign risk and consequently about the usefulness of that tool to anticipate such a crisis.

3 Data

We use 10 year-sovereign bond yields obtained from Eurostat to proxy sovereign risk obtained from Eurostat database. They are calculated as monthly averages referred to central government bond yields on the secondary market. We have gathered the data for five core countries: Austria, Belgium, France Germany and Netherlands; and five peripheral countries: Greece, Ireland, Italy, Portugal and Spain. The other main variable in our analysis is the data related to Google searches. First, we have to select the keywords. Following Da et al. [7], we begin by downloading general terms such as "European debt crisis" and "sovereign risk", and we also download the top search terms provided by Google related to those keywords. We remove terms with too few observations and keep only those keywords with representative values throughout the time period. We select *debt crisis* as the most representative keywords for the aim of our analysis.

Subsequently we have downloaded Google data from Google trends tool (https://www.google.com/trends). It is worth noting that Google Trends does not provide the total number of searches for keywords; rather, it provides an index that ranges from 0 to 100, which is usually called Google Search Volume Index (GSVI hereinafter). To build this index, Google starts by dividing the number of searches for a given keyword into the total number of searches for a given time unit (daily, weekly or monthly). Thereby, a ratio is obtained that is subsequently normalized by multiplying it by a scaling factor $F = 100/r^*$, where r^* is the fraction of highest value [1]. Thus, the numbers start at 0 on January 2004, and subsequent values denote changes from the search on that date, 100 being the point at which the number of queries has achieved the top search intensity.

4 Results

In Table 1 the Granger causality tests for the bivariate models between GSVI for the keywords debt crisis and sovereign yields are summarized.

Table 1 Granger causality tests for (GSVI) and sovereign bond yields

Country	Null hypothesis	Lags	F-statistic	p-value	β coefficients
Austria	GSVI → yields	3	3.242	0.023**	−0.0018
	Yields → GSVI	3	1.746	0.159	0.1632
Belgium	GSVI → yields	3	1.406	0.243	−0.0012
	Yields → GSVI	3	0.978	0.404	0.2936
France	GSVI → yields	3	3.236	0.023**	−0.0009
	Yields → GSVI	3	2.099	0.102	0.1819
Germany	GSVI → yields	3	3.491	0.017**	−0.0013
	Yields → GSVI	3	1.435	0.234	0.1340
Netherlands	GSVI → yields	3	3.145	0.026**	−0.0016
	Yields → GSVI	3	1.190	0.315	0.1782
Greece	GSVI → yields	9	26.98	0.000***	0.0741
	Yields → GSVI	9	2.367	0.016**	1.2234
Ireland	GSVI → yields	6	9.227	0.000***	−0.0067
	Yields → GSVI	6	4.949	0.000***	0.8900
Italy	GSVI → yields	3	2.283	0.081*	0.0005
	Yields → GSVI	3	1.221	0.304	0.2737
Portugal	GSVI → yields	4	3.782	0.006***	0.0035
	Yields → GSVI	4	7.248	0.000***	1.2197
Spain	GSVI → yields	3	1.076	0.360	0.0002
	Yields → GSVI	3	1.064	0.365	0.4805

The table shows Granger causality tests for the bivariate VAR models between sovereign bond yields and the keywords *debt crisis*. The time horizon is January 2004–October 2017. The number of lags has been selected according to Akaike's criterion. The last column indicates the cumulative value of the coefficients in the VAR model for the lags specified
***1%; **5%; *10%

We focus first on peripheral countries and observe causality between GSVI and sovereign yields for all countries but Spain. Moreover, the GSVI causes yields, but a surge in the yields causes GSVI. Therefore, an increase in sovereign bond yields increases internet attention related to debt crisis. The cumulative effect is positive for peripheral countries, with the exception of Ireland. This indicates that a surge in GSVI in the previous months leads to a surge in bond yields in peripheral countries in the contemporaneous observation for the yields.

In core countries, we also find causality between GSVI and sovereign yields. Nevertheless, this causality is not reciprocal, since it works only in one direction that is; GSVI causes sovereign yields in four of the core countries. However, the cumulative effect here is negative in all of them. This indicates that an increase in internet attention in the previous months, measured through GSVI, leads to a decline in sovereign bond yields in month *t*. According to our results, it seems that internet attention causes a surge in yields in peripheral countries but a decline in the core countries. The reason for this different behavior can be found in the flight-to-liquidity hypothesis [8, 9]. When internet attention related to the debt crisis increases, as measured through GSVI, it leads to a surge in yields in peripheral countries, especially Greece. Thus, investors move their assets to safer and more liquid countries, such as Germany [9], and that generates a decline in the yields of core countries as they become safe havens.

5 Conclusions

In this paper, we attempt to determine empirically whether internet activity, measured through Google data, matters for the evolution of sovereign bond yields in Europe. We demonstrate that in core countries, yields are negatively influenced by the Google searches, whereas in peripheral countries, this negative impact does not appear, but a positive impact does. This may be because a surge in investors' concern about the sovereign debt crisis, which leads to an increase in internet attention, generates first a surge in the yields from countries more affected by the crisis and then a subsequent decline in core countries. This result confirms the existence of a flight to quality from peripheral to core sovereign bonds.

References

1. Dergiades, T., Milas, C., Panagiotidis, T.: Tweets, Google trends, and sovereign spreads in the GIIPS. Oxf. Econ. Pap. **67**(2), 406–432 (2015)
2. Ben-Rephael, A., Da, Z., Israelsen, R.D.: It depends on where you search: institutional investor attention and underreaction to news. Rev. Financ. Stud. **30**(9), 3009–3047 (2017)
3. Da, Z., Engelberg, J., Gao, P.: In search of attention. J. Financ. **66**(5), 1461–1499 (2011)
4. Siganos, A.: Google attention and target price run ups. Int. Rev. Financ. Anal. **29**, 219–226 (2013)

5. Vozlyublennaia, N.: Investor attention, index performance, and return predictability. J. Bank. Financ. **41**, 17–35 (2014)
6. Rose, A.K., Spiegel, M.M.: Dollar illiquidity and central bank swap arrangements during the global financial crisis. J. Int. Econ. **88**(2), 326–340 (2012)
7. Da, Z., Engelberg, J., Gao, P.: The sum of all FEARS investor sentiment and asset prices. Rev. Financ. Stud. **28**(1), 1–32 (2015)
8. Beber, A., Brandt, M.W., Kavajecz, K.A.: Flight-to-quality or flight-to-liquidity? Evidence from the Euro-area bond market. Rev. Financ. Stud. **22**(3), 925–957 (2009)
9. De Santis, R.A.: The euro area sovereign debt crisis: identifying flight-to-liquidity and the spillover mechanisms. J. Empir. Financ. **26**, 150–170 (2014)

The Contribution of Usage-Based Data Analytics to Benchmark Semi-autonomous Vehicle Insurance

Montserrat Guillen and Ana M. Pérez-Marín

Abstract Semi-autonomous vehicles will have a significant impact for the automobile insurance industry. We analyze telematics information and present methods for Usage-Based-Insurance to identify the effect of driving patters on the risk of accident. These results can be used as a starting point and a benchmark for addressing risk quantification and safety for semi-autonomous vehicles. Automatic speed control devices, which allow the driver to keep the vehicle at a predetermined constant speed and can ensure that the speed limit is not violated, could be considered a first example of semi-autonomy. We show scenarios for a reduction of speed limit violations and the consequent decrease in the expected number of accident claims. If semi-autonomous vehicles would completely eliminate the excess of speed, the expected number of accident claims could be reduced to almost one third its initial value in the average conditions of our data. We also note that an advantage of automatic speed control is that the driver does not need to look at the speedometer which may contribute to safer driving.

Keywords Telematics · Poisson model · Predictive modelling · Pricing

1 Introduction

Semi-autonomous vehicles are capable of sensing their environment and navigating without human input. One of the first examples of partial autonomous driving is an automatic speed control device that can be activated by the driver and then keeps the vehicle at a predetermined constant speed. The result is that this device ensures that the speed limit is not going to be violated. This paper is about the effect of speed reduction on the risk of accidents. We pre-sent a real case study where the impact of speed control is evaluated in different scenarios by using a pay-how-you-

M. Guillen (✉) · A. M. Pérez-Marín
Department of Econometrics, Riskcenter-IREA, University of Barcelona, Barcelona, Spain
e-mail: mguillen@ub.edu; amperez@ub.edu

M. Corazza et al. (eds.), *Mathematical and Statistical Methods for Actuarial Sciences and Finance*, https://doi.org/10.1007/978-3-319-89824-7_75

drive (PHYD) insurance database provided by a Spanish company. The effect of speed reduction on the expected number of accidents is found significant. Based on these data we also see that there are other factors that can also influence accident risk, but they are not discussed here. We only provide a reference for the role of speed reduction that can be taken as a reference and a baseline for the analysis of automated control on road safety. A recent article [9] provides a fast overview of usage-base-insurance (UBI) schemes and concludes that there is evidence that UBI implementation implies lower insurance costs for less risky and less exposed drivers. The authors claim that monitoring the performance of drivers and giving feedback is a motivation for some drivers to improve their skills. There are several authors, in quite some other recent articles, that advocate for the introduction of telematics in insurance. The results in [4] indicate that information on driving behavior (how-you-drive) can improve the risk selection process. Moreover, another approach described in [8] estimates aggregate claims losses for factors that can either be controlled by the driver or by the vehicle. The authors then show that automated features that remove driver error and reduce accident risks could imply a reduction of up to 90% of the costs. They largely base their conclusions on scenario analysis. Our contribution is centered on the role of speed control and it is based on the premise that automated procedures can reduce and eventually eliminate the violation of speed limits on the road. Based on real data we then calculate the reduction in the frequency of accidents and its impact on safety. We finish our presentation with a summary of the conclusions.

2 Data and Methods

The sample consists of 9614 drivers who had a PHYD insurance policy in force during 2010, the whole year. We model the total number of claims at fault from these drivers and show its association with the total distance traveled during the year in kilometers (km, mean = 13,063.71, standard deviation(sd) = 7715.80), percentage of kilometers traveled at speeds above the mandatory limits (speed, mean = 9.14, sd = 8.76), the percentage of kilometers traveled on urban roads (urban, mean = 26.29, sd = 14.18) and the age of the driver (age, mean = 24.78, sd = 2.82). The average number of claims per policy is 0.10, with a driver having had at most 3 claims. In Fig. 1 we have grouped the observations in subsets according to the percent distance driven above the speed limit. We see that the mean number of claims in-crease when the percent distance driven above the speed limit increases. There were some outliers in the original sample, which were removed here. We also limit our observation plot to the range from 0% to 20%. There is also some more variability when the percent is large, compare to lower values. This is due to the size of the groups. Since the average distance in the whole sample is 9.14%, we note that the median is 6.14%, which means that half of the drivers never driver more than 6.14% of the total distance driven in 1 year, above the speed limits. A classical Poisson model is fitted giving an excellent goodness-of-fit. The results are

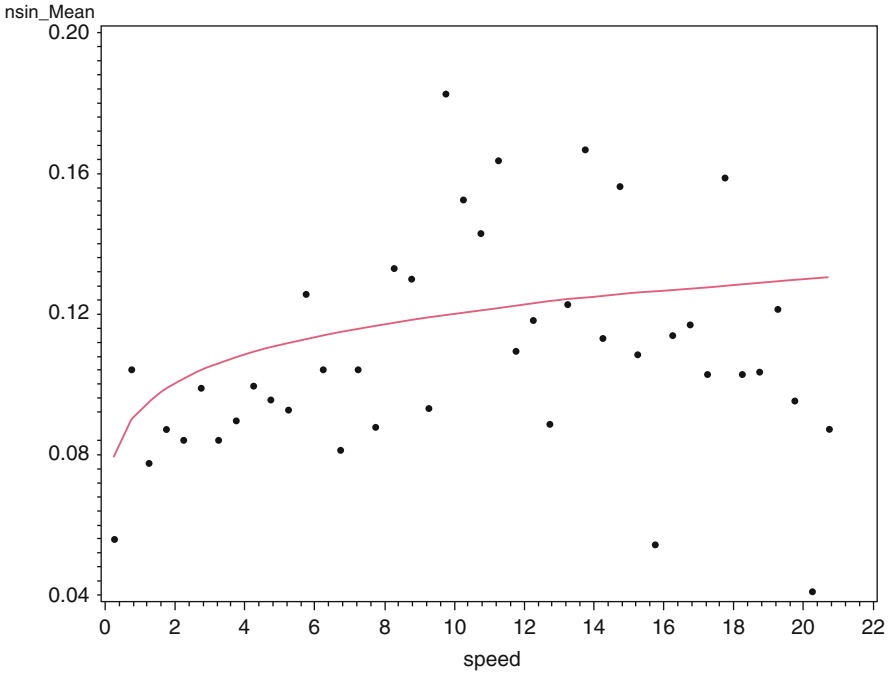


Fig. 1 The expected frequency of claims as a function of the percentage of kilometers traveled at speeds above the limit. The dots represent the average frequency of claims when the insureds are grouped according to their speed violations by intervals of 0.5%. The line represents the fitted claim frequency as a function of the percentage of kilometers traveled at speeds above the limit

available from the authors upon request and are not presented here for the sake of brevity. Similar analyses can be found in several papers [1–3, 6, 7]. They all show that telematics information can substantially complement classical insurance risk factors. They also show that one crucial driving characteristic that significantly influences the risk of an accident is the percent distance driven above the speed limit. Note that in this paper we consider that all claims are reported so that the risk of an accident and the risk of a claim are equivalent (see [5]).

3 Speed Reduction and Safety

We have compared what happens if there is a change in the average number of kilometers driven above the speed limits. Table 1 shows how the expected number of claims changes per 1000 drivers. The rows indicate the initial state and the columns the final state. Thus, the last row of the first column means that an insurer would expect 87 less claims for 1000 drivers that exceed their speed limit 20% of the distance driven if they reduced this speed limit violation to zero.

Table 1 Scenarios of the number of claims reduction per 1000 drivers

	0	1	5	10	15	20
0	0.0	50.1	68.5	77.5	83.1	87.3
1	−50.1	0.0	18.4	27.4	33.0	37.1
5	−68.5	−18.4	0.0	9.0	14.6	18.7
10	−77.5	−27.4	−9.0	0.0	5.6	9.7
15	−83.1	−33.0	−14.6	−5.6	0.0	4.1
20	−87.3	−37.1	−18.7	−9.7	−4.1	0.0

Row values indicate the initial distance (%) and column values are the final distance (%) driven with excess speed

4 Conclusions

Our real data have allowed to producing some scenarios for a reduction of speed limit violations and its impact on the decrease in the expected number of accident claims. The reduction in the number of claims per 1000 drivers is substantial and it is larger obviously if the initial state refers to a set of drivers that has a high percent of distance driven above the speed limit, compared to those that have a smaller initial value. If semi-autonomous vehicles could completely eliminate the possibility that drivers could exceed speed limits, or reduce this to a minimum, then the expected number of accident claims would be reduced. The benefits of this reduction would translate to a reduction in the number of victims on the road and an increase of overall safety.

Acknowledgements We thank support from the Spanish Ministry of Economy and Competitiveness and FEDER grant ECO2016-76203-C2-2-P.

References

1. Ayuso, M., Guillen, M., Pérez-Marín, A.M.: Time and distance to first accident and driving patterns of young drivers with pay-as-you-drive insurance. Accid. Anal. Prev. **73**, 125–131 (2014)
2. Ayuso, M., Guillen, M., Pérez-Marín, A.M.: Telematics and gender discrimination: some usage-based evidence on whether men's risk of accidents differs from women's. Risks **4**(2), 1–10 (2016)
3. Ayuso, M., Guillen, M., Pérez-Marín, A.M.: Using GPS data to analyse the distance travelled to the first accident at fault in pay-as-you-drive insurance. Transp. Res. Part C: Emerg. Technol. **68**, 160–167 (2016)
4. Baecke, P., Bocca, L.: The value of vehicle telematics data in insurance risk selection pro-cesses. Decis. Support. Syst. **98**, 69–79 (2017)
5. Boucher, J.P., Denuit, M., Guillen, M.: Number of accidents or number of claims? an approach with zero-inflated Poisson models for panel data. J. Risk Insur. **76**(4), 821–846 (2009)
6. Boucher, J.P., Pérez-Marín, A.M., Santolino, M.: Pay-as-you-drive insurance: the effect of the kilometers on the risk of accident. Anales del Instituto de Actuarios Espaoles, 3a Epoca **19**, 135–154 (2013)

7. Boucher, J.-P., Còté, S., Guillen, M.: Exposure as duration and distance in telematics motor insurance using generalized additive models. Risks **5**(4), 54 (2017)
8. Sheehan, B., Murphy, F., Ryan, C., Mullins, M., Liu, H. Y.: Semi-autonomous vehicle mo-tor insurance: a Bayesian Network risk transfer approach. Transp. Res. C **82**, 124–137 (2017)
9. Tselentis, D.I., Yannis, G., Vlahogianni, E.I.: Innovative motor insurance schemes: a re-view of current practices and emerging challenges. Accid. Anal. Prev. **98**, 139–148 (2017)

Some Empirical Evidence on the Need of More Advanced Approaches in Mortality Modeling

Asmerilda Hitaj, Lorenzo Mercuri, and Edit Rroji

Abstract Recent literature on mortality modeling suggests to include in the dynamics of mortality rates the effect of time, age, the interaction of the latter two terms and finally a term for possible shocks that introduce additional uncertainty. We consider for our analysis models that use Legendre polynomials, for the inclusion of age and cohort effects, and investigate the dynamics of the residuals that we get from fitted models. Obviously, we expect the effect of shocks to be included in the residual term of the basic model.

The main finding here is that there is persistence in the residual term but the autocorrelation structure does not display a negative exponential behavior. This empirical result suggests that the inclusion of the additional shock term requires an appropriate model that displays a more flexible autocorrelation structure than the Ornstein-Uhlenbeck employed in existing models.

Keywords Legendre polynomials · Mortality shocks · Autocorrelation

1 Mortality Models Based on Legendre Polynomials

We review in this section the model proposed in [5] and its extensions available in literature. The force of mortality, $\mu(x,t)$ for age x in calendar year t is modeled as:

$$\mu(x,t) = \exp\left[\beta_0 + \sum_{j=1}^{s}\beta_j L_j(x') + \sum_{i=1}^{r}\alpha_i t'^i + \sum_{i=1}^{\bar{r}}\sum_{j=1}^{\bar{s}}\gamma_{ij} L_j(x') t'^i\right] \quad (1)$$

A. Hitaj · E. Rroji (✉)
University of Milano-Bicocca, Milano, Italy
e-mail: asmerilda.hitaj1@unimib.it; edit.rroji@unimib.it

L. Mercuri
University of Milan, Milano, Italy
e-mail: lorenzo.mercuri@unimib.it

M. Corazza et al. (eds.), *Mathematical and Statistical Methods for Actuarial Sciences and Finance*, https://doi.org/10.1007/978-3-319-89824-7_76

where $x \in [x_1, x_2]$, $t \in [y_1, y_2]$, $\bar{s} \leq s$, $\bar{r} \leq r$ and $L_i(x')$ are the Legendre polynomials whose elements are computed, for $n \geq 1$ as:

$$L_0(x) = 1; \quad L_1(x) = x; \quad (n+1)L_{n+1}(x) = (2n+1)xL_n(x) - nL_{n+1}(x). \quad (2)$$

The quantities x' and t' are the transformed ages and the transformed calendar years that map x and t onto the interval $[-1, 1]$ following the rule:

$$x' = \frac{2x - \max(x) - \min(x)}{\max(x) - \min(x)} \quad t' = \frac{2t - \max(t) - \min(t)}{\max(t) - \min(t)}. \quad (3)$$

The force of mortality given in (1) depends on three terms:

1. $\sum_{j=1}^{s} \beta_j L_j(x')$ which captures the age effect,
2. $\sum_{i=1}^{r} \alpha_i t'^i$ that captures the time effect
3. $\sum_{i=1}^{\bar{r}} \sum_{j=1}^{\bar{s}} \gamma_{ij} L_j(x') t'^i$ that captures the interaction between age and time effects.

In [2] an additional autoregressive term with gaussian innovations is added while [1] considers other type of innovations. In order to get the generalized approach proposed in [1] we should denote with $D(x, t)$ the number of deaths for people aged x in the calendar year t assumed to be a Poisson response variable with mean:

$$\mathsf{E}[D(x,t)] = r(x,t)\mu(x,t) \quad (4)$$

where $r(x, t)$ is the central exposure to risk that refers to the number of people alive aged x in the middle of the calendar year t. The main assumption states that:

$$\mathsf{E}[D(x,t)\,|Y_{t'}] = r(x,t)\exp\left[\beta_0 + \sum_{j=1}^{s}\beta_j L_j(x') + \sum_{i=1}^{r}\alpha_i t'^i + \sum_{i=1}^{\bar{r}}\sum_{j=1}^{\bar{s}}\gamma_{ij}L_j(x')t'^i + Y_{t'}\right], \quad (5)$$

$Y_{t'}$ is assumed to be a mean-reverting process whose dynamics is the solution of the SDE:

$$\begin{cases} \mathrm{d}Y_{t'} = -aY_{t'}\mathrm{d}t' + \mathrm{d}Z_{t'} \text{ for } t' > -1 \\ Y_{-1} = 0, \end{cases} \quad (6)$$

where a is an unknown constant and $(Z_{t'} : t' \geq -1)$ is a Lévy process with initial value $Z_{-1} = 0$. Ahmadi and Gaillardetz [1] consider the class of Tempered α- Stable subordinators among positive Lévy processes. In particular, they obtain

closed form formulas in the case where $Y_{t'}$ is a Variance Gamma Ornstein-Uhlenbeck solution of Eq. (6) where $(Z_{t'} : t' \geq -1)$ is a Variance Gamma process. It seems now obvious that the extensions of model in (1) in [2] and [1] are both based on the idea that the additional term has an exponentially decaying autocorrelation structure associated to the choice of an Ornstein-Uhlenbeck process.

2 Empirical Investigation

We want to investigate empirically the soundness of the choice of the Ornstein-Uhlenbeck for the additional term that should be added to model (1) in order to capture mortality shocks. For our analysis we employ data from USA that we downloaded from the Human Mortality Database website [4]. In particular we consider the male tables for the number of deaths and exposure to risk. Data are collected yearly for $t = 1947, \ldots, 2000$. We consider ages from 65 to 100 since for this range modeling of future mortality rates is crucial for the pricing of life products that refer to the post-retirement period.

We recall that m_{xt} is the central death rate whose empirical estimator is $\hat{m}_{xt} = \frac{d(x,t)}{r^c(x,t)}$. We will assume throughout the paper that $\mu(xt) \cong m_{xt}$ and consider a constant mortality rate over a calendar year for each fixed age x. The estimation procedure is based on the maximization of the quasi-likelihood function as described in [5] while the selection model is obtained using the Likelihood Ratio Test (LR-Test).

First we report in Table 1 the fitted parameters and then for each fixed age x we get the sequence $\hat{Y}_t = \log \mu(xt) - \log \hat{\mu}(xt)$ from the fitted mortality rates $\hat{\mu}(xt)$.

The time series of residual terms for ages 65, 66, 67 and 68 are plotted in Fig. 1. For all considered ages, the residuals seem to move together and thus we are comfortable with the idea that the shock term depends only on time as suggested in the models in [2] and [1]. In Fig. 2 we observe that the autocorrelation structure suggests persistence but it does not display an exponential decay behavior. This finding suggests that the Ornstein-Uhlenbeck considered in [2] and [1] is not suitable for the description of the term that captures the shock dynamics. Non-monotonocity in the autocorrelation function paves the way to the use of models that can reproduce more flexibility. For example, as shown in [3], CARMA models are able to deal with this kind of autocorrelation structure.

Table 1 Fitted parameters of the model for US male tables for the period 1947–2000 for ages ranging in the interval [65:100]

α_1	α_2	α_3	α_4	α_5	α_6	α_7	α_8	α_9	α_{10}	α_{11}	α_{12}
−0.440	−0.392	2.698	2.583	−10.037	−10.086	17.769	18.289	−14.985	−14.987	4.820	4.544
β_0	β_1	β_2	β_3	β_4	β_5	β_6	β_7				
−2.0746	1.3134	−0.0695	−0.0653	−0.0170	0.0015	−0.0109	−0.0235				
$\gamma_{1,2}$	$\gamma_{1,1}$	$\gamma_{2,1}$	$\gamma_{2,2}$								
0.1399	0.2028	0.0174	0.0297								
loglik											
53637482.5											

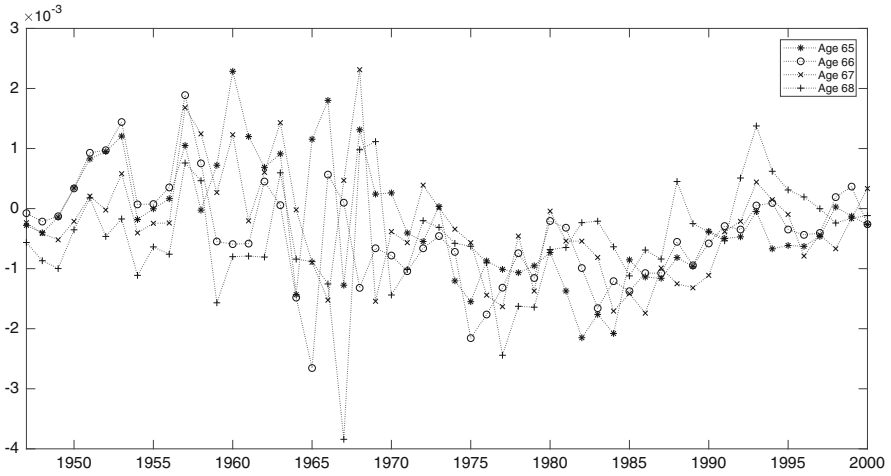


Fig. 1 Time series of residual term Y_t in the dynamics of $\mu(x,t)$ for ages 65, 66, 67 and 68

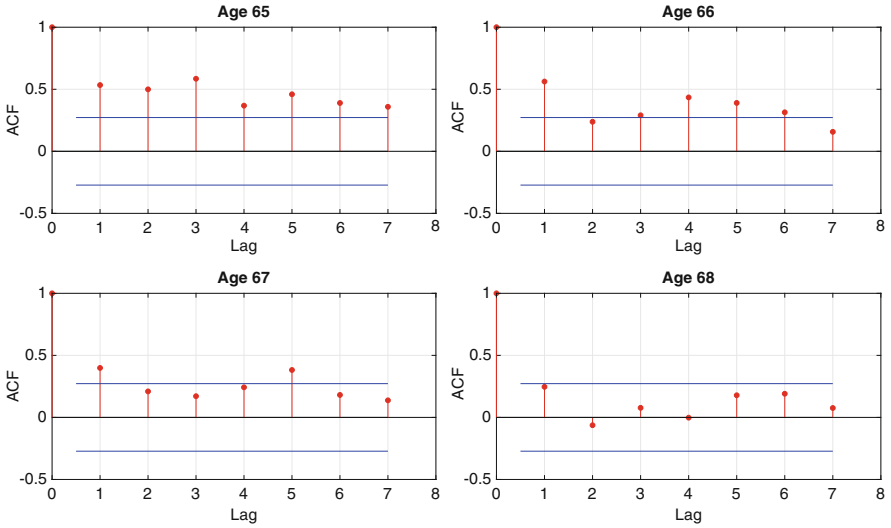


Fig. 2 Autocorrelation function for the residual terms of $\mu(x,t)$ for ages 65, 66, 67 and 68

References

1. Ahmadi, S.S., Gaillardetz, P.: Modeling mortality and pricing life annuities with Lévy processes. Insur. Math. Econ. **64**, 337–350 (2015)
2. Ballotta, L., Haberman, S.: The fair valuation problem of guaranteed annuity options: the stochastic mortality environment case. Insur. Math. Econ. **38**, 195–214 (2006)
3. Brockwell, P.J.: Representations of continuous-time ARMA processes. J. Appl. Prob. **41**, 375–382 (2004)
4. Human Mortality Database: University of California, Berkeley (USA), and Max Planck Institute for Demographic Research (Germany). www.mortality.org or www.humanmortality.de
5. Renshaw, A.E., Haberman, S., Hatzoupoulos, P.: The modelling of recent mortality trends in United Kingdom male assured lives. Brit. Actuar. J. **2**, 449–477 (1996)

Could Machine Learning Predict the Conversion in Motor Business?

Lorenzo Invernizzi and Vittorio Magatti

Abstract The aim of the paper is to estimate the *Conversion Rate* by means of three Machine Learning (*ML*) algorithms: *Classification and Regression Tree* (*CART*), *Random Forest* (*RF*) and *Gradient Boosted Tree* (*BOOST*). The *Generalized Linear Model* (*GLM*), benchmark model in the framework, is used as frame of reference.

The *RF* model has the highest *Recall*, while the *BOOST* is the most precise model. The *RF* is able to outperform the *GLM* benchmark model in terms of *Log Loss error*, *Precision*, *Recall* and *F Score*.

Variable Importance and *Strength* index, computed from the *ML* models and the *GLM* respectively, describe how the different algorithms are coherent on choosing the most relevant features.

Keywords Machine learning · Conversion rate · Generalized linear model · CART · Random forest · Gradient boosting · Motor insurance

1 Introduction

The *Conversion Rate* is defined as the ratio between insurance policies and quote request. A good prediction of it produces at least two main advantages for an Insurer:

- *Increase in Competitiveness*: this is especially important when the underwriting cycle shows a softening period;

L. Invernizzi (✉)
Zurich Insurance Company Ltd, Zürich, Switzerland
e-mail: lorenzo.invernizzi@it.zurich.com

V. Magatti
Sapienza Universita' di Roma, Roma, Italy
e-mail: vittorio.magatti@uniroma1.it

M. Corazza et al. (eds.), *Mathematical and Statistical Methods for Actuarial Sciences and Finance*, https://doi.org/10.1007/978-3-319-89824-7_77

- *Effective price changes*: a Company could identify rate changes or dedicated discounts coherently with the estimated conversion and profitability calculated for each potential client, both needed to develop a pricing optimisation tool.

The short paper starts with the description of the data set, then the *ML* models, the calibration techniques and results & conclusions are presented.

2 Description of the Data Set

The selected perimeter is the *Motor Third Party Liability* (*MTPL*) for cars. It is based on a real data set provided by *Zurich Insurance Company Ltd*. The train set is composed by 205'439 quote requests and the 1.325% of them are converted into policies, while the test set consists of 51'739 quote requests with a probability of conversion of 1.289%.

For each quote request 26 features are considered. These are the most common features known in the insurance market to affect client's choices. Among them there are: *premium range, age of the client, power-to-weight ratio, Bonus Malus class, engine power, vehicle age, years of car possession, age of vehicle when bought, client's occupation, guide style, age patent qualification, housing density, horse powers, Italian region, number of non insured years, marital status, fuel source and title study.*

3 Machine Learning Models

Below the models under analysis are introduced, highlighting their main properties.

3.1 Generalized Linear Model (GLM)

The *GLM* represents the state of the art algorithm extensively used to predict the conversion rate. The Binomial family distribution is considered as the error distribution, in association with its canonical Logit link function. See Chapter 2 of [5] for a thorough discussion of the statistical model.

3.2 Classification and Regression Tree (CART)

Let $\{A_i\}_i$ be a partition of the 26 dimensional space of the features, the *CART* is defined as a linear combination of indicator functions

$$CART(x) = \sum_i c_i \mathbf{1}\{x \in A_i\} \tag{1}$$

that is equivalent to the tree structure. The model fits by minimizing a specified loss function and is able to capture non-linear and complex relationships. In contrast there is a high risk of overfitting. See Section 9.2 of [4] for more details.

3.3 Random Forest (RF)

The *Random Forest* consists of an average of K *CART* models

$$RF(x) = \frac{\sum_k CART_k(x)}{K} \tag{2}$$

where each *CART* is estimated by means of two random effects: Bootstrapping ($\sim$70%) and Feature Bagging ($\sim\lfloor\sqrt{26}\rfloor$). Both hyperparameters, important in preventing overfitting, are subject to fine tuning. See [1] for a complete argumentation.

3.4 Gradient Boosted Tree (BOOST)

The *Gradient Boosted Tree* is defined as a linear combination of *CART* models

$$BOOST(x) = sigmoid\Big(\sum_k \alpha_k CART_k(x)\Big) \tag{3}$$

where the sigmoid function is used to map the combination of *CARTs* into [0, 1]. The model is estimated using the Gradient Boosting Algorithm. See [2] for more details.

4 Models Calibration

The fine tuning of the hyperparameters is the most crucial aspect for obtaining accurate predictions on the test set.

The *GLM* is estimated using the *Emblem Software* through a stepwise procedure consisting of: correlation analysis, χ^2 tests, Forward Stepwise Regression, introduction of Custom Factors, Variates, Interactions and test of time consistency. The Iterative Weighted Least Squares Algorithm is the fitting algorithm.

The *CART*, *RF* and *BOOST* are estimated using the *h2o Library*, interface in *R*.

The stratified 5 cross-validation technique is exploit to tune the number of trees K for the *RF* and *BOOST*. The *RF* implements the stratified bootstrapping. The *BOOST* uses the tuned $learn\_rate = 0.01$.

To reduce the observed overfitting of the *BOOST*, the Bootstrapping (as in *RF*) is introduced (see [3]), obtaining no improvements. While the introduction of the Feature Bagging results in a higher precision on the test set, maintaining the same level of recall. Hence it is considered in the *BOOST*.

The *CART*, performing poorly, is not considered in the analysis of the results.

5 Analysis of Results

All models are estimated varying the depth of trees from 1 to 11[1] and the optimal depth is chosen by maximizing the *F Score*, maintaining a good trade-off between *Precision* and *Recall*.

Figures 1 and 2 show *Precision* and *Recall* varying the depth. The optimal depths for *RF* and *BOOST* are 10 and 6 respectively. Table 1 compares the measures of comparison of the selected models. *RF* is the model with the highest recall, while *BOOST* is the most precise. *RF* is able to outperform the *GLM* benchmark model.

Being *BOOST* iteratively fit on residuals, the optimal one is composed of a linear combination of many shallow trees.

Defined the *Strength* Index for each feature (*GLM*) as the normalized weighted average of the βs coefficients, and the *Variable Importance* (see [1]), all models choose the *premium range* and the *age of the client* as the most predictive features.

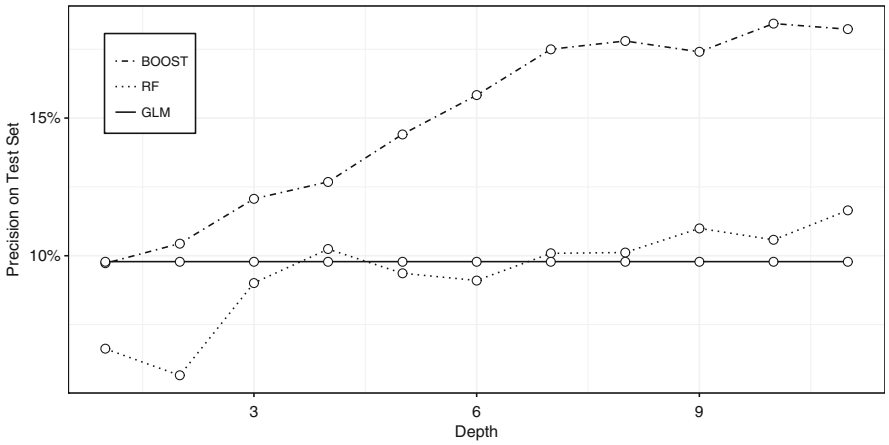


Fig. 1 *BOOST* is the most precise model as expected, the precision increases as the depth, making the model hard to tune due to overfitting problems. *RF* beats *GLM* for depths greater than 6

[1] Due to computational machine limits. From the results we conclude that it is sufficient.

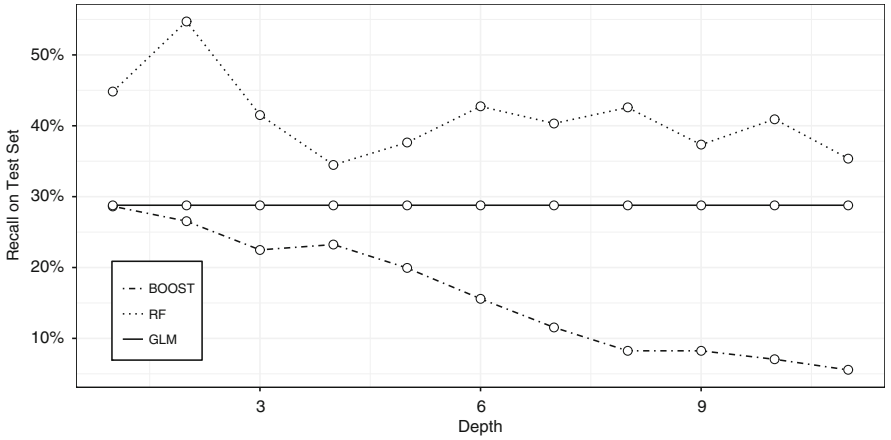


Fig. 2 *RF* is the model with the highest recall. *BOOST* performs poorly, in contrast to a high value of precision. Hence *BOOST* is a more conservative model than *RF* and *GLM*

Table 1 Performance of the selected models on test set

Model	*Depth*	*Log Loss Error*	*Accuracy* (%)	*Precision* (%)	*Recall* (%)	*F Score* (%)
GLM	n.d.	0.0137	95.6	9.8	28.8	14.6
RF	10	0.0135	94.8	10.6	41.0	16.8
BOOST	6	0.0135	97.8	15.9	15.6	15.7

References

1. Breiman, L.: Random forests. J. Mach. Learn. Arch. **45**(1), 5–32 (2001)
2. Friedman, J.: Greedy function approximation: a gradient boosting machine. Ann. Stat. **29**(5), 1189–1232 (2001)
3. Friedman, J.: Stochastic gradient boosting. J. Comput. Stat. Data Anal. Nonlinear Methods Data Min. Arch. **38**(4), 367–378 (2002)
4. Hastie, T., Tibshirani, R., Friedman, J.: The Elements of Statistical Learning: Data Mining, Inference, and Prediction. Springer Series in Statistics. Springer, New York (2009)
5. McCullagh, P., Nelder, J.A.: Generalized Linear Models, 2nd edn. Chapman & Hall/CRC Monographs on Statistics & Applied Probability. Chapman & Hall/CRC, Boca Raton (1989)

European Insurers: Interest Rate Risk Management

Francisco Jareño, Marta Tolentino, María de la O González, and María Ángeles Medina

Abstract This paper studies the interest rate risk of some relevant European insurers during the period 2003–2015, using the Quantile Regression (QR) methodology and including the state of the economy. The results show that, in general, the European insurers' returns have a statistically significant sensitivity to interest rates, although there are relevant differences between the different companies analyzed, the different subperiods and between quantiles. Thus, the sensitivity of the European insurers to movements in the European interest rates tends to be more pronounced in extreme market conditions (with upward or downward fluctuations).

Keywords Interest rate risk management · European insurers · Stock market

1 Introduction and Methodology

1.1 Aim

This research focuses on analysing the sensitivity and behaviour of some some relevant European insurers to changes in benchmark interest rates. The methodology used is the Quantile Regression (QR) approach and the period analysed covers from 2003 to 2015. In addition, this research analyzes this sensitivity, decomposing the

F. Jareño (✉) · M. de la O González
University of Castilla-La Mancha, Albacete, Spain
e-mail: Francisco.Jareno@uclm.es; Marta.Tolentino@uclm.es

M. Tolentino
University of Castilla-La Mancha, Ciudad Real, Spain
e-mail: Mariao.Gonzalez@uclm.es

M. Á. Medina
University of Alcalá, Alcalá de Henares, Spain

M. Corazza et al. (eds.), *Mathematical and Statistical Methods for Actuarial Sciences and Finance*, https://doi.org/10.1007/978-3-319-89824-7_78

entire period into different subsamples depending on the economic cycle, which allows to obtain better results according to the nature of the methodology used (QR).

1.2 Data and Methodology

This paper uses the stock prices of eight insurance companies listed on different stock markets in the Eurozone: Spain, Germany, France, Italy, the Netherlands and Switzerland. The sample period ranges from January 2003 to December 2015. The data have been obtained from the website "investing.com" using the closing prices to calculate the monthly stock returns as the relative change in the closing price for the first trading day observed for each month between two consecutive months.

In turn, the market portfolio is approached through the EuroStoxx 50 index as it is the benchmark in the Eurozone.[1] The series of the monthly market returns has been obtained as the weighted arithmetic average of the monthly stock returns of the member companies. In order to measure the sensitivity of the insurance companies' returns to changes in interest rates, the 10-year German bond yields are used as proxy variables.[2] Germany has been chosen as the reference country since it is considered the strongest, largest and most stable economy in the Eurozone. It has been decided to use a long-term interest rate, since stock prices are usually linked to the forecasts of the level of long-term interest rates and therefore company returns tend to be more sensitive to changes in long-term interest rates than to changes in short-term rates [2].

The main features of the European insurers are shown in Table 1:

Table 1 Main features of the European insurers

	Country	Year founded	Area	2016 Premium volume
Generali	Italy	1831	Life and non-life	63.837[a]
Swiss Re	Switzerland	1863	Life and non-life	17.768[a]
Catalana occidente	Spain	1864	Life and non-life	4.235[a]
Munich Re	Germany	1880	Life and non-life	48.900[a]
Allianz	Germany	1890	Life and non-life	76.331[a]
Mapfre	Spain	1933	Life and non-life	22.813[a]
AXA	France	1982	Life and non-life	94.220[a]
Aegon	Holland	1983	Life	23.453[a]

Source: Own preparation
[a]Data expressed in millions of euros

[1]Munich Re, Axa and Allianz are included in the EuroStoxx 50 index, but they represent a low percentage of the index.

[2]According to Sevillano and Jareño [1], this paper tries to study the impact of this international factor. Another option would be to include the risk premium for non-German companies.

The Quantile Regression methodology (QR) is used in this research. This approach was introduced by Koenker and Basset [3] as an alternative method of classical OLS estimation [4]. The QR approach is an estimation method based on the minimization of absolute deviations weighted with asymmetric weights that are not affected by extreme data. The QR methodology is used when the main objective of the study focuses on the estimation of different percentiles (quantiles) of a particular sample, since this technique allows to estimate different quantile functions of the conditional distribution. Each quantile function characterizes a particular point of the conditional distribution, so that by combining different quantile regressions a more complete distribution of the underlying conditional distribution is obtained [5].

The estimation of the parameters in the case of the Quantile Regression is carried out through the minimization of the absolute deviations weighted with asymmetric weights:

$$\left[\sum_{i:y_i \geq x_i'\beta_i} \theta \left|y_i - x_i'\beta\right| + \sum_{y_i < x_i'\beta_i} (1-\theta) \left|y_i - x_i'\beta\right|\right] \tag{1}$$

According to Jareño [6], González et al. [7] and Jareño et al. [8], among others, this research applies the two-factor model introduced by Stone [9]. Therefore, two explanatory factors are included: (1) the stock market return, and (2) changes in interest rates.

The two-factor asset pricing model consists of adding an interest rate change factor as an additional explanatory variable to the CAPM (Capital Asset Pricing Model). Therefore, the proposed version in this paper of the Stone two-factor model would be as follows:

$$R_{it} = \alpha_i + \beta_1 R_{mt} + \beta_2 R_t + \varepsilon_{it} \tag{2}$$

where R_{it} is the return of insurer i in time t, R_{mt} is the stock market return in t, R_t shows changes in interest rates, and, finally, ε_{it} is a random error.

The inclusion of the market portfolio is to control macroeconomic factors that impact on stock prices and correlate with changes in interest rates. The coefficient associated with the stock market return, β_1, measures the sensitivity of corporate returns to general changes in the stock market return and, therefore, is an indicator of market risk. In turn, the coefficient of changes in interest rates, β_2, reflects the sensitivity of insurers' returns to fluctuations in interest rates beyond that contained in the market return. Therefore, it can be interpreted as a measure of the exposure to interest rates [2].

2 Main Results and Concluding Remarks

The results show that, in general, the European insurers' returns have a statistically significant sensitivity to interest rates, although there are relevant differences between the different companies analyzed, the different subperiods and between quantiles.

The company that shows the highest sensitivity to interest rate changes in all periods is Munich Re (from Germany). However, the sign of the coefficients is different for the two subperiods analyzed. In the period of economic expansion (2003–2007) the coefficients are positive, unlike in the whole period (2003–2017) and in the crisis period (2008–2017), which means that variations in European interest rates result in a negative effect on this company returns.

In the whole sample period the coefficients of three out of the eight companies analyzed (Allian, Munich RE and Swiss Re) have negative signs in all the quantiles studied. At this stage, most of the coefficients that are statistically significant are found at the ends of the distribution, especially at the lowest quantiles (0.1 and 0.2), related to crisis periods.

In the first crisis subperiod, companies that show statistically significant sensitivity to changes in European interest rates (Allianz, Axa, Mapfre and Aegon) are characterized by showing positive coefficients in the last quantiles of the distribution, which shows that the effect of changes in this risk factor is more intense in good stock market performance scenarios [8].

In the second subperiod coefficients that show a statistically significant sensitivity are found mainly in the lowest quantiles of the distribution. The coefficients that are significant in the highest quantiles have a negative sign, indicating that an increase in interest rates would lead to a decrease in the European insurer returns and would show the significant impact that the crisis has had in this sector.

In general, for the different companies analyzed in the insurance sector, the sensitivity to movements in the European interest rates tends to be more intense in extreme market conditions, that is, when the stock market is experiencing significant upward or downward fluctuations. This relevant result has been detected due to the use of the Quantile Regression (QR) methodology, which helps reveal the relationships between the European interest rates and the European insurer returns that could not have been detected using more traditional techniques such as the OLS methodology [8].

References

1. Sevillano, M.C., Jareño, F.: The impact of international factors on spanish company returns: a quantile regression approach. Risk Manage. **20**(1), 51–76 (2018)
2. Ferrando, L., Ferrer, R., Jareño, F.: Interest rate sensitivity of Spanish industries: a quantile regression approach. Manch. Sch. **85**(2), 212–242 (2017)
3. Koenker, R., Basset, G.: Regression quantiles. Econometrics. **46**(1), 33–50 (1978)

4. Xiao, Z.: Quantile cointegrating regression. J. Econ. **150**(2), 248–260 (2009)
5. Koenker, R.: Quantile Regression. Econometric Society Monographs. Cambridge University Press, New York (2005)
6. Jareño, F.: Sensibilidad de los rendimientos sectoriales a tipos de interés reales e inflación. Investig. Econ. **30**(3), 577–610 (2006)
7. González, M., Jareño, F., Skinner, F.: Interest and inflation risk: investor behavior. Front. Psychol. **7**(390), 1–18 (2016)
8. Jareño, F., Ferrer, R., Miroslavova, S.: US stock market sensitivity to interest and inflation rates: a quantile regression approach. Appl. Econ. **48**(26), 1–13 (2016)
9. Stone, B.K.: Systematic interest-rate risk in a two-index model of returns. J. Financ. Quant. Anal. **9**(5), 709–721 (1974)

Estimation and Prediction for the Modulated Power Law Process

Alicja Jokiel-Rokita and Ryszard Magiera

Abstract The modulated power law process has been proposed by Lakey and Rigdon in 1992 as a compromise between the non-homogeneous Poisson process and the renewal process model. It is useful in analyzing duration dependence in economic and financial cycles. In this paper we consider a problem of estimation and prediction for the modulated power law process. Using the estimating functions approach we propose new estimators of the parameters of the modulated power law process. The estimators proposed we apply to construct predictors of the next event time. We also present algorithms for effective calculating the values of estimators and predictors proposed. In the simulation study conducted we compare the accuracy of the estimators proposed with the maximum likelihood ones and examine the precision of predictors presented. The results obtained we apply in analysing a real data set of U.S. stock market cycles.

Keywords Inhomogeneous gamma process · Maximum likelihood estimation · Estimating equation · Stock market cycle

1 Introduction

The modulated power law process (MPLP) has been proposed in [3] as a compromise between the non-homogeneous Poisson process and the renewal process model. A definition of the MPLP is the following. Suppose that events (shocks) occur according to the non-homogeneous Poisson process (NHPP) with power law intensity $\lambda(t) = \alpha\beta t^{\beta-1}$, where $\alpha > 0$, $\beta > 0$, $t \geq 0$, and suppose that a system failure occurs not at every shock but at every κth shock, where (for now) κ is a positive integer. Let T_i be the time of the ith failure, $i = 1, 2, \ldots$.

A. Jokiel-Rokita (✉) · R. Magiera
Wrocław University of Science and Technology, Wrocław, Poland
e-mail: alicja.jokiel-rokita@pwr.edu.pl; ryszard.magiera@pwr.edu.pl

M. Corazza et al. (eds.), *Mathematical and Statistical Methods for Actuarial Sciences and Finance*, https://doi.org/10.1007/978-3-319-89824-7_79

The observed sequence $\{T_i, i = 1, 2, \ldots\}$ of occurrence times $T_1, T_2, \ldots$ forms a point process which is called the MPLP. It will be denoted by MPLP(α, β, κ). The MPLP(α, β, κ) is also defined for any positive value of κ. Namely, if the random variables $W_i = \alpha T_i^{\beta} - \alpha T_{i-1}^{\beta}$, $i = 1, 2, \ldots$, where $T_0 = 0$, are independent and identically gamma distributed with unknown shape parameter κ and unit scale parameter, then the sequence $\{T_i, i = 1, 2, \ldots\}$ forms the MPLP(α, β, κ).

The MPLP is often used to model failure data from repairable systems, when both renewal type behaviour and time trend are present. It is also useful in analyzing duration dependence in economic and financial cycles (see for example [4] and [5]).

In this paper we consider a problem of estimation and prediction for the MPLP. We propose a new method of estimation of the MPLP parameters and prediction of the next event time. The results obtained we apply in analysing a real data set of U.S. stock market cycles.

2 Estimation of the MPLP Parameters

If we observe the MPLP up to nth event time, then the log-likelihood function of the process is

$$l(\alpha, \beta, \kappa \mid t_1, \ldots, t_n) = -n \log[\Gamma(\kappa)] + n\kappa \log(\alpha) + n \log(\beta) - \alpha t_n^{\beta} + (\beta - 1) \sum_{i=1}^{n} \log(t_i) + (\kappa - 1) \sum_{i=1}^{n} \log(t_i^{\beta} - t_{i-1}^{\beta}).$$

The likelihood equations in the model considered are

$$\begin{cases} \partial l/\partial \alpha = n\kappa/\alpha - t_n^{\beta} = 0, \\ \partial l/\partial \beta = n/\beta + S(\mathbf{t}) + (\kappa - 1) W(\mathbf{t}, \beta) - \alpha t_n^{\beta} \log(t_n) = 0, \\ \partial l/\partial \kappa = -n\psi(\kappa) + n \log(\alpha) + V(\mathbf{t}, \beta) = 0, \end{cases} \tag{1}$$

where $\psi(\cdot) = \Gamma'(\cdot)/\Gamma(\cdot)$ is the di-gamma function, $S(\mathbf{t}) = \sum_{i=1}^{n} \log(t_i)$,

$$W(\mathbf{t}, \beta) = \sum_{i=2}^{n} \frac{t_i^{\beta} \log(t_i) - t_{i-1}^{\beta} \log(t_{i-1})}{t_i^{\beta} - t_{i-1}^{\beta}} + \log(t_1), \quad V(\mathbf{t}, \beta) = \sum_{i=1}^{n} \log(t_i^{\beta} - t_{i-1}^{\beta}).$$

The ML estimation of the MPLP parameters was considered in [2], where the authors noted that the numerical evaluation of the ML estimators requires the use of a complex iterative procedure based on a combined use of the Nelder-Mead simplex algorithm for approximating the location of the maximum and the Newton-Raphson method that, from a very accurate starting point, converges to the ML solution. Instead of, we have the following proposition.

Proposition 1 *The ML estimates $\hat{\alpha}_{ML}$, $\hat{\beta}_{ML}$, $\hat{\kappa}_{ML}$ of the MPLP parameters α, β, κ, respectively, are*

$$\hat{\alpha}_{ML} = \frac{n\hat{\kappa}_{ML}}{t_n^{\hat{\beta}_{ML}}},$$

$$\hat{\kappa}_{ML} = \frac{W(\mathbf{t}, \hat{\beta}_{ML}) - S(\mathbf{t}) - n/\hat{\beta}}{W(\mathbf{t}, \hat{\beta}_{ML}) - n\log(t_n)} =: \frac{\kappa_N(\mathbf{t}, \hat{\beta}_{ML})}{\kappa_D(\mathbf{t}, \hat{\beta}_{ML})} =: \kappa(\mathbf{t}, \hat{\beta}_{ML}),$$

and $\hat{\beta}_{ML}$ is a solution to the equation

$$-n\psi(\kappa(\mathbf{t}, \beta)) + n\log\left[\frac{n\kappa(\mathbf{t}, \beta)}{t_n^{\beta}}\right] + V(\mathbf{t}, \beta) = 0. \tag{2}$$

Remark 1 It can be shown that a solution to Eq. (2) for which $\hat{\kappa}_{ML}$ is positive, lies in the interval (β_0, ∞), where β_0 is the unique solution to the equation $\kappa_D(\mathbf{t}, \beta) = 0$ with respect to β. Moreover, $\beta_0 < -n/\sum_{i=1}^{n}\log(t_i/t_n) =: \beta_{start}$, which is a good starting point for searching a solution to Eq. (2).

Although Proposition 1 and Remark 1 give an efficient and reliable algorithm for finding the MLEs, we also propose an alternative method of estimation of the MPLP parameters based on some properties of the MPLP. Namely, it is known from [1] that the random variables $U_i = \Lambda(T_i)/\Lambda(T_n)$, where $\Lambda(t) = \alpha t^{\beta}$, are distributed according to the beta distribution $\mathscr{B}e(\kappa i, \kappa(n-i))$, and are independent of T_n. Thus $E(U_i) = i/n$ and $\mathrm{Var}(U_i) = i(n-i)/n^2(n\kappa+1)$. We then propose estimating (α, β, κ) by $(\hat{\alpha}_{EE}, \hat{\beta}_{EE}, \hat{\kappa}_{EE})$, where $\hat{\beta}_{EE}$ is a solution to the equation

$$\sum_{i=1}^{n-1}\left[\left(\frac{t_i}{t_n}\right)^{\beta} - \frac{i}{n}\right]\frac{1}{i(n-i)} = \sum_{i=1}^{n-1}\left[\left(\frac{t_i}{t_n}\right)^{\beta}\frac{1}{i(n-i)}\right] - \sum_{i=1}^{n-1}\frac{1}{n(n-i)} = 0,$$

$\hat{\kappa}_{EE}$ is a solution to the equation

$$\log(\kappa) - \psi(\kappa) + \log(n) - \log\left(t_n^{\hat{\beta}_{EE}}\right) + \frac{1}{n}\sum_{i=1}^{n}\log\left(t_i^{\hat{\beta}_{EE}} - t_{i-1}^{\hat{\beta}_{EE}}\right) = 0,$$

and $\hat{\alpha}_{EE} = n\hat{\kappa}_{EE}/t_n^{\hat{\beta}_{EE}}$.

In the simulation study conducted we have compared the accuracy of the estimators proposed with the maximum likelihood ones. For each chosen combination of the parameters α, β, κ, the $M = 1000$ samples of the MPLP(α, β, κ) were generated up to a fixed number $n = 30$ of jumps was reached. In most cases considered, the estimated mean squared error (MSE) of the estimator $\hat{\beta}_{EE}$ was smaller than the MSE of $\hat{\beta}_{ML}$. The MSEs of $\hat{\kappa}_{ML}$ and $\hat{\kappa}_{EE}$ were almost the same. On the other hand, the MSEs of $\hat{\alpha}_{EE}$ were always bigger than MSEs of $\hat{\alpha}_{ML}$.

3 Prediction of the Next Failure Time

Let us assume that we observe the MPLP up to nth failure and we are interested in predicting the next event time T_{n+1}.

Lemma 1 *The conditional expectation of T_{n+1} given $T_1 = t_1, \ldots, T_n = t_n$ is the following*

$$E(T_{n+1} \mid T_1 = t_1, \ldots, T_n = t_n) = E(T_{n+1} \mid T_n = t_n) = E(Z + t_n^{\beta})^{1/\beta},$$

where the random variable Z has the gamma $\mathcal{G}(\kappa, 1/\alpha)$ distribution.

On the basis of Lemma 1, we propose predicting the next event time by

$$\hat{T}_{n+1}^{ML} = \frac{1}{M}\sum_{i=1}^{M} Y_i^{ML} \quad \text{or} \quad \hat{T}_{n+1}^{EE} = \frac{1}{M}\sum_{i=1}^{M} Y_i^{EE},$$

where $Y_i^{ML} = (Z_i^{ML} + t_n^{\hat{\beta}_{ML}})^{1/\hat{\beta}_{ML}}$, $Y_i^{EE} = (Z_i^{EE} + t_n^{\hat{\beta}_{EE}})^{1/\hat{\beta}_{EE}}$, Z_i^{ML} and Z_i^{EE}, $i = 1, \ldots, M$, are random variables generated from $\mathcal{G}(\hat{\kappa}_{ML}, 1/\hat{\alpha}_{ML})$ and $\mathcal{G}(\hat{\kappa}_{EE}, 1/\hat{\alpha}_{EE})$ distribution, respectively, and M is the number of Monte Carlo repetitions. Estimative prediction limits for the event time T_{n+1} can be obtained by computing appropriate sample quantiles of $(Y_1^{ML}, \ldots, Y_M^{ML})$ or $(Y_1^{EE}, \ldots, Y_M^{EE})$ for M sufficiently large.

4 Application to Some Real Data Set

In [5] the authors showed that the Weibull renewal process model, often used in analysis of duration dependence of financial cycles, does not fit data from U.S. stock market cycles (1985 through 2000). As a solution, they proposed to fit the MPLP that relies on less restrictive assumptions, and which measures both the long term properties of bull and bear markets, as well as the short term effect, such as duration dependence. In Table 1 we present the results of our analysis of data from [5].

The predicted 95%-intervals for the next failure times based on the ML and EE estimators for the Bull data are (841.86,906.739) and (842.272,915.792), respectively. For the Bear data the corresponding predicted intervals are (476.219,507.079)

Table 1 The ML and EE estimates of α, β, κ and the predicted next failure times $\widehat{T}_{n+1}^{ML}$ and $\widehat{T}_{n+1}^{EE}$ in the MPLP(α, β, κ) for the U.S. stock market data

	$\hat{\alpha}_{ML}$	$\hat{\beta}_{ML}$	$\hat{\kappa}_{ML}$	$\hat{\alpha}_{EE}$	$\hat{\beta}_{EE}$	$\hat{\kappa}_{EE}$	T_n	T_{n+1}	$\hat{T}_{n+1}^{ML}$	$\hat{T}_{n+1}^{EE}$
Bull	0.2044	0.9140	3.2318	0.3649	0.8269	3.1877	835	909	865.789	869.203
Bear	0.1365	1.0654	3.1506	0.0635	1.1817	3.0055	473	478	487.477	486.068

and (475.773,504.237). Using the EE method the next failure time is predicted more accurate than by the ML method.

References

1. Berman, M.: Inhomogeneous and modulated gamma processes. Biometrika **68**, 143–152 (1981)
2. Black, S.E., Rigdon, S.E.: Statistical inference for a modulated power law process. J. Qual. Technol. **28**(1), 81–90 (1996)
3. Lakey, M.J., Rigdon, S.E.: The modulated power law process. In: Proceedings of the 45th Annual Quality Congress, pp. 559–563. American Society of Quality Control, Milwaukee, WI (1992)
4. Zhou, H., Rigdon, S.E.: Duration dependence in US business cycles: An analysis using the modulated power law process. J. Econ. Finan. **32**, 25–34 (2008)
5. Zhou, H., Rigdon, S.E.: Duration dependence in bull and bear stock markets. Mod. Econ. **2**, 279–286 (2011)

The Level of Mortality in Insured Populations

Josep Lledó, Jose M. Pavía, and Francisco G. Morillas

Abstract In the actuarial field, life tables are used in reserving and pricing processes. They are commonly built from aggregate data and incorporate margins as a prudent measure to ensure the insurance company's viability. Solvency II requires insurance companies to calculate technical provisions using *best-estimate* assumptions for future experience (mortality, expenses, lapses, etc) to separate (i) the risk-free component from (ii) adverse deviation of claims. Nowadays, however, the methods used by insurance companies (in most countries, included Spain) do not guarantee that these components can be separated. Many companies build their own tables from general insured population life tables, assuming certain restrictive hypotheses. In this paper, we develop a new cohort-based estimator to build life tables based on individual company experience. We apply it to a real database and find that the proposed methodology improves classical approaches. The described procedure is of application in those countries covered by the Solvency II and IFRS 17 regulatory frameworks.

Keywords Life table · Insured population · Cohort-based estimator · Best-estimate

J. Lledó (✉)
Life Actuary Reale Seguros, Madrid, Spain

Universidad de Alcalá, Madrid, Spain
e-mail: Josep.Lledo@uah.es

J. M. Pavía · F. G. Morillas
Universitat de Valencia, Valencia, Spain

M. Corazza et al. (eds.), *Mathematical and Statistical Methods for Actuarial Sciences and Finance*, https://doi.org/10.1007/978-3-319-89824-7_80

1 Introduction

In the actuarial field, the study of mortality and longevity is always a topic of great interest. Estimates on the incidence of mortality, studied by way of a life table, are used by insurance companies for reserving and pricing processes, which have an impact on the long-term profitability of insurers, conditioning their survival.

In recent years, the insurance sector has suffered significant regulatory changes. *Solvency II* imposes the separation of two basic components in the calculation of technical provisions: the *best-estimate* (BE) [1, pp. 25–50] and the *risk margin* (RM) [1, pp. 54–67]. The new international accounting standards (IFRS 17 [2]), effective as of 2020, will establish that the operating hypotheses (mortality, longevity, expenses, ...) used to calculate the technical provisions should be constructed on the basis of individual experience and will aim to reflect the risks of the insurance portfolios.

At present, however, as far as the authors are aware, more than 50% of the Spanish insurance market uses life tables that do not reflect their own risk structures and do not adapt to their own experience. A large number of insurance companies use as a BE life table a percentage or *factor (fac)*. This implies that the calculations are made jointly, for the entire portfolio, regardless of age.

The aim of this paper is twofold. On the one hand, we detail the current estimators and describe their limitations and risks. On the other hand, we propose to introduce the company's own experience into the process, adapting the cohort estimator proposed in Lledó et al. [3]. The estimator is compared with the *classical* methods through the calculation of the technical provisions for a product of annuities.

2 Mortality: Current and Future. Solvency II and IFRS 17

Under Solvency II, the calculation of technical provisions is based on separating the mortality risk into two parts: BE and RM. [4] postulates that BE can be analysed through four components: (i) the level of mortality, LM; (ii) the trend mortality, TM; (iii) the volatility, VM; and (iv) the calamity, KM. That is, under Solvency II and IFRS 17, the study of mortality requires a static analysis of death and an adequate estimation of the trend. Unlike the RM, the result of the BE technical provision is strongly conditioned by the mortality table and the study group. Therefore, an adequate risk management would entail having an individual mortality table for each company. This would be built on the basis of its own experience given that, although the life insurance products marketed by the different companies are similar, mortality and composition of each insurance portfolio differ between companies.

However, for the calculation of the LM component of a BE life table it is common to use as BE probability of death a percentage (*fac*) of the corresponding value the general life table: $q_{x,t}^{BE} = fac_t * q_x^{\mathrm{T}}$. The value fac_t usually depends on the levels of

information available, using: the ratio between the observed and expected deaths,

$$fac = \frac{\sum_x^{\omega} d_x^r}{\sum_x^{\omega} q_x^T * l_x^r}; \tag{E1}$$

or the capitals at risk,[1]

$$fac = SA_x^{dea} / SA_x^{exp}. \tag{E2}$$

Methods (E1) and (E2) provide estimates that have limitations to be considered as *BE* life tables. Among these: (i) to assume that the mortality rate of the insured population has the same "age-for-age" behaviour as in the base table, (ii) do not reflect the effects of certain commercial policies; or (iii) to be insensitive to increases in deaths or exits in short periods for certain ages. Its use entails a risk. It could imply an underestimation (overestimation) of the probabilities of death, with the corresponding impact on a greater (lower) allowance for, e.g., technical provisions.

3 Methodology and Data

To overcome the above limitations, we follow the path marked by [3, 5] and develop a new cohort estimator that allows us to obtain the raw probabilities of death, q_x, of an insured population. Specifically, according to the Laplace rule, we compare the number of deaths registered in years t and $t + 1$ (for a given cohort) with the (corrected) number of exposed to risk of that cohort that reach the exact age x throughout the year t. In particular, we propose the estimator given by Eq. (E3).

$$q_x = \frac{D_{x:t-x}^t + D_{x:t-x}^{t+1}}{l_x^t - W_{x:t-x}^{t,t+1} + \sum_{j=1}^{W_{x:t-x}^t} w_{x:j}^t + \sum_{j=1}^{W_{x:t-x}^{t+1}} w_{x:j}^{t+1} + \sum_{j=1}^{NB_{x:t-x}^t} nb_{x,j}^t + \sum_{j=1}^{NB_{x:t-x}^{t+1}} nb_{x,j}^{t+1}} \tag{E3}$$

The numerator contains the number of people who die in the years t and $t + 1$ with age x of the cohort born in the year $t - x$ $\left(D_{x:t-x}^t + D_{x:t-x}^{t+1}\right)$. The denominator the insured population belonging to the cohort born in the year $t - x$ exposed to the risk of dying, l_x^t, adjusted for entries and exits (for causes other than death) in the portfolio. To l_x^t, which represents the number of people reaching the exact age x over the year t within the portfolio, must be added (subtracted) the time

[1] Where: d_x^r denotes the number of deceased at age x, of the insurance company r; q_x^T is the probability in the base life table (PERMF 2000P) that a person of age x does not reach the age $x + 1$; l_x^r is the total of persons exposed to risk at the start of the period t for each age x in the insurance company r; and, SA_x^{dea} and SA_x^{exp} are the sums insured at age x of the deceased and exposed to risk, respectively [4].

that the newly insured (the withdrawals) are at risk (no longer at risk) of dying. This can be achieved by: (i) adding the number of years that the new insurance policies contribute with age x, from the generation born in $t-x$, $\sum_{j=1}^{NB_{x:t-x}^{t}} nb_{x:j}^{t}$ and $\sum_{j=1}^{NB_{x:t-x}^{t+1}} nb_{x:j}^{t+1}$; (ii) subtracting the total policies that have been surrendered with age x, in year t of the generation born in $t-x$, $W_{x:t-x}^{t,t+1}$; and (iii) adding the time that these policies contribute to the population at risk $\sum_{j=1}^{W_{x:t-x}^{t}} w_{x:j}^{t}$ and $\sum_{j=1}^{W_{x:t-x}^{t+1}} w_{x:j}^{t+1}$.

To compare the estimator E3 with E1 and E2, we used the microdata belonging to 2015 (n = 329,011) and 2016 (n = 325,831) of an insured population aged 65 or over corresponding to a portfolio of annuities sold through the bancassurance channel of a life-savings insurance company. The inflows and outflows of said insurance portfolio (new business, surrenders and deaths) have also been considered. We have smoothed the empirical estimates using a nonparametric Gaussian kernel with bandwidth 2.

4 Results and Conclusions

The mortality tables have been calculated by gender. For ages from 65 to 100 years, Fig. 1 shows the comparison between the probabilities of death of the base table and the estimators E1, E3 and E3-graduated. As expected, for both genders, the estimate E1 is parallel to the base table. The E3 and E3-graduated estimators show a behaviour not linked to the base table. They follow the pattern of deaths in the portfolio. In general, the probabilities of death of the portfolio are lower than those of the E1 estimator for ages 65–80. A different behaviour is observed after 85 years.

Insurance companies must periodically calculate and set up the mathematical reserve in order to meet future obligations to their insured clients. Depending on the table used, the provisions made may be different. Figure 2 shows the values of the mathematical provisions obtained for an annual annuity of 5000 euros using the E1 and E3-graduated estimators. The provisions tend to diverge 10 years after the sale of the product, the differences being greater for men than for women. The lower mortality visible in Fig. 1 for ages between 65 and 80 years results in a lower level of provision for the E3-graduated estimator. The greatest discrepancies in relative and absolute terms occur between 10 and 25 years from the beginning of the annuity.

The previous results have important implications, since a modification of the life table in the levels obtained could affect a significant proportion of insurance companies, with considerable economic effects. According to ICEA data, in 2015 Spanish insurers had a total of 69,959,911 million euros in book value liabilities.

The presented methodology is characterized by its simplicity and could be used internally in national and international insurance companies. Also, unlike current methodologies, it allows the construction of a mortality table without implicit security surcharges: an indispensable criterion under Solvency II and IFRS17.

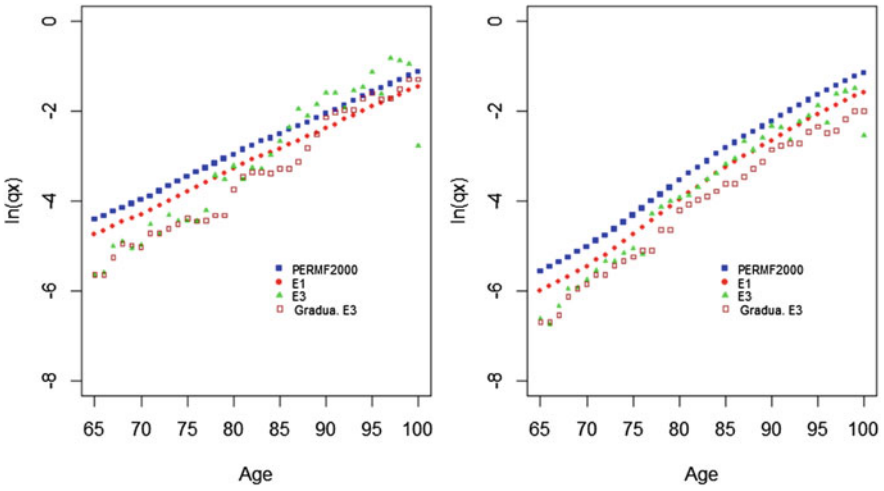


Fig. 1 Estimates of the logarithms of the probabilities of death with estimators E1, E3 and graduation of E3. Left panel: men 2015. Right panel: women 2015

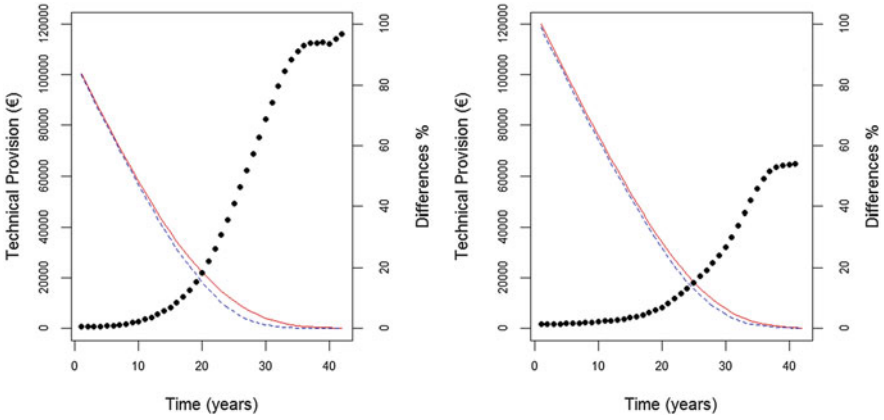


Fig. 2 Best-estimate provision of an annual annuity of 5000 euros using E1 (continuous line) and the graduation of E3 (dashed line). On the right-hand scale, the relative discrepancies between estimates are shown (dotted line): $\left|q_x^{E1} - q_x^{\text{grad E3}}\right| / q_x^{\text{grad E3}}$. Left panel: men. Right panel: women

References

1. CEIOPS: QIS5 technical specifications. Technical report, European Commission—internal market and services DG (2010)
2. IFRS 17: IFRS 17 insurance contracts. International Accounting Standards Board (2017)
3. Lledó, J., Pavía, J.M., Morillas, F.: Assessing implicit hypotheses in life table construction. Scand. Actuar. J. **6**, 495–518 (2017)

4. Van Broekhoven, H.: Market value of liabilities mortality risk. N. Am. Actuarl. J. **6**, 95–106 (2002)
5. Benjamin, B., Pollard, J.: The Analisys of Mortality and Other Actaurial Statistics. Heinemann, London (1986)

Kurtosis Maximization for Outlier Detection in GARCH Models

Nicola Loperfido

Abstract Outlier detection in financial time series is made difficult by serial dependence, volatility clustering and heavy tails. We address these problems by filtering financial data with projections achieving maximal kurtosis. This method, also known as kurtosis-based projection pursuit, proved to be useful for outlier detection but its use has been hampered by computational difficulties. This paper shows that in GARCH models projections maximizing kurtosis admit a simple analytical representation which greatly eases their computation. The method is illustrated with a simple GARCH model.

Keywords GARCH model · Kurtosis · Projection pursuit

1 Introduction

Outliers can be informally defined as observations which appears to be inconsistent with the bulk of the data [1]. They often appear in financial time series [7, 8], hampering both estimation and forecasting [5]. Moreover, outliers are more difficult to detect in time series than in independent data, since a single outlier affects subsequent observations [4].

Univariate kurtosis (i.e. the fourth standardized moment) has many merits in univariate outlier detection [3, 10]. In multivariate analysis, the problem has been satisfactorily addressed by means of projections with maximal kurtosis [6, 15]. Unfortunately, in the general case, they are computationally demanding [6, 9, 16].

Outlier detection has been successfully pursued by means of skewness-based projection pursuit, which looks for maximally skewed data projections [11, 12, 14]. However, it is useless for GARCH models [2], where both returns and innovations are assumed to be symmetric.

N. Loperfido (✉)
Università degli Studi di Urbino "Carlo Bo", Dipartimento di Economia, Società e Politica, Urbino, PU, Italy
e-mail: nicola.loperfido@uniurb.it

M. Corazza et al. (eds.), *Mathematical and Statistical Methods for Actuarial Sciences and Finance*, https://doi.org/10.1007/978-3-319-89824-7_81

In this paper, we propose to detect outliers in a financial time series by filtering it with projections achieving maximal kurtosis, which can be easily computed when the data come from a GARCH process. More precisely, let R_t be the log-return at time t, and let $\mathbf{r}_t = (R_{t-k}, \ldots, R_t, \ldots, R_{t+k})^T$, where $k = (d-1)/2$, be a random vector belonging to the GARCH process $\{R_t \; t \in \mathbb{Z}\}$ satisfying $E\left(R_t^4\right) < +\infty$. Also, let $\boldsymbol{\lambda} \circ \boldsymbol{\lambda} = \left(\lambda_1^2, \ldots, \lambda_d^2\right)^T$ be the dominant eigenvector of $E\left[(\mathbf{r}_t \circ \mathbf{r}_t)(\mathbf{r}_t \circ \mathbf{r}_t)^T\right]$, where "$\circ$" denotes the Hadamard (or elementwise) product. We shall prove that $\boldsymbol{\lambda}^T \mathbf{r}_t$ achieves maximal kurtosis among all linear combinations of $\mathbf{r}_t$. Hence in GARCH models the problem of kurtosis maximization, which in the general case is equivalent to the maximization of a fourth-order polynomial in several variables [9], is reduced to a much simpler eigenvalue problem.

2 Main Result

Generalized AutoRegressive Conditional heteroskedasticity (GARCH) models are often used to describe financial time series. A GARCH process is $\{R_t = Z_t \sigma_t, \; t \in \mathbb{Z}\}$, where $\{Z_t\}$ is a white noise process with independent normal components and $\{\sigma_t\}$ is another process whose squared t-th component is

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^{q} \alpha_i Z_{t-i}^2 + \sum_{j=1}^{p} \beta_j \sigma_{t-j}^2; \tag{1}$$

with $\alpha_0 > 0$, $\alpha_i \geq 0$, $\beta_j \geq 0$ for $i = 1, \ldots, q$ and $j = 1, \ldots, p$. The following theorem highlights a connection between GARCH processes and kurtosis maximization.

Theorem 1 *Let* $\mathbf{r} = \left(R_{t_1}, \ldots, R_{t_d}\right)^T$ *be a realization of the GARCH process* $\{R_t \; t \in \mathbb{Z}\}$*, where* $t_1, \ldots, t_d$ *are integers satisfying* $t_1 \leq \ldots \leq t_d$ *and* $E\left(R_{t_i}^4\right) < +\infty$ *for* $i = 1, \ldots, d$*. Then the kurtosis of* $\mathbf{c}^T \mathbf{r}$ *is greater than the kurtosis of any component of* $\mathbf{r}$*, provided that* $\mathbf{c}^T \mathbf{r}$ *is not proportional to any of them. Also,* $\boldsymbol{\lambda}^T \mathbf{r}$ *achieves maximal kurtosis among all linear projections of* $\mathbf{r}$ *iff* $\boldsymbol{\lambda} \circ \boldsymbol{\lambda}$ *is a dominant eigenvector of* $\mathbf{Q} = E\left[(\mathbf{r} \circ \mathbf{r})(\mathbf{r} \circ \mathbf{r})^T\right]$.

Proof For the sake of simplicity, and without loss of generality, we can assume that $t_i = i$ for $i = 1, \ldots, d$, so that $\mathbf{r} = (R_1, \ldots, R_d)^T$. The linear combination $\mathbf{c}^T \mathbf{r}$ is proportional to a component of $\mathbf{r}$ if and only if all components of $\mathbf{c}$ but one equal zero. In order to rule out this case, and without loss of generality, we can assume that $\mathbf{c}$ differs from any d-dimensional versor. We shall first recall some basic properties of GARCH processes. First, the components of $\mathbf{r}$ are centered and uncorrelated: $E(R_i) = E(R_i R_j) = 0$, for $i \neq j$ and $i, j = 1, \ldots, d$. Also, being a GARCH process stationary, all components of $\mathbf{r}$ have identical second and fourth moments: $E(R_i^2) = \gamma$ and $E(R_i^4) = \kappa$, for $i = 1, \ldots, d$. Without loss of generality we can

assume that $\gamma = 1$, so that $\mathbf{r}$ is a standardized random vector and κ is the kurtosis of R_i (that is its fourth standardized moment). The fourth moment of the projection $\mathbf{c}^T\mathbf{r}$ is

$$E\left[\left(\mathbf{c}^T\mathbf{r}\right)^4\right] = \sum_{i=1}^{d}\sum_{j=1}^{d}\sum_{h=1}^{d}\sum_{k=1}^{d} c_i c_j c_h c_k E\left(R_i R_j R_h R_k\right). \tag{2}$$

Loperfido [13] showed that the following identities hold for a GARCH process: $E\left(R_i R_j R_h R_k\right) = E\left(R_i^2 R_j R_h\right) = E\left(R_i^3 R_j\right) = 0$ whenever i, j, h, k differ from each other, for $i, j, h, k = 1, \ldots, d$. Hence the above identity might be simplified into the following one:

$$E\left[\left(\mathbf{c}^T\mathbf{r}\right)^4\right] = \kappa \sum_{i=1}^{d} c_i^4 + 3\sum_{i\neq j} c_i^2 c_j^2 E\left(R_i^2 R_j^2\right). \tag{3}$$

The kurtosis $E\left[\left(\mathbf{c}^T\mathbf{r}\right)^4\right] \Big/ \left(\mathbf{c}^T\mathbf{c}\right)^2$ of the linear projection $\mathbf{c}^T\mathbf{r}$ is greater than the kurtosis κ of any element of $\mathbf{r}$ if and only if the following inequality holds:

$$\kappa \sum_{i=1}^{d} c_i^4 + 3\sum_{i\neq j} c_i^2 c_j^2 E\left(R_i^2 R_j^2\right) > \kappa\left(\mathbf{c}^T\mathbf{c}\right)^2. \tag{4}$$

A little algebra shows that the above inequality is equivalent to the following one:

$$2\sum_{i\neq j} c_i^2 c_j^2 E\left(R_i^2 R_j^2\right) + \sum_{i\neq j} c_i^2 c_j^2\left[E\left(R_i^2 R_j^2\right) - 1\right] > 0. \tag{5}$$

The first summand in the left-hand side of the above inequality is always positive since $\mathbf{c}$ is not a null vector. The second summand is never negative since $E\left(R_i^2 R_j^2\right) - 1 = cov\left(R_i^2, R_j^2\right)$ and the covariance in a GARCH model is never negative. We have therefore proved that the kurtosis of $\mathbf{c}^T\mathbf{r}$ is greater than the kurtosis of any component of $\mathbf{x}$.

Without loss of generality we can assume that $\mathbf{c}$ is a unit-norm vector, so that the kurtosis of $\mathbf{c}^T\mathbf{r}$ is $E\left[\left(\mathbf{c}^T\mathbf{r}\right)^4\right] = (\mathbf{c}\circ\mathbf{c})^T\mathbf{M}(\mathbf{c}\circ\mathbf{c})$, where $\mathbf{c}\circ\mathbf{c} = \left(c_1^2, \ldots, c_d^2\right)^T$ and the matrix $\mathbf{M} = \left\{m_{ij}\right\}$ is defined as follows: $m_{ii} = E\left(R_i^4\right)$ and $m_{ij} = 3E\left(R_i^2 R_j^2\right)$ for $i \neq j$ and $i, j = 1, \ldots, d$. Then the d-dimensional real vector $\boldsymbol{\lambda}$ which maximizes the kurtosis of a linear projection of $\mathbf{r}$ satisfies $\boldsymbol{\lambda}\circ\boldsymbol{\lambda} = \boldsymbol{\delta}$, where $\boldsymbol{\delta}$ is the dominant eigenvector of $\mathbf{M}$. The matrices $\mathbf{M}$ and $\mathbf{Q}$ are related: $\mathbf{Q} = (1/3)(\mathbf{M} + 2\kappa\mathbf{I}_d)$, where $\mathbf{I}_d$ is the $d-$dimensional identity matrix. Hence they have the same eigenvectors and that the dominant eigenvector of $\mathbf{Q}$ is proportional to $\boldsymbol{\lambda}\circ\boldsymbol{\lambda}$. □

We shall now illustrate the proposed method with the GARCH(0,1) model $R_t = \sigma_t Z_t$, where $\sigma_t^2 = \alpha_0 + \beta_1 \sigma_{t-1}^2$, $\alpha_0 > 0$, $0 < \beta_1 < 1$ and $\{Z_t \, t \in \mathbb{Z}\}$ is a normal white noise process. Stationarity of the process implies that $E\left(R_{t+1}^4\right) = E\left(R_t^4\right) = E\left(R_{t-1}^4\right) = \psi$ and $E\left(R_t^2 R_{t+1}^2\right) = E\left(R_t^2 R_{t-1}^2\right) = \omega$, where the scalars ψ and ω satisfy the constraints $0 < \omega < \psi/\sqrt{2}$. A simple algebra also shows that $E\left(R_{t+1}^2 R_{t-1}^2\right) = 0$. Hence the matrix $\mathbf{Q}$ is

$$\begin{pmatrix} E\left(R_t^4\right) & E\left(R_t^2 a_{t+1}^2\right) & E\left(R_t^2 R_{t+2}^2\right) \\ E\left(R_t^2 R_{t+1}^2\right) & E\left(R_{t+1}^4\right) & E\left(R_{t+1}^2 R_{t+2}^2\right) \\ E\left(R_t^2 R_{t+2}^2\right) & E\left(R_{t+1}^2 R_{t+2}^2\right) & E\left(R_{t+2}^4\right) \end{pmatrix} = \begin{pmatrix} \psi & \omega & 0 \\ \omega & \psi & \omega \\ 0 & \omega & \psi \end{pmatrix}, \tag{6}$$

whose dominant eigenvector is $\left(1, \sqrt{2}, 1\right)^T$.

Hence the linear projections of $(R_{t+1}, R_t, R_{t-1})^T$ with maximal kurtosis are proportional to either $R_{t+1} + \sqrt[4]{2} R_t + R_{t-1}$, $R_{t+1} + \sqrt[4]{2} R_t - R_{t-1}$, $R_{t+1} - \sqrt[4]{2} R_t - R_{t-1}$ or $R_{t+1} - \sqrt[4]{2} R_t + R_{t-1}$. In the third linear combination subsequent filtered data are uncorrelated. More formally, let $X_t = R_{t+1} - \sqrt[4]{2} R_t - R_{t-1}$ and $X_{t-1} = R_t - \sqrt[4]{2} R_{t-1} - R_{t-2}$. The covariance between X_t and X_{t-1} is then zero, as it follows from uncorrelatedness between R_{t+1}, R_t, R_{t-1}, R_{t-2} and the identity $E\left(R_t^2\right) = E\left(R_{t-1}^2\right)$.

References

1. Barnett, V., Lewis, T.: Outliers in Statistical Data. Wiley, New York (1994)
2. Bollerslev, T.: Generalized autoregressive conditional heteroskedasticity. J. Econ. **31**, 307–327 (1986)
3. Ferguson, T.: On the rejections of outliers. In: Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability, pp. 253–287. University of California Press (1961)
4. Galeano, P., Peña D., Tsay, R.S.: Outlier detection in multivariate time series by projection pursuit. J. Am. Stat. Assoc. **101**, 654–669 (2006)
5. Grané, A., Veiga, H.: Wavelet-based detection of outliers in financial time series. Comput. Stat. Data Anal. **54**, 2580–2593 (2010)
6. Hou, S., Wentzell, P.D.: Re-centered kurtosis as a projection pursuit index for multivariate data analysis. J. Chemom. **28**, 370–384 (2014)
7. Johansen, A., Sornette, D.: Stock market crashes are outliers. J. Chemom. **1**, 141–143 (1998)
8. Johansen, A., Sornette, D.: Large stock market price drawdowns are outliers. J. Risk **4**, 69–110 (2001)
9. Kent, J.T.: Discussion on the paper by Tyler, Critchley, Dumbgen & Oja: Invariant co-ordinate selection. J. R. Stat. Soc. B **71**, 575–576 (2009)
10. Livesey, J.H.: Kurtosis provides a good omnibus test for outliers in small samples. Clin. Biochem. **40**, 1032–1036 (2007)
11. Loperfido, N.: Skewness and the linear discriminant function. Stat. Prob. Lett. **83**, 93–99 (2013)
12. Loperfido, N.: Vector-valued skewness for model-based clustering. Stat. Prob. Lett. **99**, 230–237 (2015)

13. Loperfido, N.: A new kurtosis matrix, with statistical applications. Linear Algebra Appl. **512**, 1–17 (2017)
14. Loperfido, N.: Skewness-based projection pursuit: a computational approach. Comput. Stat. Data Anal. **120**, 42–57 (2018)
15. Peña, D., Prieto, F.J.: Multivariate outlier detection and robust covariance estimation (with discussion). Technometrics **43**, 286–310 (2001)
16. Peña, D., Prieto, F.J., Viladomat, J.: Eigenvectors of a kurtosis matrix as interesting directions to reveal cluster structure. J. Multivar. Anal. **101**, 1995–2007 (2010)

Google Searches for Portfolio Management: A Risk and Return Analysis

Mario Maggi and Pierpaolo Uberti

Abstract Google search data has proven to be useful in portfolio management. The basic idea is that high search volumes are related to bad news and risk increase. This paper shows additional evidence about the use of Google search volumes in risk management, for the Standard & Poor Industrial index components, from 2004 to 2017. To overcome the (time-series and cross-section) limitations Google imposes on the data download, a re-normalization procedure is presented, to obtain a multivariate sample of volumes which preserve their relative magnitude. The results indicate that the volumes' normalization and the starting portfolio are decisive for the portfolio performances. Correctly normalized Google search volumes yield poor results. This may lead to revise the interpretation of the search volume: it can be considered a risk indicator, but when used in a *equally risk contribution* portfolio, no evidence of the improvement of the risk-return performances is found.

Keywords Online searches · Google Trends · Portfolio management

1 Introduction

In the recent years the increasing availability of web data fueled a wave of studies which analyzed the relations between web searches and many aspects of the social sciences. Since [2, 8] who first used the Google search volumes to forecast the influenza diffusion, the information contents of the Google queries has been analyzed also for other phenomena. In particular, some studies focused on the relations between web searches and financial data (for example, see [3, 5, 6]). In fact, as [1] showed, there is evidence of the information flow from media to the

M. Maggi (✉)
Department of Economics and Management, University of Pavia, Pavia, Italy
e-mail: mario.maggi@unipv.it; https://sites.google.com/a/unipv.it/mariomaggi/

P. Uberti
Department of Economics (DIEC), University of Genoa, Genoa, Italy
e-mail: uberti@economia.unige.it

M. Corazza et al. (eds.), *Mathematical and Statistical Methods for Actuarial Sciences and Finance*, https://doi.org/10.1007/978-3-319-89824-7_82

financial market. The information content of the Google search volumes has been documented, among others, by [3, 5, 6]. Recently, different works explored the possibility of exploiting the forecasting power of web data to set up asset allocation and trading strategies. For instance, in [4, 9] the Google data are used with the aim of improving the return or the risk-return combination of a financial portfolio.

This paper deepens the analysis by [4] of the contribution of Google data to the asset allocation performances. We focus on the need to obtain a multivariate series of search volumes whose sizes are proportional to the (undisclosed) real volumes. The application of re-normalized series yields different results with respect to [4]. The use of the Google search volumes as risk indicator can reduce the standard deviation of the portfolio return, rather than improve the risk-return performances.

2 Google Trends

Google collects data about every query users type on its web search engine and decided to disclose a part of these data through its service Google Trends.[1] Data about the Google search volumes are available from January 2004. The region and time window may be customized and multiple series can be downloaded in `csv` format. The data Google allows to download are not the raw volumes, but a normalized index (the Google Index, GI in the following) taking integer values between 0 and 100. The maximum volume attained on the selected window is set to 100; all the other data are normalized accordingly and rounded to the nearest integer. This way the dynamic properties are retained, but the absolute size of the volume is lost.

We remark that this rounding may lead to a strong information loss when the series' volumes have a large difference in the overall size: the resolution of the smallest series is reduced. As an example, we downloaded the series of the queries "Italy" only and the couple "Italy" and "United States" (monthly, from 2004 to present). "Italy" has a GI ranging from 32 to 100, instead when downloaded together with "United States", "Italy" ranges from 3 to 9 (only 7 different values), with a clear loss of information about the dynamics of the smallest series.

Google imposes other limitations on the disclosed data, for instance: (1) the possibility to download up to 5 multiple series; (2) the longest the time window, the lowest the frequency of the data (monthly over 5 years, weekly from 3 months to 5 years, daily from 7 to 270 days, and so on).

In order to obtain a multivariate sample of weekly data from 2004 to 2017, we follow a re-normalization procedure. First we downloaded all the series as unique queries for three periods, each period not longer than 5 years (obtaining weekly data). The three series overlap at the extreme dates. Then, we concatenate the series matching the values for the overlapping weeks. We do not round the results. Once

[1] See https://trends.google.com/trends/.

the univariate series have been composed, to obtain a multivariate sample which preserves the relative size of the search volumes, we downloaded multiple series. We include in each search up to 5 series with comparable size, to limit the information loss due to the rounding. Each query must belong to (at least) 2 different multiple series, to allow a cross normalization, similar to the one operated along time. Again, we do not round the results.

3 Asset Allocation Based on Google Search Volumes

Following [4], we use the GIs data to find the weights of a (long only) portfolio composed by the Standard and Poor Industrial index (SPI) stocks. The basic idea is that the search volume is related to bad news, so it is a risk indicator: when the interest on a given stock increases, many people look for information to trade; there is evidence of the relation between web searches and trading volumes [3, 5]; an increase in trading, produces a possible increase in the price volatility.

For this reason, it is possible to extend to search volumes the *equal risk contribution* (ERC) rule, proposed by [7] to manage the risk. Let $V_{i,t}$ be the GI and $w_{i,t}$ be the portfolio weight for stock i, at time t. The ERC rule, yields the portfolio weights

$$w_{i,t} = \frac{V_{i,t}^{-\alpha}}{\sum_{j=1}^{N} V_{j,t}^{-\alpha}}, \tag{1}$$

where α controls for the relevance given to $V_{i,t}$: for $\alpha = 0$, the portfolio is uniform, $w_i = \frac{1}{N}$, $i = 1, \ldots, N$; for $\alpha > 0$, w_i decreases with $V_{i,t}$, underweighting stocks with large GIs; for $\alpha < 0$, w_i increases with $V_{i,t}$, overweighting stocks with large GIs.

Starting from an initial portfolio $w_{i,0}$, a recursive version of (1) can be obtained applying the updating rule

$$w_{i,t} = \frac{w_{i,t-1} e^{-\alpha g_{i,t}}}{\sum_{j=1}^{N} w_{j,t-1} e^{-\alpha g_{j,t}}}, \quad t \geq 1, \tag{2}$$

where $g_{i,t} = \ln V_{i,t} - \ln V_{i,t-1}$ is the log-rate of variation of the Google search volumes (remark that $g_{i,t}$ is scale free). This way we can explicitly control for the starting portfolio: (1) and (2) are equivalent for the same $w_{i,0}$.

We set up portfolios applying (1) and (2) on the stocks listed on the SPI, from July 2004 to July 2017 (GI query = "Company Name"). The value of α ranges from -2 to 2 to vary the strength and the sign of the GIs contribution.

Figure 1 shows the averages, standard deviations and Sharpe ratios for portfolio returns based on the GIs. Remark that without the re-normalization (solid lines) the average return increases, the standard deviation decreases and the Sharpe ratio

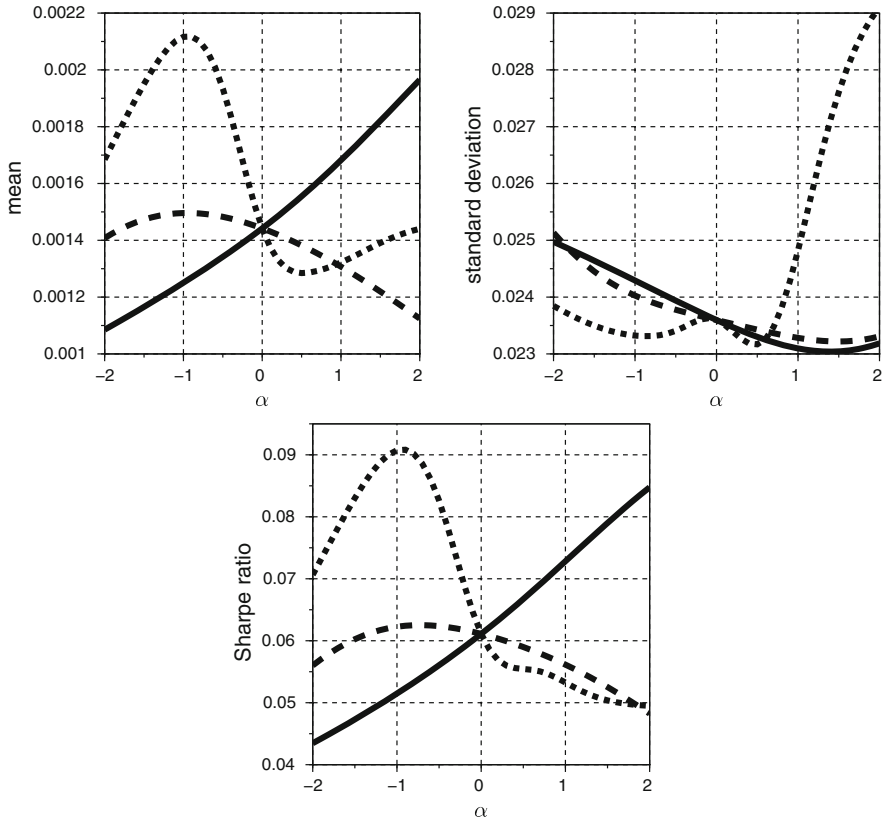


Fig. 1 Returns's average, standard deviation and Sharpe ratio of the portfolios composed through (1) and (2), for $\alpha \in [-2, 2]$: univariate GIs (solid); normalized GIs (short dashed); rule (2) starting from the uniform portfolio (long dashed)

increases with α; that is the stronger GIs are used as risk indicator, the better the portfolio performances. This result is in line with [4], who downloaded the GIs 5 by 5, without (we suppose) an overall cross normalization. Moreover, the normalized GIs and the case of uniform starting portfolio produce opposite results, suggesting that a negative α could work better.

Remark that the normalized GIs produce a very unbalanced portfolio (the weigts' Gini coefficient is 0.98). Moreover, we conjecture that the results obtained in the non normalized case may depend on chance. Therefore, we run a Monte Carlo experiment, using the GIs to update a random initial portfolio with the rule (2). We find that in $\alpha = 0$: the average return is decreasing 80.2% of times, the standard deviation is decreasing 75.4% of times, the Sharpe ratio is decreasing 77.2% of times. We conclude that if the GI is used as a risk measure, there is no evidence supporting the improvement of risk-return performances, unless GIs are used with the opposite meaning, i.e. $\alpha < 0$, so the risk is decreasing with GI.

References

1. Alanyali, M., Moat, H.S., Preis, T.: Quantifying the relationship between financial news and the stock market. Sci. Rep. **3**, 3578 (2013). https://doi.org/10.1038/srep03578
2. Ginsberg, J., Mohebbi, M.H., Patel, R.S., Brammer, L., Smolinski, M.S., Brilliant, L.: Detecting influenza epidemics using search engine query data. Nature **457**, 1012–1014. (2009). https://doi.org/10.1038/nature07634
3. Heiberger, R.H.: Collective attention and stock prices: evidence from Google trends data on standard and poor's 100. PLoS One **10**(8), (2015). https://doi.org/10.1371/journal.pone.0135311
4. Kristoufek, L.: Can Google Trends search queries contribute to risk diversification? Sci. Rep. **3**, 2713 (2013). https://doi.org/10.1038/srep02713
5. Kristoufek, L.: Power-law correlations in finance-related Google searches, and their cross-correlations with volatility and traded volume: evidence from the Dow Jones Industrial components. Phisica A **428**, 194–205 (2015). https://doi.org/10.1016/j.physa.2015.02.057
6. Li, X., Ma, J., Wang, S., Zhang, X.: How does Google search affect trader positions and crude oil prices? Econ. Model. **49**, 162–171 (2015). https://doi.org/10.1016/j.econmod.2015.04.005
7. Maillard, S., Roncalli, T., Teïletche, J.: The properties of equally weighted risk contribution portfolios. J. Port. Manag. **36**(1), 60–70 (2010)
8. Polgreen, P.M., Chen, Y., Pennock, D.M.: Using internet searches for influenza surveillance. Clin. Infect. Dis. **47**, 1443–1448 (2008). https://doi.org/10.1086/593098
9. Preis, T., Moat, H.S., Stanley, E.: Quantifying trading behavior in financial markets using *Google Trends*. Sci. Rep. **3**, 1684 (2014). https://doi.org/10.1038/srep01684

The Challenges of Wealth and Its Intergenerational Transmission in an Aging Society

André Masson

Abstract Since the beginning of the 1980s, our developed societies face an exceptional process of "patrimonialization" which is problematic in terms of unequal wealth concentration and potentially inefficient allocation of capital. This process results notably in an increasing wealth concentration by the elderly. Even if decline in mortality at old age is not the sole reason of this process, it has significantly contributed to worsen its effects. Considering various tax and social remedies, our study leads to sustain the *Taxfinh* (Tax Family Inheritance) program, an inheritance tax reform which aims to favor both gifts (to family members or of professional assets) and charitable bequests, property consumption (real estate), or long-term productive investment of elderly savings.

Keywords Rising life expectancy · Wealth inequality · Wealth transfer taxation

1 Introduction

Since the beginning of the 1980s, our developed societies face a process of 'patrimonialization', that appears both new, massive and particularly harmful for economic growth, equality of opportunities and equity between generations. Due notably but not only to the decline of mortality at old age, this process results in an increasing and unequal weight of wealth relatively to income, an "unproductive over-accumulation" by the elderly, a coming-back of inheritance and bequest, received though later and later, and young households strongly constrained in their property projects.

How to remedy such a damaging wealth situation? A brief survey of possible social and tax reforms will confirm that it is not an easy task. The most promising route would be a selective and progressive rise in inheritance taxation, while

A. Masson (✉)
CNRS-EHESS-PSE, Paris, France
e-mail: amasson@pse.ens.fr

M. Corazza et al. (eds.), *Mathematical and Statistical Methods for Actuarial Sciences and Finance*, https://doi.org/10.1007/978-3-319-89824-7_83

multiplying or facilitating the ways to sidestep the surtax introduced. More than any other reform, this new program *Taxfinh* (Tax Family Inheritance) seems to be the appropriate answer while avoiding the present unpopularity of standard wealth transfer taxation.

2 A Present Wealth Situation Which Is Both Massive and Noxious

The historical process of patrimonialization is a multi-faceted process, which takes place since the late 1970s and concerns more or less all developed societies. It can be divided, for simplicity sake in four components [1].

(1) The first one concerns the growing weight (relatively to income) and inequality of wealth and capital in our societies [2].
(2) The age-distribution of wealth has considerably changed, with a rising concentration of wealth in the hands of the elderly, who seem to "over-save" for their old age, at least in relative terms. Moreover, savings of the elderly mainly represent a potentially unproductive low-risk store of value, invested in cash savings, life insurance and retirement savings, and residential housing [3].
(3) Consequently, the relative weight of inheritance, which had considerably decreased during the 'The Glorious Thirty', has increased sizably in Europe [2]. On the other hand, inheritance is received later and later in full property (usufructs and bare property) —e.g. at an average age of almost 60 in France— due to rising longevity but also higher rights of inheritance given to the surviving spouse.
(4) Finally, young households, facing mass unemployment and precarious jobs in a number of countries and receiving inheritance far too late, are often strongly constrained in their property projects: home purchase, starting up a business, saving for retirement or for bequests. High housing prices and rents in big towns explain that most of them will begin to save for retirement only at mature age.

This massive and general wealth situation appears detrimental to: (i) economic growth, with an increasing mass of inert wealth held by the elderly; (ii) equality of opportunity, with an increasing gap between 'heirs' and 'non heirs', when success in life and well-being depends more and more on (received) wealth; and (iii) intergenerational equity or balance, at the expense of young generations.

3 'Patrimonialization' Due to Rising Longevity and Other Changes

Along with the so-called liberal turn of the 1980s, a series of more or less concomitant changes, with complex interactions, have shaped this process of patrimonialization: population aging in general, and the rise of life expectancy at old age in particular, but also slowdown of economic growth; globalization of more and more financial and mobile capital; new technologies and the emerging knowledge economy, etc.

Component (1) has been summed up rather bluntly by Piketty as "capital is back"—after the two World Wars and the catch-up economy of The Glorious Thirty in Europe. The slowdown of growth, the mobility of globalized capital entailing an international race to the bottom in capital taxation, as well as primitive accumulation in new technologies contribute to explain the re-concentration of wealth. But the high ratios of wealth to GDP are due for an important part to (latent) capital gains on housing.

Wealth is a stock with a long memory: the main driving force of the relative over-accumulation (2) of the elderly in Europe is the succession of high post WW2 past growth during The Glorious Thirty, with often high inflation easing home mortgage reimbursements, followed by low growth of the thirty 'jinxed' years (1978–2007, say) and the Great Recession since 2008, with low inflation but rising housing prices.

These changes are also assumed to be responsible for the emergence of a post-Fordist labor regime and the decline of life employment, making life more difficult for young workers in a number of countries (4), but also for older workers.

In this setting, the decline of mortality at old age is first responsible for the postponement of inheritance. Together with low economic growth penalizing labor income, population aging also explains the increasing weight of social transfers to the elderly (public pension, health and long-term care).

Under low economic growth, the main consequence of the decline of mortality at old age is finally the historical recovery of the risk of longevity understood broadly, *i.e.* the risk of becoming old and destitute with fragile or degraded health. Up to 1980 or so, this risk declined with the rise of the welfare state. It rises again for three main reasons: (i) more old days to secure and the rising risk of long-term care needs with uncertain costs; (ii) lower (in real terms, net of taxes) and more uncertain future pensions (and also health insurance), due to the problematic financial sustainability of the welfare state, and the difficulty to work longer at old age; (iii) more problematic family care owing to the increasing geographical distance from children (and more individualistic behavior of the latter). These new risks of longevity can only raise (rightly or wrongly) the desire of homeownership, especially in younger generations ("at least, we have a roof"), and may contribute to explain the unproductive over-accumulation by the elderly, driven by precautionary motives against old-age unpredictable needs.

4 Possible Tax or Social Remedies: The *Taxfinh* Program

How to remedy a present damaging wealth situation, how to reduce inequality of opportunity, allocate (elderly) households savings to productive investment, and make wealth circulate more rapidly towards needy younger generations?

Possible solutions, driven by different political or ideological motivations, are studied in more details in Masson [1]. One is a neo-liberal revolution, advocating a sizeable reduction of compulsory levies and public spending, achieved notably by lower social transfers to the elderly and incentives to work longer; savings of less-taxed young households will be encouraged. This scenario requires yet a tough and hazardous transition and would increase income and wealth inequalities. On the contrary, to preserve the 'European social model', Piketty [2] advocates a progressive annual tax on (individual, total, net) wealth, bringing tax revenue in the order of 2% of GDP. Main objections against such a wealth tax are capital mobility and tax emigration without international tax cooperation, and capital heterogeneity in rates of return, but also between rent-wealth and productive investment, or between inherited and self-accumulated wealth. Aware of these objections, current economic arguments in the European Commission advocate higher taxation of 'immovable property', especially residential housing, namely a shift from housing transaction tax to recurrent property tax—the latter being assumed to be the most growth-friendly and the least sensitive to capital mobility. This policy raises again many objections: it would be inegalitarian in a number of countries where the rich hold but a limited fraction of their wealth in housing, would require a periodic and difficult reevaluation of land and housing property, and the tax may be confiscatory, etc. [4].

A more appropriate but partial measure would be to make the 'third age' (from the legal retirement age) contribute to the financing of the needs of the 'fourth age' through the introduction of a compulsory (preferably public) and inclusive long-term care insurance. The tax base could then also include the stock of wealth owned.

Complementary to that measure, the most promising route would be a *targeted* increase in wealth transfer taxation [1, 5]. The *Taxfinh* program would hence combine two inseparable components:

- a heavier and more progressive taxation of 'family inheritances' only (for the 10–15% richest families): the relative tax advantage of *inter vivos* gifts or charitable transfers would come from a disincentive to *post-mortem* bequests to one's offspring;
- providing more numerous and easier means to sidestep this 'inheritance surtax' by encouraging parents to give early enough, consume or invest their wealth at old age, *i.e.* favoring reactions that would reduce the noxious effects of patrimonialization.

In France, reasonable tax hikes on *post-mortem* bequests could thus make the effective average tax rate on total wealth transfers roughly double, from 5 to 10%, under the assumption of unchanged bequest behavior; (additional) tax revenue could be earmarked for financing opportunity enhancing programs to the young.

Giving part of one's wealth to children sufficiently early before death (*e.g.* 10 years before, to avoid gifts being reintegrated into the estate) would allow to avoid the surtax. In addition, more freedom to bequeath should be introduced in countries such as France not only for making charitable bequests, but also for the *inter vivos* transmission of family business.

Families should be given new or increased possibilities to run down or consume their (housing) wealth at retirement, especially through new forms of reverse mortgage or 'viager' sales allowing home owners remaining in their home until death.

Investing in long-term and risky assets, conducive to economic growth, should benefit from tax exemptions or reliefs, provided that the asset is kept in the family for a long period (e.g. 25 years) by parents or beneficiary children. Such assets should be invested in dedicated funds (for financing infrastructures and productive capital, new energies and numerical transition, innovations and startups, etc.) and run by long-term (socially) responsible investors.

The implementation of the program should be gradual but *credible*, to have an immediate impact on first Baby-boomers, encouraged to prepare the transfer of their property soon enough: it would be fair as far as the surtax would be only paid by those who 'deserve' it, being too short-sighted or 'selfish' on family or social grounds.

References

1. Masson, A.: Les enjeux du patrimoine et de sa transmission dans nos sociétés vieillissantes. Revue française d'économie, **XXXIII** (2018)
2. Piketty, T.: Capital in the 21st Century. Harvard University Press, Cambridge (2014)
3. Arrondel, L., Masson, A.: Savings and Life Expectancy: Which Products and Taxation? Institut Louis Bachelier – Coll. Opinions & Débats, **14** Paris (2016)
4. Masson, A.: Resistance to reforming property taxes. In: Princen, S. (ed.) Political Economy of Tax Reforms. European Economy Discussion Paper 25, pp. 36–45 (2016)
5. Masson, A.: Taxing more post-mortem family bequests. In: Astarita, C. (ed.) Taxing Wealth: Past, Present, Future. European Economy Discussion Paper 03, pp. 48–58 (2015)

Bivariate Functional Archetypoid Analysis: An Application to Financial Time Series

Jesús Moliner and Irene Epifanio

Abstract Archetype Analysis (AA) is a statistical technique that describes individuals of a sample as a convex combination of certain number of elements called archetypes, which in turn, are convex combinations of the individuals in the sample. For it's part, Archetypoid Analysis (ADA) tries to represent each individual as a convex combination of a certain number of extreme subjects called archetypoids. It is possible to extend these techniques to functional data. This work presents an application of Functional Archetypoids Analysis (FADA) to financial time series. At the best of our knowledge, this is the first time FADA is applied in this field. The starting time series consists of daily equity prices of the S&P 500 stocks. From it, measures of volatility and profitability are generated in order to characterize listed companies. These variables are converted into functional data through a Fourier basis expansion function and bivariate FADA is applied. By representing subjects through extreme cases, this analysis facilitates the understanding of both the composition and the relationships between listed companies. Finally, a cluster methodology based on a similarity parameter is presented. Therefore, the suitability of this technique for this kind of time series is shown, as well as the robustness of the conclusions drawn.

Keywords Functional data analysis · Archetypal analysis · Stock time series

1 Introduction

Given a set of financial time series, the objective of this work is to find a representative set of them that allows to understand and to describe the entire set. Therefore,

J. Moliner
Department of Mathematics, Universitat Jaume I, Castelló de la Plana, Spain
e-mail: jmoliner@uji.es

I. Epifanio (✉)
Department of Mathematics-IMAC, Universitat Jaume I, Castelló de la Plana, Spain
e-mail: epifanio@uji.es

M. Corazza et al. (eds.), *Mathematical and Statistical Methods for Actuarial Sciences and Finance*, https://doi.org/10.1007/978-3-319-89824-7_84

the aim is to find a set of archetypal time series. Archetype Analysis (AA) [1] is an unsupervised data mining technique that allows not only discover patterns but also to reduce the dimensionality of the data set. AA describes individuals of a sample as a convex combination of certain number of elements called archetypes, which in turn, are convex combinations of the individuals in the sample. However, archetypes are not real individuals. Likewise, Archetypoid Analysis (ADA) represents each individual as a convex combination of extreme individuals called archetypoids [6]. These multivariate techniques were extended to functional data [2, 5].

Financial time series can been as functions and functional data techniques [4] are employed. Section 2 describes our data. In Sect. 3 the methodology is introduced, and results are analyzed in Sect. 4. Finally, some conclusions are given in Sect. 5.

2 Data

The starting time series consists of daily equity prices of the S&P 500 stocks. From them, measures of volatility and profitability are generated in order to characterize listed companies. A total of 496 companies are considered, each of them is characterized by two time series, returns and beta coefficients, in a 250 day time window. These series are approximated as functions with 11 Fourier basis. See [3] for details.

3 Methodology

3.1 AA and ADA for (Standard) Multivariate Data

Let $\mathbf{X}$ be an $n \times m$ matrix with n observations and m variables. The objective of AA is to find k archetypes, i.e. a $k \times m$ matrix $\mathbf{Z}$, in such a way that data can be approximated by mixtures of the archetypes. To obtain the archetypes, AA computes two matrices $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ which minimize following the residual sum of squares (RSS):

$$RSS = \sum_{i=1}^{n} \|\mathbf{x}_i - \sum_{j=1}^{k} \alpha_{ij}\mathbf{z}_j\|^2 = \sum_{i=1}^{n} \|\mathbf{x}_i - \sum_{j=1}^{k} \alpha_{ij} \sum_{l=1}^{n} \beta_{jl}\mathbf{x}_l\|^2, \tag{1}$$

under the constraints

1. $\sum_{j=1}^{k} \alpha_{ij} = 1$ with $\alpha_{ij} \geq 0$ for $i = 1, \ldots, n$ and
2. $\sum_{l=1}^{n} \beta_{jl} = 1$ with $\beta_{jl} \geq 0$ for $j = 1, \ldots, k$.

Archetypes are not necessarily actual observations. This would happen if only one β_{jl} is equal to 1 in constraint (2) for each j. In ADA, the previous constraint (2) is changed by the following one, so the continuous optimization problem of AA transforms into a mixed-integer optimization problem:

2. $\sum_{l=1}^{n} \beta_{jl} = 1$ with $\beta_{jl} \in \{0, 1\}$ and $j = 1, \ldots, k$.

3.2 AA and ADA for Functional Data

In the functional context, the values of the m variables in the standard multivariate context are replaced by function values with a continuous index t. Similarly, summations are replaced by integration to define the inner product.

In our problem, two functions characterized each company, so Functional Archetypoids Analysis (FADA) for bivariate functions must be considered. Let $f_i(t) = (x_i(t), y_i(t))$ be a bivariate function. Its squared norm is $\|f_i\|^2 = \int_a^b x_i(t)^2 dt + \int_a^b y_i(t)^2 dt$. Let $\mathbf{b_i^x}$ and $\mathbf{b_i^y}$ be the vectors of length m of the coefficients for x_i and y_i respectively for the basis functions B_h. Therefore, FADA is defined by

$$RSS = \sum_{i=1}^{n} \|f_i - \sum_{j=1}^{k} \alpha_{ij} z_j\|^2 = \sum_{i=1}^{n} \|f_i - \sum_{j=1}^{k} \alpha_{ij} \sum_{l=1}^{n} \beta_{jl} f_l\|^2 =$$

$$\sum_{i=1}^{n} \|x_i - \sum_{j=1}^{k} \alpha_{ij} \sum_{l=1}^{n} \beta_{jl} x_l\|^2 + \sum_{i=1}^{n} \|y_i - \sum_{j=1}^{k} \alpha_{ij} \sum_{l=1}^{n} \beta_{jl} y_l\|^2 = \tag{2}$$

$$\sum_{i=1}^{n} \mathbf{a^x}'_i \mathbf{W} \mathbf{a^x}_i + \sum_{i=1}^{n} \mathbf{a^y}'_i \mathbf{W} \mathbf{a^y}_i,$$

where $\mathbf{a^x}'_i = \mathbf{b^x}'_i - \sum_{j=1}^{k} \alpha_{ij} \sum_{l=1}^{n} \beta_{jl} \mathbf{b^x}'_l$ and $\mathbf{a^y}'_i = \mathbf{b^y}'_i - \sum_{j=1}^{k} \alpha_{ij} \sum_{l=1}^{n} \beta_{jl} \mathbf{b^y}'_l$, with the corresponding constraints for $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$. See [2] for computational details.

4 Results

FADA has been applied to our data set with $k = 5$. Range Resources Corp. (RRC) (Energy), Lincoln National Corp.(LNC) (Financial), eBay (EBAY) (Technology), NetApp (NTAP) (Technology), Consolidated Edison, Inc (ED) (Utilities) are the archetypoids obtained. Their economic sectors appear in brackets. ED is a company that, in comparison with the rest of the archetypoids, presents low and constant values for both returns and beta values. On the other hand, NTAP presents high returns at the beginning and at the end of the time series, while its volatility

decreases over time. LNC presents a typical financial company profile, with moderate profitability and volatility during the first three quarters of the time series. Once the crisis broke out in 2007, volatility shot up to unprecedented levels while profitability plummeted. Regarding volatilities, EBAY presents just the opposite profile, with great beta values in the first years that stabilize over time. As regards the returns, this company has a moderate profitability level compared to the other archetypoids, but with slightly higher oscillations at the first half of the time series. Finally, RRC is characterized by having bell-shaped functions, that is, with relatively low values at the extremes of the temporal domain and higher values at the centre.

It is also possible to compare the taxonomy of companies obtained by this method with the structure by sectors managed by market analysts through the matrix $\boldsymbol{\alpha}$. See [3] for detailed explanations about these results.

5 Conclusions

FADA has been applied to the time series of stock quotes in the S&P500 from 2000 to 2013. The objective was to show an unsupervised method (AA and ADA for functions) that lies somewhere between Principal Component Analysis (PCA) and clustering methods, in the sense that its modeling flexibility is higher than clustering methods and its factors are very easy to interpret, unlike PCA. At the best of our knowledge, this is the first time FADA is applied to financial time series. Furthermore, we have applied FADA for two time series for company simultaneously.

Acknowledgements This work is supported by DPI2013-47279-C2-1-R, DPI2017-87333-R from the Spanish Ministry of Economy and Competitiveness (AEI/FEDER, UE) and UJI-B2017-13 from Universitat Jaume I.

References

1. Cutler, A., Breiman, L.: Archetypal analysis. Technometrics **36**(4), 338–347 (1994)
2. Epifanio, I.: Functional archetype and archetypoid analysis. Comput. Stat. Data Anal. **104**, 24–34 (2016)
3. Moliner, J.: Bivariate functional archetypoid analysis: an application to financial time series, Master's thesis. Universitat Jaume I (2017)
4. Ramsay, J.O., Silverman, B.W.: Functional Data Analysis, 2nd edn. Springer, New York (2005)
5. Vinué, G., Epifanio, I.: Archetypoid analysis for sports analytics. Data Min. Knowl. Disc. **31**(6), 1643–1677 (2017)
6. Vinué, G., Epifanio, I., Alemany, S.: Archetypoids: a new approach to define representative archetypal data. Comput. Stat. Data Anal. **87**, 102–115 (2015)

A Note on the Shape of the Probability Weighting Function

Martina Nardon and Paolo Pianca

Abstract The focus of this contribution is on the transformation of objective probability, which in Prospect Theory is commonly referred as *probability weighting*. Empirical evidence suggests a typical inverse-S shaped function: decision makers tend to overweight small probabilities, and underweight medium and high probabilities; moreover, the probability weighting function is initially concave and then convex. We apply different parametric weighting functions proposed in the literature to the evaluation of derivative contracts and to insurance premium principles.

Keywords Cumulative prospect theory · Probability weighting function · Premium principles

1 Introduction

Cumulative Prospect Theory (CPT) has been proposed in [7] as an alternative to Expected Utility to explain actual behaviors. Formally, CPT relies on two key transformations: the value function v, which replaces the utility function for the evaluation of outcomes, and a distortion function for objective probabilities w, which models probabilistic risk behavior. Risk attitudes are derived from the shapes of these functions as well as their interaction. The focus of this contribution is on the transformation of objective probability, which is commonly referred as *probability weighting* or *probability distortion*.

A weighting function w is a strictly increasing function which maps the probability interval [0, 1] into [0, 1], with $w(0) = 0$ and $w(1) = 1$. Evidence suggests a typical *inverse-S shape*: small probabilities are overweighted, $w(p) > p$, whereas medium and high probabilities are underweighted, $w(p) < p$. The *curvature* of the

M. Nardon (✉) · P. Pianca
Ca' Foscari University of Venice, Department of Economics, Venezia, Italy
e-mail: mnardon@unive.it; pianca@unive.it

M. Corazza et al. (eds.), *Mathematical and Statistical Methods for Actuarial Sciences and Finance*, https://doi.org/10.1007/978-3-319-89824-7_85

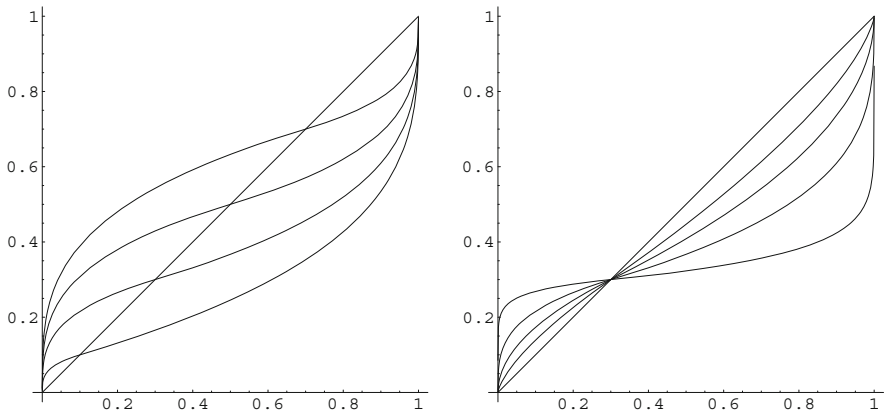


Fig. 1 Instances of the two parameter probability weighting function proposed by [1], with different elevation (left) and curvature (right)

weighting function is related to the risk attitude towards probabilities; the function is initially concave (probabilistic risk seeking or optimism) for probabilities in the interval $(0, p^*)$, and then convex (probabilistic risk aversion or pessimism) in the interval $(p^*, 1)$, for a certain value of p^*. A linear weighting function describes probabilistic risk neutrality or objective sensitivity towards probabilities, which characterizes Expected Utility. Empirical findings indicate that the intersection (*elevation*) between the weighting function and the 45 degrees line, $w(p) = p$, is for p in the interval $(0.3, 0.4)$. Figure 1 shows instances with different elevation and curvature of the two parameter *constant relative sensitivity* probability weighting function proposed by [1].

Different parametric forms for the weighting function with the above mentioned features have been proposed in the literature, and their parameters have been estimated in many empirical studies. Some forms are derived axiomatically or are based on psychological factors. Single parameter and two (or more) parameter weighting functions have been suggested; some functions have linear, polynomial or other forms, and there is also some interest for discontinuous weighting functions. Two commonly applied weighting functions are those proposed by Tversky and Kahneman [7] $w(p) = \frac{p^\gamma}{(p^\gamma+(1-p)^\gamma)^{1/\gamma}}$, with $w(0) = 0$ and $w(1) = 1$, and $\gamma > 0$ (with some constraint in order to have an increasing function); and Prelec [5] $w(p) = e^{-\delta(-\ln p)^\gamma}$, with $w(0) = 0$ and $w(1) = 1$, $0 < \delta < 1$, $\gamma > 0$. When $\gamma < 1$, one obtains the inverse-S shape.

The choice of the probability weighting function should be driven by the following motivations: its empirical properties, intuitive and empirically testable preference conditions, nonlinear behavior of the probability weighting function. Moreover, a parametric probability weighting function should be parsimonious (remaining consistent with the properties suggested by empirical evidence), in particular when we consider different parameters for the weighting of probability of gains and losses.

We analyze some applications in finance to the evaluation of derivative contracts (see [4]) and in insurance to premium principles, briefly discussed in the next section.

2 An Application to Premium Calculation

Let u denote the utility function, and W be the initial wealth; the *utility indifference price* P is the premium from the insurer's viewpoint which satisfies (if it exists) the condition:

$$u(W) = \mathbb{E}[u(W + P - X)], \tag{1}$$

where X is the possible loss, modeled with a non-negative continuous random variable. The premium P makes indifferent the insurance company about accepting the risky position and not selling the insurance policy.

Differently from Expected Utility, in Prospect Theory individuals are risk averse when considering gains and risk-seeking with respect to losses; moreover, they are more sensitive to losses than to gains of comparable magnitude (loss aversion). The final result $W + P - X$ in (1) could be positive or negative. Results are evaluated considering potential gains and losses relative to a *reference point*, rather than in terms of final wealth, hence assuming zero as reference point (the *status quo*), the relative result $P - X$ will be considered. Decision weights are differences in transformed cumulative probabilities of gains or losses.

We consider the cumulative prospect value for a continuous random variable [2]:

$$V = \int_{-\infty}^{0} \Psi^{-}[F(x)]\, f(x)\, v^{-}(x)\, dx + \int_{0}^{+\infty} \Psi^{+}[1 - F(x)]\, f(x)\, v^{+}(x)\, dx, \tag{2}$$

where $\Psi = \frac{dw(p)}{dp}$ is the derivative of the weighting function w, F is the cumulative distribution function and f is the probability density function of the outcomes.

Condition (1) under continuous CPT becomes (see also [3]):

$$0 = \int_{-\infty}^{0} \Psi^{-}[F(x)]\, f(x)\, v^{-}(P - x)\, dx + \int_{0}^{+\infty} \Psi^{+}[1 - F(x)]\, f(x)\, v^{+}(P - x)\, dx. \tag{3}$$

We also assume that decision makers are not indifferent among frames of cash flows: the framing of alternatives exerts a crucial effect on actual choices. People may keep different *mental accounts* for different types of outcomes, and when combining these accounts to obtain overall result, typically they do not simply sum up all monetary amounts, but intentionally use *hedonic framing* [6] such that the combination of the outcomes appears more favorable and increases their utility. Outcomes are aggregated or segregated depending on what leads to the highest

possible prospect value: multiple gains are preferred to be segregated (*narrow framing*), losses are preferred to be integrated with other losses (or large gains) in order to ease the pain of the loss. Mixed outcomes would be integrated in order to cancel out losses when there is a net gain or a small loss; for large losses and a small gain, they usually are segregated in order to preserve the *silver lining*. This is due to the shape of the value function in Prospect Theory, characterized by risk-seeking or risk aversion, diminishing sensitivity and loss aversion.

If we segregate the cashed premium from the possible loss, condition (3) becomes

$$0 = v^+(P) + \int_{-\infty}^{0} \Psi^-[F(x)]\, f(x)\, v^-(-x)\, dx. \tag{4}$$

A usual choice for the value function is

$$\begin{cases} v^- = -\lambda(-x)^b & x < 0 \\ v^+ = x^a & x \geq 0; \end{cases} \tag{5}$$

which leads to the following result:

$$P = \left(\lambda \int_0^{+\infty} \Psi^-[F(x)] f(x)\, x^b\, dx \right)^{1/a}. \tag{6}$$

Alternative functional forms both for the value function and the probability weighting function, embedded in CPT framework (3), yield a different model with potentially different implications for choice behavior. In particular, when the weighting function has an inverse-S shape, very low probability of extreme events are overweighted. We apply different probability weighting functions and study the effects on the premium calculation. In particular, two parameters allow for separate control of curvature and elevation, and the constant relative sensitivity probability weighting function proposed by [1], which models distinctly these two features, is of particular interest.

References

1. Abdellaoui, M., L'Haridon, O., Zank, H.: Separating curvature and elevation: a parametric probability weighting function. J. Risk Uncertain. **41**, 39–65 (2010)
2. Davies, G.B., Satchell, S.E.: The behavioural components of risk aversion. J. Math. Psychol. **51**, 1–13 (2007)
3. Kaluszka, M., Krzeszowiec, M.: Pricing insurance contracts under cumulative prospect theory. Insur. Math. Econ. **50**, 159–166 (2012)
4. Nardon, M., Pianca, P.: A behavioural approach to the pricing of European options. In: Corazza M., Pizzi C. (eds.) Mathematical and Statistical Methods for Actuarial Sciences and Finance, pp. 217–228. Springer, Milano (2014)

5. Prelec, D.: The probability weighting function. Econometrica **66**, 497–527 (1998)
6. Thaler, R.H.: Mental accounting and consumer choice. Mark. Sci. **4**, 199–214 (1985)
7. Tversky, A., Kahneman, D.: Advances in prospect theory: cumulative representation of the uncertainty. J. Risk. Uncertain. **5**, 297–323 (1992)

Disability Pensions in Spain: A Factor to Compensate Lifetime Losses

Patricia Peinado

Abstract Among the different instruments used by the welfare state to protect vulnerable population, there are the disability pensions. These pensions appear as key elements to protect disable people in the absence or breaks in their labour careers. The Spanish pay-as-you-go social security system is a good example of disability pension provision. In Spain, a safety-net combines means-tested and non-means-tested elements to guarantee a certain level of income to individuals with a given degree of disability. In this chapter attention is focused on the second type of component; that is to say, the pensions entitled to the people who having contributing for a certain period of time, have later in life, caused a disability pension. The pension entitled in this later case is computed according to the labour profile of the individual and, consequently, linked to her or his past contributions to the social security system. However, the method used to compute the main component of the pension leaves the beneficiaries of a disability pension in a disadvantageous situation compared to the beneficiaries of a regular pension. This paper discusses this loss and defines a factor to compensate that loss.

Keywords Disability pensions · Spain · Actualisation factor

1 Pension Systems: The Case of Spain

A pension system is expected to (i) provide retirees with a minimum secure income; (ii) provide a reasonable replacement rate during retirement; (iii) give the possibility to complement retirement income with private savings [1, 2]. These three objectives are concreted in the well known three "tiers" or "pillars".

P. Peinado (✉)
Department of Applied Economics V, University of the Basque Country UPV/EHU, Bilbao, Bizkaia, Spain
e-mail: patricia.peinado@ehu.eus

M. Corazza et al. (eds.), *Mathematical and Statistical Methods for Actuarial Sciences and Finance*, https://doi.org/10.1007/978-3-319-89824-7_86

In the case of Spain, the pension system may be defined as an unfunded pay-as-you-go social security system in which the final income drawn by retirees is the result of adding a means-tested minimum-income component (first pillar), a non-means-tested defined benefit component (second pillar) and private savings (third pillar). In the Spanish system, the second component represents 72% of the income drawn during retirement[1] while the private component amounts 5% of that income.[2] As a consequence, studying this second component is paramount to understand the income level enjoyed by retired population. This chapter focus attention on this non-means-tested major component of Spanish pension beneficiaries to analyse the income differences existing between standard pensioners and the individuals who have been entitled a pension for disability reasons.

2 The Main Component of Spanish Pensions: Public Non-means-tested Benefit

In order to compute the public non-means-tested component of the pension, the Spanish pension system foresees different situations. On the one hand, there are the rules for the people who had the possibility to develop a complete working career, which we could address as *beneficiaries of a standard pension* or *standard pensioners*. Standard pensioners make monthly contributions to the system during their working lives. The contributions, which are linked to their wages, are recorded by the social security administration together with the number of years contributed. By the age of retirement, work-profiles are used to compute the "effective pension", being this equal to the product of the "regulatory base" times the "coefficient" applied. The regulatory base is a weighted average of the wages or contributions made during the last 15–25 years prior to retirement while the coefficient assigns a percentage of the regulatory base according to the number of years contributed. Once the effective pension is entitled, the benefit is yearly actualised.[3]

On the other hand, there are the rules for the people who, being disable, did not have the chance to develop a regular working life. These are the *beneficiaries of a disability pension* for a permanent disability reason.[4] Four main types of permanent disabilities may be distinguished: (i) partial disability; (ii) total disability; (iii) absolute disability; and (iv) severe disability.[5]

[1] Share of gross household income for households of 65 of older people [2].

[2] Share of gross household income for households of 65 of older people [2].

[3] For further information on the rules governing the Spanish pension system see [3].

[4] There is also the partial disability. Whether the breaks in labour career have an effect on labour income and pension is an interesting discussion that relies beyond the scope of this chapter.

[5] Based on [3].

In all these cases, the Spanish social security administration defines an individual compensation as a function of the regulatory base of the individual. In this case only the last 10 years of contributions or a shorter period of the labour career is used. Under partial disability, the individual receives a fixed economic compensation equal to a 24-month payment of the regulatory base. In the cases of total, absolute or severe disability, a pension benefit is assigned to the individual. These last three cases constitute the group of *beneficiaries of a disability pension* in the present chapter. In order to compute this pension benefit the regulatory base is multiplied by a coefficient defined according to the type of disability and other personal circumstances. In the case of total disability, the coefficient is equal to 55%.[6] For absolute disability, the coefficient is equal to 100%. Finally, individuals with severe disability are entitled the corresponding total or absolute disability pension plus a complement that varies according to the specificities of the individual. After being entitled, as in the case of standard pensions, the benefits are actualised according to the factor yearly defined by the government.

3 Underlying Differences Between the Beneficiaries of a Regular Pension and the Beneficiaries of a Disability Pension

From the explanation above, it may be drawn that the Spanish pension system is design to address the specificities of the different types of individuals. However, these rules do not completely address the income differences between regular and disable pensioners.

As explained along the chapter, the regulatory base of each pensioner is computed according to the contributions made; that is to say, reflects the wages earned by individuals during their labour careers. Moreover, under the hypothesis that wages are a function of the productivity of workers, the computed regulatory base would provide information on the average value of worker's productivity during the last years of labour career before retirement. Additionally, from a life cycle perspective, as shown by empirical evidence on labour economics, productivity is expected to increase along the working life of individuals. That is to say, wages are lower at the beginning of the labour career and higher as the worker gets older and closer to the legal retirement age. As a consequence, on the light of the method used to compute the pension, it is expected for the regulatory base to be higher the closer it is computed to the legal retirement age of the individual and, accordingly, we expect the value of the regulatory base of a regular pensioner to be higher than that of a disable pensioner due to the fact that for the standard pensioner, it is computed at a more productive stage of the working career (later in life).

[6]This coefficient may be increased to 75% for those workers older than 55 for whom it is considered difficult to have access to another new job.

The question that remains is whether empirical evidence supports this hypothesis and, if this was the case, whether the difference between the regulatory bases of regular pensioners and disable pensioners could be estimated. On this direction, Peinado [4, 5] gives an estimate of the ***disability elasticity of real pension***; that is to say, the elasticity of the real regulatory base with respect to the number of years of disability using the Spanish Continued Survey of Working Lives. According to the results an additional year of disability is expected to decrease the regulatory base of pensioners a 1.5%. This implies that, as an example, a disable pensioner that caused the disability pension by the age of 55 has associated a regulatory base 15–18 percentage points lower than the one she or he would have shown, would not she or he been disable.[7] In other words, and following with the hypothesis at the beginning of the section, each year of disability is expected to generate an average productivity loss of 1.5 percentage points, which is not compensated by the system, leaving the beneficiary of a disability pension in a disadvantageous position compared to the beneficiary of a standard pension. Moreover, the sooner the disable individual starts to perceive a disability pension, the higher it is the income loss.

4 A Simple Actualisation Factor to Compensate Differences Between the Beneficiaries of a Regular Pension and the Beneficiaries of a Disability Pension

Results shed light on the existence of a situation that leaves in a disadvantageous position to the beneficiaries of a disability pension. As explained, the Spanish Social security system addresses the existence of different situations among different pensioners; however, there is no mechanism to compensate the losses in productivity that disable people have to accept after being declared disable and being entitled a disability pension.

Including a specific actualisation factor for the beneficiaries of a disability pension would address this problem. The factor could consist of increasing gradually the regulatory base of the pension of the disable individual according to productivity changes (by applying an average 1.5% increase per year of disability according to the results) each year prior to the legal age of retirement. Ideally, the policy that could be implemented may be summarised as follows: actualise all the regulatory bases of current beneficiaries of a disability pension by applying the yearly actualisation factor times the number of years of disability prior to retirement age and, from that moment on, every year actualise all the pensions drawn by disable

[7]For those retiring at the age of 65 it would be 15 percentage points lower while under the new legal retirement age of 67 the value of the loss would be equal to 18 percentage points.

people who have not yet reached the legal retirement age according to the changes in productivity registered in the economy or the corresponding estimated factor. This measure, would not only compensate the losses for disable people, but would also imply a clear step ahead to achieve non-discrimination for disability reasons.

References

1. Lanoo, K., Barslund, M., Chmelar, A., von Werder, M.: Pension schemes. Directorate general for internal policies. Policy Department A: Economic and Scientific Policy. European Parliament (2014)
2. OECD: Pensions at a Glance 2013. OECD, Paris (2013)
3. Social-Security: Home page. http://www.seg-social.es/Internet_1/index.htm. Accessed Oct 2017
4. Peinado, P.: Disability Pensions in Spain: A Factor to Compensate Life-Time Losses. Mathematical and Statistical Methods for Actuarial Sciences and Finance (MAF 2018). Carlos III University, Madrid, Spain (2018)
5. Peinado, P.: The Spanish Pension System, Disability Pensions and Vulnerability. Health Economics Seminar Series 2017/2018. Institute of Public Health, University of Cambridge, Cambridge (2017)

A Minimum Pension for Older People via Expenses Rate

Noemí Peña-Miguel, María Cristina Fernández-Ramos, and Joseba Iñaki De La Peña

Abstract For 2050, 21.8% of the world population will overcome 60 years, 16% the 65 years old and 4.4% 80 years, due partly, to the reduction of the rates of fecundity and mortality. This fact will provoke problems in the costs of public and private systems of long term care coverage.

As the elderly population increases, the expenses increase as well in health and geriatrics services. These ones imply costly technologies and sectorial inflation and is another reason to provoke problems in the cost of public and private systems of long term care coverage. To face them, the pensioners have principally their own pensions and savings. This paper proposes to develop the methodology to determine minimal pensions under the presumption of supporting the expense in vital consumption and to adapt it to the situation of severe or great dependent. A review of the literature is carried out to present the methodology used as well as the results of the application for Spain.

Keywords Sustainable benefits · Minimum pension benefit · Social security

1 Introduction

Ageing causes problems into social security [1, 2] not only in health assistance [3–6] as well in older care [7, 8]. The aging of the population produces technologies more expensive and the increase of the prices of this sector is higher than over other

N. Peña-Miguel (✉) · J. I. De La Peña
University of the Basque Country (UPV/EHU), Vizcaya, Spain

Polibienestar Research Institute, Valencia, Spain

Consolidated Research Group EJ/GV: IT 897-16, Bilbao, Spain

Faculty of Economics and Business, Bilbao, Spain
e-mail: noemi.pena@ehu.eus

M. C. Fernández-Ramos
School of Education, Junta de Castilla y León, Valladolid, Spain

M. Corazza et al. (eds.), *Mathematical and Statistical Methods for Actuarial Sciences and Finance*, https://doi.org/10.1007/978-3-319-89824-7_87

sectors [9]. Thus, the real problem to solve is how to finance care and healthcare systems. For that aim, this paper proposes a measure of the level of minimum public pension for older people via expenses rate. It develops the methodology to measure it under the presumption of supporting the expense in vital consumption.

2 Long Term Care Problem: Consume Paths

Pensions will face the change in the consumption patterns (Fig. 1) and the consumption patterns must be [10, 11] the focus to inform about the pension coverage [12]. If so, they would be a good tool for defining public policies that eliminate the risk of poverty as World Bank proposes [13], European Commission [14] or ILO [15].

Following the necessity-based and family approaches, [16] defines a minimum pension—*MP*—as a single pension based on minimum level of expenses in goods necessary for living: food; clothing and shoes; housing and public transport. Those are difficult for people under poverty line. So, this *MP* for a household "*h*" should be,

$$MP_{h;t} = MP_{n;t} \cdot n_{h;t} + \sum_{j=1}^{s} DMP_{j;t} \cdot n_{j;t}$$

where

$MP_{h;t}$: Minimum pension of the household "*h*" at year "*t*".
$MP_{n;t}$: Constant minimum pension for each individual at year "*t*".
$n_{h;t}$: Number of individuals in household "*h*" in year "*t*".
$DMP_{j;t}$: Differential of minimum pension for factor "*j*", at year "t".
$n_{j;t}$: Number of individuals in the household with factor "*j*" for year "t".

$n_{h;t}$ and $n_{j;t}$ could be equal or different depending on the number of individuals affected by the factor.

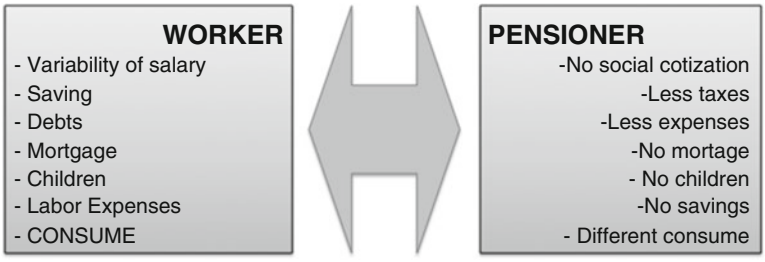


Fig. 1 Different circumstances before and after retiring. Source: Own work

This pension represents a substitution rate TS_x at age x, over the real pension RP_x ,

$$TS_x = \frac{MP_x}{RP_x}$$

Nevertheless, the expense is higher when the pensioner is older due to the sanitary and health requirements. So, the ratio must contemplate the different evolution of these expenses. Let's,

ISS_x : Sustainable replacement rate at age x.
$VAMP_x$: Actuarial value of the minimum pension at age x.
$VARP_x$: Actuarial value of the true pension at age x.
β: Long term care's expected increase.
α: Public pensions 'expected increase.
p_x^m : Probability that a person of age x is alive one year more.
v: Discount factor.

$$ISS_x = \frac{VAMP_x}{VARP_x} = \frac{\sum_{h=x}^{w} MP_x \cdot \beta^{h-x} \cdot {}_{h-x}p_x^m \cdot v^{h-x}}{\sum_{h=x}^{w} RP_x \cdot \alpha^{h-x} \cdot {}_{h-x}p_x^m \cdot v^{h-x}}$$

This measure reports how much the real cost of life is: if it is lower than the unit, the pensioner will have enough incomes from public resources.

3 An Application to Spain

Following the necessity-based and family approaches, [16] calculates the *MP* on the basis of an expense in vital goods for the pensioners in the year 2010. As average is bigger, so TS_x is less than one (Fig. 2).

The paper takes the Spanish population mortality table (PE2000NP), and positive differential in expenses $\beta > \alpha$, interest rate at long term of 2%, and changing inflation from 0% to 2% and health increasing from 0.5% to 4%. With this technical

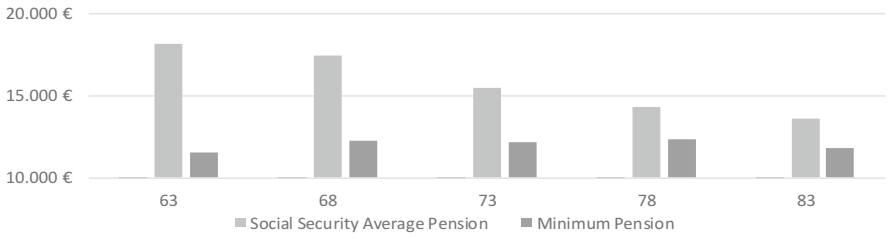


Fig. 2 Social Security average pension and minimum pension in 2010, at different ages

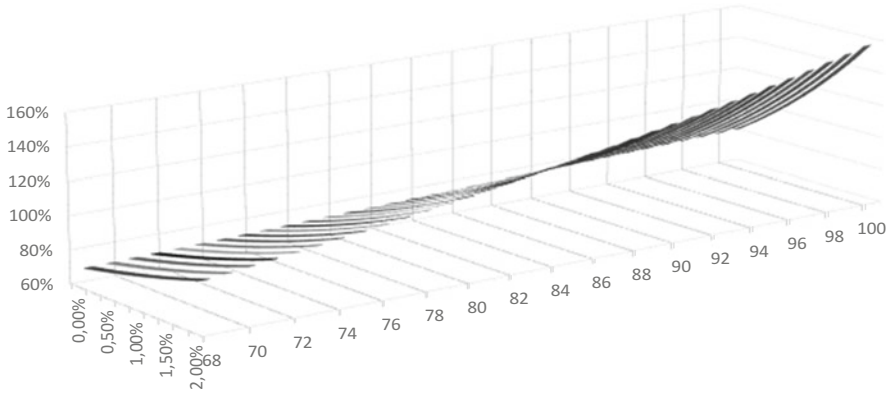


Fig. 3 *ISS* according to the age of the beneficiary and the lever of β

basis, Fig. 3 illustrates the evolution of *ISS* ratio: as the pensioner lives more, the value of the pension is not enough to face to Long Term's expenses Care.

4 Conclusion

One of the real challenges posed by the silver economy is to maintain and improve the quality of life of the elderly. In the public sector, this challenge would be possible only if public expenditure on active and healthy ageing is considered an investment and not only a cost. The Sustainable replacement rate is the measure, which takes into account the real public funds as income of the beneficiary of the pension and the cost of the ageing. This indicator permits to detect the needs of the older people taking into account their life expectancy. Achieving these objectives would require pro-active public policies designed to enable strategic investments and spending designed to foster active ageing, good health, social inclusion and independence [16].

Acknowledgements This study was supported by Polibienestar Research Institute and Consolidated Research Group EJ/GV: IT 897-16.

References

1. Muñoz, G.A., Santos, L.M.: The buildings of the elderly as supplemental retirement pensions. Rev Int Adm Finanzas. **5**(1), 17–32 (2012)
2. Kotlikoff, L., Hagist, C.: Who's going broke? Comparing growth in healthcare costs in ten OECD countries. NBER Working Paper No. w11833 (2005)

3. Abio, G.: Population ageing and sustainability of the Spanish national health system: some financial policy alternatives. Geneva Risk and Insur Issues Pract. **31**, 557–580 (2006)
4. Puga, M.D., Sancho, M., Tortosa, M.A., Malmberg, B., Sundström, G.: La diversificación y consolidación de los servicios sociales para las personas mayores en España y Suecia. Revista Española de Salud Pública. **85**(6), 525–539 (2011)
5. Blanco, A., Urbanos, R., Thuissard, I.J.: Per capita health spending by age and sex in Spain (1998–2008): changes and effects on public healthcare expenditure projections. Gaceta Sanitaria. **27**(3), 220–225 (2013)
6. Casado, D., López, G.: Vejez, dependencia y cuidados de larga duración en España. La Caixa, Barcelona (2001)
7. Ahn, N., Alonso, J., Herce, J.A.: Gasto Sanitario y envejecimiento de la población en España. Documentos de trabajo, 7. Fundación BBVA (2003)
8. Hidalgo-Vega, A., Pérez-Camarero, S., Llano, J., Pi-Corrales, G., Fernández, A., López-Díaz, J., Pérez -Veiga, J.P.: Estudio de las pautas previsibles de utilización futura de servicios sanitarios por mayores, frente a la viabilidad financiera del sistema de salud. Max Weber Institute (2011)
9. Frank, R.H.: Does growing inequality harm the middle class? East. Econ. J. **26**(3), 253–264 (2000)
10. Frank, R.H.: The mysterious disappearance of James Duesenberry. New York Times, June 9 (2005)
11. Hurd, M.D., Rohwedder, S.: The retirement-consumption puzzle: anticipated and actual declines in spending and retirement. NBER Working Paper No. 9586 (2003)
12. Holzman, R., Hinz, R.: Old-Age Income Support in the 21st Century. An International Perspective on Pension Systems and Reform. World Bank, Washington, DC (2005)
13. European Commission: Adequate and Sustainable Pensions. Publication Office, Luxembourg (2003)
14. Grech, A.: How Best to Measure Pension Adequacy. Centre for Analysis of Social Exclusion. London School of Economics, London (2013)
15. Peña-Miguel, N., De La Peña, J.I., Fernández-Sainz, A.: Main factors for a proposal for a social protection floor. Soc. Indic. Res. **123**(1), 59–76 (2015)
16. European Commission: Growing the European silver economy. Background Paper 23 (2015)

A Comparative Analysis of Neuro Fuzzy Inference Systems for Mortality Prediction

Gabriella Piscopo

Abstract Recently, Neural network (NN) and fuzzy inference system (FIS) have been introduced in the context of mortality data. In this paper we implement an Integrated Dynamic Evolving Neuro-Fuzzy Inference System (DENFIS) for longevity predictions. It is an adaptive intelligent system where the learning process is updated thanks to a preliminary clusterization of the training data. We compare the results with other neuro fuzzy inference systems, like the Adaptive Neuro Fuzzy System (ANFIS) and with the classical approaches used in the mortality context. An application to the Italian population is presented.

Keywords AR · ANFIS · DENFIS · ECM · Lee Carter model · Mortality projections

1 Introduction

In the context of mortality projection, stochastic models are very popular. The most used is the Lee-Carter (LC) model [1], whose main statistical tools are the least square estimation through the Singular Value Decomposition of the matrix of the log age specific mortality rate and the Box and Jenkins modelling and forecasting for time series. In order to capture structural changes in mortality patterns, many extensions have been proposed. Recently, Neural Network (NN) and Fuzzy Inference System (FIS) have been introduced for mortality prediction by Atsalaki et al. [2, 8]. They implement an Adaptive Neuro-Fuzzy Inference System (ANFIS) model based on a first order Takagi Sugeno (TS) type FIS [9]. They show that the ANFIS produces better results than the AR and ARIMA models for mortality projections. Piscopo [3, 4] proposes an Integrated Dynamic Evolving Neuro-Fuzzy Inference System (DENFIS) for longevity predictions. DENFIS is

G. Piscopo (✉)
Department of Economics and Statistics, University of Napoli Federico II, Napoli, Italy
e-mail: gabriella.piscopo@unina.it

M. Corazza et al. (eds.), *Mathematical and Statistical Methods for Actuarial Sciences and Finance*, https://doi.org/10.1007/978-3-319-89824-7_88

introduced by Kasabov et al. [5] for adaptive learning of dynamic time series predictions. It is an adaptive intelligent system where the learning process is updated thanks to a preliminary clusterization of the training data. The Evolving Clustering Method (ECM) is used to subdivide the input set and determine the position of each data in the input set. [5] show that DENFIS effectively describes complex data and outperforms some existing methods.

In this paper we show a comparative analysis of the performance of different FIS for mortality projections. In particular, we compare the results obtained applying the AR-DENF to the Italian mortality data with the classical LC and with other two FIS, the ANFIS and the Hybrid fuzzy system (HYFIS) [6].

The paper is organized as follow: in Sect. 2 we present the dynamic evolving neuro fuzzy procedure; in Sect. 3 we briefly describe the LC; in Sect. 4 we show a comparative application to Italian mortality dataset; final remarks are offered in Sect. 5.

2 Adaptive Neuro Fuzzy Systems

The Dynamic Evolving Neuro Fuzzy System is an adaptive learning fuzzy system for dynamic time series prediction. It differs from the ANFIS [7] because the fuzzy rules and parameters are dynamically updated as new informations come in the system; both use a TS architecture to implement learning and adaptation. DENFIS is introduced by Kasabov et al. [5] for adaptive learning of dynamic time series predictions. It is an adaptive intelligent system where the learning process is updated thanks to a preliminary clusterization of the training data. The Evolving Clustering Method (ECM) is used to subdivide the input set and determine the position of each data in the input set.

We refer to Piscopo [4] for a description of the Algorithm.

3 The Lee Carter Model

In order to model the mortality separately for each i population without considering dependence between groups, the widely used Lee Carter Model (LC) [1] describes the mortality rates at age x and time t as follows:

$$m_{xt,i} = \exp\left(\alpha_{x,i} + \beta_{x,i} k_{t,i} + u_{xt,i}\right) \tag{1}$$

The model is fitted to historical data through the Singular Value Decomposition of the matrix of the observed mortality rates. The estimated time varying parameter is modelled as a stochastic process; standard Box and Jenkins methodology are used to identify an appropriate ARIMA model according which $k_{t,i}$ are projected.

4 Application to Mortality Dataset

In this work we apply AR-DENFIS to mortality forecasts and compare the results with the classical approach of LC and with ANFIS and HYFIS.

In order to define the number of inputs of the DENFIS in the mortality dataset, we firstly apply an AR scheme. The data used are taken from the Human Mortality Database. We work on the mortality rates m_t for the Italian males aged 50, collected from $t = 1940$ up to $t = 2012$. The data, considered by single calendar year, are split into training dataset from 1940 up to 1993 and test dataset from 1994 up to 2012. The AR is fitted to the whole time series and the order equal to 2 is chosen minimizing the Akaike Information Criterion; consequently, in our DENFIS we introduce two input variable x_1 and x_2 (mortality 1 and 2 years before) and one output y (mortality one step ahead).

Firstly, we implement the DENFIS on the training dataset, setting the value of Dthr equal to 0.1, the maximum number of iteration equal to 10, the parameter d equal to 2, the step size of the gradient descent equal to 0.01. Once the DENFIS is created, the mortality rate is projected on the testing period and the results are compared with the realized mortality.

In the second step of our application, we implement the ANFIS and the HYFIS on the same training and testing dataset.

Finally, we implement the LC. We fit the model on the male population aged between 0 and 100, considering years between 1940 and 1993; the parameter k_t is derived; a random walk model is fitted on the serie of k_t and is projected from 1994 up 2012 through a Monte Carlo simulation with $n = 1000$ paths. Finally the value of projected mortality rates for male aged 50 are derived using Eq. (1).

The MSE of the LC, DENFIS, ANFIS and HYFIS are compared. The results are shown in Tables 1 and 2.

Table 1 The mortality rates realized vs projected through LC and FIS

T	Realized	DENFIS	ANFIS	HYFIS	LC
1994	0.00461	0.004458675	0.006748497	0.004606545	0.005083814
1995	0.00413	0.004593811	0.006911526	0.004609010	0.005004209
1996	0.00409	0.004354364	0.007093047	0.004600910	0.004934666
1997	0.00388	0.004143083	0.006623053	0.00460027	0.004868312
1998	0.00390	0.004003856	0.006627108	0.004598407	0.004802357
1999	0.00371	0.003933489	0.006408018	0.004598407	0.004728343
2000	0.00359	0.003829500	0.006473568	0.004597862	0.004659374
2001	0.00359	0.003684567	0.006307787	0.004597807	0.004591524
2002	0.00316	0.003637632	0.006182709	0.004597716	0.004525431
2003	0.00334	0.003384584	0.006315005	0.004597448	0.004462271
2004	0.00312	0.002780000	0.005812602	0.004597438	0.004403274
2005	0.00305	0.002780000	0.006064491	0.004597418	0.004347287
2006	0.00297	0.002780000	0.005859077	0.004597408	0.004286747
2007	0.00304	0.002780000	0.005814189	0.004597408	0.004222623
2008	0.00294	0.002780000	0.005709545	0.004597397	0.004164231
2009	0.00292	0.002780000	0.005814263	0.004597397	0.004100000
2010	0.00278	0.002780000	0.005719193	0.004597391	0.004049956
2011	0.00288	0.002780000	0.005748325	0.004597391	0.003990748
2012	0.00286	0.002780000	0.005572785	0.004597390	0.003937622

Table 2 RMSE in the DENFIS, ANFIS, HYFIS, LC

MSE	DENFIS	ANFIS	HYFIS	LC
	5.726724e-08	7.890026e-06	1.7606e-06	1.219046e-06

5 Conclusions

In this paper we have applied an integrated AR-DENFIS procedure to forecasts mortality and have compared the results with the standard LC and with other two fuzzy system, the ANFIS and the HYFIS. The results show that the DENFIS produces the lower mean square error and projected mortality rates more similar to the realized trend. We attribute this improvements to the structural features of the ANFIS: the dynamic clusterization and learning algorithm at the basis of the DENFIS allow to capture better the complexity of the mortality trend.

References

1. Lee, R.D., Carter, L.R.: Modelling and forecasting U.S. mortality. J. Am. Stat. Assoc. **87**, 659–671 (1992)
2. Atsalakis, G., Nezis, D., Matalliotakis, G., Ucenic, C.I., Skiadas, C.: Forecasting mortality rate using a neural network with fuzzy inference system. No 0806, Working Papers, Department of Economics, University of Crete (2008) http://EconPapers.repec.org/RePEc:crt:wpaper:080

3. Piscopo, G.: AR dynamic evolving neuro fuzzy inference system for mortality data. In: Skiadas, C.H., Skiadas, C. (eds.) Demography and Health Issues-Population Aging, Mortality and Data Analysis Demographic Methods and Population Analysis. Springer, Cham (2018)
4. Piscopo, G.: Dynamic evolving neuro fuzzy inference system for mortality prediction. Int. J. Eng. Res. Appl. **7**(3), 26–29 (2017)
5. Kasabov, N.K., Song, Q.: DENFIS: dynamic evolving neuro-fuzzy inference system and its application for time series-prediction. IEEE Trans. Fuzzy Syst. **10**(2), 144–154 (2002)
6. Kim, J., Kasabov, N.: HyFIS: adaptive neuro fuzzy inference systems and their application to nonlinear dynamical systems. Neural Netw. **12**, 1301–1319 (1999)
7. Jang, J.S.R.: ANFIS: adaptive-network-based fuzzy inference systems. IEEE Trans. Syst. Man Cybern. **23**, 665–685 (1993)
8. D'Amato, V., Piscopo, G., Russolillo, M.: Adaptive neuro-fuzzy inference system vs stochastic models for mortality data, in smart innovation. In: Systems and Technologies, vol. 26, pp. 251–258. Springer, Cham (2014)
9. Takagi, T., Sugeno, M.: Fuzzy identification of systems and its application to modelling and control. IEEE Trans. Syst. Man Cybern. **15**(1), 116–132 (1985)

Financial Networks and Mechanisms of Business Capture in Southern Italy over the First Global Wave (1812–1913): A Network Analysis Approach

Maria Carmela Schisani, Maria Prosperina Vitale, and Giancarlo Ragozini

Abstract The present contribution aims at analyzing the economic relationships that international actors set up around the business opportunities progressively offered by the metropolitan area of Naples over Italian political Unification in order to uncover how and in what hands power was vested. From a unique and original database, data are gathered on the whole amount of enterprises and companies operating in the city. Time varying two-mode networks are defined through the relations among economic agents and institutions and then analyzed by means of exploratory network analysis tools.

Keywords International financial relations · Temporal two-mode networks · Southern Italy

1 Introduction

Networks—at different levels (transport, communication, finance, etc.)—have been recognized as a key driving force in the rise and expansion of global waves over the time [1]. During the nineteenth century the rise of a cohesive and politically organized global financial elite and its ability to build, and capitalize on, relationships were crucial in allowing capital to move from core financial centers towards peripheral countries, this way exploiting investment opportunities

M. C. Schisani (✉)
Department of Economics and Statistics, University of Naples Federico II, Naples, Italy
e-mail: schisani@unina.it

M. P. Vitale
Department of Economics and Statistics, University of Salerno, Fisciano (SA), Italy
e-mail: mvitale@unisa.it

G. Ragozini
Department of Political Science, University of Naples Federico II, Naples, Italy
e-mail: giragoz@unina.it

M. Corazza et al. (eds.), *Mathematical and Statistical Methods for Actuarial Sciences and Finance*, https://doi.org/10.1007/978-3-319-89824-7_89

in trade and sovereign bonds as well as in industrialization, technology transfers and infrastructure development (FDI, transport networks, networked cities, etc.).

Before and after Italian political unification (1861), the peripheral area of Southern Italy was a crossroad of international capital flows. It became the hub of a complex network of relations within a space beyond its geographical and economic borders, being embedded in a wider process of integration into the “space of flows" of the developing global capitalism. In this context local actors played a role in fostering these dynamics of integration, as vehicles through which international capital flows found their way to and rooted in Southern Italy.

The present paper aims at analyzing the economic relationships that international actors set up around the business opportunities progressively offered by this peripheral area. It wants to explore how the networks of actors and of business and financial entities reorganized over Italian political unification. Network analysis tools are used in order to analyze the collected data describing temporal two-mode networks in which the actors are the subjects and the events are the business and the financial units. After basic notions about the theoretical and the methodological framework and the details about the data collection process, some first preliminary network results are here reported.

2 The South of Italy in the 1st Global Wave: Context and Data

The present work focuses on the metropolitan area of Naples in the 60-year period between 1830 and 1890. Naples, up until 1861 the capital city of the Bourbon kingdom of the Two Sicilies, was the biggest Italian city in terms of inhabitants (484,026) and, even after Unification, it remained the most important political, economic and financial centre of the entire South of Italy [2]. Indeed, the thirty-year period before 1861 almost entirely corresponds to the long reign of Ferdinand II Bourbon (1830–1859), during which the beginning of railway construction (i.e. 1839 the first Italian railway, Napoli-Portici) along with other infrastructure works (gas lighting) attracted foreign capitalists, in addition to the long-lasting presence of the Rothschilds in Naples. On the other hand, the thirty-year period after Italy’s Unification was marked by massive foreign investment, thanks to the profitable opportunities opened up by broad-ranging urban planning projects (i.e., railways, harbor, public utilities, etc.) [4, 7], on the wave of the modernizing economic policies embraced by the Italian liberal governments, firstly aimed at fostering national market integration through infrastructure investments.

Data about the economic actors and firms, and their relationships, have been directly collected from original archival sources [7]. The collected data are stored in a unique and original database containing the whole amount of enterprises and companies operating in Naples between 1812 and 1913 (source: IFESMez www.ifesmez@unina.it). The present paper is based on data coming from this database.

A large amount of information and different kind of networks can be derived from this database [5].

3 The Evolution of the Role of the Economic Actors in Naples: A Network Approach

Given the database, time-varying two-mode networks representing the relationships among economic agents and institutions over the decades have been derived. Formally, a time-varying two-mode network can be represented by a set of K two-mode networks $\{\mathscr{G}_k\}_{k=1,\ldots,K}$. The k index refers to different time points or occasions [6]. Each network $\mathscr{G}_k$ consists of two sets of relationally connected units, and can be represented by a triple $\mathscr{G}_k\,(V_{1k}, V_{2k}, \mathscr{R}_k)$ composed of two disjoint sets of nodes—V_{1k} and V_{2k} of cardinality N_k and J_k, respectively—and one set of edges or arcs, $\mathscr{R}_k \subseteq V_{1k} \times V_{2k}$. By definition, $V_{1k} \cap V_{2l} = \emptyset, \forall k$. In our case, the sets $V_{1k} = \{a_{1k}, a_{2k}, \ldots, a_{N_k}\}$ are the sets of N_k economic actors in the time occasions, whereas $V_{2k} = \{e_{1k}, e_{2k,} \ldots, e_{J_k}\}$ represents the set of J_k economic institutions in which the actors are involved in each time occasion. The edge $r_{ijk} = (a_i, e_{jk})$, $r_{ijk} \in \mathscr{R}_k$ is an ordered couple and indicates whether or not an actor a_{ik} plays one, or more then one, role in the economic institution e_{jk}. Actors' and institutions' attributes that can help in explaining the relational pattern are present in the database.

In this first phase of the analysis, two valued two-mode networks are obtained by collapsing the first three and the second three decades in two time occasions: the value associated to each tie represents how many times and how many different roles an economic actors played in a given firm. In the network 1830–1860, there are 2714 economic actors belonging to 612 institutions for 5091 total links. The 61% of the actors presents degree equal to 1, and the 90% a degree less or equal to 3. This implies that the 90% of actors are occasionally partners or shareholders of some firms during these 30 years; only 31 actors have a degree greater or equal than 10 (the maximun degree is 22). The degree distribution suggests the presence of strong core-periphery structure. In Fig. 1a the two-mode network of these 31 actors with the corresponding firms shows the presence of two communities connected by one firm. As for the second 30 years (1860–1890), the network's size increases with more dense connections. 19070 actors and 1600 firms are present in the network connected by 28969 total links. Also in this period there is an high percentage of actors (95%) with a degree less or equal than 3 links in the 30 years. In this period, 121 actors have a degree greater or equal than 10, and the maximum degree rises to 45. Thus, also in this case there is a strong core-periphery structure. In the two-mode network of these 121 actors (Fig. 1b) only one large component, divided in almost three communities, is present.

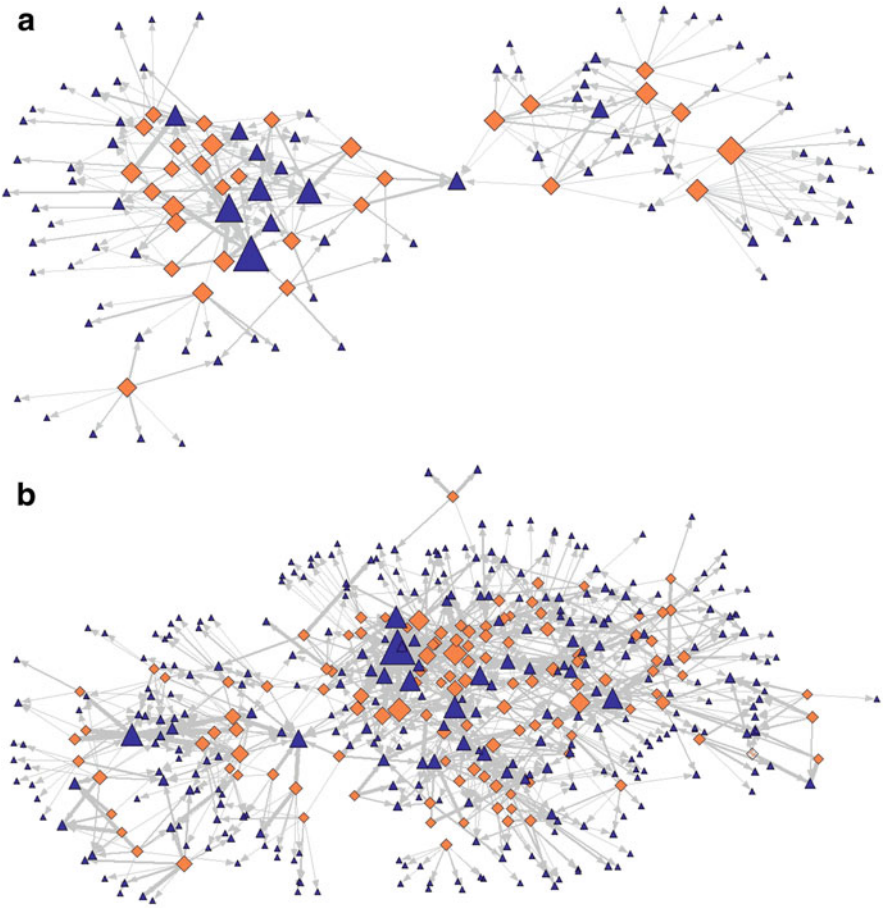


Fig. 1 Bipartite graphs for the two-mode networks: years 1830–1860 (**a**), years 1860–1890 (**b**). Node's size: weighted actor degree and event size; Node's color: orange= actors, grey= consulate, light blue= parliament, blue= society, white= state, yellow= government, black= others. Tie's size: number of actors affiliated with the same entity

4 A First Discussion

During the nineteenth century, European (and global) economy was fast transforming so did the underlying structural interdependencies of the countries across the cores and peripheries and a marked reconfiguration of power served to reshape main features of business international order. The case study of Naples well exemplifies these changing dynamics, from the point of view of a peripheral area. The two thirty-year business networks here analyzed, capture the shifting interdependencies and power relations between the local business context and foreign actors (agents and firms) featuring some of the effects of the progressive changes

in both international banking technology (i.e. since the mid 1850s, Rothschilds vs Pereires/Crédit Mobilier [3]) and Italian political balances (1861). The two analyzed networks account for increasingly dense sets of interconnections between local business environment and foreign agents when the fading-out of Rothschilds financial control (1863) brought other international Pereire-orbiting businessmen and firms to occupy a key position in the local corporate network). The pre-Unitarian network structure (the peripheral position of the Crédit Mobilier) predicts the post-Unitarian one, signaling in advance the early international interest in Southern Italy as a strategic area for realizing the ambitious Saint-Simonian French infrastructure project (railway, ports, etc.) of the Systéme de la Mediterranée (Chevalier 1832).

References

1. Casson, M., Hashimzade, N.: Large Databases in Economic History: Research Methods and Case Studies. Routledge, London (2013)
2. Davis, J.A.: Società e imprenditori nel regno borbonico, 1815–1860, vol. 816. Laterza, Bari (1979)
3. Landes, D.S.: Vieille Banque et Banque Nouvelle: la rèvolution financière du dix-neuvièeme siècle. Revue d'Histoire Moderne & Contemporaine **3**, 204–222 (1956). https://doi.org/10.2307/20527130
4. Nitti, F.S.: Il capitale straniero in Italia. Laterza, Bari (1915)
5. Primerano, I., Caiazzo, F., Giordano, G., Schisani, M. C., Vitale, M. P.: Inter-firm networks in Naples before and after Italian political unification. A network analysis perspective. Ital. J. Appl. Stat. **30**(1), 143–165 (2018)
6. Ragozini, G., De Stefano, D., D'Esposito, M. R.: Multiple factor analysis for time-varying two-mode networks. Net. Sci. **3**(1), 18–36 (2015). https://doi.org/10.1017/nws.2015.5
7. Schisani, M. C., Caiazzo, F.: Networks of power and networks of capital: evidence from a peripheral area of the first globalisation. The energy sector in Naples: from gas to electricity (1862–1919). Bus. Hist. **58**, 207–243 (2016). https://doi.org/10.1080/00076791.2015.1071796

Modeling High-Frequency Price Data with Bounded-Delay Hawkes Processes

Ali Caner Türkmen and Ali Taylan Cemgil

Abstract Hawkes processes are a recent theme in the modeling of discrete financial events such as price jumps, trades and limit orders, basing the analysis on a continuous time formalism. We propose to simplify computation in Hawkes processes via a bounded delay density. We derive an Expectation-Maximization algorithm for maximum likelihood estimation, and perform experiments on high-frequency interbank currency exchange data. We find that while simplifying computation, the proposed model results in better generalization.

Keywords Hawkes processes · Self-exciting process · High-frequency trading

1 Introduction

Detecting and predicting excitatory dynamics, how events occur stochastically while triggering each other, in (marked) point processes is a well-studied problem that recently features prominently in quantitative finance. Especially with data availability at the level of the order book and the surge in *high-frequency trading*, such analysis is now a more familiar element in modeling financial market phenomena, and is applied to price jumps, trades and limit order submissions among other discrete events [1].

Hawkes process [5] (HP) is a general form for multivariate mutually exciting point processes. Originating in seismic event modeling, it has found a range of applications including those in financial markets.

A. C. Türkmen (✉) · A. T. Cemgil
Department of Computer Engineering, Boğaziçi University, Bebek, Istanbul, Turkey
e-mail: caner.turkmen@boun.edu.tr; taylan.cemgil@boun.edu.tr

M. Corazza et al. (eds.), *Mathematical and Statistical Methods for Actuarial Sciences and Finance*, https://doi.org/10.1007/978-3-319-89824-7_90

In this work, we propose a simplification for parameter estimation by invoking a specific instantiation of HP. Namely, we place an upper bound, assumed to be known a priori, on the time where events are allowed to trigger others—reminiscent of other *fixed-lag* techniques in finance. We demonstrate that this form both simplifies computation and is better suited to modeling currency market price jumps.

2 Beta Delay Hawkes Model

A multivariate HP is a set of counting processes $\{N_k(t)\}_{k=1}^K$ with intensity functions

$$\lambda_k(t) = \lambda_k^{(0)}(t) + \sum_{k'} \int dN_{k'}(t')\phi_{(k,k')}(t-t') \tag{1}$$

That is, the intensity is equal to a base intensity, plus an "excitation" from previous discrete events in the process. We further constrain that $\phi_{(k,k')}(\tau) = 0$ for $\tau < 0$, $\phi_{(k,k')}(\tau) \geq 0$ for all k, k'. That is, the effects of prior events are causal, and additive (i.e. excitatory).

Concretely, assume we model sequences of financial events, e.g. limit orders, for different stocks. Here, k, k' index different stocks. $\lambda_k^{(0)}(t)$ is interpreted as the *background* arrival rate of limit orders for stock k, and $\phi_{(k',k)}(\tau)$ as the added intensity caused by the order flow in stock k' to stock k. This excitatory effect is assumed to decay, as time delay τ, increases.

Often a factorized kernel is assumed, with $\phi_{(k,k')}(x) = \varphi_{k,k'} g_k(x)$ where we require the delay *density* to have $\int_0^\infty g_k(x)dx = 1$ for all k. Elements $\varphi_{k,k'}$ are interpretable as the magnitude of causality from sequence k to sequence k'. $g_k(x)$ models the probability density of *any* event causing an event of type k after x time units. Common choices for $g_k(x)$ include both exponential and power-law tail densities, e.g. $g_k(x) = \alpha_k \exp(-\beta_k x)$ or $g_k(x) = \alpha_k t^{\beta_k}$.

The key problem in delay densities with semi-infinite support is that, in the general case, exact computation of the likelihood has to account for $O(N^2)$ *causality* relationships between events. Computation is simpler for exponential delay, however this form is unable to capture a realistic representation of how interactions occur.

In financial applications, one can choose to bound the maximum delay in which an event is assumed to cause further events. Imposing such a bound has several benefits. First, lower delay bounds can lead to significant improvements in computation time - crucial for high-frequency applications. Second, by means of more flexible densities in a bounded interval, it can focus inference on the shape of the delay density improving explanatory power. Finally, it may be natural to impose bounds in the application domain, as is common with fixed-lag *indicators* in quantitative finance.

Table 1 Wall-clock times to parameter estimation, number of EM iterations until convergence, and perplexity values for both training and test data

t_{max} (ms)	Time (s)	# Iter.	Perplexity	
			Train	Test
100	135	32	43.87	37.20
1000	1075	69	**38.85**	**32.87**
10000	17296	104	39.56	33.46

Better models lead to lower perplexity
Bold figures indicate best performance

In this work, we change the delay density with a non-standard Beta density and assuming that the upper bound of the delay distribution, t_{max}, is known and the first shape parameter is fixed to 1. We take $g_k(x) = \mathscr{B}_{[0,t_{max}]}(1, \beta_k) = \frac{\beta_k (t_{max}-x)^{\beta_k - 1}}{t_{max}^{\beta_k}}$, where $0 \leq x \leq t_{max}$. Hence we assume that if a previous event is to cause future events, they will occur in the next t_{max} units of time. The parameters β_k determine the shape of the tail of the bounded density.

We perform maximum likelihood parameter estimation via a Expectation Maximization (EM) algorithm [9], similar to [11]. Expectation-maximization is a well-known alternative to gradient descent methods for maximum-likelihood estimation, often offering enhanced stability and speed of convergence.

The algorithm is computationally favorable since a time delay bound is imposed, and only a limited number of causality relationships have to be accounted for in the expectation (E) step. Computation time scales linearly with respect to the total time frame of the data, T.

HPs have been applied towards modeling price jumps [4, 7], trades [3], order flows [2, 6, 8, 12] and intensity bursts related to exogenous discrete events [10] in the backdrop of equity, currency and futures exchanges of varying scale. See Bacry et al. [1] for an extensive review of HPs and applications in finance. EM algorithms similar to ours were developed in [11, 13]. A bounded logistic-normal delay density appears in [7], however the focus is on Bayesian inference with graph priors and no treatment of the effects of bounded delays is offered.

3 Experiments

We use tick-level price data of 11 currency pairs in the TrueFX[1] database. The database aggregates bid/ask quotes for pairs quoted in interbank exchanges. We model midprice (simple average of best bid and ask quotes) jumps, taking price increases and decreases as separate event sequences, for each currency pair. The data set contains a total of over 36 million price events in millisecond resolution.

[1] www.truefx.com.

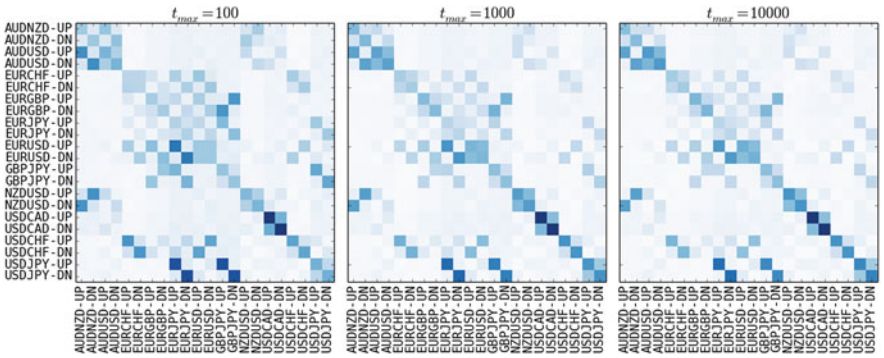


Fig. 1 Self-excitation and interaction behavior of exchange rate shifts, explored through the kernel matrix φ. Currency pairs in the study are given along axes

We posit that limiting the time horizon for interactions will lead to significant improvements in calculation time while having no major impact on model performance.

We learn parameters for three settings, with t_{max}, the maximum delay allowed, set to 100, 1000, and 10,000 ms respectively. We assume homogenous base intensities $\lambda_k(t)$. Results, along with computation times, are presented in Table 1. Model performance is given in *perplexity*, defined as $\exp(-\frac{1}{N} \log p(\mathscr{D}|\Theta))$. We measure perplexity on the training set of May 2017, and the held out test data set of June 2017. An implementation of the algorithm is made available online.[2]

Despite substantially longer time required for parameter estimation, extending the time frame for interactions has little benefit to the explanatory power. In fact, we find that it increases perplexity compared to looking for interactions only in a 1 s time frame, as a 10 s upper bound consistently results in higher perplexity in both training and test data sets.

Allowing longer time frames has the net effect of confounding our understanding of interactions among price-related events. This is intuitive from the perspective of the high-frequency trading domain, where *information* is consumed in very short time frames.

Relationships between assets are also recoverable under bounded delays. We present learned kernel matrices in Fig. 1 similar to [2]. Strong self-excitation behavior, both trend-following and mean-reverting, are observed along the diagonal. Interestingly, the trend following tendency appears stronger, contrary to findings in other markets. This may be a result of the fact that the momentum behavior in interbank markets are likely large order liquidations, and no *market takers* act in short time frames to trigger mean reversals. Meanwhile, strong off-diagonal elements are confirmed as shared currencies underlying the interacting pairs.

[2] http://github.com/canerturkmen/fasthawkes.

4 Conclusion

We introduced a bounded-delay HP model and an efficient parameter estimation algorithm. We demonstrate that imposing an upper bound on the expected delays of interactions between events results in efficient computation while preserving, or improving, explanatory power.

As immediate next steps, we will include a comparison with the standard exponential-decay Hawkes model. We will also explore whether the model and algorithm are able to deliver significant benefits in more realistic application scenarios; such as trading strategy development, optimal liquidation and price impact modeling.

References

1. Bacry, E., Mastromatteo, I., Muzy, J.F.: Hawkes processes in finance. Market Microstruct. Liquidity **1**(01), 1550005 (2015)
2. Bacry, E., Jaisson, T., Muzy, J.F.: Estimation of slowly decreasing Hawkes kernels: application to high-frequency order book dynamics. Quant. Financ. **16**(8), 1179–1201 (2016)
3. Bowsher, C.G.: Modelling security market events in continuous time: intensity based, multivariate point process models. J. Economet. **141**(2), 876–912 (2007)
4. Embrechts, P., Liniger, T., Lin, L.: Multivariate Hawkes processes: an application to financial data. J. Appl. Probab. **48**(A), 367–378 (2011)
5. Hawkes, A.G.: Point spectra of some mutually exciting point processes. J. R. Stat. Soc. Ser. B Methodol. 438–443 (1971)
6. Jaisson, T., Rosenbaum, M.: Limit theorems for nearly unstable Hawkes processes. Ann. Appl. Probab. **25**(2), 600–631 (2015)
7. Linderman, S., Adams, R.: Discovering latent network structure in point process data. In: International Conference on Machine Learning, pp. 1413–1421 (2014)
8. Muni Toke, I., Pomponio, F.: Modelling trades-through in a limited order book using Hawkes processes. Economics **6** (2012)
9. Murphy, K.: Machine Learning a Probabilistic Perspective. The MIT Press, Cambridge (2012)
10. Rambaldi, M., Filimonov, V., Lillo, F.: Detection of intensity bursts using Hawkes processes: an application to high frequency financial data (2016). arXiv:1610.05383 [q-fin]
11. Simma, A., Jordan, M.I.: Modeling events with cascades of poisson processes. In: Proceedings of the Twenty-Sixth Conference on Uncertainty in Artificial Intelligence, pp. 546–555. AUAI Press, Corvallis (2010)
12. Toke, I.M.: Market making in an order book model and its impact on the spread. In: Econophysics of Order-Driven Markets, pp. 49–64. Springer, Berlin (2011)
13. Veen, A., Schoenberg, F.P.: Estimation of space-time branching process models in seismology using an EM-type algorithm. J. Am. Stat. Assoc. **103**(482), 614–624 (2008)

Pricing Illiquid Assets by Entropy Maximization Through Linear Goal Programming

José Luis Vilar-Zanón and Olivia Peraita-Ezcurra

Abstract In this contribution we study the problem of retrieving a risk neutral probability (RNP) in an incomplete market, with the aim of pricing non-traded assets and hedging their risk. The pricing issue has been often addressed in the literature finding an RNP with maximum entropy by means of the minimization of the Kullback-Leibler divergence. Under this approach, the efficient market hypothesis is modelled by means of the maximum entropy RNP. This methodology consists of three steps: firstly simulating a finite number of market states of some underlying stochastic model, secondly choosing a set of assets—called benchmarks—with characteristics close to the given one, and thirdly calculating an RNP by means of the minimization of its divergence from the maximum entropy distribution over the simulated finite sample market states, i.e. from the uniform distribution. This maximum entropy RNP must exactly price the benchmarks by their quoted prices. Here we proceed in a different way consisting of the minimization of a different divergence resulting in the total variation distance. This is done by means of a two steps linear goal programming method. The calculation of the super-replicating portfolios (not supplied by the Kullback-Leibler approach) would then be derived as solutions of the dual linear programs.

Keywords Pricing · Hedging · Entropy · Goal programming

J. L. Vilar-Zanón (✉)
Department of Financial and Actuarial Economics & Statistics, Universidad Complutense de Madrid, Madrid, Spain
e-mail: jlvilarz@ucm.es

O. Peraita-Ezcurra
Banco de Santander, Boadilla del Monte, Spain
e-mail: olperaita@gruposantander.com

M. Corazza et al. (eds.), *Mathematical and Statistical Methods for Actuarial Sciences and Finance*, https://doi.org/10.1007/978-3-319-89824-7_91

1 Introduction

In this contribution we study the problem of retrieving a risk neutral probability (RNP) in an incomplete market, with the aim of pricing non-traded assets and hedging their risk. The pricing issue has been often addressed in the literature finding an RNP with maximum entropy by means of the minimization of the Kullback-Leibler divergence. So in this way the efficient market hypothesis is assumed and modelled by means of the maximum entropy RNP. This can be done in three steps: firstly simulating a finite number of market states from some underlying stochastic model, secondly choosing a set of assets—called benchmarks—with characteristics close to the given one, and thirdly calculating an RNP by means of the minimization of its divergence from the maximum entropy distribution over the simulated finite sample market states, i.e from the uniform distribution. This maximum entropy RNP must exactly price the benchmarks by their quoted prices.

Here we proceed in a different way consisting of the minimization of a different divergence resulting in the total variation distance. This is done by means of a linear goal programming method that calculates the minimal divergence value and returns an asset price belonging to the free arbitrage price interval defined between the buyer's and seller's prices. Different prices belonging to this interval are targeted by means of constraining the divergence being greater than its minimal value previously calculated. In a fourth step we tackle with the calculation of the super-replicating portfolios, derived as solutions of dual linear programs. These dual linear programs return the super-replicating portfolios for the asset, because in the optimal value of the dual objective the asset price returned by the primal is factored into the amounts of benchmarks composing the super-replicating portfolio plus an extra amount depending of the divergence value fixed in the primal.

The knowledge of the RNP constitutes, in the frame of Arrow-Debreu economy model and the Fundamental Pricing Theorem (for instance in [11] and [2], an interesting tool in areas such as valuing derivatives-hedging risk, see [3], portfolio management or the study of market risk aversion, see [10]. The history of the methods for risk neutral probabilities retrieval begins in [6], and the approach following the entropy maximization has been studied in [4, 5, 7, 15, 16]. The method of minimizing the Kullback-Leibler divergence has been applied in [1, 3, 8, 9, 14, 17]. Finally, linear programming is used in [13], and [12] to define in an incomplete market the buyer's and seller's prices and the free arbitrage interval of prices existing between them. In Sect. 2 we introduce the basic notation and in Sect. 2 we develop the goal programming methodology.

2 Notations and Preliminaries

We define a market model over one period with $I = \{1, 2, \ldots, N\}$ benchmarks and $J = \{1, 2, \ldots, M\}$ states. The benchmarks cash-flows are given in a matrix $\mathbf{G} = \left(g_{ij}\right)$, their respective prices in a vector $\mathbf{c} \in \mathbb{R}^{\mathbb{N}}$. A portfolio of benchmarks

is defined through $\mathbf{w} \in \mathbb{R}^{\mathbb{N}}$ with cash flows and price given respectively by $\mathbf{w}^T \mathbf{G}_{\bullet j}$, $j \in J$ and $\mathbf{w}^T \mathbf{c} = \sum_{i=1}^{N} c_i w_i$. We assume a frictionless market and the *no arbitrage hypothesis* . As a result of this last (see for instance [2]) there exist an RNP $\mathbf{q} \in [0,1]^M$ such that any asset with cash-flow $\mathbf{e} \in \mathbb{R}^{\mathbb{M}}$ can be valued through its $\mathbf{q}$-expectation discounted at the risk free interest $r : c = \frac{1}{1+r} E_{\mathbf{q}}\{\mathbf{e}\}$. In an incomplete market where $rank(\mathbf{G}) < M$, the RNP is not unique so many asset values can be calculated. Nevertheless, it is still possible to find an interval $\left(c_{\mathbf{e}}^{min}, c_{\mathbf{e}}^{max}\right)$ where the price should lie for the non-arbitrage condition still being fulfilled. Those bounds are found as solution to two linear programs that can be found in [13] and [12]. A general divergence of a probability $\mathbf{q}$ from the uniform distribution over the finite sample space J can be written as

$$D_{\psi}(\mathbf{q}|\mathbf{p}) = \sum_{j} \psi\left(\frac{q_j}{p_j}\right) q_j \tag{1}$$

for any convex function ψ (see [1, 2]). Now taking $\psi(x) = x \log(x)$ we get the Kullback-Leibler divergence, while taking $\psi(x) = |x-1|$ we get the total variation distance, so we can finally write

$$D_{\psi}(\mathbf{q}|\mathbf{p}) = \sum_{j} |\frac{q_j}{p_j} - 1| = \sum_{j} \left(q_j - p_j\right). \tag{2}$$

This is the objective function we have to minimize.

3 The Goal Programming Method

We now resume the process of pricing an illiquid asset by means of the minimization of that objective function.

Step 1: Simulating the market states. We choose some stochastic model and simulate it to obtain the M states forming the sample space J. For instance we could choose a model taking into account the extreme values phenomenon.

Step 2: Calibration of the RNP to the benchmark prices: For any market state $j \in J$ we define the *deviation variables* y_j^-, y_j^+. The following linear program must be solved to minimize the total variation distance (2):

$$\min_{\mathbf{x}=(\mathbf{q},\mathbf{y}^+,\mathbf{y}^-)} \sum_{j} \left(y_j^- + y_j^+\right), s.t. \begin{cases} q_j - p_j + y_j^- - y_j^+ = 0, & j \in J \\ \sum_j q_j g_{ij} - c_i = 0, & i \in I \\ \sum_j \left(y_j^- - y_j^+\right) = 0 & \\ q_j, y_j^-, y_j^+ \geq 0, & j \in J \end{cases}$$

Once the last step is solved, we take note of the calibrated RNP π and the minimal divergence $D^{min} \geq 0$

$$\left(\pi, \mathbf{y}^{+*}, \mathbf{y}^{-*}\right)/D^{min} = \sum_j \left(y_j^{-*} + y_j^{+*}\right)$$

Step 3: Asset valuing corresponding to a given divergence value: Now given an asset with cash-flow $\mathbf{e} \in \mathbb{R}^{\mathbb{M}}$, and fixing $D \geq D^{min}$, we solve:

$$\min_{\mathbf{x}=(\mathbf{q},\mathbf{y}^+,\mathbf{y}^-)} primal(\mathbf{x}) = \sum_j q_j e_j, s.t. \begin{cases} q_j - p_j + y_j^- - y_j^+ = 0, & j \in J \\ \sum_j \left(y_j^- + y_j^+\right) = D & \\ \sum_j q_j g_{ij} - c_i = 0, & i \in I \\ \sum_j \left(y_j^- - y_j^+\right) = 0 & \\ q_j, y_j^-, y_j^+ \geq 0, & j \in J \end{cases}$$

This could also be done with an identical linear program though maximizing the objective. Both LP would furnish two different RNP π^* and π_* that would be the same in case we choose $D = D^{min}$, when the two optimal values, i.e. the asset prices, would be the same.

Step 4: Calculating the super-replicating portfolios For a chosen divergence value $D \geq D^{min}$ in the previous step, consider that minimization (maximization) LP. Then we can write down its dual. For that sake we consider the dual variables corresponding to the prior uniform probabilities $\mathbf{w}_q^T \in \mathbb{R}^{\mathbb{M}}$, the divergence level $w_{M+1} \in \mathbb{R}$, the benchmarks portfolio $\mathbf{w}_B^T \in \mathbb{R}^{\mathbb{N}}$ and, finally, the probability constraint (the fourth one) $w_{M+N+2} \in \mathbb{R}$, so we finally have:

$$\mathbf{w}^T = \left(\mathbf{w}_q^T, w_{M+1}, \mathbf{w}_B^T, w_{M+N+2}\right).$$

Then the dual LP reads in matrix notation as:

$$\max_{\mathbf{w}}(\min) d(\mathbf{w}) = \mathbf{w}^T \begin{pmatrix} \mathbf{p} \\ D \\ \mathbf{c}_{M\times 1} \\ 0 \end{pmatrix},$$

$$s.t. : \left(\mathbf{w}_q^T, w_{M+1}, \mathbf{w}_B^T, w_{M+N+2}\right) \begin{pmatrix} \mathbf{I}_{M\times M} & \mathbf{I}_{M\times M} & -\mathbf{I}_{M\times M} \\ \mathbf{0}_{1\times M} & \mathbf{1}_{1\times M} & \mathbf{1}_{1\times M} \\ \mathbf{G}_{N\times M} & \mathbf{0}_{N\times M} & \mathbf{0}_{N\times M} \\ \mathbf{0}_{1\times M} & \mathbf{1}_{1\times M} & -\mathbf{1}_{1\times M} \end{pmatrix}$$

$$\leq (\geq) \left(\mathbf{e}_{1\times M}, \mathbf{0}_{1\times M}, \mathbf{0}_{1\times M}\right).$$

When valuing at the minimum divergence level (i.e. setting $D = D^{min}$), strong duality holds returning an optimal primal value (asset price) factored by the amounts of benchmarks composing the super-replicating portfolio plus an extra amount depending of the divergence value fixed in the primal and the prior uniform distribution $\mathbf{p}$:

$$d(\mathbf{w}^*) = \mathbf{w}_q^{*T}\mathbf{p} + w_{M+1}^* D^{min} + \mathbf{w}_B^{*T}\mathbf{c}_{M\times 1} = \mathbf{e}_{1\times M}\pi = primal(\mathbf{x}^*).$$

In the last formula we see how the optimal dual variables $\mathbf{w}_q^{*T}, w_{M+1}^*, \mathbf{w}_B^{*T}$ measure the sensibility of the asset value with respect to variations in the benchmarks prices, the prior probabilities and the divergence from its minimal value D^{min}.

4 Conclusions

We have given a summary of how we could value a non-traded asset and find its super-replicating portfolios by means of a linear goal programming method minimizing a divergence of the RNP from the maximum entropy distribution on the sample space, reducing to the total variation distance. Obtaining its solutions allows to attain all the values belonging to the interval of free arbitrage values between the seller's and the buyer's prices. If the divergence is set to its minimum D^{min}, we can measure the sensibility of the asset value with respect to changes of the benchmarks prices and the divergence value, because the dual factors the price by these components.

References

1. Avellaneda, M.: Minimum-relative-entropy calibration of asset pricing models. Int. J. Theor. Appl. Financ. **1**(4), 447–472 (1998)
2. Avellaneda, M., Lawrence, P.: Quantitative Modeling of Derivative Securities: From Theory to Practice. Chapman and Hall/CRC, London/Boca Raton (2000)
3. Avellaneda, M., Buff, R., Friedman, C., Grandechamp, N., Kruk, L., Newman, J.: Weighted Monte Carlo: a new technique for calibrating asset-pricing models. Int. J. Theor. Appl. Financ. **4**, 91–119 (2001)
4. Bose, C., Murray, R.: Maximum entropy estimates for risk-neutral probability measures with non-strictly-convex data. J. Optim. Theory Appl. **161**, 285–307 (2014)
5. Branger, N.: Pricing derivatives securities using cross entropy: an economic analysis. Working paper. London School of Economics (2003)
6. Breeden, D.T., Litzenberg, R.H.: Prices of state-contingent claims implicit in option prices. J. Bus. **51**(4), 621–651 (1978)
7. Buchen, P., Kelly, M.: The maximum entropy distribution of an asset inferred from option prices. J. Financ. Quant. Anal. **31**(1), 143–159 (1996)

8. Dhaene, J., Stassen, B., Devolder, P., Vellekoop, M.: The minimal entropy martingale measure in a market of traded financial and actuarial risks. J. Comput. Appl. Math. **282**, 111–133 (2015)
9. Frittelli, M.: The minimal entropy martingale measure and the valuation problem in incomplete markets. Math. Financ. **10**(1), 39–52 (2000)
10. Jackwerth, J.: Option-Implied Risk-Neutral Distributions and Risk Aversion. The Research Foundation of AIMR™ (2004). Available at http://www.ub.uni-konstanz.de/kops/volltexte/2008/5288/: Konstanzer Online-Publikations-System (KOPS). Cited 1 January 2016
11. Luenberger, D.: Investment Science. Oxford University Press, Oxford (1997)
12. Luenberger, D.: Arbitrage and universal pricing. J. Econ. Dyn. Control. **26**, 1613–1628 (2002)
13. Musiela, M., Rutkowsky, M.: Martingale Methods in Financial Modelling. Springer, Berlin (1998)
14. Neri, C., Schneider, L.: Maximum entropy distributions inferred from option portfolios on an asset. Financ. Stoch. **16**, 293–318 (2012)
15. Orozco-Rodríguez, J., Santosa, F.: Estimation of asset distributions from option prices: analysis and regularization. SIAM J. Financ. Math. **3**, 374–401 (2012)
16. Rockinger, M., Jondeau, E.: Entropy densities with an application to autoregressive conditional skewness and kurtosis. J. Econ. **106**(1), 119–142 (2002)
17. Samperi, D.J.: Inverse Problems, Model Selection and Entropy in Derivative Security Pricing, PhD thesis. New York University (1997)