UNIVERSITÀ DEGLI STUDI DI TORINO

Dipartimento di Statistica e Matematica Applicata

Dottorato di ricerca in ECONOMIA "VILFREDO PARETO"
Ciclo XXX

"Three Essays in Finance and Actuarial Science"

Tesi presentata da Anastasiia Novokreshchenova
Tutor prof. Elisa Luciano

Coordinatore del Dottorato prof. Fabio Bagliano

Anni Accademici 2014-2018

Settore Scientifico Disciplinare
"SECS-S/06 - metodi matematici dell'economia e delle scienze attuariali e finanziarie"

UNIVERSITY OF TURIN

DOCTORAL THESIS

Three Essays in Finance and Actuarial Science

Author:
Anastasia
NOVOKRESHCHENOVA

Supervisor:
Prof. Elisa LUCIANO
Prof. Claudio TEBALDI
Asst. Prof. Luca REGIS

A thesis submitted in fulfillment of the requirements for the degree of Doctor of Philosophy

in the

Scuola di Dottorato in Economia "V.Pareto", indirizzo in Statistica e Matematica Applicata
Dipartimento di Statistica e Matematica Applicata

April 30, 2018

Acknowledgements

I kindly thank my supervisor Elisa Luciano and co-supervisors Claudio Tebaldi and Luca Regis for their support during my Ph.D., fruitful discussions and suggestions. I am also grateful to the IGIER-Bocconi research team for their advice and for supporting me with the access to the data I use in the third Chapter of my dissertation. I would like to express my sincere gratitude to Céline Azizieh for hosting me at the Université Libre de Bruxelles during my mobility and for her helpful discussions on the second Chapter.

This work benefited of comments from participants to the 9th ASF Conference (Samos, 18-22 May 2016), the 2nd ICASQF Conference (Cartagena, 15-18 June 2016) and the 21st IME Congress (Vienna, 3-5 July 2017). In addition I would like to express my gratitude to the Fondazione Collegio Carlo Alberto and the Dipartimento di Statistica e Matematica Applicata of the University of Turin and their staffs for facilities and technical assistance. Finally, a sincere thank goes to my family for supporting me during these years.

Contents

To Galina and Vladimir

Chapter 1

Introduction

Nowadays mathematical modelling and data analytics approaches play a more and more important role in risk management of uncertainty. Stochastic differential equations and Monte Carlo simulations constitute essential parts of any risk management department both in insurance and banking sectors. In my thesis I aim to study and improve current approaches often employed in the risk management practice by using techniques from data analysis and mathematical modelling.

In mathematical terms any uncertainty about the future is modelled by stochastic processes. An important role in this area play affine processes. They have become very popular in financial applications due to their computational tractability and flexibility in capturing many of the empirical features of financial time series. The square-root process of Feller (1985) and the Ornstein-Uhlenbec process (1930) are primary examples of the affine processes which are widely applied in financial economics. They are used to model term structure of interest rates, instantaneous mortality intensity or stochastic volatility of equity prices such as in the model of Heston (1993).

In the first part of my thesis I quantitatively compare the forecasts from four different mortality models. I consider one discrete-time model proposed by Lee and Carter (1992) and three continuous-time models: the Wills and Sherris (2011) model, the Feller process and the Ornstein-Uhlenbeck (OU) process. The first two models estimate the whole surface of mortality simultaneously, while in the latter two, each generation is modelled and calibrated separately. I calibrate the models to UK and Australian population data and compare their performance in terms of accuracy and precision using backtesting procedures.

In the second part I study the problem of calibrating and modelling market price of risk in the context of stochastic volatility models. Under the Heston model it is possible to define a price of risk in a way that the state variable follows a square-root process under both an objective probability measure and an equivalent martingale measure. This is a convenient framework in terms of modelling since nowadays internal models in the context of Solvency II are required to produce both risk-neutral and real-world simulations – in particular, for calculation of the Solvency Capital Requirement (SCR). The problem of calibrating the price of risk is crucial in this setting. I propose a calibration procedure based on the maximum likelihood approximation approach of Ait-Sahalia and Kimmel (2007). The developed algorithm iteratively finds the parameters of the Heston model using two criteria – the minimal prediction error and the maximum likelihood approach. The procedure is calibrated using the data on the Eurostoxx50 index and its options for the last 17 years. This part of my thesis was developed at the department of Mathematics and Actuarial science of the Université Libre de Bruxelles under the supervision of Dr. Céline Azizieh.

The third part of the thesis, except for the notion of uncertainty in finance, also deals with the notion of interdependency. After the economic crises it became evident

that a high degree of interdependence within the financial system increases the systemic risk which can not be disregarded in credit assessment. I study how network theory can be used to improve the assessment of default credit models, both in the setting of structural and reduced-form models. In particular, I describe how to use a network effect for a bank which holds a portfolio of SME clients who are interconnected due to the trade credit relations between them. Under the reduced-form approach I have used the methods of spatial econometrics and network analysis to allow the inclusion of the network effect in the classical Z-score model. The results of the experiments proof the potential power of the network effect in the improvement of bunckrupcy prediction. In this paper I also describe other possible applications of the network constructed. Such, in the settings of the Merton (1974) model the network can be used to calculate the probability of default by pricing an option when the asset is correlated with other assets in the economy. The data was made available by a leading Italian institution and it was collected as part of a scientific project at the IGIER-Bocconi research centre where I was a part of the research team.

Chapter 2

Predicting Human Mortality: Quantitative Evaluation of Four Stochastic Models

2.1 Introduction

One of the main issues facing financial and governmental institutions, within the current economic climate, is the forecasting of mortality among an elderly population. Within a vast list of effected parties are public pension policies, private pension funds and life insurance businesses. They face the greatest risk, due to an increasing life expectancy across developed countries.

Over the last few decades it has become widely accepted that mortality can be more accurately measured by the use of stochastic models (see [25]), since they are better able to capture the uncertainty inherent within the problem. For any given individual, the probability of death naturally increases with age, however, as life expectancy increases over time, we observe improvements in mortality rates. Due to these effects, "dynamic mortality" has been introduced to produce models with age and time dependence. One of the seminal works, which became a benchmark within the industry, is the model of Lee and Carter [73] who model the central death rate as a two variable function. Since the publication of their work, several extensions of the Lee-Carter model have been proposed. For example, Renshaw-Haberman [87] considered a model that allows for a cohort effect and Blake and Dowd [26] proposed a two-factor model for mortality rates. Traditionally mortality models are used for forecasting mortality for older generations (ages over 50) since these mostly affect the uncertainty in the value of financial instruments offered by pension funds due to improvements in mortality and longer life expectancy (phenomena referred to in the literature as *longevity risk*). However, Plat [86] has recently suggested a model that can fit mortality to a wider range of ages (20–89). In [85] this model has been extended to fit even younger ages (5–89).

A fairly recent stream of actuarial literature has dealt with the phenomenon of stochastic mortality by modelling the instantaneous mortality intensity as a stochastic process. Recent works include Milevsky and Promislow [80], Dahl [32], Biffis [13], Denuit and Devolder [34], Luciano and Vigna [77], Shrager [89]. The mathematical framework in these models has been adapted from the credit risk literature to value securities subject to risk to default. Similarities between the time to default and remaining lifetime and between short-term interest rate and the force of mortality are exploited in this approach. Moreover, if the intensity process is affine, then the survival function for an individual can be derived in a closed form. This is extremely

useful when pricing mortality-linked financial products, such as endowments, annuities, variable annuities and other forms of mortality-linked financial securities.

Luciano and Vigna [77] have studied the applicability of the affine processes, such as Ornstein-Uhlenbeck and Feller, for modelling mortality intensities. The approach is focused on fitting the survival curve for which closed-form solutions are available. The future projections for survival probabilities are made, their closeness to the historical values is discussed, but not evaluated quantitatively.

Another continuous-time mortality model we consider in this work is the one proposed by Wills and Sherris (2011) [95] for the Australian population. As with the Lee and Carter (1992) [73] model, it is able to capture the whole "mortality surface" across age and period. Moreover, it takes account of the correlation structure between different generations. This is important for life offices portfolios which often have contracts written on individuals from different cohorts. The authors have shown that the multiple risk factors implied by the model reflect the actual correlation structure between generations inferred from the data and that the model is suitable for pricing financial instruments (see Wills and Sherris [95, 94]).

The advantages of continuous time mortality models mean that it is important to study how well continuous time processes can predict future mortality. There are numerous papers comparing the performance of mortality models. Nevertheless, most of them have focused on discrete stochastic mortality models. For example, Cairns et al. [24] examined the in-sample fits of eight different discrete time stochastic mortality models. However, as noted in Dowd et al. [35], it is quite possible for a model to provide a good in-sample fit to historical data and produce forecasts that appear plausible *ex ante*, but still produce poor *ex-post* forecasts, that is, forecasts that differ significantly from the subsequently realised outcomes. Consequently, a "good" model should produce forecasts that perform well out-of-sample which can be evaluated using backtesting methods.

Lee and Miller [74] evaluated the performance of the Lee-Carter model by examining the behaviour of forecast errors and plots of "percentile error distributions", although they did not report any formal test results. In contrast, Dowd et al. [35] formally evaluate the forecasting performance of six different stochastic mortality models applied to male mortality data for England and Wales. They use a backtesting procedure to test the stability of forecasts over different time horizons and conclude that the investigated models perform adequately, and that there is little difference between them.

The framework for backtesting stochastic mortality models in Dowd et al. [35] is a very general one. The "backtests" might involve the use of plots whose goodness of fit is interpreted informally, as well as formal statistical tests of predictions. The evaluation can be done for different metrics (the forecasted variable) of interest – possible metrics include mortality rates, life expectancy, future survival rates, the prices of annuities and other life-contingent financial instruments.

This paper focuses on the forecasting performance of several continuous-time models, making a novel contribution to the literature. More specifically, we concentrate on the following continuous-time mortality models: the Ornstein-Uhlenbeck process, the Feller process and the Wills and Sherris model.

To compare the performance of these models to a benchmark, we also include the Lee and Carter model in our experiments. We evaluate the in-sample goodness of fit by using statistical techniques including the BIC criteria and an analysis of the fitted residuals. To assess the forecasting performance of each model we employ out-of-sample back-testing methods using mortality rates as the metric. This is done first by computing the relative error between the forecast and actual mortality rates

and then by looking at the percentage of observed mortality rates which fall within a prediction interval. However, the same backtesting procedure using different metrics might be relevant for different purposes.

For our analysis we employ the data of the British and Australian population as they are among the countries where the market for mortality derivatives has started to emerge. According to [83], the annuity markets are relatively well developed in the UK and US. Some product innovations, such as variable annuities with guaranteed withdrawal lifetime benefits have been introduced in Australia, Japan and Europe. Multiple mortality and longevity derivatives (such as q-forwards, s-forwards, longevity and survivor bonds and swaps) have been suggested in the literature as well, see [94, 14]. In [94, 83] the authors study the securitisation of longevity risk for the Australian pension industry. In [76] natural hedging of longevity risk with application to the UK population is analysed.

This paper is organised as follows: in Section 2.2 we present some notation and description of the data that will be used in the subsequent analysis. In Section 2.3 we provide an overview of the Lee-Carter model, which we will use as a benchmark for our comparisons. Section 2.4 provides the Wills and Sherris model setup. Section 2.5 describes time-homogeneous affine processes. Section 2.6 calibrates the four models to the UK female dataset and Section 2.7 compares the results of this calibration for the four models. Section 2.8 discusses the robustness of the simulation results on the male and female datasets for the British and Australian populations. Section 2.9 concludes.

2.2 Notation and Data Description

Throughout the paper we use the following notation. Define $m(x,t)$ to be the *observed central death rate* in year t for lives initially aged x as a number of deaths divided by the population exposure:

$$m(x,t) = \frac{D(x,t)}{E(x,t)}, \tag{2.1}$$

Here $E(t,x)$ is the average size of the population aged x last birthday during year t and $D(t,x)$ is the number of deaths during year t recorded as age x last birthday at the date of death. The observed central death rate can be calculated directly from the data.

Another measure of mortality is the *force of mortality* $\mu(x,t)$. It is interpreted as the instantaneous death rate at exact time t for individuals aged exactly x at time t. The probability of death between t and $t+dt$ for small t is then approximately $\mu(x,t) \times dt$. Thus, assuming that the force of mortality remains constant over a year: $\mu(x+s,t+u) = \mu(x,t)$ for $0 \leq s,u < 1$, we can approximate the force of mortality $\mu(x,t)$ with the mortality rate $m(x,t)$.

A typical dataset consists of a number of deaths, $D(x,t)$, and the corresponding exposures, $E(x,t)$, over a range of years t and ages x. The data for the UK we use in this study contains the number of deaths and the population exposure. It was taken from the Human Mortality Database [1]. We consider female population aged 50–99 (which is relevant to the pension fund industry) during the years 1970–2009.

[1] Human Mortality Database. University of California, Berkeley (USA), and Max Planck Institute for Demographic Research (Germany)

2.3 Lee-Carter Model

In this section we describe general characteristics of the famous Lee Carter model [73] and its estimation process and forecasting technique. Lee-Carter mortality model is used widely in academia, as well as industry. It has been proposed by Lee and Carter in 1992 specifically for US mortality data covering years 1933-1987. However it has been used as a benchmark model to mortality data from many countries and time-periods. It has been shown (see [40]) that the Lee-Carter model is a special type of multivariate random walk with a drift (RWD), in which the covariance matrix depends on the drift vector. For estimation of the model parameters, principle component analysis (PCA) with a single component is applied to the census data.

Let m_{xt} denote the log of the mortality rate in an age group x $(x = 1, ..., A)$ and time t $(t = 1, ..., T)$ for one country. The mortality rate is modelled as follows:

$$m_{xt} = \alpha_x + \beta_x \kappa_t + \epsilon_{xt} \tag{2.2}$$

where ϵ_{xt} is a set of random disturbances and α_x, β_x and κ_t are parameters to be estimated:

- α_x is the average mortality curve across ages;
- β_x is a set of parameters representing the sensitivity of the mortality rate at age x to changes in κ_t;
- κ_t is a time-varying parameter representing a common risk factor;
- ϵ_{xt} is a zero mean Gaussian error $N(0, \sigma^2)$.

The parametrisation in (2.2) is not unique, since the likelihood function associated with the model above has an infinite number of equivalent maxima, each of which would produce identical forecasts, see Lee and Carter [73]. In practice, model identification implies imposing constrains. Lee and Carter adopt the constraints $\sum_t \kappa_t = 0$ and $\sum_x \beta_x = 1$.

The constraint $\sum_t \kappa_t = 0$ implies that the parameter α_x is simply the empirical average over time of the age profile in age group a: $\alpha_x = \bar{m}_x$. We can therefore rewrite the model in terms of the mean centered log-mortality rate, $\tilde{m}_{xt} = m_{xt} - \bar{m}_x$. Thus, we can rewrite Equation (2.2) as a multiplicative fixed effects model for the centered age profile:

$$\begin{aligned} \tilde{m}_{xt} &\sim N(\bar{\mu}_{xt}, \sigma^2), \\ E(\tilde{m}_{xt}) &= \bar{\mu}_{xt} = \beta_x \kappa_t. \end{aligned} \tag{2.3}$$

As a result, we use $A + T$ parameters (with A and T being the total number ages and the total number of years considered) to approximate the $A \times T$ elements of the mortality matrix, where each row represents the age of the population and each column represents the year of the observation, with the age-specific parameter β_x which is fixed over time for all x and the time-specific parameter κ_t which is fixed over age groups for all t.

The parameters β_x and κ_t in the model can be found easily using singular value decomposition (SVD) of the matrix of centered age profiles, $\tilde{m} = BLU' = Z$, which we denote by Z. Then the estimate for β_x is the first column of B, $\mathbf{b_1}$ (normalised eigenvector of the matrix ZZ') and the estimate for κ_t is $\lambda_1 \mathbf{u_1}$, where $\mathbf{u_1}$ is the first column of the matrix U (normalised eigenvector of the matrix $Z'Z$) and λ_1 is the

first element of the diagonal matrix L (the largest eigenvalue corresponding to the eigenvectors). Typically, for low-mortality populations, the approximation $Z \approx \lambda_1 \mathbf{b_1} \mathbf{u_1'}$ accounts for more than 90% of the variance of m_{xt}, see Girosi and King [40].

To forecast future mortality, Lee and Carter assume that α_x and β_x remain constant over time and the time factor γ_t is viewed as a stochastic process. They find that a random walk with drift is the most appropriate model for their data:

$$\begin{aligned} \hat{\kappa}_t &= \hat{\kappa}_{t-1} + \theta + \xi_t; \\ \xi_t &\sim N(0, \sigma^2_{rw}), \end{aligned} \tag{2.4}$$

where θ is known as the drift parameter and its maximum likelihood estimate is simply $\hat{\theta} = (\hat{\kappa}_T - \hat{\kappa_1})/(T-1)$, which only depends on the first and last components of the κ_t vector.

We can forecast $\hat{\kappa}_t$ at time $T+h$ with data available up to period T, as follow:

$$\hat{\kappa}_{T+h} = \hat{\kappa}_T + h\hat{\theta} + \sum_{l=1}^{h} \xi_{T+l-1}. \tag{2.5}$$

From this model, we can obtain forecast point estimates, which follow a straight line as a function of h with slope $\hat{\theta}$:

$$E[\hat{\kappa}_{T+h} | \hat{\kappa}_1, \ldots, \hat{\kappa}_T] = \hat{\kappa}_T + h\hat{\theta}. \tag{2.6}$$

To make a point estimate forecast for log-mortality we plug the obtained expression for $\hat{\kappa}_{T+h}$ into the vectorised version of expression (2.3):

$$\mu_{T+(\Delta t)} = \bar{m} + \hat{\beta}_x \hat{\kappa}_{T+h} = \bar{m} + \hat{\beta_x}[\hat{\kappa}_T + h\hat{\theta}], \tag{2.7}$$

where $\hat{\beta}_x = \mathbf{b_1}$ and $\hat{\kappa}_T = \lambda_1 \mathbf{u_1}$ are the estimates of β_x and κ_T respectfully obtained using SVD.

2.4 Wills and Sherris Model

Wills and Sherris suggested a stochastic longevity model where the force of mortality for age x at time t has the following dynamics (see [95, 94]):

$$\begin{aligned} d\mu(x,t) = (a(x+t)+b)\mu(x,t)dt + \sigma\mu(x,t)dW(x,t), \\ 0 < x < \omega, \quad 0 < t < \omega - x. \end{aligned} \tag{2.8}$$

In the above expression the drift parameter is an affine function of the current age $(x+t)$, while volatility function is a constant. Using Ito's formula, we find the solution to the SDE (2.8):

$$\mu(x,t) = \mu(x,0) \exp\left[\frac{a}{2}t^2 + (ax + b - \frac{1}{2}\sigma^2)t + \sigma W(x,t)\right],$$

which can be written as follows:

$$\ln\left[\frac{\mu(x,t)}{\mu(x,0)}\right] = \left[\frac{a}{2}t + ax + b - \frac{1}{2}\sigma^2\right] t + \sigma W(x,t).$$

For all ages $x_1, \dots x_N$, we consider a multivariate random vector of mortality rates:

$$\underline{\mu}(x,t) = \begin{bmatrix} \mu(x_1,t) \\ \vdots \\ \mu(x_N,t) \end{bmatrix}$$

The dynamics $\underline{d\mu}(x,t)$ are assumed to be driven by the multivariate Wiener process $\underline{dW}(x,t)$, with mean zero and the instantaneous correlation matrix given by:

$$D = \begin{bmatrix} \delta_{11} & \dots & \delta_{1N} \\ \vdots & \ddots & \vdots \\ \delta_{N1} & \dots & \delta_{NN} \end{bmatrix}$$

This means that the Wiener processes are independent between time periods, but correlated between ages and the multivariate Wiener process $\underline{dW}(x,t)$ can be expressed in terms of independent Wiener process $\underline{dZ}(x,t) = [dZ_1(t), \dots, dZ_N(t)]'$ as $dW(x,t) = DdZ(x,t)$.

Thus, the model described by Equation (2.8), becomes a system of equations where the dependence between the ages is captured by the $\delta_{x,i}$ term:

$$d\mu(x,t) = (a(x+t)+b)\mu(x,t)dt + \sigma\mu(x,t)\sum_{i=1}^{N}\delta_{x,i}dZ_i(t) \quad \forall x = x_1, \dots, x_N.$$

Using the fact that the distribution of the changes in the force of mortality follows a normal distribution, we can find the parameters $\hat{a}$, $\hat{b}$ and $\hat{\sigma}$ by means of maximum likelihood estimation. In particular,

$$\Delta\mu(x,t) \sim N((a(x+t)+b)\mu, \sigma\mu)$$

To estimate the covariance matrix of $\underline{dW}(x,t)$, we apply Principle Component Analysis (PCA) to the standardised residuals of the model. For each year, they are the realisations of the random vector $\underline{dW}(x,t)$:

$$r(x,t) = \frac{\Delta\hat{\mu}(x,t)/\hat{\mu}(x,t) - (\hat{a}(x+t)+\hat{b})}{\hat{\sigma}}$$

2.5 Time-Homogeneous Affine Processes

Mortality intensity since recently has been modelled as a stochastic process, (see Cairns [25]). In this field, an important stream of literature focuses on describing death arrival as the first jump time of a Poisson process with stochastic intensity. This approach is named doubly stochastic. Milevsky and Promislow [80] have used a stochastic force of mortality, whose expectation at any future date has a Gompertz specification. Dahl [32], Biffis [13], Denuit and Devolder [34] and Schrager [89] in modelling the stochastic force of mortality have applied the same mathematical tools used in the credit risk literature to model the time to default. Under this setting, the remaining lifetime of an individual, τ, is a doubly stochastic stopping time with intensity λ.

Let the mortality process $\mu_x(t)$ represent the mortality intensity of an individual belonging to the generation x at (calendar) time t and τ be the time of death of an individual of generation x. Then the survival probability from time t to time $T \geq t$ is

defined as a function of τ, $S_x(t,T)$, under the probability measure $\mathbb{P}$, conditional on the survivorship up to time t:

$$S_x(t,T) = \mathbb{P}(\tau \geq T | \tau > t), \tag{2.9}$$

A doubly stochastic stopping time is the analogue of the first jump time of a Poisson process, where the intensity is a stochastic process. If τ is the first jump time of a Poisson process with parameter μ, then

$$P(\tau > t) = e^{-\mu t} \tag{2.10}$$

Similarly, if τ is doubly stochastic with intensity μ, then the individual's survival function $S_x(t)$ is given by

$$S_x(t,T) = P(\tau > t | F_s) = E\left(e^{-\int_s^t \mu(u)du} | F_s\right) \tag{2.11}$$

In general, the expectation in (2.11) is not easy to calculate. However, if the intensity process is affine (see Duffie, Filipovic and Schachermayer [38]), then it is possible to provide the closed form for the survival probability:

$$S_x(t,T) = e^{\alpha(T-t)+\beta(T-t)\mu_x(t)}. \tag{2.12}$$

where the functions $\alpha(\cdot)$ and $\beta(\cdot)$ satisfy generalised Riccati ODEs, which can be solved analytically or at least numerically. The closed-from expression of survival probabilities (2.12) in affine framework allows to price financial instruments written on the underlying population, such as endowments, annuities, variable annuities and other forms of mortality-linked financial securities. Due to this result, in applications the processes selected for the mortality intensity are typically affine.

Luciano and Vigna [77] proposed and tested time-homogeneous non-mean reverting affine processes for the intensity of mortality, which are natural generalisation of the Gompertz law of mortality. They consider Ornstein Uhlenbeck process, Ornstein Uhlenbeck process with jumps and the Feller process. They provide the analytical solutions for survival function (2.12) for these processes and discuss the appropriateness of using them in modelling mortality. Calibrations on historical data show that despite their simple form, these processes fit mortality intensity dynamics very well. Another study shows how to use these processes to delta-gamma hedge mortality and interest rate risk, see Luciano, Regis and Vigna [76] .

2.5.1 The Ornstein-Uhlenbeck Processes

The SDE for the Ornstein-Uhlenbeck (OU) process without mean-reversion, with the associated solutions of the Riccati ODE $\alpha(\cdot)$ and $\beta(\cdot)$, is the following:

$$\begin{aligned} d\mu_x(t) &= a\mu_x(t)dt + \sigma dW_x(t), \\ \alpha(t) &= \frac{\sigma^2}{2a^2}t - \frac{\sigma^2}{a^3}e^{at} + \frac{\sigma^2}{4a^3}e^{2at} + \frac{3\sigma^2}{4a^3}, \\ \beta(t) &= \frac{1}{a}(1-e^{at}), \end{aligned} \tag{2.13}$$

where $a > 0$, $\sigma > 0$.

We calibrate the parameters of the OU process by means of Maximum Likelihood method applied to the mortality intensities.

Assume that the dynamics of the mortality intensity is described by the OU process without mean reversion as given by SDE (2.13). Then, the conditional probability density of an observation μ_{i+1}, given a previous observation μ_i (with a δ time step between them), has a form (here we omit x which symbolises a certain generation):

$$f(\mu_{i+1}|\mu_i; a, \hat{\sigma}) = \frac{1}{\sqrt{2\pi\hat{\sigma}^2}} e^{-\frac{(\mu_{i+1}-\mu_i e^{a\delta})^2}{2\hat{\sigma}^2}},$$

where $\hat{\sigma}^2 = \sigma^2 \frac{1-e^{2a\delta}}{2a}$.

The log-likelihood function of a set of observations $\bar{\mu} = (\mu_1, \mu_2, \ldots, \mu_n)$ can be derived from the conditional density function:

$$\begin{aligned} L(\bar{\mu}; a, \hat{\sigma}) &= \sum_{i=1}^{n} \ln f(\mu_{i+1}|\mu_i; a, \hat{\sigma}) = \\ &= -\frac{n}{2}\ln(2\pi) - n\ln(\hat{\sigma}) - \frac{1}{2\hat{\sigma}^2}\sum_{i=1}^{n}(\mu_{i+1} - \mu_i e^{a\delta})^2. \end{aligned} \tag{2.14}$$

From the Maximum Likelihood conditions we find the following equations for the parameters:

$$\begin{aligned} a &= \frac{1}{\delta}\frac{\sum_{i=1}^{n}\mu_{i+1}\mu_i}{\sum_{i=1}^{n}\mu_i^2} \\ \hat{\sigma}^2 &= \frac{\sum_{i=1}^{n}(\mu_{i+1} - \mu_i e^{-a\delta})^2}{n} \end{aligned} \tag{2.15}$$

The OU process in general can produce negative paths. The probability of λ_x turning negative is

$$P(\mu_x(t) \leq 0) = \Phi\left(-\frac{\mu_x(0)e^{at}}{\sigma\sqrt{\frac{e^{2at}-1}{2a}}}\right) = \Phi\left(\zeta(\sigma, a)\right), \tag{2.16}$$

where Φ is the distribution function of a standard normal.

In fact, the function $\zeta(\sigma, a) = -\frac{\mu_x(0)e^{at}}{\sigma\sqrt{\frac{e^{2at}-1}{2a}}}$ is increasing in σ and decreasing in a, as well as the probability of negative values of μ. In mortality modelling applications the probability that $\mu(t)$ takes negative values is very small, because in practice the obtained value of σ is small enough and the value of a, on the contrary, is high enough. In our calibration we check that the values of the obtained parameters are such that there are no negative intensities. Otherwise, to keep mortality intensity positive, it is possible to impose a restriction using Equation (2.16) during the parameter search, such that probability (2.16) is negligible (see Luciano and Vigna [77] for more details).

2.5.2 The Feller Process

The fourth model is the Feller process which is described by the following SDE with the associated solutions of the Riccati ODEs $\alpha(\cdot)$ and $\beta(\cdot)$:

$$\begin{aligned} d\mu(t) &= a\mu(t)dt + \sigma\sqrt{\mu(t)}dW(t), \\ \alpha(t) &= 0, \\ \beta(t) &= \frac{1-e^{bt}}{c+de^{bt}}, \end{aligned} \tag{2.17}$$

where $a > 0$, $\sigma \geq 0$, the boundary conditions are $\alpha(0) = 0$ and $\beta(0) = 0$, and the coefficients are:

$$\begin{aligned} b &= -\sqrt{a^2+2\sigma^2} \\ c &= \frac{b+a}{2}, \\ d &= \frac{b-a}{2}. \end{aligned} \tag{2.18}$$

The solution to the Equation (2.17) has the form:

$$\mu(t) = \mu(0)e^{at} + \sigma\int_0^t e^{a(t-u)}dW(u) \tag{2.19}$$

The Feller process is a type of the Cox, Ingersoll, Ross (1985) process [31] without mean reversion. It was proposed as a model of a short rate for financial market, referred to as the CIR model. This model is described by the following SDE:

$$dr(t) = a(b-r(t))dt + \sigma\sqrt{r(t)}dW(t), \tag{2.20}$$

where $b > 0$ is the mean-reversion level. Thus, the model suggests that the $r(t)$ is pulled towards b at a speed controlled by a. If condition $2ab > \sigma^2$ holds and $r(0) > 0$, then the CIR process remains strictly positive, almost surely, and the state (marginal) distribution of the process is steady. The marginal density is gamma-distributed. The maximum likelihood estimation of the parameter vector $\theta = (a, b, \sigma)$ is based on the transition density. Given r_t at time t the density of $r_{t+\Delta t}$ at time $t + \Delta t$ is

$$p(r_{t+\Delta t}|r_t;\theta) = ce^{-u-v}\left(\frac{v}{u}\right)^{\frac{q}{2}} I_q(2\sqrt{uv}), \tag{2.21}$$

where

$$\begin{aligned} c &= \frac{2a}{\sigma^2(1-e^{-a\Delta t})} \\ u &= cr_te^{-a\Delta t}, \\ v &= cr_{t+\Delta t}, \\ q &= \frac{2ab}{\sigma^2} - 1, \end{aligned} \tag{2.22}$$

and $I_q(2\sqrt{uv})$ is modified Bessel function of the first kind and of order q. Then the likelihood function of the time series $(r_1 \dots r_N)$ with the time between two observations

$\Delta t = 1$ is

$$L(\theta) = \prod_{i=1}^{N-1} p(r_{i+1}|r_i;\theta), \tag{2.23}$$

from which the log-likelihood function of the CIR process is derived:

$$\ln L(\theta) = (N-1)\ln c + \sum_{i=1}^{N-1}\left[-u_i - v_{i+1} + 0.5q\ln\left(\frac{v_{i+1}}{u_i}\right) + \ln(I_q(2\sqrt{u_i v_{i+1}}))\right], \tag{2.24}$$

where $u_i = cr_i e^{-a\Delta t}$ and $v_{i+1} = cr_{i+1}$.

There are two approaches to calibrating affine mortality processes to the historical data. One is to match the survival function (Equation (2.12)) using the solutions of the Riccati ODEs for the OU (Equation (2.13)) and the Feller (Equation (2.17)) processes to the set of observed survival probabilities. Another approach is to maximise the likelihood function of the transition density. In this work we employ the second approach as both for the OU process and the Feller process the transition density is known in closed-form.

2.6 Models Calibration

In this section we work with the UK female dataset which describes the mortality in population aged 50–99 for the years 1970–2009. First, we divide the data in two data sets: the estimation data set, containing 30 years of observations, from 1970 until 1999; and the backtesting data set containing the last 10 years of observations, from 2000 till 2009. First we estimate the model parameters on the estimation data set, then we make 10-years predictions of mortality rates and calculate how well the forecast is compared with actual mortality rates for the period 2000–2009.

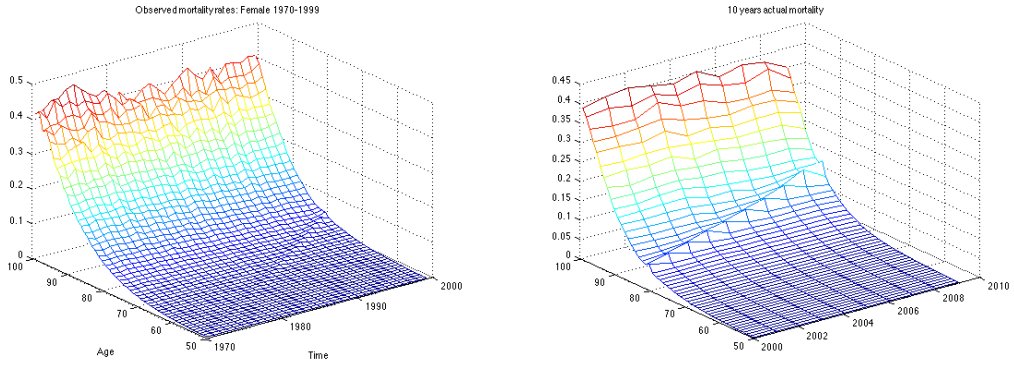
For the Lee and Carter and the Wills and Sherris models we use the whole surface of mortality to calibrate the models. Then, to compare the performance of the models between each other, we chose 19 generations. To have reliable estimation results and to make the comparison between the models which simulate the whole mortality surface (the Lee Carter and the Wills and Sherris models) and the ones which model each generation separately (the OU and the Feller processes) fairer, we take all possible generations from the data, which satisfy the criteria that the length of the backtesting period would not be less than 10 years. This results in 19 generations—aged 42–60 in the year 1970. We obtain the mortality rates for corresponding generations from the surface by taking the relative diagonal of the matrix. For the OU and the Feller processes, however, we calibrate the parameters for each of the 19 generations separately. We calculate the parameters on the estimation time period (1970–1999) and then use them to make forecasts of mortality for the next 10 years. Thus, we build forecasts for these generations and compute the relative error of prediction, as well as the percentage value of the actual mortality rates which fall within the prediction interval in the test period 2000–2009.

2.6.1 Calibration of the Lee-Carter Model

First of all, we compute the average of the log mortality m_{xt} for every age over time period 1970-1999 for the estimation dataset and subtract it in order to obtain mean centered log-mortality rates, $\tilde{m}_{xt} = m_{xt} - \bar{m}_x$. The average of the log mortality for the whole dataset is shown in Figure 3.1. Then, we perform SVD on $\tilde{m}_{xt}$ matrix and

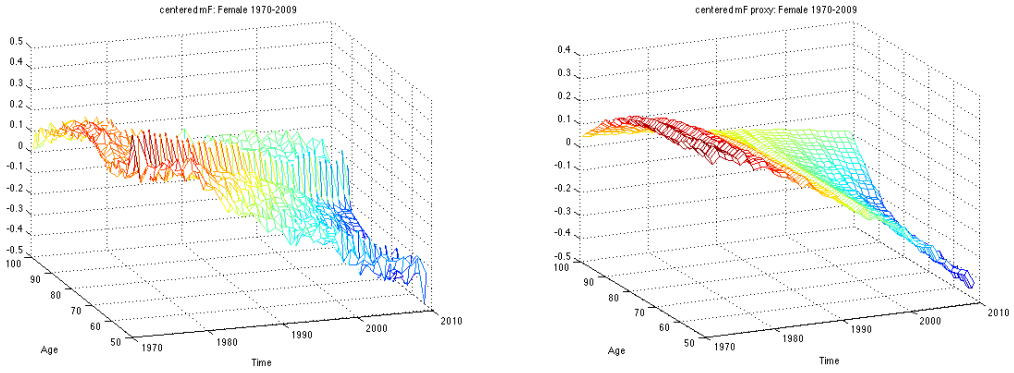
obtain estimates for parameters – two vectors $\hat{\beta}_x$ and $\hat{\kappa}_t$. The actual centered mortality and its SVD approximation are illustrated in Figure 3.2.

The obtained ML estimates for the drift and the variance of the innovations are $\hat{\theta}_{ML} = -0.5992$ and $\hat{\sigma^2}_{rw} = 0.9154$, respectively. Using these parameters we can compute the forecast for κ_t as given by Equation (2.4) and its forecast point estimate as described in Equation (2.6). In Figure 2.3a the estimated vector of κ_t and its forecast obtained for the next 10 years (in red) are shown. Then, we calculate the forecast for log-mortality as given in Equation (2.7). Figure 2.3b shows the mortality for the 10 years forecasted by the Lee-Carter. The forecast corresponds well to the observed mortality rates for the UK female population presented in Figure 3.1b. However, we can see that the cohort effect (diagonal trends in the data present in Figure 3.1b) is not captured by the Lee-Carter forecast of mortality.



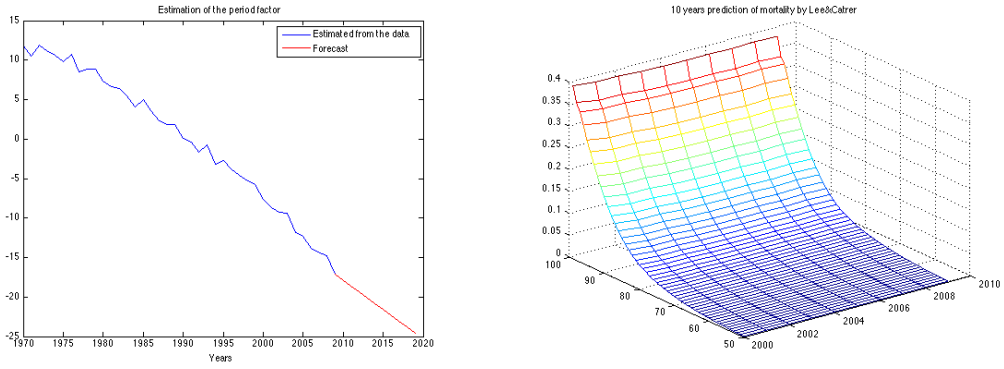
(**a**) (**b**)

FIGURE 2.1: Observed mortality rates for the UK female population. (**a**) Estimation data set, 1970–1999; (**b**) Backtesting data set, 2000–2009



(**a**) (**b**)

FIGURE 2.2: Actual centered mortality and its approximation. (**a**) Mean centered log-mortality rate; (**b**) Approximation by 1-factor SDV.



(**a**) (**b**)

FIGURE 2.3: Results of the Lee-Carter model. (**a**) Estimation and forecast of κ_t; (**b**) Lee-Carter forecast for 2000–2009.

2.6.2 Calibration of the Wills and Sherris Model

The analysis of the fit is based on the assumption that the residuals are independent and identically distributed normal variables with mean 0 and standard deviation 1. Figure 2.4a shows the graph of residuals for the UK female population aged 50–99, years 1970–1999 (the estimation dataset). The plot, together with the residuals descriptive statistics in Table 2.1, supports the hypothesis that the residuals follow a standard normal distribution with mean close to zero and standard deviation very close to one. The table also contains the value of the log likelihood function, the BIC criteria and the value of the χ-square statistics.

The Bayesian information criteria (BIC) is defined as

$$BIC = -2\ln(\hat{L}) + k\ln(n),$$

where:

n – the number of observations (sample size);

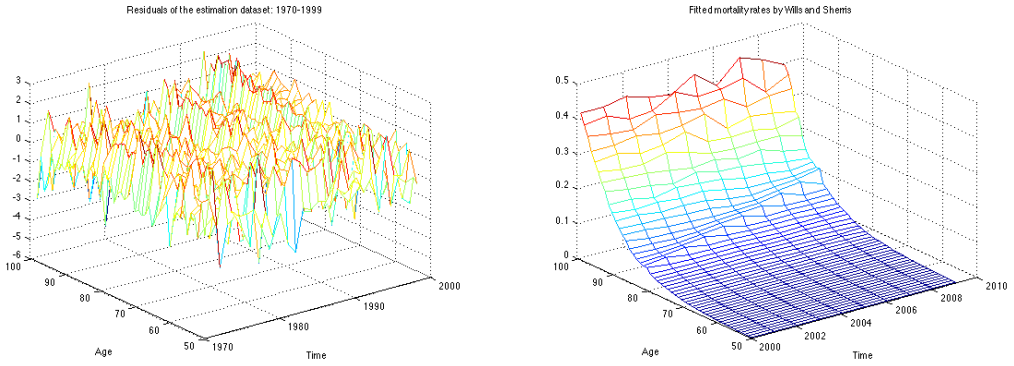
k – the number of free parameters to be estimated.

$\hat{L}$ – the maximized value of the likelihood function of the model.

Pearson's chi-square statistics, defined as $\chi^2 = \sum_{observations} \frac{(O_i - E_i)^2}{E_i}$, allows us to evaluate the goodness of fit by testing wether or not an observed frequency distribution differs from the theoretical one. We compare weather the computed value of χ^2 with the critical value of the statistic with degree of freedom defined as

$$df = \text{number of observations} - \text{number of independent parameters} - 1.$$

The obtained value of the χ^2 is 1.5835. This is compared to the chi-square distribution with 217 degrees of freedom $(49 * 29 - (49^2/2 + 3) - 1)$. Higher values of the χ^2 statistic suggest a poorer fit. Since the calculated value is very low, the test confirms a very good fit to the data.



(**a**) (**b**)

FIGURE 2.4: Results of the Wills and Sherris model. (**a**) UK female 1970-1999 fitted residual; (**b**) Mortality forecast for the test period.

Parameter Estimates	
a	0.0007032
b	0.0850
σ	0.0385
Log-likelihood	7246.4
BIC-criteria	-14480
Residual Descriptive Statistics	
mean	$1.3192 * 10^{-15}$
Minimum	-5.1384
Maximum	2.9886
Standard Deviation	1.0004
Standard Error	0.0125
Confidence Level	0.0003
χ^2	1.5835

TABLE 2.1: Parameter estimates and residual descriptive statistics for the Wills and Sherris model fit to UK female mortality rates 1970–1999.

To capture the correlation structure between ages we calculate eigenvectors of the matrix of the obtained residuals using Principal Component Analysis. Table 2.2 summarises the percentage of the observed variation explained by these vectors. The observed age-correlation matrix has a total of 49 eigenvalues. In our experiments we take first 30 eigenvectors to approximate the correlation matrix as they account for 98.9% of the variation.

Number of Eigenvectors	% of Observed Variation
1	28.1
5	55.8
10	75.4
15	86.5
20	93.1
25	96.8
30	98.9
35	100

TABLE 2.2: Percentage of the observed variation in residuals explained by the eigenvectors using PCA.

Figure 2.4b shows the mortality surface for the test period built with the Wills and Sherris model using the parameters obtained on the estimation dataset. By comparing the forecast with the actual mortality rates (Figure 3.1b), we can see that the model gives projections which are similar to the real data, although we see more variation in the simulated mortality intensities. In order to obtain a reliable prediction of mortality rates for a particular generation, we perform Monte Carlo simulations of the mortality surface for the test period and extract from the surface a diagonal corresponding to a specific generation. Then we estimate the mean of the Monte Carlo simulations for a given generation, together with the 90% prediction interval.

2.6.3 Calibration of the OU-Process

We calibrate the model on 19 generations and evaluate the goodness of fit by means of the BIC criteria and analysis of the residuals. For each generation x, having a series of length N we use $n = N - 10$ observations (first n years of the sample) to estimate the parameters a and σ and last 10 observation for backtesting the results. For instance, for the UK data, if we consider individuals who were 50 in the year 1970, and we have the data until the year 2009, we have 40 years of observations. Then the first thirty years of observations (1970–1999) is the estimation data and the last ten years of observations (2000–2009) is the backtesting data.

After obtaining the parameters we use the following simulation equation to generate paths of the mortality intensities. This expression is an exact solution of the SDE (2.13):

$$\mu_{i-1} = \mu_i e^{a\delta} + \sigma\sqrt{\frac{1 - e^{2a\delta}}{-2a}} N_{0,1}. \tag{2.25}$$

From the mortality intensities one can easily obtain the survival probabilities by means of the analytical formula 2.12 with $t = T_{j+1} - T_j = 1$. In this way the expression for a one-year survival curve with α and β being constants is:

$$\begin{aligned} S_x(T_{j+1}, T_j) &= e^{\alpha + \beta \mu_{x_j}}, \\ \alpha &= \frac{\sigma^2}{2a^2} - \frac{\sigma^2}{a^3} e^a + \frac{\sigma^2}{4a^3} e^{2a} + \frac{3\sigma^2}{4a^3}, \\ \beta &= \frac{1}{a}(1 - e^a), \end{aligned} \tag{2.26}$$

However, in our study we focus on mortality intensities. We obtain the parameters using the estimation dataset as described above, after which we use them to generate paths and to forecast mortality intensities. Finally, we calculate the error between the forecasted mortality curve and the actual mortality rates.

The residuals of the model are the realisations of the random component $dW(t)$ which should follow the standard normal distribution if the parameters are estimated correctly:

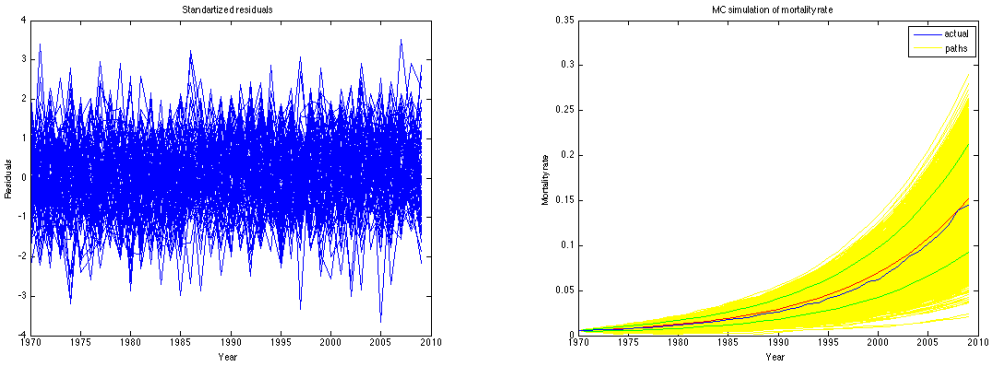
$$\frac{\Delta\mu - a\mu}{\sigma} \sim N(0,1).$$

We use Kolmogorov-Smirnov statistic to test hypothesis that the errors come from a standard normal distribution.

Taking the mortality intensities for the 7 generations we obtain the parameters presented

In Table 2.3 we report the obtained parameters for selected 7 generations. As expected, the a parameter is increasing with age, which means that the average mortality intensity is larger for older generations. The σ parameter is also growing with age. This proves the fact that there is more uncertainty in mortality rates for older ages.

Figure 3.8a represents the residuals of the model. According to the Kolmogorov-Smirnov test, the errors of the model are confirmed to be standard normal at 5% significance level. Figure 3.8b illustrates historical mortality intensities (in blue), 1000 MC simulations (in yellow), average among simulations (in red) and confidence intervals (in green) for the entire period (1970–2009). This graph is done for the generation aged 51 in the year 1970.



(**a**) (**b**)

FIGURE 2.5: Results of the OU-process for UK female generation aged 51 in the year 1970. (**a**) Fitted residuals of the model (1971–2009); (**b**) Historical mortality and simulated paths (1971–2009).

Generation Age in 1970	a	σ	MaxLogLikelihood	BIC
60	0.1024	0.0020	138.1895	−269.5766
57	0.0999	0.0015	146.5591	−286.3159
54	0.0951	0.0008	164.1945	−321.5865
51	0.0894	0.0008	165.3501	−323.8979
48	0.0845	0.0004	168.9370	−331.2095
45	0.0841	0.0004	155.7339	−305.0300
42	0.0815	0.0004	136.3726	−266.5631

TABLE 2.3: ML parameters of the OU-process and maximised log-likelihood.

2.6.4 Calibration of the Feller Process

We have maximised the log-likelihood function as it is stated in Equation (3.25) assuming that the mean-reversion parameter is zero. Table 2.4 reports the obtained parameters and the value of the maximazed log-likelihood function for each generation. The obtained parameter values correspond well with the previous work, such as in Luciano and Vigna [77] .

Generation Age in 1970	a	σ	MaxLogLikelihood	BIC
60	0.0984	0.0073	147.1586	−287.5148
57	0.0955	0.0061	156.6162	−306.4299
54	0.0922	0.0043	171.9529	−337.1034
51	0.0869	0.0046	172.7420	−338.6817
48	0.0833	0.0036	170.9023	−335.1403
45	0.0826	0.0034	155.2175	−303.9972
42	0.0801	0.0031	139.2860	−272.3900

TABLE 2.4: ML parameters of the Feller process and maximised log-likelihood.

The simulation of the future mortality is performed by discretising Equation (2.17) with time step equal to one year:

$$\mu_{t+1} = \mu_t + a\mu_t + \sigma\sqrt{\mu_t}N_{(0,1)}. \tag{2.27}$$

2.7 Comparison of the Four Models

To compare the performance of the models for the 19 generations based on their age in 1970. For each, we forecast mortality rates in the period 2000–2009 and compute the relative error of prediction, as well as the percentage of the observed mortality rates in the test period which fall within the prediction interval. The forecast and the prediction bounds are obtained using 15,000 Monte Carlo simulations.

In this section we define the tests of the mean relative error and the prediction intervals. A model can perform very well with respect to the percentage of the mortality rates which fall within the prediction bounds, while at the same time having a high relative error, if its variance grows rapidly and, therefore, the model

produces wide prediction bounds. We say that a model is *precise* if its forecasts of mortality are consistent with respect to the prediction interval and that a model is *accurate* if its mean absolute forecast errors are small. Of course, it is desirable for a good model to be both accurate and precise. To interpret the results of the experiments, it is important to understand the form of the variance implied by each model which we discuss in this section as well.

2.7.1 Relative Error

For each x, t the *relative error* is defined as follows:

$$error_x(t) = \frac{\mu_x^{predicted}(t) - \mu_x^{actual}(t)}{\mu_x^{actual}(t)} \quad \forall x, t.$$

Since the longevity risk corresponds to lower-than-expected mortality rates, we define the error so that it is positive if the forecast of mortality exceeds the historical values (actual values are lower-than-expected), and negative in the opposite case.

We compute the relative error for 19 generations—they are female aged 42–60 in the base year 1970. Thus, in the test period, for which the graphs of error are plotted, they are 30 years older—72–90 years, respectively.

The results of the experiments are presented in Figures 2.6 and 2.7 and Tables 2.5 and 2.6. The graphs of the mean absolute errors in Figure 2.7 illustrate the results shown in Tables 2.5 and 2.6. We can see that most of the errors fall in the range $[-0.1; 0.1]$. The exception is the OU process for the generation aged 60 in 1970, especially for later years of projections. The error for this generation in the Feller process forecast is also large – its absolute mean for generation aged 60 in 1970 is 0.0427 (Table 2.5). This increases the mean absolute errors for this generation for these processes shown in Table 2.5. Figure 2.7b shows the relative absolute error for each year in the test period average over 19 generations. We see that the error is smaller for younger ages . All the models show a high error for the generation aged 50 in the year 1970. This might be due to the cohort effect which is generally present in the UK data. The biggest error for this generation is produced by the Lee-Carter model. The graph of the relative absolute errors averaged over generations by year (Figure 2.7a) shows an increasing trend for all 4 models, especially for the Lee and Carter model (red line). This effect is due to the fact that the variances of the projected mortality rates increase with projection time. However, this does not happen at the same rate in different models.

The errors of the Lee-Carter model are mostly positive (Figure 2.6a) and we can observe a relative increase of the errors in time for each generation, indicating that the Lee-Carter model has tended to predict mortality rates that are too high. The errors of the Wills and Sherris model exhibit two patterns for different generations (Figure 2.6c). They are negative for the older generations (aged 60, 57, 54, 51 and 48 in 1970) and positive for the two youngest ones (45, 42 in 1970). We note that the older generations belong to the lower diagonal of the initial mortality matrix, while the younger two belong – to the upper diagonal of the matrix. Thus, it may be that Wills and Sherris model has a tendency to overestimate the mortality for younger generations and underestimates the mortality for the older ones.

According to the Tables 2.5 and 2.6, the OU process exhibits the lowest mean absolute error, followed by the Feller process, Wills and Sherris model and Lee and Carter model for the UK female data.

Age in 1970	Wills and Sherris	Lee-Carter	OU-Process	Feller
60	0.0383	0.0387	0.0837	0.0428
59	0.0346	0.0245	0.0543	0.0251
58	0.0396	0.0232	0.0562	0.0248
57	0.0457	0.0256	0.0430	0.0269
56	0.0374	0.0313	0.0573	0.0218
55	0.0420	0.0611	0.0209	0.0313
54	0.0515	0.0599	0.0220	0.0376
53	0.0379	0.0965	0.0160	0.0383
52	0.0362	0.0696	0.0260	0.0320
51	0.0320	0.0337	0.0169	0.0404
50	0.1008	0.1630	0.0919	0.1164
49	0.0388	0.0249	0.0256	0.0405
48	0.0353	0.0441	0.0383	0.0588
47	0.0258	0.0406	0.0257	0.0393
46	0.0212	0.0258	0.0251	0.0409
45	0.0279	0.0296	0.0263	0.0298
44	0.0267	0.0299	0.0193	0.0228
43	0.0312	0.0509	0.0221	0.0285
42	0.0334	0.0611	0.0232	0.0220
Mean (Rank)	0.0388 (3)	0.0492 (4)	0.0365 (1)	0.0379 (2)

TABLE 2.5: Mean (over 10 years) of the absolute errors for each generation, UK female data.

Year (Rank)	Wills and Sherris	Lee-Carter	OU-Process	Feller
2000	0.0314	0.0380	0.0379	0.0318
2001	0.0248	0.0417	0.0367	0.0247
2002	0.0267	0.0320	0.0191	0.0187
2003	0.0493	0.0316	0.0241	0.0462
2004	0.0332	0.0473	0.0321	0.0204
2005	0.0409	0.0401	0.0254	0.0311
2006	0.0326	0.0642	0.0405	0.0281
2007	0.0457	0.0537	0.0366	0.0475
2008	0.0626	0.0479	0.0483	0.0690
2009	0.0403	0.0951	0.0645	0.0614
Mean (Rank)	0.0388 (3)	0.0492 (4)	0.0365 (1)	0.0379 (2)

TABLE 2.6: Mean (over 19 generations) of the absolute errors for each year, UK female data.

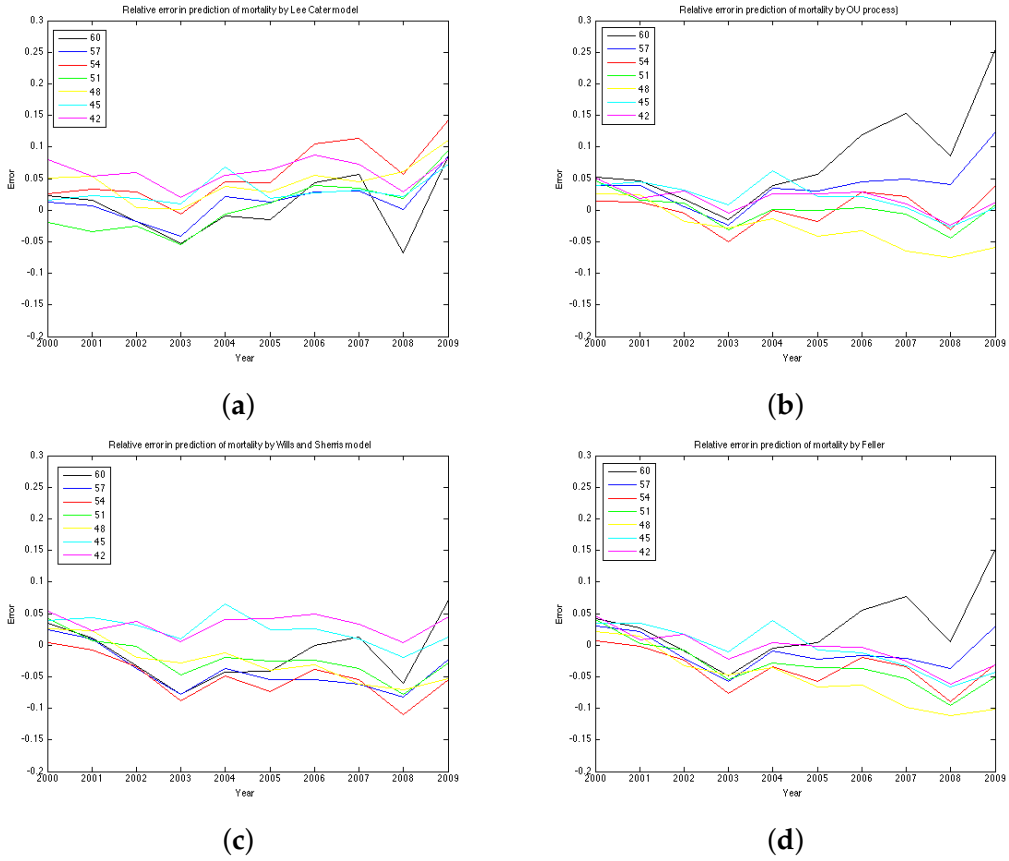


FIGURE 2.6: Relative error of each model for every generation (**a**)–(**d**). (**a**) Lee-Carte model; (**b**) OU-process; (**c**) Wills-Sherris model; (**d**) Feller process.

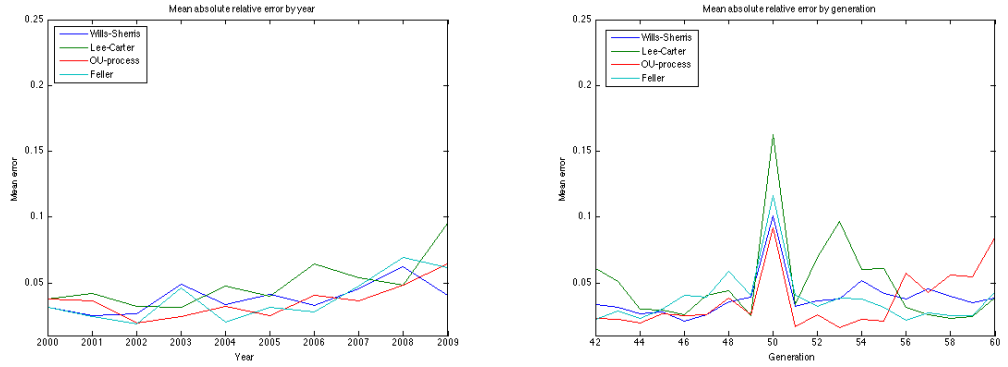


FIGURE 2.7: Mean relative absolute error of each model, UK female data. (**a**) Relative absolute error, average by year; (**b**) Relative absolute error, average by generation.

2.7.2 Discussion on the Variances

As it has been stated in the description of the model, the variance of the mortality intensity $\mu(t)$, conditional on time 0, in the OU specification has a form $\sigma^2 \frac{e^{2at}-1}{2a}$, where t is the time elapsed. For the Feller process, when intensity $\mu(t)$ is specified by

the CIR process of the form:

$$d\mu(t) = (b + a\mu(t))dt + \sigma\sqrt{\mu(t)}dW(t),$$

with $a > 0$, $b > 0$, $\sigma > 0$, the conditional distribution of the mortality intensity at time t, conditional on time 0 is given by a non-central chi-square distribution:

$$\mu(t) \sim \frac{\sigma^2(e^{at} - 1)}{4a}\chi^2_d(\nu),$$

where $\chi^2_d(\nu)$ denotes the density of a non-central chi-square random variable with d degrees of freedom:

$$d = \frac{4b}{\sigma^2},$$

and the non-centrality parameter ν is

$$\nu = \frac{4ae^{at}}{\sigma^2(e^{at} - 1)}\mu(0).$$

The $\chi^2_d(\nu)$ distribution, has a variance $Var_{\chi^2_d(\nu)} = 2(d + 2\nu)$. Thus, intensity $\mu(t)$ has a variance :

$$\frac{\sigma^2(e^{at} - 1)}{2a}\left(\frac{4b}{\sigma^2} + \frac{8ae^{at}}{\sigma^2(e^{at} - 1)}\mu(0)\right)$$

In the Feller specification, parameter b is not defined and, hence, the number of degrees of freedom d is not defined either. However, we can see that, other parameters being equal, the variance of the OU process should grow faster in time than the variance of the CIR process, as it has e^{2at} term rather than e^{at}.

The Wills and Sherris model assumes that the distribution of the changes in the force of mortality follows a normal distribution:

$$\Delta\mu(x, t) \sim N((a(x + t) + b)\mu, \sigma\mu).$$

Thus, the variance of the mortality intensity grows in time as at each time installment it is multiplied by the mortality rate from the year before.

For the Lee Carter model, the variance of the logarithm of the mortality rate at time t for each age x, is $\beta_x\hat{\sigma}^2_{rw}\sqrt{t}$, where $\hat{\sigma}_{rw}$ is the variance of the random walk process κ_t, in our case equal to 0.9154, and t is the time passed.

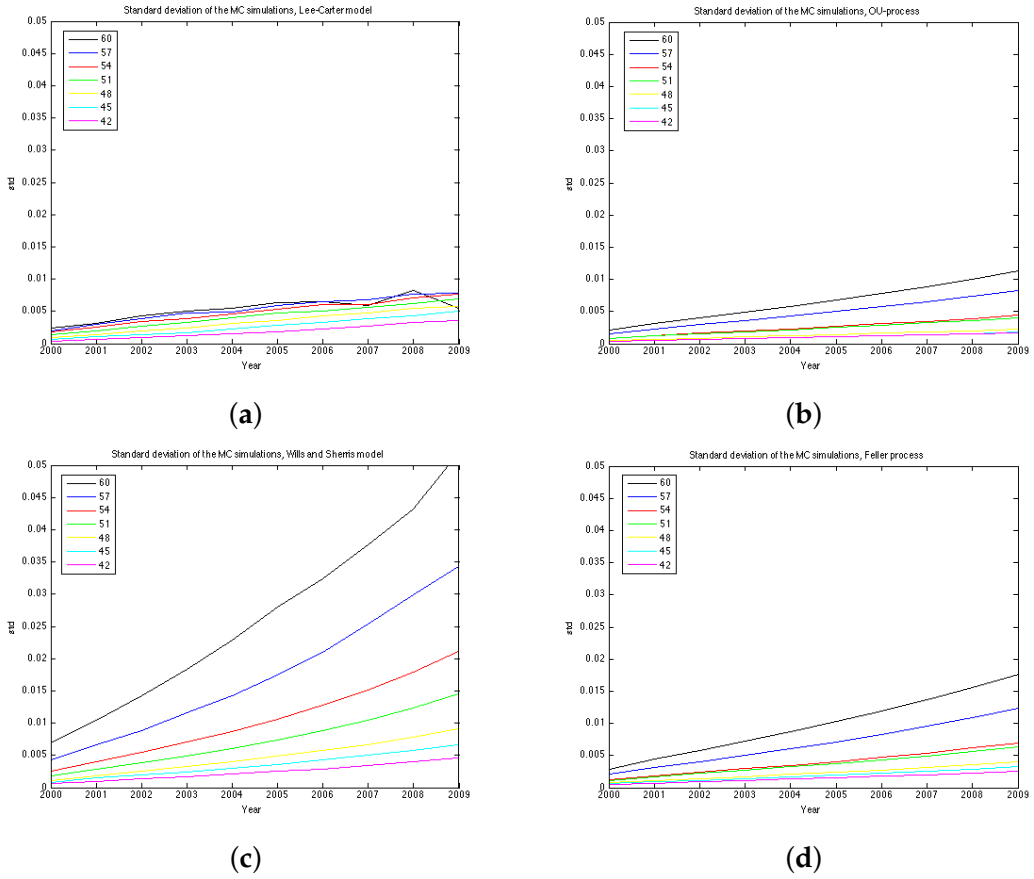


FIGURE 2.8: Standard deviation of each model for every generation (**a**–**d**). (**a**) Lee-Carte model; (**b**) OU-process; (**c**) Wills-Sherris model; (**d**) Feller process.

Regarding the covariance/correlation across generations and ages, all models employ a different structure. Lee-Carter is a one-factor model, which results in mortality improvements at all ages being perfectly correlated. Wills and Sherris model is designed to capture correlation between the ages. In practice it amplifies the effect of the variance growth over time since in reality the correlation increases with age (see Wills and Sherris [95]). In fact, in the simulation procedure the Wiener process is multiplied by the instantaneous correlation matrix D which describes the correlation structure between ages. The OU and the Feller processes in the current study do not take into account the correlation between generations. However, they can also be extended to the case of multiple generations. In [61] it is described how the OU process can be extended to the case of two generations, whose changes in mortality intensities are correlated with an instantaneous correlation coefficient.

2.7.3 Discussion on the Number of Parameters

The number of estimated parameters is different for each procedure. The Wills and Sherris model estimates only 3 parameters for the whole dataset plus eigenvectors to approximate the correlation matrix (in our case we take 30 eigenvectors), while both the OU and the Feller processes fit 2 parameters for each generation. To calibrate the Lee-Carter model, we have to estimate $A + T = 50 + 30 = 80$ parameters. To predict mortality for each generation in the dataset for which the size of the estimation part would be not be less that 10 years, we have to estimate the OU and the Feller processes

for 19 generations resulting in 38 parameters each, for the Wills and Sherris model we have used $3 + 30 = 33$ parameters.

2.7.4 Prediction Intervals

Note, that although the variances of the Lee Carter model is smaller than the ones produced by the OU-process and the Feller process, the first model shows better results.

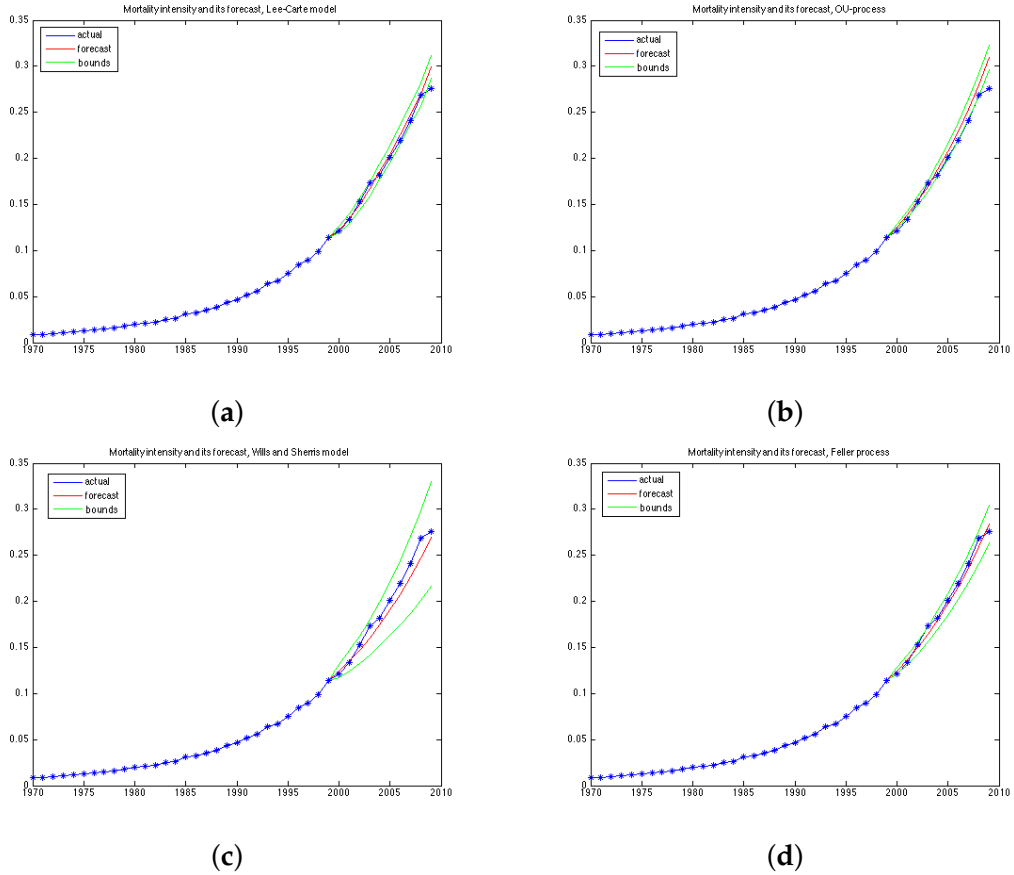


FIGURE 2.9: Actual mortality rate and 90% prediction intervals for generation aged 57 for each model. (**a**) Lee-Carte model; (**b**) OU-process; (**c**) Wills-Sherris model; (**d**) Feller process.

Age in 1970	Wills and Sherris	Lee-Carter	OU-Process	Feller
60	1.0000	0.5000	0.2000	0.7000
59	1.0000	0.9000	0.3000	0.8000
58	1.0000	0.9000	0.3000	0.8000
57	1.0000	0.9000	0.5000	0.8000
56	1.0000	0.8000	0.1000	0.9000
55	1.0000	0.3000	0.6000	0.8000
54	1.0000	0.7000	0.8000	0.7000
53	1.0000	0.2000	0.8000	0.7000
52	1.0000	0.6000	0.5000	0.7000
51	1.0000	0.9000	0.8000	0.7000
50	0.5000	0	0	0
49	0.9000	1.0000	0.7000	0.5000
48	1.0000	0.7000	0.4000	0.5000
47	1.0000	0.9000	0.8000	0.8000
46	1.0000	1.0000	0.4000	0.5000
45	1.0000	1.0000	0.6000	0.9000
44	1.0000	1.0000	0.7000	1.0000
43	0.9000	0.8000	0.7000	0.8000
42	0.9000	0.7000	0.9000	0.9000
Mean (Rank)	0.9579 (1)	0.7263 (2)	0.5316 (4)	0.7105 (3)

TABLE 2.7: Percentage of the actual mortality rates which falls within a 90% prediction interval, UK female data.

2.8 Robustness of Simulation Results

Here we evaluate the performance of the approach described above on the 4 datasets. They are:

1. UK Females
2. UK Males
3. Australian Females
4. Australian Males

The experiments in this section are made using the same time and the age periods—1970–2009 and 50–99. The generations aged 42–60 in the year 1970 are chosen in the same manner as in Sections 2.6 and 2.7.

The results of the estimation are presented in Tables 2.8 and 2.9. More detailed results of the estimation for the 4 datasets are included in the Appendix A as tables and plots of the errors for each of the 19 generations. According to the Tables 2.8 and 2.9, the results for the UK males data with regard to accuracy are the same as for the UK females—the OU and the Feller processes produce the smallest error, while the Lee-Carter and the Wills and Sherris model show the largest error. However, the mean of the absolute error for the Lee-Carter model in this case is 3 times larger in comparison to the error of the UK females estimation. This model also shows very bad result according to precision (with the percentage of the actual mortality within prediction interval being only 4.21%).

Model; Dataset	UK, Females	UK, Males	Australia, Females	Australia, Males
Wills and Sherris	0.0388 (3)	0.0580 (3)	0.1368 (4)	0.0787 (3)
Lee-Carter	0.0492 (4)	0.1252 (4)	0.0380 (1)	0.0712 (1)
OU-process	0.0365 (1)	0.0470 (2)	0.1046 (2)	0.0767 (2)
Feller	0.0379 (2)	0.0391 (1)	0.1326 (3)	0.0840 (4)

TABLE 2.8: Mean of the absolute errors for each dataset over 19 generations (rank of accuracy).

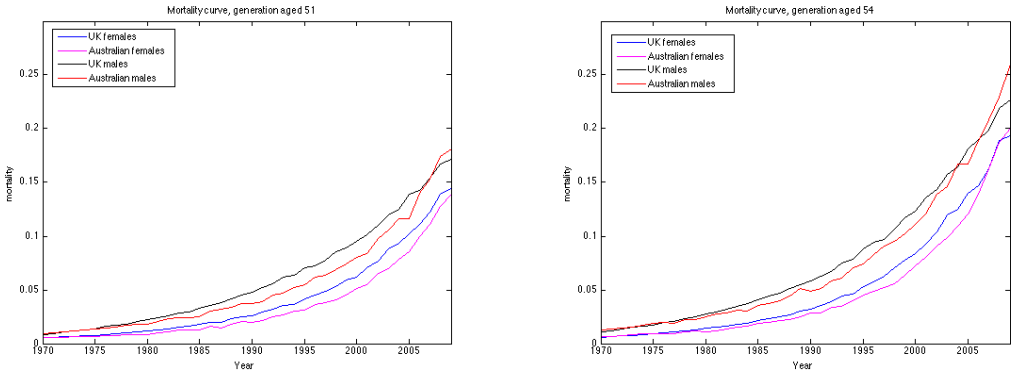
Model; Dataset	UK, Females	UK, Males	Australia, Females	Australia, Males
Wills and Sherris	0.9579 (1)	0.9105 (1)	0.7789 (2)	1.0000 (1)
Lee-Carter	0.7263 (2)	0.0421 (4)	0.8368 (1)	0.3053 (4)
OU-process	0.7105 (3)	0.4000 (3)	0.2947 (4)	0.4421 (3)
Feller	0.5316 (4)	0.7158 (2)	0.3474 (3)	0.5737 (2)

TABLE 2.9: Percentage within a 90% prediction interval for each dataset (rank of precision).

Australian females data is the only dataset which shows good results using the Lee Carter model, both according to precision and accuracy. The OU and the Feller processes, on the contrary, produce large errors for this dataset, especially for the generations aged 45–55. This may be explained by the fact that mortality in Australia is lower for people in their 40s and 50s in comparison to their UK counterparts, and, as a consequence, mortality intensities are larger for older ages. This can be seen from the plots of mortality curves for generations aged 51 and 54 in the year 1970 (Figure 2.10). More prominent convex form of the mortality curves for Australian population makes the error (which is calculated for the last 10 years of the observations) larger as the prediction of mortality underestimates the actual mortality intensity. We would suggest that the inclusion of the correlation coefficient for the OU and the Feller processes to describe the dependence between the generations could improve the calibration results for these procedures by taking into account the fact that if mortality of the generations aged 45–55 is rather low, it would imply an increase in the mortality intensity for the older ages.

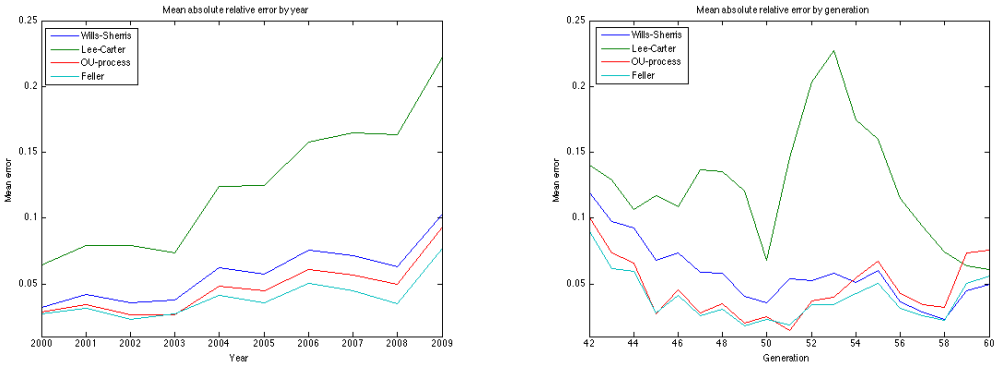
It is worth noting that for all datasets the errors produced by the Wills and Sherris model, the OU process and the Feller process exhibit similar patterns (Figure 2.7b, 2.11b, 2.13b and 2.12b), while the errors produced by the Lee-Carter model have a different pattern. This may be explained by the fact that the first three procedures model the advances (changes) in mortality intensity for a cohort, while the Lee-Carter models the central mortality rate itself.

On the whole, we can say that the results are data dependent. However, from the estimation results on the four datasets, we can conclude that the Wills and Sherris model performs best in terms of precision, but it is one of the worst in terms of accuracy. The Lee Carter model shows better fit to the Australian population dataset rather that to the British one, both for males and females. The OU process and the Feller process provide rather good results in terms of accuracy, while they have often get low ranks in terms of precision.



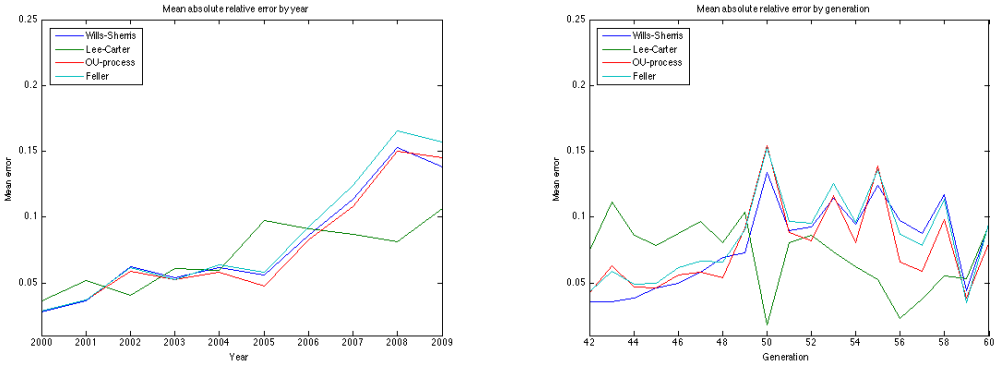
(**a**) (**b**)

FIGURE 2.10: Observed mortality curves for the UK and Australian generations aged 51 (**a**) and 54 (**b**). (**a**) Generation aged 51 in the year 1970; (**b**) Generation aged 54 in the year 1970.



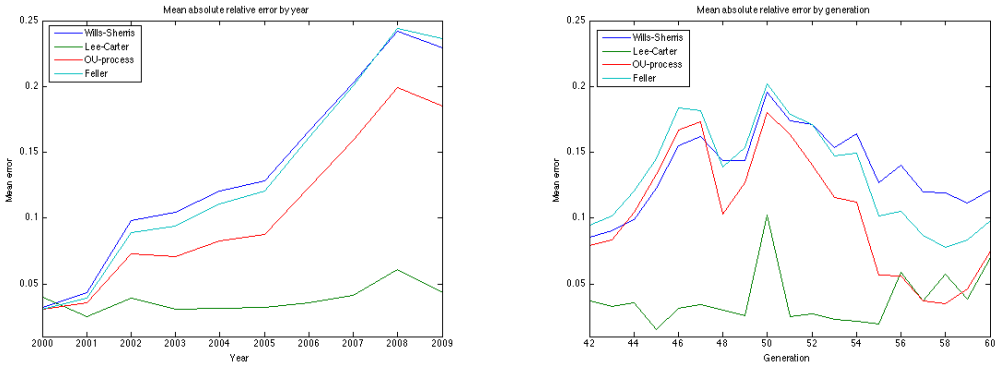
(**a**) (**b**)

FIGURE 2.11: Mean relative absolute error of each model, UK male data. (**a**) Relative absolute error, average by year; (**b**) Relative absolute error, average by generation.



(**a**) (**b**)

FIGURE 2.12: Mean relative absolute error of each model, Australian male data. (**a**) Relative absolute error, average by year; (**b**) Relative absolute error, average by generation.



(**a**) (**b**)

FIGURE 2.13: Mean relative absolute error of each model, Australian female data. (**a**) Relative absolute error, average by year; (**b**) Relative absolute error, average by generation.

2.9 Conclusions

In this study we have calibrated 4 mortality models to the UK and Australian populations and have quantitatively compared their accuracy and precision in forecasting mortality rates. To evaluate this we have used two measures – first, we looked at the relative errors between the forecasted and the observed mortality rates and second, we investigated the percentage of the observed mortality rates which fell within the projected prediction intervals. Our experiments compare one discrete-time model, proposed by Lee and Carter, and three continuous-time models—the Wills and Sherris model, the Ornstein-Uhlenbeck process and the Feller process. The first two models estimate the whole surface of mortality across ages and years simultaneously, while the latter two model each generation separately. One major advantage of the OU and the Feller processes is that they belong to the affine class of mortality models and so allow closed-form expressions for survival probabilities, which is useful for pricing many financial securities. On the other hand, the Wills and Sherris model allows the dependencies between generations to be captured, which may be useful for life offices who have portfolios written on multiple cohorts.

The choice of the model may depend on the goal and the data available. As a result of our experiments with the UK female, the Wills and Sherris model performs best in terms of the prediction interval, followed by the Lee-Carter model. In terms of the mean absolute error, the OU and the Feller processes are better. Thus, for the UK data models which capture the whole mortality surface are more precise, meaning that their forecast prediction intervals are more likely to include the observed mortality rates. Models for a single generation, on the other hand, tend to be more accurate, meaning that their mean absolute errors between the forecast and observed mortality are smaller. For the UK male data the results are rather similar—the main difference here is that the LC model in this case provides much worst result both in terms of precision and accuracy.

However, the results are different for the Australian dataset. In this case, the Lee-Carter model and the OU process are the best in terms of accuracy, both for males and females. The Wills and Sherris model shows good result with respect to the precision measure for Australia as well, followed by the LC for the females and the Feller process for the males.

Based on our experiments, different models appear to be preferred for specific generations and years. We believe that our analysis and the results discussed in this paper are useful for the insurance industry. In particular, we provide potentially useful insights into different mortality modelling frameworks and allow practitioners to chose a model that suits their specific needs.

Chapter 3

Estimation of the price of risk in the Heston mode

3.1 Introduction

In this paper we study the problem of calibrating and modelling market price of risk in the context of stochastic volatility models. The Heston model (1993) has been widely used for equity option pricing purposes which is done under a risk-neutral measure. In the context of Solvency II internal models are required to produce both risk-neutral and real-world simulations – in particular, for calculation of the Solvency Capital Requirement (SCR). Under the Heston model it is possible to define a price of risk in a way that the state variable follows a square-root process under both an objective probability measure and an equivalent martingale measure. This is a convenient framework in terms of modelling. The problem of calibrating the price of risk is crucial in this setting. We propose a calibration procedure based on the work of Ait-Sahalia and Kimmel (2007). Their methodology involves approximating the unknown likelihood function and identifying the unobserved volatility state variable by inverting option prices under the Heston model. Based on their approach we develop a procedure which allows to find a set of parameters that would be stable in terms of estimation error and coherent according to both criteria – the minimal prediction error and the maximum likelihood approach, as neither of the methods by itself can identify the price of risk parameters. Using joint observations of the Eurostoxx50 index and its options for the last 17 years we study the stability of the obtained market price of risk parameters and the coherence of the solution with respect to the risk-neutral valuation. Using Monte Carlo simulations we compute a yearly VaR which is crucial for SCR.

Stochastic volatility models were developed to overcome the drawbacks of the Black-Scholes formula in option pricing. Black-Scholes formula is based on the assumptions of log-normal stock diffusion with constant volatility. It is well-known that the model-implied volatilities for different strikes and maturities of options are not constant and tend to be smile shaped, see Weron and Wystup (2005). Such models as Hull and White [55], deterministic time-varying model for implied volatilities of Dupire [39], stochastic volatility models of Stein and Stein [91] and of Heston [88] relax the assumption of constant volatility. Heston model (1993) is a two-factor stochastic volatility model with one of the factors being responsible for the dynamics of the volatility coefficient.

Heston's model has become particularly popular in the industry for the two main reasons: (i) the process for the volatility is non-negative and mean-reverting, which is what we observe in the markets, and (ii) there exists a semi closed-form solution for vanilla options. It was one of the first models that was able to explain the smile and

simultaneously allow a front-office implementation and a market consistent valuation of many exotics.

Estimation of stochastic volatility models poses some challenges. First, due to the fact that the volatility dynamics are not entirely observable. To overcome this fact market-based perspectives are adopted. From optimization point of view estimation of stochastic volatility models is also difficult since the likelihood function is not known explicitly. Calibration can be done in two conceptually different ways. One way is to look at a historical time series of the underlying. Estimation methods such as Generalized, Simulated, and Efficient Methods of Moments (respectively GMM, SMM, and EMM), as well as Bayesian MCMC have been extensively applied, for a review see Chernov and Ghysels (2000) [29]. Fitting the empirical distributions of returns to the marginal distributions via a minimization scheme can also be used. However, none of the historical approaches allows for estimation of the market price of volatility risk, while multiple studies find evidence of a non-zero volatility risk premium. Thus, observing only the underlying spot price and estimating stochastic volatility models with this information does not deliver correct derivative security prices. The second estimation approach instead of using the spot data, calibrates the model to derivative prices. These type of methods produce more reliable results, but they still fail to estimate price of the volatility risk in explicit manner.

Ait-Sahalia and Kimmel [3] have suggested a method to estimated the Heston model under the objective measure. This procedure is based on approximating the unknown likelihood function and identifying the unobserved volatility state by inverting option prices. However, Ait-Sahalia and Kimmel [3] report large standard errors for the price of risk parameters based on the estimation performed on a synthetic dataset. Based on their approach we develop a procedure which allows to find a set of parameters that would be stable in terms of estimation error and coherent according to both criteria – the minimal prediction error and the maximum likelihood approach, as neither of the methods by itself can identify the price of risk parameters.

Calibration of the Heston model under the risk-neutral measure is generally done using joint observations of the stock and derivatives such as call options or swaps. Usually five parameters are fitted to the observations: initial variance, volatility of volatility, a long-run variance, a mean reversion, and a correlation coefficient such as in [93]. The parameters are found by an optimisation procedure which minimises the prediction error between mathematical Heston prices and the market prices. Using the characteristic function conditioned on the initial values of the underlying diffusion processes. In our case this approach is modified to take into account a time-dependent process for the unobserved variance. On each iteration, we infer the process given parameter vector considered during the estimation procedure by equating the mathematical Heston call price with the market price at each time instalment.

The main contribution of this work is the empirical. It provides an improved estimation of the price of risk parameters in the Heston model based on the Ait-Sahalia and Kimmel approach [3]. It also provides analysis of the estimation of the unobserved variance process. We find that our procedure allows to reduce the error between the actual and the model price in comparison to the standard estimation approach of the Heston model by identifying iteratively the unobserved variance process and by correcting the parameters on each step. The estimated parameters tend to converge to a stable value when the number of iterations increases. Regarding the simulation part, the estimated parameters provide more prudent estimation of the future VaR in comparison to the parametric and historical approach.

3.2 The Heston model

Heston (1993) [88] assumed that the spot price follows a diffusion process resembling a geometric Brownian motion (GBM) with a non-constant instantaneous variance Y_t while the variance itself is driven by a mean reverting stochastic process:

$$\begin{aligned} dS_t &= S_t(\mu dt + \sqrt{Y_t} d\tilde{W}_t^1), \\ dY_t &= \kappa'(\gamma' - Y_t)dt + \sigma\sqrt{Y_t} d\tilde{W}_t^2. \end{aligned} \tag{3.1}$$

The two Wiener processes are correlated with each other:

$$d\tilde{W}_t^2 d\tilde{W}_t^2 = \rho dt.$$

The model of Black and Scholes (1973) is a special case of the model of Heston (1993) [88], in which $\sigma = 0$ and $v_0 = \gamma'$ so that v_t is a constant.

The variance process $\{Y_t\}_{t\geq 0}$ is a square root mean reverting process, first introduced by Cox, Ingersoll and Ross (1985) [31] for modelling the short term interest rate. It is defined by three parameters: γ', κ' and σ, where γ' is a long-run mean, κ' is a rate of mean reversion to the long term variance and σ is referred to as the volatility of the variance (often called the *vol of vol*).

Several empirical and economical studies suggest that the Heston model fits better to the real-world financial applications. It does not have the flaw of the Black and Scholes model which assumes that an assets's log-return distribution is Gaussian. According to empirical observations, it is characterised by heavy tails (skewness) and high peaks (leptokurtosis). There is also empirical evidence and economic arguments that suggest that equity returns and implied volatility are negatively correlated (also termed as "the leverage effect"), see Cont (2001) [30]. In contrast, the Heston's model can imply a number of different distributions. ρ, which can be interpreted as the correlation between the log-returns and volatility of the asset, affects the heaviness of the tails of the distribution. Intuitively, if $\rho > 0$, then volatility will increase as the asset price/return increases. This will spread the right tail and squeeze the left tail of the distribution creating a fat right-tailed distribution. Conversely, if $\rho < 0$, then volatility will increase when the asset price/return decreases, thus spreading the left tail and squeezing the right tail of the distribution creating a fat left-tailed distribution (emphasising the fact that equity returns and its related volatility are negatively correlated). ρ, therefore, affects the skewness of the distribution. In addition, the effect of changing the skewness of the distribution has an impact on the shape of the implied volatility surface. Hence, it addresses another shortcoming of the Black-Scholes-Merton model which assumes constant volatility across different strike levels.

The volatility of the volatility parameter σ affects the kurtosis (peak) of the distribution. When σ is 0 the volatility is deterministic and, hence, the log-returns will be normally distributed. Increasing σ will then increase the kurtosis only, creating heavy tails on both sides. Again, the effect of changing the kurtosis of the distribution has an impact on the implied volatility. Higher σ makes the volatility process more volatile and the skew/smile of the implied volatility more prominent. This means that the market has a greater chance of extreme movements. In this situation writers of puts must charge more and those of calls, less, for a given strike.

The mean reversion parameter, γ', can be interpreted as the degree of "volatility clustering". This property of the asset returns has been observed in the market data

meaning that large price variations are more likely to be followed by large price variations.

A computationally convenient feature of the model is that it provides a semi-closed solution for European options, making it more tractable and easier to implement than other stochastic volatility models. However, the likelihood function of the model of Heston is not known in a closed form. System of equations (3.1) defines a two-dimensional stochastic process for the variables S_t and Y_t. The model given in (3.1) is not in the class of affine processes, whereas under the log transform for the stock, it is. By setting $s_t = \ln(S_t)$, we can express it in terms of the logarithmic asset price s_t and Y_t:

$$\begin{aligned} ds_t &= (\mu - \frac{1}{2}Y_t)dt + \sqrt{Y_t}d\tilde{W}_t^1, \\ dY_t &= \kappa'(\gamma' - Y_t)dt + \sigma\sqrt{Y_t}d\tilde{W}_t^2. \end{aligned} \tag{3.2}$$

Suppose that s_t and Y_t follow the (risk-neutral) process defined in equation (3.2). Then the drift μ of the stock process must be equal to the risk-free rate r (in case of no dividends).

A model for asset prices must specify not only the stochastic process followed by a set of underlying factors, but also the attitude of investors towards the risk of those factors. Since the pioneering work of Harrison and Kreps (1979) [49] and Harrison and Pliska (1981) [50], this task is often accomplished by specifying the behaviour of the state variable(s) under both an objective probability measure and an equivalent martingale measure. A common practice is to have the state variables follow a Feller square-root process under both probability measures, but with different governing parameters by assigning to each state variable a market price of risk process that is proportional to the square root of that state variable. Since the instantaneous volatility of each state variable is also proportional to its square root, the product of the market price of risk and volatility is proportional to the state variable itself. Subtraction of this product from the drift under the objective probability measure therefore results in a drift under the equivalent martingale measure that is also affine. The definition of the affine processes can be found in Appendix B.

We use the following specification for the market price of risk

$$\Lambda = [\lambda_1\sqrt{(1-\rho^2)Y_t}, \lambda_2\sqrt{Y_t}]'$$

The joint dynamics of s_t and Y_t under the objective measure P are then derived using the Girsanov's theorem for a two-dimensional case:

$$dW_i^P = dW_i^Q - \Lambda_i dt.$$

In our case we get:

$$\begin{aligned} dW_1^P &= dW_1^Q - \lambda_1\sqrt{(1-\rho^2)Y_t}dt, \\ dW_2^P &= dW_2^Q - \lambda_2\sqrt{Y_t}dt, \end{aligned} \tag{3.3}$$

Substituting the expressions for dW_1^P, dW_2^P obtained from 3.3 into 3.2 and assuming independence between the two Brownian motions (by using a transform $W_2(t) = \tilde{W}_2(t)$, $W_1(t) = \alpha_1\tilde{W}_1(t) + \alpha_2\tilde{W}_2(t)$ with $\alpha_1^2 + \alpha_2^2 = 1$, $\alpha_2 = \rho$), we get the

expression for the dynamics of s_t and Y_t under the objective measure P:

$$d\begin{bmatrix} s_t \\ Y_t \end{bmatrix} = \begin{bmatrix} r + bY_t \\ \kappa(\gamma - Y_t) \end{bmatrix} dt + \begin{pmatrix} \sqrt{(1-\rho^2)Y_t} & \rho\sqrt{Y_t} \\ 0 & \sigma\sqrt{Y_t} \end{pmatrix} d\begin{bmatrix} W_1^P(t) \\ W_2^P(t), \end{bmatrix} \tag{3.4}$$

where $b = \lambda_1(1-\rho^2) + \lambda_2\rho - \frac{1}{2}$, $\kappa = \kappa' - \lambda_2\sigma$, $\gamma = \frac{\kappa+\lambda_2\sigma}{\kappa}\gamma'$.

We obtain that the state variable follow a Feller square-root process under both probability measures, but with different governing parameters. Thus, Y_t is a restricted state variable as it is bounded below by zero. The boundary value zero cannot be achieved if the Feller condition $2\kappa\gamma \geq \sigma^2$ is satisfied. The log stock price has a volatility which is an affine function of Y_t, and the covariance between s_t and Y_t is also affine in Y_t. The parametrization for the price of risk vector we consider belongs to the class of *essentially affine* models where the price of risk is restricted to ensure that if the volatility of any linear combination of state variables approaches zero, the risk premium of that linear combination also approaches zero, see Duffie (2002) [38]. Such market price of risk implies a an interplay between the parameters under the objective measure and the equivalent martingale measure in a way that only the speed of mean reversion for the restricted state variables can differ independently between the two measures, while the term in the drift depends these two parameters. Now we can rewrite the system 3.2 incorporating the price of risk:

$$\begin{aligned} ds_t &= (\mu - \frac{1}{2}Y_t)dt + \sqrt{Y_t}d\tilde{W}_t^1, \\ dY_t &= (\kappa\gamma - (\kappa + \lambda_2\sigma)Y_t)dt + \sigma\sqrt{Y_t}d\tilde{W}_t^2. \end{aligned} \tag{3.5}$$

As in [88], consider any twice-differentiable function $f(s, Y, t)$ that is a conditional expectation of some function of s and Y at a later date, T, $g(s(T), Y(T))$:

$$f(s, Y, t) = E\big[g(s(T), Y(T))|s(t) = s, Y(t) = Y\big]. \tag{3.6}$$

Ito's lemma shows that

$$\begin{aligned} df = \Big(&\frac{1}{2}Y\frac{\partial^2 f}{\partial s^2} + \rho\sigma Y\frac{\partial^2 f}{\partial s\partial Y} + \frac{1}{2}\sigma^2 Y\frac{\partial^2 f}{\partial^2 Y} + (r - \frac{1}{2}Y)\frac{\partial f}{\partial s} \\ &+[\kappa\gamma - (\kappa + \lambda_2\sigma)Y]\frac{\partial f}{\partial Y} + \frac{\partial f}{\partial t}\Big)dt \\ &+(r - \frac{1}{2}Y)\frac{\partial f}{\partial s}d\tilde{W}_1 + [\kappa\gamma - (\kappa + \lambda_2\sigma)]\frac{\partial f}{\partial Y}d\tilde{W}_2. \end{aligned} \tag{3.7}$$

By iterative expectations, f must be a martingale:

$$E\big[df\big] = 0.$$

Applying this fact to equation (3.7) yields the Fokker-Planck (or forward Kolmogorov) equation:

$$\begin{aligned} &\frac{1}{2}Y\frac{\partial^2 f}{\partial s^2} + \rho\sigma Y\frac{\partial^2 f}{\partial s\partial Y} + \frac{1}{2}\sigma^2 Y\frac{\partial^2 f}{\partial^2 Y} + (r - \frac{1}{2}Y)\frac{\partial f}{\partial s} \\ &+[\kappa\gamma - (\kappa + \lambda_2\sigma)Y]\frac{\partial f}{\partial Y} + \frac{\partial f}{\partial t} = 0. \end{aligned} \tag{3.8}$$

Equation (3.6) imposes the terminal condition:

$$f(s, Y, T) = g(s, Y).$$

Thus, if $g(s, Y)$ is the Dirac delta function $g(s, Y) = \delta(s - s^*)$ (which can be presented in the integration form as $g(s, Y) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} e^{i\omega(s-s^*)} d\omega$), then the solution is the conditional probability density at time t that $s(T) = s^*$. If $g(s, Y) = \mathbf{1}_{s \geq \ln K}$, then the solution is the conditional probability at time t that $s(T)$ is greater than $\ln K$ which is interpreted as a probability of finishing in the money referring to the call option . And if $g(s, Y) = e^{i\phi s}$, then the solution is the characteristic function.

3.2.1 Option pricing

Under the risk-neutral valuation under a martingale measure Q, the price of a contingent claim is evaluated as its expected discounted payoff. For the option value, we have:

$$Call = E_t^Q[e^{r(T-t)} Payoff(T)] = E_t^Q[e^{r(T-t)}(e^{s_T} - K)^+ | s_t, Y_t], \tag{3.9}$$

where K is the strike of the option.

By analogy with the Black-Scholes formula, Heston (1993) makes a guess about the solution in the form:

$$C(S_t, Y_t, t) = S_t P_1 - K e^{-r(T-t)} P_2. \tag{3.10}$$

Here the first term is the present value of the spot asset upon optimal exercise, and the second term is the present value of the strike-price payment. Let us denote by $U(S, Y, t)$ the value of any contingent claim. Introducing the price of volatility risk, the value of the contingent claim must satisfy equation (3.8).

$$\begin{aligned} \frac{1}{2} Y \frac{\partial^2 U}{\partial s^2} + \rho\sigma Y \frac{\partial^2 U}{\partial s \partial Y} + \frac{1}{2}\sigma^2 Y \frac{\partial^2 U}{\partial^2 Y} + (r - \frac{1}{2} Y) \frac{\partial U}{\partial s} \\ + \{\kappa\gamma - (\kappa + \lambda_2 \sigma) Y\} \frac{\partial U}{\partial Y} + \frac{\partial U}{\partial t} = 0. \end{aligned} \tag{3.11}$$

In logarithmic terms, the terminal condition for a call option with a strike price K and maturity T is $C^*(s_t, Y_t, T) = max(0, e^s - K)$. This means that the PDEs for P_j are subject to the terminal condition:

$$P_j(s, Y, T) = \mathbf{1}_{s \geq \ln K}$$

Thus, they may be interpreted as "risk-neutral" probabilities. P_j corresponds to the conditional probability that the option expires in-the-money:

$$P_j(s, Y, t) = Pr[s(T) \leq \ln K | s(t) = s, Y(t) = Y].$$

The probabilities are not immediately available in closed from. However, their characteristic functions satisfy the same PDEs (3.8), subject to the terminal condition:

$$f_j(s, Y, t; \phi) = e^{i\phi s}.$$

Both of the terms P_1 and P_2 in solution (3.10) must satisfy equation (3.8):

$$\begin{aligned}\frac{1}{2}Y\frac{\partial^2 P_j}{\partial s^2}+\rho\sigma Y\frac{\partial^2 P_j}{\partial s\partial Y}+\frac{1}{2}\sigma^2 Y\frac{\partial^2 P_j}{\partial^2 Y}+(r+u_j Y)\frac{\partial P_j}{\partial s}\\ +(a_j-b_j Y)\frac{\partial P_j}{\partial Y}+\frac{\partial P_j}{\partial t}=0,\end{aligned} \tag{3.12}$$

for $j=1,2$, where

$$u_1=\frac{1}{2},\quad u_2=-\frac{1}{2},\quad a=\kappa\gamma,\quad b_1=\kappa+\lambda_2\sigma-\rho\sigma,\quad b_2=\kappa+\lambda_2\sigma.$$

Heston (1993) makes the following assumption on the form of the characteristic function, which exploits the linearity of the coefficients in the PDE (3.7) (characteristic function is conditioned on the initial values s_0, Y_0 of the underlying diffusion process):

$$f_j(s_t,Y_t,t)=\exp\{C_j(T-t)+D_j(T-t)Y_t+i\phi s_t\}, \tag{3.13}$$

Substituting this functional form (3.13) into PDE (3.7), reduces it to two ordinary differential equations:

$$\begin{aligned}-\frac{1}{2}\sigma^2\phi^2+\rho\sigma\phi iD+\frac{1}{2}D^2-u_j\phi i-b_jD+\frac{\partial D}{\partial t}=0,\\ r\phi i+\kappa\gamma D+\frac{\partial C}{\partial t}=0,\end{aligned} \tag{3.14}$$

subject to

$$C(0)=0,\quad D(0)=0.$$

Solving the system (3.14) gives the solution:

$$\begin{aligned}C_j(\tau;\phi)&=r\phi i\tau+\frac{a}{\sigma^2}\left\{(b_j-\rho\sigma\phi i+d)\tau-2\ln\left[\frac{1-ge^{d\tau}}{1-g}\right]\right\}\\ D_j(\tau;\phi)&=\frac{(b_j-\rho\sigma\phi i+d)}{\sigma^2}\left[\frac{1-e^{d\tau}}{1-ge^{d\tau}}\right],\end{aligned} \tag{3.15}$$

with

$$\begin{aligned}g&=\frac{(b_j-\rho\sigma\phi i+d)}{(b_j-\rho\sigma\phi i-d)}\\ d&=\sqrt{(\rho\sigma\phi i-b_j)^2-\sigma^2(2u_j\phi i-\phi^2)}\end{aligned} \tag{3.16}$$

The characteristic function is essentially the Fourier transform of the density generated by the underlying process (can be of diffusion or of a Lévy type) conditioned on the initial state of the process. The physical variable s has been integrated out, leaving the frequency variable ϕ in the equation.

One can invert the characteristic function to get the desired probabilities:

$$P_j(s,Y,T;\ln K)=\frac{1}{2}+\frac{1}{\pi}\int_0^\infty Re\left(\frac{e^{-\ln(K)i\phi}f_j(\phi;s_t,Y_t)}{i\phi}\right)d\phi \tag{3.17}$$

However, due to the form of the characteristic function, we can not get its inverse analytically and a numerical method for integration has to be used, see, for instance,

[27], [41], [73], [75]. We employ the approach of Carr and Madan which is described in Appendix B.

3.3 Calibration

In the financial industry calibration of the parameters for pricing products with future uncertainty is based on the market data. This "market implied approach" assumes that the current market prices are the truthful information source and that they represent the relevant information available about a stochastic model for the underlying asset. The market is assumed to be a perfect informational machine, which absorbs all the relevant information about the unknown "stock" process, and produces consistent option prices. Efficient Market Hypothesis (EMH) is at the core of the informationally efficient financial market assumption. The EMH requires that all economic agents are fully informed and perfectly rational. The objective function used to find the model parameters is defined as a minimal prediction error between mathematical model and market prices. On the first step the mathematical model of the underlying stock process is implemented and option prices are calculated based on "guessed parameters". If these computed model option prices are not coherent with the model, then the model parameters are improved by calibrating until some pre-defined criteria on the quality of fit is met.

In this setting, calibration of the Heston model under the risk-neutral measure is generally done using joint observations of the stock and derivatives such as call options or swaps. Usually five parameters are fitted to the observations: initial variance Y_0, volatility of volatility σ, a long-run variance γ', a mean reversion κ', and a correlation coefficient ρ, see [93] for a detailed calibration procedure. The parameters are found by an optimisation procedure which minimises the prediction error between mathematical Heston prices and the market prices. Using the formula (B.7) of the call price under the Heston model and using equation (3.13) for the characteristic function, we can obtain the Heston call price. In this way, the characteristic function (3.13) is conditioned on the initial values s_0, Y_0 of the underlying diffusion process. In our case this approach is modified to take into account a time-dependent Y_t process for the unobserved variance. On each iteration, we infer the Y_t process given parameter vector considered during the estimation procedure by equating the mathematical Heston call price with the market price at each time instalment t.

The objective function is defined as the Sum of Squared Errors (SSE) between the Heston model prices, which are obtained from equations (3.10), (3.13) and (3.17), and the market prices:

$$SSE(\kappa', \gamma', \sigma, \rho | Y_t) = \sum_{i=1}^{n} \{C_i^M - C_i^H(\kappa', \gamma', \sigma, \rho | Y_t)\}^2 \tag{3.18}$$

subject to

$$2\kappa'\gamma' \geq \sigma^2, \quad -1 < \rho < 1, \quad \kappa' > 0, \quad 0 < \gamma' < 1, \quad 0 < \sigma < 1.$$

where C_i^M, $C_i^H(\kappa', \gamma', \sigma, \rho | Y_t)$ are the ith option prices from the model and market, respectively. N is the number of options used for calibration. The market prices C_i^M are the option prices we obtained with the Black-Scholes formula and implied volatility values taken from the market. They are At-the-money call options with a 3 month tenor.

We calibrate the model defined by the system of equations (3.4). Equation (3.18) defines a non-linear least-squares optimisation problem which allows to identify a set of "average risk-neutral" parameters $\kappa', \gamma', \sigma, \rho$ and the unobserved Y_t process. We incapsulate the constrain of the feasible set of parameters $2\kappa'\gamma' \geq \sigma^2$ by taking the first parameter equal to $2\kappa'\gamma' - \sigma^2$ with a low bound of zero. Thus, actual vector of parameters used in the optimization is $[2\kappa'\gamma' - \sigma^2, \gamma, \sigma, \rho]$.

At the last stage of the calibration procedure we identify the price of risk parameters λ_1, λ_2 and we obtain a more precise estimation of the ρ using the likelihood expansion method proposed by Ait-Sahalia and Kimmel [3] and the extracted Y_t process coupled with the stock process S_t. The likelihood expansion technique is summarised in the next section.

Once the risk-neutral parameters are obtained by minimising SSE defined in (3.18), we optimize the likelihood function w.r.t. 3 parameters: $(\rho, \lambda_1, \lambda_2)$. The risk-neutral parameters κ'^*, γ'^* and the value of σ^* we fix with the values obtained from the previous step and we explicitly define the relationship between the parameters under the risk-neutral and the objective measures as in the system of equations (3.4): $\kappa = \kappa'^* - \lambda_2\sigma^*$ and $\gamma = \frac{\kappa+\lambda_2\sigma^*}{\kappa}\gamma'^*$.

To summarise, the calibration procedure involves the following steps:

1. Phase 1: parameters initialisation.
   - 1.1. Guess the four initial parameters $\kappa'^0, \gamma'^0, \sigma^0, \rho^0$ based on the estimates from the data
   - 1.2. For each t find Y_t^0 s.t. $C_i^M = C_i^H(\kappa'^0, \gamma'^0, \sigma^0, \rho^0, Y_t)$
2. Phase 2: parameters estimation.
   - 2.1. Given $[\kappa'^0, \gamma'^0, \sigma^0, \rho^0]$ and Y_t^0, find the four parameters as:
   $$[\kappa'^*, \gamma'^*, \sigma^*, \rho^*] = \arg\min_{\kappa', \gamma', \sigma, \rho} SSE(\kappa', \gamma', \sigma, \rho | Y_t^0).$$
   - 2.2. Find $\lambda_1^{ML}, \lambda_2^{ML}, \rho^{ML} = \arg\min_{\lambda_1, \lambda_2, \rho} -L_n(\lambda_1, \lambda_2, \rho | \kappa'^*, \gamma'^*, \sigma^*)$, where L_n is the log-likelihood function as defined by Ait-Sahalia and Kimmel [3]
   - 2.3. Obtain $\kappa^* = \kappa'^* - \lambda_2^{ML}\sigma^*$ and $\gamma^* = \frac{\kappa^*+\lambda_2^{ML}\sigma^*}{\kappa^*}\gamma'^*$
   - 2.4. Repeat the procedure from step 1.2 using the real-world parametrization:
   $$\kappa^*, \gamma^*, \sigma^*, \rho^{ML}, \lambda_2^{ML}.$$

The goal of this procedure is to find a set of parameters which would be stable in terms of estimation error and coherent according to both criteria – the minimal prediction error and the maximum likelihood approach, as neither of the methods by itself can identify the price of risk parameters in a stable manner. Such, Ait-Sahalia and Kimmel [3] report large standard errors (reported value is 3.9) for the price of risk parameters based on the estimation performed on a synthetic dataset, while the minimisation procedure itself suffers from a so-called "curse of dimensionality" prone to optimisation methods.

The standard errors of the MLE are the square roots of the diagonal elements of the inverse of the observed Fisher information matrix (the Hessian):

$$F(\theta) = \frac{\partial^2 l_X}{\partial\theta\partial\theta^T}, \quad SE(\hat{\theta}_M L) = \frac{1}{\sqrt{F(\hat{\theta}_M L)}}.$$

The optimization problem (3.18) is ill-posed since the parameter space, over which the minimization is performed, is neither convex no does it have any particular structure. This poses some complications. Finding the minimum is not as simple as finding those parameter values that turn the gradient into zero. Hence, finding a global minimum is difficult (and very dependent on the optimisation method used). Unique solutions do not need to necessarily exist, in which case only local minima can be found. To overcome this problem, we need a set of initial parameters which is close enough to the "real" ones. To get a good estimate of the initial parameters we analyse the historical data. This analysis is described in the Phase 1 of the parameters estimation procedure.

3.3.1 Likelihood expansion

Ait-Sahalia and Kimmel [3] have suggested a method to estimated the Heston model under the objective measure. This procedure is based on approximating the unknown likelihood function and identifying the unobserved volatility state by inverting option prices.

Consider the SDE describing the dynamics of the state vector X_t under the measure P, as specified by (B.1) in C. Let $p_X(\Delta, x|x_0;\theta)$ denote the transition function, that is, the conditional density of $X_{t+\Delta} = x$ given $X_t = x_0$ as a function of x, where θ denotes the vector of parameters for the model. The log-likelihood function $l_X = \ln p_X$ is approximated using closed-from expansions. It is obtained using Hermite polynomials and it takes the form of the power-series (with Δ being the time interval separating observations):

$$l_X^{(J)}(\Delta, x|x_0;\theta) = -\frac{m}{2}\ln(2\pi\Delta) - D_\nu(x;\theta) + \frac{C_X^{-1}(x|x_0;\theta)}{\Delta} + \sum_{k=0}^{J} C_X^{(k)}(x|x_0;\theta)\frac{\Delta^k}{k!}, \tag{3.19}$$

where

$$D_\nu(x;\theta) = \frac{1}{2}\ln(\det[\nu(x;\theta)]) \quad \text{and} \quad \nu(x) = \sigma(x)\sigma'(x).$$

The coefficients $C_X^{(k)}$ corresponding to Δ^k, $k = 1, \ldots, J$ are the unknowns here. First, Taylor series in $(x - x_0)$, denoted by $C_X^{j_k,k}$, of each coefficient Δ^k at order j_k are calculated. The resulting expansion is then

$$\tilde{l}_X^{(J)}(\Delta, x|x_0;\theta) = -\frac{m}{2}\ln(2\pi\Delta) - D_\nu(x;\theta) + \frac{C_X^{(j_{-1},-1)}(x|x_0;\theta)}{\Delta} + \sum_{k=0}^{J} C_X^{(j_k,k)}(x|x_0;\theta)\frac{\Delta^k}{k!},$$

The coefficients $C_X^{(j_k,k)}$ are then obtained by forcing the expansion 3.19 to satisfy, to order Δ_J, the forward and backward Fokker-Planck-Kolmogorov equations. Then the corresponding system of linear equations is solved and the coefficients $C_X^{(j_k,k)}$ are obtained. Such, the forward equation for the $\ln p_X$ has a form:

$$\begin{aligned}\frac{\partial l_X}{\partial \Delta} = &-\sum_{i=1}^{m}\frac{\partial \mu_i^P(x)}{\partial x} + \frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{m}\frac{\partial^2 v_{ij}(x)}{\partial x_i \partial x_j} - \sum_{i=1}^{m}\mu_i^P(x)\frac{\partial l_X}{\partial x_i} + \sum_{i=1}^{m}\sum_{j=1}^{m}\frac{\partial v_{ij}(x)}{\partial x_i}\frac{\partial l_X}{\partial x_j} \\ &+ \frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{m} v_{ij}(x)\frac{\partial^2 l_X}{\partial x_i \partial x_j} + \frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{m}\frac{\partial l_X}{\partial x_i}v_{ij}(x)\frac{\partial l_X}{\partial x_j}\end{aligned} \tag{3.20}$$

From there, the joint likelihood function of the observations $G_t = [S_t, C_t]'$ is obtained by multiplying the likelihood of $X_t = [S_t, Y_t]'$ by Jacobian term. The likelihood function of the observed stock and option prices $G_t = [S_t, C_t]'$ is obtained from equation (3.19) using the Jacobian term which in our case (when for each time instalment there taken only one stock and one option price) is simply the first derivative of the option price with respect to the additional state variable Y_t (it is similar to the Vega of the call option):

$$J_t = \frac{\partial C_t}{\partial Y_t}$$

Let $p_G(\Delta, g|g_0; \theta)$ similarly denote the transition function of the vector of asset prices G observed Δ units apart. Due to the Markovian property, the log-likelihood function of the asset prices g_t sampled at dates $t_0, \ldots, t_n$ has a form:

$$L_n(\theta) = n^{-1} \sum_{i=1}^{n} l_G(t_i - t_{i-1}, g_{t,i}|g_{t,i-1}; \gamma) \tag{3.21}$$

where

$$l_G(\Delta, g|g_0; \theta) = \ln p_G(\Delta, g|g_0; \theta) = -\ln J_t(\Delta, g|g_0; \theta) + l_X(\Delta, f^{-1}(g, \theta)|f^{-1}(g_0; \theta); \theta).$$

with l_X obtained from equation (3.19) and $f^{-1}(G_{t+\Delta}; \theta) = X_{t+\Delta}$ being the inverse function to express the state as a function of the observed asset prices. Thus, we aim to maximise the likelihood expressed in the equation (3.21) on the bases of the daily/weekly data.

3.4 Experiments

3.4.1 Data description

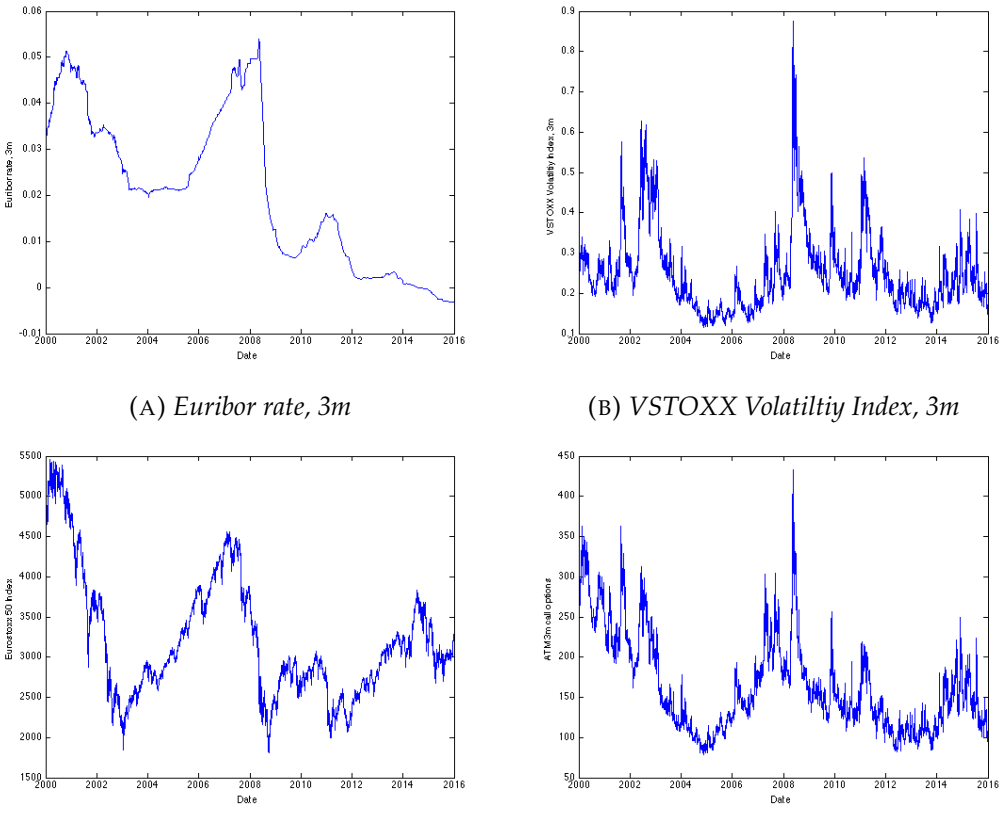
We take time series observations of the Eurostoxx 50 Index for the years 2000-2016. To obtain the time series for the prices of the corresponding call options, we calculate the Black and Scholes call prices from their implied volatilities captured by a volatility index. As an input to the Black and Scholes formula, we use 3 months Euribor rate and the Euro Stoxx 50 Volatility Index (VSTOXX), also for the maturity of 3 months. The VSTOXX Indices are based on the Euro Stoxx 50 realtime options prices and are designed to reflect the market expectations of near-term up to long-term volatility by measuring the square root of the implied variance across all options of a given time to expiration.

Since different indices are sometimes quoted on different dates, before calculating option prices, we match the dates of the Euribor interest rates, VSTOXX Indices and the Eurostoxx 50 Indices. The deduced call prices we further use as a market data in our experiments. For the 17 yeas (2000-1016) we get 4236 daily observations. It results in 847 weekly observations.

The series of the observed Eurostoxx 50 Index, Euribor rate, Euro Stoxx 50 Volatility Index (VSTOXX) and the obtained ATM call prices with 3 months maturity are shown in Figure 3.1.

3.4.2 Parameters estimation: phase 1.

To find the initial parameters for the optimisation routine, we first pre-estimate them on the historical time series of the logarithmic returns. The log return is defined



(A) *Euribor rate, 3m*

(B) *VSTOXX Volatiltiy Index, 3m*

(C) *Eurostoxx 50 Index*

(D) *ATM 3m call options*

FIGURE 3.1: *Time series of the (a) Euribor rate, (b) Volatility Index (VSTOXX), (c) Eurostoxx 50 Index and (d) deduced call option prices; 2000-2016 daily data*

as $x = \ln\left[\frac{S(t+1)}{S(t)}\right]$. Unconditional long-term variance target (reversion level of the variance), γ, is estimated (per year) as the standard deviation of the log returns multiplied by the square root of the length of the time period ($\sqrt{50}$ for the weekly data and $\sqrt{250}$ for the daily data), squared. Correlation coefficient between the log returns and the changes in volatility is estimated based on the realized volatility of the Eurostoxx50 Index. We use a window of different length (5,10 and 20) to calculate the realized volatility series. Then we find the correlation coefficient between the time series of the differences in realized volatility and the log returns. In addition we have calculated the correlation coefficient between the differences in the Volatility Index (VSTOXX) and the log returns of the Eurostoxx50. Estimated statistics are summarized in Table 3.1.

It is well known that the mean reversion level of the variance is much lower when it is estimated based on the volatility of the Eurostoxx50 log returns rather than when it is estimated based on the VSTOXX Index (based on the VSTOXX Index it is 7.23% for the weekly data and 7.21% for the daily data while the $\hat{\gamma}$ estimated based on the realised volatility is 4.79% and 5.87% for the weekly and daily data respectfully). This fact is confirmed by the graphs in Figure 3.2 – the level of the VSTOXX Index exceeds the level of the realized volatility. However, their dynamics are similar.

The correlation coefficient is much more prominent between the log returns and the VSTOXX Index rather than between the log returns and the realized volatility. Moreover, it is larger in absolute value for the weekly data rather than for the daily data.

In practice, different approaches are used for estimation of the unobserved variance process Y_t. For instance, Moody's Analytics report on equity volatility (Hibbert and Manning (2014), [52]) estimates the unobserved variance process as $Y = (C * VSTOXX)^2$, where C is a constant coefficient obtained by running a linear regression between the realized volatility and the market volatility Index. This is a very trivial approach which basically assumes that the unobserved volatility process is a multiple of the market implied volatility. Ait-Sahalia and Kimmel [3] estimate all the coefficients of the stock and volatility process simultaneously by approximating the unknown likelihood function of the joint diffusion process. This is an appealing technique and we use it in our estimations, however, as we find out in our experiments, the parameters obtained using likelihood function approximation technique, strongly depend on the initial parameters. Thus, the question of finding good initial parameters is crucial.

Estimation of the CIR parameters. To get an idea on the parameters estimates for the γ, κ and σ of the unobserved variance process Y, we use the time series of the realized volatility and the VSTOXX Index. We calibrate the parameters γ, κ and σ of the CIR process to the realized variance process and the $VSTOXX^2$. The maximum likelihood estimation of the parameter vector $[\gamma, \kappa, \sigma]$ of the CIR process is based on the transition density. Given Y_t at time t, the density of $Y_{t+\Delta t}$ at time $t + \Delta t$ is

$$p(Y_{t+\Delta t}|Y_t; [\gamma, \kappa, \sigma]) = ce^{-u-v}\left(\frac{v}{u}\right)^{\frac{q}{2}} I_q(2\sqrt{uv}), \tag{3.22}$$

Parameter	Weekly data	Daily data
$\hat{\gamma}$	0.0479	0.0587
$mean(VSTOXX^2)$	0.0723	0.0721
$\hat{\rho}_{x,VSTOXX}$	-0.7622	-0.7749
$\hat{\rho}_{x,Vol5}$	-0.1250	-0.0789
$\hat{\rho}_{x,Vol10}$	-0.1493	-0.0960
$\hat{\rho}_{x,Vol20}$	-0.1607	-0.0717
$mean(Euribor)$	0.0207	0.0207
$mean(logReturns)$	$-4.7017 * 10^{-4}$	$-9.2427 * 10^{-5}$

TABLE 3.1: Parameters' statistics

where

$$\begin{aligned} c &= \frac{2\kappa}{\sigma^2(1-e^{-\kappa\Delta t})} \\ u &= cY_t e^{-\kappa\Delta t}, \\ v &= cY_{t+\Delta t}, \\ q &= \frac{2\kappa\gamma}{\sigma^2} - 1, \end{aligned} \tag{3.23}$$

and $I_q(2\sqrt{uv})$ is a modified Bessel function of the first kind and of order q. Then the likelihood function of the time series $(Y_1 \dots Y_N)$ with the time between two observations of one unit is

$$L([\gamma,\kappa,\sigma]) = \prod_{i=1}^{N-1} p(Y_{i+1}|Y_i;[\gamma,\kappa,\sigma]), \tag{3.24}$$

from which the log-likelihood function of the CIR process is derived:

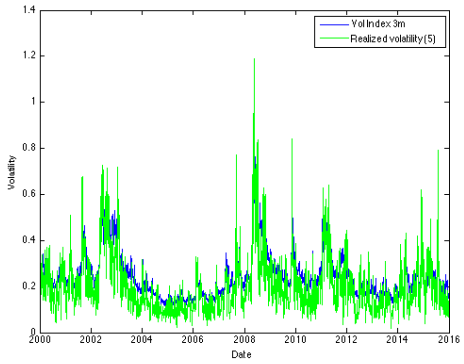
$$\ln L([\gamma,\kappa,\sigma]) = (N-1)\ln c + \sum_{i=1}^{N-1}\left[-u_i - v_{i+1} + 0.5q\ln\left(\frac{v_{i+1}}{u_i}\right) + \ln(I_q(2\sqrt{u_i v_{i+1}}))\right], \tag{3.25}$$

where $u_i = cY_i e^{-a\Delta t}$ and $v_{i+1} = cY_{i+1}$.

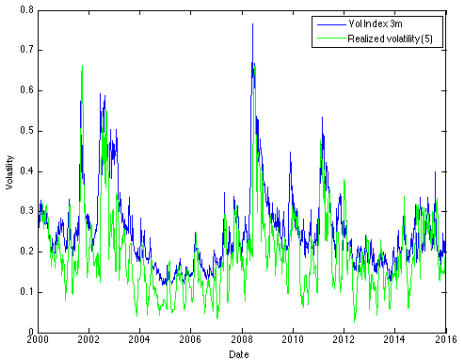
Table 3.2 and Table 3.3 summarise the calibrated parameters for the weekly and daily data. In addition, they contain the information on the moments of the Y_t. The moments are computed as follows (see [60]):

$$\begin{aligned} E[Y_t] &= \gamma + (Y_0 - \gamma)e^{-\kappa t}, \\ Var[Y_t] &= \frac{\sigma^2}{\kappa}(1-e^{-\kappa t})[Y_0 e^{-\kappa t} + \frac{\gamma}{2}(1-e^{-\kappa t})]. \end{aligned} \tag{3.26}$$

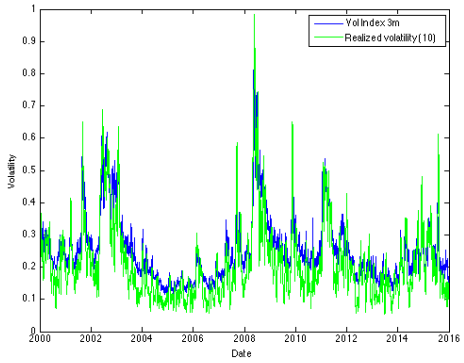
First, let us look at the parameters obtained for the $VSTOXX^2$. The mean reversion level $\hat{\gamma}$ based on the daily and the weekly data is almost the same (0.0715 and 0.0713 respectfully) with the $\hat{\gamma}$ for the daily data higher by 0.23%. As expected, the volatility of the volatility parameter $\hat{\sigma}$ is higher for the daily data in comparison to the weekly data by 14.19% (0.5769 vs. 0.5052). The speed of the mean reversion $\hat{\kappa}$ is higher for the daily data as well – by 33.88% (6.0887 vs. 4.5479). This relation between the parameters of the weekly data and the daily data holds also for the realized volatility



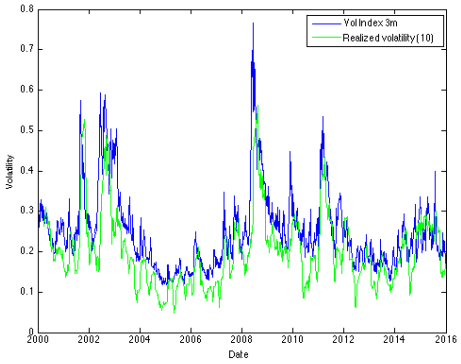
(A) *Daily data - 5 days window*



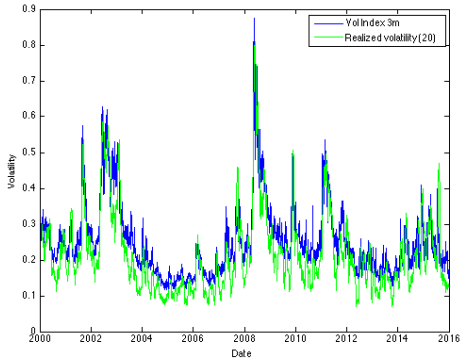
(B) *Weekly data - 5 weeks window*



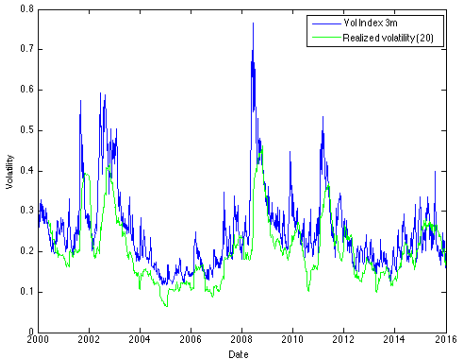
(C) *Daily data - 10 days window*



(D) *Weekly data - 10 weeks window*



(E) *Daily data - 20 days window*



(F) *Weekly data - 20 weeks window*

FIGURE 3.2: *The VSTOXX Index and the realized volatility*

Process	$(\hat{\kappa}, \hat{\gamma}, \hat{\sigma})$	$E(Y_t)$, $Var(Y_t)$	$\sigma^2 \ / \kappa$
$VSTOXX^2$	(4.5479, 0.0713, 0.5052)	0.0713, 0.0040	0.05612
Vol5	(5.3724, 0.0483, 0.6229)	0.0483, 0.0028	0.07222
Vol10	(2.1409, 0.0472, 0.3298)	0.0472, 0.0036	0.05080
Vol20	(0.8280, 0.0442, 0.1802)	0.0442, 0.0048	0.03922

TABLE 3.2: CIR-parameters of the volatility processes, weekly data

Process	$(\hat{\kappa}, \hat{\gamma}, \hat{\sigma})$	$E(Y_t)$, $Var(Y_t)$	σ^2/κ
$VSTOXX^2$	(6.0887, 0.0715, 0.5769)	0.0715, 0.0034	0.05466
Vol5	(26.9250, 0.0601, 1.59789)	0.0601, 0.0018	0.09482
Vol10	(9.2070, 0.0592, 0.8313)	0.0592, 0.0027	0.07506
Vol20	(2.9596, 0.0580, 0.4423)	0.0580, 0.0043	0.06610

TABLE 3.3: CIR-parameters of the volatility processes, daily data

given a fixed length of the window. For both datasets, the mean reversion level $\hat{\gamma}$, the speed of the mean reversion $\hat{\kappa}$ and the volatility of the volatility parameter $\hat{\sigma}$ decrease as the length of the window increases. The latter two observations can be explained by the fact that the dispersion range of the volatility for a smaller window is wider. However, if we look at the variance of the process, $Var(Y_t)$ – it, on the contrary, grows as the length of the window increases. We can conclude that the variance of the implied volatility processes is in the same range as the variance of the $VSTOXX^2$ process, both for the daily and the weekly data, while $\hat{\kappa}$ and $\hat{\sigma}$ vary more for the implied volatility process depending on the length of the window. The ratio of σ^2/κ, however, is closer to the one of the $VSTOXX^2$ process for the 10-days window for the weekly data and for the 20-days window for the daily data.

The coefficient $\hat{\beta}$ of the linear regression between the implied volatility and the VSTOXX Index is significant for each time series of the implied volatility. However, the R^2 coefficient is higher for the daily data. The result based on the daily data also corresponds better with the Moody's Analytics report on equity volatility (Hibbert and Manning (2014)) [52] where the authors obtain $\hat{\beta} = 0.92$.

We think that the presented analysis of the obtained CIR parameters gives a sound idea on the range in which the real-world parameters of the unobserved variance process should lie. While running the optimization routine we use the triples $[\kappa, \gamma, \sigma]$ obtained based on the implied volatility process (Tables 3.1, 3.2 and 3.3) as the initial ones.

3.4.3 Parameters estimation: phase 2.

After the parameters are initialised (with the results of the phase 1), we proceed with the calibration of the model. For the Step 2.1 of the calibration procedure – minimising the SSE w.r.t. four parameters, we use the MATLAB's least-squares, non-linear optimiser, a function *lsqnonlin(fun, x0, lb, ub)*. It minimises the vector-valued function, *fun*, using the vector of initial parameter values, $x0$, where the lower and upper bounds of the parameters are specified in vectors *lb* and *ub*, respectively. lsqnonlin uses an interior-reflective Newton method. The result produced by the *lsqnonlin* is dependent on the choice of $x0$, the initial estimate. This is, therefore, not a global optimiser, but rather, a local one. We have no way of knowing whether the solution is a global/local minimum, but we report the SSE value for each set of optimal parameters. Thus, *lsqnonlin* can only be used when one is confident that the solution to (3.18) is not very far from the initial estimates. We do not use the ASA algorithm on the Step 3 of the calibration procedure for the reason that it requires more time in this case than the MATLAB's least-squares lsqnonlin functions which is able to provide good results as a local optimiser.

Using parameters obtained during the Phase 1 as the initial ones with an upper bound and a lower bound taken close to them, we run the first step of the calibration

procedure and infer the variance process Y_t. The resulting option prices and the volatility process are depicted in Figure 3.3. The initial parameters used are $[\kappa = 4.5479, \theta = 0.0713, \sigma = 0.5052, \rho = -0.7622]$. We find that on the first step the parameters estimated based on the VSTOXX Index provide the best fit. The resulting error on the first step between the model prices and the market prices is $SSE = 6.7879 * 10^5$. As we can see, the fitting is already quite good with visible error occurring only in the events of distress – when the market is more volatile the options are undervalued by the model. This might be due to the variance process inferred as during such periods the inferred volatility is lower than the market has experienced.

For the Step 2.2 of the calibration procedure – minimising the minus likelihood function, we use the Adaptive Simulated Annealing algorithm. Since we do not have an idea on the range of the parameters of the risk vector λ_1 and λ_2, it is important to use an algorithm which is able to find a global minima, rather than a local one. ASA was developed by the theoretical physicist Lester Ingber [56]. The simulation of annealing as an approach that reduces a minimisation of a function of large number of variables to the statistical mechanics of equilibration (annealing) of the mathematically equivalent artificial multiatomic system. ASA is similar to SA except that it uses statistical measures of the algorithm's current performance to modify its control parameters. A proof is provided by Ingber shows that ASA is a global optimiser. He also provides arguments in favour of ASA's computational efficiency and accuracy. The initial point of the algorithm is taken and a random combination between the upper and the lower bound on ache parameter introduced by the user.

Tables 3.4 - 3.7 report the resulting parameters starting from different values of $\lambda_2^{(0)}$ and using different bounds on the parameters during the Step 3 of the optimisation procedure. The standard errors reported for λ_1 and λ_2 of the MLE are the square roots of the diagonal elements of the inverse of the observed Fisher information matrix (the Hessian):

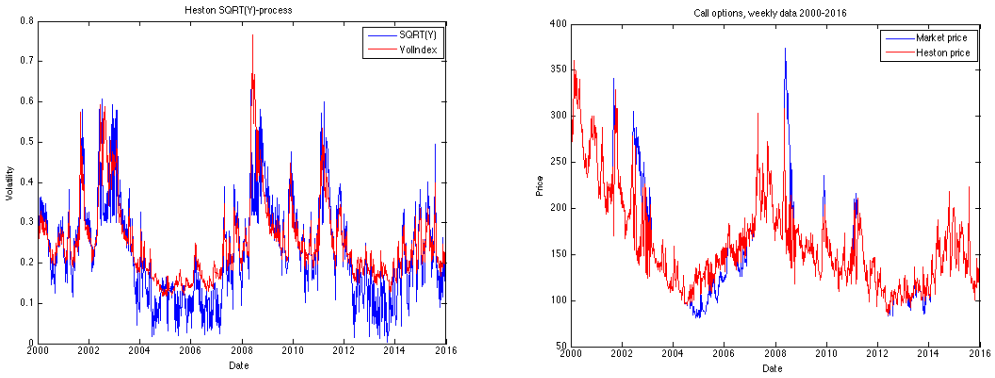
$$F(\theta) = \frac{\partial^2 l_X}{\partial\theta\partial\theta^T}, \quad SE(\hat{\theta}_M L) = \frac{1}{\sqrt{F(\hat{\theta}_M L)}}.$$

The price of risk following from the model which has an economical interpretation as a risk premium per unit of volatility is expressed as $\lambda_1(1-\rho^2) + \lambda_2\rho$. It is reported in the tables as Risk premia.

Results of the parameters estimation by the iterative procedure are depicted in figures 3.5 and 3.6 and in the Tables 3.4-3.7. We can see that the real-world parameter (κ) standing for the speed of mean reversion of the variance is smaller than the risk-neutral (κ') for all initial values of $\lambda_2^{(0)}$ tested. The parameter of the mean reversion level of the variance (γ) is on the contrary larger under the objective measure rather than under the risk-neutral measure (γ').

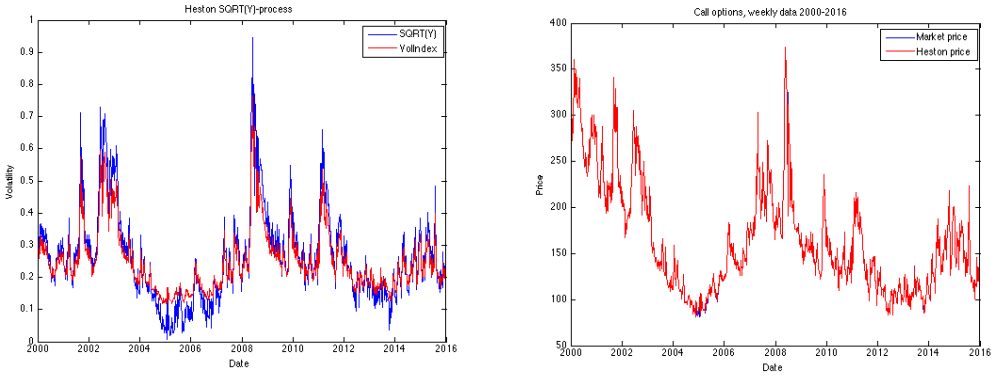
The parameters obtained are coherent with the historical estimate presented in Tables 3.1 and 3.2 – such, the correlation coefficient is close to the one estimated from the VSTOXX Index (Table 3.1 indicates that the correlation coefficient estimated from the weekly data is -0.7622). In the iterative procedure the correlation coefficient is either converging to the lower bound of -0.9 or it is estimated at the level of -0.8039. This result is coherent with the "the leverage effect" – the fact that the equity returns and implied volatility are negatively correlated.

It is interesting to note, that the risk neutral level of mean reversion γ' is close to the estimated one obtained from the log returns. The level of the real-world value of γ is close to the one obtained form the VSTOXX Index, while the level of the risk-neutral value of γ' is close to the estimates obtained from the realized volatility (Table 3.2).



(A) *Inferred* $\sqrt{Y}$*-process* (B) *Heston options prices*

FIGURE 3.3: *Results of the calibration Step 1, weekly data;* $SSE = 6.7879 * 10^5$



(A) *Inferred* $\sqrt{Y}$*-process* (B) *Heston options prices*

FIGURE 3.4: *Results of the full calibration procedure, weekly data;* $SSE = 1.9064 * 10^4$

While the former observation can be expected as the VSTOXX Index should translate the real-world dynamics, the latter might be specific to the data set employed in the experiments and further quantitative experiments would be needed to investigate if the observed dependency holds for other equity indexes. Regarding the Standard Errors of the price of risk parameters and the correlation coefficient, the standard error of the correlation coefficient is always zero and, therefore, not reported in the tables. As for the price of risk, we obtain much lower standard error for λ_1 than reported by Ait-Sahalia and Kimmel [3] – 1.2 vs. 4.3. The SE corresponding to λ_2 is not reported in [3] since no experiments are performed on the real data with λ_2. In absence of λ_2, the reported value of λ_1 for the $S\&P$ 500 Index is 3.9, while in our iterative procedure λ_1 either converges to the upper bound of 10 or it is estimated at the level of 7.8264. Therefore, the associated risk premium per unit of volatility in [3] would be approximately -1.6 (as the correlation coefficient is estimated at -0.767 level). We see our result to be coherent with the one obtained in [3] since the risk premium per unit of volatility in our case fluctuates around -1 and the equity index considered is a different one.

Parameter	Optimal	Bounds (L, U)
$\kappa'(\kappa)$	3.4865 (2.1088)	$[1.6250, 15.7500]$
$\gamma'(\gamma)$	0.0516 (0.0854)	$[0.06, 0.08]$
σ	0.4	$[0.4, 1.3]$
ρ	-0.8039	$[-0.9, -0.5]$
λ_1 (SE)	7.8264 (1.2162)	$[-10, 10]$
λ_2 (SE)	3.4437 (0.8960)	$[-10, 10]$
SSE	$1.9064 * 10^4$	-
Risk premia	$-9.1434 * 10^{-5}$	-

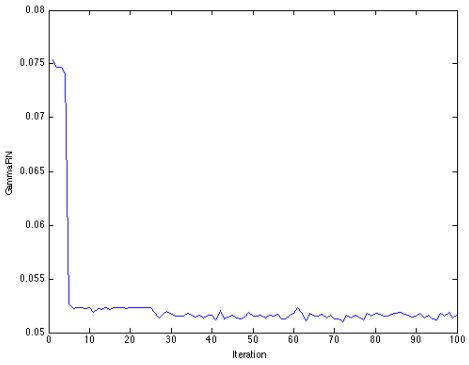
TABLE 3.4: Results of the full calibration procedure,$\lambda_2^{(0)} = -1$ weekly data;

Parameter	Optimal	Bounds (L, U)
$\kappa'(\kappa)$	2.9137 (1.9300)	$[0.0950, 15.7500]$
$\gamma'(\gamma)$	0.0543 (0.0819)	$[0.06, 1.00]$
σ	0.3409	$[0.3, 1.3]$
ρ	-0.9	$[-0.9, -0.5]$
λ_1 (SE)	10 (1.2175)	$[-10, 10]$
λ_2 (SE)	2.8855 (1.0126)	$[-10, 10]$
SSE	461.3809	-
Risk premia	-0.6969	-

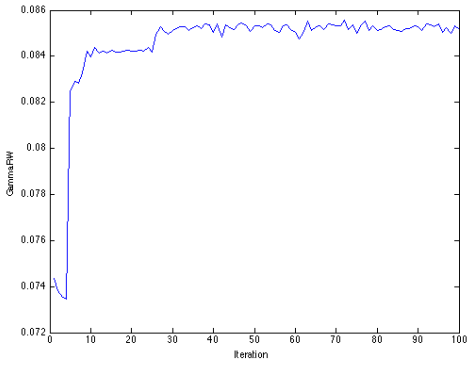
TABLE 3.5: Results of the full calibration procedure, $\lambda_2^{(0)} = -1$ weekly data;

Parameter	Optimal	Bounds (L, U)
$\kappa'(\kappa)$	3.1776 (2.0162)	$[0.0950, 15.7500]$
$\gamma'(\gamma)$	0.0531 (0.0837)	$[0.06, 0.08]$
σ	0.3709	$[0.3, 1.3]$
ρ	-0.9000	$[-0.9, -0.5]$
λ_1 (SE)	10 (1.2172)	$[-10, 10]$
λ_2 (SE)	3.1309 (1.0077)	$[-10, 10]$
SSE	755.2722	-
Risk premia	-0.9178	-

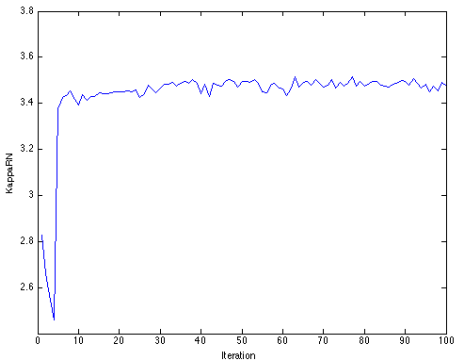
TABLE 3.6: Results of the full calibration procedure, $\lambda_2^{(0)} = 6$, weekly data;



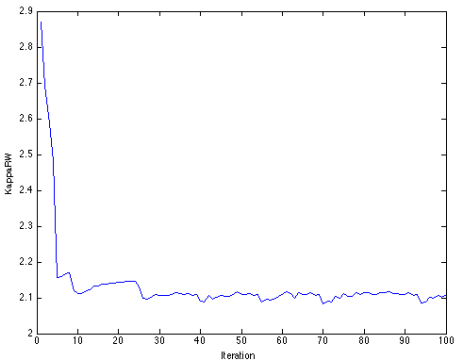
(A) γ'



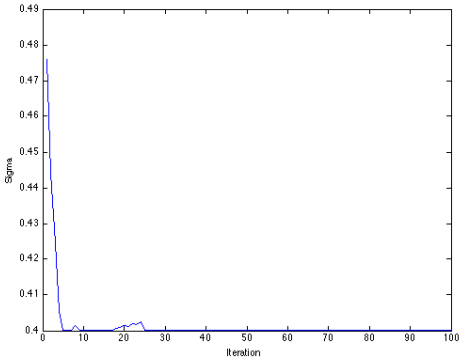
(B) γ



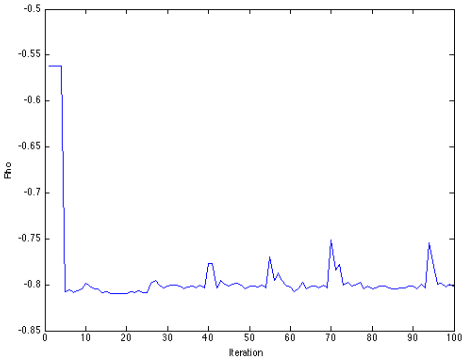
(C) κ'



(D) κ

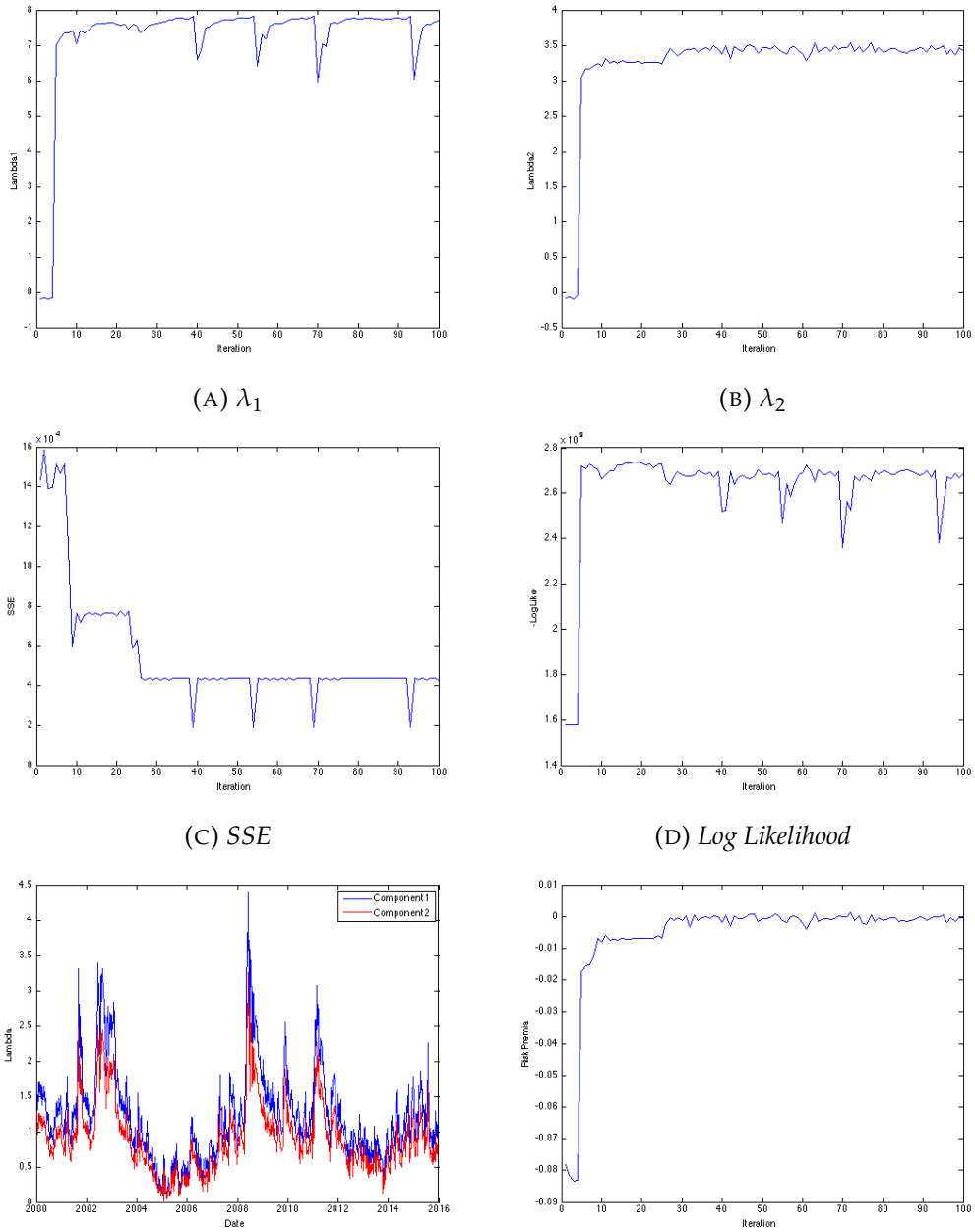


(E) σ



(F) ρ

FIGURE 3.5: *Results of the parameters estimation by the iterative procedure*



(A) λ_1 (B) λ_2

(C) *SSE* (D) *Log Likelihood*

(E) *Price of risk components* (F) *Financial price of risk*

FIGURE 3.6: *Results of the parameters estimation by the iterative procedure*

Parameter	Optimal	Bounds (L, U)
$\kappa'(\kappa)$	3.3181(2.0136)	$[0.0950, 15.7500]$
$\gamma'(\gamma)$	0.0516 (0.0850)	$[0.06, 0.08]$
σ	0.3800	$[0.3, 1.3]$
ρ	-0.9000	$[-0.9, -0.5]$
λ_1 (SE)	10 (1.2163)	$[-10, 10]$
λ_2 (SE)	3.4333 (1.0050)	$[-10, 10]$
SSE	775.0003	-
Risk premia	-1.1900	-

TABLE 3.7: Results of the full calibration procedure, $\lambda_2^{(0)} = 1$, weekly data;

3.4.4 Monte Carlo simulations

We simulate the Heston model as it is given in equation (3.4) under the objective measure P using the parameters obtained in the previous chapter. First, the state volatility process is simulated using the transition density which is known for the CIR-type processes. The distribution of Y_t given Y_u for some $u < t$ is a non-central chi-square distribution. The transition law of Y_t can be expressed as

$$Y(t) = \frac{\sigma^2(1 - e^{-\kappa(t-u)})}{4\kappa} \chi^2_d \left(\frac{4\kappa e^{-\kappa(t-u)}}{\sigma^2(1 - e^{-\kappa(t-u)})} Y(u) \right),$$

where $d = \frac{4\kappa\gamma}{\sigma^2}$ is the degree of freedom. This says, that, given $Y(u)$, $Y(t)$ is distributed as $\frac{\sigma^2(1-e^{-\kappa(t-u)})}{4\kappa}$ times a non-central chi-square random variable with d degrees of freedom and non-centrality parameter:

$$\lambda = \frac{4\kappa e^{-\kappa(t-u)}}{\sigma^2(1 - e^{-\kappa(t-u)})} Y(u).$$

Thus, we can simulate the process Y_t on a discrete time grid by sampling from a non-central chi-square distribution (see Glasserman (2003) [47] for more details).

Simulation of the logarithm of the stock process s_t under the objective measure P is done by discretising equation for s_t as it is specified in the system (3.4) with a time step δ:

$$s_{t+1} = s_t + \left[r - d + \lambda_1(1 - \rho^2) + \lambda_2\rho - \frac{1}{2} \right] Y_t\delta + [\sqrt{1 - \rho^2} Z_1 + \rho Z_2] \sqrt{Y_t}\sqrt{\delta}, \quad (3.27)$$

where Z_1 and Z_2 are two independent standard normal random variables.

We simulate trajectories of the variance process and the stock process weekly taking a horizon of one year. As a result of the simulations we compute the yearly VaR. To compare the characteristics of the distribution of the logarithmic returns with the actual one obtained from the history, we also compute the skewness and the kurtosis of the distributions. In addition we fit the Student's and the Normal distribution to the resulting distribution of returns to be able to compare the value of VaR with the standard parametric approach.

Data	VAR-05	VAR-1	VAR-5	Mean	Std	Skewness	Kurtosis
HistDaily	-0.6332	-0.5972	-0.4697	-0.0322	0.2201	-0.7458	2.7512
HistWeekly	-0.6374	-0.6046	-0.4713	-0.0322	0.2205	-0.7508	2.7536
Sim $\rho = -0.8039$	-1.0112	-0.8243	-0.5581	-0.0701	0.2702	-0.8980	4.5850
Sim Student	-0.8791	-0.7627	-0.5049	-0.0576	0.2371	-	-
Sim Normal N	-0.7993	-0.7296	-0.5394	-0.08011	0.2792	-	-
Sim $\rho = -0.7622$	-0.9093	-0.7905	-0.4876	-0.0249	0.2587	-0.8201	4.3536
SimStudent	-0.7545	-0.6549	-0.4269	-0.0096	0.2263	-	-
SimNormal N	-0.6912	-0.6267	-0.4504	-0.0249	0.2587	-	-
Sim $\rho = -0.1607$	-0.5522	-0.4868	-0.2908	0.1169	0.2494	0.0356	3.5222
SimStudent	-0.5670	-0.4874	-0.2906	0.1166	0.2325	-	-
SimNormal	-0.5257	-0.4634	-0.2934	0.1169	0.2494	-	-

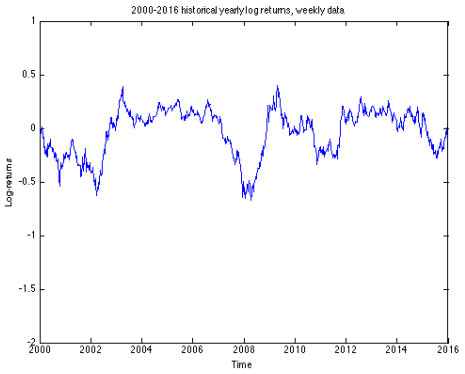
TABLE 3.8: Characteristics of the log-returns distribution, simulation for the weekly data

The results of the simulation are summarised in Table 3.8. The parameters used are from Table . However, we make additional simulations using correlation coefficient obtained from estimation on the historical data . Since the value of the correlation coefficient is always difficult to estimate, we want to see how different values of ρ effect the result. As expected, the value of ρ estimated from the realized volatility ($\rho = -0.1607$), does not produce simulations coherent with historical data (the skewness of the distribution is positive) as it is too low.

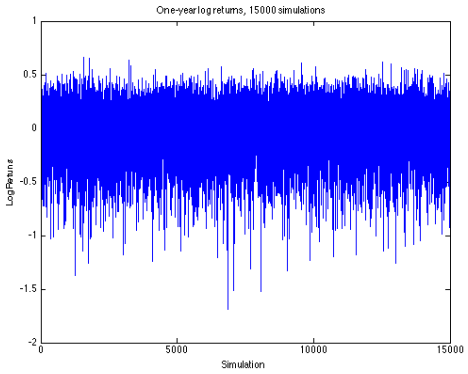
On the contrary, the value of $\rho = -0.7622$ estimated as correlation between the stock log returns and the differences in the VSTOXX Index provides results coherent with history, as well as the value of $\rho = -0.8039$ obtained via ML estimation. We can conclude that the parameters obtained with the the iterative procedure provide more prudent estimates of VaR due to fact that the resulting distribution exhibits higher level of kurtosis and skewness. As we can see from Figure 3.7 and 3.8, the resulting simulations are coherent with the history.

3.5 Conclusions

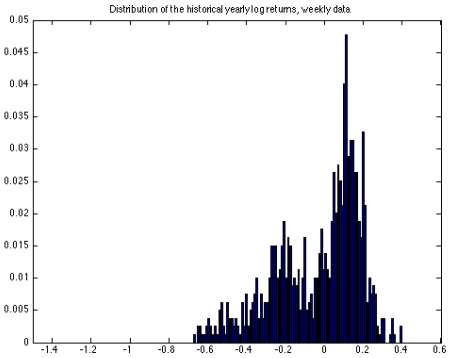
In this paper we have proposed a methodology to couple a standard technique of risk-neutral valuation of equity options based on the minimization procedure with the approach of Ait-Sahalia and Kimmel (2007) [3] which invloves approximating the unknown likelihood function. We showed that the standard errors of the estimated price of risk parameters are smaller than those reported by Ait-Sahalia and Kimmel (2007). We find that our procedure allows to reduce the error between the actual and the model price in comparison to the standard estimation approach of the Heston model by identifying iteratively the unobserved variance process and by correcting the parameters on each step. Although there are instabilities in the algorithm, with a large number of iterations the estimated parameters tend to converge to stable values. Simulations show that the calibrated parameters produce equity trajectories coherent with the history. The VaR value computed using obtained parameters tend to provide more prudent estimates compared to the historical and parametric approach. Future work could involve an extension to the state vector with an additional stochastic state variable representing an interest rate component.



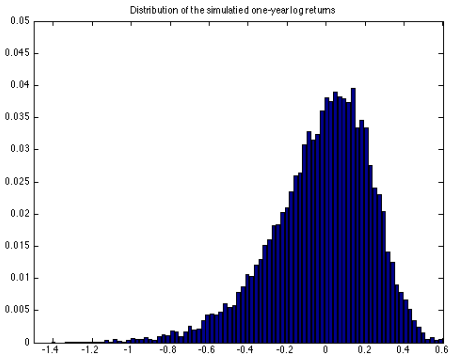
(A) *Historical yearly log-returns*



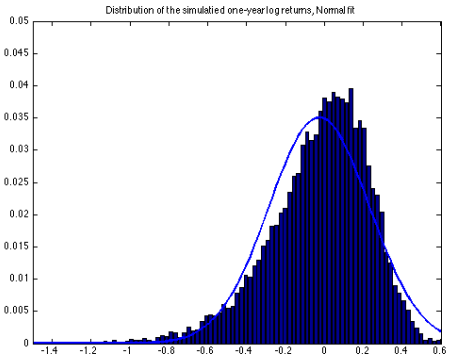
(B) *One year log-returns, 15,000 MC simulations*



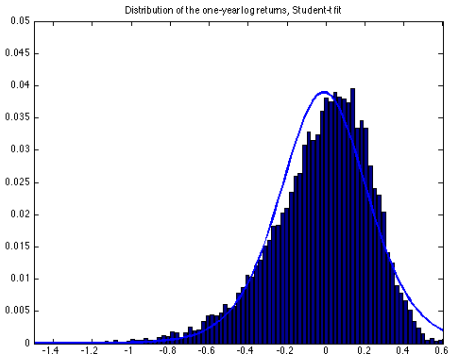
(C) *Distribution of the historical yearly log-returns*



(D) *Distribution of the one year log-returns, 15,000 MC simulations*

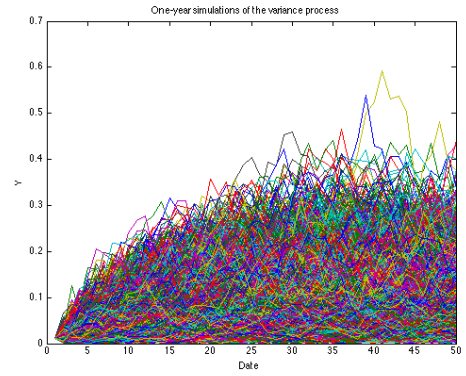


(E) *Fitted Normal distribution*

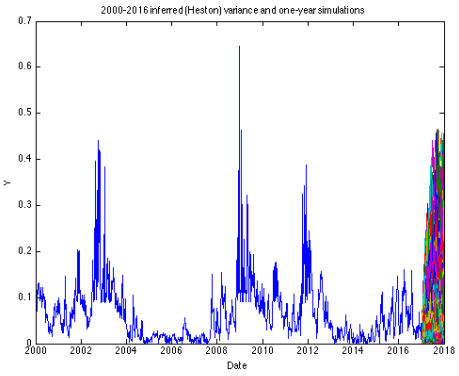


(F) *Fitted Student distribution*

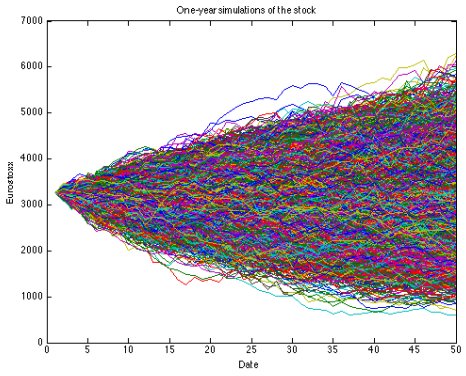
FIGURE 3.7: *Simulated and historical yearly log-returns, weekly data;*



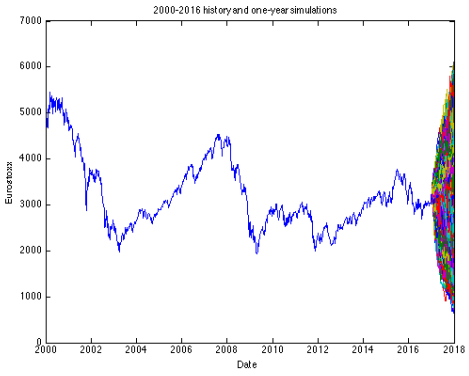
(A) *Simulated variance process*



(B) *Inferred variance process and simulation*



(C) *Simulated stock process*



(D) *History and silmulations*

FIGURE 3.8: *15,000 MC simulations, weekly data;*

Chapter 4

Network-based approach in modelling credit defaults

4.1 Introduction

This work studies how network theory can be used to improve the assessment of default credit models, both in the setting of structural and reduced-form models. In particular, we describe how to use a network effect for a bank who holds a portfolio of SME clients who are interconnected due to trade credit relations through a Supply Chain Network (SCN).

A Supply Chain Network (SCN) is an evolution of the basic supply chain which is defined as "a network used to deliver products and services from raw materials to end customers through an engineered flow of information, physical goods and cash" [90]. Nowadays organisations with a basic supply chain can develop this chain into a more complex structure involving a higher level of interdependence and connectivity between more organisations, this constitutes a supply chain network. SCNs can be used to show the flow of information and materials across organisations. Supply chain networks are typically structured with five key areas: external suppliers, production centres, distribution centres, demand zones, and transportation assets [68]. Thus, it involves direct supplier-customer connections together with indirect connections such as marketing and logistic services, financial and consultancy firms.

Trade credit is one of the most important source of external finance within SCN. It appears on every balance sheet and represented companies' short term liabilities. Firms simultaneously grant and receive trade credits, which therefore appear on both sides of their balance sheet. Trade credits are not well diversified at the firm level, as firms's customers tend to belong to a specific sector. In [64] a theoretical model of systemic risk propagation in credit chains is described. It is shown that a small liquidity shock may cause a chain reaction in which other firms get into financial difficulties, thus, generating a large persistent fall in aggregate activity. They model a set of small enterprises run by entrepreneurs who are unable to raise funds from investors, but an entrepreneur can borrow from its suppliers.

Recent studies highlight the importance of understanding the structural connections across companies and the implications of these connections for cross relations. Boissay and Gropp (2013) [17] document that trade creditors are likely to respond to late trade debtor payments by, in turn, postponing their own trade credit payments. A negative liquidity shock is transmitted along the trade credit chain until it reaches a trade creditor with access to external financing, or sufficient cash-holdings, in order to absorb the liquidity shock. Their result suggests that trade credit chains function as an insurance mechanism by allocating liquidity from unconstrained to constrained firms. Jacobson and von Schvedin (2015) [57] document the dark side of trade credit

by providing insights on its role as propagator of corporate failure. The results documented above offer a solid empirical indication about the relevance of interfirm credit relationships as drivers of distress propagation.

Given this framework, we would like to investigate three main question: *1. Can the inclusion of the network effect improve the bunckrupcy prediction ? 2. To which extend the network of firms as a whole is vulnerable to liquidity shocks? 3. What are the implications for the bank lending policy and its credit assessment strategy?*

To address these question we apply methods from graph theory, network analysis and spatial econometrics to model a default score. The Altman's Z-score model (1968) [5] is the first, pioneering approach to use financial ratios to identify or predict company bankruptcy. Since that time, it has been considered that the evaluation and apply of financial ratios has become a vital component for failure prediction techniques. In addition, Altman's Z-score model (1968) [5] is commonly utilised to assess company insolvency. His model composed of five linear combinations of business ratios, which used a multivariate approach, MDA, in order to measure the business performance or competence of a firm. For instance, financial ratios can be calculated as a criterion of company performance; those involving profitability, liquidity, capital structure and efficiency [5].

To include the interaction between the firms into assessment of the Z-score we adopt methods of spatial econometrics which is a set of statistical models and techniques which allows for the interaction of the variables within general linear regression models. In the spatial literature a widely used approach to model spatial interactions is through spatial lags. Dynamic panel data models that allow for spatial interactions in terms of spatial lags have recently been considered, see Kuersteiner and Prucha (2018) [69].

To estimate the panel spatial econometric model we reconstruct the spatial matrix of the dependencies between the firms by looking at the historical data on inferfirm transactions. The data is made available by a large anonymous bank. In order to achieve this we reconstruct relevant industrial and financial interactions among any pair of firms. To assess the use of working capital to sustain trade relationships, we analyse cash transactions between the firms and characterise them according to their volume and volatility. In time of distress, the survival of the chain will depend on the quality of the cash flows generated by the seller of the final good and by the capacity to access external financing sources by firms along the chain. This allows us to set up a simulation framework where an idiosyncratic shock originating within a certain company or industry transmits along the network in a contagious manner. The shock may be of different type. The primary type of shock we consider is a liquidity shock, since it has a direct link to transactions between the firms. Of course, the events which cause the liquidity shock may be of different origin – either systemic or idiosyncratic nature, such as a change in a macro economic environment, change in a policy or regulation related to a particular industry, technological innovation or operational shock such as a strike or change of company's management.

In order to quantify the direct and indirect relevance of firm payments within the network, we adapt methods from graph theory and network analysis. We assume that when modelling a network of interconnected firms the exposure to the idiosyncratic shock is given by the firm's position in the network. The adjacency matrix of interfirm connections is built for each period of time. The matrix itself can be constructed in multiple ways – it can reflect the intensity, relative weight or the volume of interactions.

The motivation to the banking industry is twofold. First, to assess the overall quality of its loan portfolio, a bank needs to understand the interactions among

its borrowers. In particular, the interaction between the prospective and existing borrowers should be an important factor in a bank's lending decisions. A more extended network of firms connected by cash transactions implies a higher amount of private information that the bank may use to assess the spillover effect that any decision about a credit policy toward a given firm may cause to other firms in the bank portfolio. A higher degree of connectivity will in general imply a lower level of diversification. Therefore the bank could explore the optimal tradeoff between the increase of information acquired when well connected firms become customers and the necessity to keep a well-diversified loan portfolio. Second, for risk management purposes, the effect on the expected credit looses of the portfolio can be quantified by simulating scenarios when the major clients experience financial destress. The outcomes of this experiment can be used to correspondingly adjust stand-alone credit ratings.

Computation of firm distress probability in reduced form credit risk forecasting models often relies on conditional independence assumptions. From an economic point of view this approach relies on the questionable assertion that the effects of a firm's financial distress do not reflect firms that have industrial or financial relationships with the defaulted entity. In fact there are many counterfactuals that indicate that such modeling approach needs to be improved. For example, the historical time series of observed defaults give evidence of distress events clustering, i.e. default event concentrate in time and localise over space and economic sectors. This simple observation confirms the necessity to extend current models introducing some form of dependence among the firm distress probabilities in order to improve prediction accuracy. Mainstream research in credit risk assessment, see e.g. Azizpour, Giesecke, Schwenkler (2015) [9], has identified two potential drivers of clustering: dynamic frailties or contagion. Frailty attributes clustering to the presence of some additional (unobserved) dynamic factor that drives up the distress transition intensities. Contagion channels impute clustering to the direct and indirect impact of defaults on the likelihood of other firm defaults.

In [96] the supplier-customer relation is examined from the perspective of an economic supply chain network, with each relation serving as a directed node of the network. The performance of several centrality measures is studied in capturing the major determinants of the supply chain network. It concludes that the supplier-customer centrality pair defined based on Kleinberg [66] algorithm brings the most value. They argue that the closure or production delay of a major supplier can cause significant issues for the company's production, while the changing demand of the customer influences the company's sales projection. They conclude that supplier central portfolios tend to be more volatile than customer central portfolios.

For listed companies the correlation structure is often obtained by performing PCA analysis on the matrix of returns. In [12] a PCA is performed on the returns of financial intermediaries to analyze systemic risk in the financial sector. They calculate eigenvector centrality using the variance-covariance matrix of asset returns as a measure of connectedness. For the companies which are not listed on the market, time series of returns are not available, however it is possible to use other type of data which the banks collects on its clients. In this work we use data on cash transactions happening between the companies to reconstruct the unobserved variance-covariance matrix.

This paper primary relates to recent work that studies networks in finance. Most of the work in this area focuses on interbank contagion or dependencies within financial institutions, such as in [11] and [12], while trade credit is an important part of commercial life and constitutes a significant fraction of total borrowing and

lending. Systemic risk is just as important for inter-firm lending as it is for interbank lending. The papers which study the propagation of distress among SME segment focus on the set of listed companies such as in [18] or [9] or use publicly available aggregate statistics [1], while we focus on the corporate segment of the bank which includes both listed and non-listed companies.

The innovation of this paper is to apply methods of graph theory and network analysis to a SME segment in order to provide a quantitative insight into the consequences of chain reactions. We aim to study how the network effect can be used to improve bank's lending decisions policy and credit risk assessment. To do so, we aim to calibrate the Altman's Z-score model in the setting of a spatial econometric approach using, first, standard financial ratios and, second, a historical data of the bank on the transactions between the companies. The results of the experiments statistically proof the potential of the network effect in prediction of financial distress since all the aggregate measures of the models performance (R^2, AUC, Accuracy and F-score) are higher for the models which include the network effect.

Although the experimental part of the paper covers only the prediction of bankruptcy in settings of the reduced from models, we also describe how the network effect can be incorporated into the Structural form models. Finally, we discuss possible applications of the constructed network useful for risk management purposes to support the bank in its lending decisions. Simulation of the shock is then equivalent to modelling network traffic [16] with an intensity of the shock decreasing with respect to distance from its origination.

4.2 Credit default models

For modelling credit risk two classes of models have been studied in the literature: structural and reduced-form. Structural models view a firm's liabilities as put options on the firm's assets. The firm's liability structure and the firm's asset value process are the key variables modelled in this approach. This methodology originated with Black and Scholes [15] and Merton [78].

In these models, the default time is usually characterized as the first hitting time of the firm's asset value to a given boundary determined by the firm's liabilities. As such, if the firm' asset value process follows a diffusion, then the default time is usually a predictable stopping time. The difficulties with the structural approach are twofold: first, the firm's asset value process is not directly observable, making empirical implementation difficult; and second, a predictable default time implies credit spreads should be near zero on short maturity debt. This second implication is well known to be inconsistent with historical market credit spread data, see [58].

The reduced-form approach was developed to avoid modeling the firm's unobservable asset value process. This approach was originated by Jarrow and Turnbull [59] and Duffie and Singleton [38]. Typically, reduced-form models characterize default as the first jump time of a point process, often a Cox process (i.e., a doubly stochastic Poisson process). As such, the default time is usually a totally inaccessible stopping time, implying non-zero credit spreads for short maturity debt.

The two types of models are viewed as competing with a reference to default prediction and/or hedging performance. However, it is noted that they are actually the same model containing different informational assumptions [58]. Such, structural models assume complete knowledge of a set comparable to that held by the firm's managers which implies that a firm's default time is predictable. In contrast, reduced form models assume knowledge of a less detailed information set as that observed

by the market. Which model is preferred depends on the purpose for which the model is being used. If one is using the model for risk management purposes, such as pricing and hedging, then the reduced form perspective is the correct one to take [58]. In marking-to-market, or judging market risk, reduced form models are the preferred modeling methodology. Instead, if one represents the management within a firm, judging its own firm's default risk for capital considerations, then a structural model may be preferred, but this is not the approach one wants to take for pricing a firm's risky debt or related credit derivatives. If one is interested in pricing a firm's risky debt or related credit derivatives, then reduced form models are the preferred approach.

Assume we are in a settings of a standard Black-Scholes model, i.e. we analyse a market with continuous trading which is frictionless and competitive in the sense that

- agents are price takers, i.e. trading in assets has no effect on prices,
- there are no transactions costs,
- there is unlimited access to short selling and no indivisibilities of assets, and
- borrowing and lending through a money-market account can be done at the same riskless, continuously compounded rate r.

Merton's model assumes the firm's assets at time t to follow a geometric Brownian motion:

$$dV_t = V_t \mu dt + \sigma V_t dW_t.$$

with drift μ and volatility σ, W_t is a standard Brownian motion.

Assume that there exists a money-market account with a constant riskless rate r whose price evolves deterministically as $\beta_t = e^{rt}$. In an economy consisting of these two assets, the price C_0 at time 0 of a contingent claim paying $C(V_T)$ at time T is equal to

$$C_0 = E^Q[e^{-rt} C_T],$$

where Q is the equivalent martingale measure under which the dynamics of V are given as

$$dV_t = V_t r dt + \sigma V_t dW_t^Q.$$

Here, W^Q is a Brownian motion under which the drift μ has been replaced by r .

A critical assumption is that this asset-value process is given and will not be changed by any financing decisions made by the firm's owners. Now assume that the firm at time 0 has issued two types of claims: debt and equity. In the simple model, debt is a zero-coupon bond with a face value of K and maturity date T. Hence, the default time τ is a discrete random variable given by:

$$\tau = \begin{cases} T, & \text{if } V_T < K \\ \infty, & \text{if else} \end{cases}$$

Figure 4.1 depicts the situation graphically as described in [45].

Since W_T is normally distributed with mean zero and variance T, default probabilities $p(T)$ are given by

$$p(T) = P[V_T < K] = P\Big[\sigma W_T < \log \frac{K}{V_0} - (r - \frac{1}{2}\sigma^2)T\Big] = \Phi\Big(\frac{\log \frac{K}{V_0} - (r - \frac{1}{2}\sigma^2)T}{\sigma\sqrt{T}}\Big).$$

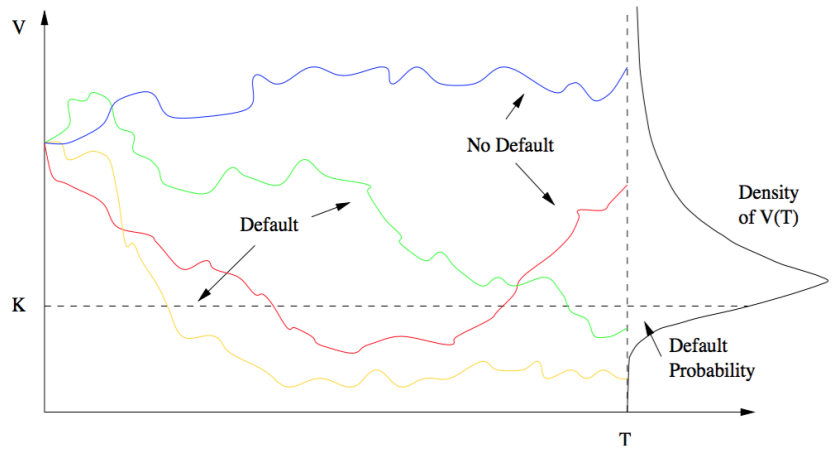


FIGURE 4.1: *Default in a classical approach*

Assets	Bonds	Equity	
No Default	$V_T \leq K$	K	$V_T - K$
Default	$V_T < K$	V_T	0

TABLE 4.1: Payoffs at maturity under the classical approach

With this assumption, the payoffs to debt, B_T , and equity, S_T , at date T are given as

$$\begin{aligned} B_T = min(K, V_T) = K - max(K - V_T, 0), \\ S_T = max(V_T - K, 0). \end{aligned} \tag{4.1}$$

If assets are worth less than K, equity owners do not want to pay K, and since they have limited liability they do not have to either. Bond holders then take over the remaining asset and receive a "recovery" of V_T instead of the promised payment K. The firm's equity is viewed as a European call option on the firm's assets with maturity T and a strike price equal to the face value of the debt. The debt can be viewed as the difference between a riskless bond and a put option. The payoffs to the firm's liabilities at debt maturity T are as summarized in Table 4.1.

Applying the Black-Scholes formula to price these options, we obtain the Merton model for risky debt. The values of debt and equity at time t are

$$\begin{aligned} S_t = C^{BS}(V_t, K, \sigma, r, T-t) = V_0\Phi(d_+) - e^{-rT}K\Phi(d_-), \\ B_t = De^{-r(T-t)} - P^{BS}(V_t, K, \sigma, r, T-t) = V_0 - V_0\Phi(d_+) + e^{-rT}K\Phi(d_-), \end{aligned} \tag{4.2}$$

where C^{BS}, P^{BS} are the Black-Scholes European call and put option formulas and

$$d_\pm = \frac{(r \pm \frac{1}{2}\sigma^2)T - \log\frac{K}{V_0}}{\sigma\sqrt{T}}$$

Based on the Merton model there has been developed a wide range of modifications incorporating stochastic interest rates, jumps in the assets dynamics or different levels of the debt seniority [70].

While the option-based approach provides a consistent way of thinking about default probabilities and prices of corporate bonds, it seems implausible that a single

value, the value of the firm's assets, is the sole determinant of default probabilities. Such indicators as the liquidity of assets and restrictions on asset sales are the key factors as well, see [70]. Reduced form models have been constructed specifically to be based on the information available to the market. In the reduced-form models the default time is modelled as a stopping time generated by a Cox process with a specified intensity. In this formulation, the stopping time is totally inaccessible. The liability structure of the firm is usually not continuously observable, whereas the resulting recovery rate process is. Thus, the recovery process is exogenously supplied. In this setting, many statistical methods which can be applied to the market data have been developed, including logistic regression, discriminant analysis, hazard regression, survival methods and Markov chains [70].

From a legal point of view, the concept of business failure can be defined in different ways. The bankruptcy process begins when firms are incapable of paying back their obligations to banks, suppliers, tax authorities and employees. When aggregate liabilities of firm override the face value of the company's assets, this leads to bankruptcy, whereupon the assets are utilised to repay a portion of outstanding debt. In contrast, insolvency is a case in which the company is no longer able to meet its financial obligations when debts become payable. Insolvency happens when current assets are less than current liabilities (Ahn, 2001 [2]).

Some examples of business failure are bankruptcy, bond defaults, bank loan defaults, insolvency, the delisting of a firm, liquidation and government interference through special financing (Altman and Narayanan, 2007 [7]). Financial distress is a term that is utilised excessively in the financial studies available. Levratto (2013) [71] defines it as whenever a company's liabilities exceed its book value of assets which predominantly leads to financial distress.

From a statistical point of view, firm default is an example of a qualitative response. We simply observe a firm's characteristics and whether it defaults or not. Thus, in this work, we do not differentiate between different types of financial failures and assume that any type of financial distress is considered to be a default state.

To find the most accurate model for prediction of default, a variety of financial ratios and failure prediction models have been applied. Multiple discriminate analysis (MDA) approach, the logit regression analysis (LRA), the artificial neural networks (ANN) model are the most popular models in this area. Among these models, the Altman Z-score has become a benchmark since it is widely used among researchers, academics and practitioners in various countries, see, for instance, [20], [92]. It is accepted to be the most accurate and most reliable model for predicting corporate failure. The Altman Z-score model consists of five financial rations based on the multivariate approach, Multiple Discriminate Analysis (MDA)

4.2.1 The Altman Z-score

The Altman Z-score model (1968) was the first study that identified companies as failed and non-failed companies. Altman's study (1968) [5] analysed 33 inactive and 33 active companies using MDA. Ultimately, Altman's Z-score results found that the accuracy of the first and the second year prior to failure was 95% and 72%, respectively. In the original Altman Z-score model (1968) for predicting bankruptcy public manufacturing firms were utilised. Later, private manufacturing firms were employed in the revised Altman Z-score model (1983) [6]. The accuracy of this last model was demonstrated by the 95% and 73% accuracy at year one and year two prior to failure, respectively.

Five financial variables (ratios) have been chosen by Altman based on their capacity to predict for company bankruptcy. They are liquidity, profitability, leverage, solvency and activity. Altman's original Z-score model (1968) [5] equation is:

$$Z = 0.012X1 + 0.014X2 + 0.033X3 + 0.006X4 + 0.999X5 \tag{4.3}$$

Based upon Altman's formula, the firms were classified into three categories according to the company's sustainability. For instance, if the firm is in the distress area then there is a strong probability of failures when the Z-score index of the company is below 1.8. On the other hand, when the Z-score index exceeds 2.99, it is considered that the enterprise is in the safe zone, with a low percentage of company failure. Moreover, when the value of the Z-score index is greater than 1.80 and less than 2.99, there is no strong evidence to specify the financial condition of the company; that is, the results cannot precisely ascertain whether the company is in the safe or distressed zone [5].

$$\begin{aligned} Z &< 1.80 - DistressZone, \\ Z &> 2.99 - SafeZone, \\ 1.8 &< Z < 2.99 - GreyZone. \end{aligned} \tag{4.4}$$

The Altman Z-score model (1968) [5] and the Ohlson's model (1980) [84] are two models that are well accepted and commonly used at present [92]. Below we describe the financial ratios of the Z-score regression:

X1, Working Capital to Total Assets The working capital to total assets ratio is one of the commonly found ratios in the research of firm issues. It is a measure of the net liquid assets of the corporate in comparison to the overall capitalisation. The differences between current liabilities and current assets are considered as working capital. Obviously, size and liquidity features should be taken into consideration. Generally speaking, current assets are found to be low in comparison to total assets, when a company undergoing consistent operations fails.

X2, Retained Earnings to Total Assets The overall amount of reinvested losses or/and earnings of a corporate during its whole life can be obtained by retaining earnings. This is also called earned surplus. It is worth noting that an earned surplus account is subject to manipulation by stock dividend announcements. This ratio is found to be implicitly affected by the age of a company and an old company might have higher retained earnings/total assets ration than a young company. This is because the younger company has not had enough time to increase its cumulative profits. Therefore, this analysis is argued not to be appropriate for young companies because their chance of being classified as a failed company is higher compared to the chance of older company.

X3, Earnings Before Interest and Taxes toTotal Assets The true productivity of a company's assets is measured by the EBITDA/TA ratio without taking into consideration leverage or tax factors. This ratio is believed to be extremely appropriate for investigating firm bankruptcy because the ultimate existence of the company depends on earning power [5].

X4, Equity to Book Value of Total Liabilities Liabilities is the measuring of both the long and current term, while equity is found to be the market value of all the shares of common, preferred and stock. This measure demonstrates how much

the firm's assets might decline in value before the assets become lower than liabilities and the company goes bankrupt. In other words, a firm with a market value of its debt of 500 and its equity of 1000 might experience a two third decrease in asset value prior to bankruptcy. Nevertheless, if assets decrease one third in value, the same company with 250 equity will failed.

X5, Sales/Total Assets Ratio is the well-known ratio showing the sales generating efficiency of the company's assets. It is widely used for dealing with competitive situations. This ratio is considered to be the least considerable ratio on an individual basis. Consequently, it is found to be quite an important ratio. It is ranked as the second most important ratio for contributing to the total discriminate ability of the model because it has a unique and quite significant association to other variables in the model [5].

4.2.2 Spatial Linear Regression Model

This chapter discusses different specifications of linear spatial econometric models that can be considered once the hypothesis of no spatial autocorrelation in the disturbances is violated. A general form to take into account the violation of the ideal conditions for the applicability of OLS is given by the following set of equations [8]:

$$\begin{aligned} y &= \rho W y + Z\beta + u \\ u &= \alpha W u + \epsilon, \end{aligned} \tag{4.5}$$

where X is a matrix of regressors, W is a weight matrix exogenously given, $\epsilon|X \sim N(0, \sigma_\epsilon^2 I)$, β, ρ and λ are the parameters to be estimated with $|\lambda| < 1$ and $|\rho| < 1$. $Z = [X, WX]$ the matrix of all regressors, current and spatially lagged. The first equation considers the spatially lagged variable of the dependent variable y as one of the regressors and may also contain spatially lagged variables of some or all of the exogenous variables (the term WX). The second equation considers a spatial model for the stochastic disturbances.

This model was termed SARAR (acronym for Spatial AutoRegressive with additional AutoRegressive error structure) by Kelejian and Prucha (1998) [63] and encompasses several spatial econometric models. In particular, we have five cases:

1. $\beta = 0$ and either ρ or $\lambda = 0$, known as the pure spatial autoregressive model
2. $\rho = \lambda = 0$, known as the Lagged independent variable model
3. $\rho = 0, \lambda \neq 0$ known as Spatial Lag Model (SLM)
4. $\rho \neq 0, \lambda = 0$ known as Spatial Error Model (SEM)
5. $\rho \neq 0, \lambda \neq 0$ the complete model (SARAR)

With respect to out data, the y variable is the credit rating of the firm, X is the matrix of independent Z-score variables described in 4.3 and W is the matrix describing connections between the firms.

4.2.3 Relationship between Supply Chain Network, credit chains and cash transactions

Here we describe how the data on transactions between the firms relates to the the credit chains in the customer-supply network.

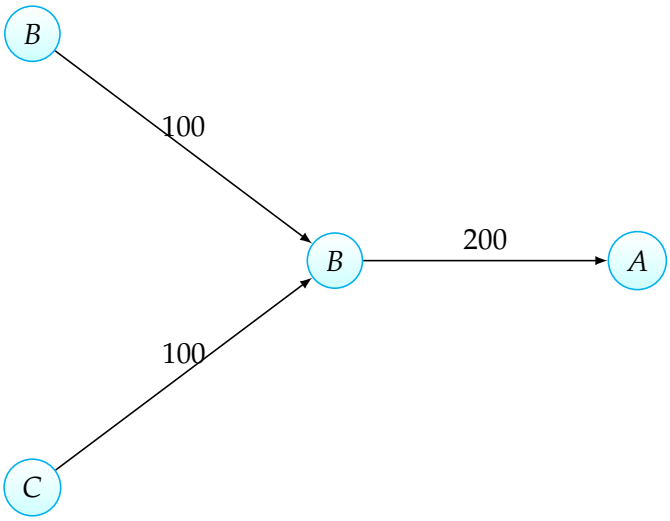


FIGURE 4.2: A supply chain: D and C supply to B; B supplies to A

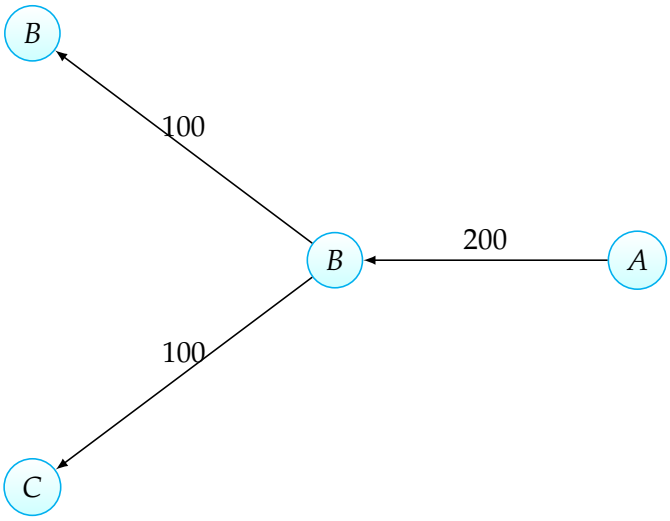


FIGURE 4.3: A credit chain: A is in debt to B; B is in debt to D and C

Each entrepreneur simultaneously lends to his customers (other entrepreneurs) and borrows from his suppliers. Because his balance sheet has financial assets (account receivable from his customers) and liabilities (account payable to his supplies), he is exposed to credit risk of his debtors/customers. Such, if his customers experience a negative liquidity shock and default, he himself may run into financial difficulties and may have to default agains his suppliers causing further difficulties along the credit chain.

To illustrate the concept, consider the following example in spirit of [64]. Suppose an entrepreneur A has ordered 200 units of specific goods from an entrepreneur B at 1 dollar per unit. That is, A now owes 200 to B to be paid at the delivery date. At the delivery date A expects to have 200 to pay to be. At the same time B has ordered 100 units of specific input from entrepreneur C and 100 – from entrepreneur D: B owes 100 to each at delivery. B has no cash of his own, but expects to use the 200 that A owes to him to pay his debt to C and D. This may be a link in a longer supply chain, which in turn might be part of a larger network. The supply chain of intermediate product is illustrated in Figure 4.2 and the corresponding credit chain in Figure 4.3.

Now suppose that A finds her cash holding only at 120 rather than the 200 she

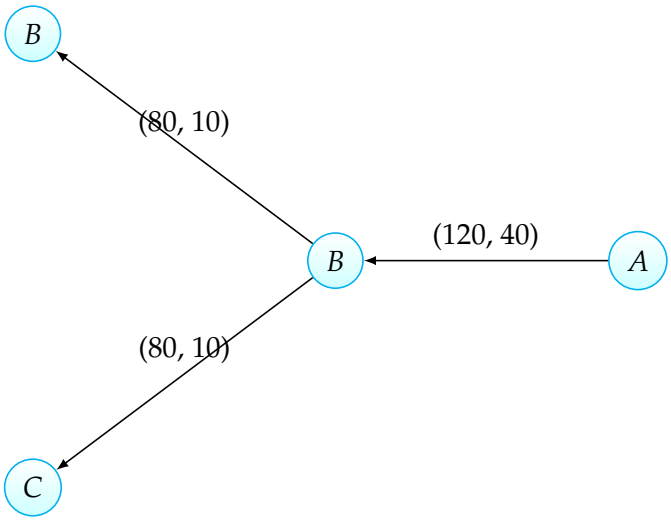


FIGURE 4.4: Transactions happened between the firms

expected. She is unable to meet her obligations to B. If, following A's default, B continues to charge 1 dollar for each unit, A can only afford to take 120 units. B can liquidate the remaining units, but since they are specific to A, they fetch less than 1 dollar, but 0.5 for instance. As a result, B get 160 in total: 120 in cash from A plus 40 in liquidation receipts. 160 is less than 200 B had to pay to his supplies D and C. He pays 80 to D and 80 to C. If each of them liquidates the remaining 20 at 0.5 per unit, D receives 90 and C receives 90. If D and C are debtors, there may be further default along the credit chain. The transactions happened in the trade chain are illustrated in Figure 4.4

These links are part of an intricate network of such credit/supply relationship: B has many customers like A and many suppliers like C and D. Note that the figures here may be misleading as the actual credit/supply network consists of combinations of bilateral trade. They are used only for illustrative purposes.

If we now assume that the four companies have a long history of trade between each other with a stable flow – in our illustrative examples it would mean, that A in fact needs the same input goods from B and it orders the same 200 amount every month, while B orders every month 100 form C and 100 from 100. Thus, having time series on cash transactions between the firms we would be able to identify that at time t, when A paid only 120 in cash instead of 200, there was a liquidity shock.

Note, that the initial lack of liquidity may be caused by both systemic or idiosyncratic risk factors.

Example of the spread of industrial risk is the recent hurricanes in the US (Sandy and Irma): the storm basically stopped and damaged the oil production and processing in the Gulf of Mexico. This affected the energy industry and the insurance industry. The energy industry doesn't have any oil supplies so needs constant influx of oil from the oil companies themselves. While the insurance industry had to pay out 95 billion US dollars in insurance claims to the oil industry in the Gulf of Mexico. Therefore, the all needed to re-insure themselves at SwissRE. Because of the premiums they had to pay on their re-insurance, the premiums of their customers had to go up, including construction companies and retailers who have to rebuild all the damaged properties. This had lead to a cool-down in the economic growth of the US, especially the Southern states along the Mexican Gulf. Half of the *S&P* 500 listed firms reported losses in Q3 of 2017.

Firm-level shock may happen due to changes into a company's management or a shock on operation, such as a strike.

4.2.4 Incorporating the network effect

While modelling credit risk, except for providing an adequate description of default arrival at the single firm level, one needs to take into account dependence modelling. Research on default dependence has been developed in the recent years. It is well known that the defaults of different issuers are correlated through time [45]. Two patterns have been found in the time series of spreads – that the spreads vary smoothly with general macro-economic factors and that there are jumps in spreads. The first factor is due to a common dependence on the economic environment, while the second describes dynamics common to several firms or even entire market. This suggests that the credit risk of one issuer can propagate to other issuers as well. The reasoning behind this is that economic distress is contagious and propagate from firm to firm. A typical channel for these effects are borrowing and lending chains. Thus, the financial health of a firm depends on a status of other firms. Important role in this setting play liquidity shocks in credit chains.

The importance of trade credit chains for the propagation of corporate bankruptcy is quantified in [57] using an exhaustive data set on claims held by Swedish trade creditors (suppliers) on failed trade debtors (customers). The propagation mechanism explains a significant part of the aggregate bankruptcy frequency, especially during economic downturns. More specifically, they show that the propagation mechanism increased the overall bankruptcy frequency by around 13 percent during the Swedish banking crisis in the early 1990s. In [17] it is documented that trade creditors are likely to respond to late trade debtor payments by, in turn, postponing their own trade credit payments. A negative liquidity shock is transmitted along the trade credit chain until it reaches a trade creditor with access to external financing, or sufficient cash-holdings, in order to absorb the liquidity shock. Their result suggests that trade credit chains function as an insurance mechanism by allocating liquidity from unconstrained to constrained firms. In [96] the performance of several centrality measures is studied in capturing the major determinants of the supply chain network. It concludes that the supplier-customer centrality pair defined based on Kleinberg [66] algorithm brings the most value. The supplier centrality of a company is defined as the sum of the customer centralities of all its customers and the customer centrality of a company is defined as the sum of the supplier centrality of all its suppliers. The centralities can be solved as the leading eigenvector of the products of the supplier-customer network matrix. They argue that the closure or production delay of a major supplier can cause significant issues for the company's production, while the changing demand of the customer influences the company's sales projection. They conclude that supplier central portfolios tend to be more volatile than customer central portfolios.

Below we briefly explain how the network information can be embedded in both reduces and structural form models:

Reduced Form Models. Suppose, the distress probability is forecasted using a classical Z-score model. Then the Spatial Econometric model setup would allow to include the network effect in the prediction of the financial distress. If the connectivity matrix is built from the cash transactions flowing from a firm i to firm j, it would capture the exposure of a firm j to any of its customers/debtors. The inclusion of the lag of the Rating variable into a Z-score regression would mean that the rating of the customers/debtors of a firm j influences its own credit rating and the intensity of the impact of a debtor/customer i is proportional to the weight w_{ij} of the

matrix of connectivity. The inclusion of the lag of order two would mean that the debtors/customers of the debtors/customers of the firm j have also an effect on the credit rating of a firm j. This interpretation can be extended to the lag of order n. The same holds for the independent variables X. Such, an inclusion of the lag of order one, would mean, for instance for X1, that the decrease in Working Capital to Total Assets ratio of the debtor/customer i of a firm j would decrease the same ratio of the firm j proportionally to the weight w_{ij} of the matrix of connectivity.

Structural Form models. According to the Merton [78] model, debtors of a firm own a claim on assets of the firm whose face value is equal to the face value of debt. Firm payable and receivable modify the size and the riskiness of the assets and of the defaultable claims of a firm. In fact, because accounts receivable are junior claims their value is sensitive to the financial strength of the debtor firm. A distress shock on these account receivables will unavoidably affect the creditworthiness of the creditor firm. In other terms, the assets of a firm will have a composition depending also on the receivable claims. Similarly, the downstream firms' accounts payable is a liability of the firm backed by its assets, including its own accounts receivable against customers yet further down the chain. We may think of a firm's payable accounts as defaultable debt backed by its assets.

Computation of the probability of distress can be performed analyzing the probability that the asset value is sufficient to repay the claim at maturity. Standard sensitivity analysis can be used to quantify the rise in the distress probability of Firm A if Firm B is hit by a distress shock that lowers the probability to repay receivables due to Firm A from Firm B. The higher the share of assets of a firm A depends on receivables from firm B, the higher is the increase in distress probability of the firm A implied by an increase in the probability of distress of firm B.

4.3 An introduction to network theory

Theory of networks is a relatively recent brach of science which has been rapidly developing during the last decades [81]. Its subject of study are various systems the elements of which are interconnected. Such systems are called networks. The study of networks pervades all of science, from neurobiology to statistical physics. Its aim is to understand how an enormous network of interacting dynamical systems – be they neurones, power stations, lasers, humans or financial institutions – will behave collectively, given their individual dynamics and coupling architecture.

Theory of social networks emerged in the 1920's, initially as a brach of sociology which studied personal ties among individuals. The sociologists were interested in such questions as "Who has the most connections, and thus, the most influence?" or "Who is transferring the information in the network and how fast?". A famous hypothesis in this area is the theory of Six degrees of separation. It is an idea that all living things and everything else in the world are six or fewer steps away from each other so that a chain of "a friend of a friend" statements can be made to connect any two people in a maximum of six steps. This characteristic corresponds to the average path length in the graph.

Initially emerging in sociology, theory of networks found a wide range of applications in the modern world [81]. Such, the World Wide Web is a network of web pages connected with hyperlinks. The network theory is applied in biology, epidemiology, for studying the spread of computer viruses and properties of natural languages.

In the financial area the network science found an application in risk management, especially after the economic crises, when it became evident that a high degree of

interdependence within the financial system increases the systemic risk. However, recent studies show that while more links between banks might be expected to increase the risk of contagion, the banking systems with a more complete set of connections may be less susceptible to contagion than those with an incomplete structure [67].

A network approach to financial systems is particularly important for assessing financial stability and can be instrumental in capturing the externalities that the risk associated with a single institution may create for the entire system. Thus, what matters is the level of heterogeneity of the financial system – not to have too many big players (such as "too big to fail" banks) and not to have too many isolated firms. To measure these properties of the financial system networks metrics and techniques for identifying central, vulnerable or systemically important institutions and markets are employed. As a results, social network analysis may be useful for financial institutions to make credit and investment decisions, to analyse the counterparties' risk exposure from their interconnectivity within the financial system, as well as to develop regulatory strategies in order to improve financial stability.

Formally, a network is usually defined with a graph $G = (V, E)$, containing a finite set of vertices (or nodes) V and a set of directed or undirected edges E. The elements of the network are then the nodes of the graph, while the relationships between the elements are edges of the graph. A mathematical model of the network is a matrix $A = a_{ij}$ of connections between the elements of a set V. An element of a matrix $a_{ij} = 1$, if a node i is connected to the node j, and equals to zero, otherwise. Thus, in a case of undirected edges the matrix A will be symmetrical. By convention, entries on the main diagonal of A are set to equal to zero (self-loops are excluded).

In the last decades network theory gained a lot on the quantitative part – due to physicists, who contrary to sociologists were looking at the macroproperties of the networks such as average length of path in the graph, average number of connections of an element, distribution of the indegree or outdegree of the network, centrality. We state some of them.

Path length is the average length between any two nodes of the graph.

Degree of a node $d(v_i)$ is the number of edges adjacent to it (neighbourhood of node v_i). In a directed graph a node has in-degree and out-degree. Thus, $d(v_i)^{in}$ denotes the nodes pointing to v_i, while $d(v_i)^{out}$ denotes the set of nodes v_i points to.

Degree distribution is a distribution function which for each k defines a probability that any chosen node in a graph has k edges adjacent to it[**?**]:

$$P(k) = \frac{N_k}{N}.$$

Here N_k is the number of nodes in the graph with a degree equal to k.

In a large number of networks degree distribution follows the Power Law (or Pareto Distribution Function) [82]:

$$P(k) = Ck^{\alpha}.$$

A large number of patterns in physics, biology, studies about the planets, economics and finance, informatics, as well as social sciences and demographics are described by the Power law. Such, the distribution of the size of the cities is described by this law[82]: there are much less big cities in the world like London, New York or Moscow, rather than smaller cities, the number of which is large. It is well known as well that the income of the population also follows the Pareto law. The Pareto principle is sometimes stated in popular expositions by saying 20% of the population

has 80% of the income. This corresponds to the $\alpha = 1.16$, whereas the 70-30 rule corresponds to $\alpha = 1.42$.

4.4 Experiments setup

The main objective of this experiment is, first, to verify the accuracy of the Altman Z'-score model (1968) [5] in order to determine whether it is an optimal model for predicting corporate failure using the data of Italian SME companies in the period 2013-2015 and, second, to study if the inclusion of the network effect in the classical Z-score model improves the predictive power of the Z-score model.

The beginning of this chapter describes the construction process of the data used for panel regression and the construction of the matrix of connectivity based on the transactions data. Then, we start our experiment by first identifying a coherent classical Z-score regression model, after which we proceed with an inclusion of the network effect as a spatial dependency in the regression. Finally, we analyse the performance of the models based on the information retrieval measures commonly used in estimating performance of classification models.

4.4.1 Data description

The general architecture of the database is based on an innovative scheme where data are categorized in two classes called node based and edge based information. The first traditional class provides all the information relevant to a single firm on a stand alone basis, which conceptually represents the nodes of the hypothetical network of firms. Edge based classification of data qualifies all the available information regarding the relationships between any fixed pair of firms present in the database. This information is useful to reconstruct the relevant industrial and financial interactions among any pair of firms. The constructed graph is a directed one where edge may correspond to different types of relations. Typical data associated to edges are the bank wire transfers and their factoring transactions, accounts payables and receivables.

The process of database construction involves several steps. First, a set of companies which are classified as SMEs is identified. The companies are selected according to the type of portfolio they belong to. Note, that the classification of companies as SMEs according to the definition of the Basel II capital accords is that the reported yearly sales for the consolidated group the firm belongs to are less that 50 million euros [22]. We collect the data for three consecutive years – 2013-2015. Such, for the year 2015 we obtain 555,727 firms. However, the number of firms present in each of the three years reduces to approximately 200 thousand.

On the second step potentially useful areas of the bank internal database were identified and classified as a node or edge information. The financial balance sheet data is commonly used to build a single-firm indicator. For this study we choose a selection of the financial variables which are relevant for the estimation of the Z-score model such as total assets and liabilities, short term assets and liabilities, retained earnings, sales, EBITDA and a book value of equity. We extent this set of indicators with information describing the use of credit by the firms – in particular, the information provided by the Bank of Italy on the total outstanding credit exposure in the economy. The credit registry from the Bank of Italy provides information on the indebtedness to banks and financial companies (intermediaries) of the customers. Intermediaries report monthly to the Bank of Italy the total amount of receivables or credit open to their customers – in particular, receivables equal to or greater than

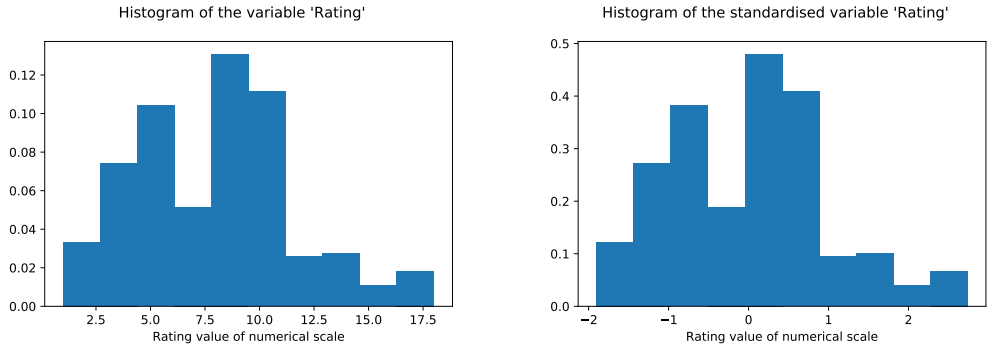
Numeric Rating variable	Equivalent S&P grade
1	AAA
2	AA+
3	AA
4	AA-
5	A+
6	A-
7	BBB
8	BB+
9	BB
10	BB-
11	B+
12	B
13	B-
14	CCC
15	CC
16	C
17	R
18	D

TABLE 4.2: Obtained numerical scale for the Rating variable

30,000 euros and non-performing loans of any amount. As a return information flow, the Bank of Italy monthly provides banks and financial companies with information on the total indebtedness towards the credit system of each customer reported.

Construction of the reliable dataset imposes several challenges such as data clearing, keeping coherent time stamps and excluding externalities. To obtain a reliable time-series for a period of three years, out of the identified set of companies and variables, we keep only those firms which have the information on the credit line utilisation present on a monthly basis. The balance sheet data is collected on the yearly basis. Regarding the balance data, we select all companies which have each of the cZ-score financial indicators present in each of the three years. Then, the data is duplicated on a monthly basis for each year. The final number of firms used in the experiments is 51,023. These firms satisfy a condition of the Rating variable present on at least a quarterly bases and a condition of the information on the credit line utilisation present on a monthly basis. Considering that we use the monthly data in the estimation, we obtain 1,785,805 records. To exclude extreme values during the estimation we replace any observation below the 1st percentile with the 1st percentile, and any observation above the 99th percentile with the 99th percentile. The variable Rating is the independent variable. Since not all the companies had all values of the Rating variable present for each month, we impute the missing values. The missing values for the Rating variable are filled with the rating of the company from the closest previous month for which the rating is available. Since the internal rating scale of the bank is a categorical one, we convert it to a numerical scale. The conversion of the internal rating scale is performed based on the equivalent $S\&P$ rating gradation. The resulting numerical scale is constructed in such a way that the higher is the numerical Rating value, the riskier is the firm. The obtained numerical scale for rating, together with its equivalent $S\&P$ gradation, is described in Table 4.2.

According to the internal definition of the bank, the firm is considered to be in



(A) *Histograms of the numerical Rating variable* (B) *Histograms of the standartised Rating variable*

FIGURE 4.5: *Obtained numerical scale for the Rating variable*

financial distress (we refer to this state as a default state) if its rating is CCC or worse. Thus, we do not differentiate between different types of financial failures – it may be bankruptcy or insolvency. The fraction of the firms which has been assigned a default rating during the three years period is approximately 6.2185%. This suggests that the collected data may be biased towards less risky firms since the companies in distress might disclose less data to the bank. Moreover, out of firms which appear in the network of payments, the rate of default is only 3.1829%. This fact suggests that the connected companies are less risky than those which are not connected.

To estimate the Z-score we standatrise the numerical Rating variable and all independent variable. As a result the percentage of the firms in distress correspond to the 1.39 quantile of the standartised Rating variable. Note, that in the original Altman Z-score regression, the cut-off point for the high risk of bunckrupcy is 1.80 [5], while in the revised [6] it is 1.23. The histograms of the numerical Rating variable and standatrised Rating variable are illustrated in Figure 4.5. The reader can explore the corresponding histograms of the independent variables in Appendix C. Note that in the classical Z-score regression, the risky firms appear in the left tail of the distribution, while in our rating scale they appear in the right tail. Thus, the coefficients of the Z-score regression are expected to be all negative.

To explore the discriminative power of the independent variables, we report the statistics of the means of the initial and standartised variables within default and non-default classes. This is summarised in Table 4.3. The first five variables are the standard Z-score variables, while the last variable corresponds to the ratio between the used and the allowed credit. As it is expected, statistics on the average of the means within each class, proof that lower values of the Z-score ratios indicate financial distress, while on the contrary, lower values of the UA Ratio indicate financial strength of the firm.

4.4.2 Matrix of connectivity

There exist different types of transactions which connect the firms. Here we concentrate on the wire transactions and describe how this data can be incorporated into a Z-score regression model. The network of interactions between customers of the bank corporate segment is constructed using the information collected by observing quarterly aggregated cash account transactions for each firm. Each transaction is characterised by:

1. ID

Variable	Mean	Mean of the default class	Mean of the non-default class
Ratio1	0.10627	-0.16554	0.12296
Ratio2	-0.00940	-0.00946	-0.00940
Ratio3	0.05673	-0.02389	0.06168
Ratio4	0.22600	0.08774	0.23449
Ratio5	4.72113	2.329446	4.86806
UA	58.28591	744.77433	16.11250
Ratio1 std	0.00000	-0.25472	0.01565
Ratio2 std	0.00000	-0.00084	0.00005
Ratio3 std	0.00000	-0.14236	0.00875
Ratio4 std	0.00000	-0.63591	0.03907
Ratio5 std	0.00000	-0.01732	0.00106
UA std	0.00000	0.048981	-0.00300

TABLE 4.3: Statistics on the independent variables (initial variables and standartised)

2. source
3. target
4. bonifico transactions (number of transactions)
5. bonifico total (total amount).
6. other product transaction information

Based on this, the information we collect on each firm's transactions on a quarterly basis is the following:

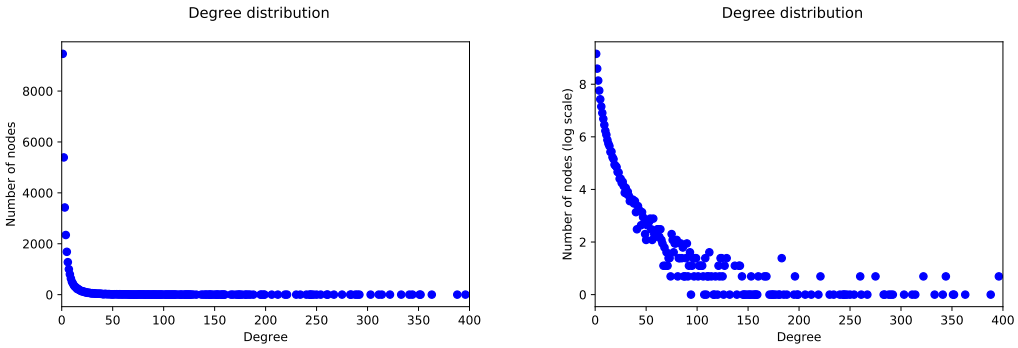
1. transactions out (number of operations)
2. transactions in (number of operations)
3. total out (total amount)
4. total in (total amount)

The final set of firms amenable to the reconstruction of the network is selected using a minimum score threshold on its total incoming or outgoing amount. The firms belong to the SME segment according to the internal bank classification. The nature of the rough data describing connections between the firms requires a transformation prior to the estimation. In particular, we need to transform it into a weighted sparse matrix where the weights quantify the strength of links between any two firms. In doing so, to each network edge connecting firm i to firm j, we assign a weight corresponding to the monetary values of the payments flowing from the cash accounts of firm i to the cash accounts of firm j over the total payments received by firm j. The connection structure is estimated using the sum of the quarterly aggregated payments.

The resulting weight, w_{ij}, corresponds to the ratio between the sum of monetary cash accounts from a firm i to a firm j divided to the total monetary value of the cash

Id	N firms	Mean Degree	Std(Degree)	Max Degree	N links	Density (%)
1	15,685	3.2022	13.475	811	50,227	$2 * 10^{-4}$
2	15,758	3.1776	13.464	877	50,072	$2 * 10^{-4}$
3	15,406	3.1048	13.184	830	47,832	$2 * 10^{-4}$
4	15,699	3.1226	13.123	833	49,022	$2 * 10^{-4}$
5	15,532	3.1841	13.224	808	49,456	$2 * 10^{-4}$
6	15,456	3.1040	13.159	853	47,976	$2 * 10^{-4}$
7	14,963	3.0647	13.018	781	45,857	$2 * 10^{-4}$
8	14,509	2.8611	11.482	674	41,512	$2 * 10^{-4}$
9	15,127	3.1398	13.948	797	47,496	$2 * 10^{-4}$
10	15,533	3.1453	13.455	804	48,856	$2 * 10^{-4}$
11	15,377	3.0908	13.116	758	47,527	$2 * 10^{-4}$
12	15,742	3.1223	13.189	770	49,151	$2 * 10^{-4}$
Full	30,877	7.3205	35.264	2,892	226,033	$2 * 10^{-4}$

TABLE 4.4: Statistics on the networks



(A) *Original scale* (B) *Log scale*

FIGURE 4.6: *In-degree distribution of the aggregate network*

transactions received by a firm j:

$$w_{ij} = \frac{A_{ij}}{\sum_k A_{kj}}, \tag{4.6}$$

here A is the adjacency matrix with an element A_{ij} being a sum of monetary cash accounts from a firm i to a firm j over a quarter. Thus, this measure reflects relative importance of a firm j to a firm i with respect to all the interactions of a firm i.

In this way we construct 12 matrices of connectivity corresponding to the 12 quarters of the three years. To obtain an aggregated matrix for the whole period of three years we sum the normalized weights over the 12 periods and row-normalise the final matrix. Table 4.4 summarises the statistics on the 12 quarterly networks and the aggregated one. The aggregated network contains only 30,877 firms. Figure 4.6 illustrates the in-degree distribution of the aggregate network. As expected, it follows a power law which proofs that the network is scale-free.

Dependent Variable :	Rating
Mean dependent var :	0.0000
S.D. dependent var :	1.0000
Number of Observations:	1785805
Number of Variables :	6
Degrees of Freedom :	1785799
R-squared :	0.2803
Adjusted R-squared :	0.2803
Sum squared residual:	1285302.511
Sigma-square :	0.720
S.E. of regression :	0.848
Sigma-square ML :	0.720
S.E of regression ML:	0.8484
F-statistic :	139079.6077
Prob(F-statistic) :	0
Log likelihood :	-2240294.051
Akaike info criterion :	4480600.102
Schwarz criterion :	4480674.475

Variable	Coefficient	Std.Error	t-Statistic	Probability
CONSTANT	0.000	$6.3 * 10^{-4}$	0.000	1.000
Ratio1	-0.162	$6.9 * 10^{-4}$	-233.998	0.000
Ratio2	0.059	$6.5 * 10^{-4}$	90.152	0.000
Ratio3	-0.162	$6.6 * 10^{-4}$	-243.068	0.000
Ratio4	-0.364	$7.0 * 10^{-4}$	-515.909	0.000
Ratio5	-0.172	$6.4 * 10^{-4}$	-264.436	0.000

TABLE 4.5: Results of the Z-score regression, estimation with OLS

4.4.3 Results of the estimation

As a first step, we estimate the classical Z-score regression with 5 ratios. For experiment we use a python Pysal package [1] specifically developed for spatial regression models. For the general OLS regression use the two-stage least squares algorithm which allows very large sample sizes. For the combined model we use the GMM algorithm (based on Kelejian and Prucha (1998) [63].

Estimating the OLS regression on the monthly data corresponding to 3 years and 51,023 firms (1,785,805 records), we obtain the results presented in Table 4.5 Although all the variables are significant, the coefficient in front of the Ratio2 is positive. This does not align with the expected result of the classical Z-score regression and, thus, we omit this variable in the further analysis assuming that the data obtained for this variable is not reliable enough.

Therefore, on the next step we estimate the regression with the four Z-score variables and the Ratio UA. The results are presented in Table 4.6. On the whole, the F-test is highly significant and leads to the acceptance of the model. Furthermore, all parameters are significant at the usual confidence level. The output also provides the results of several statistical tests. Both the JB (Jarque-Bera) and the BP (Breusch-Pagan)

[1] *http://pysal.readthedocs.io/en/latest/*

tests are significant, thus leading to the rejection of the two hypotheses of normality and homescedasticity. Lagrange multiplier test for a spatial lag model is significant indicating the appropriateness of the Spatial Lag Model. The Moran I test statistic for the hypothesis of spatial correlation of the residuals is also significant showing positive residual spatial correlation. All in all, the model is satisfactory, however, given the evidence of a residual positive and significant spatial correlation, we have clear indications of a Spatial Error model as an alternative framework, while the Lagrange multiplier test allows to conclude that the Spatial Lag Model should be used as well.

Based on the obtained tests' statistics we decide to estimate SLM, SEM and a combined SARAR model to assess which model provides better results in terms of fit and accuracy of predictions. To assess the performance of the models we compute the standard validation measures. Confusion matrix together with the Receiver Operating Characteristic curve, the Accuracy Ratio and the F-score are commonly used to estimate the results of the forecasts [2]. One should note though, that the unbalanced distribution of the classes in the dataset has an effect on the TPR and FPR.

Table 4.7 summarises the results highlighting the two best models according to each measure. It includes, for each model, the obtained statists on R^2, Area Under ROC Curve, True Positive Rate, False Positive Rate, Accuracy, Precision, Recall, F-score and Optimal Threshold – a cut-off on the distribution of the standartised Rating variable. The optimal threshold in information retrieval is identified by maximising the difference between TPR and FPR.

We remind that the initial Altman's Z-score provided the accuracy of the first and the second year prior to failure of 95% and 72%, respectively [5]. The accuracy of our Z-score regression is smaller with respect to the usual Z-score result. This can be explained by the use of the monthly data.

The results of the 6 models considered are comparable to those of the general OLS model estimated on the first step (in terms of the significance and of the sign of the variables). The parameter ρ related to the residual spatial autocorrelation is positive and significantly different from zero for all the spatial models considered. For all the 6 models we observe an improvement in the combined measures – Adjusted R^2, AUC, Accuracy and F-score. Although the improvement is relatively small – it quantifies in 2.1% increase in the Adjusted R^2, 0.5% increase in the AUC, 2.26% increase in the accuracy and 1.36% – in the F-measure, the result allows to infer that the network effect is potentially useful if the prediction of default. Additionally, we can conclude, that among the models considered, the combined models perform better in terms of Accuracy, Precision and F-score. However, in terms of R^2 and AUC, the SLM model with 3 lags provides as good result as the SARAR model with 3 lags. In Table 4.8 we report the resulting parameters of the SARAR model with 3 lags.

4.4.4 Illustration of the distress propagation in the network

In this section we illustrate with the data of the bank a process of distress propagation through a local customer-supply network of a firm which goes bankrupt. As it has

[2] We recall that the Receiver Operating Characteristics is a graphical plot that illustrates the performance of a binary classifier system as its discrimination threshold is varied. The curve is created by plotting the true positive rate (TPR) against the false positive rate (FPR) at various threshold settings. The true-positive rate is also known as sensitivity. The false-positive rate is also known as the fall-out or probability of false alarm. The ROC curve is thus the sensitivity as a function of fall-out. A measure that combines precision and recall is the harmonic mean of precision and recall, the traditional F-score.

Dependent Variable:	Rating
Mean dependent var :	0.0000
S.D. dependent var :	1.0000
Number of Observations:	1785805
Number of Variables :	6
Degrees of Freedom :	1785799
R-squared :	0.4421
Adjusted R-squared :	0.4420
Sum squared residual:	996387.621
Sigma-square :	0.558
S.E. of regression :	0.747
Sigma-square ML :	0.558
S.E of regression ML:	0.747
F-statistic :	282970.3494
Prob(F-statistic) :	0
Log likelihood :	-2012949.443
Akaike info criterion :	4025910.886
Schwarz criterion :	4025985.258

Variable	**Coefficient**	**Std.Error**	**t-Statistic**	**Probability**
CONSTANT	0.000	$5.6 * 10^{-4}$	0.000	1.000
Ratio1	-0.071	$6.2 * 10^{-4}$	-114.629	0.000
Ratio3	-0.116	$5.7 * 10^{-4}$	-201.616	0.000
Ratio4	-0.250	$6.4 * 10^{-4}$	-389.168	0.000
Ratio5	-0.104	$5.7 * 10^{-4}$	-180.254	0.000
UA	-0.453	$6.2 * 10^{-4}$	726.841	0.000

REGRESSION DIAGNOSTICS

MULTICOLLINEARITY CONDITION NUMBER	1.794

TEST ON NORMALITY OF ERRORS

TEST	DF	VALUE	PROB
Jarque-Bera	2	183551.601	0.000

DIAGNOSTICS FOR HETEROSKEDASTICITY
RANDOM COEFFICIENTS

TEST	DF	VALUE	PROB
Breusch-Pagan test	5	35045.575	0.000
Koenker-Bassett test	5	21155.703	0.000

SPECIFICATION ROBUST TEST

TEST	DF	VALUE	PROB
White	20	66716.799	0.0000

DIAGNOSTICS FOR SPATIAL DEPENDENCE

TEST	MI/DF	VALUE	PROB
Moran's I (error)	0.1318	157.003	0.000
Lagrange Multiplier (lag)	1	27919.008	0.000
Robust LM (lag)	1	5372.940	0.000
Lagrange Multiplier (error)	1	24648.915	0.000
Robust LM (error)	1	2102.847	0.000

TABLE 4.6: Results of the regression with Z-score four variables and the ratio of the Allowed to Utilised credit, estimation with OLS

Variable	OLS	SLM-1	SLM-2	SLM-3	SARAR 1	SARAR-2	SARAR-3
R^2	0.4420	0.4504	0.4510	**0.4512**	0.4505	0.4510	**0.4512**
AUC	0.836485	0.840081	0.840414	**0.840562**	0.840102	0.840381	**0.840526**
TPR	0.738316	0.740009	**0.748996**	**0.748816**	0.736056	0.74805	0.738577
FPR	0.220036	0.216193	0.224985	0.22486	**0.212401**	0.22418	**0.214789**
ACC	0.773302	0.773899	0.773396	0.777065	**0.790799**	**0.784756**	0.774295
PR	0.978453	**0.978882**	**0.978976**	0.978705	0.977817	0.978224	0.978896
RE	0.775344	0.77564	0.775015	0.779238	**0.794961**	**0.788026**	0.77606
F-s	0.865137	0.865489	0.865137	0.867655	**0.876959**	**0.872884**	0.865756
OPT	0.45709	0.464271	0.449221	0.449857	0.4773	0.455547	0.471767

TABLE 4.7: Results of the models estimation

Variable	Coefficient	Std.Error	t-Statistic	Probability
CONSTANT	0.0286	$6.1 * 10^{-4}$	46.732	1.000
Ratio1	-0.072	$6.1 * 10^{-4}$	-117.089	0.000
Ratio3	0.115	$5.7 * 10^{-4}$	-199.938	0.000
Ratio4	-0.247	$6.3 * 10^{-4}$	-388.607	0.000
Ratio5	-0.101	$5.7 * 10^{-4}$	-174.785	0.000
UA	-0.444	$6.2 * 10^{-4}$	100.539	0.000
Instrumented:	W-Rating			
Instruments:	W2-Ratio1, W2-UA, W3-Ratio5, W-Ratio4,	W2-Ratio3, W3-Ratio1, W3-UA, W-Ratio5,	W2-Ratio4, W3-Ratio3, W-Ratio1, W-UA	W2-Ratio5, W3-Ratio4, W-Ratio3,

TABLE 4.8: Results of the Z-score regression, estimation with OLS

been explained in Section 4.2.3 with theoretical examples, in the same manner we built a credit chain network from our transactions data and visualise it with a graph. The state of the local network of the selected firm in three consecutive months is illustrated with three graphs in Figure 4.7. For illustration we have selected a firm which has gone bankrupt and which does not have a large number of connections – otherwise, the graphs would become incomprehensive as the number of firms in the local network grows exponentially with the degree of the vertex. On the figures, the node labels are the Z-score values computed with the SLM-3 model, while the edge labels are the weights of the connectivity matrix.

Since the suppliers of the defaulted firm are more likely to be hit by a distress of its customer, for illustrative purposes we include in the visualisations only outgoing cash transactions from the defaulted firm. Thus, the graphs provide a partial view on the network of a firm. First level connections are the firms to which the defaulted firm transfers money – thus, they are the suppliers of the defaulted firm; second level connections are then the suppliers of the suppliers, etc. The illustrated graphs include the connections of the firm up to a third degree.

Figure 4.7a illustrates the state of a credit chain network of the firm which had a Z-score of 1.16 in November 2014. From the graph we can see that it has two suppliers – one with a Z-scores of 0.64 to which it is the only customer (since the weight of the edge is 100) and the other with a Z-score of -1.39 to which it is one of many customers (since the weight of the edge is very small – it is only 0.306). The last firm is also the only customer of the defaulted firm. We remind that our network by construction includes only connections with the total amount over each quarter exceeding a threshold and, therefore, represents only the most strong links between customers and suppliers.

Figure 4.7b illustrates the state of the same credit chain network in December 2014 when the considered firm had been hit by default – its rating increased from 13 in November to 18 in December. The defaulted firm is highlighted on the graph together with its 1st and 2nd degree connections. Comparing the network in November 2014 and December 2014, we can see from the graphs that the Z-score of the defaulted firm has increased from 1.16 to 1.24. The Z-scores of its suppliers have increased in the next month after the default has happened. This can bee seen in Figure 4.7c in comparison to 4.7b. The supplier to which the defaulted firm is the only customer has been hit mostly – its Z-score has increased from 0.5 in December 2014 to 0.89 in January 2015. The Z-score of the firm to which the defaulted firm is only one of many customers, has increased relatively smaller – from -1.24 in December 2014 to -1.08 in January 2015. This firm itself has 5 suppliers (excluding the defaulted firm). The Z-score of the supplier with the highest weight among the five of them has also increased (the Z-score of the firm connected with the weight of 7.6 has increased from -0.31 to -0.1).

The effects described above proof that the Z-score model with 3 lags captures the propagation of distress through the supply-chain network. Moreover, we observe that the distress indeed affects stronger the firms connected to the defaulted one (or its direct neighbours) with a higher weight. The former proofs that the connectivity matrix constructed from the cash transactions data reflects the actuals links between the firms. Therefore, suppose, a certain firm defaults in the current month, then the bank can use the constructed connectivity matrix to predict future decrease in rating of the firms which are most strongly connected to a defaulted one or its first and second degree connections. On the other hand, suppose, a firm which holds its accounts in the bank requests an extension of its disposable credit or a new credit line. Then, knowing the position of this firm in the supply-customer network –

in particular, with respect to the recently defaulted companies, would enhance its classical approach to credit assessment and support the bank in its lending decision policy.

4.5 Possible applications of the network

The weight matrix can be used in different ways – first, for enhancing the prediction of bunckrupcy as we have described in the experiments, and, second, for simulation. In this chapter we discuss possible applications of the network.

Simulation of the assets dynamics

After the network has been built different types of stress tests useful for risk management purposes can be performed. Further we discuss how simulations can be done.

Suppose each firm i, $i = 1, \ldots, n$ has a single zero coupon debt, with face value K_i which expires at maturity T as described above in the Merton model, then the dynamics of the assets are describe by the system of SDE in a matrix form as

$$d\overline{V_t} = diag[V_t](\overline{\mu}_t dt + \Sigma_t d\overline{W_t}), \tag{4.7}$$

where $\Sigma = SDS$ is a variance-covariance matrix with D being a correlation matrix and S being a diagonal matrix :

$$D = \begin{bmatrix} \sigma^{11} & \ldots & 0 \\ & \ldots & \\ 0 & \ldots & \sigma^{nn} \end{bmatrix}$$

The discrete version of the process for S_i is

$$V^i_{t+\delta t} = V^i_t + \mu^i V^i_t \delta t + \sigma^i V^i_t \epsilon^i \sqrt{\delta t}, \tag{4.8}$$

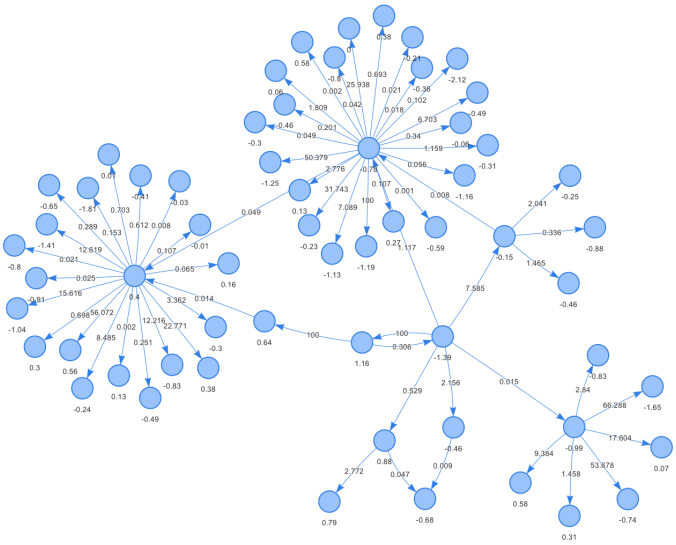
where ϵ^i is a random sample from a standard normal distribution with correlation ρ^{ij} between ϵ^i and ϵ^j.

The value of each asset V^i and the growth rate of the company μ^i we obtain from the financial data, as well as the value of the debt. The volatility of the asset is assumed to be proportional to the volatility of the cash flow going through node i. The Brownian motions at each time t are characterized by a set of instantaneous correlation coefficients incorporated in the correlation matrix. To simulate correlated Brownian motion we must generate standard normal variables which are correlated. We estimate the centrality c^i of each company on the weights matrix built from the cash transactions. Assuming that c is a normalised vector of centrality, such that $|c| = 1$, then the correlated random variables are simulated as

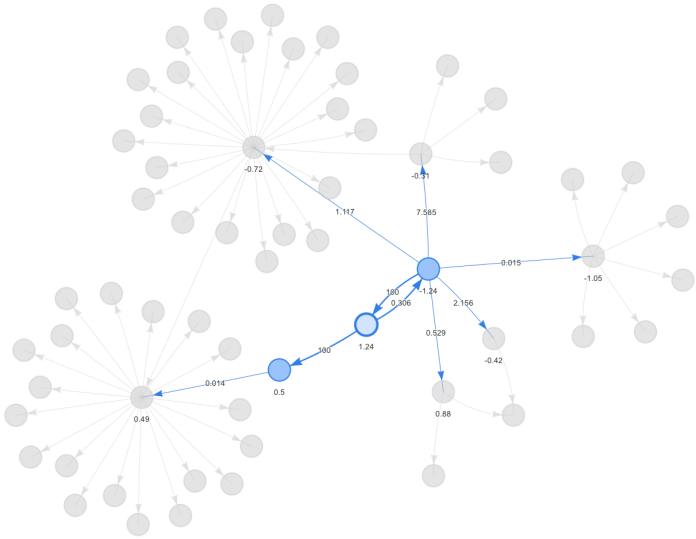
$$\epsilon^i = c^i \eta^i,$$

where $\bar{\eta} \sim iidN(0, I)$.

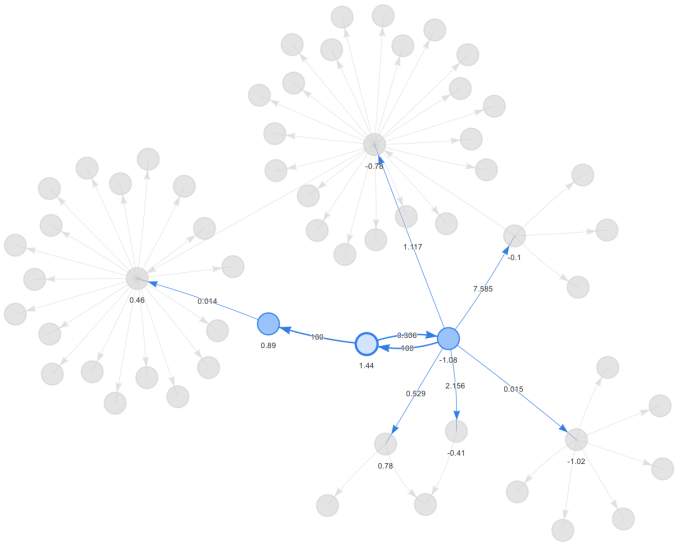
Since economic shocks that transmit across an economy do not have final recipients and are unlikely to follow the shortest path between industries, eigenvector centrality is the most appropriate measure for intersectoral trade networks according to [16]. Eigenvector centrality is calculated as the principal eigenvector of the network's adjacency matrix. Nodes are more central if they are connected to other nodes that are themselves more central.



(A) *Local network of a firm in November 2014*



(B) *Local network of a firm hit by a default in December 2014*



(C) *Local network of a firm after a default in January 2015*

FIGURE 4.7: *Illustration of the distress propagation though a credit chain network*

To test the hypothesis that firms with higher average level of transactions flow and lower level of volatility are more resistant to shocks to idiosyncratic and systemic shocks we simulate the shock propagation through the network and we check for the stability characteristics of the firms which would survive and the firms which would be destroyed by the shock.

Simulation of the shock propagation in the network

Simulations can be done in two conceptually different ways – taking a global/-macro view on the network or taking a local/micro view. By the macro view we mean the evaluation of how a shock on a single (or several) major client would propagate through the network and effect the default probabilities of all the other firms. This approach fits well with the reduced-form models. For a bank it could be useful to access the overall level of systemic risk in the portfolio. Simulating an attack on a client would provide a view on the resistance capacity or vulnerability of the portfolio to liquidity shocks. For performing such stress tests we would need to consider an entire graph with the prior estimated vector of the single-firm probabilities of default and define a function of the change in default probability as a function of a measure of centrality.

Economic shocks that transmit across an economy do not have final recipients and are unlikely to follow the shortest path between industries. In addition, economic shocks are likely to have feedback effects. A supply shock in one industry could affect the supply of downstream industries, which eventually could flow back to the original industry. For instance, an oil shock could affect the cost of gasoline, which affects transportation costs, which could then affect the oil industry. Thus, economic shocks are unlikely to be restricted to follow paths or trails, in which nodes and links are not repeated. In addition, network flow can spread through multiple paths at the same time.

Networks containing important nodes, or hubs, are called “scale-free" [10] in the sense that some hubs have a huge number of links and no node is typical of the other. However, scale-free networks have a common structure and properties to a certain extent. First, the probability that a node has exactly k links following a power-law distribution. Another is that they are remarkably robust against accidental failures, but extremely vulnerable to coordinated attacks. In application to the bank data, we can expect that an attack (which is equivalent to a stress-test performed in the financial domain) on a firm which is not strongly connected to other firms (such node would have a low corresponding Hub and Authority measures) would not effect many firms, while an attack on a hub will propagate through the network and effect a large number of firms.

An attack on a hub will spread the shock through the network [10]. However, the type of node and its properties should also be taken into account. In the financial world such hubs will most likely have a stronger financial position and have a better access to credit and other types of financing. This may have an opposite effect resulting in financial hubs absorbing liquidity shocks. The outcome of such an experiment – which effect would prevail (liquidity shock would destroy the firm or the company would absorb the shock) would depend on the intensity of the shock and the number of debtors affected, as well as on a general economical situation which would effect the credit conditions and capabilities of the company. Nevertheless, even if the company will survive, the shock will to a certain extend propagate in the network though the company’s trade counterparties – since the firm may postpone its own credit payments or reduce its production demands.

By the micro view, we mean a closer look at a certain client of the bank and the trade chain it is involved into. In a simple settings we assume that the value

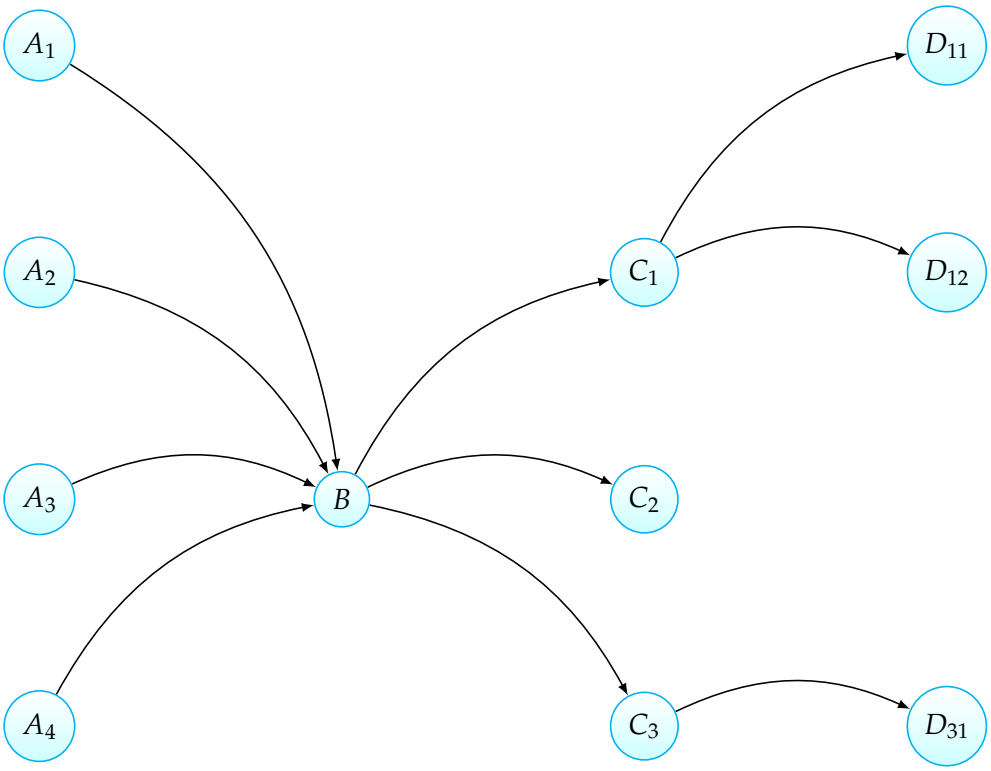


FIGURE 4.8: Illustration of a trade chain of the firm B

of each company at time t is V_t and that the balance sheet can be represented as in the Merton model. Assume a company B has $A_1, \ldots, A_n$ debtors and $C_1, \ldots, C_m$ creditors (as depicted in Figure 4.8). Then a liquidity shock on any of A_i would decrease its capability to repay the receivables due to B in time. This could potentially propagate to any of C_j and further down the trade chain. Depending on a scenario, one or several A_i can be affected by the shock. Then only a certain percentage of the receivables of firm B would be paid. This view is more in line with the structural models setup since it assumes that we have the information on the balance sheet of the firms. To perform the simulations technically we would only need a function which can reconstruct a local network up to a certain degree for a given node and fetch all the balance sheet information about those firms. Such assessment could be useful to support the bank in its lending decisions – when, for instance, a company B which is already a client of the bank is requesting to open a new credit.

4.6 Conclusions

In this work we have described how the network effect can be embedded into default prediction models both in the settings of reduced-form and structural approaches. Under the reduced-form approach we have used the methods of spatial econometrics and network analysis to allow the inclusion of the network effect in the classical Z-score model. The experiments performed show that the network effect improves the quality of predictions based on the AUC, R^2, F-score and Accuracy measures with respect to the general OLS approach. Although the increase in the measures is marginal – around 2%, the results proof the potential relevance of the network effect in the bunckrupcy prediction. Within the models with a spatial effect, we observe that the AUC measure increases as we extend the number of lags considered.

However, each additional lag provides less marginal improvement. The best models with respect to the AUC and R^2 are the SLM model with 3 lags and the combined SARAR model with 3 lags. The inclusion of the spatial error in addition to the SLM model does not improve the result for our dataset.

We have illustrated that the 3-lag SLM model captures the propagation of distress through a local credit chain network of a defaulted firm. We have discussed how the bank can use the constructed connectivity matrix to predict future decrease in rating or embed the information on the network position of a firm in its lending policy.

It is worth mentioning that the data quality and the process of its aggregation play a major role in the estimation of the models. In the current approach the firms have been chosen on the criteria of the amount of information which can be collected – both on a stand alone basis and based on the level of its connectivity in the network. While this approach allows to identify strong financial and trade relations between the firms, the constructed database is likely to be biased toward less risky firms since more risky companies may disclose less information. In particular, we observe that the default rate within the companies which appear in the network of payments is less than the default rate of the total sample of firms. Extending the database would require a more accurate treatment of missing values and a thorough analysis of the companies with less interactions.

To conclude, we think that use of BigData technologies offers a number of potential insights that can be profitably exploited to improve the probability of default prediction, and has the potential to improve the management of lending policies and the loan pricing.

Appendix A

Appendix to Chapter 2

A.1 Results of the estimation for the 4 datasets

Age in 1970	Wills and Sherris	Lee-Carter	OU-Process	Feller
60	0.0494	0.0612	0.0758	0.0560
59	0.0445	0.0634	0.0737	0.0506
58	0.0227	0.0739	0.0320	0.0222
57	0.0288	0.0941	0.0342	0.0257
56	0.0360	0.1148	0.0425	0.0315
55	0.0602	0.1601	0.0670	0.0506
54	0.0511	0.1746	0.0549	0.0423
53	0.0583	0.2275	0.0399	0.0340
52	0.0527	0.2031	0.0368	0.0344
51	0.0536	0.1455	0.0143	0.0188
50	0.0356	0.0676	0.0249	0.0228
49	0.0408	0.1205	0.0204	0.0178
48	0.0579	0.1350	0.0347	0.0306
47	0.0585	0.1369	0.0280	0.0260
46	0.0738	0.1086	0.0456	0.0410
45	0.0680	0.1169	0.0274	0.0278
44	0.0924	0.1065	0.0659	0.0598
43	0.0976	0.1291	0.0735	0.0614
42	0.1196	0.1404	0.1007	0.0904
Mean (Rank)	0.0580 (3)	0.1252 (4)	0.0470 (2)	0.0391 (1)

TABLE A.1: Mean (over 10 years) of the absolute errors for each generation, UK male data.

Year (Rank)	Wills and Sherris	Lee-Carter	OU-Process	Feller
2000	0.0318	0.0643	0.0283	0.0268
2001	0.0418	0.0790	0.0342	0.0312
2002	0.0354	0.0793	0.0268	0.0226
2003	0.0378	0.0735	0.0267	0.0271
2004	0.0621	0.1237	0.0485	0.0413
2005	0.0576	0.1248	0.0449	0.0356
2006	0.0759	0.1575	0.0611	0.0503
2007	0.0713	0.1646	0.0566	0.0449
2008	0.0632	0.1632	0.0494	0.0349
2009	0.1027	0.2224	0.0932	0.0768
Mean (Rank)	0.0580 (4)	0.1252 (1)	0.0470 (3)	0.0391 (2)

TABLE A.2: Mean (over 19 generations) of the absolute errors for each year, UK male data.

Age in 1970	Wills and Sherris	Lee-Carter	OU-Process	Feller
60	1.0000	0.1000	0.1000	0.5000
59	1.0000	0.1000	0.1000	0.6000
58	1.0000	0	0.6000	0.9000
57	1.0000	0	0.7000	1.0000
56	1.0000	0	0.3000	0.7000
55	1.0000	0	0	0.5000
54	1.0000	0	0.4000	0.7000
53	1.0000	0	0.6000	0.9000
52	1.0000	0	0.4000	0.7000
51	1.0000	0.1000	1.0000	1.0000
50	1.0000	0.4000	0.8000	0.9000
49	1.0000	0	0.7000	0.9000
48	1.0000	0	0.5000	1.0000
47	1.0000	0	0.5000	0.6000
46	0.9000	0	0.2000	0.6000
45	1.0000	0	0.6000	1.0000
44	0.8000	0.1000	0.1000	0.6000
43	0.5000	0	0	0.5000
42	0.1000	0	0	0
Mean (Rank)	0.9105 (1)	0.0421 (4)	0.4000 (3)	0.7158 (2)

TABLE A.3: Percentage of the actual mortality rates which falls within a 90% prediction interval, UK male data.

Age in 1970	Wills and Sherris	Lee-Carter	OU-Process	Feller
60	0.1215	0.0697	0.0746	0.0983
59	0.1113	0.0386	0.0461	0.0834
58	0.1191	0.0573	0.0347	0.0777
57	0.1201	0.0368	0.0372	0.0869
56	0.1405	0.0591	0.0561	0.1049
55	0.1270	0.0198	0.0570	0.1018
54	0.1637	0.0215	0.1123	0.1490
53	0.1533	0.0229	0.1158	0.1469
52	0.1713	0.0269	0.1402	0.1709
51	0.1736	0.0253	0.1633	0.1791
50	0.1955	0.1026	0.1802	0.2019
49	0.1438	0.0255	0.1269	0.1533
48	0.1434	0.0300	0.1030	0.1390
47	0.1622	0.0344	0.1735	0.1814
46	0.1551	0.0311	0.1665	0.1839
45	0.1227	0.0151	0.1331	0.1453
44	0.0986	0.0356	0.1044	0.1206
43	0.0905	0.0331	0.0833	0.1017
42	0.0850	0.0373	0.0787	0.0944
Mean (Rank)	0.1368 (3)	0.0380 (4)	0.1046 (1)	0.1326 (2)

TABLE A.4: Mean (over 10 years) of the absolute errors for each generation, Australian female data

Year (Rank)	Wills and Sherris	Lee-Carter	OU-Process	Feller
2000	0.0322	0.0400	0.0303	0.0308
2001	0.0436	0.0253	0.0354	0.0391
2002	0.0978	0.0394	0.0728	0.0887
2003	0.1041	0.0309	0.0710	0.0941
2004	0.1207	0.0316	0.0825	0.1110
2005	0.1284	0.0322	0.0877	0.1205
2006	0.1664	0.0358	0.1231	0.1612
2007	0.2025	0.0412	0.1588	0.2007
2008	0.2421	0.0606	0.1992	0.2441
2009	0.2296	0.0434	0.1849	0.2361
Mean (Rank)	0.1368 (4)	0.0380 (1)	0.1046 (2)	0.1326 (3)

TABLE A.5: Mean (over 19 generations) of the absolute errors for each year, Australian female data.

Age in 1970	**Wills and Sherris**	**Lee-Carter**	**OU-Process**	**Feller**
60	1.0000	0.4000	0.3000	0.4000
59	1.0000	0.7000	0.4000	0.4000
58	1.0000	0.4000	0.6000	0.4000
57	1.0000	0.7000	0.6000	0.5000
56	1.0000	0.6000	0.2000	0.2000
55	1.0000	1.0000	0.4000	0.3000
54	0.7000	1.0000	0.1000	0.2000
53	0.7000	0.9000	0.2000	0.2000
52	0.5000	1.0000	0.1000	0.1000
51	0.5000	1.0000	0.1000	0.2000
50	0.3000	0.4000	0	0.1000
49	0.7000	0.9000	0.1000	0.2000
48	0.6000	1.0000	0.2000	0.5000
47	0.6000	0.9000	0.2000	0.2000
46	0.6000	1.0000	0.1000	0.2000
45	0.8000	1.0000	0.2000	0.4000
44	0.8000	1.0000	0.4000	0.5000
43	1.0000	1.0000	0.7000	0.7000
42	1.0000	1.0000	0.7000	0.9000
Mean (Rank)	0.7789 (2)	0.8368 (1)	0.2947 (4)	0.3474 (3)

TABLE A.6: Percentage of the actual mortality rates which falls within a 90% prediction interval, Australian female data.

Age in 1970	Wills and Sherris	Lee-Carter	OU-Process	Feller
60	0.0940	0.0930	0.0801	0.0945
59	0.0439	0.0534	0.0386	0.0346
58	0.1171	0.0553	0.0980	0.1133
57	0.0879	0.0376	0.0586	0.0783
56	0.0974	0.0231	0.0657	0.0867
55	0.1238	0.0522	0.1385	0.1357
54	0.0949	0.0622	0.0805	0.0958
53	0.1144	0.0737	0.1163	0.1257
52	0.0928	0.0860	0.0817	0.0955
51	0.0900	0.0806	0.0880	0.0968
50	0.1338	0.0179	0.1543	0.1531
49	0.0725	0.1036	0.0916	0.0906
48	0.0693	0.0807	0.0541	0.0658
47	0.0581	0.0967	0.0578	0.0663
46	0.0499	0.0874	0.0562	0.0617
45	0.0460	0.0784	0.0458	0.0500
44	0.0387	0.0860	0.0470	0.0486
43	0.0356	0.1111	0.0628	0.0591
42	0.0357	0.0742	0.0422	0.0432
Mean (Rank)	0.0787 (3)	0.0712 (1)	0.0767 (2)	0.0840 (4)

TABLE A.7: Mean (over 10 years) of the absolute errors for each generation, Australian male data.

Year (Rank)	Wills and Sherris	Lee-Carter	OU-Process	Feller
2000	0.0279	0.0365	0.0287	0.0285
2001	0.0362	0.0516	0.0370	0.0367
2002	0.0623	0.0404	0.0590	0.0615
2003	0.0536	0.0609	0.0521	0.0525
2004	0.0615	0.0597	0.0580	0.0640
2005	0.0561	0.0970	0.0478	0.0578
2006	0.0858	0.0913	0.0825	0.0926
2007	0.1132	0.0866	0.1076	0.1239
2008	0.1526	0.0814	0.1497	0.1652
2009	0.1380	0.1067	0.1448	0.1571
Mean (Rank)	0.0787 (3)	0.0712 (1)	0.0767 (2)	0.0840 (4)

TABLE A.8: Mean (over 19 generations) of the absolute errors for each year, Australian male data.

Age in 1970	**Wills and Sherris**	**Lee-Carter**	**OU-process**	**Feller**
60	1.0000	0.2000	0.4000	0.6000
59	1.0000	0.4000	0.7000	1.0000
58	1.0000	0.2000	0.2000	0.3000
57	1.0000	0.3000	0.4000	0.8000
56	1.0000	0.6000	0.4000	0.5000
55	1.0000	0.1000	0	0
54	1.0000	0.2000	0.4000	0.5000
53	1.0000	0.1000	0.2000	0.3000
52	1.0000	0	0.3000	0.4000
51	1.0000	0.2000	0.3000	0.3000
50	1.0000	1.0000	0	0.2000
49	1.0000	0	0.4000	0.7000
48	1.0000	0.2000	0.6000	0.8000
47	1.0000	0.1000	0.4000	0.6000
46	1.0000	0.5000	0.7000	0.7000
45	1.0000	0.6000	0.7000	0.8000
44	1.0000	0.4000	0.7000	0.8000
43	1.0000	0.1000	0.7000	0.7000
42	1.0000	0.6000	0.9000	0.9000
Mean (Rank)	1.0000 (1)	0.3053 (4)	0.4421 (3)	0.5737 (2)

TABLE A.9: Percentage of the actual mortality rates which falls within a 90% prediction interval, Australian male data.

Appendix B

Appendix to Chapter 3

B.1 Market price of risk in affine processes

Broadly, a stochastic volatility model for a stock price is defined as a model in which the price is a function of a vector of state variable X_t that follows a multivariate diffusion process. Here we state the definition of the **affine model** as specified in [28].

1. Assume the state variable X_t under the objective probability measure P follows the diffusion process:

$$dX_t = \mu^P(X_t)dt + \sigma(X_t)dW_t^P, \tag{B.1}$$

   Here X_t is an N-vector of state variables, dW_t^P is an N-dimensional Brownian motion under the objective measure P, $\mu^P(X_t)$ is an N-dimensional function of X_t and $\sigma(\cdot)$ is an $N \times N$ matrix-valued function of X_t.

2. The instantaneous drift (under the measure P) of each state variable is an affine function of X_t:

$$\mu^P(X_t) = a^P + b^P X_t,$$

   for some N-vector a^P and some $N \times N$ matrix b^P.

3. The instantaneous covariance between any pair of state variables is an affine function of X_t

$$[\sigma(X_t)\sigma^T(X_t)]_{i,j} = \alpha_{ij} + \beta_{ij}^T X_t,$$

   where $[\sigma(X_t)\sigma^T(X_t)]_{i,j}$ denotes the element in row i and column j of the product $\sigma(X_t)\sigma^T(X_t)$, α_{ij} is a constant and β_{ij}^T is an N-vector for each $1 < i, j \leq N$.

4. There exists a probability measure Q, equivalent to P, such that X_t is a diffusion under Q:

$$dX_t = \mu^Q(X_t)dt + \sigma(X_t)dW_t^Q, \tag{B.2}$$

   where dW_t^Q is an N-dimensional Brownian motion under the martingale measure Q which guarantees the absence of arbitrage, $\sigma(\cdot)$ is an $N \times N$ matrix-valued function of X_t and the drift of each state variable is an affine function of the state vector

$$\mu^Q(X_t) = a^Q + b^Q X_t,$$

   for some N-vector a^Q and some $N \times N$ matrix b^Q.

Moving from the real-world measure to an equivalent martingale measure is achieved by Girsanov's Theorem:

Given a specification of the $\mu^P(X_t)$ and $\sigma(X_t)$ such that a solution to the equation (B.1) exists, an equivalent martingale measure Q is defined as:

$$Q = \exp\left(-\int_0^T \Lambda^T(X_u) dW_u^P - \frac{1}{2}\int_0^T \Lambda^T(X_u)\Lambda^T(X_u) du \right) P$$

by specifying a market price of risk process $\Lambda(X_t)$ that satisfies the condition:

$$E^P\left[-\int_0^T \Lambda^T(X_u) dW_u^P - \frac{1}{2}\int_0^T \Lambda^T(X_u)\Lambda^T(X_u) du \right] = 1$$

It follows from Girsanov's theorem that the process $W_t^Q = W_t^P + \int_0^t \Lambda(X_s) ds$ is an N-dimensional Brownian motion under Q, and the drift in the diffusion (B.2) is given by:

$$\mu^Q(X_t) = \mu^P(X_t) - \sigma(X_t)\Lambda(X_t).$$

Thus, affine diffusion models allow to specify the behaviour of the state variable(s) under both an objective probability measure and an equivalent martingale measure with the drift and the diffusion coefficients being affine functions of the state variable itself under both measures. The market price of risk is then defined as in [28]:

$$\Lambda_t = [\sigma(X_t)]^{-1}\left[\mu^P(X_t) - \mu^Q(X_t)\right]$$

The characteristic function of an affine diffusion process has also an affine form as stated in equation (3.13). The dimensionality of the ODE system (3.14) corresponds to the dimensionality of the state vector X_t.

B.2 Carr and Madan Inversion

Equations (3.10), (3.13) and (3.17) give a solution for european call options. However, numeric integration in equation (3.17) can be difficult as the characteristic function of the option is singular at the origin. Carr and Madan (1998) [27] present an alternative Fourier transform procedure that avoids this difficulty.

Due to (3.9) we can write the value of a European call as a function of $k = lnK$:

$$C_T(k) = e^{-rT}\int_k^\infty (e^{s_T} - e^k) q_T(s) ds_T, \tag{B.3}$$

where $q_T(s)$ is the risk-neutral density function of s_t. The characteristic function of s_T has a form $f_T(u) = E[e^{s_T iu}]$. Thus, the characteristic function of a density is defined as

$$f_T(\phi) = \int_{-\infty}^\infty e^{i\phi s} q_T(s) ds, \tag{B.4}$$

The call function in equation (B.3) is not square-integrable. To obtain a square-integrable function, Carr and Madan (1998) consider a modified call price $c_T(k)$ with a damping factor

$$c_T(k) = e^{\alpha k} C_T(k). \tag{B.5}$$

For $\alpha > 0$ this function will be square-integrable on the entire real line.

Consider now the Fourier transform of $c_T(k)$:

$$\psi_T(u) = \int_{-\infty}^{\infty} e^{ius} c_T(k) dk, \tag{B.6}$$

The call prices are obtained numerically using the inverse transform and the analytical expression for the ψ:

$$C_T(k) = \frac{e^{-\alpha k}}{\pi} \int_0^{\infty} e^{-iuk} \psi_T(u) du, \tag{B.7}$$

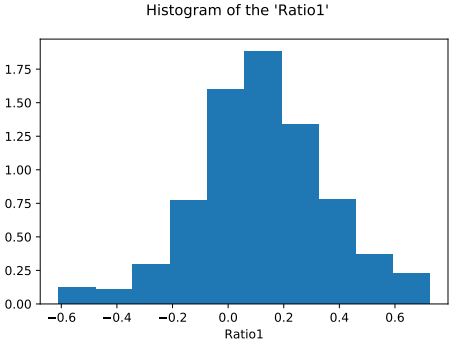
$$\psi_T(u) = \frac{e^{rT} \phi_T(u - (\alpha + 1)i)}{\alpha^2 + \alpha - u^2 + i(2\alpha + 1)u} \tag{B.8}$$

The values are obtained by substituting (B.8) into (B.7) and performing the required integration. The integral is numerically computed using the trapezoid rule and the Simpson's weighting is applied for the summation. For more details see Carr and Madan (1998) [27].

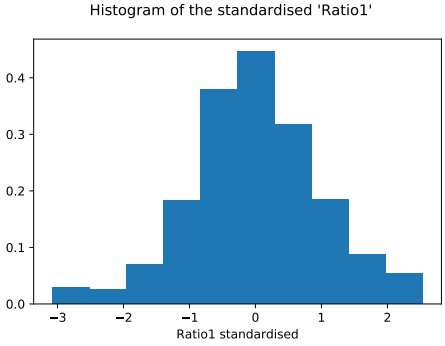
Appendix C

Appendix to Chapter 4

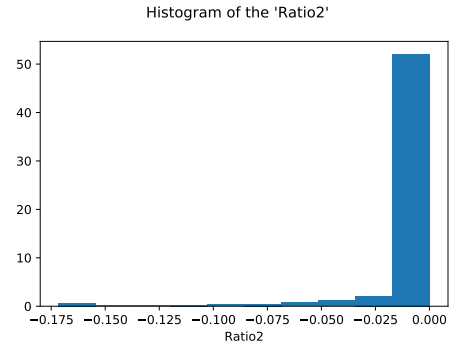
C.1 Histograms of the independent variables



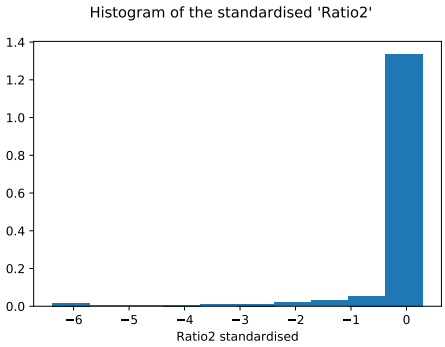
(A) *Ratio1*



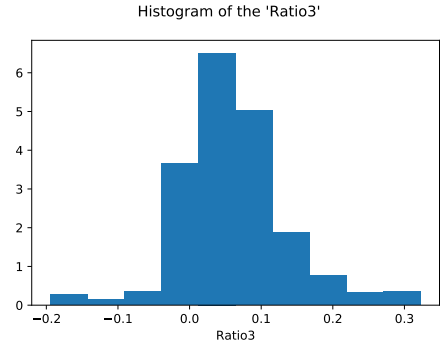
(B) *Ratio1 standartised*



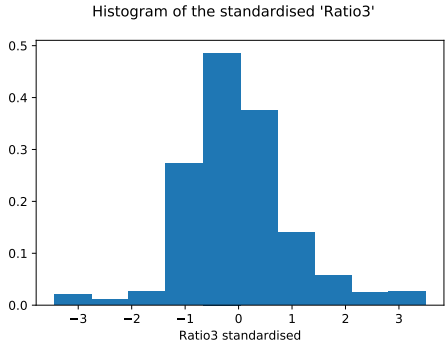
(C) *Ratio2*



(D) *Ratio2 standartised*

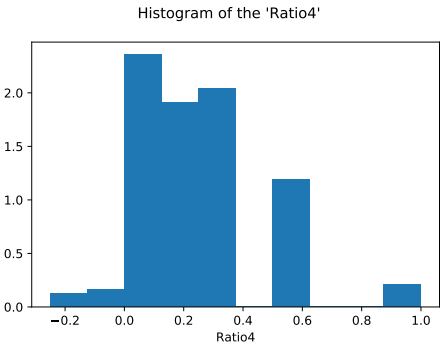


(A) *Ratio3*

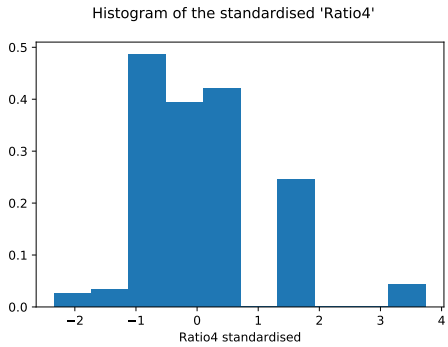


(B) *Ratio3 standartised*

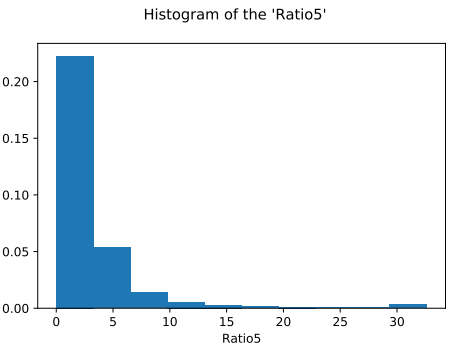
FIGURE C.2: *Histograms of the Z-score variables*



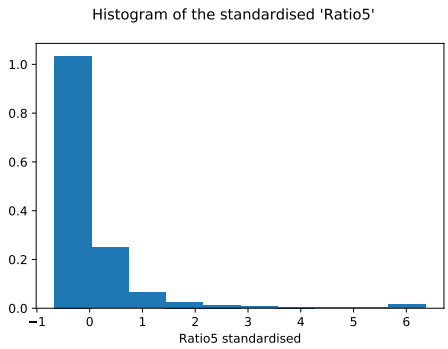
(A) *Ratio4*



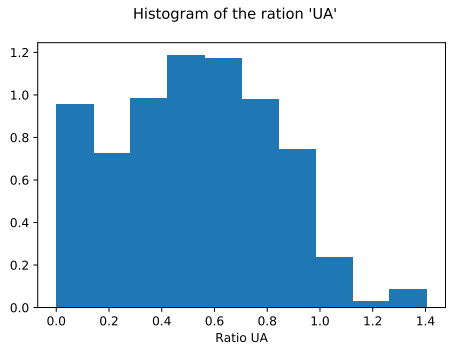
(B) *Ratio4 standartised*



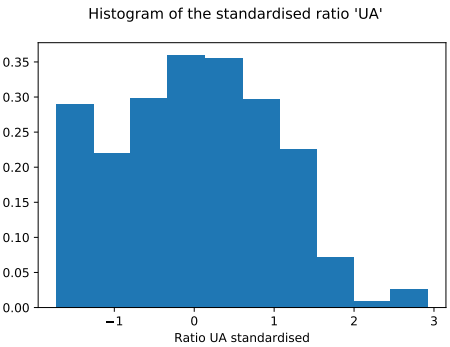
(C) *Ratio5*



(D) *Ratio5 standartised*



(E) *Ratio UA*



(F) *Ratio UA standartised*

FIGURE C.3: *Histograms of the Z-score variables*

References

[1] Ahern K.R. Network Centrality and the Cross Section of Stock Returns. *SSRN Electronic Journal* **2013**.

[2] Ahn, C. Y. Financial and corporate sector restricting in South Korea: accomplishments and unfinished agenda. *Japanese Economic Review* **2001**, *52*, 452-470.

[3] Ait-Sahalia Y., Kimmel R. Maximum likelihood estimation of stochastic volatility models. *Journal of Financial Economics***2007**, *83*, 413-452

[4] Ait-Sahalia Y. Closed-form likelihood expansions for multivariate diffusions. Working paper, Princeton University **2001**

[5] Altman, E.I Financial ratios, discriminant analysis and the prediction of corporate failure, *Journal of Finance* **1968**, *23*, 589-609.

[6] Altman E.I. Corporate Financial Distress. *Wiley Interscience* **1983**.

[7] Altman, E. and Narayanan, P. An International Survey of Business Failure Classification Models. *Financial Markets, Institutions and Instruments* **1997**, *6*,1-57.

[8] Arbia G. A Primer for Spatial Econometrics, Palgrave Macmillan **2014**.

[9] Azizpour S., Giesecke K., Schwenkler G., (2015) Exploring the sources of default clustering, preprint Stanford University.

[10] Barabasi A-L. and Bonabeau E. Scale Free Networks. *Scientific American* **2003**, 50-59.

[11] Barucca P., Bardoscia M., Caccioli F., D'Errico M. Network valuation in financial systems. *Cornell University Library* **2016**.

[12] Billio M., Getmansky M., Lo A. W., Pelizzon L. Econometric measures of connectedness and systemic risk in the finance and insurance sectors. *Journal of Financial Economics* **2012**, *104*, 535-559.

[13] Biffis E. Affine processes for dynamic mortality and actuarial valuations. *Insurance: Mathematics and Economics* **2005**, *37*, 443-468.

[14] Blake, D. and W. Burrows. Survivor bonds: helping to hedge mortality risk. *The Journal of Risk and Insurance* **2001**, *68*, 339-348.

[15] Black F. and Scholes M. The pricing of options and corporate liabilities. *Journal of Political Economy* **1973**, *81*, 637-654.

[16] Borgatti S.P. Centrality and network flow. *Social Networks* **2005**, *27*, 55-71.

[17] Boissay F. and Gropp R. Payment defaults and interfirm liquidity provisions. *Review of Finance* **2012**, *17*, 1853-1894.

[18] Boissay F. Credit chains and the propagation of financial distress. *ECB working Paper Series* **2006**, *573*.

[19] Bonacich, P. Power and centrality: A family of measures. *American Journal of Sociology* **1987**, *92*, 1170-1182.

[20] Boritz, J., Kennedy, D. and Sun, J. Predicting Business Failures in Canada La PréDiction des Faillites D'entreprise au Canada. *Accounting Perspectives* **2007**, *6*, 141- 165.

[21] Brin, S., Page, L. The anatomy of a large-scale hypertextual web search engine. *Computer Networks and ISDN Systems* **1998**, *30*, 107-117.

[22] B.C. on Banking Supervision. International convergence of capital measurement and capital standards: a revised framework. *Bank for International Settlements* **2006**.

[23] Campbell, J. Y., Hilscher, J., Szilagyi, J. In search of distress risk. *The Journal of Finance* **2008**, *63*, 2899-2939.

[24] 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. *North American Actuarial Journal* **2009**, *13*, 1-35.

[25] Cairns A. J. G., Blake D., and Dowd K. Pricing death: Framework for the valuation and securitization of mortality risk. *ASTIN Bulletin* **2005**, *36*, 79-120.

[26] Cairns A. J. G., Blake D., and Dowd K. A two-factor model for stochastic mortality with parameter uncertainty: theory and calibration. *The Journal of Risk and Insurance* **2006**, *73*, 687-718.

[27] Carr P., Madan D. B. Option valuation using the fast Fourier Transform. *Journal of Computational Finance* **1998**, *2*, 61-73.

[28] Cheridito P., Filipovic D., Kimmel R. L. Market price of risk specification for affine models: theory and evidence. Journal of Financial Economics **2007**, *83*, 123-170.

[29] Chernov, M. and Ghysels, E. Estimation of the Stochastic Volatility Models for the Purpose of Options Valuation, *Computational Finance - Proceedings of the Sixth International Conference*, MIT Press, Cambridge **2000**.

[30] Cont, R. Empirical properties of asset returns: stylized facts and statistical issues. *Quantitative Finance* **2000**, *1*, 223-236.

[31] Cox J.C., Ingersoll J.E and Ross S.A. A Theory of the Term Structure of Interest Rates. *Econometrica* **1985**, *53* .

[32] Dahl, M. Stochastic mortality in life insurance: market reserves and mortality-linked insurance contracts. *Insurance: Mathematics and Economics* **2004**, *35*, 113-136.

[33] Dai, Q., Singleton, K.J. Specification analysis of affine term structure models. Journal of finance **2000**, *55*, 1943-1978.

[34] Denuit, M. and Devolder, P. Continuous time stochastic mortality and securitization of longevity risk. Institut des Sciences Actuarielles, Universite' Catholique de Louvain, Louvain-la-Neuve Working Paper 06-02, *2006*.

[35] Dowd K., Cairns A.J.G., Blake D., Coughlan G. D., Epstein D., Khalaf-Allah M. Evaluating the goodness of fit of stochastic mortality models. *Insurance: Mathematics and Economics* **2010**, *47*, 255-265.

[36] Dowd K., Cairns A.J.G., Blake D., Coughlan G.D., Epstein D., and Khalaf-Allah, M. Backtesting stochastic mortality models: an ex-post evaluation of multi-period-ahead density forecasts. *North American Actuarial Journal* **2010**, *14*, 281-298.

[37] Duffie, D., Filipović, D., Schachermayer, W. Affine processes and applications in finance. *Annals of applied probability* **2003**, *13*, 984-1053.

[38] Duffie, D. and Singleton, K. Modeling Term Structures of Defaultable Bonds. *Review of Financial Studies* **1999**, *12*, 197-226.

[39] Dupire B. Pricing with a smile, Risk **1994**, *7*, 18-20.

[40] Girosi F., King G Understanding the Lee-Carter Mortality Forecasting Method **2007**.

[41] Fang F., Oosterlee C.W. A Novel Pricing Method for European Options Based on Fourier-Cosine Series Expansions. *SIAM Journal on Scientific Computing* **2008**, *31*, 826-848.

[42] Feldhutter P. and Schaefer, S. The Myth of the Credit Spread Puzzle, preprint **2016**.

[43] Fonseca J. D., Grasselli M., Tebaldi C.Option pricing when correlations are stochastic: an analytical framework, *Review of Derivatives Research* **2007**, *10*, 151-180.

[44] Gao, J. The Effects of Firm Network on Banks Portfolio Consideration. *SSRN Electronic Journal* **2015**.

[45] Giesecke K. Credit modelling and valuation: an introduction. *SSRN Electronic Journal* **2004**, *2*, 1-67.

[46] Giacometti R, Bertocchi M., Rachev S.T., Fabozzi F.J. A comparison of the Lee-Carter model and AR-ARCH model for forecasting mortality rates **2009**.

[47] Glasserman P. Monte Carlo methods in financial engineering. *Springer* **2003**.

[48] Grzelak L., Oosterlee K. On the Heston model with stochastic interest rates. *Munich Personal RePEc Archive* **2010**, *20620*.

[49] Harrison J.M., Kreps D.M. Martingale and arbitrage in multiperiod securities markets. *Journal of Economics* **1979**, *20*, 381-408.

[50] Harrison J.M., Pliska S.R. Martingales and stochastic integrals in the theory of continuous trading. *Stochastic Processes and their Applications* **1981**, *11*, 215-260.

[51] Heston S.L. A closed-form solution for options with stochastic volatility with applications to bond and currency options. *A Review of Financial Studies* **1993**, 327-343.

[52] Hibbert H., Manning B. Real-world equity volatility assumptions. Moody's Analytics Research, Assumption update **2014**.

[53] Hout K, Bierkens J, Ploeg A. A semi closed-form analytic pricing formula for call options in a hybrid Heston-Hull-While model. *Proceedings of the 58th European Study Group Mathematics with Industry* **2007**.

[54] Hull J., White A. Using Hull-White interest rate trees. *Journal of Derivatives* **1996**, *4*, 26-36.

[55] Hull J., White A. The pricing of options on assets with stochastic volatilities. *Journal of Finance* **1987**, *42*, 281-300.

[56] Ingber, A. L. Adaptive simulated annealing (ASA): Lessons learned. *Control and Cybernetics* **1995**, 1-27.

[57] Jacobson T. and von Schvedin S. Trade credit and the propagation of corporate failure: an american analysis. *Econometrica* **2015**, *83*, 1315-1371.

[58] Jarrow R. and P. Protter P. Structural versus reduced-form models: A new information based perspective. *Journal of Investment Management* **2004**, *2*, 34-43.

[59] Jarrow, R. and Turnbull, S. Credit Risk: Drawing the Analogy. *Risk Magazine* **1992**, *5*.

[60] Jeanblanc M., Yor M., Chesney M. Mathematical methods for financial markets. *Springer Finance* **2009**.

[61] Jevtic P., Luciano E., Vigna E. Mortality surface by means of continuous time cohort models. *Insurance: Mathematics and Economics* **2013**, *53*, 122-133

[62] Kalemli-Ozcan S., Kim S-J, Shin H.S., Sorensen B. E., Yesiltas S. Financial Shocks in Production Chains. Princeton University preprint **2013**.

[63] Kelejian, H.H. and Prucha, I. A Generalized Spatial Two-Stage Least Squares Procedure for Estimating a Spatial Autoregressive Model with Autoregressive Disturbances. *Journal of Real Estate Finance and Economics* **1998**, *17*, 99-121.

[64] Kiyotaki, N. and MooreJ. Credit Chains. *Discussion Paper Series of the Edinburgh School of Economics* **1997**, *118*.

[65] Kladivko K. Maximum likelihood estimation of the Cox-Ingersoll-Ross process: the Matlab implementation. Working paper of the Technical Computing Conference, Prague **2007**.

[66] Kleinberg, J. M. Authoritative sources in a hyperlinked environment. *Journal of the ACM* **1999**, *46*, 604-632.

[67] Kleindorfer P. R., Wind Y. J. The network challenge. Strategy, profit and risk in an interlinked world *Wharton School Publishing* **2009**.

[68] Klibi, Walid K., Alain M. Scenario-based Supply Chain Network risk modeling. *European Journal of Operational Research* **2012**, *223*, 644-658.

[69] Kuersteiner G.M. and Prucha R. Dynamic Spatial Panel Model: Networks, Common Shocks, and Sequential Exogeneity. *Cornell University Library* **2018**.

[70] Lando D. Credit risk modelling. *Princeton University Press* **2004**.

[71] Levratto, N. From failure to corporate bankruptcy: a review. *Journal of Innovation and Entrepreneurship* **2013**, *2*, 20-42.

[72] Lee R. Option pricing by transform methods: extensions, unification, and error control. *Journal of Computational Finance* **2004**, *7(3)*: 51-86.

[73] Lee R.D., Carter, L.R. Modelling and forecasting U.S. mortality. *Journal of the American Statistical Association* **1992**, *87*, 659-671.

[74] Lee R.D., Miller T. Evaluating the performance of the Lee-Carter method for forecasting mortality. *Demography* **2001**, *38*, 537-549.

[75] Lewis A. Option valuation under stochastic volatility. *Finance Press Newport Beach* **2001**.

[76] Luciano E., Regis L., Vigna E. Delta-Gamma hedging of mortality and interest rate risk. *Insurance: Mathematics and Economics* **2012**, *50*, 402-412.

[77] Luciano E., Vigna E. Mortality risk via affine stochastic intensities: calibration and empirical relevance. *Belgian Actuarial Bulletin* **2008**, *8*.

[78] Merton R. C. On the pricing of corporate debt: the risk structure of interest rates. *Finance* **1974**, *29*, 449-70.

[79] Milgram S. The Small World Problem. *Psychology Today* **1967**, *2*, 60-67.

[80] Milevsky M.A. and Promislow. Mortality Derivatives and the Option to Annuitize. *Insurance: Mathematics and Economics* **2001**, 29, 299-318.

[81] Newman M. The physics of networks. *Physics today* **2008**, 33-38.

[82] Newman M. Power laws, Pareto distributions and Zipf's law. *Contemporary Physics* **2005**, *46*, 323-351.

[83] Ngai N. and Sherris M. Longevity risk management for life and variable annuities: The effectiveness of static hedging using longevity bonds and derivatives. *Insurance: Mathematics and Economics* **2011**, 49, 100-114.

[84] Ohlson, J. A. Financial ratios and the probabilistic prediction of bankruptcy. *Journal of accounting research* **1980**, 109-131.

[85] O'Hare C., Li Y. Explaining young mortality. *Insurance: Mathematics and Economics* **2012**, 50, 12-25.

[86] Plat R. On stochastic mortality modelling. *Insurance: Mathematics and Economics* **2009**, 45, 393-404.

[87] Renshaw, A.E., and Haberman, S. A cohort-based extension to the Lee- Carter model for mortality reduction factors. *Insurance: Mathematics and Economics* **2006**, *38*, 556-570.

[88] Rouah F.D., Heston S.L. The Heston model and its extensions in Matlab and *C#*. *Wiley Finance Series* **2013**.

[89] Schrager, D. Affine stochastic mortality. *Insurance. Mathematics and Economics* **2006**, *38*, 81-97.

[90] Schönsleben, P. Integral Logistics Management. *Auerbach Publications* **2007**.

[91] Stein E.M., Stein J.C. Stock price distributions with stochastic volatility: an analytical approach. *Review of Financial Studies* **1993**, *4*, 727-752.

[92] Wang, Y., Campbell, M. Financial ratios and the prediction of bankruptcy: the Ohlson model applied to Chinese publicly traded companies. *The Journal of Organizational Leadership and Business* **2010**, 1-15.

[93] Weron, R., and Wystup. U. Heston's model and the smile. Statistical Tools for Finance and Insurance. *Springer* **2005**. bibitemWesserman Wesserman S., Faust K. Social network analysis: methods and applications. *Cambridge University Press* **1994**. White, H., 1994, Estimation, Inference and Specification Analysis, Cambridge University Press, New York.

[94] Wills S., Sherris, M. Securitization, Structuring and Pricing of Longevity Risk. *Insurance Math. Econom.* **2010** *46*, 173-185.

[95] Wills S., Sherris M. Integrating Financial and Demographic Longevity Risk Models: An Australian Model for Financial Applications. UNSW Australian School of Business Research Paper No. 2008ACTL05 **2011**.

[96] Wu L. Centrality of the supply chain network. Working paper of the Chicago Booth School of Business **2015**.

UNIVERSITÀ DEGLI STUDI DI TORINO

n. file contenuti: 1
spazio su disco: 700 MB

(Spazio riservato all'Ufficio:)

prot. n. _____ del ___ /___ /____