Netspar

Network for Studies on Pensions, Aging and Retirement

NETSPAR ACADEMIC SERIES

Parameter Uncertainty In The Lee-Carter Model

Koen Steeghs

MSc 08/2020-014

Parameter Uncertainty In The Lee-Carter Model

A thesis submitted to
Maastricht University
School of Business and Economics

in partial fulfillment of the requirements
for the

Master's Degree

in

Econometrics & Operations Research

by

Koen Steeghs
Matriculation number: i6140387

Supervised by
Prof. Dr. Antoon Pelsser

August 14, 2020

Acknowledgement

I would like to start with expressing my deepest gratitude to Prof. Dr. Antoon Pelsser who has guided me throughout the journey of writing my master thesis. I want to thank him for sharing part of his invaluable knowledge, and for the suggestion of investigating the parameter uncertainty in the Lee-Carter model. In addition, I want to extend my sincere thanks to Marie Ternes and David Paulissen for reviewing my work on linguistic matters. Lastly, I would like to acknowledge the unwavering support and encouragement of my friends and family members along the way of this demanding project.

Abstract

This paper investigates the parameter uncertainty in the Lee-Carter model, which is a two-step procedure for estimating and forecasting mortality rates and hence life expectancies (Lee and Carter, 1992). Although the Lee-Carter model is frequently applied, actuaries often tend to ignore the uncertainty that arises when estimating the model parameters. More specifically, when executing the two-stage procedure of the Lee-Carter model for estimating and forecasting mortality, the estimated κ_t index which represents the improvement in mortality over the years is often regarded as if it is a known quantity. Yet, this simplification does not represent the fact that the κ values are prone to estimation uncertainty. Neglecting the parameter uncertainty may have consequences for risk management, especially under the Solvency II legislation. Therefore, this paper proposes a simulation study to investigate the true underlying parameter uncertainty in the Lee-Carter model, and focuses in particular on the identification of the κ_t process. The simulation study enabled us to link the parameter uncertainty problem that arises when using the Lee-Carter model with the theory on errors-in-variables models, as these are the regression models that elucidate the measurement errors in the independent variables. The results in all simulation settings indicate that there is a noticeable estimation error. In two out of four simulation cases, the estimation error did influence the time series analysis procedure for identifying the right model specification of the κ_t index.

Keywords: Mortality, Lee-Carter model, SVD, Parameter uncertainty, Time series analysis, ARIMA, Stochastic mortality modelling, Simulation study, Errors-in-variables models.

Contents

1 Introduction

The Solvency II legislation that is being applied in Europe since January 2016, serves as a framework for prudential supervision of reinsurance and insurance companies. This new supervisory framework which replaces the former Solvency I aims to protect policyholders by laying out several rules to which (re)insurers have to adhere. In order to reduce the risk of insolvency, the legislation for instance demands that an insurer should hold enough buffer capital such that it can meet all of its obligations over a one-year horizon with a probability of at least 99.5%. In the Netherlands, insurers failing to comply with the Solvency II rules will be confronted with formal recovery actions enforced by De Nederlandsche Bank (DNB).

This new solvency regime leads to a more pragmatic approach in assessing and modelling all types of risks to which insurance companies are exposed. One of the risks insurance companies frequently face is the non-diversifiable longevity trend risk. Longevity trend risk represents the uncertainty over the future path of mortality rates for policyholders. In addition, it is generally acknowledged that the management of annuity and pension portfolios is affected by this uncertainty over the future mortality trends. As a consequence, the prediction of future mortality rates became a vital part of assessing life insurance portfolios and pension plans.

Mortality forecasts are regularly made with the aid of the famous Lee-Carter model introduced by Lee and Carter in 1992. The use of stochastic mortality models and their projections as instruments to study the uncertainty over longevity trend risk is even encouraged under the Solvency II regime as described in the paper of Börger (2010). However, there exists a large spectrum of other available models. Some popular models that are used by actuaries are summarized in the paper of Cairns et al. (2009), who compared the performance of the Lee-Carter model with seven other stochastic mortality models. The conclusion drawn is that there is no model that stands out. Moreover, Richards and Currie (2009) demonstrated that a different choice of model within the Lee-Carter family applied to the same data set can change the direction of mortality projections. This in turn affects the best-estimate reserves, which brings us to another risk insurers face, known as model (selection) risk. In order to reduce the model risk, it is crucial that insurers and analysts use more than one stochastic projection model when evaluating longevity trend risk.

Kleinow and Richards (2017) addressed this facet by comparing the projections arising from a random walk model and various other auto regressive integrated moving average (hereafter, referred to as ARIMA) models. They managed to decompose the overall uncertainty over future mortality rates, which can be measured as the variance of mortality forecast values, into two parts. The two components of uncertainty are described as the uncertainty over the trend, and the year-on-year fluctuations referred to as temporary volatility. The Lee-Carter model has been used to demonstrate this decomposition of the uncertainty over the future mortality rates.

However, in the paper of Kleinow and Richards (2017) it is assumed that the trend risk is equivalent with the uncertainty over the model parameters. Supplementary, the estimated κ values of the Lee-Carter model which represent the improvement in mortality over the years, are regarded

as if they are known quantities. Yet, this simplification does not represent the fact that the κ values are themselves prone to estimation uncertainty. In reality, there exists uncertainty over the true underlying κ values. Kleinow and Richards (2017) indeed notice this and state that the true κ values can be considered as a hidden process. This is because the κ values are not directly visible and one can only deduce possible values given the random variation from actualized deaths in a confined population. In case the κ_t index follows from a mortality data set that is based on a population which is tiny in size, then the uncertainty is even larger.

Ignoring the uncertainty of the parameter estimates may have an impact on the risk management. To this end, we try to relax the simplifications and propose a simulation study that investigates the true underlying parameter uncertainty that arises when estimating the parameters in the the Lee-Carter model. In particular, it focuses on the identification of the κ_t process. In addition, the simulation study helps to identify the impact the parameter uncertainty may have on the prediction uncertainty and thus risk management.

In the simulation study, several true but hidden κ_t processes are simulated, such as a κ_t process that follows from a random walk without or with drift. This made it possible to construct the central death rates that are modelled by the Lee-Carter model in a reverse manner. The simulation procedure enabled us to link the parameter uncertainty problem that arises when using the Lee-Carter model with the theory on errors-in-variables models. This is due to the fact that the Lee-Carter model can be seen as a regression setting with no observable quantities on the right hand side of the model, and errors-in-variables models are the regression models that elucidate the measurement errors in the independent variables.

The results of the simulation study indicate that there is a noticeable estimation error. Moreover, it is highlighted how the parameter uncertainty affects the time series techniques for selecting the right ARIMA model. In two simulation settings, the estimation error did not seem to influence the time series analysis procedure for identifying the right ARIMA process much. However, in the other settings it did influence the time series analysis procedure for identifying the right ARIMA process.

1.1 Overview

The remainder of this dissertation is structured as follows. Section 2 reviews the literature. It provides an overview of mortality modelling with the Lee-Carter model in general. Section 3 lays out the methodology. It explains the Lee-Carter model, the Singular Value Decomposition, and how the κ_t index is usually modelled as a random walk. Section 4 serves as an application of the Lee-Carter model where we still regard the κ values as if they are known quantities, and do not yet take account for the fact that the κ_t index follows from a hidden process. Section 5 introduces the simulation study. Here, the link with the theory on errors-in-variables models will be made. It also explains how the true parameter uncertainty in the Lee-Carter model is investigated and how the κ_t process is identified. Next, the simulation results are presented. Section 6 completes the dissertation with a conclusion on all theoretical and numerical findings and proposes further research.

2 Literature Review

In 1992, Lee and Carter developed a two-step procedure for estimating and forecasting mortality rates and hence life expectancies. The Lee-Carter model became one of the most widely used methods in both the academic literature and practical applications (Girosi and King, 2007). The model was first fitted on a data set containing the U.S. age-specific mortality rates in the period 1933-1987. Consecutively, in a second-step time series methods were used to forecast mortality rates over the period from 1990-2065. Their model accounted for nearly all variance over time in age-specific death rates (Lee and Carter, 1992).

According to Lee and Carter the greatest benefit of their model is that it links a rich yet parsimonious demographic model with statistical time series methods. The reason for this is that their method does not need to incorporate any knowledge about medical, behavioral or social influences on mortality changes. With these features, they were convinced that their method had considerable advantages over other extrapolative procedures.

Albeit the simplicity of the Lee-Carter model, it demonstrated to yield good results in fitting mortality rates for different countries. Their work has for example been used by Wilmoth (1996) who used the Lee-Carter model to construct mortality projections for Japan, by Tuljapurkar et al. (2000) who examined the mortality over five decades in the G7 countries (Canada, France, Germany, Italy, Japan, UK, and the US), and by Koissi et al. (2005) who applied the Lee-Carter model to data from various Scandinavian countries.

On the other hand, apart from the attention and acceptance, their method also has some shortcomings and received some considerable criticism (e.g. Lee and Miller, 2001). Likewise, Girosi and King (2008) for instance argued that the Lee-Carter model does not include important covariates like tobacco consumption or comorbidity patterns. It is also argued that researchers who use the Lee-Carter model implicitly assume that all information about the future is contained in the past observed values of the log-mortality rate (Girosi and King, 2008). In addition, an exogenous shock to mortality, such as the discovery of new medical technologies, an economic crisis, a pandemic, or public health interventions are ignored when using the Lee-Carter model (Gutterman and Vanderhoof, 1998). This shows that the implicit assumption that all information about the future is contained in the past is not always realistic.

Furthermore, Keilman and Pham (2006) state a few reasons why predictions derived from the Lee-Carter model are uncertain. Firstly, they argue that forecasts are uncertain because the predictions are based on observed death counts and empirical death rates which are subject to Poisson variability. Secondly, there is uncertainty because of the extrapolated values of the model's time index. Thirdly, the predictions are build upon the estimates of the parameters of the Lee-Carter model. Keilman and Pham also observed that many authors who have computed the prediction intervals for future mortality rates usually only consider the second source of uncertainty, but ignore the first or the third source, or both. They deduce that the extent to which various sources contribute to overall uncertainty is an empirical issue. Kleinow and Richards (2017) also notice this, but neglect the

problem when producing the forecasts. Therefore, in this dissertation we particularly shed a light upon the aforementioned problem of neglecting the parameter uncertainty.

3 Methodology

3.1 The Lee-Carter model

In order to investigate the parameter uncertainty a proper understanding of the Lee-Carter model is necessary. For that reason, the fundamentals of the Lee-Carter model will be explained in detail. To improve the readability of the paper, some common definitions that are useful in understanding mortality modelling in general, such as the central death rate, are explained in Appendix 7.1.

Let $m_{x,t}$ denote the central death rate for age x in year t, with $x \in \{1, 2, ..., X\}$ and $t \in \{1, 2, ..., T\}$. The matrix M containing the central death rates is then given by:

$$M = \begin{pmatrix} m_{1,1} & m_{1,2} & \dots & m_{1,T} \\ m_{2,1} & m_{2,2} & \dots & m_{2,T} \\ \vdots & \vdots & \ddots & \vdots \\ m_{X,1} & m_{X,2} & \dots & m_{X,T} \end{pmatrix} \tag{3.1}$$

where M has dimensions $X \times T$. Then according to Lee and Carter, the central death rates of matrix (3.1) are fitted to the model:

$$m_{x,t} = e^{\alpha_x + \beta_x \kappa_t + \varepsilon_{x,t}} \tag{3.2}$$

$$\ln(m_{x,t}) = \alpha_x + \beta_x \kappa_t + \varepsilon_{x,t} \tag{3.3}$$

where α_x, β_x, and κ_t are the parameters that need to be estimated. The parameters α_x and β_x depend only on age, while κ_t is a stochastic process depending only on the year of observation. Thus, there will be X amount of α and β parameters that need to be estimated, and T amount of κ parameters that need to be estimated. Furthermore, the disturbances $\varepsilon_{x,t}$ represent the independent identically distributed normal errors with mean 0 and variance σ_ε^2, such that $\varepsilon_{x,t} \overset{iid}{\sim} \mathrm{N}(0, \sigma_\varepsilon^2)$. In addition, if one applies a logarithmic transformation to (3.2), then the Lee-Carter model can be seen as a log-linear model. The α_x parameters captures the general pattern of mortality by age, while the β_x parameters describe the speed or sensitivity at which mortality rates change due to changes in κ_t. Lastly, the values for κ_t can be understood as the index of mortality at year t. Hence, this index represents the improvement in mortality over years. In detail, the partial derivative $\frac{\partial \ln(m_{x,t})}{\partial t} = \beta_x \frac{\partial \kappa_t}{\partial t}$, implies that for higher values of β_x, the central death rate at age x fluctuates more when the index of mortality changes (Lee and Miller, 2001). This phenomenon can for instance be observed for infant mortality ($x = 1$). On the other hand, for small values of β_x, the central death rate at age x fluctuates less when the index of mortality changes. This occurs for example with mortality at older ages.

In order to use the Lee-Carter model (3.3) for forecasting, we first have to fit it using historical mortality rate data. That is, estimating the parameters α_x, β_x, and κ_t for all x and t. Notice that standard regression techniques cannot be used to fit the model, because the right hand side of the Lee-Carter model in (3.3) only contains parameters that need to be estimated. Nevertheless, there are several ways to fit the Lee-Carter model. Originally, Lee and Carter used the Singular Value Decomposition (hereafter, referred to as SVD) to fit the model. Apart from the SVD (more on this in Section 3.2) there are several other approaches to fit the model like the weighted least squares method, or maximum likelihood estimation as adopted by Wilmoth (1993) who tested the performance of fitting the Lee-Carter model using these methods. Koissi et al. (2006) compared the performance of these three approaches for estimating the parameters of the Lee-Carter model on Scandinavian data. They conclude that the values for α_x and β_x obtained by the three methods are almost identical. Moreover, the estimated mortality index κ_t generally displayed a linear decreasing trend for all three approaches, indicating that age-specific mortality declined yearly in an almost exponential fashion. Therefore, we decided to stick to the original SVD method which is used by Lee and Carter.

Nonetheless, for all these methods there exists an identification issue, as the solution to the system is not unique. In the pursuance of finding an unique solution to this system one can impose additional conditions. These constraints are known as identification constraints. The constraints used by Lee and Carter are:

$$\sum_{x=1}^{X} \beta_x = 1, \text{ and } \sum_{t=1}^{T} \kappa_t = 0 \tag{3.4}$$

However, in the literature there are several other constraint combinations which ensure the identifiability of the Lee-Carter model. Girosi and King (2008) for instance used $\sum_{x=1}^{X} \beta_x^2 = 1$ instead of $\sum_{x=1}^{X} \beta_x = 1$. At the same time, Richards and Currie (2009) used $\sum_{t=1}^{T} \kappa_t = 0$ and $\sum_{t=1}^{T} \kappa_t^2 = 1$. Note that different constraint systems generate distinct parameter estimates. Despite the fact that different constrains generate distinct parameter estimates, the generated fitted values for $\ln(m_{x,t})$ will be identical.

After estimating α_x, β_x and κ_t in the first stage, the second stage proceeds with modelling the fitted κ_t values as an ARIMA(p, d, q) process. Here p stands for the number of autoregressive parameters, d for the number of differencing terms needed to make the process stationary, and q for the number of moving-average parameters. Following the classical approach of Box et al. (2015), the κ_t values are then extrapolated through the fitted ARIMA model. The predicted κ_t values serve to compute the estimated future central death rates $m_{x,t}$. According to Lee and Carter, the κ_t process is well described by a random walk with drift. In Section 3.3, we consider in detail how to model the κ_t process as a random walk with drift. In addition, this section explains how Kleinow and Richards, who apart from the simple random walk model also considered various other ARIMA models, decomposed the overall uncertainty over the future mortality rates into two parts. All together, the above steps define the entire two-step procedure of the Lee-Carter model that

actuaries often adopt for risk management as highlighted in the paper of Li et al. (2009).

3.2 Singular Value Decomposition

The SVD is by itself an important concept within linear algebra, and has many applications within engineering and statistics, like least squares fitting of data. In general, the SVD factorizes or decomposes a matrix into a product of matrices. That is, if the SVD is applied to a real $m \times n$ matrix A, we get the factorization: $A = U\Sigma V^*$. Here, Σ represents a $m \times n$ rectangular diagonal matrix containing non-negative values on the diagonal in decreasing order ($\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_n$). The diagonal entries σ_i of Σ are called the singular values and are used to describe the characteristics of the random matrix A. In addition, the number of non-zero singular values matches with the rank of A (Golub and Reinsch, 1971). Furthermore, U and V^* are both orthogonal matrices with dimensions $m \times m$ and $n \times n$, respectively. The columns of U contain the left-singular vectors, whereas the columns of V^* represent the right-singular vectors. The orthogonality implies that both matrices satisfy the property that $U^*U = I$, and $V^*V = I$. Note that I denotes the identity matrix, and U^* and V^* denote the transpose of U and V.

In order to find the parameters in the Lee-Carter model, one can specifically apply the SVD to a matrix A with dimension $X \times T$ such that:

$$A = \begin{pmatrix} \ln(m_{1,1}) - \hat{\alpha}_1 & \ln(m_{1,2}) - \hat{\alpha}_1 & \dots & \ln(m_{1,T}) - \hat{\alpha}_1 \\ \ln(m_{2,1}) - \hat{\alpha}_2 & \ln(m_{2,2}) - \hat{\alpha}_2 & \dots & \ln(m_{2,T}) - \hat{\alpha}_2 \\ \vdots & \vdots & \ddots & \vdots \\ \ln(m_{X,1}) - \hat{\alpha}_X & \ln(m_{X,2}) - \hat{\alpha}_X & \dots & \ln(m_{X,T}) - \hat{\alpha}_X \end{pmatrix} \tag{3.5}$$

where $\hat{\alpha}_x = \frac{1}{T}\sum_{t=1}^{T} \ln(m_{x,t})$. Hence, A represents a matrix where the averages over time of the logarithmic age-specific rates have been subtracted from the logarithm of the central death rates.

The parameter estimate $\hat{\alpha}_x$ follows from the derivation:

$$\begin{aligned}
\ln(m_{x,t}) &= \alpha_x + \beta_x \kappa_t + \varepsilon_{x,t} \\
\sum_{t=1}^{T} \ln(m_{x,t}) &= \sum_{t=1}^{T} (\alpha_x + \beta_x \kappa_t + \varepsilon_{x,t}) \\
&= \sum_{t=1}^{T} \alpha_x + \sum_{t=1}^{T} \beta_x \kappa_t + \sum_{t=1}^{T} \varepsilon_{x,t} \\
&= \sum_{t=1}^{T} \alpha_x + \beta_x \sum_{t=1}^{T} \kappa_t + \sum_{t=1}^{T} \varepsilon_{x,t} \\
&\approx T\alpha_x + \beta_x \sum_{t=1}^{T} \kappa_t \\
&\overset{(3.4)}{=} T\alpha_x \\
\hat{\alpha}_x &= \frac{1}{T} \sum_{t=1}^{T} \ln(m_{x,t})
\end{aligned} \tag{3.6}$$

where we made use of the identification constraint $\sum_{t=1}^{T} \kappa_t = 0$ as displayed in (3.4).

Consequently, U, Σ and V^* follow from the SVD that is being applied to the matrix A of (3.5). Additionally, $\hat{\beta}_x$ can then be computed as $\hat{\beta}_x = \frac{1}{c} \cdot U_{x,1}$, where $U_{x,1}$ corresponds to the first column of U. Moreover, $\hat{\kappa}_t$ can then be computed as $\hat{\kappa}_t = c \cdot \Sigma_{1,1} \cdot V^*_{t,1}$, where $\Sigma_{1,1}$ is the largest singular value of Σ and $V^*_{t,1}$ corresponds to the first column of V^*. Note that the normalization factor c in both cases is defined as the sum over the first column of the left singular vector U. Thus, $c = \sum_{x=1}^{X} U_{x,1}$, which guarantees that $\sum_{x=1}^{X} \hat{\beta}_x = 1$ and $\sum_{t=1}^{T} \hat{\kappa}_t = 0$. However, if $\sum_{x=1}^{X} U_{x,1} = 0$, then it is not possible to normalize $\hat{\beta}_x$.

3.3 Modelling the κ_t process as a random walk

According to Lee and Carter, the κ_t process that follows from the SVD that is being applied to a historical table containing the central death rates, is well described by a random walk with drift. Therefore, if we suppose that κ_t process follows a random walk with drift it can be denoted as:

$$\kappa_{t+1} = \kappa_t + d + \epsilon_{t+1} \tag{3.7}$$

where d is an unknown drift term, and ϵ_{t+1} is an error term with mean zero and finite constant variance, σ^2_ϵ.

Supplementary, Kleinow and Richards (2017) showed how to decompose the overall uncertainty over the future mortality rates into two parts. In order to do so, the notation used by Kleinow and Richards (2017) is followed. Let κ_{t+h} denote a future realized or "yet-to-be-observed" value of κ at time $t+h$, and let $\kappa_t(h)$ denote a forecasted value at time $t+h$ where the forecast is h years ahead

from time t and where the underlying parameters are known. Lastly, let $\hat{\kappa}_t(h)$ be identical to $\kappa_t(h)$, however for $\hat{\kappa}_t(h)$ the underlying process parameters are estimated. Consequently, Kleinow and Richards (2017) then claim that the difference between $\hat{\kappa}_t(h)$ and $\kappa_t(h)$ represents the parameter risk, whereas the difference between $\kappa_t(h)$ and κ_{t+h} represents the accumulated random error over h years. Nonetheless, this does not take into account that the κ_t values follow from a hidden process.

Neglecting this for the moment, the decomposition of the two uncertainty factors can be derived as follows. First, under the random walk model one finds that the future realized values are given by:

$$\kappa_{t+h} = \kappa_t + h \cdot d + \sum_{j=1}^{h} \epsilon_{t+j} \tag{3.8}$$

Given the fitted value of κ_t, and under the assumption that d is known, one can acquire the central projection h years ahead by setting all error terms to zero:

$$\kappa_t(h) = \kappa_t + h \cdot d \tag{3.9}$$

Replacing d with the estimate $\hat{d}$ relaxes the previous assumption that d is known. The estimate of d is given by:

$$\hat{d} = \frac{1}{t-1} \sum_{i=2}^{t} (\kappa_i - \kappa_{i-1}) = \frac{\kappa_t - \kappa_1}{t-1} \tag{3.10}$$

Combining equations (3.9) and (3.10) yields:

$$\hat{\kappa}_t(h) = \kappa_t + h \cdot \hat{d} \tag{3.11}$$

which is the projection estimator (h years into the future). The variance of $\hat{d}$ is:

$$\mathbb{V}ar(\hat{d}) = \mathbb{V}ar\left(\frac{\kappa_t - \kappa_1}{t-1}\right) = \frac{\sigma_\epsilon^2}{t-1} \tag{3.12}$$

In fact, σ_ϵ^2 is also a parameter being estimated, and there exists parameter uncertainty over this estimate as well. Regardless of the latter, Kleinow and Richards (2017) use the sample variance $\hat{\sigma}_\epsilon^2$ to estimate σ_ϵ^2, which is given by:

$$\hat{\sigma}_\epsilon^2 = \frac{1}{t-2} \sum_{i=2}^{t} (\kappa_i - \kappa_{i-1} - \hat{d})^2 \tag{3.13}$$

Lastly, for decomposing the overall uncertainty over the future mortality rates into two parts, Kleinow and Richards used the mean squared prediction error, which is given by:

$$\mathbb{E}[(\hat{\kappa}_t(h) - \kappa_{t+h})^2] = \underbrace{\frac{h}{t-1} h\sigma_\epsilon^2}_{\text{parameter uncertainty}} + \underbrace{h\sigma_\epsilon^2}_{\text{volatility}} \tag{3.14}$$

According to Kleinow and Richards the first component in equation (3.14) corresponds to the parameter uncertainty, whereas the second component corresponds to the volatility. Moreover, from this formulation they note that the parameter uncertainty is equal to $\mathbb{V}ar(h \cdot \hat{d}) = h^2\mathbb{V}ar(\hat{d})$. Also the relation between the parameter uncertainty and volatility can be established from (3.14), which is given by:

$$\text{parameter uncertainty} = \frac{h}{t-1}\text{volatility} \tag{3.15}$$

This means that the parameter uncertainty is the bigger component of the overall uncertainty if the forecasting horizon is long and there are not many historical observations available. In practice, the aforementioned is usually the case, as actuaries often try to forecast over a long period with a relatively short data set. However, in the opposite direction, when the forecasting horizon is short and a large data set is used, it implies that the volatility component is the bigger component of the overall uncertainty. Moreover, from (3.15) one can deduce that the crossover point for the parameter uncertainty component to equal the volatility component is reached when $\frac{h}{t-1} = 1$. To summarize, if the κ_t process follows a random walk with drift, then the overall forecast uncertainty can be split up into two components: parameter uncertainty and volatility. We refer to Appendix 7.2 for a more detailed description of this decomposition that has been adopted by Kleinow and Richards (2017).

Modelling the κ_t process as a random walk with drift is the most common approach, but the process is not always modelled as a random walk with drift. Kleinow and Richards also for instance showed that other ARIMA(p, d, q) processes, such as an ARIMA$(1, 1, 2)$ perform better than the random walk with drift. These ARIMA processes showed to have superior properties such as better behaved residuals or lower information criteria when modelling the fitted κ values that follow from a historical mortality data set.

Nevertheless, as mentioned, this decomposition of risks and these time series techniques to model the κ_t process do not take into account that the κ values come from a hidden process. As a matter of fact, the κ values that follow from the SVD that is being applied to a historical mortality table are themselves prone to estimation error, and thus there exist parameter uncertainty over this estimated index. Hence, if that parameter uncertainty is not taken into account when forecasting the κ_t index, then this could cause scientist such as Kleinow and Richards (2017) to model the "wrong" κ_t process very accurately. Therefore, in Section 5 we propose a simulation study for the identification of the κ_t process and the investigation of the true parameter uncertainty in the Lee-Carter model. The simulation study also empowers us to verify how the parameter uncertainty affects the time series techniques for selecting the right ARIMA model. In addition, Section 5.1 discusses the theory on errors-in-variables models as this addresses the fundamental problem we are facing of not observing the true κ_t process.

4 Empirical Application

The following section serves as an illustration of the Lee-Carter model where we apply the SVD and time series techniques for forecasting the κ values. In addition, it demonstrates how Kleinow and Richards (2017) decomposed the overall uncertainty of the future mortality rates into the uncertainty over the trend and a volatility component. In this part, we still regard the κ values as if they are known quantities. Thus, we do not take into account the fact the κ values follow from a hidden process.

4.1 Data

The data used for the illustration and application of the Lee-Carter model is retrieved from the "Human Mortality Database" (www.mortality.org). We decided to investigate the life table for England and Wales (hereafter, referred to as the EWA data set). The EWA data set contains information on the central death rate between ages x and $x+n$, the number of deaths between ages x and $x+n$, the number of survivors at exact age x, the probability of death between ages x and $x+n$, and so on. Therefore, the data set is suitable to demonstrate the two-stage procedure of the Lee-Carter model. The data set covers the total population of England and Wales and contains both civilian and military deaths regardless whether death occurred abroad. The entire span of the EWA data set ranges from the year 1841 to 2016, covers the ages 0-110+, and can be subdivided by gender. Yet, insurers and actuaries are often concerned with annuity and pension liabilities. Therefore, in order to construct an example time series which we want to forecast, we will restrict ourselves to the ages 50-100 over the years 1971-2013 for the male data set. A comprehensive description of the this data set is provided in the paper of Cairns et al. (2016) where several limitations are pointed out when working with population data. Moreover, for the illustration, as well as the simulations which are performed later on, the statistical programming language R is being used.

4.2 Results application

First of all, the general log-mortality development in England and Wales over age and time is displayed in Figure 1. From this picture, it is clear that the logarithmic central death rates decrease over time for a given age, and increase for older ages for a given year. In the twentieth century the mortality improvements can in particular be attributed to better medical health care (diagnostics, therapies, medicine), better hygiene (drinking water, maternity care), and less dangerous life (dull work, safer cars, less drinking/smoking). Moreover, it can be observed that mortality decreases quickly in the first year, and starts to rise as of age 10 until age 100. A typical U-shaped pattern is created for the age specific death rates at younger ages due to infant mortality, and for male data a typical hump shaped feature for the general pattern of mortality is apparent around the age of 18. The hump is larger for males than for females, reflecting the higher number of male deaths due to car accidents. The aforementioned logarithmic central death rates will be reconstructed in a reverse

manner for the ages 50-100 in the simulation study that is presented in Section 5.

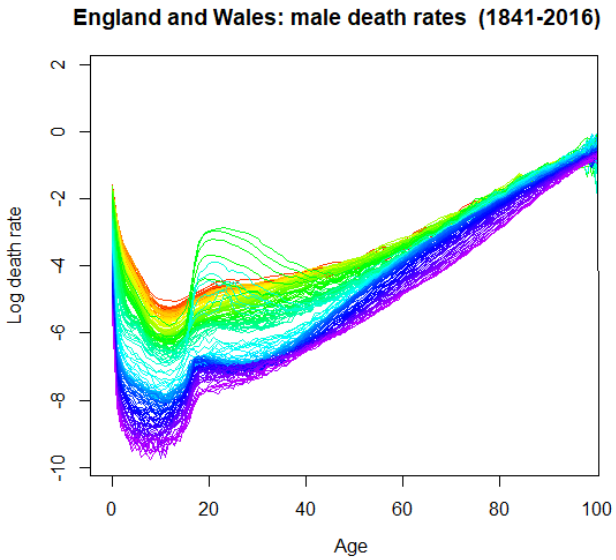

Figure 1: The historical logarithmic central death rates for the EWA male data set for the ages 0-100 over the time span from 1841-2016. Note that warmer colors (red-green) represent the more bygone years, whereas colder colors (blue-purple) represent the more recent years.

Consecutively, the resulting estimated parameters where the Lee-Carter model is being applied to the EWA data set are presented in Figure 2. In fact, notation wise there should already be a "hat" on top of these fitted α, β and κ values, because it is at this stage that the parameter uncertainty comes into being. However, in most actuarial literature these values are just denoted as α, β and κ. Therefore, throughout this section we will stick to this notation. Later in Section 5, the simulation study will be proposed that investigates the true parameter uncertainty. Nevertheless, from Figure 2 it is clear that the estimated α_x parameter for ages 50-100 displays a regular linear pattern. Usually α_x reflects the typical pattern of mortality which is relatively high around infancy, low throughout childhood, and increasing after the age of 10. In addition, α_x displays a hump around the age of 18 for male data and starts exhibiting a linear pattern until the age of 100. However, the ages 0-49 are cut-off, as a consequence we do not observe the typical U-shaped pattern at younger ages, or the typical hump shaped feature at the age of 18. The estimated β_x parameter for ages 50-100 displays a concave pattern. This indicates that the central death rates for the ages 50-80 are more sensitive to changes in κ_t. This observation makes sense, as older people for instance live longer due to advances in healthcare over time. The decreasing pattern for β_x after age 80 describes that the central death rate at higher ages fluctuates less when the index of mortality changes. In addition, in practical applications actuaries often tend to smooth the α_x and β_x parameters. However, as we are investigating the true parameter uncertainty later on, we refrained from doing so. The estimated κ_t

index, which represents the improvement in mortality over the years, displays a regular decreasing trend. This estimated κ_t index is projected in the second step of the Lee-Carter model. To this end, the estimated κ values are also included in Table 1. Although, it should once more be stressed that the fitted κ values follow from a hidden process, and it is at this point where actuaries often tend to forget that these fitted values are prone to estimation uncertainty. We refer to Appendix 7.3 for the tables containing the estimated values for α_x and β_x. These are included in the Appendix because we are not only interested in the parameter uncertainty around κ_t, but are equally interested in the joint parameter uncertainty that arises when estimating α_x, β_x, and κ_t together.

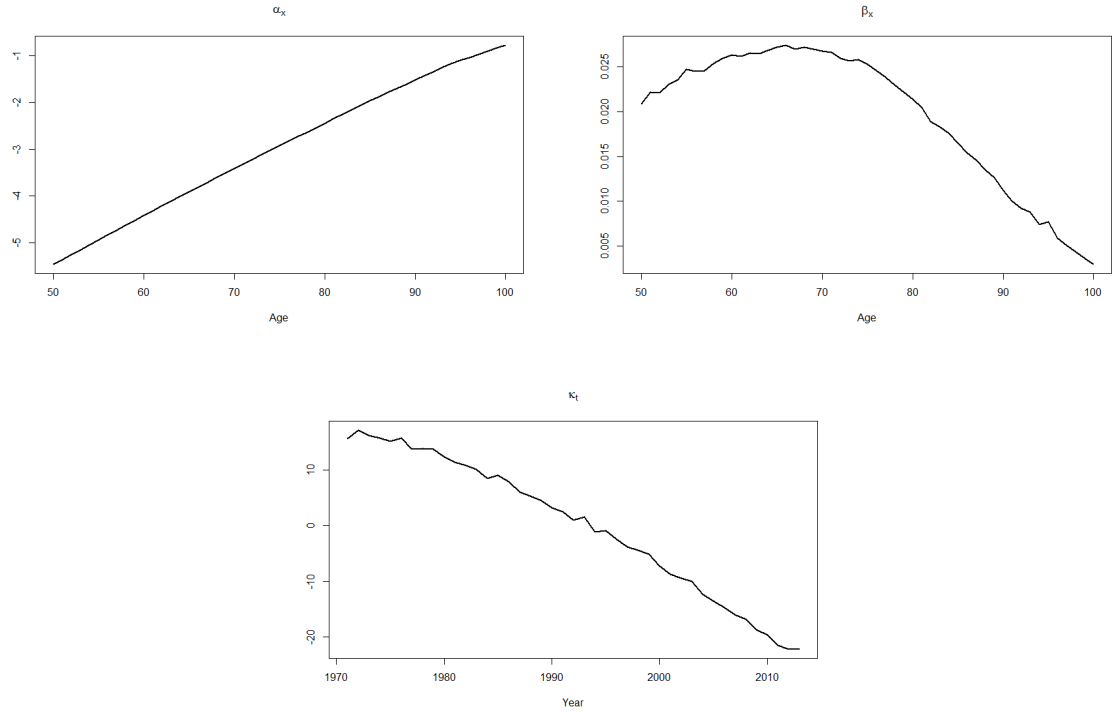

Figure 2: Parameter estimates for the Lee-Carter model fitted to the EWA male data set for the ages 50-100 over the period 1971-2013 ($x = 1, \ldots, 51$; $t = 1, \ldots, 43$).

Table 1: Estimates of κ_t from Figure 2, with the identification constraints $\sum_{x=1}^{X} \beta_x = 1$, and $\sum_{t=1}^{T} \kappa_t = 0$.

Year	κ	Year	κ	Year	κ	Year	κ
1971	15.6703967	1982	10.7949406	1993	1.5445717	2004	-12.3918436
1972	17.1831947	1983	10.0562556	1994	-1.1518479	2005	-13.5808887
1973	16.1668012	1984	8.4171405	1995	-0.8959225	2006	-14.7704749
1974	15.6845245	1985	9.1185252	1996	-2.4064719	2007	-15.9656000
1975	15.1624895	1986	7.8867891	1997	-3.7747665	2008	-16.7826076
1976	15.7304360	1987	6.0272380	1998	-4.4476806	2009	-18.6777034
1977	13.7213590	1988	5.2875482	1999	-5.1360600	2010	-19.6668690
1978	13.8978434	1989	4.4434553	2000	-7.2627270	2011	-21.4825830
1979	13.7650461	1990	3.2137841	2001	-8.7148207	2012	-22.2246769
1980	12.3789109	1991	2.5361984	2002	-9.4779281	2013	-22.2406759
1981	11.3681604	1992	1.0037582	2003	-10.0072188		

After estimating the model parameters one can further continue with the risk decomposition as described in Section 3.3. First, the results based on a ten year ahead projection are shown. In Figure 3, the estimated κ_t values from Table 1 are forecasted through a random walk with drift. Moreover, from Table 1 it follows that $\hat{d} = -0.903$. In addition, the projected line is extended backwards. This led Kleinow and Richards (2017) to argue that the random walk model with drift is a poor representation of the behaviour of the estimated κ_t process. That is why Kleinow and Richards (2017) also still modelled the κ_t process as various other ARIMA models.

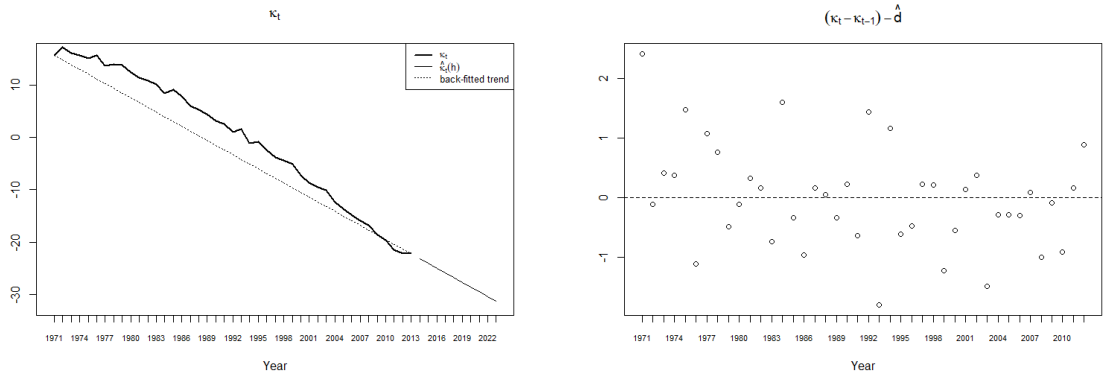

Figure 3: Left panel: the κ_t values from Table 1 and the $\hat{\kappa}_t(h)$ values which represents the random walk forecast from equation (3.11) with a forecasting horizon of $H = 10$. Right panel: the residuals following from equation (3.7) with $\hat{d} = -0.903$.

With the ease of equations (3.13) and (3.14), one can still further decompose the overall uncertainty over the future mortality rates into two parts as done in the paper of Kleinow and Richards (2017). Using equation (3.13) it follows that the sample variance equals $\hat{\sigma}_\epsilon^2 = 0.751$. The estimated standard error for $\hat{d}$ is then given by $\frac{\hat{\sigma}_\epsilon}{\sqrt{T-1}} = 0.134$. The resulting decomposition is shown in Fig-

ure 4. Figure 4 illustrates that for a small forecasting horizon ($H = 10$) the volatility component dominates the parameter uncertainty component. In addition, from panel (d) in Figure 4 it is clear that combining both sources of uncertainty results in a wider overall range.

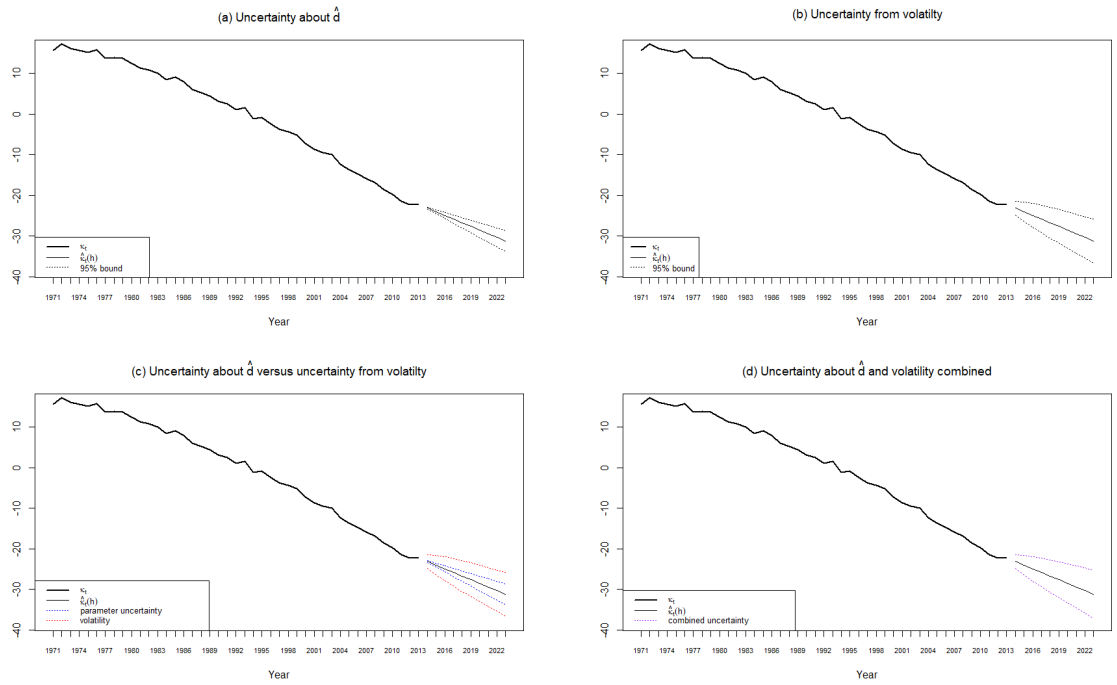

Figure 4: Panel (a): the κ_t index and the $\hat{\kappa}_t(h)$ values with the 95% bounds for the parameter/drift uncertainty over a forecasting horizon of 10 years. Panel (b): idem, but now with the 95% bounds for the stochastic volatility. Panel (c): idem, but now showing the relation between the two sources of uncertainty. Panel (d): idem, but now showing the interval arising from including both sources of uncertainty.

Next, it is considered what happens if we increase the forecasting horizon, which is something insurers typically do in practice, as they are concerned with risk management. Therefore, we now choose a forecasting horizon of $H = 60$ years. From panel (c) in Figure 5, it is clear that when the forecasting horizon increases the parameter uncertainty starts dominating the volatility component. Moreover, it illustrates the relation that is expressed in equation (3.15). The crossover point is in 2055, which is 42 years ahead from the last estimated κ_t term.

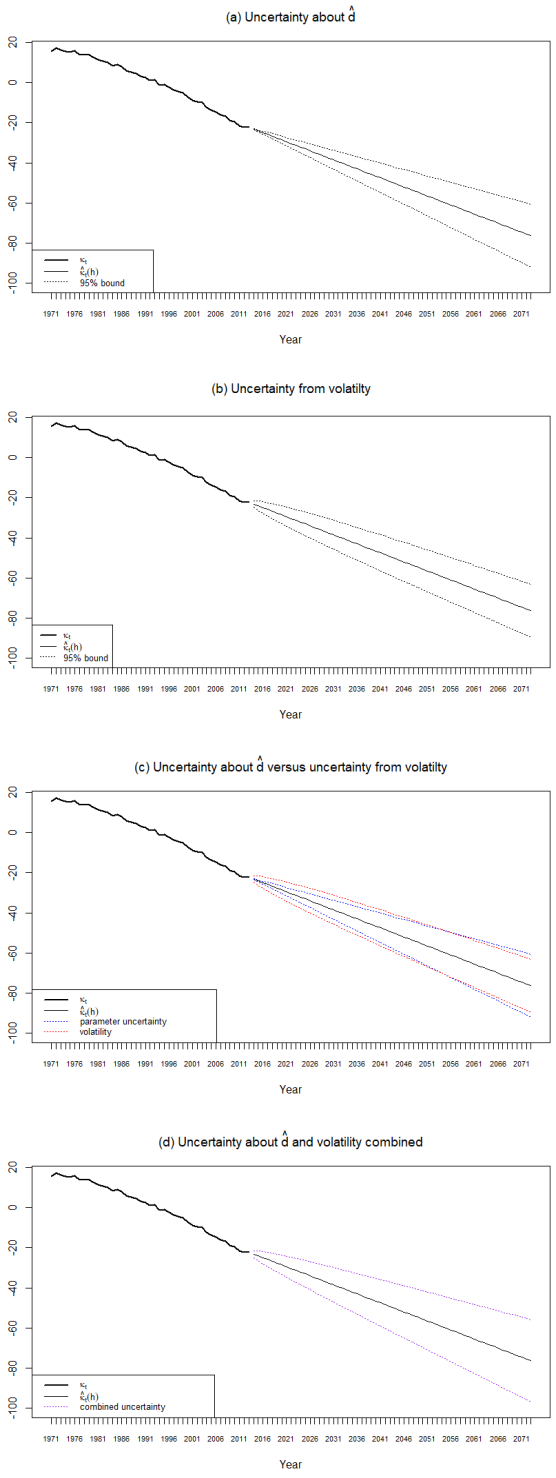

Figure 5: A similar decomposition as in Figure 4, but now with a forecasting horizon of 60 years.

5 Simulation Study

A simulation study has been performed in order to investigate the parameter uncertainty that arises when estimating the parameters of the Lee-Carter model in the first place. We are particularly interested in the κ_t parameter uncertainty, as actuaries often adopt a two-stage approach for mortality forecasting as highlighted in the paper of Li et al. (2009). As previously explained, in this two-stage approach one first estimates the α_x, β_x and κ_t parameters by applying the SVD decomposition to historical mortality data, and in the second stage the fitted κ_t values are modelled by an ARIMA process. The ARIMA model is then used to extrapolate the κ_t index. The extrapolated κ_t values serve us to get an idea of how the future central death rates might develop. However, we often tend to forget that these κ_t values in fact follow from a hidden process, and are prone to estimation uncertainty themselves. Therefore, we developed a simulation that will examine this parameter uncertainty. In addition, due to the SVD, one will obtain a joint probability distribution over the estimated $\hat{\alpha}_x$, $\hat{\beta}_x$, and $\hat{\kappa}_t$ parameters. This leads us to consider the combined parameter uncertainty of the $\hat{\alpha}_x$, $\hat{\beta}_x$, and $\hat{\kappa}_t$ parameters as well. The simulation study also investigates to which extent the joint probability distribution will have an impact on the time series techniques for finding the best fitting ARIMA model for the $\hat{\kappa}_t$ parameter. The simulation design consists out of several steps that are explained in Section 5.2. However, first the theory on error-in-variable models will be discussed in Section 5.1, as this addresses the problem we are facing of not observing the true κ_t process.

5.1 Errors-in-variables models

Standard regression models only take the errors in the dependent variable into account and suppose that the independent variables or regressors have been measured exactly or are observed without error. Yet, estimation based on the aforementioned assumption leads to inconsistent estimates in the event that the independent variables have been measured with errors. This implies that, regardless of the sample size, the parameter estimates will never converge to the true values. If one performs a simple linear regressions then this problem is known to cause an underestimate of the coefficients. This phenomenon is also known as attenuation bias, regression dilution or the "Iron Law of Econometrics" (Hausman, 2001). For non-linear models it is more complex to specify the direction of the bias. For the Lee-Carter model the same problem might arise as this is a regression framework with no observable quantities on the right hand side of the model. In particular, it holds that the κ values are not directly visible. Nevertheless, in contrast to the standard regression models, errors-in-variables models are the regression models in the field of statistics that elucidate the measurement errors in the independent variables. Therefore, the following section elaborates on the theory of error-in-variable models, where we follow the notation used in Green (2012).

Take a simple linear regression model:

$$y_t = \alpha + \beta\kappa_t^* + \varepsilon_t, \quad \text{for } t = 1, ..., T \tag{5.1}$$

where y_t represents the perfectly measured dependent variable and κ_t^* the true but unobserved independent variable. Green (2012) then describes that under the errors-in-variables problem one observes this regressor with an error term such that:

$$\kappa_t = \kappa_t^* + \eta_t \tag{5.2}$$

Here it is supposed that the measurement error η_t is independent from the true value κ_t^*. If (5.2) enters back into (5.1) one obtains:

$$\begin{aligned} y_t &= \alpha + \beta\kappa_t + \varepsilon_t - \beta\eta_t \\ y_t &= \alpha + \beta\kappa_t + v_t \end{aligned} \tag{5.3}$$

where $v_t = \varepsilon_t - \beta\eta_t$. From (5.3), it can be deduced that $\mathbb{C}ov(\kappa_t, v_t) = \mathbb{C}ov(\kappa_t^* + \eta_t, \varepsilon_t - \beta\eta_t) = -\beta\sigma_\eta^2$, which indicates that the regressor is correlated with the disturbance. As a consequence, it causes least squares regression to be inconsistent, because it violates the assumption of the classical regression model. In addition, if one now performs a simple linear regressions (y_t's on κ_t's), then the estimator for β is given by:

$$\hat{\beta} = \frac{\frac{1}{T}\sum_{t=1}^{T}(\kappa_t - \bar{\kappa})(y_t - \bar{y})}{\frac{1}{T}\sum_{t=1}^{T}(\kappa_t - \bar{\kappa})^2} \tag{5.4}$$

Subsequently, the estimator for the slope coefficient converges in probability as the sample size T increases to:

$$\hat{\beta} \xrightarrow{p} \frac{\mathbb{C}ov(\kappa_t, y_t)}{\mathbb{V}ar(\kappa_t)} = \frac{\beta\sigma_{\kappa^*}^2}{\sigma_{\kappa^*}^2 + \sigma_\eta^2} = \frac{\beta}{1 + \frac{\sigma_\eta^2}{\sigma_{\kappa^*}^2}} \tag{5.5}$$

where $\sigma_{\kappa^*}^2$ is the variance of κ_t^*. Expression (5.5) represents the regression dilution as described in Green (2012). Hence, it becomes apparent that regression dilution is the effect of biasing the estimate towards zero, due to the fact that variances in the denominator are always non-negative. All in all, the ordinary least squares estimator would be inconsistent and is biased towards zero in case the explanatory variables are observed with an error term.

Next, in Section 5.2, we will construct a simulation design where we simulate the true but unobserved independent variable κ_t^*. Therefore, expression (5.5) leads us to formulate the following hypothesis. If the unobserved but true κ_t^* values in the Lee-Carter model are measured with errors, then the parameter estimates of the Lee-Carter model will never converge to their true values. In addition, we expect that the larger the measurement error variance (σ_η^2), the worse the bias will be. Moreover, in the second stage of the Lee-Carter model, we also expect that this could result in finding an estimated drift term that is biased towards zero.

5.2 Simulation design

Steps for simulation:

1. Create N times a κ_t index, where the κ_t indices are simulated as a random walk without drift or with drift. Consecutively, make sure that each simulated κ_t index is normalized properly such that it satisfies $\sum_{t=1}^{T} \kappa_t = 0$. Besides this, fix the α_x and β_x parameters. This can be done by fitting a Lee-Carter model to the central death rates that come from an existing life table as is done in Section 4. This guarantees that the α_x and β_x parameters have a sufficient regular shape. Similar to the empirical application, we choose to fix the α_x and β_x parameters for the ages 50-100, such that both have length 51. In addition, the simulated κ_t indices have length 43. Consequently, one can use these parameters together with the simulated κ_t indices to construct the central death rates $\ln(m_{x,t})$ in a reverse manner, where $\ln(m_{x,t}) = \alpha_x + \beta_x\kappa_t + \varepsilon_{x,t}$. The error terms $\varepsilon_{x,t}$ are normally distributed and have mean 0 and variance σ_ε^2. We choose $\sigma_\varepsilon^2 = 0.001$, which is in line with the fact the variances over time of age-specific components of $\varepsilon_{x,t}$ should not differ too much as they represent deviations from the logarithmic of the central death rates (Lee and Carter, 1992). A reasonable order of σ_ε^2 can likewise be determined from the empirical application of the Lee-Carter model on the EWA data set as $\varepsilon_{x,t} = \ln(m_{x,t}) - (\alpha_x + \beta_x\kappa_t)$. In the case where the κ_t indices come from a random walk with drift we specify that the drift is equal to -0.903, which corresponds to the drift term we found in the empirical application.

2. The previous procedure defines our true data generating process (hereafter, referred to as DGP). Consecutively, we fit a Lee-Carter model to the $\ln(m_{x,t})$ that have been constructed according to our DGP. This is done by applying the SVD to the matrix A as is described in Section 3.2.

3. From this point we can obtain the estimated $\hat{\alpha}_x$, $\hat{\beta}_x$, and $\hat{\kappa}_t$ parameters.

4. Next, the estimated parameters in step 3 are compared with the parameters that defined our true DGP in step 1. This is done by computing the bias or estimation errors, such that for each simulation (N=100,000) we observe the estimation errors arising from: $\alpha_x - \hat{\alpha}_x$, $\beta_x - \hat{\beta}_x$, and $\kappa_t - \hat{\kappa}_t$. In addition, the joint estimation errors arising from $(\alpha_x, \beta_x, \kappa_t) - (\hat{\alpha}_x, \hat{\beta}_x, \hat{\kappa}_t)$ are calculated. This in turn enables us to compute the variance-covariance matrices of the estimation errors, and the correlation matrices of the estimation errors. We refer to the Appendix (7.2) for a more detailed description on how to obtain the variance-covariance and correlation matrices of the estimation errors.

5. In the last stage, using the "auto.arima()" function in R, the best ARIMA(p, d, q) models are selected according to various information criteria for the simulated κ_t indices as well as for the estimated $\hat{\kappa}_t$ indices. The Akaike information criterion (hereafter, referred to as AIC), and the Bayesian information criterion (hereafter, referred to as BIC) are used. As the models are selected according to these two information criteria it is briefly stated how they are computed:

$$\begin{aligned} \text{AIC} &= 2k - 2\ln(\hat{L}) \\ \text{BIC} &= k\ln(n) - 2\ln(\hat{L}) \end{aligned} \tag{5.6}$$

In both cases $\hat{L}$ denotes the maximum value of the likelihood function for the ARIMA model, and k corresponds to the number of estimated parameters in the ARIMA model. Lastly, n denotes the number of observations. Hence, both the AIC as the BIC assess the goodness of fit measured by the maximum value of the likelihood function, but add a penalty for including a higher amount of estimated parameters. So, given a set of candidate ARIMA models for the simulated κ_t and estimated $\hat{\kappa_t}$ indices, the "auto.arima()" function conducts a search over possible models and selects the model with the lowest AIC or BIC value. Moreover, the "auto.arima()" function mimics the approach of Box et al. (2015). Therefore, this allows one to answer how the parameter uncertainty affects the time series techniques for finding the best fitting ARIMA model for the $\hat{\kappa}_t$ parameter. Sequentially, the hypothesis as formulated in Section 5.1 is tested. To do so, the "Arima()" function in R is used. This function allows one to fit a specific ARIMA(p, d, q) with drift to the simulated κ_t and estimated $\hat{\kappa}_t$ indices. Comparing the estimated drift terms that arise from this procedure, enabled us to answer whether the drift terms that follow from the estimated $\hat{\kappa}_t$ indices are biased towards zero. Note that a model based on the "Arima()" function might be outperformed in terms of AIC or BIC by the model that is selected according to the "auto.arima()" function, as the latter function will select the best model according to the AIC and BIC, respectively.

5.3 Simulation results

In this section the results obtained from the simulation study will be described. First, the results where the κ_t indices come from a random walk without drift will be discussed. Subsequently, the results for the case where the κ_t indices come from a random walk with drift will be shown. These results are followed by a short discussion, after which we present the results obtained by various alternative cases that are available for reconstructing the mortality rates. Note that all results are based on 100,000 simulations.

5.3.1 Random walk without drift

We first consider the design where the κ_t index is simulated according to a random walk without drift. Using the simulated κ_t indices we were able to construct the central death rates $\ln(m_{x,t})$ in a reverse manner. Figure 6 displays an illustration of some of the simulated paths that define our DGP in this setting.

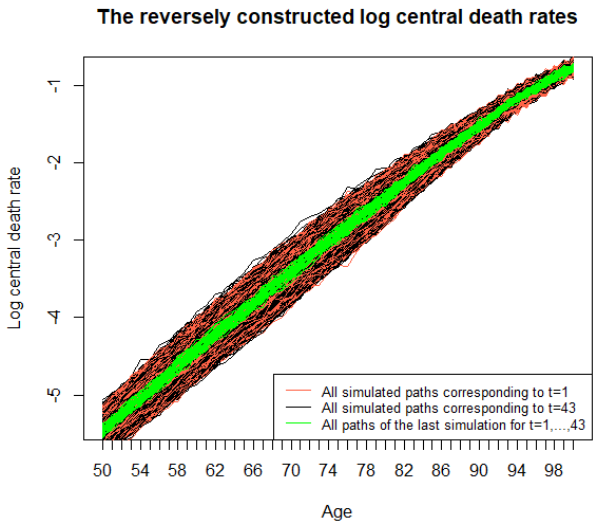

Figure 6: The simulated logarithmic central death rates resembling the ages 50-100 and the years 1971-2013 in case there is no drift.

Next, the Lee-Carter model is fitted to the $\ln(m_{x,t})$ from Figure 6. As a consequence, we obtained the estimated parameters $\hat{\alpha}_x$, $\hat{\beta}_x$, and $\hat{\kappa}_t$. Figure 7 serves to illustrate the estimation error that arises. It displays the estimated $\hat{\alpha}_x$, $\hat{\beta}_x$, and $\hat{\kappa}_t$ parameter values that correspond to the last simulation, and they are being compared to the true α_x, β_x, and κ_t parameter values. From these figures it appears that for this single case there is almost no estimation error for the α_x parameter, whereas for the β_x and κ_t parameters there is a certain degree of estimation error.

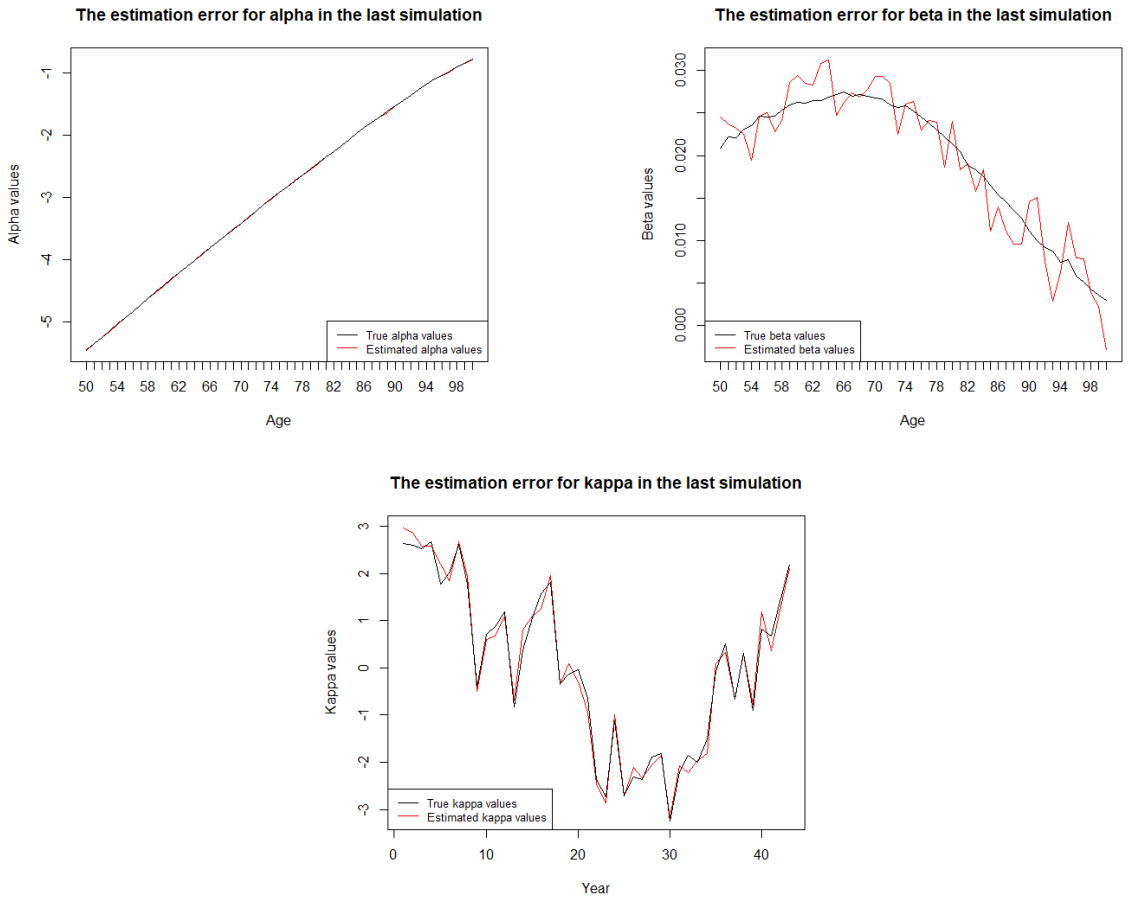

Figure 7: An illustration of the estimation error in the last simulation for all three parameters.

Subsequently, we obtained four variance-covariance matrices of the estimation errors for the α_x, β_x, κ_t, and the joint parameter terms. These variance-covariance matrices are in the form:

$$cov(E_j) = \begin{pmatrix} var(e_1) & cov(e_1, e_2) & \dots & cov(e_1, e_M) \\ cov(e_2, e_1) & var(e_2) & \dots & cov(e_2, e_M) \\ \vdots & \vdots & \ddots & \vdots \\ cov(e_M, e_2) & cov(e_M, e_2) & \dots & var(e_M) \end{pmatrix} \tag{5.7}$$

where $cov(E_j)$ has dimensions 51×51, 51×51, 43×43 or 145×145 for the α_x, β_x, κ_t or joint parameter terms, respectively (see Appendix 7.2). From these variance-covariance matrices four correlation matrices were obtained. These correlation matrices are used to construct heatmaps showing the correlation between the estimation errors for all parameters which are displayed in Figure 8. From Figure 8, it is apparent that the estimation errors for α_x are uncorrelated, whereas

the estimation errors for β_x and κ_t are negatively correlated. For the latter parameter this indicates that there is serial correlation present. So, over various time intervals, the estimation errors for the κ terms are negatively correlated with lagged versions of the estimation errors for the κ terms. Also interesting are the negative block structures that one can clearly observe in the heatmap that displays the correlation between the joint estimation errors. These observations could cause troubles when forecasting the mortality rates. Nevertheless, the colour palette scale for the correlation values had to be adjusted to $\rho \in [-0.1; 0.1]$ instead of the regular $\rho \in [-1; 1]$. This was done to make the small correlations more visible, otherwise the heatmaps would appear completely white. The highest and lowest off-diagonal element for the correlation of the estimation error for α_x is 0.01 and -0.01, respectively. For both β_x and κ_t this corresponds to 0.00 and -0.04. Although negative correlation block structures are clearly present, the aforementioned indicates that the correlation between the estimation errors is relatively small when the κ_t index is simulated as a random walk without drift.

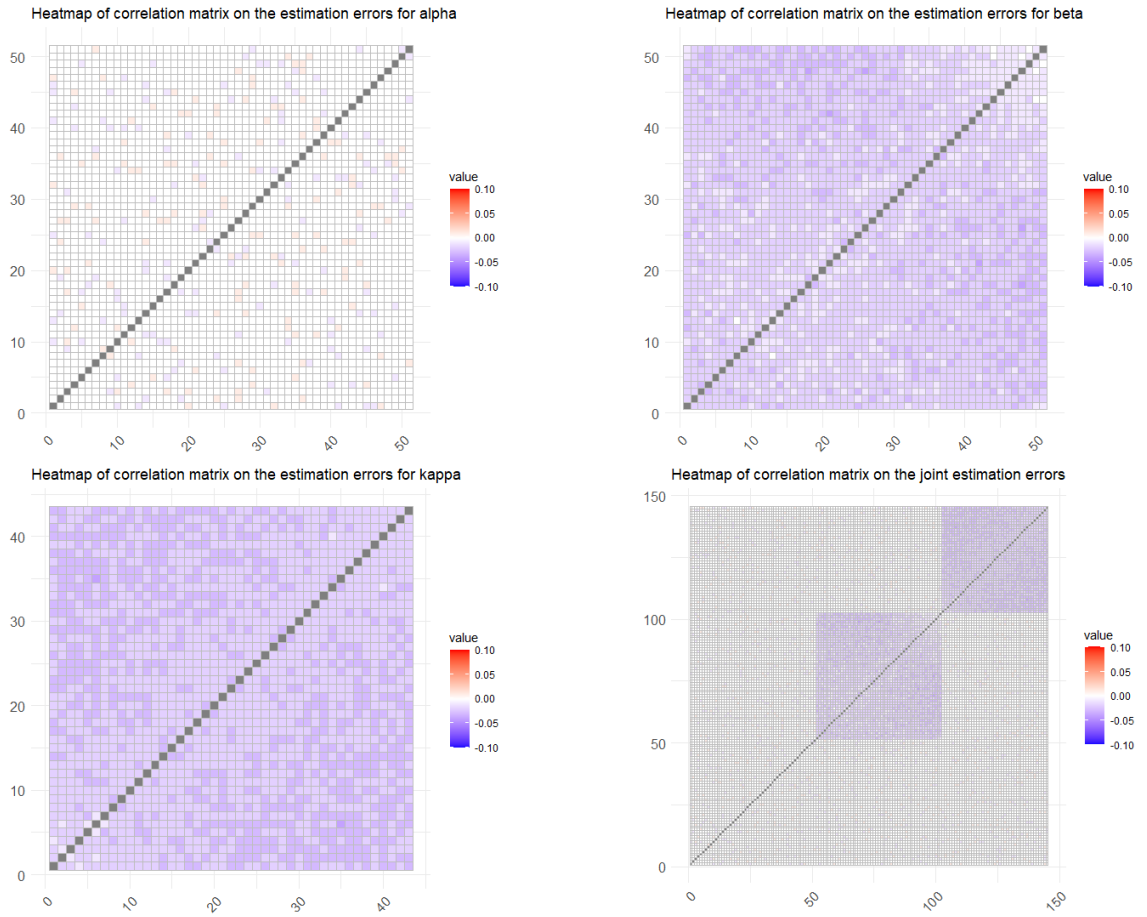

Figure 8: Heatmaps of correlation matrices on the estimation errors for α_x, β_x, κ_t, and all parameters jointly. The case where the κ_t index is simulated as a random walk without drift.

Next, it is verified how the parameter uncertainty affects the time series techniques for selecting the right ARIMA model. Table 2 displays the percentage of times a certain ARIMA(p, d, q) model without drift is selected out of 100,000 simulations for the as random walk without drift simulated κ_t indices, and for the estimated $\hat{\kappa}_t$ indices according to AIC, and BIC, respectively. We observe that when using the AIC in 76.9% of the cases an ARIMA(0,1,0) without drift is found for the simulated κ_t indices, and in 75.5% of the cases for the estimated $\hat{\kappa}_t$ indices. So, there is a difference of 1.4%, demonstrating that for the estimated $\hat{\kappa}_t$ indices we find slightly less ARIMA(0,1,0) models without drift. Instead, slightly more often an ARIMA(1,1,0) or an ARIMA(0,1,1) is selected for the $\hat{\kappa}_t$ indices. A similar pattern can be observed when using the BIC, with a difference of 1.1% between the simulated and estimated indices. Hence, for both information criteria the differences in model selection between the κ_t and $\hat{\kappa}_t$ indices are quite small. However, when using the BIC one finds in general way more often the right model for simulated as well as for estimated indices than when using the AIC. This difference can be explained, as the BIC tends to select more sparse models in comparison to the AIC. Also from equation (5.6) it is clear that AIC penalizes less for model complexity than BIC, which explains that the BIC more often finds an ARIMA(0,1,0) without drift. All in all, the small correlation between the estimation errors that is present seems to have a negligible effect on the identification of the right ARIMA models for the estimated $\hat{\kappa}_t$ indices.

Table 2: The percentage of times a certain ARIMA(p, d, q) model without drift is selected out of 100,000 simulations for the as random walk without drift simulated κ_t indices, and for the estimated $\hat{\kappa}_t$ indices according to AIC, and BIC, respectively.

	Fitted model for simulated κ_t (AIC/BIC)	Fitted model for estimated $\hat{\kappa}_t$ (AIC/BIC)
ARIMA$(0, 1, 0)$	76.9% / 92.5%	75.5% / 91.4%
ARIMA$(1, 1, 1)$	2.1% / 0.2%	2.2% / 0.2%
ARIMA$(1, 1, 0)$	6.9% / 2.9%	7.5% / 3.4%
ARIMA$(0, 1, 1)$	8.4% / 3.9%	9.0%/ 4.5%
Other ARIMA(p, d, q)	5.7% / 0.5%	5.8% / 0.5%

5.3.2 Random walk with drift

The second design we consider is where the κ_t index is simulated according to a random walk with drift. After simulating the κ_t indices, we were able to construct the central death rates $\ln(m_{x,t})$ in a reverse manner. Figure 9 displays an illustration of some of the simulated paths that define our DGP. In comparison to Section 5.3.1, we now also incorporated the downward shift in logarithmic central death rates over time. This resembles the pattern that is also visible in practice, see Figure 1. Clearly, the difference is that in the current setting one knows which underlying processes define the DGP.

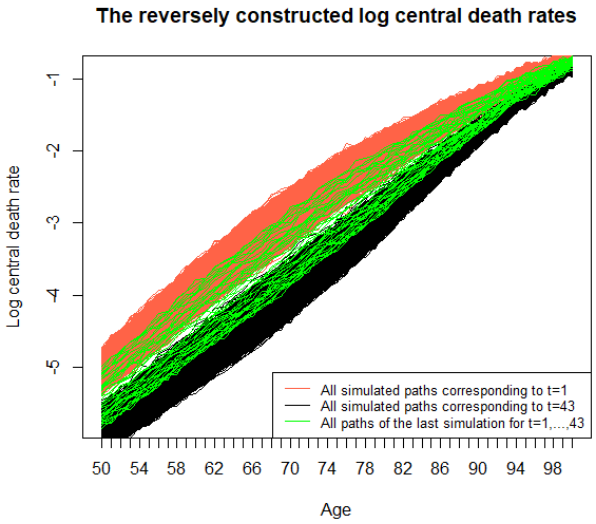

Figure 9: The simulated logarithmic central death rates resembling the ages 50-100 and the years 1971-2013 in the case where there is a drift.

Sequentially, following steps 2-4 from the simulation design, we obtained four correlation matrices of the estimation errors for the α_x, β_x, κ_t and joint parameter terms. The corresponding heatmaps are displayed in Figure 10. Similarly to the case where the κ_t index is simulated without drift we observe that the estimation errors for α_x are uncorrelated, whereas the estimation errors for β_x and κ_t are negatively correlated. Thus, also in this case there exists autocorrelation between the estimation errors for the κ terms. Moreover, the negative block structures are again visible in the heatmap representing the negative correlation between the joint estimation errors. In addition, we now also observe positive correlation between the β_x and κ_t estimation errors in this heatmap. These observations could again cause problems when forecasting mortality rates in the second-stage of the Lee-Carter model. Nevertheless, the highest and lowest off-diagonal element for the correlation of the estimation error for α_x is 0.01 and -0.01, respectively. For both β_x and κ_t this corresponds to 0.00 and -0.04. Despite the fact that clear structures in the heatmaps are visible, the latter demonstrates

that the correlation between the estimation errors is minor when the κ_t indices are simulated as a random walk with drift.

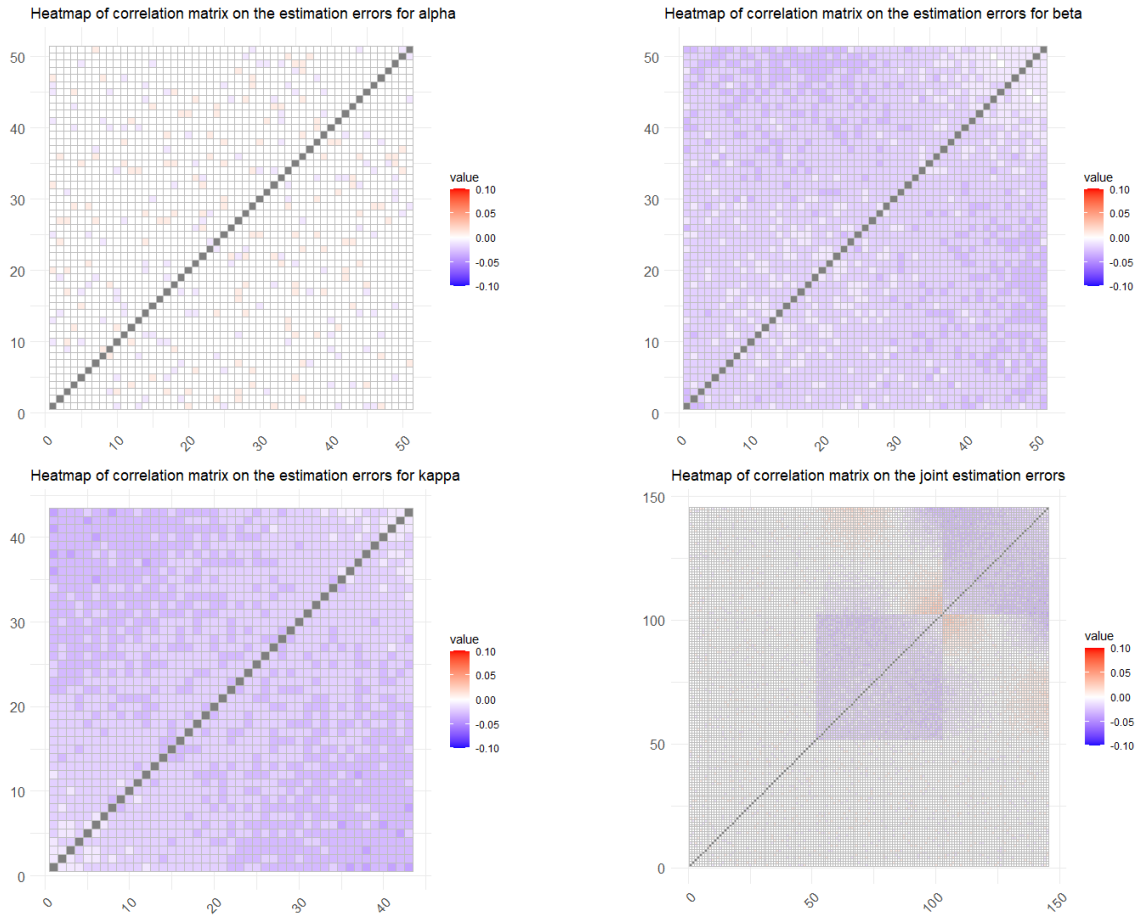

Figure 10: Heatmaps of correlation matrices on the estimation errors for α_x, β_x, κ_t, and all parameters jointly. The case where the κ_t index is simulated as a random walk with drift.

In the following, the effects of not taking the parameter uncertainty into account are highlighted when forecasting mortality rates. Table 3 displays the percentage of times a certain ARIMA(p, d, q) model with drift is selected out of 100,000 simulations for the as random walk with drift simulated κ_t indices, and for the estimated $\hat{\kappa}_t$ indices according to AIC, and BIC, respectively. From this table it is evident that not for all κ_t indices with underlying random walk with drift we find an ARIMA(0,1,0) with drift. Note that when using the AIC in approximately 76.6% of the cases the time series techniques indeed return a proper ARIMA(0,1,0) with drift for the simulated κ_t indices, whereas for the BIC this holds for approximately 91.6% of the cases. However, for the estimated $\hat{\kappa}_t$ indices this is approximately 73.6%, and 89.4%, respectively. Hence, in comparison with the simulated κ_t indices, it is apparent that the percentage of times the right ARIMA(0,1,0) model with

drift is returned for the estimated $\hat{\kappa}_t$ indices is slightly less for both information criteria. In specific, 3.0% less often when AIC is used, and 2.2% less often when BIC is used. Instead, more often an ARIMA(1,1,0) with drift or an ARIMA(0,1,1) with drift is selected for the $\hat{\kappa}_t$ indices. Nevertheless, for both information criteria the differences in model selection between the κ_t and $\hat{\kappa}_t$ indices due to the parameter uncertainty remains limited. In addition, the BIC performs better than the AIC when identifying the right ARIMA process for the simulated κ_t indices. This difference between AIC and BIC can again be explained by the fact that the BIC tends to select more sparse models than the AIC, see equation (5.6). So, when using the AIC, more often a less sparse ARIMA(p, d, q) with drift is selected for forecasting the κ_t and $\hat{\kappa}_t$ indices than the ARIMA(0,1,0) with drift. Some of the models that appeared under the "Other ARIMA(p, d, q)" enumerator in Table 3 when using the AIC are for instance an ARIMA(2,1,2), an ARIMA(0,1,2), and an ARIMA(2,1,2) model. In short, the small correlation between the estimation errors that is present, seems to have a limited effect on the identification of the right ARIMA models for the estimated $\hat{\kappa}_t$ indices.

Table 3: The percentage of times a certain ARIMA(p, d, q) model with drift is selected out of 100,000 simulations for the as random walk with drift simulated κ_t indices, and for the estimated $\hat{\kappa}_t$ indices according to AIC, and BIC, respectively.

	Fitted model for simulated κ_t (AIC/BIC)	Fitted model for estimated $\hat{\kappa}_t$ (AIC/BIC)
ARIMA$(0, 1, 0)$	76.6% / 91.6%	73.6% / 89.4%
ARIMA$(1, 1, 1)$	1.2% / 0.1%	1.2% / 0.1%
ARIMA$(1, 1, 0)$	6.3% / 2.8%	7.5% / 3.6%
ARIMA$(0, 1, 1)$	10.6% / 5.1%	12.2% / 6.4%
Other ARIMA(p, d, q)	5.3% / 0.4%	5.5% / 0.5%

Next, we continue with presenting the results on testing the hypothesis as formulated in Section 5.1. In order to do so, we follow the procedure as explained in step 5 of the simulation design. That is, we fit an ARIMA(0,1,0) with drift in each simulation to the simulated κ_t index and to the estimated $\hat{\kappa}_t$ index. Thereafter, we compute the average drift terms that follow from this procedure. The average drift term corresponding to fitting an ARIMA(0,1,0) model with drift to the simulated κ_t indices equals -0.903 (0.151), whereas the average drift term corresponding to fitting an ARIMA(0,1,0) model with drift to the estimated $\hat{\kappa}_t$ indices equals -0.903 (0.158). The standard deviations (SD) are in parentheses. Hence, the average drift term that is based on the estimated $\hat{\kappa}_t$ processes is not lower than the average drift term that is based on the simulated κ_t processes. Moreover, reminisce that we used a drift term of -0.903 to simulate the true κ_t processes. These observations are not in line with the magnitude of bias that we expected for the drift term when formulating the hypothesis in Section 5.1. Hence, we cannot support the hypothesis that the

estimated drift term for the $\hat{\kappa}_t$ processes is biased towards zero. Nevertheless, it is noticeable that the SD for the average drift term of the estimated $\hat{\kappa}_t$ indices is slightly bigger than the SD for the average drift term of the simulated κ_t indices, which results in a wider prediction interval for the estimated $\hat{\kappa}_t$ indices. Further, in view of equation (5.5), we found that σ_η^2 on average equals 0.091, and that $\sigma_{\kappa^*}^2$ on average equals 135.671. Recall that σ_η^2 denotes the measurement error variance, and that $\sigma_{\kappa^*}^2$ represents the variance of the true but unobserved independent variable. We expected that the larger the measurement error variance the worse the bias will be. On the contrary, we found that the measurement error variance is small. Moreover, the fraction $\frac{\sigma_\eta^2}{\sigma_{\kappa^*}^2}$ is on average smaller than 0.001. Therefore, it is concluded that the parameter estimates do converge to their true values.

Summary results. For the cases discussed in Section 5.3.1 and 5.3.2 there is a certain degree of correlation between the estimation errors. In both cases there is no correlation between the estimation errors for the α_x parameter. Contrariwise, for the β_x and κ_t parameters mainly negative correlation between the estimation errors is observed, causing block structures for the estimation errors when all parameters are considered jointly. However, in both cases the parameter uncertainty seems to have a limited effect on the time series techniques for identifying the right ARIMA model. In the no drift case one finds 1.4% less often an ARIMA(0,1,0) for the estimated $\hat{\kappa}_t$ indices when AIC is used, and 1.1% less often when BIC is used. In the drift case one finds 3% less often an ARIMA(0,1,0) for the estimated $\hat{\kappa}_t$ indices when AIC is used, and 2.2% less often when BIC is used. Lastly, for the drift case we find that the estimated drift term for the estimated $\hat{\kappa}_t$ indices is not biased towards zero, and that the measurement error variance is small.

Discussion results. It can be argued that the κ values do not need to follow from a random walk without drift or with drift. The κ values could equally come from any other ARIMA(p, d, q) process. Therefore, we also still considered an ARIMA$(1, 1, 0)$ process with drift to simulate the κ_t indices and to construct the central death rates in a reverse manner. The ARIMA$(1, 1, 0)$ also corresponds to one of the additional models that have been used to produce mortality forecasts in the paper of Kleinow and Richards (2017). Similar features and block structures were observed in the heatmap, indicating that there is again a certain degree of correlation between the estimation errors. Despite the aforementioned, the percentage of times the time series techniques failed to identify the correct ARIMA model for the estimated $\hat{\kappa}_t$ indices remained small. For those reasons we do not again display the corresponding heatmap and table for this case. Instead, we refer to Appendix 7.3 and 7.4.

Alternatively, it could be that the errors of the Lee-Carter model in equation (3.3) are not normally distributed. Therefore, we still considered the same simulation design, but now the Student's t-distribution with 4 degrees of freedom will be used in first step of the design to reconstruct the central death rates. Section 5.3.3 still addresses the results where the errors are not normally distributed.

Besides this, the central death rates do not need to be constructed in a reverse manner according

to the Lee-Carter model. Instead, they could follow from any other higher order stochastic mortality model such as a multivariate version of the Lee-Carter model. This indicates that actuaries in practice are not only exposed to longevity trend risk, but to model selection risk as well. Therefore, Section 5.3.4 still addresses the results where we reconstruct the mortality rates using a bivariate Lee-Carter model. Doing so, we were able to mimic a situation in which an actuary uses a Lee-Carter for forecasting mortality rates, but where the underlying central death rates actually followed from a bivariate Lee-Carter model.

For this purpose, the theory behind the multivariate and thus the bivariate Lee-Carter model is still briefly explained. Let the multivariate Lee-Carter approach for mortality modelling be given by:

$$\ln(m_{x,t}) = \alpha_x + \sum_{i=1}^{\tau} \beta_x^{(i)} \kappa_t^{(i)} + \varepsilon_{x,t} \quad \tau < \min(X, T) \tag{5.8}$$

where $\beta_x^{(i)}$ and $\kappa_t^{(i)}$ correspond to the left and right singular vectors that are subject to the identification constraints:

$$\sum_{x=1}^{X} \beta_x^{(i)} = 1, \text{ and } \sum_{t=1}^{T} \kappa_t^{(i)} = 0, \quad \forall_i \tag{5.9}$$

Note that if $\tau = 1$ we find the original Lee-Carter model. Consecutively, if $\tau = 2$, one obtains a bivariate Lee-Carter model that is also known as the LC2 model (Renshaw and Haberman, 2003), such that:

$$\ln(m_{x,t}) = \alpha_x + \beta_x^{(1)} \kappa_t^{(1)} + \beta_x^{(2)} \kappa_t^{(2)} + \varepsilon_{x,t} \tag{5.10}$$

Hence, in a multivariate Lee-Carter setting, the dynamics of mortality are driven by multiple stochastic time indexes $\kappa_t^{(i)}$, and multiple age related factors $\beta_x^{(i)}$. As a consequence, the forecasting of mortality rates then requires the modelling of multiple time indexes using time series techniques. In actuarial literature it is often assumed that all the time indexes $\kappa_t^{(i)}$ follow a random walk with drift, such that:

$$\kappa_{\mathbf{t}} = \delta + \kappa_{\mathbf{t-1}} + \mathbf{\Psi_t^{\kappa}}, \quad \kappa_{\mathbf{t}} = \left(\kappa_{\mathbf{t}}^{\mathbf{1}}, \ldots, \kappa_{\mathbf{t}}^{\tau}\right)^{\mathbf{T}}, \quad \mathbf{\Psi_t^{\kappa}} \sim \mathbf{N}(\mathbf{0}, \mathbf{\Sigma}) \tag{5.11}$$

Here δ is an τ-dimensional vector containing the drift parameters, and Σ is the $\tau \times \tau$ variance-covariance matrix of the multivariate white noise Ψ_t^{κ}. In a more complex setting, it is sometimes supposed that the time indices follow a certain ARIMA(p_i, d_i, q_i) with drift.

5.3.3 Results non normal errors

This section addresses the results obtained when the errors are not normally distributed, and the κ_t values are simulated as a random walk with drift. Instead, when constructing the central death rates in a reverse manner the Student's t-distribution with 4 degrees of freedom has been used to simulate the errors. Figure 11 displays the heatmap of the correlation matrix on the estimation

errors for all parameters jointly. The blue coloured block structures are again visible. This indicates that the estimation errors for β_x and κ_t are negatively correlated, and that there is serial correlation between the estimation errors for the κ_t terms. The estimation errors for α_x remain uncorrelated. We also detect positive correlation between the estimation errors of the β_x and κ_t terms. The highest and lowest off-diagonal element for the correlation between the estimation errors in this heatmap is 0.03 and -0.05, respectively. Despite the fact that we used a Student's t-distribution to simulate the errors, the correlation between the estimation errors did not drastically change. Nevertheless, the correlation between the estimation errors remains prominently visible. This could again cause troubles when forecasting mortality rates in the second-stage of the Lee-Carter model.

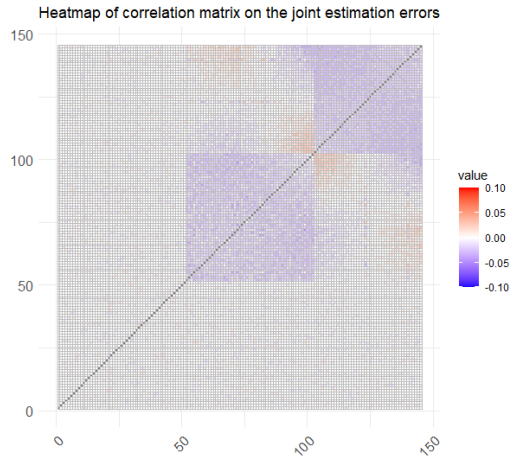

Figure 11: Heatmap of the correlation matrix on the estimation errors for all parameters jointly. The case where the κ_t index is simulated as a random walk with drift. The setting where the errors are not normally distributed.

Thereupon, it is verified how the parameter uncertainty affects the time series techniques for selecting the right ARIMA model. From Table 4 it is evident that when using the AIC in approximately 76.6% of the cases the time series techniques indeed return a proper ARIMA(0,1,0) with drift for the simulated κ_t indices, whereas for BIC this holds for approximately 91.6% of the cases. For the estimated $\hat{\kappa}_t$ indices this is approximately 69.1%, and 86.1%, respectively. In particular, the difference in model selection becomes 7.5% when AIC is used, and 5.5% when BIC is used. Instead, more frequently an ARIMA(1,1,0) or ARIMA(0,1,1) is selected. Thus, in comparison to Section 5.3.2 the difference in model selection between the κ_t and $\hat{\kappa}_t$ indices becomes more substantial in case the errors are simulated using a t-distribution with 4 degrees of freedom. Put differently, under these circumstances the time series techniques fail more frequently to identify the right ARIMA model for the underlying κ_t process. As a result, in more cases it causes the estimated prediction intervals that are based on the $\hat{\kappa}_t$ indices to look different than the prediction intervals that would have followed from the true underlying κ_t process. Thus, in this setting there are also more consequences for the

prediction uncertainty. For the previous cases in Section 5.3.1 and 5.3.2 this problem did not yet arise as the differences in model selection were limited.

Table 4: The number of times a certain ARIMA(p, d, q) model with drift is selected out of 100,000 simulations for the as random walk with drift simulated κ_t indices, and for the estimated $\hat{\kappa}_t$ indices according to AIC, and BIC, respectively. The setting where the errors are not normally distributed.

	Fitted model for simulated κ_t (AIC/BIC)	Fitted model for estimated $\hat{\kappa}_t$ (AIC/BIC)
ARIMA$(0, 1, 0)$	76.6% / 91.6%	69.1% / 86.1%
ARIMA$(1, 1, 1)$	1.2% / 0.1%	1.2% / 0.1%
ARIMA$(1, 1, 0)$	6.3% / 2.8%	9.0% / 4.8%
ARIMA$(0, 1, 1)$	10.6% / 5.1%	14.8% / 8.4%
Other ARIMA(p, d, q)	5.3% / 0.4%	5.9% / 0.6%

5.3.4 Results bivariate Lee-Carter model

This section discusses the results obtained when the mortality rates are reconstructed using a bivariate Lee-Carter model. To do so, first the $\beta_x^{(1)}$ and $\beta_x^{(2)}$ terms had to be fixed. Therefore, we fitted a bivariate Lee-Carter model to the EWA male data set. We refer to Appendix 7.4 for the resulting estimated parameters. Consecutively, we simulated 100,000 times a $\kappa_t^{(1)}$ process and a $\kappa_t^{(2)}$ process that both follow a random walk. The drift term for the simulated $\kappa_t^{(1)}$ process is still specified to equal -0.903, whereas we specified the simulated $\kappa_t^{(2)}$ process to have no drift term. Simulating such two κ_t processes allowed us to construct the $\ln(m_{x,t})$ mortality rates in a reverse manner according to a bivariate Lee-Carter model.

Figure 12 views the heatmap of the correlation matrix on the estimation errors for all parameters jointly. Note that we did not adjust the colour palette scale for the heatmap in this case, such that $\rho \in [-1; 1]$. The estimation errors for α_x remain uncorrelated. However, the correlation between the estimation errors for β_x changes drastically in this situation. Also the pattern for the correlation between the estimation errors for κ_t becomes different. The estimation errors for the κ_t terms are no longer slightly negatively correlated. Instead, the highest and lowest off-diagonal element for the correlation of the estimation error for the κ_t terms is 0.31 and -0.16, respectively. The positively correlated estimation error terms are observed closer to the diagonal, whereas the negatively correlated estimation error terms are observed further away from the diagonal. Moreover, the correlation between the estimation errors for the β_x terms and the κ_t terms vary way more than in the previous settings. The highest and lowest off-diagonal element for the correlation between the estimation errors in the entire heatmap is 0.98 and -0.97, respectively. This shows that the estimation errors for the β_x terms are almost perfectly positively and or negatively correlated.

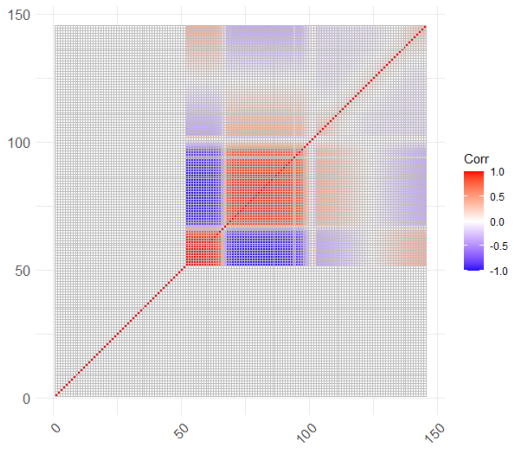

Figure 12: Heatmap of the correlation matrix on the estimation errors for all parameters jointly. The case where the $\kappa_t^{(1)}$ index is simulated as a random walk with drift and the $\kappa_t^{(2)}$ is simulated as a random walk without drift. The bivariate Lee-Carter model setting.

Thereafter, it is verified how the time series techniques for selecting the right ARIMA model are affected. From Table 5, it is noticeable that when using the AIC in approximately 76.6% of the cases the time series techniques indeed return a proper ARIMA(0,1,0) with drift for the simulated $\kappa_t^{(1)}$ indices, whereas for BIC this holds for approximately 91.6% of the cases. For the estimated $\hat{\kappa}_t$ indices this is approximately 70.7%, and 85.6%, respectively. In particular, the difference in model selection becomes 5.9% when AIC is used, and 6.0% when BIC is used. Instead, more frequently "Other ARIMA(p, d, q)" models are selected. This is a remarkable observation, as the "Other ARIMA(p, d, q)" models now contribute for the bigger part, whereas in the previous settings mainly ARIMA(1,1,0) or ARIMA(0,1,1) were selected. Hence, in comparison to Section 5.3.2 the difference in model selection between the $\kappa_t^{(1)}$ and $\hat{\kappa}_t$ indices becomes more substantial in case the logarithmic central death rates are simulated using a bivariate Lee-Carter model. In other words, the time series techniques fail more frequently to identify the right ARIMA model in case there are multiple underlying true κ_t processes. Thus, in some cases this causes the estimated prediction intervals that are based on the $\hat{\kappa}_t$ indices to look different than those prediction intervals that would have followed from the true underlying κ_t processes. The aforementioned contributes to the prediction uncertainty.

Table 5: The percentage of times a certain ARIMA(p, d, q) model with drift is selected out of 100,000 simulations for the as random walk with drift simulated $\kappa_t^{(1)}$ indices, and for the estimated $\hat{\kappa}_t$ indices according to AIC, and BIC, respectively. The setting where the central death rates are constructed in a reverse manner according to a bivariate Lee-Carter model.

	Fitted model for simulated κ_t (AIC/BIC)	Fitted model for estimated $\hat{\kappa}_t$ (AIC/BIC)
ARIMA$(0, 1, 0)$	76.6% / 91.6%	70.7% / 85.6%
ARIMA$(1, 1, 1)$	1.2% / 0.1%	1.1 % / 0.1%
ARIMA$(1, 1, 0)$	6.3% / 2.8%	7.1 % / 3.5%
ARIMA$(0, 1, 1)$	10.6% / 5.1%	12.0 % / 6.3%
Other ARIMA(p, d, q)	5.3% / 0.4%	9.1 % / 4.5%

6 Conclusion

This dissertation performed a thorough analysis on the parameter uncertainty in the Lee-Carter model. We have seen that the Lee-Carter model can be regarded as a regression setting with no observable quantities on the right hand side of the model. To obtain the model parameters the SVD is being applied to a matrix containing the central death rates. The fitted values that follow from this procedure are often regarded as if they are known quantities. Yet, this simplification does not represent the fact that the fitted values, in particular the κ values which represent the improvement in the mortality over the years, are prone to estimation uncertainty. Our aim was to shed a light on this problem, and to identify the impact the parameter uncertainty may have on the prediction uncertainty and thus risk management. In order to investigate this, a simulation study has been developed. The simulation study enabled us to construct the central death rates in a reverse manner under various settings. This also led us to link the parameter uncertainty problem in the Lee-Carter model with the theory on errors-in-variables models, as these are the regression models that elucidate the measurement errors in the independent variables. In addition, the simulation study empowered us to verify how the parameter uncertainty affects the time series techniques for selecting the right ARIMA model for the underlying simulated processes.

Our simulation study unveiled that in all simulation settings there is a noticeable degree of estimation error when using the Lee-Carter model. However, in the setting where the κ_t process is simulated as a random walk with drift the estimated drift parameter is not biased towards zero. In addition, the measurement error variance was small. These observations were not in line with the magnitude of bias that we expected when formulating the hypothesis that followed from the theory on errors-in-variables models. Moreover, it has been demonstrated how the estimation error affects the time series techniques for identifying the right ARIMA model. In the cases where the κ_t processes

are generated as a random walk without and or with drift the estimation error does not cause the time series techniques to wrongly identify the true underlying κ_t process. Therefore, in these settings the parameter uncertainty seems to have a limited effect on the identification procedure of the κ_t process. However, in the settings where the errors are simulated using a Student's t-distribution with 4 degrees of freedom, and in the setting where the central death rates are constructed using a bivariate Lee-Carter model, the estimation error plays a bigger role in the identification. We have seen that under these circumstances, the time series techniques fail more frequently to identify the true underlying κ_t processes. In these settings we must conclude that the parameter uncertainty is not negligible when producing predictions, and thus for risk management. Therefore, this dissertation hopes to raise the awareness once more that the parameter uncertainty should not be neglected or forgotten when forecasting mortality rates.

It has not been the objective of this dissertation to solve the problem for the parameter uncertainty that arises when using the Lee-Carter model. Therefore, further research could focus on finding a suitable solution for incorporating the parameter uncertainty when producing predictions. Nonetheless, the above findings have implications for actuaries who use the Lee-Carter model in practice. Our findings indicate that neglecting the parameter uncertainty can cause researchers such as Kleinow and Richards (2017), who attempt to find the best fitting ARIMA model for the estimated κ_t index, to model the wrong hidden process very accurately. Therefore, in view of the solvency II requirements, we advice insurers to reconsider how they structure their buffer capital in order to reduce the risk of insolvency instead of purely trying to find the best fitting model for forecasting the κ_t index.

Lastly, despite the shortcomings of the Lee-Carter model, this does not take away that the Lee-Carter model is an extremely practical and parsimonious tool for forecasting mortality rates. In essence, "all models are wrong, but some are useful" — George E.P. Box.

References

Börger, M. (2010). Deterministic shock vs. stochastic value-at-risk—an analysis of the solvency ii standard model approach to longevity risk. *Blätter der DGVFM*, 31(2):225–259.

Box, G. E., Jenkins, G. M., Reinsel, G. C., and Ljung, G. M. (2015). *Time series analysis: forecasting and control*. John Wiley & Sons.

Cairns, A. J., Blake, D., Dowd, K., Coughlan, G. D., Epstein, D., Ong, A., and Balevich, I. (2009). A quantitative comparison of stochastic mortality models using data from england and wales and the united states. *North American Actuarial Journal*, 13(1):1–35.

Cairns, A. J., Blake, D., Dowd, K., and Kessler, A. (2016). Phantoms never die: living with unreliable population data. *Journal of the Royal Statistical Series A*.

Gerber, H. U. (2013). *Life insurance mathematics*. Springer Science & Business Media.

Girosi, F. and King, G. (2007). Understanding the lee-carter mortality forecasting method. *Gking. Harvard. Edu*.

Girosi, F. and King, G. (2008). *Demographic forecasting*. Princeton University Press.

Golub, G. H. and Reinsch, C. (1971). Singular value decomposition and least squares solutions. In *Linear Algebra*, pages 134–151. Springer.

Green, W. H. (2012). Econometric analysis (seventh edition).

Gutterman, S. and Vanderhoof, I. T. (1998). Forecasting changes in mortality: a search for a law of causes and effects. *North American Actuarial Journal*, 2(4):135–138.

Hausman, J. (2001). Mismeasured variables in econometric analysis: problems from the right and problems from the left. *Journal of Economic perspectives*, 15(4):57–67.

Keilman, N. and Pham, D. Q. (2006). Prediction intervals for lee-carter-based mortality forecasts. In *European Population Conference*, pages 21–24.

Kleinow, T. and Richards, S. J. (2017). Parameter risk in time-series mortality forecasts. *Scandinavian Actuarial Journal*, 2017(9):804–828.

Koissi, M. C., Shapiro, A., and Högnäs, G. (2005). Fitting and forecasting mortality rates for nordic countries using the lee-carter method. *Actuarial Research Clearing House*, 1:21.

Koissi, M.-C., Shapiro, A. F., and Högnäs, G. (2006). Evaluating and extending the lee–carter model for mortality forecasting: Bootstrap confidence interval. *Insurance: Mathematics and Economics*, 38(1):1–20.

Lee, R. and Miller, T. (2001). Evaluating the performance of the lee-carter method for forecasting mortality. *Demography*, 38(4):537–549.

Lee, R. D. and Carter, L. R. (1992). Modeling and forecasting us mortality. *Journal of the American statistical association*, 87(419):659–671.

Li, J. S.-H., Hardy, M. R., and Tan, K. S. (2009). Uncertainty in mortality forecasting: an extension to the classical lee-carter approach. *ASTIN Bulletin: The Journal of the IAA*, 39(1):137–164.

Renshaw, A. E. and Haberman, S. (2003). Lee–carter mortality forecasting with age-specific enhancement. *Insurance: Mathematics and Economics*, 33(2):255–272.

Richards, S. and Currie, I. (2009). Longevity risk and annuity pricing with the lee-carter model. *British Actuarial Journal*, 15(2):317–343.

Tuljapurkar, S., Li, N., and Boe, C. (2000). A universal pattern of mortality decline in the g7 countries. *Nature*, 405(6788):789–792.

Wilmoth, J. R. (1993). Computational methods for fitting and extrapolating the lee-carter model of mortality change. Technical report, Technical report, Department of Demography, University of California, Berkeley.

Wilmoth, J. R. (1996). 13 mortality projections for japan. *Health and mortality among elderly populations*, page 266.

7 Appendices

7.1 Appendix A Definitions

Over the past years the jargon that is used within actuarial sciences quickly expanded. This appendix serves to review some frequently used terms and concepts within mortality modelling in general, and explains how the central death rates are established. The notation that is used closely follows the notation used in the book of Gerber (2013).

- ▷ Let $T(x)$ be a random variable that describes the future lifetime of an individual with age x. $T(x)$ has a probability distribution function: $G(t) = \Pr(T(x) \leq t), \quad t \geq 0$. Thus, the function $G(t)$ represents the probability that an individual aged x will die within t years, for any fixed t.

- ▷ Let ${}_tq_x$ denote the probability of an individual with age x dying within t years. Hence, ${}_tq_x = G(t)$.

- ▷ Let ${}_tp_x$ denote the probability of an individual with age x surviving at least t years. Moreover, this implies that ${}_tp_x = 1 - G(t) = 1 - {}_tq_x$.

- ▷ Let μ_{x+t} denote the force of mortality at age $x+t$, where $\mu_{x+t} = \frac{g(t)}{1-G(t)} = -\frac{d}{dt}\ln[1-G(t)]$. This shows that the force of mortality can also be represented as $\mu_{x+t} = -\frac{d}{dt}\ln[{}_tp_x]$. It describes that force of mortality can be seen as the the probability of someone with age x dying between time t and $t+\Delta t$, with $\Delta t \to 0$. Moreover, if we integrate the latter we find that ${}_tp_x = e^{-\int_0^t \mu_{x+s}ds}$.

- ▷ Let $m_{x,t}$ denote the central death rate for age x and time t. Furthermore, $m_{x,t} = \frac{D_{x,t}}{E_{x,t}}$. Here, $D_{x,t}$ denotes the number of death among people with an age between $[x, x+1)$ in the period $[t, t+1)$. $E_{x,t}$ denotes the average amount of people that lived with an age between $[x, x+1)$ in the period $[t, t+1)$. One can consider $m_{x,t}$ as an estimate for the force of mortality, and the relation between the central death rate and the force of mortality is roughly given by the expression: $m_{x,t} \approx \mu_{x,\frac{t+1}{2}}$. Hence, the central death rate is approximately equal to the force of mortality in the middle of a time period. The probability ${}_tq_x$ is then estimated as ${}_t\hat{q}_x = 1 - e^{-\frac{D_{x,t}}{E_{x,t}}} = 1 - e^{-m_{x,t}}$.

7.2 Appendix B Derivations

7.2.1 The decomposition

The decomposition of the forecast uncertainty for a random walk with drift of equation (3.14) according to Kleinow and Richards (2017) in detail:

$$\begin{aligned}
\mathbb{E}[(\hat{\kappa}_t(h) - \kappa_{t+h})^2] &= \mathbb{E}[(\hat{\kappa}_t(h) - \kappa_t(h) + \kappa_t(h) - \kappa_{t+h})^2] \\
&= \mathbb{E}[(\hat{\kappa}_t(h) - \kappa_t(h))^2] + 2\mathbb{E}[\hat{\kappa}_t(h) - \kappa_t(h)]\mathbb{E}[\kappa_t(h) - \kappa_{t+h}] + \mathbb{E}[(\kappa_t(h) - \kappa_{t+h})^2] \\
&= \mathbb{E}[(\hat{\kappa}_t(h) - \kappa_t(h))^2] + 2\mathbb{E}[\hat{\kappa}_t(h) - \kappa_t(h)]\mathbb{E}[-\sum_{j=1}^{h} \epsilon_{t+j}] + \mathbb{E}[(\kappa_t(h) - \kappa_{t+h})^2] \\
&= \underbrace{\mathbb{E}[(\hat{\kappa}_t(h) - \kappa_t(h))^2]}_{\text{parameter uncertainty}} + \underbrace{\mathbb{E}[(\kappa_t(h) - \kappa_{t+h})^2]}_{\text{volatility}} \\
&= \underbrace{\frac{h}{t-1} h\sigma_\epsilon^2}_{\text{parameter uncertainty}} + \underbrace{h\sigma_\epsilon^2}_{\text{volatility}}
\end{aligned}$$

To arrive at this expression, Kleinow and Richards (2017) made use of the conditional independence of $\hat{\kappa}_t$ and κ_{t+h} given κ_t. This independence follows from the fact that $\hat{\kappa}_t$ is a function of error terms $\{\epsilon_1, \ldots, \epsilon_t\}$ whereas κ_{t+h} depends on $\{\epsilon_{t+1}, \ldots, \epsilon_{t+h}\}$. Moreover, note that in the first step a telescope trick is being applied, which is followed by expanding the expression in the second step. In the third step one can make use of equations (3.8) and (3.9) to substitute the expression within the expectation operator. The expression in the fourth step can still further be decomposed into the final result. However, in order to improve the readability of the derivation, those calculations are shown separately.

First we calculate the expectation of $\hat{\kappa}_t(h)$, which yields:

$$\begin{aligned}
\mathbb{E}[\hat{\kappa}_t(h)] &= \mathbb{E}[\kappa_t + h\hat{d}] \\
&= \kappa_t + h\mathbb{E}[\hat{d}] \\
&= \kappa_t + hd \\
&= \kappa_t(h)
\end{aligned}$$

to arrive Kleinow and Richards (2017) made use of expressions (3.9) and (3.11), and the fact that $\hat{d}$ from (3.10) is an unbiased estimator that coincides with the maximum likelihood estimator for the mean of the normal distribution. This in addition implies that:

$$\begin{aligned}
\mathbb{E}[\hat{\kappa}_t(h) - \kappa_t(h)] &= \mathbb{E}[\hat{\kappa}_t(h)] - \mathbb{E}[\kappa_t(h)] \\
&= \kappa_t(h) - \kappa_t(h) \\
&= 0
\end{aligned}$$

Thus, for the parameter uncertainty component we have:

$$\begin{aligned}\underbrace{\mathbb{E}[(\hat{\kappa}_t(h) - \kappa_t(h))^2]}_{\text{parameter uncertainty}} &= \mathbb{V}ar(\hat{\kappa}_t(h) + \underbrace{\mathbb{E}[(\hat{\kappa}_t(h) - \kappa_t(h))]^2}_{=0} \\ &= h^2 \mathbb{V}ar(\hat{d}) \\ &= h^2 \frac{\sigma_\epsilon^2}{t-1} \\ &= \frac{h}{t-1} h\sigma_\epsilon^2\end{aligned}$$

Moreover, again making use of (3.8), (3.9) and the fact that the error terms $\{\epsilon_{t+1}, \ldots, \epsilon_{t+h}\}$ are distributed with mean 0 and variance σ_ϵ^2, one finds that for the volatility component it holds that:

$$\begin{aligned}\underbrace{\mathbb{E}[(\kappa_t(h) - \kappa_{t+h})^2]}_{\text{volatility}} &= \mathbb{E}[(\sum_{j=1}^{h} \epsilon_{t+j})^2] \\ &= h\sigma_\epsilon^2\end{aligned}$$

However, as mentioned in Section 3.3, the derivation of Kleinow and Richards (2017) ignores the model risk and the fact that the κ_t index is estimated, and not directly observed.

7.2.2 The variance-covariance and correlation matrices

This extra section serves to demonstrate how we obtained the variance-covariance matrices and correlation matrices arising from of the estimation errors in our simulation design. In each simulation (N=100,000), we computed the bias or estimation errors arising from: $\alpha_x - \hat{\alpha}_x$, $\beta_x - \hat{\beta}_x$, and $\kappa_t - \hat{\kappa}_t$. In addition, the joint estimation errors arising from $(\alpha_x, \beta_x, \kappa_t) - (\hat{\alpha}_x, \hat{\beta}_x, \hat{\kappa}_t)$ are computed, where the true and estimated parameters are stacked in a vector. Then, for each simulation we observe a vector of estimation errors for each corresponding parameter. Hence, we obtain a matrix E_α with dimension $100,000 \times 51$ capturing the estimation errors for the α_x parameter in each simulation, such that:

$$E_\alpha = \begin{pmatrix} \alpha_1 - \hat{\alpha}_1 & \alpha_2 - \hat{\alpha}_2 & \ldots & \alpha_{51} - \hat{\alpha}_{51} \\ \alpha_1 - \hat{\alpha}_1 & \alpha_2 - \hat{\alpha}_2 & \ldots & \alpha_{51} - \hat{\alpha}_{51} \\ \vdots & \vdots & \ddots & \vdots \\ \alpha_1 - \hat{\alpha}_1 & \alpha_2 - \hat{\alpha}_2 & \ldots & \alpha_{51} - \hat{\alpha}_{51} \end{pmatrix}. \tag{7.1}$$

Similarly, we obtain a matrix E_β with dimension $100,000 \times 51$ capturing the estimation errors for the β_x parameter in each simulation, such that:

$$E_\beta = \begin{pmatrix} \beta_1 - \hat{\beta}_1 & \beta_2 - \hat{\beta}_2 & \dots & \beta_{51} - \hat{\beta}_{51} \\ \beta_1 - \hat{\beta}_1 & \beta_2 - \hat{\beta}_2 & \dots & \beta_{51} - \hat{\beta}_{51} \\ \vdots & \vdots & \ddots & \vdots \\ \beta_1 - \hat{\beta}_1 & \beta_2 - \hat{\beta}_2 & \dots & \beta_{51} - \hat{\beta}_{51} \end{pmatrix}. \tag{7.2}$$

Likewise, we obtain a matrix E_κ with dimension $100,000 \times 43$ capturing the estimation errors for the κ_t parameter in each simulation, such that:

$$E_\kappa = \begin{pmatrix} \kappa_1 - \hat{\kappa}_1 & \kappa_2 - \hat{\kappa}_2 & \dots & \kappa_{43} - \hat{\kappa}_{43} \\ \kappa_1 - \hat{\kappa}_1 & \kappa_2 - \hat{\kappa}_2 & \dots & \kappa_{43} - \hat{\kappa}_{43} \\ \vdots & \vdots & \ddots & \vdots \\ \kappa_1 - \hat{\kappa}_1 & \kappa_2 - \hat{\kappa}_2 & \dots & \kappa_{43} - \hat{\kappa}_{43} \end{pmatrix}. \tag{7.3}$$

Lastly, we obtain a matrix E_{joint} with dimension $100,000 \times 145$ capturing the estimation errors for all parameters jointly in each simulation, such that:

$$E_{\text{joint}} = \begin{pmatrix} \alpha_1 - \hat{\alpha}_1 & \dots & \alpha_{51} - \hat{\alpha}_{51} & \beta_1 - \hat{\beta}_1 & \dots & \beta_{51} - \hat{\beta}_{51} & \kappa_1 - \hat{\kappa}_1 & \dots & \kappa_{43} - \hat{\kappa}_{43} \\ \alpha_1 - \hat{\alpha}_1 & \dots & \alpha_{51} - \hat{\alpha}_{51} & \beta_1 - \hat{\beta}_1 & \dots & \beta_{51} - \hat{\beta}_{51} & \kappa_1 - \hat{\kappa}_1 & \dots & \kappa_{43} - \hat{\kappa}_{43} \\ \vdots & \vdots & \ddots & \vdots & & & & & \\ \alpha_1 - \hat{\alpha}_1 & \dots & \alpha_{51} - \hat{\alpha}_{51} & \beta_1 - \hat{\beta}_1 & \dots & \beta_{51} - \hat{\beta}_{51} & \kappa_1 - \hat{\kappa}_1 & \dots & \kappa_{43} - \hat{\kappa}_{43} \end{pmatrix}. \tag{7.4}$$

Consecutively, one can calculate the variance, covariance and correlation between the columns of matrix E_j with $j = \{\alpha, \beta, \kappa, \text{joint}\}$. Let e_i denote a certain column of matrix E_j for $i = 1, \dots, M$. The four variance-covariance matrices and four correlation matrices with dimensions 51×51, 51×51, 43×43, and 145×145 (depending on the dimension of E_j) are then in the form:

$$cov(E_j) = \begin{pmatrix} var(e_1) & cov(e_1, e_2) & \dots & cov(e_1, e_M) \\ cov(e_2, e_1) & var(e_2) & \dots & cov(e_2, e_M) \\ \vdots & \vdots & \ddots & \vdots \\ cov(e_M, e_2) & cov(e_M, e_2) & \dots & var(e_M) \end{pmatrix} \tag{7.5}$$

$$cor(E_j) = \begin{pmatrix} 1 & cor(e_1, e_2) & \dots & cor(e_1, e_M) \\ cor(e_2, e_1) & 1 & \dots & cor(e_2, e_M) \\ \vdots & \vdots & \ddots & \vdots \\ cor(e_M, e_2) & cor(e_M, e_2) & \dots & 1 \end{pmatrix} \tag{7.6}$$

7.3 Appendix C Tables

Table 6: Estimates of α_x from Figure 2, with the identification constraints $\sum_{x=1}^{X} \beta_x = 1$, and $\sum_{t=1}^{T} \kappa_t = 0$.

Age	α	Age	α	Age	α	Age	α	Age	α
50	-5.4557483	61	-4.3094604	72	-3.2168840	83	-2.1562231	94	-1.1736403
51	-5.3500166	62	-4.2088233	73	-3.1131741	84	-2.0621854	95	-1.0981243
52	-5.2506859	63	-4.1089610	74	-3.0164564	85	-1.9692951	96	-1.0385078
53	-5.1443804	64	-4.0123752	75	-2.9172968	86	-1.8776904	97	-0.9685334
54	-5.0343988	65	-3.9053209	76	-2.8263790	87	-1.7848464	98	-0.9014365
55	-4.9350316	66	-3.8070708	77	-2.7265975	88	-1.7003705	99	-0.8372900
56	-4.8330265	67	-3.7131193	78	-2.6385178	89	-1.6133548	100	-0.7761462
57	-4.7272061	68	-3.6078609	79	-2.5438698	90	-1.5242579		
58	-4.6200838	69	-3.5110044	80	-2.4463753	91	-1.4338564		
59	-4.5198444	70	-3.4114910	81	-2.3447565	92	-1.3519362		
60	-4.4177894	71	-3.3153989	82	-2.2567348	93	-1.2550508		

Table 7: Estimates of β_x from Figure 2, with the identification constraints $\sum_{x=1}^{X} \beta_x = 1$, and $\sum_{t=1}^{T} \kappa_t = 0$.

Age	β	Age	β	Age	β	Age	β	Age	β
50	0.02089169	61	0.02620634	72	0.02592137	83	0.01829314	94	0.00741386
51	0.02220350	62	0.02650630	73	0.02566892	84	0.01755514	95	0.00772677
52	0.02207057	63	0.02645957	74	0.02582740	85	0.01647895	96	0.00588205
53	0.02307068	64	0.02685588	75	0.02527313	86	0.01540129	97	0.00508461
54	0.02354906	65	0.02719305	76	0.02456566	87	0.01461610	98	0.00432892
55	0.02468329	66	0.02743161	77	0.02381726	88	0.01353723	99	0.00361735
56	0.02452023	67	0.02695781	78	0.02297009	89	0.01265103	100	0.00295296
57	0.02459090	68	0.02718725	79	0.02217653	90	0.01116390		
58	0.02534749	69	0.02699674	80	0.02140189	91	0.01003658		
59	0.02593778	70	0.02672387	81	0.02045172	92	0.00923629		
60	0.02628338	71	0.02663403	82	0.01887321	93	0.00877567		

Table 8: The percentage of times a certain ARIMA(p, d, q) model with drift is selected out of 100,000 simulations for the as ARIMA(1,1,0) with drift simulated κ_t indices, and for the estimated $\hat{\kappa}_t$ indices according to AIC, and BIC, respectively.

	Fitted model for simulated κ_t (AIC/BIC)	Fitted model for estimated $\hat{\kappa}_t$ (AIC/BIC)
ARIMA$(1, 1, 0)$	40.2% / 39.7%	39.1% / 39.1%
ARIMA$(0, 1, 0)$	14.0% / 28.0%	12.1% / 25.1%
ARIMA$(1, 1, 1)$	2.8% / 0.8%	2.5% / 0.7%
ARIMA$(0, 1, 1)$	33.7% / 29.8%	36.9% / 33.3%
Other ARIMA(p, d, q)	9.3% / 1.7%	9.4% / 1.8%

7.4 Appendix D Figures

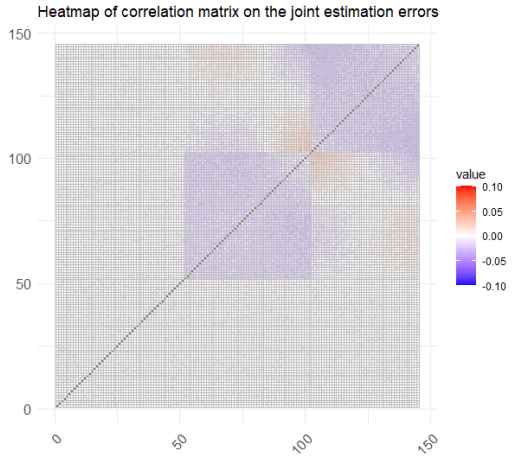

Figure 13: Heatmap of the correlation matrix on the estimation errors for all parameters jointly. The case where the κ_t index is simulated as an ARIMA(1,1,0) with drift.

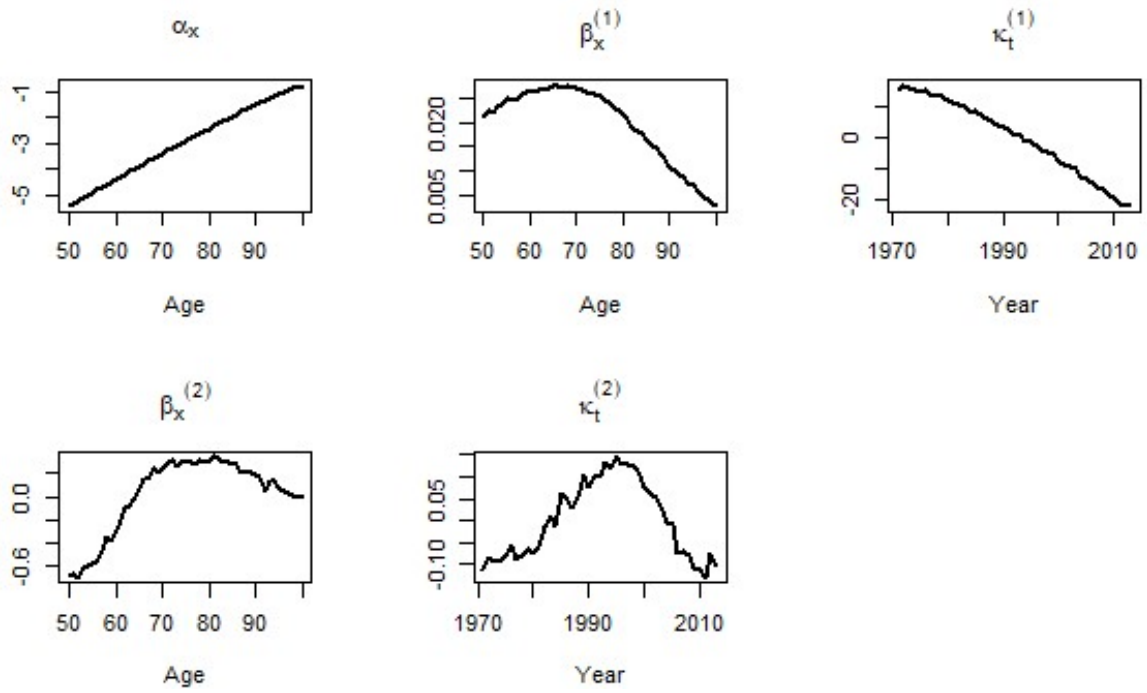

Figure 14: Parameter estimates for the bivariate Lee-Carter model fitted to the EWA male data set for the ages 50-100 over the period 1971-2013 ($x = 1, \ldots, 51$; $t = 1, \ldots, 43$).