https://doi.org/10.15388/vu.thesis.155
https://orcid.org/0000-0002-8510-8533

VILNIUS UNIVERSITY

Rokas
GYLYS

Application of the Lee-Carter Mortality Projection Model and Its Modifications in the Modelling of an Insurance Company's Solvency Capital

DOCTORAL DISSERTATION

Natural Sciences,
Mathematics (N 001)

VILNIUS 2021

This dissertation was written between 2016 and 2020 at Vilnius University.

The research was funded by grant No. S-MIP-20-16 from the Research Council of Lithuania.

Academic supervisor: Prof. Habil. Dr. Jonas Šiaulys (Vilnius University, natural sciences, mathematics – N 001).

Academic consultant: Prof. Habil. Dr. Remigijus Leipus (Vilnius university, natural sciences, mathematics – N 001).

https://doi.org/10.15388/vu.thesis.155
https://orcid.org/0000-0002-8510-8533

VILNIAUS UNIVERSITETAS

Rokas
GYLYS

Lee ir Carterio mirtingumo prognozavimo modelis ir jo modifikacijos draudimo įmonės mokumui skaičiuoti

DAKTARO DISERTACIJA

Gamtos mokslai,
Matematika (N 001)

VILNIUS 2021

Disertacija rengta 2016-2020 metais Vilniaus universitete.

Mokslinius tyrimus rėmė Lietuvos mokslo taryba, projekto Nr. S-MIP-20-16.

Mokslinis vadovas – prof. habil. dr. Jonas Šiaulys (Vilniaus universitetas, gamtos mokslai, matematika – N 001).

Mokslinis konsultantas – prof. habil. dr. Remigijus Leipus (Vilniaus universitetas, gamtos mokslai, matematika – N 001).

CONTENTS

LIST OF ABBREVIATIONS

Abbreviation	Explanation
APC	Age-Period-Cohort (Model)
CBD	Cairns-Blake-Dowd (Model)
DLM	Dynamic Linear Model
EIOPA	European Insurance and Occupational Pensions Authority
EU	European Union
FFBS	Forward Filtering Backward Sampling
GLM	Generalized Linear Model
IRLS	Iteratively Weighted Least Squares
MCMC	Markov Chain Monte Carlo
OLS	Ordinary Least Squares
RWD	Random Walk with Drift
SCR	Solvency Capital Requirement
SSM	State Space Model
SSM with switching	State Space Model with Regime Switching
SVD	Singular Value Decomposition
Var(X)	The variance of random variable X
VaR	Value-at-Risk

INTRODUCTION

Solvency II was a major driver for European insurers to upgrade their quantitative risk assessment models both for the assessment of the regulatory solvency position and internal risk management needs. Major European insurance groups have developed and implemented full or partial internal models, which value risks from the perspective of a company-specific risk profile.

As demonstrated in Table 1, the majority of the European insurers, predominantly small and medium-sized firms, calculate the regulatory solvency capital requirement using the standard methods and stress scenarios prescribed by the legislation (Standard Formula). Standard Formula prescribes that, firstly, an insurer's regulatory Solvency Capital Requirement (SCR) is estimated on an individual risk module/sub-module level. Secondly, the individual risk assessments at the risk module and sub-module level are aggregated to the total SCR by taking into account the diversification between the risks.

Table 1. The number of Solvency II submissions (solo undertakings) in 2017 by the method used for calculation of the Solvency Capital Requirement (SCR). Data source: European Insurance and Occupational Pensions Authority (EIOPA).

Type of insurer	**Internal Model**	**Partial Internal Model**	**Standard Formula**	**Total**
Life insurance	21	29	545	595
Non-life insurance	37	42	1,519	1,598
Reinsurance	15	4	253	272
Composite insurance	8	30	368	406
Total	81	105	2,685	2,871

In this dissertation, we deal with SCR for the mortality risk sub-module, which, according to the Standard Formula, is assessed by calculating the amount of loss of insurer's available Solvency II regulatory capital (Basic Own Funds) resulting from an immediate and permanent 15% increase in mortality probabilities used for the calculation of the Solvency II technical provisions. SCR for mortality risk is aggregated with SCR's for other life underwriting risks, such as lapse risk, longevity risk, expense risk, and life

catastrophic risk, to derive the SCR for life underwriting risk module, which is used as an input for calculations of the overall SCR.

Solvency II Directive requires that SCR is calibrated to the Value-at-Risk (VaR) of the basic own funds subject to a confidence level of 99.5% over a one-year period. However, as argued by Richards et al. [69]:

. . . some risks do not fit naturally into a one-year VaR framework and it would be excessively dogmatic to insist that the longevity trend risk can only be measured over a one-year horizon.

It seems that this approach is shared by the European Insurance and Occupational Pensions Authority (EIOPA), who used a run-off approach for the calibration of the uniform mortality shocks for mortality and longevity risks during the review of the Standard Formula [23]. The run-off VaR methodology is well-established, but to perform solvency calculations strictly with the requirements of Solvency II Directive, in this dissertation, we develop an alternative one-year VaR methodology and provide a comparison of the results derived using both VaR approaches.

In its essence, the Standard Formula is designed to suit the risk profile of the "average" European insurer. However, insurers in the European Union (EU) are highly heterogeneous, and the "average" insurer is difficult to define. For example, life insurers may be exposed to different levels of mortality risk due to the differences in products offered, distribution channels used, term to maturity of insurance portfolios, prevailing policy terms and conditions, the volatility of mortality in a country where the insurer operates, and several other factors. Thus, the Standard Formula mortality risk assessment may not reflect the actual risk stemming from a particular insurer's portfolio.

The calculation of mortality VaR requires simulations from a stochastic mortality model as an input. A model proposed by Lee and Carter [53] started a new generation of the extrapolative mortality projection methods, and in its original or modified form is widely used both in academic research and in practice. Several improvements of the Lee-Carter model have been proposed, both from the point of view of model specification and the statistical estimation of parameters. In this dissertation, we shall explore three different modifications of the Lee-Carter model: the classical model, Poisson regression version, firstly developed by Brouhns et al. [10], and state space Lee-Carter model and show that different fitting methods can result in different confidence intervals of mortality projections and, consequently, in different VaRs.

In particular, the classical stochastic Lee-Carter model, which was used for calibrating the Standard Formula, assumes a constant mortality trend and constant variance over the fitting and forecasting periods. In the real world,

this is not always the case. For example, significant one-off fluctuations in mortality can occur due to pandemics, wars, and other events. Many Central and Eastern European countries, after the transformation to free market economies, experienced a major one-time change in socio-economic development, which had a major impact on mortality development. In addition, the classical Lee-Carter model does not allow for the uncertainty of parameters, which is an important source of uncertainty in the projections. The alternatives of the classical Lee-Carter model can be used to provide the required flexibility to overcome these shortcomings.

State space models (SSM) offer a natural opportunity to extend the Lee-Carter model both in terms of alternative fitting procedure and additional model flexibility. We use two types of SSM: dynamic linear models (DLM), which have already been applied in mortality modeling, and SSM with regime switching (SSM with switching), which so far were mainly used for modeling economic processes. Such models are often referred to as multi-process models and are DLMs conditionally on a sequence of indicator variables, which represent varying regimes. In the context of mortality modeling, switching between stable and volatile socio-economic development conditions can positively or negatively affect mortality development. In this dissertation, we use the Gibbs sampler Markov Chain Monte Carlo (MCMC) method to estimate parameters of SSMs.

The above sets the background for the key objectives of this dissertation:

- Develop a mathematical stochastic mortality projection model which allows for changes in mortality trend, parameter uncertainty, and varying levels of volatility.
- Develop detailed calculation methods of one-year VaR for mortality risk, using one-year and run-off approaches, for application by insurance companies for solvency assessment and risk management.
- Perform a mortality analysis of Lithuania and Sweden and check the calibration of the Standard Formula in the area of mortality risk. Highlight the key areas where the Standard Formula mortality stress parameter deviates from the assessed VaR.

The theses of this dissertation can be summarized as follows:

- State space models, such as dynamic linear models and state space models with regime switching, and the Gibbs sampler fitting method provide a good alternative to modeling changes of mortality trends and switching of variance regimes.

- Goodness of fit of alternative Lee-Carter model modifications, such as classical, Poisson bilinear regression, and state space, varies depending on the model's flexibility and the source data used.
- Confidence intervals of mortality projections, derived using various Lee-Carter model modifications, significantly vary due to different goodness of fit, allowance for the uncertainty of parameters, and distribution applied for modelling of errors.
- One-year VaR approach can be used for the assessment of mortality risk as an alternative to run-off VaR. Depending on the parametrization, the results of VaR calculations using the two approaches are significantly different.
- VaR that was calculated by taking into account specific features of the country's mortality, insurance product, and policy terms is significantly different from the VaR calibrated using "average" European experience.

This dissertation follows the following structure. Section 1 discusses the research performed so far in the area of mortality VaR and stochastic mortality modeling, including the application of state space models. Section 2 discusses in detail the three modifications of the Lee-Carter model and their fitting methods. We provide the details of the singular value decomposition method used in the classical Lee-Carter model, discussed the estimation of overdispersed Poisson regression with quasi-likelihood functions, and introduce the Gibbs sampler, which is the Bayesian Markov Chain Monte Carlo method used for fitting SSMs. The output from the Gibbs sampler also allows us to take into account parameter uncertainty in performing the simulations. At the end of the section, a brief overview of few other alternative stochastic mortality models is provided. In this section, we develop and describe several algorithms used for model fitting and simulation of projections. A summary of the algorithms is provided in Table 2. Algorithms No. 7-10 were specifically developed for the models proposed in this dissertation. Section 3 provides an overview of the data and the model fitting results. We provide and discuss the parameter estimates, their uncertainty, as well as provide model diagnostics statistics. For the comparison of alternative state space models, we calculate the marginal likelihood using sequential Monte Carlo methods (particle filter). Section 4 discusses the results of mortality projections and provides detailed VaR calculation methodology and calculation results. In this section, we develop two alternative VaR methodologies: run-off VaR, which considers the possible losses over the who

policy term, and one-year VaR, which considers the losses at the end of the first projection year. The section

Table 2. Summary of algorithms used/ developed in the dissertation.

No.	Name	Brief description
1.	Simulation using classical Lee-Carter model	The algorithm is used to simulate random mortality projections when the fitting is performed using classical Lee-Carter model.
2.	Quasi-likelihood IRLS	The iteratively weighted least squares (IRLS) algorithm is used to estimate parameters of quasi-likelihood GLM model.
3.	Quasi-Poisson Lee-Carter criss-cross ILRS algorithm	The algorithm allows us to estimate parameters of quasi-Poisson GLM model with the bilinear term. In particular, it is used to fit Poisson Lee-Carter model.
4.	Simulation using Poisson Lee-Carter model	The algorithm is used to simulate random mortality projections when the fitting is performed using Poisson Lee-Carter model.
5.	Kalman filter recursion	The classical recursive algorithm is used to estimate unobservable state parameter vector.
6.	Forward Filtering Backward Sampling	The algorithm used to sample unobservable state parameter vectors conditionally on estimated state space model parameters.
7.	Gibbs sampler for DLM Lee-Carter model	The algorithm implements the Gibbs sampler method for estimation of parameters of linear state space Lee-Carter model.
8.	Gibbs sampler for SSM Lee-Carter with switching	The algorithm implements the Gibbs sampler method for estimation of parameters of SSM Lee-Carter with switching.
9.	Simulation using state space Lee-Carter model	The algorithm is used to simulate random mortality projections when the fitting is performed using state space Lee-Carter model.
10.	Simulation of SSM Lee-Carter	The algorithm implements auxiliary particle filter to simulate state space Lee-Carter with

Continued table.

No.	Name	Brief description
	with switching log likelihood	switching model log likelihood conditional on the estimated collection of parameters.
11.	Simulation of log density of parameter posterior	The algorithm performs a simulation of log density of posterior of parameters fitted with Gibbs sampler.

is concluded with the results of VaR calculations using three Lee-Carter model modifications as fitted in Section 3, and comparisons with the Standard Formula calculation results. The dissertation is concluded with a discussion of the results and their implications for the modeling of insurers' solvency capital.

A significant part of work in preparing this dissertation was devoted to the development of the *R* software code, used for model fitting, simulation of mortality projections, and calculation of VaR. As the building blocks to this this code, in some cases we used packages *gnm* and *dlm*, which enabled us to fit a generalized linear model with the bi-linear term and to run the basic Kalman filter recursions, respectively. The rest of the code was developed originally for the models proposed and used in this dissertation. Specifically, we developed the code for the following:

- The classical Lee-Carter model fitting using SVD method.
- Simulation of mortality projections using the classical Lee-Carter model (Algorithm 1).
- The Poisson Lee-Carter model fitting using quasi Poisson method (Algorithm 3). Package *gnm* was used as a building block.
- Simulation of mortality projections using the Poisson Lee-Carter model (Algorithm 4).
- The state space Lee-Carter model fitting using Gibbs sampler (Algorithm 7). Package *dlm* was used as a building block for basic Kalman filter and Forward Filtering Backward Sampling (FFBS) recursions.
- The state space Lee-Carter with switching model fitting using Gibbs sampler (Algorithm 8). Package *dlm* was used as a building block for basic Kalman filter and FFBS recursions.

- Simulation of mortality projections using the state space Lee-Carter model, with or without switching (Algorithm 9).
- Simulation of log-likelihood, conditional on estimated parameters, of the state space Lee-Carter model, with or without switching (Algorithm 10).
- Simulation of log-density of posterior of parameters fitted with Gibbs sampler of the state space Lee-Carter model, with or without switching (Algorithm 11). The code uses codes of Algorithm 7 and Algorithm 8 as an input and consequently uses package *dlm* as a building block for basic Kalman filter and FFBS recursions.
- Taking simulated mortality rates as an input, calculation of run-off VaR for the single premium term life assurance policy, with level or decreasing benefits.
- Taking simulated mortality rates as an input, calculation of one-year VaR for the single premium term assurance life policy, with level or decreasing benefits.

In this dissertation, we use the following notation. Capital letters in bold are used to denote matrices, e.g. $\boldsymbol{A}$. Lowercase letters in bold are used to denote vectors, e.g. $\boldsymbol{a}$. By default, vectors are assumed to be column vectors. If row vectors are used, they are denoted as a transpose of a column vector, e.g. $\boldsymbol{a}'$.

1. OVERVIEW OF RESEARCH

In this section we provide an overview of the research so far in the areas of VaR for mortality risk, stochastic generation of mortality probabilities, and use of state space models in the mortality modeling.

1.1. Mortality VaR and calibration of the Standard Formula

We start the overview with the description of the analysis performed by EIOPA, which sets the basis for the stress parameters used in the Standard Formula.

EIOPA calibrates VaR using temporary life expectancies:

$$e_{x:n,t} = \frac{1}{2} + \sum_{k=1}^{N-1} {}_kp_{x,t} + {}_Np_{x,t},$$

where ${}_kp_{x,t}$ is k year survival probability of life aged $x \in \{1, \dots, N\}$ at time t. By setting stressed temporary life expectancy to the life expectancy derived from a stochastic mortality model, appropriate mortality shock level is determined [23].

By using this approach, EIOPA implicitly assumes that benefits payable under life assurance policy decrease with time. Decreasing benefits are common in life assurance products linked to mortgages or other credit instruments. In addition, mortality sum at risk is decreasing with policy duration for some risk and savings products (e.g., traditional endowment insurance) where the total benefit payable on death or survival is fixed, and the insurer is able to recoup part of the losses by reversing the accumulated savings amount. However, a significant part of life assurance products has fixed sums assured payable on death. In this dissertation we examine the effect on VaR of both benefit formulas: level (fixed) sum assured and sum assured which is decreasing linearly with time.

EIOPA calibrated both mortality and longevity modules at the same time; therefore, the data for ages 40-120 years was used [23]. For model fitting EIOPA used data of the major European countries, covering the relatively recent time period, which excluded major socio-economic development disturbances such as wars, epidemics, etc. Historically, human mortality behavior was much less steady, and the impact of the inclusion of less steady periods of mortality development is also investigated in this dissertation.

There are two main approaches to mortality VaR calculation: run-off VaR and one-year VaR. Run-off VaR, the approach also used by EIOPA, takes into

account the volatility of cash flows until policy expirity. Run-off VaR is based on the basic principles, following from the definition of VaR, see [58], and the key modeling task is the derivation of realistic mortality projections.

One-year mortality VaR, in contrast, poses more methodological challenges and was subject to academic research. One-year VaR aims to replicate the behavior of the solvency balance sheet during the valuation year and its changes due to mortality fluctuations. Several models have been proposed to model mortality/longevity risk using the one-year VaR approach, where the key difficulty is to model the risk of variation of technical provisions in one year's time after the valuation date. Börger et al. [8] and Plat [64] developed models which explicitly model the mortality trend risk which is used as the proxy to capture the risk of variation of technical provisions. For example, Plat stochastically models mortality trend parameter $\lambda_{x,t}$ of life aged x at time t, represented by the formula:

$$\lambda_{x,t} = \kappa_t^1 + \kappa_t^2(\bar{x} - x) + \kappa_t^2(\bar{x} - x)^+,$$

where $(\bar{x} - x)^+ = \max(\bar{x} - x, 0)$ and κ_t^1, κ_t^2, κ_t^3 are estimated parameters. Stochastic volatility of mortality trend parameter $\lambda_{x,t}$ drives the uncertainty of projected mortality rates, which can be used to approximate the basis of calculation of technical provision in one-year VaR calculations.

Richards et al. [69] took into account the risk of changes in technical provisions by simulating one year's portfolio mortality and subsequently refitting the stochastic mortality model, for each simulation scenario, based on the results of the first-year simulations. The model operates as follows:

- Simulate first projection year mortality rates.
- The projected rates are treated as observations and are used to augment the data set. The model is refitted using the augmented data set.
- Project the mortality rates for the second and further years using the refitted model output.

Such model avoids assumptions about the specific parametric form of mortality trend and, therefore, is more simple than the models by Börger et al. [8] and Plat [64]. However, it's results very much depends on the size of the data set: the longer history is available, the lower sensitivity of the estimated parameters to the additional observation set.

A similar approach to Richards et al. [69] was used by Jarner and Møller [45] who proposed a model designed specifically for the Danish system of

reserving for longevity risks. Olivieri and Pitacco [60] developed a model, which included a Bayesian procedure for updating the parameters of projected distributions of deaths, and formulated conditions for few alternative definitions of solvency capital rules, however, their model did not necessarily strictly follow a one-year VaR approach. Munroe et al. [59] developed a one-year VaR model for non-life insurance risks.

1.2. Simulation of mortality rates

Calculation of mortality VaR requires the simulations from a stochastic mortality model as an input. We note that there are many alternative stochastic mortality models, see [7], [27] for an overview. In this summary, we focus on the evolution of the extrapolatory stochastic mortality models used in this dissertation.

We also note that with the purpose to avoid repetition we keep this overview brief as full details of the models are provided in Section 2.

A model proposed by Lee and Carter [53], see Subsection 2.1. for details, started a new generation of the extrapolative mortality projection methods, and in its original or modified form is widely used both in academic research and in practice. Several improvements of the Lee-Carter model have been proposed, both from the point of view of model specification and the statistical estimation of parameters. Lee and Miller [54] proposed an alternative method of the second stage re-estimation of time varying index parameters. Booth et al. [6], Renshaw and Haberman [66] analyzed the effect of including higher order terms of the singular value decomposition with the purpose of obtaining a better fit. Kleinov and Richards [50] experimented with different specifications of time series models for fitting time varying index, which in the classical Lee-Carter model is assumed to follow a simple Random Walk with Drift (RWD). Girosi and King [36] developed a Bayesian framework for fitting Lee-Carter model parameters. Ignatavičiūtė et al. [44] applied the classical Lee-Carter model with the modification for higher order terms to the Lithuanian data.

A significant development in stochastic mortality models was the application of the generalized linear model framework for more realistic variance modeling, especially in high ages. Poisson regression version of the Lee-Carter model was firstly developed by Brouhns et al. [10], see Subsection 2.2. for details. In comparison with the classical Lee-Carter, where it is assumed that errors are additive on the logarithm of mortality rates, the Poisson Lee-Carter model assumes that a number of deaths is distributed

according to the Poisson law. Thus, for example, we can model a higher variance of mortality rates in high ages, where the expected mortality rates are higher.

The matter of parameter uncertainty in simulated projections was initially addressed by Lee and Carter [53] who identified time varying index as the key source of uncertainty and suggested simple analytical and bootstrap procedures to estimate the uncertainty of other parameters. Lee and Miller [54] suggested a simple method how to take into account the uncertainty of the drift parameter of time varying index.

Brouhns et al. [11] developed a bootstrap method to assess the effect of uncertainty of parameter estimates of the Poisson Lee-Carter model. The bootstrap samples were obtained by applying Poisson noise to the observed number of deaths. For each bootstrap sample model is refitted and parameters are re-estimated.

The alternative bootstrap method was developed by Koissi et al. [52] who sampled with replacements deviance residuals derived from the initial fitting of the Lee-Carter model. Deviance residuals are then converted to the numbers of death, which can be used to refit the model and to re-estimate the parameters.

Renshaw and Haberman [68] and Li [55] performed detailed comparative studies of various bootstrap approaches in simulations.

Czado et al. [19] developed an alternative Bayesian approach for forecasting mortality rates, where the MCMC fitting method allowed to estimate parameter uncertainty. In contrast to the state space models discussed below, Czado et al. [19] did not use Kalman filter and used the Metropolis-Hastings sampling algorithm instead.

1.3. State space models

State space models (SSM) can be traced back to a highly influential paper by Kalman [46]. Although initially these methods were applied predominantly in engineering and signal processing, over time it became a recognized statistical modeling technique.

The general form of state space model can be defined by the following equations:

$$\overline{\boldsymbol{y}}_t \sim p(\overline{\boldsymbol{y}}_t \mid \overline{\boldsymbol{x}}_t),$$

$$\overline{\boldsymbol{x}}_t \sim p(\overline{\boldsymbol{x}}_t \mid \overline{\boldsymbol{x}}_{t-1}),$$

where vector $\overline{\boldsymbol{y}}_t$ at time $t = 1,2, \ldots$ is observed, vector $\overline{\boldsymbol{x}}_t$ is unobserved state parameter vector, and p is arbitrary distribution, and $\overline{\boldsymbol{x}}_0$ is constant.

The advantage of state space models is their high flexibility and relative simplicity of calculations. Many examples of the application of the SSMs are presented by West and Harrison [77], Durbin and Koopman [22], Särkkä [70], Shumway and Stoffer [71] among others. Forward Filtering Backward Sampling Algorithm, which is a prerequisite for the efficient application of SSMs for mortality studies, was developed by Carter and Kohn [14], Frühwirth-Schnatter [24], and de Jong and Shephard [21].

A linear state space model, often called Dynamic Linear Model (DLM), can be represented by the following two equations:

$$\boldsymbol{y}_t = \boldsymbol{Z}_t \boldsymbol{x}_t + \boldsymbol{\varepsilon}_t, \tag{1}$$

$$\boldsymbol{x}_t = \boldsymbol{U}_t \boldsymbol{x}_{t-1} + \boldsymbol{\eta}_t, \tag{2}$$

where vector $\boldsymbol{y}_t$ at time $t = 1,2, \ldots$ is observed, vector $\boldsymbol{x}_t$ is unobserved state parameter vector, $\boldsymbol{x}_0$ is constant, $\boldsymbol{\varepsilon}_t \sim N(\boldsymbol{0}, \boldsymbol{H}_t)$ and $\boldsymbol{\eta}_t \sim N(\boldsymbol{0}, \boldsymbol{Q}_t)$ are the Gaussian errors and $\boldsymbol{Z}_t$, $\boldsymbol{H}_t$, $\boldsymbol{U}_t$, $\boldsymbol{Q}_t$ are the parameter matrices. Thus, we can use state space representation to represent Lee-Carter model by modeling time varying index, including the processes driving its drift, with the state equation.

There were several examples of the application of SSM to fit the Lee-Carter model. For example, Pedroza [61] used DLM to model US mortality data. Her model was defined as follows:

$$\boldsymbol{y}_t = \boldsymbol{\alpha} + \boldsymbol{\beta}\kappa_t + \boldsymbol{\varepsilon}_t,$$

$$\kappa_t = \kappa_{t-1} + \mu + \eta_t,$$

where vector $\boldsymbol{y}_t$ at time $t = 1, 2, \ldots$ is observed vector by age group of log mortality rates, $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are parameter vectors with the same interpretation as in the classical Lee-Carter model, κ_t is unobserved time series process with the drift μ, κ_0 is constant, and $\boldsymbol{\varepsilon}_t$ and η_t are error terms.

A similar model was applied by Kogure and Kurachi [51] to model Japanese mortality and used the estimation results to develop a model for pricing longevity risk. Fung et al. [27] developed a practical methodology to estimate linear state space Lee-Cater model parameters using maximum likelihood and Bayesian approaches and illustrated the results using the Danish mortality data. In addition, Fung et al. [27] developed a non-linear stochastic volatility state space Lee-Carter model, which could be considered as an alternative to SSM with regime switching model used in this dissertation. The model is specified using the following equations:

$$\boldsymbol{y}_t = \boldsymbol{\alpha} + \boldsymbol{\beta}\kappa_t + \boldsymbol{\varepsilon}_t,$$

$$\kappa_t = \kappa_{t-1} + \mu + \eta_t, \qquad \eta_t|\gamma_t \sim N(0, exp\{\gamma_t\}),$$

$$\gamma_t = \theta\gamma_{t-1} + \lambda + \omega_t,$$

where γ_t, θ and λ are additional parameters in comparison with the state space Lee-Carter model of Pedroza [61].

An advantage of stochastic volatility in the time varying index model is the ability to model changes in variance in time. A disadvantage is its large number of additional parameters, which creates a risk of overfitting, especially in cases when the available data set is relatively short. Therefore, in this dissertation, we apply more restrictive SSM with regime switching as described below.

Another possible extension of the state space Lee-Carter is to improve mortality forecasts by incorporating exogenously measured covariates, for example, data of a country's economic development, which could be considered as a driver or a predictor of the modeled country's mortality. Girosi and King [36, Ch. 3] suggested including the exogenously measured covariates into the Poisson Lee-Carter model. Similarly, Toczydlowska et al. [72] used a similar approach to extend the state space Lee-Carter model. Toczydlowska's et al. [72] model additionally allows for the cohort effect of mortality; therefore, the unobserved state parameter vector includes both parameters driving the time varying index and cohort parameters. The model is as follows:

$$\boldsymbol{y}_t = \boldsymbol{\alpha} + \boldsymbol{B}\boldsymbol{\varphi}_t + \boldsymbol{\varepsilon}_t,$$

$$\boldsymbol{\varphi}_t = \boldsymbol{\Lambda}\boldsymbol{\varphi}_t + \boldsymbol{\Theta} + \eta_t,$$

where, $\boldsymbol{\varphi}_t$, $t = 1,2, \dots$, is unobserved state parameter vector, and $\boldsymbol{B}$, $\boldsymbol{\Lambda}$, $\boldsymbol{\Theta}$ are parameter vectors supplemented with exogenously measured data input.

The incorporation of exogenously measured data can enable to improve the credibility of mortality projections, provide means of dealing with missing data, and improve model fit. On the other hand, it introduces significant subjectivity in deciding which external data to incorporate, as the selection of possible external covariates is vast. Therefore, in this dissertation, we remain with the extrapolatory mortality projection models and leave models with exogenously measured covariates as an area for future research.

SSM with regime switching, often referred to multi-process models, are DLMs conditionally on a sequence of indicator variables, which represent

varying regimes. West and Harrison [77] and Kim and Nelson [49] provide an introduction to such models. In the context of the mortality modeling, switches between regimes represent changes between stable and volatile socio-economic development conditions, which positively or negatively affect mortality development. Such models were applied for the modeling of business cycles; see for instance Hamilton [37], Kaufmann [46], and Kim and Nelson [49] among others. We have not identified the application of SSM with regime switching in modeling of mortality.

Bayesian SSM parameter estimation method called Gibbs sampler was proposed by Geman and Geman [32]. Gelfand and Smith [30] showed the relationship of Gibbs sampler to other Bayesian sampling algorithms and discussed the implications to the estimation of posterior densities. A significant amount of research was devoted to analysis of the label switching phenomenon applicable when Gibbs sampler is used with the SSM with switching, see [15], [25], [26], [32], which identified several issues with the application of Gibbs sampler and indicated different methods of dealing with label switching.

The method of estimation of the marginal likelihood for state space models was suggested by Chib [16]. Pitt and Shephard [63] proposed a sequential Monte Carlo method, called the Auxiliary Participle Filter, which provides a convenient method of estimation of conditional likelihood to be used in combination with Chib's method. Further details on the method can be found in [4] and [22]. Kaufmann [47] applied the model for estimation of the marginal likelihood of SSM with switching.

2. STOCHASTIC MORTALITY MODELS

In this section, we describe in detail the stochastic mortality models, their fitting, and projection methods used in this dissertation. Three different mathematical specifications (classical, Poisson bilinear regression, and state space model) of the Lee-Carter model are analyzed in this section. We also provide a brief overview of several alternative stochastic mortality models is provided. The section is concluded with a summary.

2.1. Classical Lee-Carter model

The classical Lee-Carter model is based on the approach presented in the paper by Lee and Carter [53]. The model parameters are estimated using two stage procedure using singular value decomposition and time series analysis methods.

2.1.1. Classical Lee-Carter model definition

According to Lee and Carter [53] mortality rates can be modeled with the following formula:

$$\log(m_{x,t}) = \alpha_x + \kappa_t \beta_x + e_{x,t}, \tag{3}$$

where: $m_{x,t}$ is the central mortality rate for age group $x \in \{1, \dots, N\}$ in year $t \in \{1, \dots, T\}$, α_x is a parameter, which represents the general shape of changes in log mortality rates with age, κ_t is a time varying index, which represents the general trend of changes in mortality rates with time, β_x is a parameter, which determines the impact of the time varying index on age specific log mortality rates, and $e_{x,t}$ are errors. It is assumed that random variables $e_{x,t}$ are independent and have zero mean and fixed and equal variances.

Age specific parameter α_x is estimated for each age group as follows:

$$\alpha_x = \sum_{t=1}^{T} \frac{\log(m_{x,t})}{T}.$$

The remaining parameters are estimated in two stages. Firstly, κ_t and β_x are estimated using singular value decomposition (SVD) of a matrix of centered log mortality rates (log mortality rates minus α_x). Secondly, time

series analysis methods are used to model and forecast the dynamics of parameter κ_t.

As it can be easily seen, the model is over-parametrized. In order to ensure that the model is determined, the following constraints are introduced:

$$\sum_t \kappa_t = 0, \tag{4}$$

$$\sum_x \beta_x = 1. \tag{5}$$

In the original paper, Lee and Carter [53] have used the second step reestimation of parameter κ_t. After fixing α_x and β_x, κ_t was re-estimated to ensure that the actual historic total number of deaths in a fitting year equals to the fitted number of death. In subsequent applications, second step reestimation was often debated. Lee and Miller [54] suggested setting the reestimated κ_t in such a way that the periodic life expectancy is matched for the year in question. Booth et al. [5] suggested setting the adjusted κ_t by fitting a Poisson regression model to the annual number of deaths at each age. However, as noted by Girosi and King [36, Ch. 2] the second stage reestimation of the classical model does not always result in a unique solution and is skipped by some researchers. Due to possible bias consideration as discussed in the next subsection and considering that the second stage reestimation of parameter κ_t makes the classical model results less comparable with alternative specifications of the Lee-Carter model, in this dissertation the reestimated parameters are provided for comparison, but for forecasting, the initial SVD estimates of parameter κ_t are used.

2.1.2. Singular value decomposition

The SVD method is based on the idea of principal component analysis developed by Hotelling [41]. The method relies on the fact that the data matrix may be decomposed into the sum of k matrices of the same dimensions, where k is the rank of the original data matrix. Thus, we can model the data matrix with one or a few of the component matrices, which reproduce the largest part of the variation in the data matrix and discharge the remaining component matrices as a model simplification.

The method relies on the following two key results from the matrix algebra, see [2]:

Theorem 1. (the Spectral Decomposition of a symmetric matrix) Given any symmetric matrix $\boldsymbol{B}$, there exists an orthogonal matrix $\boldsymbol{C}$ such that

$$\boldsymbol{C'BC} = \boldsymbol{D} = \begin{bmatrix} d_1 & 0 & \cdots & 0 \\ 0 & d_2 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & d_p \end{bmatrix}.$$

If $\boldsymbol{B}$ is positive semidefinite, then $d_i \geq 0$ for $i = 1, \ldots, p$; if **B** is positive definite[1] than $d_i > 0$ for $i = 1, \ldots, p$.

Proof: see [2, Ch. 11].

Theorem 2. (the singular value decomposition) Given an $n \times m$ matrix $\boldsymbol{X}$, $n \geq m$, there exists an $n \times n$ orthogonal matrix $\boldsymbol{P}$, an $m \times m$ orthogonal matrix $\boldsymbol{Q}$, and $n \times p$ matrix $\boldsymbol{D}$ consisting of $m \times m$ diagonal positive semidefinite matrix and a $(n - m) \times m$ zero matrix such that

$$\boldsymbol{X} = \boldsymbol{PDQ}.$$

Proof, see [2, p. 634].

Householder and Young [42] proved that if matrix $\boldsymbol{X}$ is of rank r, the sum of the first $j<r$ terms of the SVD, corresponding to the largest diagonal elements d_i provides the best approximation of matrix $\boldsymbol{X}$ when the error is measured as the sum of squares of the residuals. In addition, they showed that the approximation is unique unless $d_j = d_{j+1}$. As stated by Good [34]:

Proposition 3. The sum of the first *j* terms $j = 1,2, \ldots, k$ of the SVD of a $n \times m$ matrix $\boldsymbol{X}$ of rank *k* gives the matrix $\boldsymbol{Z}$ of rank *j* that best approximates $\boldsymbol{X}$ in the sense of least squares, i.e. for which we minimize

$$d^2 = \sum_{i=1}^{n} \sum_{j=1}^{m} \left(x_{ij} - z_{ij}\right)^2 = \text{trace}[(\boldsymbol{X} - \boldsymbol{Z})(\boldsymbol{X'} - \boldsymbol{Z'})]. \qquad (6)$$

[1] Symmetric matrix A and its quadratic form $\boldsymbol{x'Ax} = \sum_{i,j=1}^{p} a_{ij} x_i x_j$ is positive semidefinite if $\boldsymbol{x'Ax} \geq 0$ for all $\boldsymbol{x}$. If is $\boldsymbol{x'Ax} > 0$ for all $\boldsymbol{x} \neq \boldsymbol{0}$, than $\boldsymbol{A}$ and the quadratc form are called positive definite.

Proof: The matrix $\boldsymbol{Z}$ is constructed by selecting first the largest j diagonal elements of martix $\boldsymbol{D}$ and replacing the remaining elements with zeros, thus obtaining matrix $\boldsymbol{D}^*$. Inserting $\boldsymbol{Z} = \boldsymbol{P}\boldsymbol{D}^*\boldsymbol{Q}$ into Equation (6) and by considering that $\boldsymbol{P}$ and $\boldsymbol{Q}$ are orthogonal, we obtain:

$$d^2 = \text{trace}[(\boldsymbol{X} - \boldsymbol{P}\boldsymbol{D}^*\boldsymbol{Q})(\boldsymbol{X}' - (\boldsymbol{P}\boldsymbol{D}^*\boldsymbol{Q})')] = \text{trace}[\boldsymbol{P}\boldsymbol{D}^+\boldsymbol{Q}(\boldsymbol{P}\boldsymbol{D}^+\boldsymbol{Q})'] = \sum_{i=j+1}^{\boldsymbol{k}} d_i^2, \tag{7}$$

where $\boldsymbol{D}^+ = \boldsymbol{D} - \boldsymbol{D}^*$.
Thus d^2 is minimized when the d_i excluded from matrix $\boldsymbol{D}^*$ are the smallest.

Using Equation (7) it is possible to derive the measure of the proportion of variance of matrix $\boldsymbol{X}$ explained by matrix $\boldsymbol{Z}$. We denote this measure as $V\%$ and define as:

$$V\% = \frac{\sum_{i=1}^{j} {d_i}^2}{\sum_{i=1}^{k} {d_i}^2}, \tag{8}$$

where j is the number of SVD terms used to construct matrix $\boldsymbol{Z}$, k is the rank of matrix $\boldsymbol{X}$ and d_i are defined as in Equation (7).

Gabriel (1978) showed that SVD in combination with ordinary least squares fit can be used to provide a solution to the mixed additive and multiplicative models:

$$x_{i,j} = a_i + b_j + d_i^{(1)}c_j^{(1)} + \cdots + d_i^{(m)}c_j^{(m)}, \tag{9}$$

where a_i and b_j are additive terms and the remaining terms model the multiplicative effects.

Theorem 4. For any real matrix $n \times m$ matrix $\boldsymbol{X}$ and $n \times k$ matrix $\boldsymbol{P}$ and any integer r:

$$\min_{\boldsymbol{D}_{(n\times h)}} \min_{\boldsymbol{C}_{(h\times m)}} \min_{\boldsymbol{B}_{(k\times m)}} \|\boldsymbol{X} - \boldsymbol{P}\boldsymbol{B} - \boldsymbol{D}\boldsymbol{C}\|^2 = \min_{\boldsymbol{D}_{(n\times h)}} \min_{\boldsymbol{C}_{(h\times m)}} \left\|\tilde{\boldsymbol{X}} - \boldsymbol{D}\boldsymbol{C}\right\|^2,$$

where:

$$\tilde{\boldsymbol{X}} = \left(\boldsymbol{I} - \boldsymbol{P}^{\boldsymbol{0}}\right)\boldsymbol{X},$$

$$\boldsymbol{P^0 = P(P'P)^{-1}P'}.$$

Proof, see [28, 1978].

Theorem 4 enables us to solve the models as formulated in Equation (9), by firstly using least squares to estimate the fixed effects and as the second stage to estimate the multiplicative terms with SVD. The fitting process of the original Lee-Carter model follows this approach by making the additional assumption that fixed time effects are zero.

An important consideration when using the SVD method is how many components to use in constructing the approximating matrix $\boldsymbol{Z}$ from Proposition 3. Lee and Carter [53], when applying their model for the US data, used the first component corresponding to the largest eigenvalue. Tuljapurkar et al. [74], applied SVD to the mortality rates (starting from 1950) of G7 countries and showed that for all analyzed countries the first singular value component explains over 94% variation. As we shall see further in this dissertation, the percentage may not be so high if the analyzed mortality rates include periods of less steady development. Some authors, e.g., Booth et al. [6] and Renshaw and Haberman [66], analyzed the effect of inclusion of second order SVD terms to model age-time interactions and even higher order terms. The inclusion, as expected, resulted in a better fit of the model. However, the model with higher order terms proved to be less suitable for forecasting or did not result in significantly different forecasts. Therefore, for the purpose of the models of this dissertation, we shall use only the first singular value.

In the light of the theoretical background of the SVD we can also give the interpretation of the second step reestimation of parameter κ_t described in the previous subsection. The second stage reestimation procedure effectively is the upload on the eigenvector of matrix $\boldsymbol{Q}$, related to the largest singular value, with the residual net effect of the variation contained in higher order terms. Such modification reduces the residuals in the observation Equation (3) at the expense of higher volatility of parameter κ_t. Thus, it could be argued, that by such procedure we are introducing a bias in the first principal component which is supposed to capture the overall trend in mortality improvements.

2.1.3. Time series analysis of time varying index

After carrying the standard time series model identification procedures, Lee and Carter [53] proposed to suppose κ_t to be a random walk with drift (RWD):

$$\kappa_t = \kappa_{t-1} + \mu + \varepsilon_t, \qquad t = 1, 2, \dots, \tag{10}$$

where μ is drift parameter, independent identically distributed $\varepsilon_t \sim N(0, \sigma^2)$, and κ_0 is a constant.

$ARIMA(p, d, q)$ models with $d = 1$, predominantly RWD, were the main model for κ_t in subsequent applications of the Lee-Carter model using the data of various countries. Therefore, our primary hypothesis is that κ_t derived from SVD can be modeled as a differenced time series with $d = 1$. In order to accept or reject this hypothesis statistical testing of unit roots, using Phillips-Perron and Augmented Dickey-Fuller tests as described by Hamilton [39, Ch. 17], can be performed.

Phillips-Perron test is based on the following regression, which in our case is fitted using the estimated historic κ_t values by the ordinary least squares (OLS):

$$\kappa_t = \rho\kappa_{t-1} + \mu + \delta t + v_t.$$

In the equation above, v_t for $t = 1, 2, \dots$ are zero mean random variables, which may be heterogeneously distributed or serially correlated, and ρ, μ, δ are estimated parameters. To test the hypothesis about the presence of unit root, the following two statistics are calculated:

$$Z_\rho = T(\hat{\rho} - 1) - \frac{T^2 \hat{\sigma}_{\hat{\rho}}^2}{2\hat{s}^2}\left(\hat{\lambda}^2 - \hat{\gamma}_0\right),$$

$$Z_\tau = \sqrt{\frac{\hat{\gamma}_0}{\hat{\lambda}^2}}\frac{\hat{\rho} - 1}{\sigma_{\hat{\rho}}} - \frac{T\hat{\sigma}_{\hat{\rho}}}{2\hat{s}\hat{\lambda}}\left(\hat{\lambda}^2 - \hat{\gamma}_0\right),$$

where in the sequel $\hat{\gamma}_m$, $m \in \{0,1, \dots, q\}$ are estimated sample autocovariances with m^{th} lag, q is maximum lag considered in the test, $\hat{s}^2$ is sample variance, $\hat{\sigma}_{\hat{\rho}}$ is estimated OLS standard error of $\hat{\rho}$, and estimated long run variance $\hat{\lambda}^2$ is calculated using the following formula:

$$\hat{\lambda}^2 = \hat{\gamma}_0 + 2\sum_{j=1}^{q}\left(1 - \frac{j}{q+1}\right)\hat{\gamma}_j.$$

The Augmented Dickey-Fuller test is based on the following regression using OLS:

$$\kappa_t = \rho\kappa_{t-1} + \mu + \delta t + \sum_{i=1}^{p-1}\zeta_i(\kappa_{t-i} - \kappa_{t-i+1}) + w_t.$$

where p is the number of lags considered in the test, w_t, $t = 1, 2, \dots$, are i.i.d. zero mean random variables, fixed variance and finite fourth moments, and $\mu, \rho, \delta, \zeta_i, \dots, \zeta_{p-1}$ are the model parameters. To test the hypothesis about the presence of unit root the following two statistics are used:

$$Z_{DF}(\rho) = \frac{T(\hat{\rho} - 1)}{1 - \hat{\zeta}_1 - \hat{\zeta}_2 - \cdots - \hat{\zeta}_{p-1}},$$

$$Z_{DF} = \frac{(\hat{\rho} - 1)}{\hat{\sigma}_{\hat{\rho}}^2},$$

where $\hat{\sigma}_{\hat{\rho}}$ is estimated OLS standard error of $\hat{\rho}$, and $\hat{\rho}$, $\hat{\zeta}_1, \dots, \hat{\zeta}_{p-1}$ denote estimates of corresponding parameters.

2.1.4. Forecasting and simulation

The Lee-Carter model projections are derived using the basic model structure defined by formula (3). In the original paper, Lee and Carter [53] assumed that model parameters are fixed and changes in the mortality rates are driven by innovations of time varying index κ_t. Although such an approach is simple to implement, it might be argued that it produces too narrow confidence intervals. Lee and Miller [54] proposed a simple adjustment to take into account the uncertainty in the trend parameter μ, when κ_t is modelled as RWD. Various bootstrap based approaches have been tried primarily with Poisson bi-linear specification of Lee-Carter model (see [55] and [68] for an overview), and their results are often applicable to SVD model specification. However, in the majority of applications of Lee-Carter model κ_t has remained the key source of uncertainty. Thus, also aiming to maintain consistency with the forecasting approach applied to state space models, we

shall simulate mortality rates for periods $T+1, \dots, T+K$, where T is the length of time period used for model fitting and K is the term of forecast, using the allowing algorithm:

Algorithm 1. *Simulation using classical Lee-Carter model*
For iteration s $= 1, \dots, S$:

i. Draw μ from $N\left(\hat{\mu}, \frac{\hat{\sigma}^2}{T}\right)$, where $\hat{\mu}$ and $\hat{\sigma}^2$ are parameter estimates of RWD model specified by Equation (10).
ii. For each $t = T+1, \dots, T+K$:
    a. Draw κ_t from $\kappa_t \sim N(\kappa_{t-1} + \mu, \hat{\sigma}^2)$ where the notation of parameters is as in Equation (10).
    b. For each $x = 1, \dots, N$ draw $y_{x,t}$ from $y_{x,t} \sim N(\alpha_x + \beta_x \kappa_t, \sigma_e^2)$, where σ_e^2 is the variance of the residual $e_{x,t}$ and the notation of parameters is as in Equation (3).
iii. Calculate $m_{x,t} = e^{y_{x,t}}$.

2.2. Poisson Lee-Carter model

It can be argued, that estimates provided by the classical Lee-Carter model are not optimum, because the model's assumption that the errors are distributed homoskedastically is not realistic. The motivation for fitting the model with Poisson regression is to apply a different and more realistic variance structure of the errors, as it can be expected that the variance of mortality rates increase with age due to the growing count of deaths. Thus we can use the generalized linear model (GLM) framework for such alternative specification of the Lee-Carter model which initially was proposed by Alho [1].

2.2.1. Definition of the Poisson Lee-Carter model

Brouhns et al. [10] developed a method of fitting the Lee-Carter model as Poisson regression:

$$D_{x,t} \sim Poiss\left(E_{x,t}\lambda_{x,t}\right) \text{ with } \lambda_{x,t} = e^{\alpha_x + \kappa_t \beta_x}, \tag{11}$$

where $D_{x,t}$ and $E_{x,t}$ are a number of deaths and exposed to risk at age group $x \in \{1, \dots, N\}$ and time $t \in \{1, \dots, T\}$ respectivelly. As the original Lee-Carter model, this model is overparametrized and the identifiability Constraints (4)

and (5) are required. Such model specification has the following likelihood function:

$$\ell = \sum_{t=1}^{T}\sum_{x=1}^{N}\left[D_{x,t}(\alpha_x + \kappa_t\beta_x) - E_{x,t}e^{\alpha_x+\kappa_t\beta_x}\right] + C, \qquad (12)$$

where C is constant which does not depend on estimated parameters α_x, β_x and κ_t.

It is noted by McCullagh and Nelder [57] that in practice, when modeling data with Poisson regression, it is not uncommon for the actual variance of the response variable to exceed the variance implied by the Poisson model, i.e., in our case $\mathrm{Var}(D_{x,t}) > \mathrm{E}(D_{x,t})$. This phenomenon is called *overdispersion* and may occur due to a number of reasons, one of which is the heterogeneity of the modeled population due to features other than factors explicitly taken into account by the model. For example, the mortality of the same age population may further vary with income level, urban/rural living location, etc. Li et al. [56] allowed for overdispersion in mortality data by expressing the Lee-Carter model as negative binomial regression. Such an approach allows estimating different overdispersion parameters for each age group. However, the negative binomial version of the Lee-Carter model suffers from overparameterization, as we need to estimate N instead of one overdispersion parameter. As shown in [56] overdispersion is very high in young ages and starts to increase for a population starting from 80 years old, however, for the mid-aged population (which is the target group for our research) the overdispersion is relatively stable. Therefore, we choose a simpler approach and fit the overdispersed quasi-Poisson regression.

The quasi-Poisson regression, given the independent observations

$$y_i,\ i \in \{1, \dots, M\},$$

can be defined as GLM with the following properties:

$$\mathrm{E}(y_i) = \mu_i = e^{\eta_i}, \qquad \mathrm{Var}(y_i) = \varphi\mathrm{Var}(\mu_i) = \varphi\mu_i,$$

where $\eta_i = x^T\beta$ is linear predictor with parameter vector β, and φ is the overdispersion parameter. Such model does not have closed form solution for density or log-likelihood, but may be fitted by maximizing quasi-likelihood function, which is defined by McCullagh and Nelder [57, p. 325] as follows:

$$\frac{\partial Q(y_i,\mu_i)}{\partial \mu_i} = \frac{y_i - \mu_i}{\varphi\mathrm{Var}(\mu_i)}, \qquad Q(y_i,\mu_i) = \int_{y_i}^{\mu_i}\frac{y_i - t}{\varphi\mathrm{Var}(t)}dt + C.$$

As proved by Wedderburn [76] the properties of quasi-likelihood function Q are similar to those of log likelihood ℓ, in particular:

$$\mathrm{E}\left(\frac{\partial Q}{\partial \mu}\right) = 0, \qquad \mathrm{Var}\left(\frac{\partial Q}{\partial \mu}\right) = \frac{1}{\varphi \mathrm{Var}(\mu)} = -\mathrm{E}\left(\frac{\partial^2 Q}{\partial \mu^2}\right), \tag{13}$$

which justifies the practice to use quasi-likelihoods instead of likelihoods in fitting GLMs where likelihood is not available in a closed form. Thus, in the case of the overdispersed Poisson Lee-Carter model we shall work with the following quasi-likelihood function:

$$\frac{\partial Q\left(D_{x,t}, E_{x,t}\lambda_{x,t}\right)}{\partial\left(E_{x,t}\lambda_{x,t}\right)} = \frac{D_{x,t} - E_{x,t}\lambda_{x,t}}{\varphi E_{x,t}\lambda_{x,t}}.$$

2.2.2. Maximum likelihood fitting of Poisson bilinear regression

The basic estimation algorithm used for fitting of GLMs is called iteratively weighted least squares (IRLS), which can be applied both for cases when the estimation is based on the likelihood function and quasi-likelihood function. For the Poisson Lee-Carter model, we have another technical complication arising from the presence of bilinear term, which models the age and time interaction. In this subsection, we show how to fit the bilinear term using so-called "criss-cross" algorithm.

Algorithm 2. *Quasi-likelihood IRLS*

The following discussion is based on McCullagh and Nelder [57, Ch. 9].

Using Expression (12) and noting the independence of the observations, we aim to maximize quasi-likelihood function $Q = \sum_{i=1}^{M} Q(y_i, \mu_i)$ with respect to $\boldsymbol{\beta} = \{\beta_1, \dots, \beta_P\}$ using the following score function:

$$\frac{\partial Q}{\partial \beta} = \frac{\partial Q}{\partial \mu}\frac{\partial \mu}{\partial \beta} = \boldsymbol{D}'\boldsymbol{V}^{-1}(\boldsymbol{y} - \boldsymbol{\mu})\varphi^{-1},$$

where $\boldsymbol{D}$ is $M \times P$ matrix with elements $D_{ij} = \frac{\partial \mu_i}{\partial \beta_j}$, $i \in \{1, \dots, M\}$ and $j \in \{1, \dots, P\}$, $\boldsymbol{V}$ is $M \times M$ diagonal matrix with elements $V_{ii} = \mathrm{Var}(\mu_i)$, $i \in \{1, \dots, M\}$, $\boldsymbol{y}$ is the vector of observations and $\boldsymbol{\mu}$ is vector of estimates of the

means. Using the properties of the quasi-likelihood, see Equation (13) we obtain approximate covariance matrix $\left\{-\mathrm{E}\left(\frac{\partial^2 Q}{\partial \boldsymbol{\beta}^2}\right)\right\}^{-1}$, see Wedderburn [76]:

$$-\mathrm{E}\left(\frac{\partial^2 Q}{\partial \boldsymbol{\beta}^2}\right) = \boldsymbol{D}'\boldsymbol{V}^{-1}\boldsymbol{D}\varphi^{-1},$$

$$\mathrm{cov}(\boldsymbol{\beta}) \approx \varphi(\boldsymbol{D}'\boldsymbol{V}^{-1}\boldsymbol{D})^{-1}.$$

By applying Newton-Raphson algorithm with Fisher scoring, the following recursion is derived:

$$\boldsymbol{\beta}^{(s)} = \boldsymbol{\beta}^{(s-1)} + (\boldsymbol{D}'\boldsymbol{V}^{-1}\boldsymbol{D})^{-1}\boldsymbol{D}'\boldsymbol{V}^{-1}(\boldsymbol{y} - \boldsymbol{\mu}), \tag{14}$$

which leads us to the following algorithm:
For iteration $s = 1,2, \ldots$:

i. Initiate the algorithm at arbitrary parameter vector $\boldsymbol{\beta}^{(0)}$.
ii. As described above calculate matrices $\boldsymbol{D}$, $\boldsymbol{V}$ and vector $\boldsymbol{\mu}$.
iii. Using Equation (14) calculate the updated parameter vector $\boldsymbol{\beta}^{(s)}$.
iv. Go to the next step (ii).

Stop the algorithm when the pre-defined convergence criteria is satisfied, e.g. further iterations do not result in an increase in quasi-likelihood above a certain threshold.

It should be noted that the estimates of β_n do not depend on the overdispersion parameter φ, which can be estimated separately, see Davison [20, Ch. 10]:

$$\varphi = \frac{1}{N-P}\sum_{i=1}^{N}\frac{(y_i - \hat{\mu}_i)^2}{V_i(\hat{\mu}_i)}.$$

When fitting Poisson Lee-Carter model, we are dealing with GLM matrix:

$$\mathrm{E}(\boldsymbol{Y}) = \log(\boldsymbol{M}) = \boldsymbol{X}\boldsymbol{B}'. \tag{15}$$

where $\boldsymbol{Y}$ and $\boldsymbol{M}$ are $N \times T$ matrices. As noted by Gabriel [29], GLM matrix can be converted to the usual GLM vector using the relationship $\langle \boldsymbol{CDE} \rangle = (\boldsymbol{E}' \otimes \boldsymbol{C})\langle \boldsymbol{D} \rangle$, where $\langle\ \rangle$ is a concatenation of columns operator and $\otimes$ is the Kroenecker product. After the conversion the model of Equation (15) becomes:

$$\mathrm{E}(\langle \boldsymbol{Y} \rangle) = \log(\langle \boldsymbol{M} \rangle) = (\boldsymbol{I}_m \otimes \boldsymbol{X})\langle \boldsymbol{B}' \rangle, \ \mathrm{Var}(\langle \boldsymbol{Y} \rangle) = \varphi \mathrm{diag}[\log(\langle \boldsymbol{M} \rangle)], \tag{16}$$

where diag[$\langle \boldsymbol{M} \rangle$] denotes diagonal matrix with diagonal elements are equal to the elements of vector $\langle \boldsymbol{M} \rangle$, $\boldsymbol{I}_m$ is the identity matrix, and m is the number of columns of $\boldsymbol{Y}$.

Finally, we need to deal with the estimation of the bilinear term in the linear predictor of GLM. In the original paper, Brouhns et al. [10] used the iterative procedure of Goodman [35], which is based on the univariate Newton's method. We use a slightly different algorithm based on the multivariate Newton-Raphson method, which should result in faster convergence. The idea of the algorithm is that the IRLS algorithm is initially run over the rows by fixing the bilinear term representing the column effects and treating it as a known covariate. Secondly, the bilinear term representing the row effects is fixed and IRLS is run over the columns. The algorithm (also called "criss-cross algorithm") is repeated until convergence is achieved. See van Eeuwijk [75] and Gabriel [29] for details.

The following algorithm implements these ideas for the fitting of quasi-Poisson Lee-Carter.

Algorithm 3. *Quasi-Poisson Lee-Carter criss-cross ILRS algorithm*

The algorithm uses the following notation:

$N \times T$ matrices of observations (number of deaths) $\boldsymbol{Y}$ and exposed to risk $\boldsymbol{E}$ have elements $y_{ij} = D_{i,j}$ and $e_{ij} = E_{i,j}$ respectively, where $i \in \{1, \dots, N\}$ is the age group and $j \in \{1, \dots, T\}$ is the observation year.

$N \times T$ matrix of expectations $\boldsymbol{M}$ has elements

$$\mu_{ij} = E_{i,j}\lambda_{i,j},\ i \in \{1, \dots, N\}\ ,\ j \in \{1, \dots, T\},$$

where

$$\log(\lambda_{i,j}) = \alpha_i + \kappa_j \beta_i.$$

We denote the $N \times T$ matrix of $\lambda_{i,j}$ by $\boldsymbol{\Lambda}$.

Estimated parameter collections, at iteration s:

$$\alpha^{(s)} \in \left\{\alpha_1^{(s)}, .., \alpha_N^{(s)}\right\},\ \beta^{(s)} \in \left\{\beta_1^{(s)}, .., \beta_N^{(s)}\right\},\ \kappa^{(s)} \in \left\{\kappa_1^{(s)}, .., \kappa_T^{(s)}\right\}.$$

By converting to vector GLM as in Equation (16) and using the Equation (14), we can update the estimate of age (row) effects $\boldsymbol{A} = [\boldsymbol{\alpha} \quad \boldsymbol{\beta}]$ by running IRLS Algorithm 2 and treating $\boldsymbol{B} = [\mathbf{1} \quad \boldsymbol{\kappa}]$ as fixed covariates:

$$\langle \boldsymbol{A}' \rangle^{(s)} = \langle \boldsymbol{A}' \rangle^{(s-1)} + (\boldsymbol{D}'\boldsymbol{V}^{-1}\boldsymbol{D})^{-1}\boldsymbol{D}'\boldsymbol{V}^{-1}(\langle \boldsymbol{Y}' \rangle - \langle \boldsymbol{M}' \rangle).$$

Here:

$$\langle \boldsymbol{M}' \rangle = \mathrm{E}(\langle \boldsymbol{Y}' \rangle) = \log(\langle \boldsymbol{E}' \rangle) + \log(\langle \boldsymbol{\Lambda} \rangle) = \log(\langle \boldsymbol{E} \rangle) + (\boldsymbol{I}_N \otimes \boldsymbol{B})\langle \boldsymbol{A}' \rangle,$$

$$\boldsymbol{V} = \mathrm{Var}(\langle \boldsymbol{Y}' \rangle) = \mathrm{diag}[\langle \boldsymbol{M}' \rangle],$$

and $\boldsymbol{D}$ is $NT \times 2N$ matrix with elements $d_{ij} = \frac{\partial \langle \boldsymbol{M} \rangle_i}{\partial \langle \boldsymbol{A}' \rangle}$, where index j represents a derivative with respect to the j element of $\langle \boldsymbol{A}' \rangle$.

Inserting the values of parameters and after performing matrix multiplication we obtain the following expression:

$$\begin{bmatrix} \alpha_1 \\ \beta_1 \\ \vdots \\ \alpha_N \\ \beta_N \end{bmatrix}^{(s)} = \begin{bmatrix} \alpha_1 \\ \beta_1 \\ \vdots \\ \alpha_N \\ \beta_N \end{bmatrix}^{(s-1)} + \begin{bmatrix} \Sigma\mu_{1,j} & \Sigma\mu_{1,j}\kappa_j & \dots & 0 & 0 \\ \Sigma\mu_{1,j}\kappa_j & \Sigma\mu_{1,j}{\kappa_j}^2 & \dots & 0 & 0 \\ \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & \dots & \Sigma\mu_{N,j} & \Sigma\mu_{N,j}\kappa_j \\ 0 & 0 & \dots & \Sigma\mu_{N,j}\kappa_j & \Sigma\mu_{N,j}{\kappa_j}^2 \end{bmatrix}^{-1} \begin{bmatrix} \Sigma(y_{1,j} - m_{1,j}) \\ \Sigma(y_{1,j} - m_{1,j})\kappa_j \\ \vdots \\ \Sigma(y_{N,j} - m_{N,j}) \\ \Sigma(y_{N,j} - m_{N,j})\kappa_j \end{bmatrix}. \quad (17)$$

where summations are performed by $j \in \{1, \dots, T\}$.

By converting to vector GLM as in Equation (16) and using Equation (14), we can estimate column (time) effects $\boldsymbol{\kappa}$ by running IRLS Algorithm 2 with $\boldsymbol{\beta}$ as fixed covariates and $\boldsymbol{\alpha}$ as fixed offsets:

$$\boldsymbol{\kappa}^{(s)} = \boldsymbol{\kappa}^{(s-1)} + (\boldsymbol{D}'\boldsymbol{V}^{-1}\boldsymbol{D})^{-1}\boldsymbol{D}'\boldsymbol{V}^{-1}(\langle \boldsymbol{Y} \rangle - \langle \boldsymbol{M} \rangle).$$

Here:

$$\begin{aligned} \langle \boldsymbol{M} \rangle = \mathrm{E}(\langle \boldsymbol{Y} \rangle) &= \log(\langle \boldsymbol{E} \rangle) + \log(\langle \boldsymbol{\Lambda} \rangle) \\ &= \log(\langle \boldsymbol{E} \rangle) + (\boldsymbol{I_T} \otimes \boldsymbol{\alpha})1_T + (\boldsymbol{I_T} \otimes \boldsymbol{\beta})\bar{\kappa}, \end{aligned}$$

$$\boldsymbol{V} = \mathrm{Var}(\langle \boldsymbol{Y} \rangle) = \mathrm{diag}[\langle \boldsymbol{M} \rangle],$$

and $\boldsymbol{D}$ is $NT \times N$ matrix with elements $d_{ij} = \frac{\partial \langle \boldsymbol{M} \rangle_j}{\partial \kappa_i}$. 1_T is column vector of length T, with all elements equal to 1.

Inserting the values of parameters and performing matrix multiplication we obtain the following expression:

$$\begin{bmatrix}\kappa_1\\ \vdots\\ \kappa_T\end{bmatrix}^{(s)} = \begin{bmatrix}\kappa_1\\ \vdots\\ \kappa_T\end{bmatrix}^{(s-1)} + \begin{bmatrix}\Sigma{\mu_{i,1}\beta_i}^2 & \dots & 0\\ \vdots & & \vdots\\ 0 & \dots & \Sigma{\mu_{i,T}\beta_i}^2\end{bmatrix}^{-1} \begin{bmatrix}\Sigma(y_{i,1}-m_{i,1})\beta_i\\ \vdots\\ \Sigma(y_{i,N}-m_{i,N})\beta_i\end{bmatrix}, \quad (18)$$

where summations are performed with respect to $i \in \{1, \dots, N\}$.

The algorithm may be summarized as follows:

For iteration $s = 1,2, \dots$

i. Initiate the algorithm at arbitrary parameter vector $\boldsymbol{\alpha}^{(0)}$, $\boldsymbol{\beta}^{(0)}$ and $\boldsymbol{\kappa}^{(0)}$.
ii. Run IRLS Algorithm 2 for row effects using Equation (17), until convergence is achieved.
iii. Fix parameter values $\boldsymbol{\alpha}^{(s)}$ and $\boldsymbol{\beta}^{(s)}$ and run IRLS Algorithm 2 for column effects using Equation (18), until convergence is achieved.
iv. Fix parameter values $\boldsymbol{\kappa}^{(s)}$ and go to step (ii).

Stop the algorithm when the pre-defined convergence criteria is satisfied, e.g. further iterations do not result in an increase in quasi-likelihood above the certain threshold.

2.2.3. Forecasting and simulation

In the Poisson Lee-Carter model κ_t is modeled in the same way as in the classical Lee-Carter model, see Subsection 2.1.3. The simulation of projections takes into account different variance structure and uses Normal approximation of Poisson distribution. Simulations are performed according to the following algorithm:

Algorithm 4. *Simulation using Poisson bilinear Lee-Carter model*
For each simulation $s = 1, \dots, S$:

i. Draw μ from $N\left(\hat{\mu}, \frac{\hat{\sigma}^2}{T}\right)$, where $\hat{\mu}$ and $\hat{\sigma}^2$ are parameter estimates of RWD model specified by Equation (10),
ii. For each $t = T+1, \dots, T+K$, where T is the length of time period used for model fitting and K is the term of forecast:
    a. Draw κ_t from $\kappa_t \sim N(\kappa_{t-1} + \mu, \hat{\sigma}^2)$ where the notation of parameters is as in Equation (10).
    b. For each $x = 1, \dots, N$ draw $D_{x,t}$ from $D_{x,t} \sim N\left(E_{x,t}\lambda_{x,t}, \varphi E_{x,t}\lambda_{x,t}\right)$, where $\lambda_{x,t} = e^{\alpha_x + \kappa_t \beta_x}$ and φ is

the estimated overdispersion parameter and the notation of parameters is as in Equation (11).

iii. Calculate $m_{x,t} = \frac{D_{x,t}}{E_{x,t}}$.

2.3. State space Lee-Carter model

State space time series models (SSM) (see [22], [40], [77]) offer a natural extension of the Lee-Carter model both in terms of the alternative coherent fitting procedure and the possibility to flex the model by introducing additional parameters to ensure a better fit. In Subsection 2.3.1. we start with the basic linear Dynamic Linear Model (DLM) as defined by Equations (1) and (2). Using the flexibility of SSMs and allow, if necessary, for one time change in mortality trend. Thus, instead of one trend parameter μ of time varying index κ_t we use

$$\boldsymbol{\mu} = \begin{bmatrix} \mu^{(I)} \\ \mu^{(II)} \end{bmatrix},$$

where we denote $\mu^{(I)}$ the drift over the whole fitting period and $\mu^{(II)}$ the additional drift from the year of change in drift.

In Subsection 2.3.2. we proceed with another modification where we apply non-linear SSM with regime switching to model changes in the volatility of the time varying index.

2.3.1. State space Lee-Carter model definition

We use two forms of the state space Lee-Carter model: DLM and state space model with regime switching (SSM with switching). In this subsection, we describe the DLM version of Lee-Carter model.

By using the general Expressions (1) and (2) of DLM, we can formulate the DLM Lee-Carter model by the following two equations:

$$\boldsymbol{y}_t = [\boldsymbol{\alpha} \quad \mathbf{0} \quad \mathbf{0} \quad \boldsymbol{\beta}]\boldsymbol{x}_t + \boldsymbol{\varepsilon}_t, \quad \boldsymbol{\varepsilon}_t \sim N(\mathbf{0}, \boldsymbol{H}_t), \tag{19}$$

$$\boldsymbol{x}_t = \begin{bmatrix} 1 \\ \mu_t^{(I)} \\ \mu_t^{(II)} \\ \kappa_t \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & \mathbb{I}_{\{t>T_0\}} & 1 \end{bmatrix} \boldsymbol{x}_{t-1} + \boldsymbol{\eta}_t, \qquad \boldsymbol{\eta}_t \sim N(\mathbf{0}, \boldsymbol{Q}_t). \tag{20}$$

In the observation Equation (19), symbol $\boldsymbol{y}_t$, $t \in \{1,2,\dots,T\}$ denotes a vector with N coordinates, where N is the total number of age groups. Each coordinate of this vector $x \in \{1,2,\dots,N\}$ is the specific log-mortality rate $y_{x,t} = \log(m_{x,t})$. Parameters $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are vectors of N age specific parameters α_x and β_x which have the same interpretation as in the classical Lee-Carter model. Matrix $\boldsymbol{H}_t$ is a diagonal $N \times N$ matrix with all diagonal elements equal to σ_H^2.

In the state Equation (20), element κ_t is the time varying index, which represents the general trend of changes in mortality rates with time, which has the same interpretation as in the classical Lee-Carter model. Parameter $\mu^{(I)}$ represents the drift of κ_t applicable to the whole fitting period and $\mu^{(II)}$ represents the additional drift element in the periods $t > T_0$ and symbol $\mathbb{I}$ denotes the standard indicator function. Thus, during the periods $t > T_0$ the model assumes that the total drift is $\mu^{(I)} + \mu^{(II)}$. As in the classical Lee-Carter model, we assume that both drift parameters stay constant during the fitting period and we apply the following covariance matrix of errors of the state vector $\boldsymbol{x}_t$:

$$\boldsymbol{Q}_t = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \sigma_Q^2 \end{bmatrix}.$$

The above model maintains many of the features of the classical Lee-Carter model. The time varying index κ_t follows the RWD. Contrary to the classical Lee-Carter model, if needed we can allow the one-time change in the drift at some time moment T_0, but otherwise, drifts are not allowed to vary in time. Such structure ensures that the model is sufficiently rigid, which is important for making long-term forecasts. As in the classical Lee-Carter model, the general trend of mortality modeled with parameter κ_t is translated to age-specific log mortality rates via the parameters β_x, and the random errors $\varepsilon_{x,t}$ are assumed to be i.i.d. The model is unidentified, and we impose the same identifiably Constrains (4) and (5) as in the classical Lee-Carter model.

SSMs enable us to estimate the distribution of each of the unobservable parameters $\boldsymbol{x}_t$, by using the observed data $\boldsymbol{y}_1, \dots, \boldsymbol{y}_t$. The recursive algorithm, called Kalman filter, which is based on the method of orthogonal projections and uses the features of the conditional Normal distribution, updates the estimate of the mean and variance of $\boldsymbol{x}_{t-1}$ with the new data contained in $\boldsymbol{y}_t$ to derive the mean and variance of $\boldsymbol{x}_t$.

***Algorithm 5**. Kalman filter recursion*

Consider the system defined for $k \geq 0$ by Equations (1) and (2). Suppose that $\boldsymbol{x}_0$ has the mean $\boldsymbol{a}_0$ and the covariance matrix $\boldsymbol{P}_0$ and is uncorrelated with $\{\boldsymbol{\varepsilon}_1, .., \boldsymbol{\varepsilon}_t\}$ and $\{\boldsymbol{\eta}_1, \dots, \boldsymbol{\eta}_t\}$.

Using the properties of Normal conditional distribution, we can recursively calculate the distribution of unobserved state vector $\boldsymbol{x}_t$, conditionally on $\boldsymbol{y}_t$ $\boldsymbol{x}_{t|t} \sim N(\boldsymbol{a}_{t|t}, \boldsymbol{P}_{t|t})$ and it's one step ahead prediction $\boldsymbol{x}_{t+1} \sim N(\boldsymbol{a}_{t+1}, \boldsymbol{P}_{t+1})$ as follows:

$$\boldsymbol{a}_{t|t} = \boldsymbol{a}_t + \boldsymbol{P}_t \boldsymbol{Z}_t' \boldsymbol{F}_t^{-1} \boldsymbol{v}_t, \qquad \boldsymbol{P}_{t|t} = \boldsymbol{P}_t - \boldsymbol{P}_t \boldsymbol{Z}_t' \boldsymbol{F}_t^{-1} \boldsymbol{Z}_t \boldsymbol{P}_t,$$

$$\boldsymbol{a}_{t+1} = \boldsymbol{U}_t \boldsymbol{a}_t + \boldsymbol{K}_t \boldsymbol{v}_t, \qquad \boldsymbol{P}_{t+1} = \boldsymbol{U}_t \boldsymbol{P}_t (\boldsymbol{U}_t - \boldsymbol{K}_t \boldsymbol{Z}_t)' + \boldsymbol{Q}_t,$$

$$\boldsymbol{v}_t = \boldsymbol{y}_t - \boldsymbol{Z}_t \boldsymbol{a}_t, \qquad \boldsymbol{F}_t = \boldsymbol{Z}_t \boldsymbol{P}_t \boldsymbol{Z}_t' + \boldsymbol{H}_t,$$

$$\boldsymbol{K}_t = \boldsymbol{U}_t \boldsymbol{P}_t \boldsymbol{Z}_t' \boldsymbol{F}_t^{-1}.$$

See [22, Ch. 4] for proof.

As shown by Anderson and Moore [3, Ch. 3], assuming that errors are uncorrelated and zero mean, the Kalman filter estimator is the minimum variance estimator within a certain restricted class of filters, whether or not the Normality assumption is made. Therefore, we can justify the application of the recursions in practical situations where the distribution of errors is not Normal.

Therefore, in fitting the DLM Lee-Carter model, the Kalman filter provides us an easy way to estimate the time varying index $\{\kappa_1, .., \kappa_T\}$. Estimation of the collection of the remaining model parameters $\boldsymbol{\psi} = \{\boldsymbol{\alpha}, \boldsymbol{\beta}, \sigma_Q, \sigma_H, \boldsymbol{\mu}\}$ is described in Subsection 2.3.3.

2.3.2. State space Lee-Carter model with regime switching

The basic equations of the state space Lee-Carter model with regime switching remain (19) and (20). The key difference from the previous model is that we assume that the error terms $\boldsymbol{\eta}_t$, $t \in \{1, \dots, T\}$ follow a mixture of two zero mean Normal distributions with the following covariance matrices:

$$\boldsymbol{Q}_t^{(0)} = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \left(\sigma_Q^{(0)}\right)^2 \end{bmatrix}, \qquad \boldsymbol{Q}_t^{(1)} = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \left(\sigma_Q^{(1)}\right)^2 \end{bmatrix},$$

each of them corresponding to one of the two regimes. We restrict $\left(\sigma_Q^{(0)}\right)^2 \leq \left(\sigma_Q^{(1)}\right)^2$ and use a binary index s_t: low variance regime $s_t = 0$ and high variance regime $s_t = 1$.

Switching between the two regimes is assumed to follow the homogeneous Markov process with constant transition probabilities as introduced by Hamilton [38]. We define two probabilities

$$\pi^{(0)} = \mathbb{P}(s_t = 0 | s_{t-1} = 0), \;\; \pi^{(1)} = \mathbb{P}(s_t = 1 | s_{t-1} = 1)$$

which imply that:

$$\mathbb{P}(s_t = 1 | s_{t-1} = 0) = 1 - \pi^{(0)}, \;\; \mathbb{P}(s_t = 0 | s_{t-1} = 1) = 1 - \pi^{(1)}.$$

Consequently, the parameter collection of the model with regime switching is $\boldsymbol{\psi}^{(s)} = \{\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\sigma}_Q, \sigma_H, \boldsymbol{\mu}, \boldsymbol{\pi}\}$, where $\boldsymbol{\pi} = \begin{bmatrix} \pi^{(0)} \\ \pi^{(1)} \end{bmatrix}$ and $\boldsymbol{\sigma}_Q = \begin{bmatrix} \sigma_Q^{(0)} \\ \sigma_Q^{(1)} \end{bmatrix}$.

The key implication of the modeling of errors as a mixture is that the resulting distribution of errors

$$\sum_{i=0}^{1} \mathbb{P}(s_t = i) N\left(0, \left(\sigma_Q^{(i)}\right)^2\right),$$

depending on its parametrization may have significantly heavier tails than the tails of a Normal distribution.

Conditionally on values of κ_t, $t \in \{1, \dots, T\}$, we can recursively estimate probabilities of $s_t = 0$ and $s_t = 1$, by taking into account the corresponding probabilities at time moment $t - 1$, as well as the likelihood of being in a particular state, which is driven by the deviation of the change $\kappa_t - \kappa_{t-1}$ from its expected value $\mu^{(I)} + \mathbb{I}_{\{t > T_0\}} \mu^{(II)}$: the higher the deviation is, the more likely is that the process has moved to high variance regime. For details see step (ii) of Algorithm 8.

Conditionally on values of s_t, $t \in \{1, \dots, T\}$, we can apply the basic Kalman filter with matrix $\boldsymbol{Q}_t$ varying depending on the regime.

The key benefit from using the SSM with regime switching is that it enables us to segregate the total variance of κ_t, which is generally recognized to be the key source of variability of mortality rates, into the variance during the periods of the "normal" socio-economic development and the variance during the periods such as epidemics, wars, changes economic and political systems, etc. As we shall see in Section 3, the development of the general mortality trend may be far from constant, therefore, the classical Lee-Carter approach, which averages the estimated variance of κ_t over the fitting periods can miss important features of the real life development, especially when the purpose of the analysis is the estimation of the confidence intervals of the mortality forecasts rather than just the central forecast.

In this disseration, we limit our analysis to a two-regime model. It would be possible to develop a model where switching occurs between three or a higher number of regimes. However, in such a case we would need to estimate more parameters which could have a major impact on their credibility and robustness. In addition, we would depart from the initial assumption that there are two socio-economic development modes. Therefore, we leave this as a possible model extension for further research.

2.3.3. Gibbs sampler

There are two main approaches to the fitting SSMs: maximum likelihood (MLE) and Bayesian Markov Chain Monte Carlo (MCMC). For estimation of the parameter collection $\boldsymbol{\psi}$ of DLM Lee-Carter model we can easily apply MLE, see e.g. [27]. However, as noted by Frühwirth-Schnatter [26], in the case of SSM with regime switching the marginal likelihood where both latent processes $\boldsymbol{x}$ and $\boldsymbol{s}$ are integrated out, is not available in the closed form. In addition, MLE does not explicitly allow for uncertainty in model parameters, which is an important element in satisfying our aim to estimate realistic confidence intervals of the forecasts.

Due to these reasons, for model fitting, we apply the MCMC method, called the Gibbs sampler, which was also used in some earlier applications of the state space Lee Carter model without regime switching by Fung et al. [27], Kogure and Kurachi [51], Pedroza [61]. The Gibbs sampler draws samples of parameter values from sequentially updated conditional distributions. Thus, if we have a parameter vector $\boldsymbol{\theta} = \{\theta_1, \theta_2, \dots, \theta_M\}$, the Gibbs sampler at each iteration $j = 1,2, \dots$ would sequentially draw samples of θ_i^j for $i \in \{1,2, \dots, M\}$ from $\left(\theta_i^j | \theta_1^j, \dots, \theta_{i-1}^j, \theta_{i+1}^{j-1}, \dots, \theta_M^{j-1}\right)$, where θ_i^j represents draw of parameter θ_i at j iteration. As shown by Geman and Geman [32] under

certain mild conditions the distribution of draws $\boldsymbol{\theta}^j$ converge to the true distribution of $\boldsymbol{\theta}$ if $j \to \infty$, independently of the starting values of $\boldsymbol{\theta}^1 = (\theta_1^1, \theta_2^1, \dots, \theta_M^1)$.

The key complication in constructing and using the Gibbs sampler is the requirement at each simulation step to have conditional distributions of the sampled parameters. In our case, the complication arises in implementing a proper sampler of the unobserved state parameter $\boldsymbol{x}_t$ and in the case of model with regime switching the collection vector of regimes $\boldsymbol{s} = \{s_1, s_2, \dots, s_T\}$. This issue was successfully resolved by the Forward Filtering Backward Sampling (FFBS) algorithm, see [14].

Algorithm 6. *Forward Filtering Backward Sampling (FFBS)*

In *FFBS* we firstly run the Kalman filter and sample $\widetilde{\boldsymbol{x}}_T$ by supposing that $\boldsymbol{x}_T \sim N(\boldsymbol{a}_{T|T}, \boldsymbol{P}_{T|T})$. Sampled value is treated as an observation which allows us to perform one step backward sampling of $\boldsymbol{x}_{T-1}, \boldsymbol{x}_{T-2}, \dots, \boldsymbol{x}_1$ using the following recursions:

$$\boldsymbol{a}_{t|t,\widetilde{x}_{t+1}} = \boldsymbol{a}_{t|t} + \boldsymbol{P}_{t|t}\boldsymbol{U}_t'\boldsymbol{P}_{t+1}^{-1}(\widetilde{\boldsymbol{x}}_{t+1} - \boldsymbol{a}_{t+1}), \tag{21}$$

$$\boldsymbol{P}_{t|t,\widetilde{x}_{t+1}} = \boldsymbol{P}_{t|t} - \boldsymbol{P}_{t|t}\boldsymbol{U}_t'\boldsymbol{P}_{t+1}^{-1}\boldsymbol{U}_t\boldsymbol{P}_{t|t}. \tag{22}$$

See [62, Ch. 4] for details.

Thus, FFBS algorithm can be formulated as follows:

i. For each $t = 1,2, \dots, T$:
   a. Run Algorithm 5 (Kalman filter) to derive the distribution of $\boldsymbol{x}_T$.
   b. Sample $\widetilde{\boldsymbol{x}}_T$.

ii. For each $t = T-1, T-2, \dots, 1$:
   a. Calculate $\boldsymbol{a}_{t|t,\widetilde{x}_{t+1}}$ and $\boldsymbol{P}_{t|t,\widetilde{x}_{t+1}}$ using Equations (21) and (22).
   b. Sample $\widetilde{\boldsymbol{x}}_t$ by supposing that $\boldsymbol{x}_t \sim N(\boldsymbol{a}_{t|t,\widetilde{x}_{t+1}}, \boldsymbol{P}_{t|t,\widetilde{x}_{t+1}})$.

2.3.4. Gibbs sampler algorithm for estimation of parameters of state space Lee-Carter model

In this section detailed algorithms used for estimation of state space Lee-Carter model parameters are provided.

Algorithm 7. *Gibbs sampler for DLM Lee-Carter model*

i. Initiate the sampler using an arbitrary starting parameter collection $\boldsymbol{\psi} = \{\boldsymbol{\alpha}, \boldsymbol{\beta},\ \sigma_Q,\ \sigma_H,\ \boldsymbol{\mu}\}$ and collection of priors $\boldsymbol{\Theta} = \{\boldsymbol{m}_0, \boldsymbol{M}_0, g_0, G_0, \boldsymbol{b}_0, \boldsymbol{B}_0\}$.

ii. Sample

$$\tilde{\boldsymbol{x}}_{\boldsymbol{t}} = [1 \quad \mu^{(I)} \quad \mu^{(I)} \quad \kappa_t]',$$

conditionally on ψ and store vector $\boldsymbol{\kappa}$, by applying FFBS Algorithm 6, with the initial distribution

$$\boldsymbol{x}_{\boldsymbol{1}} \sim N\left(\begin{bmatrix} 1 \\ \mu^{(I)} \\ \mu^{(II)} \\ \kappa_0 \end{bmatrix}, \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \sigma_0^2 \end{bmatrix}\right).$$

Simulation results are usually insensitive to selection of κ_0, provided σ_0^2 is sufficiently large (we use $\sigma_0^2 = 10$).

iii. Sample $\boldsymbol{\mu} = [\mu^{(I)} \quad \mu^{(II)}]'$, conditionally on $\boldsymbol{\kappa}$ and σ_Q^2, from the bivariate Normal distribution:

$$N\left(\left(\boldsymbol{M}_0^{-1} + \frac{1}{\sigma_Q^2}\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\left(\boldsymbol{M}_0^{-1}\boldsymbol{m}_0 + \frac{1}{\sigma_Q^2}\boldsymbol{X}'\Delta\boldsymbol{\kappa}\right), \left(\boldsymbol{M}_0^{-1} + \frac{1}{\sigma_Q^2}\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\right),$$

where $\Delta\boldsymbol{\kappa}$ is a vector of changes in κ_t, and $\boldsymbol{X}$ is a matrix of dimensions $T\times 2$ having the form:

$$\boldsymbol{X} = \begin{bmatrix} 1 & 0 \\ \vdots & \vdots \\ 1 & 0 \\ 1 & 1 \\ \vdots & \vdots \\ 1 & 1 \end{bmatrix},$$

where number of first lines $T_0 \leq T$ is the number of periods before change in drift of κ_t occurred.

iv. Sample σ_Q^2, conditionally on $\boldsymbol{\kappa}$ and $\boldsymbol{\mu}$, from the inverse Gamma distribution:

$$IG\left(G_0 + \frac{T}{2}, g_0 + \frac{1}{2}\sum_{t=1}^{T}\left(\kappa_t - \kappa_{t-1} - \mu^{(I)} - \mathbb{I}_{\{t>T_0\}}\mu^{(II)}\right)^2\right),$$

where $\mathbb{I}_{\{t>T_0\}}$ is the standard indicator function.

v. Sample $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$, conditionally on $\boldsymbol{\kappa}$ and σ_H^2, from the bivariate Normal distribution. We assume that the matrix of measurement errors $\boldsymbol{H}_t$ is diagonal, therefore, we can sample α_x and β_x for each age group $x \in \{1,2,\dots,N\}$ separately:

$$\begin{bmatrix}\alpha_x\\ \beta_x\end{bmatrix} \sim N\left(\left(\boldsymbol{B}_0^{-1} + \frac{1}{\sigma_H^2}[\mathbf{1} \quad \boldsymbol{\kappa}]'[\mathbf{1} \quad \boldsymbol{\kappa}]\right)^{-1}\left(\boldsymbol{B}_0^{-1}\boldsymbol{b}_0 + \frac{1}{\sigma_H^2}[\mathbf{1} \quad \boldsymbol{\kappa}]'\boldsymbol{y}_x\right), \left(\boldsymbol{B}_0^{-1} + \frac{1}{\sigma_H^2}[\mathbf{1} \quad \boldsymbol{\kappa}]'[\mathbf{1} \quad \boldsymbol{\kappa}]\right)^{-1}\right),$$

where $\boldsymbol{y}_x$ is a vector of centered observations for age group x, and $[\mathbf{1} \quad \boldsymbol{\kappa}]$ is a matrix of dimensions T×2 where all elements of the first column are equal to 1.

vi. Sample σ_H^2, conditionally on $\boldsymbol{\kappa}$, $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$, from the inverse Gamma distribution:

$$IG\left(G_0 + \frac{TN}{2}, g_0 + \frac{1}{2}\sum_{x=1}^{N}\sum_{t=1}^{T}\left(y_{x,t} - \alpha_x - \beta_x\kappa_t\right)^2\right).$$

vii. Reweight $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ and related parameters by implementing Lee-Carter identifiability Constraints (4) and (5). We proceed as follows:

$$\boldsymbol{\beta}^{adj} = \frac{\boldsymbol{\beta}}{\sum_{x=1}^{N}\beta_x}, \quad \boldsymbol{\alpha}^{adj} = \boldsymbol{\alpha} + \bar{\kappa}\boldsymbol{\beta},$$

$$\boldsymbol{\mu}^{adj} = \sum_{x=1}^{N}\beta_x\left(\boldsymbol{\mu} - \begin{bmatrix}\bar{\kappa}\\ \bar{\kappa}\end{bmatrix}\right), \quad {\sigma_Q^2}^{adj} = \sigma_Q^2 \cdot \left(\sum_{x=1}^{N}\beta_x\right)^2,$$

where, $\bar{\kappa}$ is a mean κ_t.

viii. Collect the updated parameters to collection $\boldsymbol{\psi}$ and proceed to step (ii). Repeat the algorithm for the predetermined number of runs.

Algorithm 8. *Gibbs sampler for state space Lee-Carter model with switching*

i. Initiate the sampler using an arbitrary starting parameter collection $\boldsymbol{\psi}^{(s)} = \{\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\sigma}_Q, \sigma_H, \boldsymbol{\mu}, \boldsymbol{\pi}\}$ and collection of priors $\boldsymbol{\Theta}^{(s)} = \left\{\boldsymbol{m}_0, \boldsymbol{M}_0, g_0, G_0, \boldsymbol{b}_0, \boldsymbol{B}_0, p_0^{(1)}, p_0^{(2)}\right\}$ and an arbitrary vector $\boldsymbol{s} = [s_0, s_1, \dots, s_T]'$ of indexes 0 and 1.

ii. Sample

$$\tilde{\boldsymbol{x}}_t = [1 \quad \mu^{(I)} \quad \mu^{(I)} \quad \kappa_t]',$$

conditionally on $\boldsymbol{\psi}^{(\boldsymbol{s})}$ and regime vector $\boldsymbol{s}$, and stored vector $\boldsymbol{\kappa}$ of values κ_t, by applying FFBS Algorithm 6, with the initial distribution

$$\boldsymbol{x}_1 \sim N\left(\begin{bmatrix} 1 \\ \mu^{(I)} \\ \mu^{(II)} \\ \kappa_0 \end{bmatrix}, \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \sigma_0^2 \end{bmatrix}\right).$$

Simulation results are usually insensitive to selection of κ_0, provided σ_0^2 is sufficiently large (we use $\sigma_0^2 = 10$).

iii. Sample vector $\boldsymbol{s}$, conditionally on $\boldsymbol{\kappa}$, $\boldsymbol{\sigma}_Q$, $\boldsymbol{\mu}$ and $\boldsymbol{\pi}$, in the following way:

a. Calculate the initial probabilities of the two regimes by assuming that the filter is initialized with the stationary Markov chain probabilities:

$$\mathbb{P}(s_0 = 0) = \frac{1 - \pi^{(1)}}{2 - \pi^{(0)} - \pi^{(1)}},$$

$$\mathbb{P}(s_0 = 1) = \frac{1 - \pi^{(0)}}{2 - \pi^{(0)} - \pi^{(1)}}.$$

b. For $t \in \{1,2, \dots, T\}$ use the recursive procedure described in [14]:

- calculate priors

$$\mathbb{P}(s_t = 0|\boldsymbol{x}_{t-1}) = \mathbb{P}(s_{t-1} = 0|\boldsymbol{x}_{t-1})\pi^{(0)} + \mathbb{P}(s_{t-1} = 1|\boldsymbol{x}_{t-1})\left(1 - \pi^{(1)}\right),$$

$$\mathbb{P}(s_t = 1|\boldsymbol{x}_{t-1}) = \mathbb{P}(s_{t-1} = 1|\boldsymbol{x}_{t-1})\pi^{(1)} + \mathbb{P}(s_{t-1} = 0|\boldsymbol{x}_{t-1})\left(1 - \pi^{(0)}\right).$$

- calculate posteriors

$$\mathbb{P}^*(s_t = 0|\boldsymbol{x}_t) = \mathbb{P}(s_t = 0|\boldsymbol{x}_{t-1})\mathbb{P}(\boldsymbol{x}_t = \boldsymbol{X}_t|s_t = 0),$$

$$\mathbb{P}^*(s_t = 1|\boldsymbol{x}_t) = \mathbb{P}(s_t = 1|\boldsymbol{x}_{t-1})\mathbb{P}(\boldsymbol{x}_t = \boldsymbol{X}_t|s_t = 1).$$

In our case for $i \in \{0,1\}$ probability $\mathbb{P}(\boldsymbol{x}_t = \boldsymbol{X}_t|s_t = i)$ should be calculated by a Normal distribution with

mean $\kappa_t - \kappa_{t-1} - \mu^{(I)} - \mathbb{I}_{\{t>T_0\}}\mu^{(II)}$ and variance of $\left(\sigma_Q^{(i)}\right)^2$.

- Standardize the posteriors

$$\mathbb{P}(s_t = 0|\boldsymbol{x}_t) = \frac{\mathbb{P}^*(s_t = 0|\boldsymbol{x}_{t-1})}{\mathbb{P}^*(s_t = 0|\boldsymbol{x}_{t-1}) + \mathbb{P}^*(s_t = 1|\boldsymbol{x}_{t-1})},$$

$$\mathbb{P}(s_t = 1|\boldsymbol{x}_t) = \frac{\mathbb{P}^*(s_t = 1|\boldsymbol{x}_{t-1})}{\mathbb{P}^*(s_t = 0|\boldsymbol{x}_{t-1}) + \mathbb{P}^*(s_t = 1|\boldsymbol{x}_{t-1})}.$$

c. Sample s_T from the Bernoulli distribution with success probability $P(s_T = 1|\boldsymbol{x}_T)$.

d. For $t = T-1, T-2, \ldots, 1$ sample s_t recursively from the Bernoulli distribution with the one of the following success probabilities depending on the value of s_{t+1}

$$\mathbb{P}(s_t = 1|\boldsymbol{x}_t, s_{t+1} = 0) = \frac{\left(1-\pi^{(1)}\right)\mathbb{P}(s_t = 1|\boldsymbol{x}_t)}{(1-\pi^{(1)})\mathbb{P}(s_t = 1|\boldsymbol{x}_t) + \pi^{(0)}\mathbb{P}(s_t = 0|\boldsymbol{x}_t)},$$

$$\mathbb{P}(s_t = 1|\boldsymbol{x}_t, s_{t+1} = 1) = \frac{\pi^{(1)}\mathbb{P}(s_t = 1|\boldsymbol{x}_t)}{\pi^{(1)}\mathbb{P}(s_t = 1|\boldsymbol{x}_t) + (1-\pi^{(0)})\mathbb{P}(s_t = 0|\boldsymbol{x}_t)}.$$

e. Collect the sampled regime indicators to vector $\boldsymbol{s}$.

We remark here that steps (iv)–(vi) of the presented algorithm are constructed following results in [49].

iv. Sample $\boldsymbol{\pi} = [\pi^{(0)} \quad \pi^{(1)}]'$, conditionally on $\boldsymbol{s}$, from the beta distributions

$$\pi^{(0)} \sim Beta\left(p_0^{(1)} + n_{00}, p_0^{(2)} + n_{01}\right),$$

$$\pi^{(1)} \sim Beta\left(p_0^{(1)} + n_{11}, p_0^{(2)} + n_{10}\right),$$

where $p_0^{(1)}$ and $p_0^{(2)}$ are prior parameters and symbol n_{ij} denotes the number of transitions from the regime *i* to the regime *j* in the regime vector $\boldsymbol{s}$.

v. Sample $\boldsymbol{\mu} = [\mu^{(I)} \quad \mu^{(II)}]'$, conditionally on $\boldsymbol{s}$, $\boldsymbol{\kappa}$ and $\boldsymbol{\sigma_Q^2}$, from the bivariate Normal distribution:

$$N((\boldsymbol{M}_0^{-1} + \boldsymbol{X}^{*\prime}\boldsymbol{X}^{*})^{-1}(\boldsymbol{M}_0^{-1}\boldsymbol{m}_0 + \boldsymbol{X}^{*\prime}\Delta\boldsymbol{\kappa}^{*}), (\boldsymbol{M}_0^{-1} + \boldsymbol{X}^{*\prime}\boldsymbol{X}^{*})^{-1}),$$

where $\Delta\boldsymbol{\kappa}^{*}$ is a vector of changes in κ_t, divided by $\sigma_Q^{(i)}$ where the index *i* corresponds to regimes in vector $\boldsymbol{s}$, and $\boldsymbol{X}^{*}$ is a $T\times 2$ matrix of the form:

$$\boldsymbol{X}^{*} = \begin{bmatrix} 1/\sigma_Q^{(i)} & 0 \\ \vdots & \vdots \\ 1/\sigma_Q^{(i)} & 0 \\ 1/\sigma_Q^{(i)} & 1/\sigma_Q^{(i)} \\ \vdots & \vdots \\ 1/\sigma_Q^{(i)} & 1/\sigma_Q^{(i)} \end{bmatrix},$$

where number of first lines $T_0 \leq T$ is the number of periods before change in drift of κ_t occurred.

vi. Sample $\boldsymbol{\sigma_Q^2} = \left[\left(\sigma_Q^{(0)}\right)^2 \quad \left(\sigma_Q^{(1)}\right)^2\right]'$, conditionally on $\boldsymbol{s}$, $\boldsymbol{\kappa}$ and $\boldsymbol{\mu}$, from the inverse Gamma distribution.

a. Sample $\left(\sigma_Q^{(0)}\right)^2$ from the inverse Gamma distribution

$$IG\left(G_0 + \frac{T}{2}, g_0 + \frac{1}{2}\sum_{t=1}^{T}\left(\frac{\kappa_t - \kappa_{t-1} - \mu^{(I)} - \mathbb{I}_{\{t>T_0\}}\mu^{(II)}}{\sqrt{1+hs_t}}\right)^2\right),$$

and suppose that $h = \frac{\left(\sigma_Q^{(1)}\right)^2}{\left(\sigma_Q^{(0)}\right)^2} - 1$.

b. Sample $1 + h$ from the inverse Gamma distribution

$$IG\left(\begin{array}{c} G_0 + \frac{1}{2}\max\{t \leq T: s_t = 1\}, \\ g_0 + \frac{1}{2}\sum_{t:s_t=1}\left(\frac{\kappa_t - \kappa_{t-1} - \mu^{(I)} - \mathbb{I}_{\{t>T_0\}}\mu^{(II)}}{\sigma_Q^{(0)}}\right)^2 \end{array}\right),$$

where summation is performed over the moments $t \in \{1,2,\dots,T\}$ for which $s_t = 1$.

c. If the sampled $h > 0$, define $\left(\sigma_Q^{(1)}\right)^2 = (1+h)\left(\sigma_Q^{(0)}\right)^2$. Otherwise leave *h* from the previous run.

vii. Sample $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$, conditionally on $\boldsymbol{\kappa}$ and σ_H^2, from the bivariate Normal distribution. We assume that the matrix of measurement errors $\boldsymbol{H}_t$ is diagonal, therefore, we can sample α_x and β_x for each age group $x \in \{1,2,\dots,N\}$ separately:

$$\begin{bmatrix}\alpha_x \\ \beta_x\end{bmatrix} \sim N\left(\left(\boldsymbol{B}_0^{-1} + \frac{1}{\sigma_H^2}[\mathbf{1} \quad \boldsymbol{\kappa}]'[\mathbf{1} \quad \boldsymbol{\kappa}]\right)^{-1}\left(\boldsymbol{B}_0^{-1}\boldsymbol{b}_0 + \frac{1}{\sigma_H^2}[\mathbf{1} \quad \boldsymbol{\kappa}]'\boldsymbol{y}_x\right), \left(\boldsymbol{B}_0^{-1} + \frac{1}{\sigma_H^2}[\mathbf{1} \quad \boldsymbol{\kappa}]'[\mathbf{1} \quad \boldsymbol{\kappa}]\right)^{-1}\right),$$

where $\boldsymbol{y}_x$ is a vector of centered observations for age group *x*, and $[\mathbf{1} \quad \boldsymbol{\kappa}]$ is a matrix of dimensions *T*×*2* where all elements of the first column are equal to 1.

viii. Sample σ_H^2, conditionally on $\boldsymbol{\kappa}$, $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$, from the inverse Gamma distribution:

$$IG\left(G_0 + \frac{TN}{2}, g_0 + \frac{1}{2}\sum_{x=1}^{N}\sum_{t=1}^{T}\left(y_{x,t} - \alpha_x - \beta_x\kappa_t\right)^2\right).$$

ix. Reweight $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ and related parameters by implementing Lee-Carter identifiability Constraints (4) and (5). We proceed as follows:

$$\boldsymbol{\beta}^{adj} = \frac{\boldsymbol{\beta}}{\sum_{x=1}^{N}\beta_x}, \quad \boldsymbol{\alpha}^{adj} = \boldsymbol{\alpha} + \bar{\kappa}\boldsymbol{\beta},$$

$$\boldsymbol{\mu}^{adj} = \sum_{x=1}^{N}\beta_x\left(\boldsymbol{\mu} - \begin{bmatrix}\bar{\kappa} \\ \bar{\kappa}\end{bmatrix}\right), \quad {\sigma_Q^2}^{adj} = \sigma_Q^2 \cdot \left(\sum_{x=1}^{N}\beta_x\right)^2,$$

where, $\bar{\kappa}$ is a mean κ_t.

x. Collect the updated parameters to collection $\boldsymbol{\psi}^{(s)}$ and proceed to step (ii). Repeat the algorithm for the predetermined number of runs.

2.3.5. Forecasting and simulation

As noted in the previous subsection, the Gibbs sampler, after its convergence, produces draws from the posterior joint distribution of parameters. Thus, after discharging a certain number of initial draws (called burn or warm-up phase), we can use the draws as an input for the forecasting

simulations of the state space models. For both the linear model and the state space model with regime switching we simulate the projected mortality rates for periods $T+1, T+2, \dots, T+K$, where T is the length of time period used for model fitting and K is the term of the forecast, in the following way.

Algorithm 9. *Simulation using the state space Lee-Carter model*
For each iteration $j = 1, \dots, J$:

i. Depending on the model, DLM or SSM with regime switching, sample parameter collection $\boldsymbol{\psi}$ or $\boldsymbol{\psi}^{(s)}$ from the matrix of Gibbs sampler draws obtained during the model fitting phase after the warm-up. In this step, we are sampling the parameter values from their joint posterior distributions.

ii. For SSM with regime switching, draw a sample of regime indicators s for the forecasted period using step (iii) of Algorithm 8. As κ_t are not available for time moments $t = T+1, T+2, \dots, T+K$ multiplication by the probability $\mathbb{P}(\boldsymbol{x}_t|s_t)$ in position (iii.b) of Algorithm 8 may be disregarded.

iii. For SSM with regime switching, using sampled parameters $\boldsymbol{\mu}$ and $\boldsymbol{\sigma_Q^2}$ and regime vector $\boldsymbol{s}$ from step (ii) simulate κ_t for time moments $t = T+1, T+2, \dots, T+K$ by supposing that

$$\kappa_t \sim N\left(\kappa_{t-1} + \mu^{(I)} + \mu^{(II)}, \left(\sigma_Q^{(i)}\right)^2\right),$$

where $\left(\sigma_Q^{(i)}\right)^2$ is switching between $\left(\sigma_Q^{(0)}\right)^2$ and $\left(\sigma_Q^{(1)}\right)^2$ depending on simulated $\boldsymbol{s}$. In case of DLM, a fixed variance σ_Q^2 is used.

iv. Using sampled parameters $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$, σ_H^2 and sampled vector $\boldsymbol{\kappa}$ from step (iii), as well as allowing that errors are i.i.d., simulate $y_{x,t}$ for time moments $t = T+1, T+2, \dots, T+K$ and age groups $x = 1, \dots, N$ by supposing that

$$y_{x,t} \sim N(\alpha_x + \beta_x \kappa_t,\ \sigma_H^2).$$

v. Calculate mortality rates, having in mind that $m_{x,t} = e^{y_{x,t}}$ for all possible *x* and *t*.

2.3.6. Likelihood evaluation

For a comparison of state space models, the marginal log likelihood approach can be used. The comparison aims to decide which of the state space

Lee-Carter models should be used: with regime switching or more simple linear model. For this purpose the following formula of Chib [16] is applied:

$$m(\boldsymbol{Y}_{1:T}) = m(\boldsymbol{Y}_{1:T}|\widehat{\boldsymbol{\psi}}) + m(\widehat{\boldsymbol{\psi}}) - m(\widehat{\boldsymbol{\psi}}|\boldsymbol{Y}_{1:T}),$$

where $\boldsymbol{Y}_{1:T} = (\boldsymbol{y}_1, \boldsymbol{y}_2, \dots, \boldsymbol{y}_T)$, $m(\boldsymbol{Y}_{1:T})$ is the marginal log likelihood, $m(\boldsymbol{Y}_{1:T}|\widehat{\boldsymbol{\psi}})$ is log likelihood conditional on the estimated collection of parameters, $m(\widehat{\boldsymbol{\psi}})$ is log density of prior at $\widehat{\boldsymbol{\psi}}$ and $m(\widehat{\boldsymbol{\psi}}|\boldsymbol{Y}_{1:T})$ is the log density of posterior at $\widehat{\boldsymbol{\psi}}$.

The term $m(\boldsymbol{Y}_{1:T}|\widehat{\boldsymbol{\psi}})$ can be calculated in the closed form for linear state space models, but for SSM with regime switching, the likelihood function would be a mixture of 2^T Normal distributions, the direct assessment of which is not practical. Therefore, we use the sequential Monte Carlo method based on the auxiliary particle filter of Pitt and Shephard [63], which was adapted to SSM with regime switching by Kaufmann [47]. The idea of the method is for $r = 1,2, \dots, R$ separate chains ("particles") to sample sequentially $\left\{S_t^{(r)}|\boldsymbol{x}_{t-1}^{(r)}, S_{t-1}^{(r)}, \boldsymbol{y}_t\right\}$ and $\left\{\boldsymbol{x}_t|S_t^{(r)}, \boldsymbol{x}_{t-1}^{(r)}, \boldsymbol{y}_t\right\}$, and given the sampled values for each particle to calculate the likelihood. The $m(\boldsymbol{Y}_{1:T}|\widehat{\boldsymbol{\psi}})$ is estimated as the average of log likelihoods over the *R* particles. The details are provided in the following algorithm.

Algorithm 10. *Simulation of the state space Lee-Carter with switching model log likelihood conditional on the estimated collection of parameters*

i. With the parameter collection $\widehat{\boldsymbol{\psi}}^{(s)} = \{\widehat{\boldsymbol{\alpha}}, \widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{\sigma}}_Q, \hat{\sigma}_H, \widehat{\boldsymbol{\mu}}, \widehat{\boldsymbol{\pi}}\}$, initiate the filter by drawing *R* times with replacement κ_0 from the output of the Gibbs sampler and generating *R* draws of s_0 from the Bernoulli distribution with probability

$$P(s_0 = 0) = \frac{1 - \pi^{(1)}}{2 - \pi^{(0)} - \pi^{(1)}}.$$

ii. For each $r \in \{1, \dots, R\}$ and $t \in \{1, \dots, T\}$, conditionally on $\left\{s_{t-1}^{(r)}, \kappa_{t-1}^{(r)}\right\}$ sequentially sample $\left\{s_t^{(r)}, \kappa_t^{(r)}\right\}$ in the following way:

   a. Using the basic Kalman filter, for $i \in \{0,1\}$ derive one step ahead prediction

$$\mathbb{P}\left(\kappa_{t|t-1}^{(r)}|s_t = i, \kappa_{t-1}^{(r)}\right) \sim N\left(\bar{\kappa}_{t|t-1}^{(r)}, P_{t|t-1}^{(r)}(s_t = i)\right),$$

where $\bar{\kappa}_{t|t-1}^{(r)} = \kappa_{t-1}^{(r)} + \mu^{(I)} + \mu^{(II)}$ and $P_{t|t-1}^{(r)}(s_t = i) = P_{t-1|t-1}^{(r)} + \left(\sigma_Q^{(i)}\right)^2$. Obviously, this filter only considers the last state line of Equation (20) as the drifts are fixed.

b. For $i \in \{0,1\}$ find then matrix $\boldsymbol{F}_t^{(r)}(s_t = i) = \boldsymbol{\beta} P_{t|t-1}^{(r)}(s_t = i)\boldsymbol{\beta}' + \boldsymbol{H}_t$ and $\boldsymbol{v}_t^{(r)} = \boldsymbol{y}_t - \boldsymbol{\alpha} - \boldsymbol{\beta}\bar{\kappa}_{t|t-1}^{(r)}$.

c. For $i \in \{0,1\}$ calculate and standardize probabilities

$$\mathbb{P}\left(s_t^{(r)} = i | \boldsymbol{x}_{t-1}^{(r)}, s_{t-1}^{(r)}, \boldsymbol{y}_t\right) \propto$$

$$\left|\boldsymbol{F}_t^{(r)}(s_t = i)\right|^{-1/2} e^{-\frac{1}{2}\boldsymbol{v}_t^{(r)\prime}\left(\boldsymbol{F}_t^{(r)}(s_t=i)\right)^{-1}\boldsymbol{v}_t^{(r)}} \mathbb{P}\left(S_t = i | s_{t-1}^{(r)}\right).$$

d. Sample $s_t^{(r)}$ using the probabilities derived in step (ii.c).

e. Given the sampled $s_t^{(r)}$ with the basic Kalman filter, calculate the distribution of the prediction given the observation $\boldsymbol{y}_t$:

$$\mathbb{P}\left(\kappa_{t|t}^{(r)} | s_t^{(r)} = i, \kappa_{t|t-1}^{(r)}, \boldsymbol{y_t}\right) \sim$$

$$N\left(\begin{array}{c} \bar{\kappa}_{t|t-1}^{(r)} + P_{t|t-1}^{(r)}(s_t = i)\boldsymbol{\beta}'\left(\boldsymbol{F_t^{(r)}}(s_t = i)\right)^{-1}\boldsymbol{v_t^{(r)}}, \\ P_{t|t-1}^{(r)}(s_t = i) - P_{t|t-1}^{(r)}(s_t = i)\boldsymbol{\beta}'\left(\boldsymbol{F_t^{(r)}}(s_t = i)\right)^{-1}\boldsymbol{\beta} P_{t|t-1}^{(r)}(s_t = i) \end{array}\right).$$

f. Sample $\kappa_t^{(r)}$ and proceed to step (ii.a).

iii. From the draws $\left\{s_t^{(r)}, \kappa_t^{(r)}\right\}, t \in \{1,2,\dots,T\}$ calculate the one step likelihood by formula

$$f\left(\boldsymbol{y}_t | \boldsymbol{y}_{t-1}, \widehat{\boldsymbol{\psi}}^{(s)}\right) = \frac{1}{R}\sum_{r=1}^{R} f\left(\boldsymbol{y}_t | s_{t-1}^{(r)}, \kappa_{t-1}^{(r)}, \widehat{\boldsymbol{\psi}}^{(s)}\right),$$

where

$$f\left(\boldsymbol{y}_t | s_{t-1}^{(r)}, \kappa_{t-1}^{(r)}, \widehat{\boldsymbol{\psi}}^{(s)}\right) = \sum_{i=0,1} N\left(\boldsymbol{\beta}\kappa_{t|t-1}^{(r)}, \boldsymbol{F}_t^{(r)}(s_t = i)\right) \mathbb{P}\left(s_t = i | s_{t-1}^{(r)}\right),$$

follows from the model assumption that errors σ_H^2 do not depend on $\boldsymbol{y_{t-1}}$.

iv. Using the Markov property of the model, find the log likelihood conditional on $\widehat{\boldsymbol{\psi}}^{(s)}$

$$m\left(\boldsymbol{Y}_{1:T}|\widehat{\boldsymbol{\psi}}^{(s)}\right) = \sum_{t=1}^{T} \log f\left(\boldsymbol{y}_t|\boldsymbol{y}_{t-1}, \widehat{\boldsymbol{\psi}}^{(s)}\right).$$

The term $m\left(\widehat{\boldsymbol{\psi}}|\boldsymbol{Y}_{1:T}\right)$ we estimate similarly as by Chib [16] by running sequentially Gibbs sampler for each of the parameters in the collection $\boldsymbol{\psi} = \{\psi_1, \dots, \psi_L\}$ and estimating their posterior likelihood at the mean points $\widehat{\boldsymbol{\psi}}$ from MCMC runs using the following algorithm.

Algorithm 11. *Simulation of log density of posterior of parameters fitted with the Gibbs sampler*

i. For each of the parameters in the collection $\boldsymbol{\psi} = \{\psi_1, \dots, \psi_L\}$ perform the following:
   a. Estimate $m(\psi_1|\boldsymbol{Y}_{1:T})$ as a mean of the appropriate likelihood function of the parameter ψ_1 evaluated over the parameters of MCMC runs of the Gibbs sampler sampled as described in Algorithms 7 and 8.
   b. Estimate $m\left(\psi_2|\boldsymbol{Y}_{1:T}, \hat{\psi}_1\right)$ by fixing the parameter ψ_1 at the value $\hat{\psi}_1$ and by performing additional run of Gibbs sampler as in step (i.a).
   …
   z. Estimate $m\left(\psi_L|\boldsymbol{Y}_{1:T}, \hat{\psi}_1, \hat{\psi}_2, \dots, \hat{\psi}_{L-1}\right)$ by fixing the parameters $\psi_1, \psi_2, \dots, \psi_{L-1}$ at the values $\hat{\psi}_1, \hat{\psi}_2, \dots, \hat{\psi}_{L-1}$ and by performing an additional run of Gibbs sampler as in step (i.a).

ii. Calculate the total posterior log likelihood

$$m\left(\widehat{\boldsymbol{\psi}}|\boldsymbol{Y}_{1:T}\right) = \boldsymbol{m}(\psi_1|\boldsymbol{Y}_{1:T}) + \boldsymbol{m}\left(\psi_2|\boldsymbol{Y}_{1:T}, \hat{\psi}_1\right) + \cdots + \boldsymbol{m}\left(\psi_F|\boldsymbol{Y}_{1:T}, \hat{\psi}_1, \dots, \hat{\psi}_{F-1}\right).$$

2.4. Other specifications of stochastic mortality models

In this section, we provide a brief overview of other alternative stochastic mortality models and explain their applicability to the analysis performed in this dissertation.

2.4.1. Age-Period-Cohort model

The Age-Period-Cohort (APC) model (see [13], [17]) models log mortality rates for ages $x \in \{1,2,\dots,N\}$ and time periods $t \in \{1,2,\dots,T\}$ according to the following formula

$$\log(m_{x,t}) = \alpha_x + \frac{1}{N}\kappa_t + \frac{1}{N}\gamma_{t-x} + e_{x,t},$$

where α_x is the parameter of age effects, κ_t is the parameter of time effect and γ_{t-x} is the parameter of a cohort effect. Independent errors $e_{x,t}$ may have a Normal distribution, however, the APC model can be easily modeled using the Poisson regression specification. As in the case of the Lee-Carter model, the model is unidentified and it needs three location identifiability constraints.

The motivation of the model is that in addition to age and time effects, cohort effect on mortality rates is taken into account. Cohort (generation) effects, i.e., the dependence of mortality rates on a particular year of birth, were observed in England and Wales and the USA, in particular in studies covering higher ages. As discussed by Tuljapurkar and Boe [73], cohort effects may be attributable to a particular lifestyle or habits of different generations of a population as well as health conditions in childhood. However, cohort effects may be difficult to separate where data availability is limited (as in the Lithuanian case) and as noted by Hunt and Blake [43] inappropriate model specification may result in period effects be wrongly attributed to the cohort effects. Therefore, in this dissertation, we do not fit the cohort effects.

In comparison with the Lee-Carter model, the APC model does not allow for interaction between age and period effects, and it can be fitted using the standard GLM techniques. Thus, in this respect, it is a more simple model than the Lee-Carter model. As we shall see in the next section, the Lee-Carter model parameter β_x varies substantially between the age groups; therefore, the implicit assumption of the APC model that the Lee-Carter parameter β_x is the same for all age groups may result in undue simplification of the model.

2.4.2. Renshaw-Haberman model

A model proposed by Renshaw and Haberman [67] used the following overdispersed Poisson regression expressions for modeling of a number of deaths $D_{x,t}$ at ages $x \in \{1,2,\dots,N\}$ and time periods $t \in \{1,2,\dots,T\}$

$$\mathrm{E}(D_{x,t}) = E_{x,t}\mu_{x,t}, \qquad \mathrm{Var}(D_{x,t}) = \varphi\,\mathrm{E}(D_{x,t}),$$

with the log link and linear predictor

$$\eta_{x,t} = \log(E_{x,t}\mu_{x,t}) = \log(E_{x,t}) + \log(\mu_{x,t_c}) + \gamma_t(t - t_c) + \beta_x(t - t_c),$$

where $E_{x,t}$ is exposed to risk, t_c is time period at which the fitting internal is centered, γ_t is the parameter of time effects and β_x is the parameter of age effects. The other modification of the model includes an additional parameter to model one time change in age effects and in such a way to achieve a better fit.

In comparison with the Poisson Lee-Carter model, the Renshaw-Haberman model avoids the bilinear term and can be fitted using the standard IRLS Algorithm 2. The model replaces one bilinear term of the Lee-Carter model with two separate parameters, one of which, β_x, models age specific changes in mortality with time, and another, γ_t, the overall mortality trend with time. The term $\log(\mu_{x,t_c})$ plays a similar role as the parameter α_x in the Lee-Carter model. Thus, the key difference of the model is an approach to modeling of time: Renshaw and Haberman model time as a known covariate, not as a factor as in the Lee-Carter model.

Despite a more simple fitting procedure, the model has its shortcomings, such as the implicit simplification of modeling of age and time interaction as well as lower flexibility to forecasting. Therefore, for our analysis the Lee-Carter model is a preferred model.

2.4.3. Cairns-Blake-Dowd model

The basic Cairns-Blake-Dowd (CBD) model proposed by Cairns et al. [12] assesses log mortality initial mortality probabilities $q_{x,t}$ for ages $x \in \{1,2,\dots,N\}$ and time periods $t \in \{1,2,\dots,T\}$ as

$$\mathrm{logit}(q_{x,t}) = \log\left(\frac{q_{x,t}}{1 - q_{x,t}}\right) = \kappa_t^{(1)} + \kappa_t^{(2)}(x - \bar{x}) + e_{x,t},$$

where $\kappa_t^{(1)}$ is a random process which models the overall dynamics of mortality at all ages, factor $\kappa_t^{(2)}$ is a random process that mostly affects mortality dynamic at higher ages. The processes $\kappa_t^{(1)}$ and $\kappa_t^{(2)}$ were assumed to have the multivariate Normal distribution. Alternatively, the equation can

be easily modeled in the Poisson or Binomial GLM settings, see [17]. Later generalizations of the model, see [13], include cohort effect modeling as well as a quadratic term of the age effect.

The key shortcoming of the CBD model is that it does not allow modeling of age specific dynamics of the mortality rates. Therefore, the model is more suitable for application on data sets covering relatively short age spans. The age span analyzed in this dissertation is relatively wide and as demonstrated in Section 2, the sensitivity of mortality rates to time does not change linearly with age, therefore, the CBD model would give a poor fit for our data set.

2.4.4. Currie two dimensional spline model

The final model in this overview was proposed by Currie et al. [18] and is based on the Poisson GLM regression as in Equation (11), with the linear predictor modeled using two dimensional B-splines

$$\eta_{x,t} = \log\left(E_{x,t}\mu_{x,t}\right) = \log\left(E_{x,t}\right) + \sum_{i,j} \theta_{ij} B_{ij}(x,t),$$

where θ_{ij} are weights and $B_{ij}(x,t)$ are cubic B-spline basis functions. Additional smoothness in the model is introduced with the help of the penalty function which is applied as a modification to log likelihood.

The advantage of the model is its flexibility, as there is no need to assume a certain functional form of linear age and time effects, as well as their interactions. The disadvantage is reduced robustness for long term forecasts, in particular due to sensitivity of the forecasts to the choice of the penalty function, as well as the technical complexity of parameter estimation.

2.5. Summary of the section

Stochastic mortality models provide the basic input for the calculation of VaR for mortality risk. In this section, we provide a detailed overview of three alternative Lee-Carter stochastic mortality models: classical, Poisson, and state space. We provide the model specifications, model fitting methods, and algorithms as well as algorithms for simulation of mortality projections. For comparison, a brief overview of four stochastic mortality models alternative to the Lee-Carter model is also provided.

The classical Lee-Carter model is fitted using the singular value decomposition of the data matrix. We state the key results related to the SVD

model fitting method and discuss our key selections regarding the model specification. In the next sections, the classical Lee-Carter model will be mainly used as a benchmark for comparison with the other Lee-Carter model specifications.

The Poisson Lee-Carter model utilizes the generalized linear model framework with the aim of more realistic modeling of the variance of errors. In this section, we develop a model fitting algorithm adapted to the overdispersed Poisson version of the Lee-Carter model. The algorithm is based on the multivariate Newton-Raphson method and fits the bilinear term of the Lee-Carter model by subsequently running over columns and rows of the model matrix. In each run, the iteratively weighted least squares algorithm is executed.

Our key contribution is in the area of the state space Lee-Carter model. To overcome some common limitations of the classical Lee-Carter model, we proposed two new modifications of the state space Lee-Carter model: the state space Lee-Carter model with one-time change in drift and the state space Lee-Carter model with regime switching. We develop Gibbs sampler algorithms used to estimate model parameters, as well as algorithms used for the estimation of marginal likelihood, which can be used for model comparison. Mortality projection algorithms, which use parameter simulation results from MCMC runs, are developed as well.

3. MODEL FITTING RESULTS

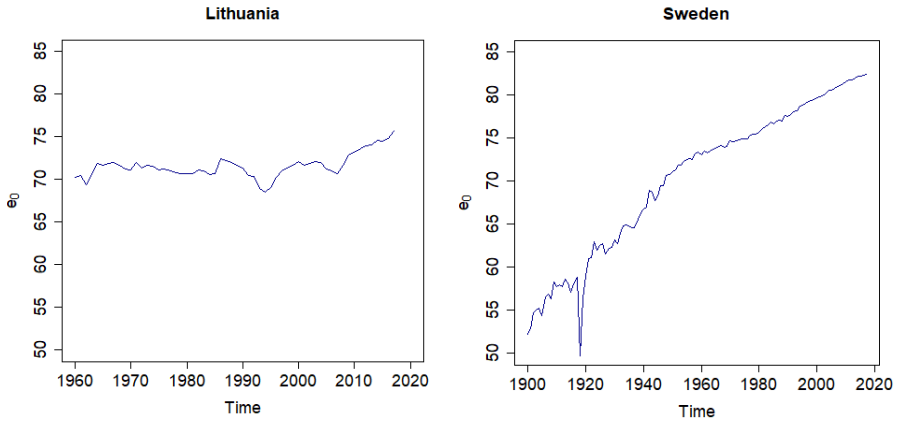
This section provides an overview of the data used in this dissertation, the results of the estimation of model parameters, as well as model fitting diagnostics.

3.1. Overview of the data

The study covers ages from 25 to 74 and is primary aimed at the assessment of mortality risk at insurance companies. However, the methods used can be extended to an analysis of mortality of older ages and can be used for assessment of longevity risk, which is particularly important for pension products providers. For the analysis we use the general population data from Human Mortality Database for ages 25-74 grouped in 5 year age groups for the following periods: Lithuania 1959-2017, Sweden 1900-2017. The Human Mortality Database (https://www.mortality.org) is maintained by the University of California, Berkeley (USA), and the Max Planck Institute for Demographic Research (Germany). Data for our research was downloaded in February 2020.

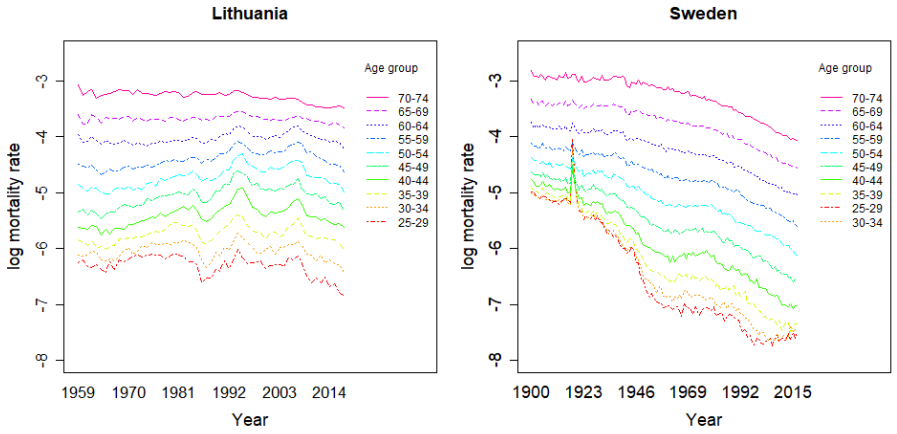
The initial overview of the data in Figure 1 shows that the development of periodic life expectancy on birth in Lithuania was very different from Sweden. Periodic life expectancy is the estimated life expectancy on birth based on the period's mortality tables, i.e., it does not take into account the projected changes in future mortality. Sweden experienced a consistent and almost linear decline in log mortality rates and the corresponding increase in life expectancy on birth, which correspond to the experience in G7 countries, see [74]. The major deviation from the trend was during the Spanish flu pandemics in 1918-1920, which led to a dramatic increase in the mortality rates, especially for younger age groups, see Figure 2.

The development of mortality rates in Lithuania was different: slight improvement of life expectancy in the 1960s was followed by stagnation in 1970s and 1980s. At the beginning of the 1990s Lithuania, like many other countries of the former Soviet Union, experienced a sharp increase in mortality rates. As illustrated in Figure 2, the increase coincided with the collapse of the Soviet Union and the start of the transition to a market economy and was especially high for younger age groups.



(a) **(b)**

Figure 1. Development of periodic life expectancy at birth in Lithuania (a), Sweden (b) according to data from the Human Mortality Database.



(a) **(b)**

Figure 2. Development of log mortality rates in Lithuania (a) during 1959-2017 and Sweden (b) during 1900-2017 according to data from the Human Mortality Database.

Several studies investigated this phenomenon in different countries of the former Soviet Union. For example, Brainerd and Cutler [9] investigated six factors that possibly had led to the increase in Russian mortality and concluded that the increase in alcohol consumption and stress caused by the transition are likely to be the key causes of this phenomenon.

In analyzing Lithuanian mortality statistics we should be careful about the possible data quality issues. Statistical systems and processes experienced a major development during the analyzed period. Therefore, we note the possible limitations related to the quality of data for further analysis and interpretation of the results.

3.2. Estimation of the parameters and model diagnostics

In this subsection, we provide estimates of models' parameters and describe model diagnostics. We start with the estimates of the classical and Poisson Lee-Carter models and later proceed to the state space Lee-Carter models.

For model diagnostics, we use the analysis of residuals, statistical unit root tests, analysis of parameter estimates, and model specific diagnostic tests, such as MCMC convergence tests.

3.2.1. Estimation of classical and Poisson Lee-Carter models parameters

We performed parameter estimation of classical Lee-Carter and Poisson Lee-Carter models using two data sets: full dataset and dataset trimmed to the latest experience, which is assumed to reflect the latest trends. In such a way we can compare the estimates and understand the impact on parameters introduced by the additional data. In the Lithuanian case, the trimmed data set comprises data of years 1995-2017, and in the Swedish case, years 1960-2017. Trimmed data sets are mainly used to develop alternative estimates for comparison and illustration. In selecting the periods for trimmed datasets we have performed visual analysis, used period selections in other studies [74], and performed an assessment of the sensitivity of selection of the initial years.

Table 3 provides a summary of statistics on the fit of the models.

Table 3. The proportion of variance explained by the first principal component of SVD fit and estimated overdispersion parameter of the Poisson regression fit. The results are provided for estimates performed using full and trimmed data sets. The overdispersion parameter was estimated using Equation (8).

	Lithuania		**Sweden**	
	1959-2017	**1995-2017**	**1901-2017**	**1960-2017**
SVD: proportion of variance 1st principal component	0.73	0.91	0.97	0.97
Poisson: estimated over-dispersion	9.18	3.51	31.91	3.34

The Lithuanian model fitted to the full data set has the lowest percentage of variance explained by the first principal component. This may be explained by the impact of high volatility during the early 1990s and the change in the trend of mortality development. For Sweden and Lithuanian models fitted to a trimmed data set, the proportion explained by the first principal component is adequately high. Similarly, overdispersion parameters, when estimated from full data sets (especially in the Swedish case) indicate significant additional variability, unexplained by the Poisson model. The overdispersion is much lower when derived using trimmed datasets.

To check the goodness of fit the analysis of residuals was performed, which comprised an analysis of Normality and autocorrelation of residuals.

A normality check was done by visual analysis quantile/ quantile (Q/Q) plots which are provided in Figures 3 and 4. The plots for Lithuania did not reveal any major deviations. For Sweden, the plots for the model fitted using full data set indicate the Spanish flu pandemic years as a clear outlier in both state and observation equations.

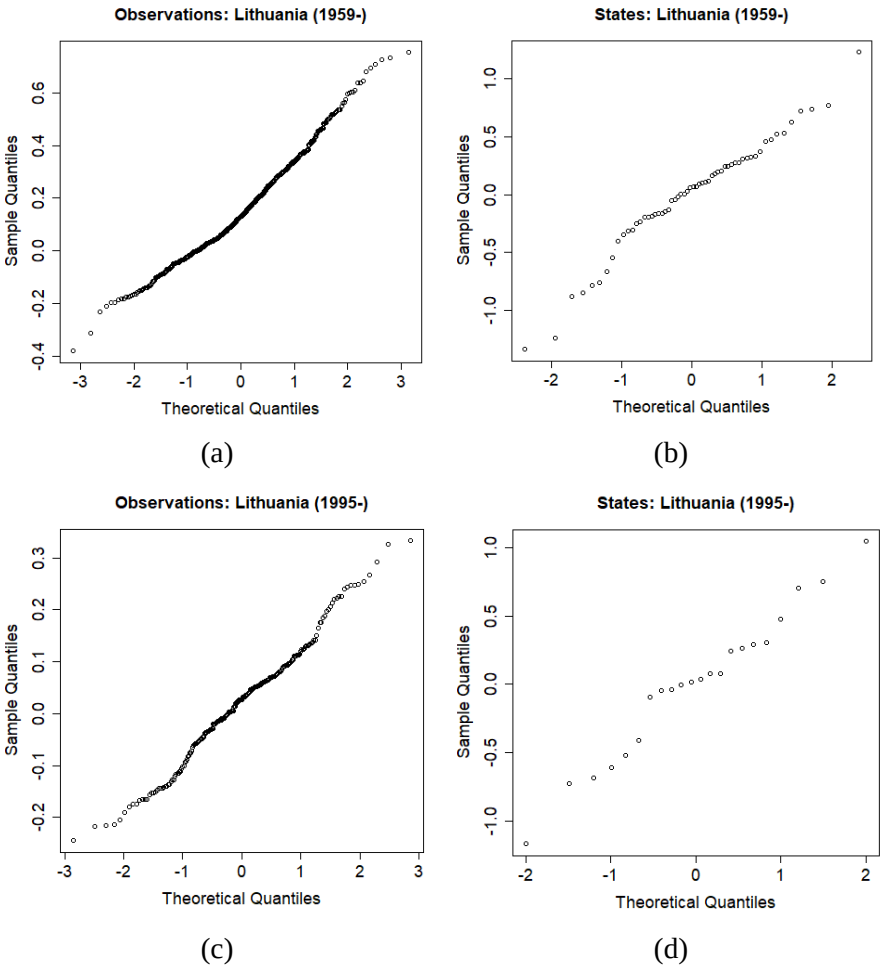


Figure 3. Q/Q plots of residuals for observation and state equations to check the Normality assumption for models fit using the full **Lithuanian data set** (panels (a) and (b)) and trimmed dataset (panels (c) and (d))

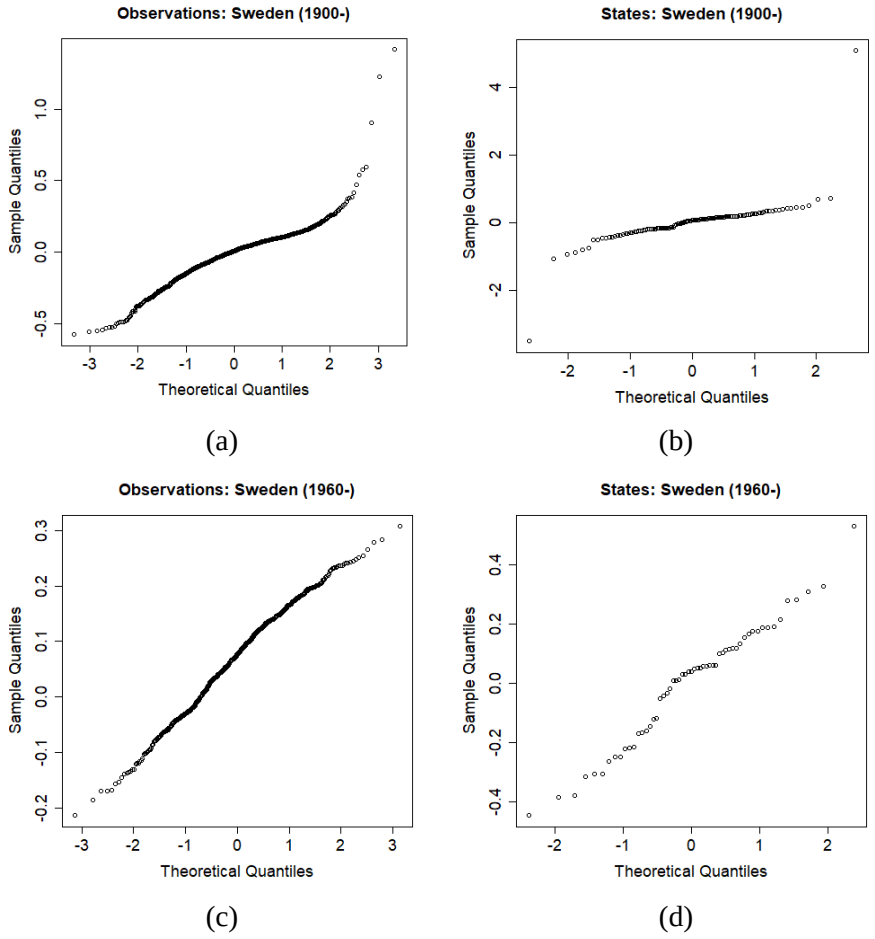


Figure 4. Q/Q plots of residuals for observation and state equations to check the Normality assumption for models fit using the full **Swedish data set** (panels (a) and (b)) and trimmed dataset (panels (c) and (d)).

The analysis of autocorrelations of the residuals did not reveal any significant autocorrelations.

For Poisson bilinear regression, the analysis of Pearson residuals, as illustrated in Figure 5, revealed major outliers when the model was fitted using the full Swedish dataset. We also performed the analysis of residuals of state equation of parameter κ_t, derived using Poisson bilinear regression. Its results are comparable to the results of an analysis of state equation residuals from the SVD model as illustrated in Figures 3 and 4, therefore, they are not provided. Overall, the Poisson bilinear regression provides a reasonable fit for both Lithuanian datasets and Swedish trimmed dataset.

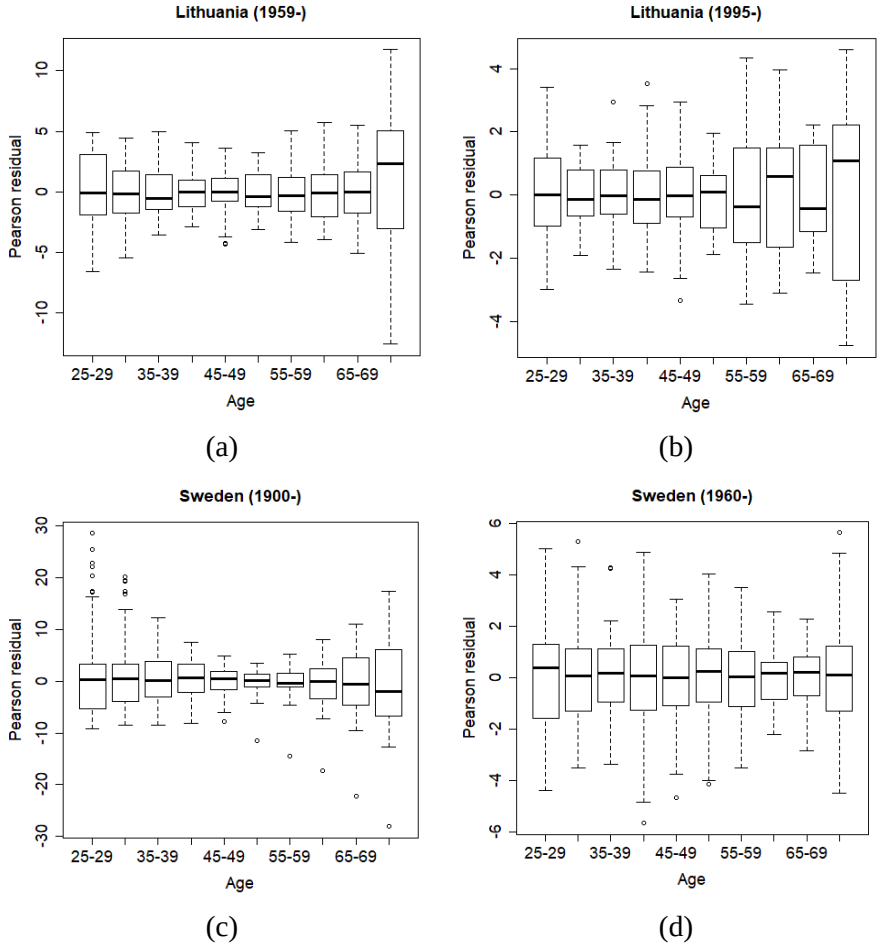


Figure 5. Plots of Pearson residuals by age for observation equations for Poisson bilinear regression models fit using the Lithuanian full and trimmed data sets (panels (a) and (b)) and the Swedish full and trimmed data sets (panels (c) and (d)).

To assess the reasonableness of the RWD model assumption for the time varying index κ_t we performed tests as described in Subsection 2.1.3. A summary of the test results is provided in Table 4. The Phillips-Perron test statistics were calculated using the pp.test function of R package tseries. The 2-lag model was used for the Lithuanian (1995-2017) model and the 3-lag model was used for the remaining models. Critical values at 95% confidence level are provided in the brackets are from [39], (see Tables B5 and B6).

Overall, the results of the tests support the hypotheses that κ_t can be modeled as time series with unit root.

Table 4. Statistical testing for unit-roots for the time series parameter of κ_t. Brackets [ ] provide the 95% confidence level critical values from [39, tables B5 and B6].

	Lithuania		**Sweden**	
	1959-2017	**1995-2017**	**1901-2017**	**1960-2017**
SVD model:				
Phillips-Perron test				
Z_ρ	-4.64 [-19.8]	-6.70 [-17.9]	-19.2 [-20.7]	-3.8 [-19.8]
Z_τ	-1.07 [-3.5]	-1.69 [-3.6]	-3.22 [-3.45]	-2.0 [-3.5]
Augmented Dickey-Fuller test				
$Z_{DF}(\rho)$	-7.20 [-19.8]	-8.36 [-17.9]	-9.8 [-20.7]	-6.4 [-19.8]
Z_{DF}	-1.05 [-3.5]	-1.34 [-3.6]	-2.1 [-3.45]	-3.1 [-3.5]
Poisson model:				
Phillips-Perron test				
Z_ρ	-4.15 [-19.8]	-6.77 [-17.9]	-16.5 [-20.7]	-2.5 [-17.9]
Z_τ	-0.95 [-3.5]	-1.72 [-3.6]	-2.9 [-3.45]	-1.5 [-3.5]
Augmented Dickey-Fuller test				
$Z_{DF}(\rho)$	-8.76 [-19.8]	-5.80 [-17.9]	-7.4 [-20.7]	-3.7 [-17.9]
Z_{DF}	-1.11 [-3.5]	-1.08 [-3.6]	-1.8 [-3.45]	-2.8 [-3.5]

3.2.2. Estimation of the state space Lee-Carter model parameters

To obtain the state space model estimates, we run Algorithms 7 and 8 5000 times for each of five different starting values, and we discharge the first 1000 draws as a warm-up. For the basic Kalman filter and FFBS calculations, we use the functions from *R* package ***dlm***, see [62]. After recording the results of the runs we split each chain into two equal parts and perform MCMC

diagnostics by assessing convergence and the efficiency of MCMC sampling according to Gelman et al. [31].

Convergence is assessed by calculating for each sampled parameter:

$$\hat{R} = \sqrt{\frac{\widehat{\mathrm{Var}}(\Lambda|\boldsymbol{Y})}{W}},$$

where:

$$\widehat{\mathrm{Var}}(\Lambda|\boldsymbol{Y}) = \frac{n-1}{n}W + \frac{1}{n}B, \tag{21}$$

is an estimate of the parameter's Λ marginal posterior variance, conditional on the data $\boldsymbol{Y}$, W is the average of sample variances of a parameter within each of m half-chains of length n, and B is the sample variance of a parameter between the half-chains. If $\hat{R}$ is close to 1 than B is small relatively to W, it implies that there are no major consistent deviations due to different starting values or between early/late samples in a chain and supports the assumption of adequate convergence.

Sampling efficiency is assessed by calculating effective sample size:

$$n_{eff} = \frac{mn}{1 + 2\sum_{t=1}^{T}\rho_t},$$

where ρ_t are autocorrelations of parameter samples in the chains. To limit the effect of noise of sample correlations we used T to be the first positive integer for which $\rho_{T+1} + \rho_{T+2}$ is negative.

In the Lithuanian case, we allow for the additional drift element since 1995 when the country finalized the major reforms and started independent European development. We consider the country's independence a major change in socio-economic development, which significantly altered the trend of mortality development. In the Swedish case, no such changes occurred during the fitting period, thus variation in the drift is not modeled. In summary, the fitted drift parameters are summarized in Table 5.

Table 5. Time periods modeled by drifts of time varying index when fitting with the Lithuanian and Swedish data.

	Lithuania	**Sweden**
$\mu^{(I)}$	1959-2017	1900-2017
$\mu^{(II)}$	1995-2017	Not used

For the DLM and SSM with switching we use the following collections of priors:

$$\boldsymbol{\Theta} = \begin{Bmatrix} \boldsymbol{m}_0 = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \boldsymbol{M}_0 = \begin{pmatrix} 5 & 0 \\ 0 & 5 \end{pmatrix}, g_0 = 0.1, G_0 = 2.1, \boldsymbol{b}_0 = \begin{pmatrix} -4 \\ 0.1 \end{pmatrix}, \\ \boldsymbol{B}_0 = \begin{pmatrix} 5 & 0 \\ 0 & 5 \end{pmatrix} \end{Bmatrix},$$

$$\boldsymbol{\Theta}^{(s)} = \begin{Bmatrix} \boldsymbol{m}_0 = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \boldsymbol{M}_0 = \begin{pmatrix} 5 & 0 \\ 0 & 5 \end{pmatrix}, g_0 = 0.1, G_0 = 2.1, \boldsymbol{b}_0 = \begin{pmatrix} -4 \\ 0.1 \end{pmatrix}, \\ \boldsymbol{B}_0 = \begin{pmatrix} 5 & 0 \\ 0 & 5 \end{pmatrix}, \ p_0^{(1)} = p_0^{(2)} = 1 \end{Bmatrix}.$$

We set priors with the aim, as suggested in [31 p. 55], that the information they provide is intentionally weaker than the actual knowledge that is available. For example, for Normally distributed parameters, we set priors with sufficiently large variances, not to introduce an undue bias towards the selected prior mean.

The calculated convergence and sampling efficiency statistics for models fitted using Gibbs sampler are provided in Table 6.

The statistics $\hat{R}$ shows adequate convergence for both models and both countries. Effective sample size indicates the high efficiency of the Gibbs sampler for most of the parameters, with lower efficiency for SSM with switching variances and transition probabilities, especially in the Lithuanian case. Overall, the performance of the Gibbs sampler is considered appropriate in our case.

In some cases, for validating goodness of fit, back-testing methods are used. Such methods involve model fitting using the dataset with the latest observations excluded and a comparison of the resulting projections with the actual experience. In our case, the possibility to apply such methods for the newly developed state space Lee-Carter models was limited due to allowance in the change of trend. With respect to the SSM with regime switching, back-testing is also problematic, as it is very difficult to assign the actual outcome to the specific chain of modeled regimes. Nevertheless, we have compared the actual mortality experience in 2019 with the corresponding confidence intervals derived using the models. Overall, the actual rates lie within the confidence intervals, just the lowest and the highest age groups in Lithuania were marginally below the lower threshold, which can be explained by relativelly higher volatility at lower ages and different dynamics of mortality development of higher age population.

Table 6. Estimates of convergence statistics $\hat{R}$ and effective sample size statistics n_{eff} (presented in thousands) for DLM and SSM with switching model parameters, calculated for five MCMC chains of 4000 iterations after burn-in (total of 20 thousand iterations).

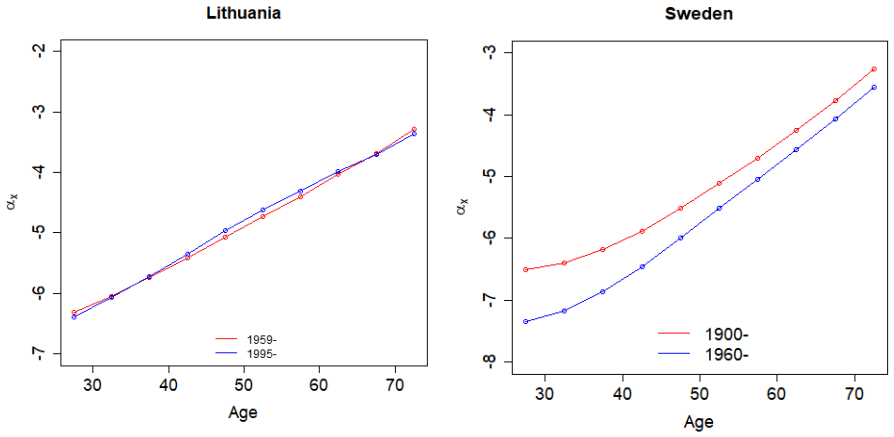
Parameter	Lithuania				Sweden			
	DLM		SSM with switching		DLM		SSM with switching	
	$\widehat{R}$	$\widehat{n}_{eff}$	$\widehat{R}$	$\widehat{n}_{eff}$	$\widehat{R}$	$\widehat{n}_{eff}$	$\widehat{R}$	$\widehat{n}_{eff}$
α_{25-29}	1.0	20.1	1.0	20.0	1.0	20.0	1.0	18.3
α_{30-34}	1.0	20.4	1.0	19.5	1.0	20.0	1.0	20.1
α_{35-39}	1.0	18.6	1.0	20.0	1.0	19.6	1.0	19.7
α_{40-44}	1.0	20.0	1.0	19.7	1.0	20.1	1.0	20.0
α_{45-49}	1.0	20.1	1.0	20.1	1.0	19.7	1.0	19.9
α_{50-54}	1.0	20.0	1.0	20.3	1.0	20.4	1.0	19.0
α_{55-59}	1.0	19.8	1.0	19.6	1.0	19.9	1.0	20.0
α_{60-64}	1.0	19.8	1.0	20.0	1.0	19.4	1.0	19.2
α_{65-69}	1.0	19.3	1.0	20.0	1.0	18.9	1.0	20.2
α_{70-74}	1.0	20.4	1.0	20.3	1.0	20.4	1.0	19.1
β_{25-29}	1.0	12.8	1.0	11.8	1.0	18.7	1.0	19.8
β_{30-34}	1.0	16.8	1.0	17.8	1.0	19.6	1.0	20.3
β_{35-39}	1.0	19.9	1.0	19.5	1.0	18.0	1.0	19.4
β_{40-44}	1.0	17.2	1.0	18.6	1.0	20.2	1.0	19.9
β_{45-49}	1.0	16.0	1.0	17.6	1.0	20.2	1.0	19.5
β_{50-54}	1.0	18.5	1.0	16.2	1.0	20.2	1.0	19.9
β_{55-59}	1.0	17.9	1.0	17.4	1.0	20.0	1.0	20.2
β_{60-64}	1.0	18.0	1.0	19.5	1.0	20.1	1.0	20.0
β_{65-69}	1.0	19.9	1.0	19.5	1.0	19.6	1.0	20.2
β_{70-74}	1.0	16.8	1.0	17.6	1.0	16.9	1.0	19.9
$\left(\sigma_Q^{(0)}\right)^2$	1.0	8.5	1.0	1.0	1.0	3.4	1.0	1.8
$\left(\sigma_Q^{(1)}\right)^2$	n/a	n/a	1.0	2.6	n/a	n/a	1.0	7.2
σ_H^2	1.0	15.7	1.0	15.8	1.0	13.1	1.0	17.7
$\mu^{(I)}$	1.0	18.7	1.0	13.9	1.0	19.4	1.0	13.5
$\mu^{(II)}$	1.0	18.8	1.0	12.8	n/a	n/a	n/a	n/a
$\pi^{(0)}$	n/a	n/a	1.0	1.0	n/a	n/a	1.0	3.9
$\pi^{(1)}$	n/a	n/a	1.0	0.7	n/a	n/a	1.0	8.4

3.2.3. Summary of parameter estimates

In this subsection we provide and discuss estimates of the Lee-Carter model parameters. The following subsection provides more details on regime switching parameters.

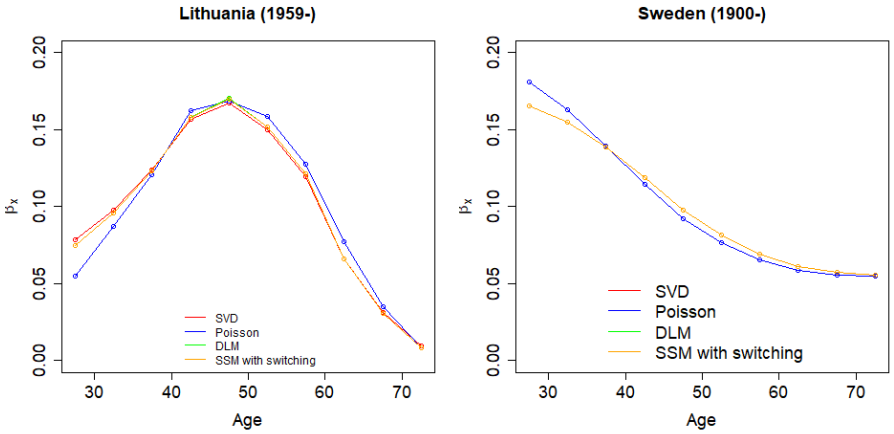
Figures 6 and 7 summarize the results of the estimation of parameters α_x and β_x. The estimates of parameters of α_x, derived using different models, coincide, therefore, the plots are not provided. When parameters are derived using data of different periods, they reflect the average mortality by age group of that period. We see that for Lithuania, the average mortality did not change significantly, while for Sweden a major decrease in the overall level was observed. The estimates of parameter β_x is slightly more model-dependent. In particular, estimates derived using Poisson model provides slightly different estimates due to different variance structure assumption.

Overall, the level and shape of parameter estimates do not depend significantly on the model used, although the level of mortality differs from country to country and from period to period.



(a) (b)

Figure 6. Lee-Carter model parameter α_x estimated for various fitting periods for Lithuania (a) and Sweden (b).



(a) (b)

Figure 7. Lee-Carter model parameter β_x estimated using SVD, Poisson, DLM and SSM with switching models for Lithuania (a) and Sweden (b).

Tables 7(a) and 7(b) summarize the estimates of variance and drift parameters derived using different models. The estimates of the drift parameter $\mu^{(I)}$ are similar for different models in the Swedish case but are different in the Lithuanian case due to allowance for the change in drift in state-space models. Estimates of variance $\left(\sigma_Q^{(0)}\right)^2$ and σ_H^2 are generally lower for state-space models than for SVD or Poisson models which indicates a better fit of the state space models.

Table 7(a). Estimates of variance and drift parameters derived using SVD, Poisson, DLM, and SSM with switching models for Lithuania.

Para-meter	SVD (1995-)	SVD (1959-)	Poisson (1995-)	Poisson (1959-)	DLM	SSM with switch.
$\left(\sigma_Q^{(0)}\right)^2$	0.268	0.227	0.265	0.195	0.176	0.113
$\left(\sigma_Q^{(1)}\right)^2$	n/a	n/a	n/a	n/a	n/a	0.250
σ_H^2	0.013	0.058	n/a	n/a	0.007	0.007
$\mu^{(I)}$	-0.243	-0.021	-0.247	-0.021	0.094	0.099
$\mu^{(II)}$	n/a	n/a	n/a	n/a	-0.267	-0.273

Table 7(b). Estimates of variance and drift parameters derived using SVD, Poisson, DLM and SSM with switching models for Sweden.

Para-meter	SVD (1960-)	SVD (1900-)	Poisson (1960-)	Poisson (1900-)	DLM	SSM with switch.
$\left(\sigma_Q^{(0)}\right)^2$	0.040	0.425	0.025	0.474	0.167	0.049
$\left(\sigma_Q^{(1)}\right)^2$	n/a	n/a	n/a	n/a	n/a	5.442
σ_H^2	0.014	0.026	n/a	n/a	0.013	0.012
$\mu^{(I)}$	-0.145	-0.153	-0.158	-0.184	-0.152	-0.146

Table 8 provides estimates of parameters of the state space models, together with the estimated standard deviations. The majority of the marginal posterior standard deviations are relatively low. As expected, higher variances are recorded for parameters that are estimated from a smaller number of input variables, such as the Lithuanian drifts and the regime-specific parameters in the SSM with switching model.

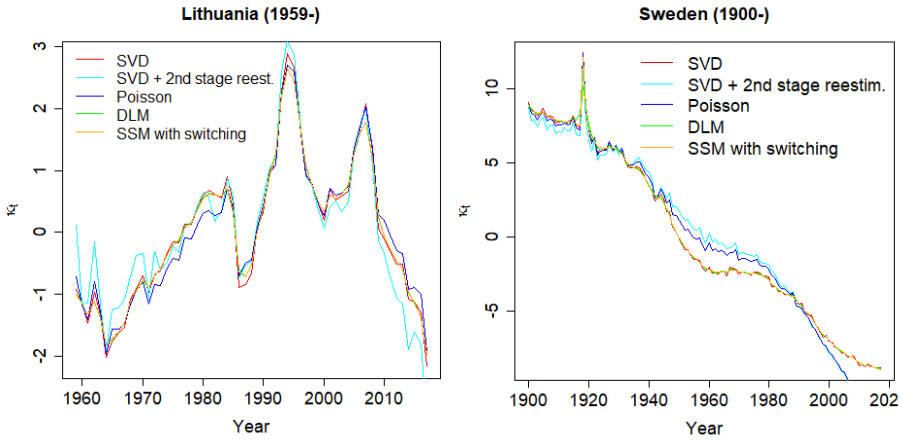
For SSM with switching the variance σ_Q^2 is split into two parts: $\left(\sigma_Q^{(0)}\right)^2$ is applicable to low variance regimes and $\left(\sigma_Q^{(1)}\right)^2$ is applicable to high variance regime. We observe that for Sweden there is a big difference between $\left(\sigma_Q^{(0)}\right)^2$ and $\left(\sigma_Q^{(1)}\right)^2$, which indicates that the two regimes are very different in terms of the volatility of parameter κ_t, but this is not the case of Lithuania. Parameters $\pi^{(0)}$ and $\pi^{(1)}$ are applicable to SSM with switching models only and they indicate that for Sweden, regime 0 has high frequency and regime 1 has low frequency. In the Lithuanian case, both regimes have a similar frequency. Thus, we can conclude that in the Swedish case SSM with switching has converged to two distinct regimes, while in the Lithuanian case the movements in κ_t were not sharp enough to achieve this. For more details see Subsection 3.2.4.

Table 8. Estimates of parameters using DLM and SSM with switching. Marginal posterior standard deviations are estimated using Formula (21).

Parameter	Lithuania				Sweden			
	DLM		SSM with switching		DLM		SSM with switching	
	mean	s.d.	mean	s.d.	mean	s.d.	mean	s.d.
α_{25-29}	-6.306	0.011	-6.306	0.011	-6.512	0.010	-6.513	0.010
α_{30-34}	-6.046	0.011	-6.046	0.011	-6.403	0.010	-6.403	0.010
α_{35-39}	-5.742	0.011	-5.742	0.011	-6.186	0.010	-6.186	0.010
α_{40-44}	-5.412	0.011	-5.412	0.011	-5.882	0.010	-5.882	0.010
α_{45-49}	-5.066	0.011	-5.066	0.011	-5.510	0.010	-5.510	0.010
α_{50-54}	-4.729	0.011	-4.728	0.011	-5.109	0.010	-5.109	0.010
α_{55-59}	-4.401	0.011	-4.401	0.011	-4.695	0.010	-4.695	0.010
α_{60-64}	-4.033	0.011	-4.033	0.011	-4.245	0.010	-4.245	0.010
α_{65-69}	-3.683	0.011	-3.683	0.011	-3.766	0.010	-3.767	0.010
α_{70-74}	-3.276	0.011	-3.276	0.011	-3.263	0.010	-3.263	0.010
β_{25-29}	0.075	0.010	0.075	0.010	0.165	0.002	0.165	0.002
β_{30-34}	0.096	0.010	0.096	0.010	0.155	0.002	0.155	0.002
β_{35-39}	0.123	0.010	0.123	0.010	0.139	0.002	0.139	0.002
β_{40-44}	0.158	0.010	0.158	0.010	0.119	0.002	0.119	0.002
β_{45-49}	0.170	0.010	0.170	0.010	0.098	0.002	0.098	0.002
β_{50-54}	0.152	0.010	0.152	0.010	0.082	0.002	0.082	0.002
β_{55-59}	0.121	0.010	0.121	0.010	0.069	0.002	0.069	0.002
β_{60-64}	0.066	0.010	0.066	0.010	0.061	0.002	0.061	0.002
β_{65-69}	0.030	0.010	0.030	0.010	0.057	0.002	0.057	0.002
β_{70-74}	0.008	0.010	0.008	0.010	0.056	0.002	0.056	0.002
$\left(\sigma_Q^{(0)}\right)^2$	0.176	0.043	0.113	0.053	0.167	0.040	0.049	0.014
$\left(\sigma_Q^{(1)}\right)^2$	n/a	n/a	0.250	0.117	n/a	n/a	5.442	4.112
σ_H^2	0.007	0.001	0.007	0.001	0.013	0.001	0.012	0.001
$\mu^{(I)}$	0.094	0.093	0.099	0.094	-0.152	0.074	-0.146	0.068
$\mu^{(II)}$	-0.267	0.129	-0.273	0.130	n/a	n/a	n/a	n/a
$\pi^{(0)}$	n/a	n/a	0.507	0.275	n/a	n/a	0.973	0.020
$\pi^{(1)}$	n/a	n/a	0.562	0.283	n/a	n/a	0.574	0.200

Figure 8 provides the comparison of parameters κ_t derived using different models applied to the full dataset. Overall, the estimates are similar, except

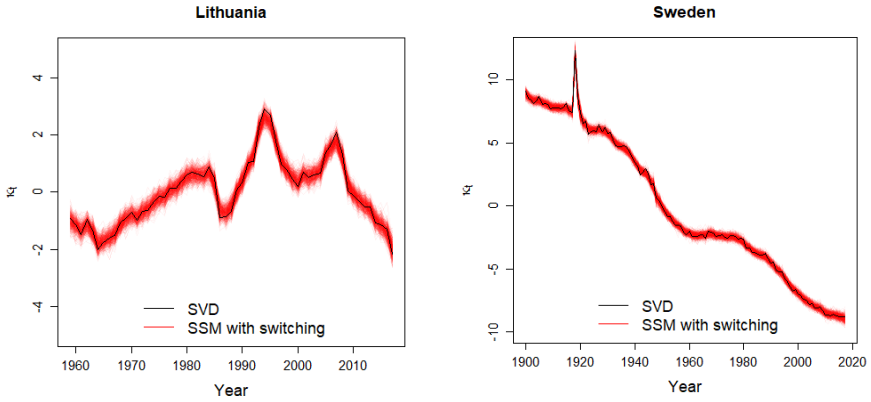
for the classical Lee-Carter model with the second stage re-estimation of the parameter. The results support the assumption that the second stage re-estimation may introduce a bias, therefore, is not applied in our analysis.



(a) (b)

Figure 8. Estimation of the parameter κ_t for Lithuania (a) and Sweden (b) using different models.

Figure 9 provides the comparison of estimates of time varying index κ_t obtained from SVD and from SSM with switching MCMC runs. We observe that in both cases the shape of the basic κ_t curve is similar, but in the MCMC runs we allow for uncertainty of κ_t under various samples of the parameter collection $\psi^{(s)}$, thus the values of κ_t fluctuate.

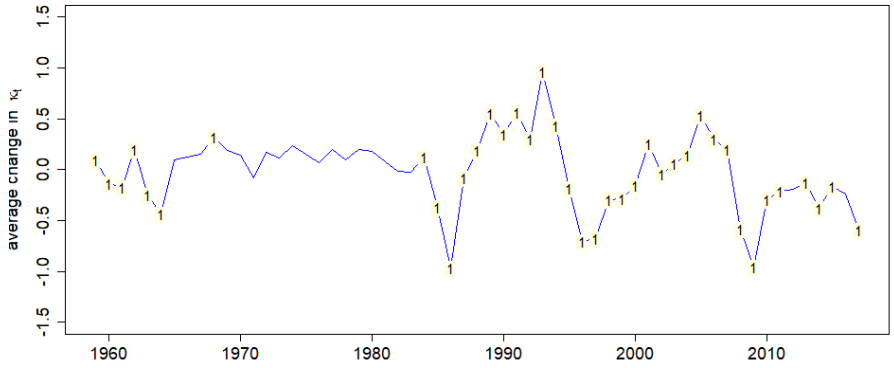


(a) (b)

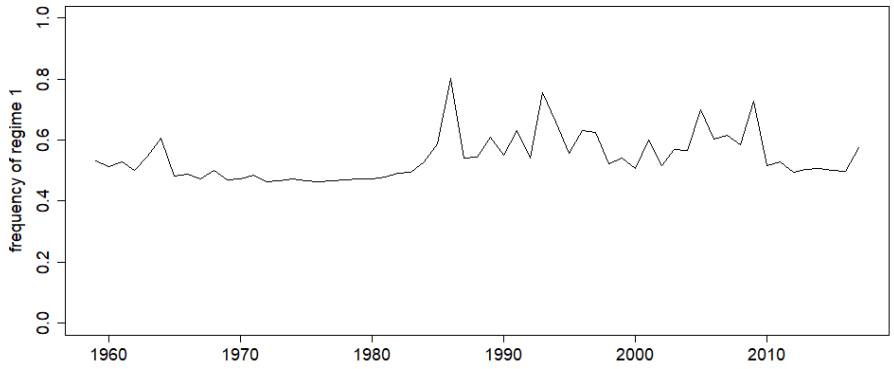
Figure 9. Estimation of the parameter κ_t for Lithuania (a) and Sweden (b). The black line represents the classical Lee-Carter model SVD estimate and the red lines are FFBS samples of 500 MCMC iterations of the SSM with switching.

3.2.4.Regime switching estimates

Figures 10 and 11 provide a comparison of changes in the sampled κ_t versus the average frequencies of regime 1 in MCMC runs of the SSM with switching. The figures show that, as expected, the large fluctuations in κ_t increase the probability of high variance regime 1.

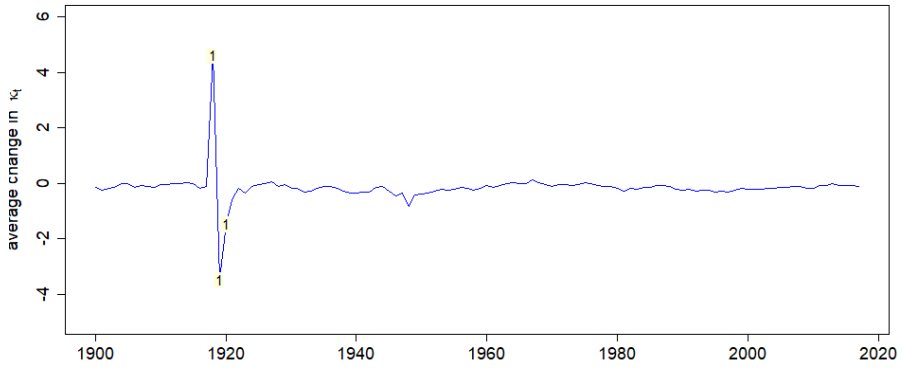


(a)

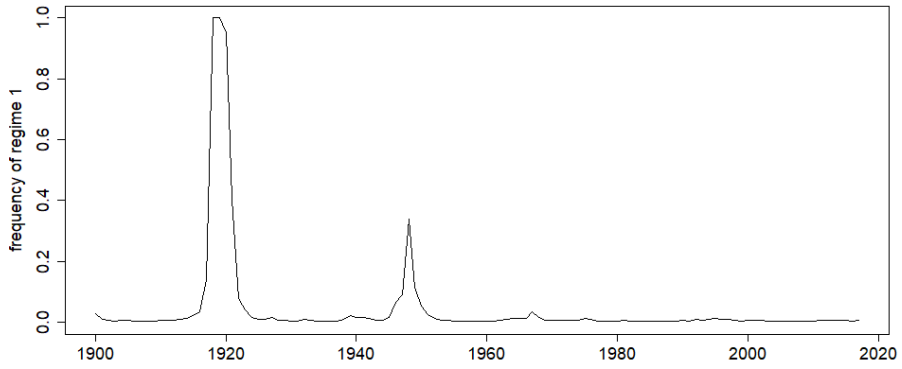


(b)

Figure 10. Th average change in κ_t time varying index over the recorded MCMC runs (a) and the average frequency regime 1 (b) for **Lithuania**. Index "1" denotes the periods where the label of regime 1 is assigned in more than 50% of the recorded MCMC runs.



(a)



(b)

Figure 11. The average change in κ_t time varying index over the recorded MCMC runs (a) and the average frequency regime 1 (b) for **Sweden**. Index "1" denotes the periods where the label of regime 1 is assigned in more than 50% of the recorded MCMC runs.

In the Lithuanian case, the model has managed to label with "1" the key disturbances in mortality, such as a large increase in mortality during the country's transformation to the market economies at the beginning of 1990s, large spike in mortality during the economic crisis of 2008-2009 and the subsequent recovery. However, the average frequency of MCMC runs when a high variance regime label is assigned in those periods and barely exceeds 0.5. The poor ability of the model to identify high variance scenarios even after the identifying restriction $\left(\sigma_Q^{(0)}\right)^2 \leq \left(\sigma_Q^{(1)}\right)^2$ was imposed indicates that this condition is not sufficient to ensure the unique labeling of regimes. In the

MCMC runs there is a lot of label switching which indicates that the constraint is poor, see [26]. Thus, in the Lithuanian case, our specification of SSM with switching does not provide adequate performance.

In the Swedish case, the model in most cases stays in regime 0. Regime 1 is visited just during the years of Spanish flu pandemic of 1918-1920 and with much lower frequency in the mid-1940s, when an improvement in mortality was observed after WWII. Contrary to Lithuania, the regimes are labeled consistently in the MCMC runs and the label switching is almost not observed. This indicates that the identifying restriction $\left(\sigma_Q^{(0)}\right)^2 \leq \left(\sigma_Q^{(1)}\right)^2$ is suitable for this model and that the estimated variances of the state parameter are acceptable.

3.3. Marginal likelihood comparison of the state space models

We compare the models used by performing the assessment of the marginal log-likelihood $m(\boldsymbol{Y}_{1:T})$ as described in Subsection 2.3.6. The results of the calculations are summarized in Table 9.

Table 9. Estimates of the marginal log-likelihood for DLM and SSM with switching (SSM w.s.).

Model	**Lithuania**			**Sweden**		
	$m(\boldsymbol{Y}_{1:T}\|\widehat{\boldsymbol{\psi}})$	$m(\widehat{\boldsymbol{\psi}}) - m(\widehat{\boldsymbol{\psi}}\|\boldsymbol{Y}_{1:T})$	$m(\boldsymbol{Y}_{1:T})$	$m(\boldsymbol{Y}_{1:T}\|\widehat{\boldsymbol{\psi}})$	$m(\widehat{\boldsymbol{\psi}}) - m(\widehat{\boldsymbol{\psi}}\|\boldsymbol{Y}_{1:T})$	$m(\boldsymbol{Y}_{1:T})$
DLM	585	-125	460	846	-137	709
SSM w.s.	585	-126	459	893	-155	738

The results show that SSM with switching has the highest marginal likelihood in the Swedish case. For Lithuania, both DLM and SSM with switching models have similar marginal likelihoods, thus a simpler DLM is preferred.

3.4. Summary of the section

In this section, we have reviewed the data used to fit the models of the previous section, estimated and analyzed the parameters, and performed model fitting diagnostics and the state space model comparison.

For the analysis, we used the data of Lithuania and Sweden. The Swedish data shows a significantly different mortality development from Lithuania. Lithuanian mortality development was volatile, in particular during the transitional period, at the end of which a major change in mortality trend was observed. In Sweden, contrarily, the only major exception from steady mortality improvement were the Spanish flu pandemics.

Model diagnostics did not reveal any major unexpected model deficiencies. As expected, both classical Lee-Carter and Possion Lee-Carter models were not able to handle increased volatility caused by the Spanish flu pandemics in Sweden. A high increase in Lithuanian mortality in the early 1990s and subsequent change in trend also was reflected in the results of model diagnostics. In the case of the state space models, no significant deficiencies were identified during the tests of goodness of fit.

Parameter estimates, overall, are comparable and differences are explainable across the models. Major differences relate to additional parameters introduced and different levels of model flexibility.

Comparison of marginal likelihoods indicated that in the Lithuanian case more simple DLM model is preferable. Meanwhile, for Sweden SSM with regime switching has a higher marginal log-likelihood and is preferred.

4. ESTIMATION OF SOLVENCY CAPITAL

In this section, the results of mortality forecasts are provided for three versions of the Lee-Carter model discussed in Section 2. In the second part of the section, we develop the methodology for the calculation of VaR and provide the results of VaR calculations for Lithuania and Sweden using different insurance benefit formulas, terms to maturity, and ages at policy inception.

4.1. Simulation of mortality rates

By taking exponentials on both sides of Equation (3), we can represent the modeled mortality rates as follows:

$$m_{x,t} = e^{\alpha_x} e^{\kappa_t \beta_x} e^{e_{x,t}},$$

which shows that the variability in projected mortality rates can be considered to arise multiplicatively from three sources: variation in constant, which represents the increase in mortality with age (which we ignore due to its insignificance as a simplification), variation in trend, mainly driven by the parameter κ_t, and the remaining random variation. In the remainder of this sub-section, we first discuss the modeled uncertainty of κ_t, and later we consider how this uncertainty translates to the variation of the overall mortality rate.

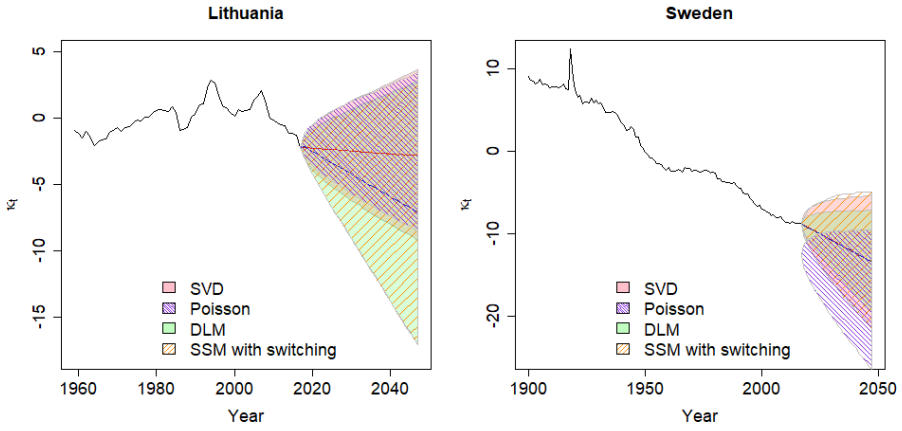
4.1.1. Simulation of time varying index

As can be seen from Equation (10), the uncertainty of the parameter κ_t is driven by variation in the drift parameter μ and the residual error. The results of the simulations of κ_t are provided in Figure 12 for a confidence level of 95%, and in Figure 13 for a confidence level 99.5%.

In the Lithuanian case, due to unsettled historic experience, the confidence intervals are wide for both SVD and state-space models. Allowance for a one-time change in trend results in a more reasonable central forecast; however, it adds additional uncertainty due to the second drift parameter $\mu^{(II)}$. Therefore, although the estimated error variance σ_Q^2 is lower in the SSM case, the overall confidence interval of κ_t is even slightly wider than the one derived by using the classical Lee-Carter model.

The SSM with switching does not converge to two clearly expressed low and high variance regimes for Lithuania. Therefore, the distribution of error terms modeled as a mixture can be well approximated with a Normal

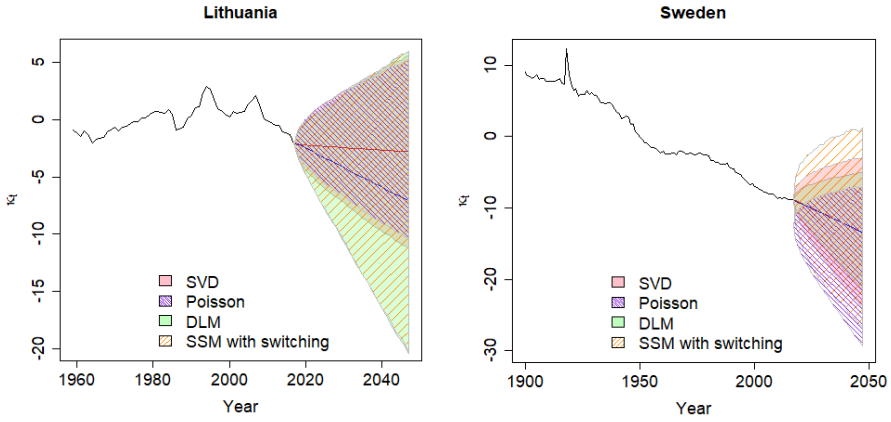
distribution. For this reason, the simulated confidence intervals for both models are very similar.



(a) (b)

Figure 12. Results of estimation of the Lee-Carter model parameter κ_t for Lithuania (a) and Sweden (b). The shaded region provides 95% simulated confidence intervals of the κ_t parameter fitted using SVD, Poisson, DLM, and SSM with switching models. In the Lithuanian case, DLM and SSM with switching allow for a one-time change in trend in 1995; therefore, the direction of the trend differs from the SVD estimate. The red line is used to denote the Lithuanian central forecast estimated with the SVD and the blue line is used to denote the SSM central forecast.

In the Swedish case, DLM gives a narrower estimate of confidence interval with respect to SVD generally due to the better fit and the lower error variance. The Poisson model confidence interval is shifter down due to lower estimate of κ_t in the latest years using Poisson regression in comparison with SVD, DLM, and SSM with switching methods, see Figure 8. For Sweden SSM with switching converges into two different regimes: low probability and high variance pandemic regime and high probability low variance usual development regime. Thus, the distribution of the error terms when modeled as a mixture has much heavier tails than when it is approximated with a single Normal distribution. This results in significantly wider confidence intervals for SSM with switching, especially at a 99.5% confidence level. We note that the width on the difference in confidence intervals is driven by model differences, not the poor fit. As shown in model comparison Subsection 3.3, despite having the widest confidence intervals SSM with switching has a higher log-likelihood than DLM and is a preferred model for Sweden.

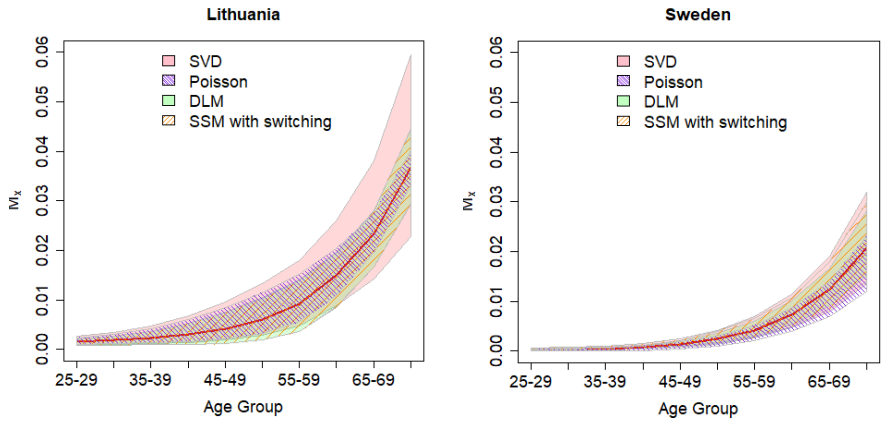


(a) (b)

Figure 13. Results of estimation of the Lee-Carter model parameter κ_t for Lithuania (a) and Sweden (b). The shaded region provides 99.5% simulated confidence intervals of the κ_t parameter fitted using SVD, Poisson, DLM, and SSM with switching models. In the Lithuanian case, DLM and SSM with switching allow for a one-time change in trend in 1995; therefore, the direction of the trend differs from the SVD estimate. The red line is used to denote the Lithuanian central forecast estimated with the SVD and the blue line is used to denote the SSM central forecast.

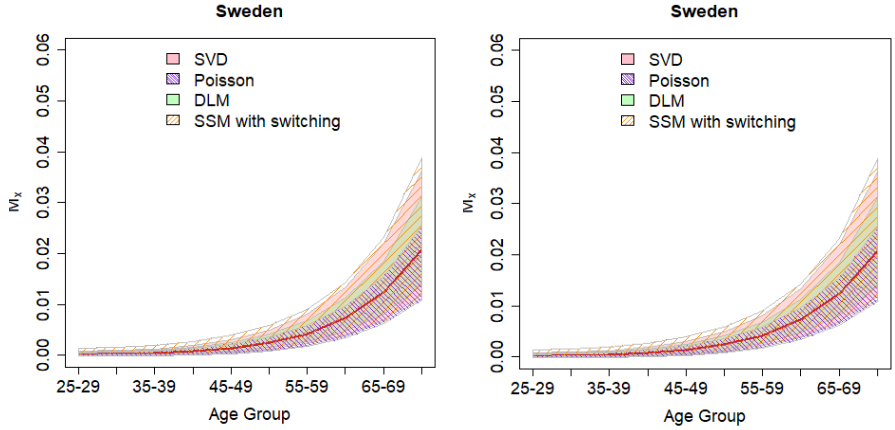
4.1.2. Derivation of simulated mortality rates

The results of the simulations of $m_{x,t}$ mortality rates are provided in Figure 14 for a confidence level of 95%, and Figure 15 for confidence level 99.5%. Considering that in practical applications we usually need a set of mortality rates for each of the future years in the forecasting period, in our illustration we provide the results for the 15 year projection, which is in the middle of the total forecasting period.



(a) (b)

Figure 14. Results of simulation of the 15 year projection of mortality rates for ages from 25 to 74 for Lithuania (a) and Sweden (b). The shaded region provides 95% simulated confidence intervals of the projected mortality rates fitted using SVD, Poisson, DLM, and SSM with switching models. The red line is used to denote the central SVD forecast.



(a) (b)

Figure 15. Results of simulation of the 15 year projection of mortality rates for ages from 25 to 74 for Lithuania (a) and Sweden (b). The shaded region provides 99.5% simulated confidence intervals of the projected mortality rates fitted using SVD, Poisson, DLM, and SSM with switching models. The red line is used to denote the central SVD forecast.

As we can see from Equation (3), the variation in κ_t is translated to changes in the mortality rates via the parameter β_x. The higher β_x for a certain age

group is, the more sensitive are the projected mortality rates of that age group to changes in κ_t, and vice versa. As shown in Figure 7, for Lithuania the highest β_x are for mid-ages, and therefore in the Poisson and the state space models mid-agers get the widest confidence intervals. The Lithuanian SVD calculations, due to relatively poor fit, have a large residual variance, which dominates in the classical Lee-Carter model simulations and results in very wide simulated mortality rate confidence intervals. Thus, although we did not manage to achieve with Poisson and SSMs a significant improvement in the confidence intervals of the parameter κ_t, the width of the simulated confidence intervals of mortality rates are reasonable and comparable to the Swedish result.

In the Swedish case, the results are consistent with the results of the simulation of the parameter κ_t. Due to better fit, at 95% confidence level the confidence intervals of Poisson model and SSMs are narrower than in case of SVD. At 99.5% confidence level, as expected, the heavy tails of SSMs with switching prevail, resulting in the widest confidence intervals.

4.2. Calculation of VaR

In this subsection, we show how to translate the simulated mortality rates to Value-at-Risk (VaR). According to [58] given a certain confidence level $\alpha \in (0, 1)$ VaR can be defined as:

$$\mathrm{VaR}_\alpha(NP) = \inf\{l \in \mathbb{R}: \mathbb{P}(NP > l) \leq \alpha\}, \tag{22}$$

where *NP* is a random variable of a net profit. Throughout the dissertation, we applied the convention that *L* is negative when the company makes a loss, and positive when the company is profitable. Consistently with the Solvency II legislation, we calculated VaR using 99.5% confidence level, that is, set $\alpha = 0.995$.

In the context of mortality risk in the Solvency II framework, *NP* is a loss/gain of the Solvency II capital (Basic Own Funds) due to variation of mortality rates. Thus, $\mathrm{VaR}_{0.995}(NP)$ defines extra capital needed to cover the unreserved mortality losses arising from the increase in the future mortality rates.

As noted in the Introduction, we consider two alternative approaches to VaR calculation: run-off VaR and one-year VaR. Taking the run-off approach, we consider the fluctuations in mortality rates until the maturity of a policy. In contrast, if we take a one-year view, the key drivers of risk become fluctuation of mortality rates in the first projection year and the risk of changes

in reserving (Best Estimate) assumptions at the end of the first projection year. A graphical illustration of the contrast between run-off and one-year simulations of mortality rates is provided in Figure 16.

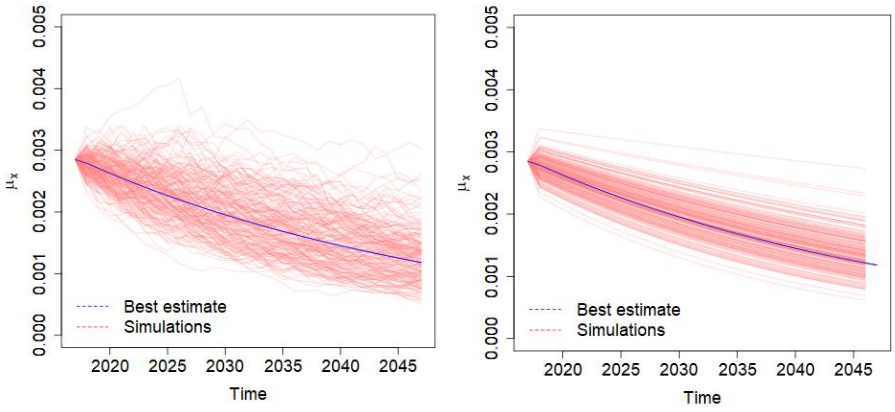


Figure 16. Example of simulations of central mortality rates (for a fixed age) used for the purpose of calculation of run-off VaR (see (a)) and one-year VaR (see (b)). The calculations are based on the Lithuanian data for a life aged 40 years. In calculations of one-year VaR reserve the credibillity factor $\delta = 10\%$ was used.

As illustrated in the charts, in the calculation of run-off VaR, we apply simulated mortality rates that vary stochastically for the whole period of insurance coverage. In contrast, in the calculation of one-year VaR, we use stochastically simulated mortality rates for modeling of death benefits in the first projection year, and for the remaining years, we apply the curves of projected Best Estimate mortality rates, the level and trend of which depend on the outcome of the first year simulations. For example, if a certain simulation scenario stochastically results in a sharp increase of mortality rates in the first projection year, we assume that a curve of the projected Best Estimate mortality rates for that simulation scenario is shifted upwards as well. In addition, the results of the first year simulations affect the trend of the Best Estimate mortality rates and the projected curves are not parallel. The bigger is the variation in the level and trend, the higher risk is associated with the changes in the Best Estimate technical provisions at the end of the first year, which results in the higher one-year VaR.

4.2.1. Run-off VaR methodology

We used the following formula for the calculation of run-off VaR:

$$\mathrm{VaR}_{0.995}\left(\sum_{i=1}^{K} CF_{T+i}\right) = \inf\left\{l \in \mathbb{R}: \mathbb{P}\left(BE_T - \sum_{i=1}^{K} CF_{T+i} > l\right) \leq 0.995\right\}. \quad (23)$$

The formula implies that run-off VaR is simply the quantile of the distribution of the deviation of the sum of total estimated losses in year *t*, denoted by CF_t, till maturity in *K* years minus the Best Estimate at the end of year *T* which is calculated as the expected value of future mortality cash flows derived using the central projection of the mortality probabilities

$$BE_T = \mathrm{E}\left(\sum_{i=1}^{K} CF_{T+i}\right).$$

In this dissertation, we assess VaR for two different insurance benefit formulas: level benefits and decreasing benefits. Decreasing benefits are common in life assurance products linked to mortgages or other credit instruments. In addition, mortality sum at risk is decreasing with policy duration for some risk and savings products (e.g., traditional endowment insurance), where the total benefit payable on death is fixed and the insurer is able to recoup part of the losses by reversing the accumulated savings amount. However, a significant part of life assurance products has fixed sums assured. In this paper we examine the effect on VaR of both benefit formulas: level (fixed) sum assured and sum assured which is decreasing linearly with time.

Level benefits

Since we assume that there is no discounting, lapses, future premiums and the death benefit does not change during the policy term, we can ignore the timing of death benefits during the period covered by an insurance policy. Therefore, for a policy with a sum assured of 1 monetary unit we can calculate the sum of expected cash outflows (death benefits) as the estimated mortality probability over the remaining policy term *K*:

$$\sum_{i=1}^{K} CF_{T+i} = {}_K q_x(T),$$

where ${}_K q_x(T)$ is a probability that a life aged *x*, which is alive at the end of year *T*, will die during the *K* year period. ${}_K q_x(T)$ can be calculated using the following formulas:

$$ {}_Kq_x(T) = 1 - {}_Kp_x(T), $$

$$ {}_Kp_x(T) = \prod_{s=0}^{K-1} p_{x+s}(T+s) = \prod_{s=0}^{K-1} \big(1 - q_{x+s}(T+s)\big). $$

where ${}_Kp_x(T)$ is a probability that a life aged *x*, which is alive at the end of year *T*, will survive the *K* year period, $p_x(T)$ is one year survival probability of a life aged *x*, which is alive at the end of year *T*, and $q_x(T)$ is one year death probability of a life aged *x*, which is alive at the end of year *T*.

Inserting this expression into Formula (23) and denoting the Best Estimate of term assurance policy with level benefits as $BE_T^{(L)}$, we obtain

$$ \mathrm{VaR}_{0.995}^{(L)}\left(\sum_{i=1}^{K} CF_{T+i}\right) = \inf\Big\{l \in \mathbb{R}\colon \mathbb{P}\big(BE_T^{(L)} - {}_Kq_x(T) > l\big) \le 0.995\Big\}. $$

Decreasing benefits

Consider *n* year term life assurance policy with initial sum assured of 1 which, at the end of each policy year, decreases linearly by $\frac{1}{K}$. Thus, the insurance company is expected to pay 1 if the policyholder has not survived the first policy year, $\frac{K-1}{K}$ if the policyholders died in the second year, and so on. In such a case, we have

$$ \sum_{i=1}^{K} CF_{T+i} = \sum_{j=0}^{K-1} \frac{K-j}{K} q_{x+j}(T+j)\,{}_jp_x(T). \tag{24} $$

Proposition 5. The sum of cash flows resulting from a term assurance policy with linearly decreasing benefits can be calculated using the following equation:

$$ \sum_{i=1}^{K} CF_{T+i} = \frac{1}{K} \sum_{j=0}^{K-1} {}_{j+1}q_x(T). \tag{25} $$

Proof

The Result (25) is obtained from Formula (24) by changing the order of summation and noting that

$$q_x(T) + q_{x+1}(T+1)p_x(T) + q_{x+2}(T+2)_2p_x(T) + \cdots + q_{x+j}(T+j)_jp_x(T) = {}_{j+1}q_x(T).$$

Inserting Expression (25) into Formula (23), and denoting the Best Estimate of a term assurance policy with decreasing benefits by $BE_T^{(D)}$, we get that the run-off VaR for a term assurance policy with decreasing benefits can be calculated as follows:

$$\mathrm{VaR}_{0.995}^{(D)}\left(\sum_{i=1}^{K} CF_{T+i}\right) = \inf\left\{l \in \mathbb{R}: \mathbb{P}\left(BE_T^{(D)} - \frac{1}{K}\sum_{j=0}^{K-1} {}_{j+1}q_x(T) > l\right) \leq 0.995\right\}.$$

4.2.2. One-year VaR methodology

In deriving one-year VaR we use the basic Solvency II principle, which requires the calibration of Solvency Capital Requirement (SCR) to VaR of the (loss of) Basic Own Funds subject to a confidence level of 99.5% over a one-year period. Basic Own Funds at the end of year *t*, BOF_T, is defined as the difference between insurer's assets A_T and liabilities L_T, valued according to the Solvency II requirements, plus subordinated liabilities, which we will ignore in our analysis as a simplification. Liabilities L_T may be disaggregated into technical provisions, consisting of Best Estimate BE_T and Risk Margin RM_T, and other liabilities OL_T. Therefore we suppose that

$$BOF_T = A_T - BE_T - RM_T - OL_T.$$

In practice, there are many factors which could contribute to the change in the Basic Own Funds over the year. Considering that our focus is on the mortality risk, the analysis was performed for a simplified portfolio of a single premium (payable in advance) fixed term life assurance policies. We assumed that there are no lapses, no expenses, no new business, no changes in other liabilities, and that claims are paid immediately after they occur. We also assumed that there are no cash flows to or from a shareholder, such as dividends, capital injections, and that the investment return on assets and the discount rate used for calculation of technical provisions both are zero. Depending on the size of a discount rate, discounting would have a similar effect as a reduction of benefits with time, that is, based on the results presented later, would generally reduce VaR. Finally, we assumed that the

Risk Margin is also fixed. In practice, changes in the Best Estimate and the Risk Margin are likely to be positively dependent and the Risk Margin is likely to affect VaR. However, we have accepted this simplification, considering that due to its relative size, the Risk Margin is likely to have a significantly smaller effect on VaR than the Best Estimate.

Under the above assumptions a change in assets during year $T+1$ is caused only by a cash outflow CF_{T+1} due to payment of death claims. Therefore, the random variable of a change in Basic Own Funds during the year $T+1$ can be expressed as follows:

$$\Delta BOF_{T+1} = A_{T+1} - A_T - (BE_{T+1} - BE_T) = BE_T - CF_{T+1} - BE_{T+1}. \qquad (26)$$

Under our model, the expectation of Best Estimate for valuation date at the end of the year $T+1$, assessed using the information $\mathcal{F}_T$ available at the end of year T, can be calculated as

$$\mathrm{E}(BE_{T+1}|\mathcal{F}_T) = \sum_{i=2}^{K} \mathrm{E}(CF_{T+i}|\mathcal{F}_T).$$

Inserting this equality into Equation (26), and assuming that we assess ΔBOF_{T+1} using information available at the end of year T we obtain the following expression

$$\Delta BOF_{T+1} = BE_T - \big(CF_{T+1} + E(BE_{T+1}|\mathcal{F}_T)\big),$$

i.e. random variable of the change in the Basic Own Funds during the next valuation year is equal to the difference between the initial Best Estimate and the sum random variables of the claims cash flow during the year $T+1$ and the Best Estimate at the end of year $T+1$.

Substituting the expression of the change in the Basic Own Funds into Formula (22), we get the VaR formula for the loss of the Basic Own Funds over one year (one-year VaR), with the confidence level of 99.5%, as assessed using the information available at time *t*:

$$\begin{aligned}\mathrm{VaR}_{0.995}(\Delta BOF_{T+1}) &\\ = \inf\big\{l \in \mathbb{R}\colon \mathbb{P}\big\{BE_T - \big(CF_{T+1} + E(BE_{T+1}|\mathcal{F}_T)\big) > l\big\} &\\ \leq 0.995\big\}.&\end{aligned} \qquad (27)$$

Thus, SCR is the capital required to cover the excess of the total of claims payments during the first projection year and the Best Estimate at the end of the first projection year over the initial Best Estimate with 99.5% probability. We note that random variables CF_{T+1} and $E(BE_{T+1}|\mathcal{F}_T)$ are not independent.

Actuaries usually reconsider the assumptions used to calculate Best Estimate based on the recent actual mortality experience. For example, significantly higher than expected incurred mortality losses are likely to lead to upwards revision of the Best Estimate mortality assumptions. Therefore, we can presume positive dependence between CF_{T+1} and $E(BE_{T+1}|\mathcal{F}_T)$ and our modeling challenge is to estimate the conditional expectation

$$\mathrm{E}\left(\sum_{i=2}^{K} CF_{T+i}|CF_{T+1}\right).$$

There are several ideas on how to model this relationship. For example, Richards et al. [69] for each simulation run added the simulated mortality experience in year $T+1$ to the already available historic data and used the total data set to refit the stochastic mortality model, which enabled to quantify variation in the mortality trend parameter. Similarly, Börger et al. [8] and Plat [64] used simulated mortality experience for each stochastic run to re-estimate the mortality trend parameter at $T+1$. We applied a similar approach, and for each run used simulated time varying index κ_{T+1} to adjust the RWD drift parameter:

$$\mu_{T+1} = \mu + \delta(\kappa_{T+1} - \kappa_T - \mu),$$

where δ is chosen credibility factor and μ, κ_T are parameters estimated during the initial Lee-Carter model fitting.

Our approach requires setting explicit credibility factor δ, which is avoided in the methods mentioned above. However, the above methods also rely on certain additional model parameters which control the level of trend risk produced by the simulation. Börger et al. [8], for the purpose of introducing variability in the slope of the linear trend, reestimates with weighted least squares the trend using a new stochastic realization of mortality trend parameter. In this model, trend sensitivity is dependent on the choice of the length of the fitting period and the weights used. Plat [64] assumed a specific functional form of the trend parameter and the selected fitting period has a significant effect on its volatility. Similarly, using the approach of Richards et al. [69], if applied together with the Lee-Carter model, trend sensitivity would depend on the selected length of the fitting period. For the purpose of our analysis, considering that fitting periods for Lithuania and Sweden differ, we found it more convenient to set an explicit parameter controlling the level of trend risk.

Let us consider what kind of practical reserving behavior would be consistent with the one-year VaR model described above. Firstly, the model

implies that actuaries in the reserving assumptions allow for future mortality improvements (mortality reduction factors), which are projected with the Lee-Carter model. In practice, actuaries often reserve life assurance products without allowing for future mortality improvements, that is, some prudence is left in the Best Estimate, although strictly speaking this is not in line with the legislation. The overstatement of the Best Estimate means that the risk of its insufficiency due to fluctuation in mortality rates and consequently VaR is lower. Thus, if the Best Estimate for life assurance products is calculated without allowance for future mortality improvements, the company specific one-year VaR would be lower than the one-year VaR assessed by us.

Secondly, consistently with the RWD model applied for modeling of the Lee-Carter model time varying index, we assume that the variation of the latest year mortality rates is fully translated to shifts in projected Best Estimate mortality curves. This implies that actuaries base the Best Estimate mortality assumptions on the mortality level implied by the last year's experience. In practice, actuaries often use various smoothing, averaging and similar methods, as a result of which the variation driven by the mortality level of the last observed year may be taken into account only partially. The application of such techniques may result in a lower risk of changes in Best Estimate and lower VaR.

Finally, even if the reserving methodology allows for mortality improvements, the sensitivity of the mortality reduction factors (trend risk) to the last year's mortality experience may vary from insurer to insurer. In our calculations, we use two levels of credibility factor δ (5% and 10%), which represents variation in reserve risk due to differences in actuarial methodologies applied. The credibility parameter may be interpreted as the proportion of evidence related to last year's experience accounted for in setting the assumed mortality reduction factors. Thus, in the case of $\delta = 10\%$ mortality reduction factors are more sensitive to the last year's experience than in the case of $\delta = 5\%$, and the modeled trend risk increases with the growth of δ.

Overall, the above considerations imply that our approach results in a rather conservative estimate of one-year VaR for two levels of selected trend risk parameters.

Below, the detailed VaR formulas are derived for the two benefit formulas considered in this dissertation.

Level benefits

In the case of one-year VaR, in addition to the first year's death benefits, we calculate the projected benefits starting from year $T+2$, multiplied by the probability of survival in the first year. Inserting to Formula (27), one-year VaR for term life assurance with level benefits can be calculated as follows:

$$\mathrm{VaR}_{0.995}^{(L)}(\Delta BOF_{T+1}) = \\ = \inf\left\{l \in \mathbb{R}: \mathbb{P}\left\{BE_T^{(L)} - q_x(T) - \widetilde{BE}_{T+1}^{(L)} > l\right\} \leq 0.995\right\},$$

where:

$$\widetilde{BE}_{T+1}^{(L)} = p_x(T)\mathrm{E}({}_{K-1}q_{x+1}(T+1)|\mathcal{F}_T).$$

Decreasing benefits

For the decreasing benefits in order to derive the simulated mortality probabilities, we use Result (25) to calculate the expected benefits starting from year $T+1$

$$p_x(T)\sum_{i=1}^{K} CF_{T+i} = p_x(T)\frac{1}{K}\sum_{j=1}^{K-1} {}_jq_{x+1}(T+1). \tag{28}$$

Substituting Result (28) into Formula (27) for the one-year VaR for term life assurance with decreasing benefits, we get that

$$\mathrm{VaR}_{0.995}^{(D)}(\Delta BOF_{T+1}) = \\ = \inf\left\{l \in \mathbb{R}: \mathbb{P}\left\{BE_T^{(D)} - q_x(T) - \widetilde{BE}_{T+1}^{(D)} > l\right\} \leq 0.995\right\},$$

where:

$$\widetilde{BE}_{T+1}^{(D)} = p_x(T)\frac{1}{K}E\left(\sum_{j=1}^{K-1} {}_jq_{x+1}(T+1)\,|\mathcal{F}_T\right).$$

4.2.3. Results of VaR calculations

In this section we provide results VaR assessments, derived from the simulated mortality rates, described in Subsection 3.1. For comparability of our results with the Standard Formula mortality stress parameter, we have converted the calculated VaRs to equivalent VaR rates. We define VaR rate as g which satisfies the following condition:

$$BE_T + \mathrm{VaR}_{0.995}(L) = \sum_{i=1}^{K} CF_{T+i}^{(g)},$$

where $CF_{T+i}^{(g)}$ are projected mortality benefits calculated according to the applicable benefit formula by applying the following k year stressed mortality probabilities

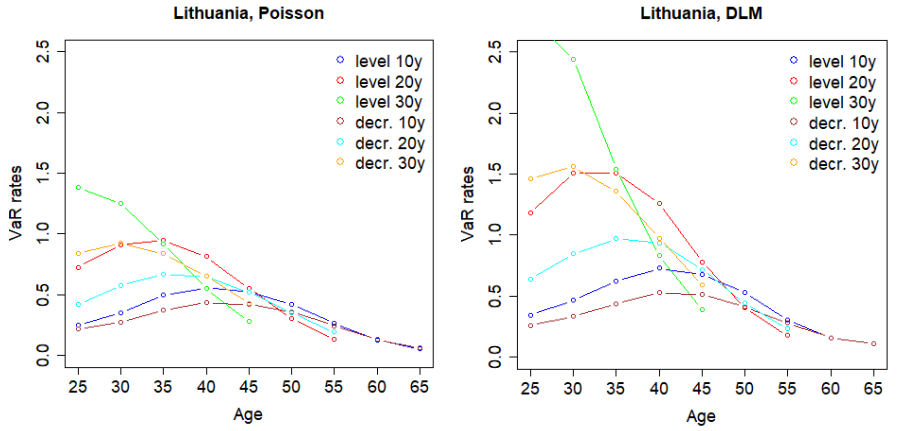
$${}_k q_x^{(g)}(t) = 1 - \prod_{s=0}^{k-1} \big(1 - (1+g)\hat{q}_{x+s}(t+s)\big).$$

In the formula above, $\hat{q}_x(t)$ denotes the Best Estimates of one year mortality probabilities and $k \in \{1,2,\dots,K\}$. In such a way, we search for rate g, which makes the stressed Best Estimate equal to the initial Best Estimate plus VaR.

We calculate VaRs using two models for each of the two countries. We note that the results of calculations depend on the analyzed country's mortality experience. Overall, Lithuania and Sweden provide a good proxy of different mortality development in the EU: very stable development typical to Western European countries, see [74], and volatile development with the change in trend applicable to Central and Eastern European countries.

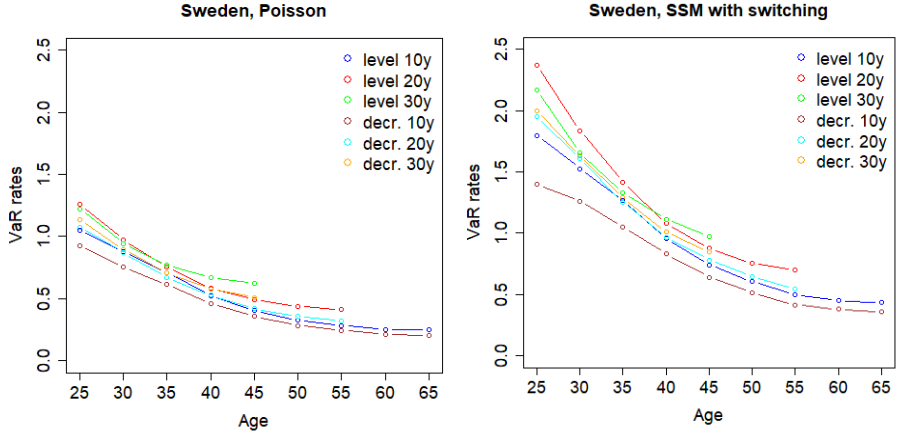
We do not use the SVD model due to high residual variance. We use the Poisson model for both of the countries, but note that the goodness of fit tests identified outliers in the Swedish case. From the SSMs for each country, we choose a model that fits better: DLM in the Lithuanian case and SSM with switching in the Swedish case.

The results of the calculation of run-off VaR rates are presented in Figures 17 and 18. The results indicate that run-off VaR rates tend to increase with the policy term. Uncertainty about the future mortality rates grows with time, which leads to wider confidence intervals of mortality rates for long term projections. Similarly, for the policies with decreasing sum assured, higher sums assured are paid in early policy years when the mortality uncertainty is lower than in later policy years, which implies that run-off VaR rates are generally lower for policies with decreasing sum assured than for equivalent policies with a fixed sum assured.



(a) (b)

Figure 17. Calculated run-off VaR rates at policy inception for **Lithuania** for different ages at policy inception, three different terms to maturity (10 years, 20 years, and 30 years), and two benefit formulas (level sum assured and sum assured decreasing with time) where mortality rates are simulated using Poisson (a) and DLM (b) models.



(a) (b)

Figure 18. Calculated run-off VaR rates at policy inception for **Sweden** for different ages at policy inception, three different terms to maturity (10 years, 20 years, and 30 years), and two benefit formulas (level sum assured and sum assured decreasing with time), where mortality rates are simulated using Poisson (a) and DLM with switching (b) models.

The shape of the VaR rate by age curves is related to the shape of the Lee-Carter parameter β_x which determines the sensitivity of projected age specific mortality rates to the changes in the time varying parameter and drives the relative level of age-specific projected volatility of mortality rates. As demonstrated in Figure 7, Sweden has a declining parameter β_x over the modeled ages. Consequently, for Sweden run-off VaR rate curves also declined with age. For Lithuania, higher volatility of mortality rates at mid ages increases the run-off VaR rates for policies issued to younger and mid-age policyholders.

Comparing the results with the standard Solvency II Standard Formula stress level of 15%, the calculated VaR rates are significantly higher. Figure 19 provides VaR rates calculated using the Poisson model fitted with Swedish data for the years 1960-2017. As it can be seen, the calculated VaR rates are comparable with the Standard Formula stress level. This is the expected result considering the mortality stress level calibration method applied by EIOPA, see Section 1. Thus, the differences in VaR rates can be explained by the fact that in our analysis we allow for the effect of disturbances in mortality development as well as take into account parameter uncertainty, which is not the case in the analysis performed by EIOPA.

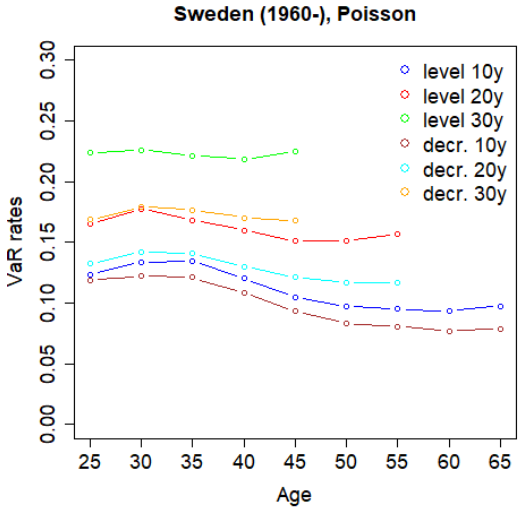
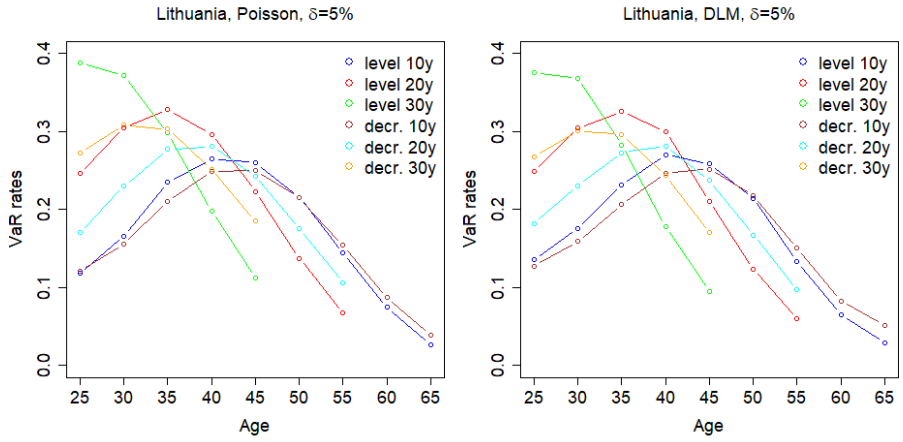


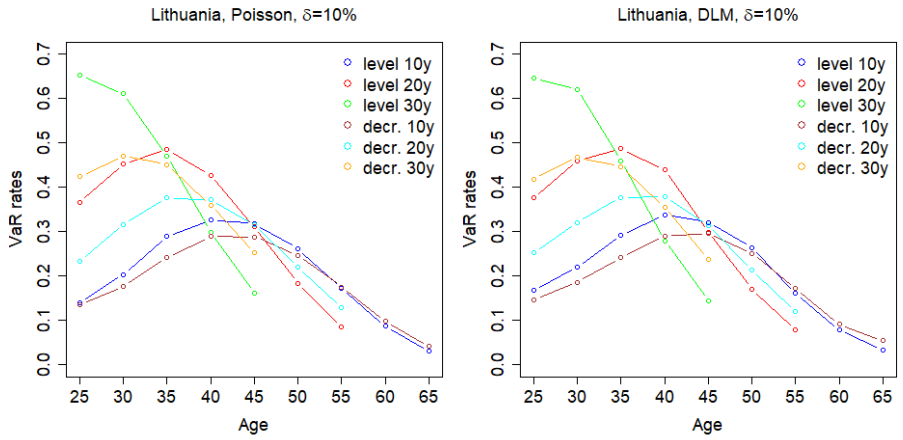
Figure 19. Calculated run-off VaR rates at policy inception for **Sweden** for different ages at policy inception, three different terms to maturity (10 years, 20 years, and 30 years), and two benefit formulas (level sum assured and sum assured decreasing with time), where mortality rates are simulated using Poisson models fitted using trimmed data set (1960-2017).

The results of calculation of one-year VaR for two levels of the credibility parameter δ are presented in Figures 20-23.



(a) (b)

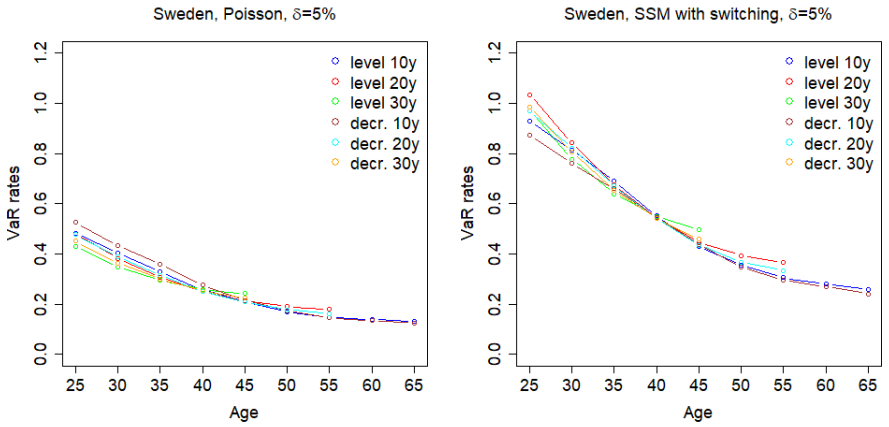
Figure 20. Calculated one-year VaR rates **($\delta = 5\%$)** at policy inception for **Lithuania** for different ages at policy inception, three different terms to maturity (10 years, 20 years, and 30 years), and two benefit formulas (level sum assured and sum assured decreasing with time), where mortality rates are simulated using Poisson (a) and DLM (b) models.



(a) (b)

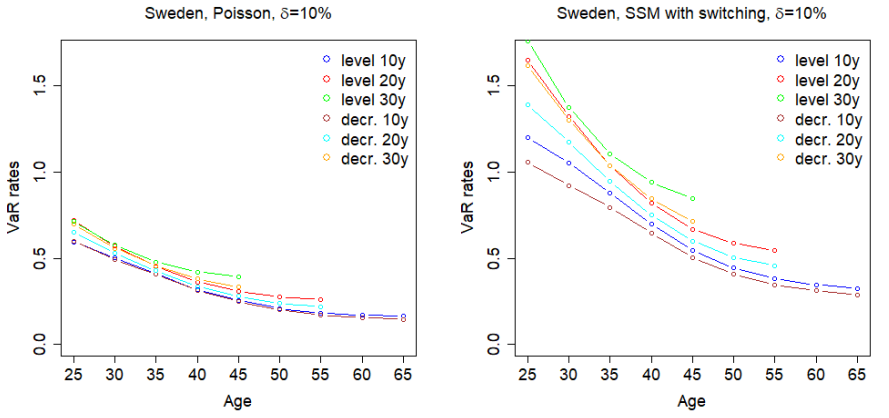
Figure 21. Calculated one-year VaR rates **($\delta = 10\%$)** at policy inception for **Lithuania** for different ages at policy inception, three different terms to maturity (10 years, 20 years, and 30 years), and two benefit formulas (level sum assured and sum assured decreasing with time), where mortality rates are simulated using Poisson (a) and DLM (b) models.

The shape of the one-year VaR rate by age curves is similar to the shapes of the run-off VaR rate curves of the corresponding countries. The overall level of the curves to a large extend depends on the value of the credibility parameter δ, especially for longer term policies. Larger parameter δ implies larger best estimate sensitivity at the end of the year and higher risk.



(a) (b)

Figure 22. Calculated one-year VaR rates **($\delta = 5\%$)** at policy inception for **Sweden** for different ages at policy inception, three different terms to maturity (10 years, 20 years, and 30 years), and two benefit formulas (level sum assured and sum assured decreasing with time), where mortality rates are simulated using Poisson (a) and DLM (b) models.



(a) (b)

Figure 23. Calculated one-year VaR rates **($\delta = 10\%$)** at policy inception for **Sweden** for different ages at policy inception, three different terms to maturity (10 years, 20 years, and 30 years), and two benefit formulas (level sum assured and sum assured decreasing with time), where mortality rates are simulated using Poisson (a) and DLM (b) models.

Comparing the impact on VaR results to the model used, we observe that for Lithuania, VaRs derived using the Poisson model and DLM are similar. For Sweden, VaRs derived with the Poisson model are significantly lower than VaRs derived using SSM with switching. The key reason for the difference is the ability to capture heavier tails of the mixture distrubution with SSM with switching model, therefore, we consider SSM with switching to be a preferred model for the calculation of VaR for Sweden.

4.3. Summary of the section

In this section, we have compared confidence intervals of mortality projections derived using different stochastic mortality models and used the selected projections as an input to VaR calculations.

Application of various models results in different widths of confidence intervals of forecasted mortality rates. The classical Lee-Carter model projections have wide confidence intervals due to relatively poor fit, especially in Lithuania. For the Swedish SSM with regime switching model, confidence intervals are wide as expected due to heavy tails of the mixture distribution. We also observe a much more reasonable modeled mortality trend in Lithuania due to allowance for change in drift in the 1990s.

For VaR calculations, two models were developed: run-off VaR and one-year VaR. Run-off VaR model definition comes from the basic principles of cash flow projections. One-year VaR is more complex and requires the projection of an insurer's solvency capital over a one-year period. We proposed and developed a model which uses a credibility factor reflecting to what extend actuaries take into account the latest mortality experience in setting the mortality assumptions for calculation of Solvency II technical provisions. Both VaR models were developed for two term life assurance benefit formulas: level benefits and decreasing benefits.

We performed VaR calculations for both VaR calculation models and both benefit formulas using several different terms to maturity and different ages at policy inception. The results of the calculation show varying levels of VaR depending on the model, country, and insurance cover details. In most of the cases, VaR was above the regulatory solvency stress level.

CONCLUSIONS

In this section, we provide an overview of the results and conclusions derived from this dissertation.

- We developed two new state space Lee-Carter model modifications which provided us with the higher flexibility to deal with changes in the mortality trend and temporary increase in volatility. For model fitting, we developed algorithms based on the Gibbs sampler, and for model comparison we developed algorithms based on the auxiliary particle filter.
- We fitted the new state space models and classical Lee-Carter and Poisson Lee-Carter models using the Lithuanian and Swedish data. State space models enabled us to achieve a better fit and to derive more reasonable projected mortality rates in comparison with the classical Lee-Carter and Poisson Lee-Carter models. Under model diagnostics, we determined that in the Swedish case, the pandemic effect was well captured by SSM with switching, and in the Lithuanian case a simpler DLM was the preferred model.
- When performing an analysis of confidence intervals of projected mortality rates, we noted significant differences between the models. In particular, due to its relatively poor fit, the classical Lee-Carter model resulted in a very wide confidence intervals for Lithuania, but the DLM enabled us to achieve much more reasonable results. In the Swedish case, the SSM with the switching method allowed us to model the effect of pandemics, which was not possible to achieve with the classical Lee-Carter model.
- We developed a detailed calculation method of VaR for mortality risk using the following two approaches: run-off VaR and one-year VaR. Our one-year VaR model included explicit allowance for reserve sensitivity, driven by internal insurer's reserving practices. Our methodologies suit two different insurance benefit formulas: level benefits and decreasing benefits.
- We performed VaR calculations using both approaches and showed, that VaR levels can be significantly affected by policy term to maturity, benefit formula, and policyholder age. In case of one-year VaR, the reserve sensitivity factor has a significant impact as well. Based on the results of this dissertation, insurance companies can

perform a more realistic assessment of mortality risk VaR by taking into account the specific exposures in their insurance portfolios and use it for solvency calculations and risk management.

- When comparing the estimated Lithuanian and Swedish VaR rates to the Solvency II Standard Formula mortality stress, on average, our derived VaR rates are higher. The key reasons for the difference are the inclusion of higher volatility periods during the model fitting and allowance for parameter uncertainty.

REFERENCES

[1] J. M. Alho, "The Lee-Carter Method for Forecasting Mortality, with Various Extensions and Applications", Ronald Lee, January 2000, *North American Actuarial Journal* **4** (2000), no. 1, 91-93.

[2] T.W. Anderson, *An Introduction to Multivariate Statistical Analysis, Third Edition*, New Jersey: Willey&Sons, 2003.

[3] B. D. O. Anderson and J. B. Moore, *Optimal Filtering*, Englewood Cliffs: Prentice-Hall, 1979.

[4] S. M. Arulampalam, S. Maskell, N. Gordon, and T. Clapp, A Tutorial on Particle Filters for Online Nonlinear/ Non-Gaussian Bayesian Tracking, *IEEE Transactions on Signal Processing* **50** (2002), no. 2, 174-207.

[5] H. Booth, J. Maindonald, and L. Smith, Applying Lee-Carter under conditions of variable mortality decline, *Population Studies* **56** (2002), no. 3, 325-336.

[6] H. Booth, J. Maindonald, and L. Smith, Age-time interactions in mortality projection: applying Lee-Carter to Australia, *Working paper* (2002), https://openresearch-repository.anu.edu.au/handle/1885/41457.

[7] H. Booth and L. Tickle, Mortality Modelling and Forecasting: a Review of Methods, *Annals of Actuarial Science* **3** (2008), no. 1-2, 3-43.

[8] M. Börger, D. Fleischer, and N. Kuksin, Modeling the mortality trend under modern solvency regimes, *ASTIN Bulletin* 44 (2014), 1-38.

[9] E. Brainerd and D. M. Cutler, Autopsy on the empire: Understanding mortality in Russia and the former Soviet Union, *The Journal of Economic Perspectives* **19** (2005), 107–130.

[10] N. Brouhns, M. Denuit, and J. K. Vermunt, A Poisson log-bilinear regression approach to the construction of projected lifetables, *Insurance: Mathematics and Economics* **31** (2002), no. 3, 373-393.

[11] N. Brouhns, M. Denuit, and I. Van Keilegom, Bootstrapping the Poisson log-bilinear model for mortality forecasting, *Scandinavian Actuarial Journal* (2005), no. 3, 212-224.

[12] A. J. G. Cairns, D. Blake and K. Dowd, A Two-Factor Model for Stochastic Mortality with Parameter Uncertainty: Theory and Calibration, *The Journal of Risk and Insurance*, **73** (2006), no. 4, 687-718.

[13] A. J. G. Cairns, D. Blake, K. Dowd, G. D. Coughlan, D. Epstein, A. Ong, and I. Balevich, A Quantitative Comparison of Stochastic Mortality Models Using Data From England and Wales and the United States, *North American Actuarial Journal*, **13** (2009), no. 1, 1-35.

[14] C. K. Carter and R. Kohn, On Gibbs sampling for state space models, *Biometrika* **81** (1994), 541-553.
[15] G. Celeux, M. Hurn, and C. P. Robert, Computational and Inferential Difficulties with Mixture Posterior Distributions, *Journal of the American Statistical Association* **95** (2000), no. 451, 957-970.
[16] S. Chib, Marginal Likelihood from the Gibbs Output, *Journal of the American Statistical Association* **90** (1995), no. 432, 1313-1321.
[17] I. D. Currie, On fitting generalized linear and non-linear models of mortality, *Scandinavian Actuarial Journal* (2016), 356-383.
[18] I. D. Currie, M. Durban, and P. HC. Eilers, Smoothing and forecasting mortality rates, *Statistical Modelling* **4** (2004), 279-298.
[19] C. Czado, A. Delwarde, and M. Denuit, Bayesian Poisson log-bilinear mortality projections, *Insurance: Mathematics and Economics* **36** (2005), 260-284
[20] A. C. Davison, *Statistical Models (Cambridge Series in Statistical and Probabilistic Mathematics)*. Cambridge: Cambridge University Press, 2003.
[21] P. de Jong and N. Shephard, The Simulation Smoother for Time Series Models, *Biometrika* **82** (1995), no. 2, 339-350.
[22] J. Durbin and S. J. Koopman, *Time Series Analysis by State Space Methods: Second Edition*, Oxford: Oxford University Press, 2012.
[23] EIOPA, EIOPA's Second Set of Advice to the European Commission on Specific Items in the Solvency II Delegated Regulation, EIOPA-BoS-18/075 (2018), https://eiopa.europa.eu/Publications/Consultations/EIOPA-18-075-EIOPA_Second_set_of_Advice_on_SII_DR_Review.pdf
[24] S. Frühwirth-Schnatter, Data Augmentation and Dynamic Linear Models, *Journal of Time Series Analysis* **15** (1994), 183-202.
[25] S. Frühwirth-Schnatter, Fully Bayesian Analysis of Switching Gaussian State Space Models, *Annals of the Institute of Statistical Mathematics* **53** (2000), 31-49.
[26] S. Frühwirth-Schnatter, Markov Chain Monte Carlo Estimation of Classical and Dynamic Switching and Mixture Models, *Journal of the American Statistical Association* **96** (2001), no. 453, 194-209.
[27] MC. Fung, G. W. Peters, and P. V. Shevchenko, A unified approach to mortality modelling using state-space framework: characterization, identification, estimation and forecasting, *Annals of Actuarial Science* **11** (2017), no. 2, 343-389.

[28] K. R. Gabriel, Least Squares Approximation of Matrices by Additive and Multiplicative Models, *Journal of the Royal Statistical Society. Series B (Methodological)* **40** (1978), no. 2, 186-196.

[29] K. R. Gabriel, Generalised Bilinear Regression, *Biometrika* **85** (1998), no. 3, 689-700.

[30] A. E. Gelfand and A. F. M. Smith, Sampling-Based Approaches to Calculating Marginal Densities, *Journal of the American Statistical Association* **85** (1990), no. 410, 398-409.

[31] A. Gelman, J. B. Carlin, H. S. Stern, D. B. Dunson, and D. B. Rubin, *Bayesian Data Analysis, 3rd ed.*, Boca Raton, FL: Chapman and Hall/CRC, 2013.

[32] S. Geman and D. Geman, Stochastic Relaxation, Gibbs Distributions and the Bayesian Restoration of Images, *IEEE Transactions on Pattern Analysis and Machine Intelligence* **6** (1984), 721-741.

[33] J. Geweke, Interpretation and inference in mixture models: Simple MCMC works, *Computational Statistics & Data Analysis* **51** (2007), 3529-3550.

[34] I. J. Good, Some Applications of the Singular Decomposition of a Matrix, *Technometrics* **11** (1969), no. 4, 823-831.

[35] L. A. Goodman, Simple Models for the Analysis of Association in Cross-Classifications having Ordered Categories, *Journal of the American Statistical Association* **74** (1979), no. 367, 537-552.

[36] F. Girosi and G. King, *Demographic Forecasting*, Princeton: Princeton University Press, 2008, http://j.mp/2nxZqJD.

[37] S. Haberman and A. Renshaw, A comparative study of parametric mortality projection models. *Insurance: Mathematics and Economics* **48** (2011), no. 1, 35-55.

[38] J. D. Hamilton, A New Approach to the Economic Analysis of Nonstationary Time Series and the Business Cycle. *Econometrica* **57** (1989), 357-384.

[39] J. D. Hamilton, *Time series analysis*, Princeton, N.J: Princeton University Press, 1994.

[40] A. C. Harvey, *Forecasting, structural time series models and the Kalman filter*, Cambridge: Cambridge University Press, 1989.

[41] H. Hotelling, Analysis of a complex of statistical variables into principal components, *Journal of Educational Psychology* **24** (1933), 417-441 and 498-520.

[42] A. S. Householder and G. Young, Matrix Approximation and Latent Roots, *The American Mathematical Monthly* **45** (1938), no. 3, 165-171.

[43] A. Hunt and D. Blake, A General Procedure for Constructing Mortality Models, *North American Actuarial Journal*, **18** (2014), no. 1, 116-138.

[44] E. Ignatavičiūtė, R. Mikalauskaitė-Arminienė, and J. Šiaulys, Lee-Carter mortality forecasting. *Lithuanian Journal of Statistics* **51** (2012), 22-35.

[45] SF. Jarner and T. Møller, A partial internal model for longevity risk, *Scandinavian Actuarial Journal* (2015), 352-382.

[46] R. E. Kalman, A new approach to linear filtering and prediction problems, *Transactions of the ASME – Journal of Basic Engineering* (Series D) **82** (1960), 35-45.

[47] S. Kaufmann, Measuring business cycles with a dynamic Markov switching factor model: An assessment using Bayesian simulation methods, *Econometrics Journal* **3** (2000), no. 1, 39-65.

[48] CJ. Kim and C. R. Nelson, Dynamic Factor Model with Regime Switching, *The Review of Economics and Statistics* **80** (1998), no. 2, 188-201.

[49] CJ. Kim and C. R. Nelson, *State Space Models with Regime Switching: Classical and Gibbs Sampling Approaches with Applications*, Cambridge, MA: MIT Press, 1999.

[50] T. Kleinov and S. J. Richards, Parameter risk in time-series mortality forecasts, *Scandinavian Actuarial Journal* (2017), 804-828.

[51] A. Kogure and Y. Kurachi, A Bayesian approach to pricing longevity risk based on risk-neutral predictive distributions, *Insurance: Mathematics and Economics* **46** (2010), 162-172.

[52] MC. Koissi, A. F. Shapiro, and G. Högnäs, Evaluating and extending the Lee-Carter model for mortality forecasting: Bootstrap confidence interval, *Insurance: Mathematics and Economics* **38** (2006), 1-20.

[53] R. D. Lee and L. R. Carter, Modeling and forecasting US mortality, *Journal of the American Statistical Association* **87** (1992), no. 419, 659-671.

[54] R. D. Lee and T. Miller, Evaluating the performance of the Lee-Carter method for forecasting mortality, *Demography* **38** (2001), no. 4, 537-549.

[55] J. Li, A quantitative comparison of simulation strategies for mortality projection. *Annals of Actuarial Science* **8** (2014), no. (2): 281-297.

[56] J. S. H. Li, M. R. Hardy, and K. S. Tan, Uncertainty in Mortality Forecasting: An Extension to the Classical Lee-Carter Approach, *ASTIN Bulletin* **39** (2009), no. 1, 137-164.

[57] P. McCullagh and J. Nelder, *Generalized Linear Models, Second Edition*, London: Chapman & Hall, 1989.

[58] A. J. McNeil, R. Frey, and P. Embrechts, *Quantitative Risk Management: Concepts, Techniques and Tools*, Princeton: Princeton University Press, 2005, p. 38.

[59] D. Munroe, B. Zehnwirth, and I. Goldenberg, Solvency capital requirement and claims development result, *British Actuarial Journal* 23 (2018), 1-16.

[60] A. Olivieri and E. Pitacco, Stochastic mortality: The impact on target capital, *ASTIN Bulletin* **39** (2009), 541-563.

[61] C. Pedroza, A Bayesian forecasting model: predicting U.S. male mortality, *Biostatistics* **7** (2006), no. 4, 530-550.

[62] G. Petris, S. Petrone, and P. Campagnoli, *Dynamic Linear Models with R*, New York: Springer-Verlag, 2009.

[63] M. K. Pitt and N. Shephard, Filtering via Simulation: Auxiliary Particle Filters, *Journal of the American Statistical Association* **94** (1999), no. 446, 590-599.

[64] R. Plat, One-year Value-at-Risk for longevity and mortality, *Insurance: Mathematics and Economics* **49** (2011), 462-470.

[65] R Core Team, R: A language and environment for statistical computing, R Foundation for Statistical Computing, Vienna, Austria, 2020, https://www.R-project.org/.

[66] A. E. Renshaw and S. Haberman, Lee-Carter mortality forecasting with age-specific enhancement, *Insurance: Mathematics and Economics* **33** (2003), 255-272.

[67] A. E. Renshaw and S. Haberman, Lee-Carter mortality forecasting: a parallel generalized linear modelling approach for England and Wales mortality projections, *Applied Statistics* **52** (2003), no. 1, 119-137.

[68] A. E. Renshaw and S. Haberman, On simulation-based approaches to risk measurement in mortality with specific reference to Poisson Lee-Carter modelling, *Insurance: Mathematics and Economics* **42** (2008), 797-816.

[69] S. J. Richards, I. D. Currie, and G. P. Ritchie, A Value-at-Risk framework for longevity trend risk, *British Actuarial Journal* **19** (2013), 116-39.

[70] S. Särkkä, *Bayesian Filtering and Smoothing*, Cambridge: Cambridge University Press, 2013.

[71] R. H. Shumway and D. S. Stoffer, *Time Series Analysis and Its Applications with R Examples, Fourth Edition*, New York: Springer-Verlag, 2016.

[72] D. Toczydlowska, G.W. Peters, M.C. Fung, P.V. Shevchenko, Stochastic Period and Cohort Effect State-Space Mortality Models Incorporating Demographic Factors via Probabilistic Robust Principal Components. *Risks* **5** (2017), 42.

[73] S. NL. Tuljapurkar and C. Boe, Mortality Change and Forecasting, *North American Actuarial Journal* **2** (1998), no. 4, 13-47.

[74] S. NL. Tuljapurkar and C. Boe, A universal pattern of mortality decline in the G7 countries, *Nature* **405** (2000), 789-792.

[75] F. A. Van Eeuwijk, Multiplicative Interaction in Generalized Linear Models, *Biometrics* **51** (1995), no. 3, 1017-1032.

[76] R. W. M. Wedderburn, Quasi-Likelihood Functions, Generalized Linear Models, and the Gauss-Newton Method, *Biometrika* **61** (1974), no. 3, 439-447.

[77] M. West and J. Harrison, *Bayesian Forecasting and Dynamic Models, Second Edition,* New York: Springer-Verlag, 1999.

LIST OF PUBLICATIONS

The results presented in this disseration were published in the following publications:

R. Gylys, and J. Šiaulys. Revisiting Calibration of the Solvency II Standard Formula for Mortality Risk: Does the Standard Stress Scenario Provide an Adequate Approximation of Value-at-Risk? *Risks* 7 (2019), 58.
https://www.mdpi.com/2227-9091/7/2/58

R. Gylys, and J. Šiaulys. Estimation of Uncertainty in Mortality Projections Using State-Space Lee-Carter Model. *Mathematics* 8 (2020), 7.
https://www.mdpi.com/2227-7390/8/7/1053

CONFERENCES

The results derived in this disseration were presented at the following conferences:

R. Gylys, Modelling uncertainty in human mortality projections using state-space Lee-Carter, *StatMod2020: Statistical Modeling with Applications* "Gheorghe Mihoc-Caius Iacob" Institute of Mathematical Statistics and Applied Mathematics, Romania and Laboratory of Mathematics Raphael Salem, University of Rouen-Normandy, France, Online conference, November 2020.

R. Gylys, State space representation of Lee-Carter stochastic mortality model: application to modelling of insurers' solvency capital, *Data Analysis Methods for Software Systems*, Vilnius University Institute of Data Science and Digital Technologies, the Lithuanian Academy of Sciences and the Lithuanian Computer Society. Druskininkai, November 2019.

R. Gylys, Draudimo įmonės mokumo kapitalo reikalavimo mirtingumo rizikai vertinimas VaR metodu, *Lietuvos matematikų draugijos LX konferencija*, Vilnius, June 2019.

NOTES

NOTES

NOTES

Vilnius University Press
9 Saulėtekio Ave., Building III, LT-10222 Vilnius
Email: info@leidykla.vu.lt, www.leidykla.vu.lt
Print run copies 15