ANNE PUUSTELLI

Bayesian Methods in Insurance Companies' Risk Management

ACADEMIC DISSERTATION
To be presented, with the permission of
the board of the School of Information Sciences
of the University of Tampere,
for public discussion in the Auditorium Pinni B 1097,
Kanslerinrinne 1, Tampere, on December 10th, 2011, at 12 o'clock.

UNIVERSITY OF TAMPERE

UNIVERSITY OF TAMPERE

ACADEMIC DISSERTATION
University of Tampere
School of Information Sciences
Finland

Distribution
Bookshop TAJU
P.O. Box 617
33014 University of Tampere
Finland

Tel. +358 40 190 9800
Fax +358 3 3551 7685
taju@uta.fi
www.uta.fi/taju
http://granum.uta.fi

Cover design by
Mikko Reinikka

Acta Universitatis Tamperensis 1681
ISBN 978-951-44-8635-7 (print)
ISSN-L 1455-1616
ISSN 1455-1616

Acta Electronica Universitatis Tamperensis 1145
ISBN 978-951-44-8636-4 (pdf)
ISSN 1456-954X
http://acta.uta.fi

Tampereen Yliopistopaino Oy – Juvenes Print
Tampere 2011

Acknowledgments

I would like to express my deepest gratitude to my supervisor Dr. Arto Luoma for all the guidance, support, patience and encouragement he has given me during this project. Dr. Luoma has wide expertise in Bayesian analysis and it has been a great privilege and pleasure to have him as my teacher.

I have also been privileged to have Dr. Lasse Koskinen as my second supervisor. He was the one who introduced me to this interesting research area and organized a joint project *Insurance company's risk management models* between the Insurance Supervisory Authority of Finland and the Department of Mathematics and Statistics in the University of Tampere. For this I am deeply grateful, as it was a cornerstone for this thesis to materialize. The two-year project started in January 2007 and was also a starting-point for the preparation of this thesis. Dr. Koskinen has given me support, enthusiasm and encouragement as well as pleasant travelling company in the conferences we attended together.

I would like to thank Mr. Vesa Ronkainen for introducing me to the challenges of life insurance contracts. His fruitful discussions and crucial suggestions made a notabe contribution to this work. A special thankyou goes to Dr. Laura Koskela for reading and commenting on the introduction to the thesis and for being not only an encouraging colleague but also a good friend through many years.

I would like to express my graditude to Professor Erkki P. Liski for encouraging me during my postgraduate studies and for letting me act as a course assistant on his courses. I would also like to thank all the other personnel in the Department of Mathematics and Statistics in the University of Tampere. Especially I want to thank Professor Tapio Nummi, who hired me as a researcher on his project. The experience I gained from the project made it possible for me to become a graduate school student.

I owe thanks to Mr. Robert MacGilleon, who kindly revised the language of this thesis. I would also like to thank Jarmo Niemelä for his help with LaTeX.

For financial support I wish to thank the Tampere Graduate School in Information Science and Engineering (TISE), the Vilho, Yrjö and Kalle Väisälä Foundation in the Finnish Academy of Science and Letters, the Insurance Supervisory Authority of Finland, the Scientific Foundation of the City of Tampere and the School of Information Sciences, University of Tampere. I am also grateful to the Department of Mathematics and Statistics, University of Tampere, for supporting me with facilities.

Finally, I want to thank my parents, close relatives and friends for their support and encouragement during my doctoral studies. And lastly, I wish to express my warmest gratitude to my husband, Janne, for his love, support and patience, and to our precious children Ella and Aino, who really are the light of our lives.

Espoo, November 2011 *Anne Puustelli*

Abstract

In this thesis special issues emerging from insurance companies' risk management are considered in four research articles and in a brief introduction to concepts examined in the articles. The three main topics in the thesis are financial guarantee insurance, equity-linked life insurance contracts, and mortality modeling.

Common to all of the articles is the utilization of Bayesian methods. With these the model and parameter uncertainty can be taken into account. As demonstrated in this thesis, oversimplified models or oversimplified assumptions may cause catastrophic losses for an insurance company. As financial systems become more complex, risk management needs to develop at the same time. Thus, model complexity cannot be avoided if the true magnitude of the risks the insurer faces is to be revealed. The Bayesian approach provides a means to systematically manage complexity.

The topics studied here serve a need arising from the new regulatory framework for the European Union insurance industry, known as Solvency II. When Solvency II is implemented, insurance companies are required to hold capital not only against insurance liabilities but also against, for example, market and credit risk. These two risks are closely studied in this thesis. Solvency II also creates a need to develop new types of products, as the structure of capital reguirements will change. In Solvency II insurers are encouraged to measure and manage their risks based on internal models, which will become valuable tools. In all, the product development and modeling needs caused by Solvency II were the main motivation for this thesis.

In the first article the losses ensuing from the financial guarantee system of the Finnish statutory pension scheme are modeled. In particular, in the model framework the occurrence of an economic depression is taken into account, as losses may be devastating during such a period. Simulation results show that the required amount of risk capital is high, even though depressions are an infrequent phenomenon.

In the second and third articles a Bayesian approach to market-consistent valuation and hedging of equity-linked life insurance contracts is introduced. The framework is assumed to be fairly general, allowing a search for new insurance savings products which offer guarantees and certainty but in a capital-efficient manner. The model framework includes interest rate, volatility and jumps in the asset dynamics to be stochastic, and stochastic mortality is also incorporated. Our empirical results support the use of elaborated instead of stylized models for asset dynamics in practical applications.

In the fourth article a new method for two-dimensional mortality modeling is proposed. The approach smoothes the data set in the dimensions of cohort and age using Bayesian smoothing splines. To assess the fit and plausibility of our models we carry out model checks by introducing appropriate test quantities.

Key words: Equity-linked life insurance, financial guarantee insurance, hedging, MCMC, model error, parameter uncertainty, risk-neutral valuation, stochastic mortality modeling

Contents

List of original publications

I. Puustelli, A., Koskinen, L., Luoma, A., 2008. Bayesian modelling of financial guarantee insurance. *Insurance: Mathematics and Economics*, 43, 245–254.

The initial version of the paper was presented in AFIR Colloquium, Stockholm, Sweden, 12.–15.6.2007.

II. Luoma, A., Puustelli, A., Koskinen, L., 2011. Bayesian analysis of equity-linked savings contracts with American-style options. Submitted.

The initial version of the paper titled 'Bayesian analysis of participating life insurance contracts with American-style options' was presented in AFIR Colloquium, Rome, Italy, 30.9.–3.10.2008.

III. Luoma, A., Puustelli, A., 2011. Hedging equity-linked life insurance contracts with American-style options in Bayesian framework. Submitted.

The initial version of the paper titled 'Hedging against volatility, jumps and longevity risk in participating life insurance contracts – a Bayesian analysis' was presented in AFIR Colloquium, Munich, Germany, 8.–11.9.2009.

IV. Luoma, A., Puustelli, A., Koskinen, L., 2011. A Bayesian smoothing spline method for mortality modeling. Conditionally accepted in *Annals of Actuarial Science*, Cambridge University Press.

Papers I and IV are reproduced with the kind permission of the journals concerned.

1 Introduction

This thesis consists of four research papers which consider specific problems arising from insurance companies' risk management. In particular, the papers propose models and techniques for risk mitigation in financial guarantee insurance, equity-linked life insurance contracts, and mortality modeling. In all steps throughout the research articles Bayesian methods are utilized. In this chapter a brief introduction to risk management in the insurance industry is provided and Chapter 2 introduces Bayesian analysis. Chapter 3 serves as an introduction to derivative pricing, which is a fundamental cornerstone in Papers II and III. Summaries of original publications discuss our proposed models in financial guarantee insurance, equity-linked life insurance policies and mortality modeling.

Risk management has become a matter of fundamental importance in all sectors of the insurance industry. Various types of risks need to be quantified to ensure that insurance companies have adequate capital, solvency capital, to support their risks. Over 30 years ago Pentikäinen (1975) argued that actuarial methods should be extended to a full-scale risk management process. Later Pentikäinen et al. (1982) and Daykin et al. (1994) suggested that solvency should be evaluated through numerous sub-problems which jeopardize solvency. These include, for example, model building, variations in risk exposure and catastrophic risks.

Better risk management is a focus in the new regulatory framework for the European Union insurance industry, known as Solvency II, which is expected to be implemented by the end of 2012 (see European Commission, 2009). At the moment mainly insurance risks are covered by the EU solvency requirements, which are over 30 years old. As financial and insurance markets have recently developed dramatically, wide discrepancy prevails between the reality of the insurance business and its regulation.

Solvency II is designed to be more risk-sensitive and sophisticated compared to current solvency requirements. The main improvement consists in requiring companies to hold capital also against market risk, credit risk and operational risk. In other words, not only liabilities need to be taken into account, but also, for example, the risks of a fall in the value of the insurers' investments, of third parties' inability to repay their debts and of systems breaking down or of malpractice. Recent developments in financial reporting (IFRS) and banking supervision (Basel II) have also undergone similar changes. This thesis focuses on market risk, which affects equity-linked life insurance policies. In addition, credit risk is studied in the context of financial guarantee insurance.

Solvecy II will increase the price of more capital-intensive products such as equity-linked life insurance contracts with capital guarantees. This creates a need to develop new types of products to fulfill the customer demands for traditional life contracts but in a capital-efficient manner (Morgan Stanley and Oliver Wyman, 2010). One important objective in this thesis was to address this need.

In Solvency II insurers are encouraged to measure and manage their risks based on internal models (see, e.g., Ronkainen et al., 2007). The Groupe Consultatif defines the internal model in its Glossary on insurance terms as "Risk management system of an insurer for the analysis of the overall risk situation of the insurance undertaking, to quantify risks and/or to determine the capital requirement on the basis of the company specific risk profile." Hence, internal models will become valuable tools, but are also subject to model risk. A model risk might be caused by a misspecified model or by incorrect model usage or implementation. In particular, the true magnitude of the risks the insurer faces may easily go unperceived when oversimplified models or oversimplified assumptions are used.

As Turner et al. (2010) point out, the recent financial crisis, which started in the summer of 2007, showed the danger of relying on oversimplified models and increased the demand for reliable quantitative risk management tools. Generally, unnecessary complexity is undesirable, but as the financial system becomes more complex, model complexity cannot be avoided. The Bayesian approach provides tools to easily extend the analysis to more complex models. Bayesian inference is particularly attractive from the insurance companies' point of view, since it is exact in finite samples. An exact characterization of finite sample uncertainty is critical in order to avoid crucial valuation errors. Another advantage of Bayesian inference is its ability to incorporate prior information in the model.

In general, uncertainty in actuarial problems arises from three principal sources, namely, the underlying model, the stochastic nature of a given model and the parameter values in a given model (see, e.g., Draper, 1995; Cairns, 2000). To quantify parameter and model uncertainty in insurance Cairns (2000) has also chosen the Bayesian approach. His study shows that a contribution to the outcome of the modeling exercise was significant when taking into account both model and parameter uncertainty using Bayesian analysis. Likewise Hardy (2002) studied model and parameter uncertainty using a Bayesian framework in risk management calculations for equity-linked insurance.

In this thesis model and parameter uncertainty is taken into account by following the Bayesian modeling approach suggested by Gelman et al. (2004, Section 6.7). They recommend constructing a sufficiently general, continuously parametrized model which has models in interest as its special cases. If a generalization of a simple model cannot be constructed, then model comparison is suggested to be done by measuring the distance of the data to each of the models in interest. The criteria which may be used to measure the discrepancy between the data and the model are discussed in Section 2.

As insurance supervision is undergoing an extensive reform and at the same time the financial and insurance market is becoming more complex, risk management in insurance is required to improve without question. However, more advanced risk management will become radically more complicated to handle, and complicated systems have a substantial failure risk in system management. The focus in this thesis is on contributing statistical models using the Bayesian approach for insurance companies' risk management. This approach is chosen since it provides means to systematically manage complexity. Computational methods in statistics play the primary role here, as the techniques used require high computational intensity.

2 Bayesian analysis

Bayesian data analysis provides practical methods for drawing inferences regarding unobserved quantities from a set of data using probability models for both observed and unobserved quantities. The explicit use of probability to quantify uncertainty is the essential characteristic of Bayesian methods. Gelman et al. (2004) divide the process of Bayesian data analysis into three steps:

1. Setting up a full probability model – a joint probability distribution for all observable and unobservable quantities in the problem. The model should be consistent with knowledge about the underlying scientific problem and the data collection process.

2. Conditioning on observed data: calculating and interpreting the appropriate posterior distribution – the conditional probability distribution of the unobserved quantities of ultimate interest, given the observed data.

3. Evaluating the fit of the model and the implication of the resulting posterior distribution: does the model fit the data, are the substantive calculations reasonable, and how sensitive are the results to the modeling assumptions in step 1? If necessary, one can alter or expand the model and repeat the three steps.

These three steps are taken in all the articles in this thesis.

In Bayesian inference the name Bayesian comes from the use of the theorem introduced by the Reverend Thomas Bayes in 1764. Bayes' theorem gives a solution to the inverse probability problem, which yields the posterior density:

$$p(\theta|y) = \frac{p(\theta, y)}{p(y)} = \frac{p(\theta)p(y|\theta)}{p(y)},$$

where θ denotes unobservable parameters of interest and y denotes the observed data. Further, $p(\theta)$ is referred to as the prior distribution and $p(y|\theta)$ as the sampling distribution or the likelihood function. Now $p(y) = \sum_\theta p(\theta)p(y|\theta)$ in the case of discrete θ and $p(y) = \int p(\theta)p(y|\theta)d\theta$ in the case of continuous θ. With fixed y the factor $p(y)$ does not depend on θ and can thus be considered as a constant. Omitting $p(y)$ yields the unnormalized posterior density

$$p(\theta|y) \propto p(\theta)p(y|\theta),$$

which is the technical core of Bayesian inference.

The prior distribution can be used to incorporate the prior information in the model. Uninformative prior distributions (for example, uniform distribution) for parameters can be used in the absence of the prior information or when information derived only from the data is chosen. The choice of the uninformative prior is not unique, and hence to some extent controversial. However, the role of prior distribution decreases and becomes insignificant in most cases as the data set becomes larger.

2.1 Posterior simulation

In applied Bayesian analysis inference is typically carried out by simulation. This is due simply to the fact that closed form solutions of posterior distributions exist only in special cases. Even if the posterior distribution in some complex special cases were solved analytically, the algebra would become extremely difficult and a full Bayesian analysis of realistic probability models would be too burdensome for most practical applications. By simulating samples from the posterior distribution, exact inference may be conducted, since sample summary statistics provide estimates of any aspect of the posterior distribution to a level of precision which can be estimated. Another advantage in simulation is that a potential problem with model specification or parametrization can be detected from extremely large or small simulated values. These problems might not be perceived if estimates and probability statements were obtained in analytical form.

The most popular simulation method in the Bayesian approach is Markov chain Monte Carlo (MCMC) simulation, which is used when it is not possible or computationally efficient to sample directly from the posterior distribution. The MCMC methods have been used in a large number and wide range of applications also outside Bayesian statistics, and are very powerful and reliable when cautiously used. A useful reference for different versions of MCMC is Gilks et al. (1996).

MCMC simulation is based on creating a Markov chain which converges to a unique stationary distribution which is the desired target distribution $p(\theta|y)$. The chain is created by first setting the starting point θ^0 and then iteratively drawing θ^t, $t = 1, 2, 3, \ldots$, from a transition probability distribution $T(\theta^t|\theta^{t-1})$. The key is to set the transition distribution such that the chain converges to the target distribution. It is important to run the simulation long enough to ensure that the distribution of the current draws is close enough to the stationary distribution. The Markov property of the distributions of the sampled draws is essential when the convergency of the simulation result is assessed.

Throughout all articles, our estimation procedure is one of the MCMC methods called a single-component (or cyclic) Metropolis-Hastings algorithm or two of its special cases, Metropolis algorithm and Gibbs sampler. The Metropolis-Hastings algorithm was introduced by Hastings (1970) as a generalization of the Metropolis algorithm (Metropolis et al., 1953). Also the Gibbs sampler proposed by Geman and Geman (1984) is a special case of the Metropolis-Hastings algorithm. The Gibbs sampler assumes the full conditional distributions of the target distribution to be such that one is able to generate random numbers or vectors from them. The Metropolis and Metropolis-Hastings algorithms are more flexible than the Gibbs sampler; with them one only needs to know the joint density function of the target distribution with density $p(\theta)$ up to a constant of proportionality.

With the Metropolis algorithm the target distribution is generated as follows: first a starting distribution $p_0(\theta)$ is assigned, and from it a starting-point θ^0 is drawn such that $p(\theta^0) > 0$. For iterations $t = 1, 2, \ldots$, a proposal θ^* is generated from a jumping distribution $J(\theta^*|\theta^{t-1})$, which is symmetric in the sense that $J(\theta_a|\theta_b) = J(\theta_b|\theta_a)$ for all θ_a and θ_b. Finally, iteration t is completed by calculating the ratio

$$r = \frac{p(\theta^*)}{p(\theta^{t-1})} \tag{2.1}$$

and by setting the new value at

$$\theta^t = \begin{cases} \theta^* & \text{with probability } \min(r, 1) \\ \theta^{t-1} & \text{otherwise.} \end{cases}$$

Thus, under the Metropolis algorithm, the transition distribution $T(\theta^t|\theta^{t-1})$ of the Markov chain is a mixture of a point mass at $\theta^t = \theta^{t-1}$ and the weighted version of the jumping distribution $J(\theta^t|\theta^{t-1})$.

The Metropolis-Hastings algorithm generalizes the Metropolis algorithm by removing the assumption of symmetric jumping distribution. The ratio r in (2.1) is replaced by

$$r = \frac{p(\theta^*)/J(\theta^*|\theta^{t-1})}{p(\theta^{t-1})/J(\theta^{t-1}|\theta^*)}$$

to correct for the asymmetry in the jumping rule.

In the single-component Metropolis-Hastings algorithm the simulated random vector is divided into components or subvectors which are updated one by one. Besides being parameters in the model, these components or subvectors might also be latent variables in it. If the jumping distribution for a component is its full conditional posterior distribution, the proposals are accepted with probability one. In the case where all the components are simulated in this way, the algorithm is the Gibbs sampler. It can be shown that these algorithms produce an ergodic Markov chain whose stationary distribution is the target distribution.

It is absolutely necessary to check the convergence of the simulated sequences to ensure the distribution of the current draws in the process is close enough to the stationary distribution. In particular, two difficulties are involved in inference carried out by iterative simulation.

First, the starting approximation should not affect the simulation result under regularity conditions, which are irreducibility, aperiodicity and positive recurrence. The chain is irreducible if it is possible to get to any value of the parameter space from any other value of the parameter space; positively recurrent if it returns to the specific value of the parameter space at finite times; and aperiodic if it can return to the specific value of the parameter space at irregular times. By simulating multiple sequences with starting-points dispersed throughout the parameter space, and discarding early iterations of the simulation runs (referred to as a burn-in period), the effect of the starting distribution may be diminished.

Second, the Markov property introduces autocorrelation in the within-sequence. Aside from any convergence issues, the simulation inference from correlated draws is generally less precise than that from the same number of independent draws. However, at convergency, serial correlation in the simulations is not necessarily a problem, as the order of simulations is in any case ignored when preforming the inference. The concept of mixing describes how much draws can move around the parameter space in each cycle. The better the mixing is, the closer the simulated values are to the independent sample and the faster the autocorrelation approaches zero. When the mixing is poor, more cycles are needed for the burn-in period as well as to attain to a given level of precision for the posterior distribution.

To monitor convergency, the variations between and within simulated sequences are compared until within-variation roughly equals between-variation. Simulated sequences can only approximate the target distribution when the distribution of each simulated sequence is close to the distribution of all the sequences mixed together.

Gelman and Rubin (1992) introduce a factor by which the scale of the current distribution for a scalar estimand ψ might be reduced if the simulation were continued in the limit $n \to \infty$. Denote the simulation draws as ψ_{ij} ($i = 1, \ldots, n$; $j = 1, \ldots, m$), where the length of the sequence is n (after discarding the first half of the simulations as burn-in period) and the number of parallel sequences is m. Further, let B and W denote the between- and within-sequence variances, respectively, computed as

$$B = \frac{n}{m-1} \sum_{j=1}^{m} \left(\bar{\psi}_{.j} - \bar{\psi}_{..}\right)^2, \quad \text{where} \quad \bar{\psi}_{.j} = \frac{1}{n} \sum_{i=1}^{n} \psi_{ij}, \quad \bar{\psi}_{..} = \frac{1}{m} \sum_{j=1}^{m} \bar{\psi}_{.j},$$

and

$$W = \frac{1}{m} \sum_{j=1}^{m} s_j^2, \quad \text{where} \quad s_j^2 = \frac{1}{n-1} \sum_{i=1}^{n} \left(\psi_{ij} - \bar{\psi}_{.j}\right)^2.$$

The marginal posterior variance of the estimand can be estimated by a weighted average of B and W, namely

$$\widehat{\text{var}}^{+}(\psi|y) = \frac{n-1}{n} W + \frac{1}{n} B.$$

Finally, the potential scale reduction is estimated by

$$\widehat{R} = \sqrt{\frac{\widehat{\text{var}}^{+}(\psi|y)}{W}},$$

which declines to 1 as $n \to \infty$. If $\widehat{R}$ is high, then proceeding the simulation can presumably improve the inference about the target distribution of the associated scalar estimand.

2.2 Model checking

Assessing the fit of the model to the data and to our substantive knowledge is a fundamental step in statistical analysis. In the Bayesian approach replicated data sets produced by means of posterior predictive simulation may be used to check the model fit. In detail, a replicated data set is produced by first generating the unknown parameters from their posterior distribution and then, given these parameters, the new data values. Once several replicated data sets y^{rep} have been produced, they may be compared with the original data set y. If they look similar to y, the model fits.

The discrepancy between the data and the model may be measured by defining an arbitrary test quantity which is a scalar summary of parameters and the data. The value of the test quantity is computed for each posterior simulation using both original and replicated data sets. The same set of parameters is used in both cases. If the test quantity depends only on data and not on parameters, then it is said to be a test statistic. The Bayesian p-value is defined to be the posterior probability that the test quantity computed from a replication, $T(y^{rep}, \theta)$, exceeds that computed from the original data, $T(y, \theta)$. This test may be illustrated by a scatter plot of $(T(y, \theta), T(y^{rep}, \theta))$, where the same scale is used for both coordinates. Further details on this approach can be found in Chapter 6 of Gelman et al. (2004) or Chapter 11 of Gilks et al. (1996).

If one wishes to compare the fit of different models, nested or nonnested, over the same set of data, a deviance $D(y, \theta) = -2 \log p(y|\theta)$ may be used. The expected deviance – computed by averaging the deviance over the true sampling distribution – has a connection to the Kullback-Leibler information measure and is thus a reasonable measure of overall model fit. The model with the lowest Kullbach-Leibler information and thus the lowest expected deviance has the highest posterior probability. To obtain a summary which depends only on y, the average discrepancy may be used. It is defined as $D_{\text{avg}}(y) = \mathrm{E}(D(y, \theta)|y)$, where the discrepancy is averaged over the posterior distribution and is estimated as $\widehat{D}_{\text{avg}}(y) = \sum_{l=1}^{L} D(y, \theta_l)/L$, where the vectors θ_l are posterior simulations.

2.3 Computational aspect

In this thesis fairly general and complex models allowed by the Bayesian approach are used. These models require high computational intensity and thus, the computational aspects are in the primary role throughout all papers. All the computations in this thesis were performed using the R computing environment (see R Development Core Team, 2009). A special R library called LifeIns was developed for computations used in Paper III, and the entire code used in other papers is available in http://mtl.uta.fi/codes.

In Paper I a Markov regime-switching model or more precisely, a Hamilton model (Hamilton, 1989), is used to model the latent economic business cycle process. The posterior simulations of this model are used as an explanatory variable in a transfer function model which models the claim amounts of a financial guarantee insurance. As the business cycle process is assumed to be exogenous in the transfer function model, it can be estimated separately. For both models the Gibbs sampler is used in the estimation. The posterior simulations of the transfer function model are used to simulate the posterior predictive distribution of the claim amounts. A number of model checks introduced earlier in this chapter were performed to assess the fit and quality of the models. In particular, both models were checked by means of data replications, test statistics and residuals. The average discrepancy was calculated to compare the model fit of the Hamilton against the AR(2) model, and for competing transfer function models. Further, robustness and sensitivity analyses were also made.

In Papers II and III the use of the Bayesian approach on pricing and hedging equity-linked life insurance contracts is particularly attractive, since it can link the uncertainty of parameters and several latent variables to the predictive uncertainty of the process. The estimation guidelines provided by Bunnin et al. (2002) are used in Paper II, and in Paper III the guidelines provided by Jones (1998) are followed. Metropolis and Metropolis-Hastings algorithms are used to estimate the unknown parameters of the stock index, volatility and interest rate models as well as to estimate the latent volatility and jump processes. The major challenge in estimation is its high dimensionality, which results from the need to estimate latent processes. In paper III we effectively apply parameter expansion to work out issues in estimation. Further, the contract includes an American-style path-dependent option which is priced using a regression method (see, e.g., Tsitsiklis and Van Roy, 1999). The code also includes valuation of the lower and the upper limit of the price for such a contract. In Paper III a stochastic mortality is incorporated in the framework and we construct a replicating portfolio to study dynamic hedging strategies. In both papers the most time-consuming loops are coded in C++ to speed up computations.

Paper IV introduces a new two-dimensional mortality model utilizing Bayesian smoothing splines. Before estimating the model special functions are developed to form a smaller estimation matrix from the large original data matrix. The estimation is carried out using Gibbs sampler with one Metropolis-Hastings step. Two Bayesian test quantities are developed to test the consistency of the model with historical data. Also the robustness of the parameters as well as the accuracy and robustness of the forecasts are studied.

3 Principles of derivative pricing

In Papers II and III methods of financial mathematics, in particular derivative pricing, are used extensively to price and hedge an equity-linked life insurance contract. Glasserman (2004) describes the three most important principles of derivative pricing as follows:

1. If a derivative security can be perfectly replicated (equivalently, hedged) through trading in other assets, then the price of the derivative security is the cost of the replicating trading strategy.

2. Discounted (or deflated) asset prices are martingales under a probability measure associated with the choice of discount factor (or numeraire). Prices are expectations of discounted payoffs under such a martingale measure.

3. In a complete market, any payoff (satisfying modest regularity conditions) can be synthesized through a trading strategy, and the martingale measure associated with a numeraire is unique. In an incomplete market there are derivative securities that cannot be perfectly hedged; the price of such a deriative is not completely determined by the prices of other assets.

The first principle says the foundation of derivative pricing and hedging, and introduces a principle of arbitrage-free pricing. Arbitrage is a practice of profiting by exploiting the price difference of identical or similar financial instruments, on different markets or in different forms. However, the principle does not give strong tools to evaluate the price in practice. In contrast, the second principle offers a powerful tool by decribing how to represent prices as expectations. This leads to the use of Monte Carlo and other numerical methods.

The third principle describes conditions under which the price of a derivative is determined. In a complete market all risks which affect derivative prices can be perfectly hedged. This is attained when the number of driving Brownian motions of the derivative is less than or equal to the number of instruments used in replication. However, jumps in asset prices cause incompleteness in that the effect of discontinuous movements is often impossible to hedge. In Paper II our set-up is in the complete market, while in Paper III we work in the incomplete market set-up.

Let us describe the dynamics of asset prices S_t by a stochastic differential equation

$$\mathrm{d}S_t = \mu(S_t, t)S_t\mathrm{d}t \ + \ \sigma(S_t, t)S_t\mathrm{d}B_t, \tag{3.1}$$

where B_t is a standard Brownian motion, and $\mu(S_t, t)$ and $\sigma(S_t, t)$ are deterministic functions depending on the current state S_t and time t. These dynamics describe the empirical dynamics of asset prices under a real world probability measure $\mathbb{P}$. We may introduce a risk-neutral probability measure $\mathbb{Q}$ which is a particular choice of equivalent martingale measure to $\mathbb{P}$. These equivalent probability measures agree as to which events are impossible.

The asset dynamics under the risk-neutral probability measure may be expressed as

$$\mathrm{d}S_t = rS_t\mathrm{d}t \;+\; \sigma(S_t,t)S_t\mathrm{d}B_t^o, \tag{3.2}$$

where B_t^o is a standard Brownian motion under $\mathbb{Q}$ and r is a constant risk-free interest rate. The processes (3.1) and (3.2) are consistent if $\mathrm{d}B_t^o = \mathrm{d}B_t + \nu_t\mathrm{d}t$ for some ν_t satisfying $\mu(S_t,t) = r + \sigma(S_t,t)\nu_t$. It follows from the Girsanov Theorem (see, e.g., Glasserman, 2004, Appendix B) that the measures $\mathbb{P}$ and $\mathbb{Q}$ are equivalent if they are related through a change of drift in the driving Brownian motion. To employ a model of the form (3.2) is simpler than a model of the form (3.1), because the drift can be set equal to the risk-free rate rather than to a potentially complicated drift in (3.1). Further, under $\mathbb{P}$ and $\mathbb{Q}$ the diffusion terms $\sigma(S_t,t)$ must be the same. This is important from the estimation point of view, since the parameters describing the dynamics under the risk-neutral measure may be estimated based on the real-world data.

The derivative pricing equation

$$V_t = \exp\left(-r(T-t)\right)\mathrm{E}_{\mathbb{Q}}\left(V_T\right), \quad t < T, \tag{3.3}$$

expresses the current price of the derivative V_t as the expected terminal value V_T discounted at the risk-free rate r. The expectation must be taken under $\mathbb{Q}$. Here V_t is European-style derivative, meaning it can be exercised only on the expiration date. However, in this thesis we have utilized American-style derivatives which can be exercised at any time. In articles II and III it is explained how this type of derivative is priced.

Equation 3.3 is the cornerstone of derivative pricing by Monte Carlo simulation. Under $\mathbb{Q}$ the discounted price process $\tilde{S}_t = \exp(-rt)S_t$ is a martingale. If the constant risk-free rate r is replaced with a stochastic rate r_t, the pricing formula continues to apply and we can express the formula as

$$V_t = \mathrm{E}_{\mathbb{Q}}\left(\exp\left(-\int_t^T r_s ds\right)V_T\right).$$

In Paper II we utilize the constant elasticity of variance (CEV) model introduced by Cox and Ross (1976) to model the equity index process. This generalizes the geometric Brownian motion (GBM) model, which underlies the Black-Scholes approach to option valuation (Black and Scholes, 1973). Although a generalization, the CEV process is still driven by one source of risk, so that option valuation and hedging remain straightforward.

In the case of a stochastic interest rate, we assume the Chan-Karolyi-Longstaff-Sanders (CKLS) model (see Chan et al., 1992), which generalizes several commonly used short-term interest rate models. Now there are two stochastic processes which affect the option valuation and hedging. Perfect hedging would now require two different hedging instruments, but in Paper III we have ignored the risk arising from the stochastic interest rate and used only one instrument to hedge.

The dynamics of the stock index S_t and the riskless short-term rate r_t are described by the following system of SDEs:

$$\begin{aligned} \mathrm{d}S_t &= \mu S_t \mathrm{d}t \ + \ \nu S_t^{1-\alpha} \mathrm{d}B_t^{(1)}, \\ \mathrm{d}r_t &= \kappa(\xi - r_t)\mathrm{d}t + \sigma r_t^{\gamma} \mathrm{d}B_t^{(2)}, \end{aligned}$$

with $B_t^{(1)}$ and $B_t^{(2)}$ two standard Brownian motions, correlated through $B_t^{(1)} = \rho B_t^{(2)} + \sqrt{1-\rho^2} B_t^{(3)}$, where $B_t^{(2)}$ and $B_t^{(3)}$ are independent standard Brownian motions under $\mathbb{P}$.

In Paper III we allow not only the interest rate but also the volatility and jumps in the asset dynamics to be stochastic. For a stochastic interest rate and volatility we assume a square-root diffusion referred to as the Cox-Ingersoll-Ross (CIR) model (Cox et al., 1985). The dynamics of stock index S_t, variance V_t and riskless short-term rate r_t are assumed to be described by the following system of SDEs:

$$\begin{aligned} \mathrm{d}\log S_t &= \mu \mathrm{d}t \ + \ \sqrt{V_t} \mathrm{d}B_t^{(1)} + U_t \mathrm{d}q_t \\ \mathrm{d}V_t &= (\alpha_1 + \beta_1 V_t)\mathrm{d}t + \sigma_V \sqrt{V_t} \mathrm{d}B_t^{(2)} \\ \mathrm{d}r_t &= (\alpha_2 + \beta_2 r_t)\mathrm{d}t + \sigma_r \sqrt{r_t} \mathrm{d}B_t^{(3)}, \end{aligned}$$

where $B_t^{(1)}$, $B_t^{(2)}$ and $B_t^{(3)}$ are standard Brownian motions, and q_t is a jump process with jump size U_t. We further assume that these Brownian motions have the correlation structure

$$\mathrm{Cor}\left(B_t^{(1)}, B_t^{(2)}, B_t^{(3)}\right) = \begin{pmatrix} 1 & \rho_{12} & \rho_{13} \\ \rho_{12} & 1 & \rho_{23} \\ \rho_{13} & \rho_{23} & 1 \end{pmatrix},$$

and q_t is a Poisson process with intensity λ, that is, $\Pr(\mathrm{d}q_t = 1) = \lambda \mathrm{d}t$ and $\Pr(\mathrm{d}q_t = 0) = 1 - \lambda \mathrm{d}t$. Conditional on a jump occurring, we assume that $U_t \sim \mathrm{N}(a, b^2)$. In addition, we assume that q_t is uncorrelated with U_t or with any other process.

Euler discretization is used in the estimation of the unknown parameters for all the models, since the transition densities of the multivariate processes described above do not have a closed form solution. Accordingly, the simulation is carried out using the discretized risk-neutral process.

In paper III examine dynamic hedging strategies to control for various risks by utilizing a replicating portfolio. We study hedges in which only a single instrument (i.e., the underlying stock index) is employed, in particular, a partial delta-neutral hedge and a minimum-variance hedge. Delta-neutral hedging is a trading strategy where the number of shares in the replication portfolio is given by

$$N_t^S = \frac{\partial V_t(S_t)}{\partial S_t} \doteq \Delta_t^{(S)} \geq 0,$$

where $V_t(S_t)$ is the value of the derivative at time t. Here it should be noted that the only source of risk arises from S_t. Delta-neutral hedge is employed to the model introduced in Paper II.

Minimum-variance hedging relies on the underlying asset as a single hedging instrument and we follow the work of Bakshi et al. (1997) when deriving the hedge. When the minimum-variance hedge is employed the variance of a hedging error is minimized. This type of hedge can also take into account risks arising from asset volatility and jumps. Hence, we employ the hedge for the model introduced in Paper III. However, even this type of single-instrument hedge can only be partial. Nonetheless, as argued by Ross (1997), such factors as untraded risks, model misspecification or transaction costs make this type of hedge more feasible compared to a perfect delta-neutral hedge.

Summaries of original publications

I. Financial guarantee insurance

Financial guarantee insurance covers a lender from losses due to the default or delinquency of the borrower. It is a country-specific business, since there are differences in laws and regulations (see, e.g., Swiss Re, 2006). In Paper I a specific financial guarantee insurance administrated by the Finnish Centre for Pensions (FCfP) is considered. The special feature of the statutory earnings-related pension scheme in the private sector is that client employers have a legal right to reborrow a specific amount of pension payments. In order to use this right, clients are obliged to take out a guarantee to secure these so-called premium loans.

Losses in financial guarantee insurance may reach catastrophic dimensions for several years when a country experiences an economic depression. During that time the number of claims may be extraordinarily high and, more importantly, the proportion of excessive claims may be much higher than in usual periods. A mild and short downturn in the national economy increases the losses suffered by financial guarantee insurers only moderately, whereas severe downturns are crucial. Hence, financial guarantee insurance is characterized by long periods of low loss activity punctuated by short severe spikes. This indicates that the economic business cycle, and in particular depressions, should be incorporated into the modeling framework.

To model financial guarantee insurance we propose a simple transfer function model where a dichotomic business cycle model is incorporated. The latent business cycle is modeled with a Markov regime-switching model, or more precisely a Hamilton model, where the two states represent the depression period state and its complement state consisting of both boom and mild recession periods. We use the Finnish real GNP to estimate the business cycle. The prediction of claim amounts is obtained by posterior predictive simulation, and based on predictions, the requisite premium and initial risk reserve are determined.

The results in Paper I reveal that an economic depression constitutes a substantial risk to financial guarantee insurance. A guarantee insurance company should incorporate the business cycle covering a depression period in its risk management policy and when adjusting the premium and the risk reserve. According to our analysis the pure premium level based on the gross claim process should be at minimum 2.0%. Moreover, our analysis shows the 95% value at risk for a five-year period to be $2.3 - 2.9$ times the five-year premium. The corresponding 75% value at risk is only $0.17 - 0.29$ times the five-year premium. Thus, a financial guarantee insurer should have a fairly substantial risk reserve in order to get through a long-lasting depression. Moreover, reinsurance contracts are essential in assessing the risk capital needed.

II & III. Equity-linked life insurance contracts

Papers II and III describe in detail how the Bayesian framework is applied to value and hedge an equity-linked life insurance contract. The contract is defined to have fairly general features, in particular, an equity-linked bonus, an interest rate guarantee for the accumulated savings, a downside protection and a surrender (early exercise) option. These properties make the contract a path-dependent American-style derivative which we price in a stochastic, market-consistent framework.

We denote the amount of savings in the insurance contract at time t_i by $A(t_i)$. Its growth during a time interval of length $\delta = t_{i+1} - t_i$ is given by

$$\log \frac{A(t_{i+1})}{A(t_i)} = g\,\delta \; + b \max\left(0, \log \frac{X(t_{i+1})}{X(t_i)} - g\,\delta\right), \tag{3.3}$$

where $X(t_i) = \sum_{j=0}^{q} S(t_{i-j})/(q+1)$ is a moving average of the total return equity index $S(t_i)$. Further, g is the guarantee rate dependent on the riskless short-term interest rate and b is the bonus rate, the proportion of the excessive equity index yield returned to the customer.

The price of an option depends on the assumption of the model describing the behaviour of the underlying instrument. In our framework the price is assumed to depend on an equity index and a riskless short-term interest rate. We assume these to follow a fairly complex stochastic process, and furthermore, the price to depend on the path of the underlying asset. As a closed form solution for the price does not exist, we use Monte Carlo simulation methods (see, e.g., Glasserman, 2004).

In Paper II we utilize the constant elasticity of variance (CEV) model introduced by Cox and Ross (1976) to model the equity index process, and for the stochastic interest rate, we assume the Chan-Karolyi-Longstaff-Sanders (CKLS) model (see Chan et al. (1992)). In Paper III we allow not only the interest rate but also the volatility and jumps in the asset dynamics to be stochastic. For stochastic interest rate and volatility we assume a square-root diffusion referred to as the Cox-Ingersoll-Ross (CIR) model (Cox et al., 1985).

As our contract is a path-dependent American-style option, an optimal exercising strategy needs to be found. It is attained by finding a stopping time maximizing the expected discounted payoff of the option. The decision to continue is based on comparing the immediate exercise value with the corresponding continuation value. We use the regression approach in pricing (see, e.g., Tsitsiklis and Van Roy, 1999), in which the continuation value is expressed as a linear regression of the discounted future value on known functions of the current state. The sample paths needed in the method are simulated using the posterior predictive distribution under risk-neutral dynamics, as suggested by Bunnin et al. (2002).

In Paper II we introduce a method to evaluate a fair bonus rate b such that the risk-neutral price of the contract is equal to the initial investment. The problem of determining b is a kind of inverse prediction problem, and we need to estimate the option values for various values of b. Since we also wish to estimate the variance of the Monte Carlo error related to the regression method, we repeat the estimation several times for fixed values of b. A scatter plot is produced from the values of the bonus rates and the option price estimates, and a third-degree polynomial curve is fitted to the data. Thereafter we solve the bonus rate b for which the option price is equal to 100, which we assume to be the initial amount of savings.

These types of insurance policies involve not only risks arising from financial factors, but also a risk related to mortality. With a stochastic mortality model we do not need to make an assumption of a large insurance portfolio, and we avoid invoking the law of large numbers. This again is significant from the risk management point of view. In Paper II we ignore the risk from mortality and interpret the policy-holder as surviving until the maturity of the contract. The model thus provides an upper bound for insurance liabilities. However, in Paper III stochastic mortality is incorporated in the framework.

In Paper III we also study dynamic hedging strategies to control for various risks by utilizing a replicating portfolio. As a hedging strategy we employ minimum-variance hedging, which relies on the underlying asset as a single hedging instrument. This type of hedge is needed, since a perfect delta-neutral hedge is not feasible due to untraded risks. However, a single-instrument hedge can only be partial, since in our set-up there is more than one source of risk. We also construct a conventional delta-neutral hedge which uses a simpler model for asset dynamics, and compare the performance of the hedges.

The most important finding in Papers II and III is the following. When pricing equity-linked life insurance contracts, the model risk should be incorporated in the insurance company's risk management framework, since the use of an unrealistic model might lead to catastrophic losses. In particular, different model choices imply significant differences in bonus rates. There is a difference in bonus rate estimate even when fixing the index model and using either a stochastic or a fixed interest rate model, but only when the initial interest rate is exceptionally low or high. However, including stochastic mortality has only a slight effect on the estimated bonus rate. Moreover, we assessed the accuracy of the bonus rate estimates. Although the confidence interval of the bonus rate was in some cases fairly long, the spread between the estimate and the lower confidence limit was reasonably small. This is a good result, since the insurance company would probably set the bonus rate close to its lower limit in order to hedge against the liability.

The hedging performances of the minimum-variance hedge and the delta-neutral hedge turned out to be similar. Probably the effect of the imperfectness of single-instrument hedging is vanishingly small compared to other sources of error. Such are, for example, discretization errors and estimation errors of the deltas obtained with the regression method. Our study showed that the most significant factor producing large hedging errors is the duration of the contract. In contrast, the mortality and updating interval of the hedge have only a minor effect on hedging performance.

Our results suggest the following two-step procedure to choose a sensible bonus rate: first, the theoretical fair bonus rate is determined, and second, it is adjusted so that the Value at Risk (VaR) of the hedging error becomes acceptable for the insurance company.

IV. Mortality modeling

Mortality forecasting is a problem of fundamental importance for the insurance and pensions industry, and in recent years stochastic mortality models have become popular. In Paper IV we propose a new Bayesian method for two-dimensional mortality modeling which is based on natural cubic smoothing splines. Compared to other splines approaches, this approach has the advantage that the number of knots and their locations do not need to be optimized. Our method also has the advantage of allowing the cohort data set to be imbalanced, since more recent cohorts yield fewer observations. In our study we used Finnish mortality data for females, provided by the Human Mortality Database.

Let us denote the logarithms of observed death rates as $y_{kt} = \log(m_{kt})$ for ages $k = 1, 2, ..., K$ and cohorts (years of birth) $t = 1, 2, ..., T$. The observed death rates are defined as

$$m_{kt} = \frac{d_{kt}}{e_{kt}},$$

where d_{kt} is the number of deaths and e_{kt} the person years of exposure. In our preliminary set-up we model the observed death rates, while in our final set-up we model the theoretical, unobserved death rates μ_{kt}. Specifically, we assume that

$$d_{kt} \sim \text{Poisson}(\mu_{kt} e_{kt}),$$

where d_{kt} is the number of deaths at age k for cohort t, μ_{kt} is the theoretical death rate (also called intensity of mortality or force of mortality) and e_{kt} is the person years of exposure. This is an approximation, since neither the death rate nor the exposure is constant during any given year.

To smooth and predict logarithms of unobserved death rates, we fit a smooth two-dimensional curve $\theta(k, t)$, and denote its values at discrete points as θ_{kt}. Smoothing is carried out in the dimensions of cohort and age, and the smoothing effect is obtained by giving a suitable prior distribution for $\theta(k, t)$.

We perform a number of model checks and follow the mortality model selection criteria provided by Cairns et al. (2008) to assess the fit and plausibility of our model. The checklist is based on general mortality characteristics and the ability of the model to explain historical patterns of mortality. The Bayesian framework allows us to easily assess parameter and prediction uncertainty using the posterior and posterior predictive distributions, respectively. By introducing two test quantities we may assess the consistency of the model with historical data.

We find that our proposed model meets the mortality model checklist fairly well, and is thus a notable contribution to stochastic mortality modeling. A minor drawback is that we cannot use all available data in estimation but must restrict ourselves to a relevant subset. This is due to the huge matrices involved in computations when many ages and cohorts are included in the data set. However, this problem can be alleviated using sparse matrix computations. Besides, for practical applications the use of "local" data sets should be sufficient.

References

Bakshi, G., Cao, C., Chen, Z., 1997. Empirical performance of alternative option pricing models. *Journal of Finance*, 52, 2003–2049.

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

Bunnin, F.O., Guo, Y., Ren, Y., 2002. Option pricing under model and parameter uncertainty using predictive densities. *Statistics and Computing*, 12, 37–44.

Cairns, A.J.G., 2000. A discussion of parameter and model uncertainty in insurance. *Insurance: Mathematics and Economics*, 27, 313–330.

Cairns, A.J.G., Blake, D., Dowd, K., 2008. Modelling and management of mortality risk: a review. *Scandinavian Actuarial Journal*, 2, 79–113.

Chan, K.C., Karolyi, G.A., Longstaff, F.A., Sanders, A.B., 1992. An empirical comparison of alternative models of the short-term interest rate. *The Journal of Finance*, 47, 1209–1227.

Cox, J.C., Ingersoll, J.E., Ross, S.A., 1985. A theory of term structure of interest rates. *Econometrica*, 53, 385–407.

Cox, J.C., Ross, S.A., 1976. The value of options for alternative stochastic processes. *Journal of Financial Economics*, 3, 145–166.

Daykin, C., Pentikäinen, T., Pesonen, M., 1994. *Practical Risk Theory for Actuaries.* Chapman & Hall.

Draper, D., 1995. Assessment and propagation of model uncertainty (with discussion). *Journal of the Royal Statistical Society Series B*, 57, 45–97.

European Commission, 2009. Directive 2009/138/EC of the European Parliament and of the Council of 25 November 2009 on the taking-up and pursuit of the business of Insurance and Reinsurance (Solvency II).

Gelman, A., Carlin, J.B., Stern, H.S., Rubin, D. B., 2004. *Bayesian Data Analysis.* Chapman & Hall/CRC, Second ed.

Gelman, A. and Rubin, D.B., 1992. Inference from iterative simulation using multiple sequences. *Statistical Science*, 7, 457–511.

Geman, S. and Geman, D., 1984. Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 6, 721–741.

Gilks, W. R., Richardson, S., Spiegelhalter, D. (eds.), 1996. *Markov Chain Monte Carlo in Practice.* Chapman & Hall.

Glasserman, P., 2004. *Monte Carlo Methods in Financial Engineering.* Springer.

Hamilton, J., 1989. A new approach to the economic analysis of nonstationary time series and the business cycle. *Econometrica*, 57, 357–384.

Hardy, M.R., 2002. Bayesian risk management for equity-linked insurance. *Scandinavian Actuarial Journal*, 3, 185–211.

Hastings, W.K., 1970. Monte Carlo sampling methods using Markov chains and their applications. *Biometrika*, 57, 97–109.

Jones, C., 1998. Bayesian estimation of continuous-time finance models. Working paper, Simon School of Business, University of Rochester, New York.

Metropolis, N., Rosenbluth, A.W., Rosenbluth, M.N., Teller, A.H., Teller, E., 1953. Equations of state calculations by fast computing machines. *Journal of Chemical Physics*, 21, 1087–1092.

Morgan Stanley and Oliver Wyman, 2010. Solvency 2: Quantitative & strategic impact, The tide is going out.

Pentikäinen, T., 1975. A model of stochastic-dynamic prognosis: An application of risk theory to business planning. *Scandinavian Actuarial Journal*, 24, 29–53.

Pentikäinen, T. (ed.), 1982. *Solvency of Insurers and Equalization Reserves, Vol I.* The Insurance Publishing Company Ltd, Helsinki.

R Development Core Team, 2009. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org.

Ronkainen, V., Koskinen L., Berglund R., 2007. Topical modelling issues in Solvency II. *Scandinavian Actuarial Journal*, 2, 135-146.

Ross, S., 1997. Hedging long-run commitments: Exercises in incomplete market pricing. *Economic Notes*, 2, 385–420.

Swiss Re, 2006. Credit insurance and surety; solidifying commitments. *Sigma*, 6/2006.

Tsitsiklis, J., and Van Roy, B., 1999. Optimal stopping of Markov processes: Hilbert space theory, approximation algorithms, and an application to pricing high-dimensional financial derivatives. *IEEE Transactions on Automatic Control*, 44, 1840–1851.

Turner, A., Haldane, A., Woolley, P., Wadhwani, S., Goodhart, C., Smithers, A., Large, A., Kay, J., Wolf, M., Boone, P., Johnson, S., Layard, R., 2010. The future of finance: the LSE report. London School of Economics and Political Science.

Insurance: Mathematics and Economics 43 (2008) 245–254

Contents lists available at ScienceDirect

Insurance: Mathematics and Economics

journal homepage: www.elsevier.com/locate/ime

Bayesian modelling of financial guarantee insurance

Anne Puustelli[a], Lasse Koskinen[b,c], Arto Luoma[a,*]

[a] Department of Mathematics and Statistics, Statistics Unit, FIN-33014 University of Tampere, Finland

[b] Insurance Supervisory Authority of Finland, P.O. Box 449, FIN-00101 Helsinki, Finland

[c] Helsinki School of Economics, P.O. Box 1210, FIN-00101 Helsinki, Finland

ARTICLE INFO

Article history:
Received July 2007
Received in revised form
July 2008
Accepted 8 July 2008

PACS:
C11
G22
G32
IM22
IM41

Keywords:
Business cycle
Gibbs sampler
Hamilton model
Risk capital
Surety insurance

ABSTRACT

In this paper we model the claim process of financial guarantee insurance, and predict the pure premium and the required amount of risk capital. The data used are from the financial guarantee system of the Finnish statutory pension scheme. The losses in financial guarantee insurance may be devastating during an economic depression (i.e., deep recession). This indicates that the economic business cycle, and in particular depressions, must be taken into account in modelling the claim amounts in financial guarantee insurance. A Markov regime-switching model is used to predict the frequency and severity of future depression periods. The claim amounts are predicted using a transfer function model where the predicted growth rate of the real GNP is an explanatory variable. The pure premium and initial risk reserve are evaluated on the basis of the predictive distribution of claim amounts. Bayesian methods are applied throughout the modelling process. For example, estimation is based on posterior simulation with the Gibbs sampler, and model adequacy is assessed by posterior predictive checking. Simulation results show that the required amount of risk capital is high, even though depressions are an infrequent phenomenon.

1. Introduction

A guarantee insurance (surety insurance) is typically required when there is doubt as to the fulfilment of a contractual, legal or regulatory obligation. It is designed to protect some public or private interest from the consequences of a default or delinquency of another party. Financial guarantee insurance covers losses from specific financial transactions. Due to differences in laws and regulations guarantee insurance is a country-specific business (see, for example, Sigma (2006)).

When a country experiences an economic depression (that is, deep recession), losses in financial guarantee insurance may reach catastrophic dimensions for several years. During that time the number of claims may be extraordinarily high and, more importantly, the proportion of excessive claims may be much higher than in usual periods (see, for example, Romppainen (1996) and Sigma (2006)). As the future growth of the economy is uncertain, it is important to consider the level of uncertainty one can expect in the future claim process. A mild and short downturn in the national economy increases the losses suffered by financial guarantee insurers only moderately, whereas severe downturns are crucial. History knows several economic depressions. These include the Great Depression in the 1930s, World Wars I and II, and the oil crisis in the 1970s. More recently, the Finnish experience from the beginning of the 1990s and the Asian crisis in the late 1990s are good examples. An interesting statistical approach in analyzing the timing and effects of the Great Depression is the regime switching method presented in Coe (2002).

There is no single "best practice" model for credit risk capital assessment (Alexander, 2005). The main approaches are structural firm-value models, the option-theoretical approach, rating-based methods, macroeconomic models and actuarial loss models. In contrast to market risk, there has been little detailed analysis of the empirical merits of the various models. A review of commonly used financial mathematical methods can be found for example in McNeil et al. (2005). Since we here study special guarantee loans, so-called premium loans, which are not traded in Finland, we adopt an actuarial approach. More specifically, we model the claim process of financial guarantee insurance in the economic business cycle context. However, this approach is also demanding, since a depression is a particularly exceptional event.

* Corresponding author. Tel.: +358 3 35516428; fax: +358 3 35516157.
E-mail addresses: anne.puustelli@uta.fi (A. Puustelli), lasse.koskinen@vakuutusvalvonta.fi (L. Koskinen), arto.luoma@uta.fi (A. Luoma).

0167-6687/$ – see front matter © 2008 Elsevier B.V. All rights reserved.
doi:10.1016/j.insmatheco.2008.07.001

We build on the following three studies of the financial guarantee system of the Finnish pension scheme. Rantala and Hietikko (1988) modelled solvency issues by means of linear models, their main objective being to test methods for specifying bounds for the solvency capital. The linear method combined with data not containing any fatal depression period – Finland's depression in the early 1990s struck after the article was published – underestimated the risk. Romppainen (1996) analyzed the structure of the claim process during the depression period. Koskinen and Pukkila (2002) also applied the economic cycle model. Their simple model gives approximate results but lacks sound statistical grounding. We use modern statistical methods which offer advantages for assessing uncertainty.

From the methodological point of view, we adopt the Bayesian approach, recommended for example by Scollnik (2001). Simplified models or simplified assumptions may fail to reveal the true magnitude of the risks the insurer faces. While undue complexity is generally undesirable, there may be situations where complexity cannot be avoided. Best et al. (1996) explain how Bayesian analysis can generally be used for realistically complex models. An example of concrete modelling is provided by Hardy (2002), who applies Bayesian techniques to a regime-switching model of the stock price process for risk management purposes. Another example can be found in Smith and Goodman (2000), who present models for the extreme values of large claims and use modern techniques of Bayesian inference. Here, Bayesian methods are used throughout the modelling process. For example, estimation is based on posterior simulation with the Gibbs sampler, and model adequacy is assessed by posterior predictive checking. The proposed actuarial model is used for simulation purposes in order to study the effect of the economic cycle on the requisite pure premium and initial risk reserve.

We apply the Markov regime-switching model to predict the frequency and severity of depression periods in the future. Prediction of claim amounts is made using a transfer function model where the predicted growth rate of the real GNP is an explanatory variable. More specifically, we utilize the business cycle model introduced by Hamilton (1989). In this method, all the dating decisions or, more correctly, the probabilities that particular time periods will be recession periods, are based on observed data. The method assumes that there are two distinct states (regimes) in the business cycle – one for expansion and one for recession – which are governed by a Markov chain. The stochastic nature of GNP growth depends on the prevailing state.

Financial guarantee insurance is characterized by long periods of low loss activity, punctuated by short severe spikes (see Sigma (2006)). As such, conventional dichotomic business models are inadequate, since severe recessions constitute the real risk. We propose a model where the two states represent (1) the depression period state and (2) its complement state consisting of both boom and mild recession periods. We use Finnish real GNP data to estimate our model. The claim data are from the financial guarantee insurance system of the Finnish pension scheme. Combining a suitable business cycle model with a transfer function model provides a new way to analyze the solvency of a financial guarantee provider with respect to claim risk.

The paper is arranged as follows. In Section 2 the Finnish credit crisis in the 1990s is described. Section 3 introduces the business cycle model and Section 4 presents the transfer function model and predictions. Model checks are presented in Section 5. Section 6 concludes.

2. The Finnish experience in the 1990s

During the years 1991–1993 Finland's GNP dropped by 12%. The period was inevitably harmful to all sectors of the economy and society as a whole. However, the injuries suffered in the insurance sector were only moderate, at least compared with the problems of the banking sector at the same time. An important exception was financial guarantee insurance related to the statutory earnings-related pension scheme in the private sector. At a general level, Norberg (2006) describes the risk for pension schemes under economic and demographic developments.

The administration of the statutory earnings-related pension scheme of the private sector is decentralized to numerous insurance companies, company pension funds and industry-wide pension funds. The central body for the pension scheme is the Finnish Centre for Pensions (FCfP). The special feature of the pension scheme is that client employers have a legal right to reborrow a specific amount of pension payments. The loans are called premium loans. In order to use this right, clients are obliged to take a guarantee to secure the loans. FCfP administrated a special financial guarantee insurance for this purpose. Competitive alternatives were the requirement of safe collateral, guarantee insurance from another insurance company, or a bank guarantee. The claim event was a failure of the borrowing employer to fulfil his commitment. A more detailed description of the case of the FCfP can be found in Romppainen (1996).

The business was initiated in 1962, and continued successfully until Finland was hit by depression in the 1990s. The consequent losses reached catastrophic dimensions, and the financial guarantee insurance activity of the FCfP was closed. Claims paid by the FCfP are shown in Fig. 1. As may also be seen from this figure, the claim recoveries after the realization process of collaterals has typically been about 50%. The cost losses were levied on all employers involved in the mandatory scheme and hence, pension benefits were not jeopardized. Subsequently the FCfP's run-off portfolio was transferred to a new company named "Garantia".

In order to promote the capital supply, the FCfP was under a legal obligation to grant financial guarantee insurance to client employers. It therefore employed fairly liberal risk selection and tariffs, which probably had an influence on the magnitude of the losses. Hence, the data reported by Romppainen (1996) and used here cannot be expected, as such, to be applicable in other environments. The risks would be smaller in conventional financial guarantee insurance, which operates solely on a commercial basis.

It is interesting to note that there are also, at present, similar problems in the USA. The corresponding US institute is the Pension Benefit Guaranty Corporation (PBGC), a federal corporation created by the Employee Retirement Income Security Act of 1974. It currently protects the pensions of nearly 44 million American workers and retirees in 30,330 private single-employer and multiemployer defined benefit pension plans. The Pension Insurance Data Book (2005) (page 31) reveals that total claims on the PBGC have increased rapidly from about 100 million dollars in 2000 to 10.8 billion dollars in 2005. This increase cannot be explained by nation-wide depression, but may be related to the problems of special industry sectors (for example aviation).

3. National economic business cycle model

Our first goal is to find a model by which we can forecast the growth rate of the GNP. We will use annual Finnish real GNP data from 1860 to 2004, provided by Statistics Finland. We are particularly interested in the frequency and severity of depression periods. For this purpose we will utilize the Markov regime-switching model introduced by Hamilton (1989). The original Hamilton model envisages two states for the business cycle: expansion and recession. In our situation, however, it is more important to detect depression, since this is the phase when financial guarantee insurance will suffer its most severe losses. We will therefore define the states in a slightly different

I

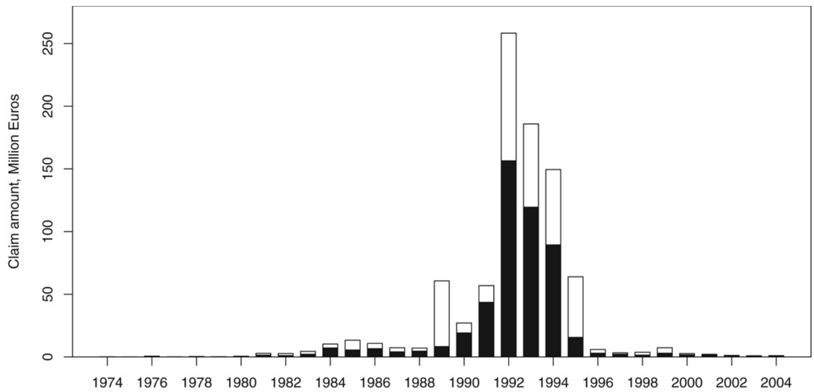


Fig. 1. The total amount of claims paid from the financial guarantee insurance by the Finnish Centre for Pensions between 1974 and 2004. The lower dark part of the bar describes the final loss after recoveries from collaterals and reinsurance by December 2006.

way in our application. Specifically, we use a two-state regime-switching model in which the first state covers both expansion and recession periods and the second depression. Our estimation results correspond to this new definition, since depression periods are included in our data set. By contrast, Hamilton used quarterly US data from 1951 to 1984, which do not include years of depression.

The Hamilton model may be expressed as $y_t = \alpha_0 + \alpha_1 s_t + z_t$, where y_t denotes the growth rate of the real GNP at time t, s_t the state of the economy and z_t a zero-mean stationary random process, independent of s_t. The parameters α_0 and α_1 and the state s_t are unobservable and must be estimated. A simple model for z_t is an autoregressive process of order r, denoted by $z_t \sim \mathrm{AR}(r)$. It is defined by the equation $z_t = \phi_1 z_{t-1} + \phi_2 z_{t-2} + \cdots + \phi_r z_{t-r} + \epsilon_t$, where $\epsilon_t \sim \mathrm{N}(0, \sigma_\epsilon^2)$ is an i.i.d. Gaussian error process. After some initial analysis, we found that AR(2) was sufficient to capture the autocorrelation of z_t, and we therefore used it in estimation. The growth rate at time t is calculated as $y_t = \log(\mathrm{GNP}_t) - \log(\mathrm{GNP}_{t-1})$.

We define the state variable s_t to be 0, when the economy is in expansion or recession, and 1, when it is in depression. The transitions between the states are controlled by the first-order Markov process with transition probabilities

$$\Pr(s_{t+1} = 0|s_t = 0) = p,$$
$$\Pr(s_{t+1} = 1|s_t = 0) = 1 - p,$$
$$\Pr(s_{t+1} = 0|s_t = 1) = 1 - q,$$
$$\Pr(s_{t+1} = 1|s_t = 1) = q.$$

Thus, the transition matrix is given by

$$\mathbf{P} = \begin{pmatrix} p & 1-p \\ 1-q & q \end{pmatrix}.$$

The stationary probabilities $\boldsymbol{\pi} = (\pi_0, \pi)'$ of the Markov chain satisfy the equations $\boldsymbol{\pi}'\mathbf{P} = \boldsymbol{\pi}'$ and $\boldsymbol{\pi}'\mathbf{1} = 1$, where $\mathbf{1} = (1, 1)'$.

The Hamilton model was originally estimated by maximizing the marginal likelihood of the observed data series y_t. The probabilities of the states were then calculated conditional on these maximum likelihood estimates. The numerical evaluation was made by a kind of nonlinear version of the Kalman filter. By contrast, we use Bayesian computation techniques throughout, the advantage being that we need not rely on asymptotic inference, and the inference on the state variables is not conditional on parameter estimates. The Hamilton model will be estimated using the Gibbs sampler introduced by Geman and Geman (1984) in the context of image restoration. Examples of Gibbs sampling can be found in Gelfand et al. (1990) and Gelman et al. (2004). Carlin et al. (1992) provide a general approach to its use in nonlinear state-space modelling.

Gibbs sampling, also called alternating conditional sampling, is a useful algorithm for simulating multivariate distributions, for which the full conditional distributions are known. Let us assume that we wish to simulate the random vector $\boldsymbol{\theta} = (\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \ldots, \boldsymbol{\theta}_p)$ whose subvectors $\boldsymbol{\theta}_i$ have known conditional distributions $p(\boldsymbol{\theta}_i|\boldsymbol{\theta}_{(-i)})$, where $\boldsymbol{\theta}_{(-i)} = (\boldsymbol{\theta}_1, \ldots \boldsymbol{\theta}_{i-1}, \boldsymbol{\theta}_{i+1}, \ldots \boldsymbol{\theta}_p)$. In each iteration the Gibbs sampler goes through $\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \ldots, \boldsymbol{\theta}_p$ and draws values from their conditional distributions $p(\boldsymbol{\theta}_i|\boldsymbol{\theta}_{(-i)})$ where the conditioning subvectors have been set at their most recently simulated values. It can be shown that this algorithm produces an ergodic Markov chain whose stationary distribution is the desired target distribution of $\boldsymbol{\theta}$. In Bayesian inference one can use the Gibbs sampler to simulate the posterior distribution, if one is able to generate random numbers or vectors from all the full conditional posterior distributions.

To simplify some of the expressions, we will use the following notation: $\mathbf{y} = (y_1, y_2, \ldots, y_T)'$, $\mathbf{s} = (s_1, s_2, \ldots, s_T)'$ and $\mathbf{z}_{t-1} = (z_{t-1}, z_{t-2}, \ldots, z_{t-r})'$. Furthermore, we denote the vector of autoregressive coefficients by $\boldsymbol{\phi} = (\phi_1, \phi_2, \ldots, \phi_r)'$ and the vector of all parameters by $\boldsymbol{\eta} = (\alpha_0, \alpha_1, \boldsymbol{\phi}', \sigma_\epsilon^2, p, q)'$. Using these notations the density of $\mathbf{y}$, conditional on $\mathbf{s}$ and the parameters $\boldsymbol{\eta}$, can be written as

$$p(\mathbf{y}|\mathbf{s}, \boldsymbol{\eta}) = \prod_{t=1}^{T} \frac{1}{\sqrt{2\pi\sigma_\epsilon^2}} \exp\left(-\frac{1}{2\sigma_\epsilon^2}(y_t - \alpha_0 - \alpha_1 s_t - \boldsymbol{\phi}'\mathbf{z}_{t-1})^2\right).$$

In order to facilitate computations, we assume that the prior distributions are independent, and are given as follows:

$$p \sim \mathrm{Beta}(\alpha_p, \beta_p),$$
$$q \sim \mathrm{Beta}(\alpha_q, \beta_q),$$
$$p(\boldsymbol{\phi}) \propto 1,$$
$$p(\alpha_0) \propto 1,$$
$$p(\sigma_\epsilon^2) \propto \frac{1}{\sigma_\epsilon^2},$$
$$p(\alpha_1) \propto \mathrm{N}(\alpha_1|\mu_0, \sigma_0^2) \times I(\alpha_1 < -0.03).$$

We obtained noninformative prior distributions for p and q by specifying as prior parameters $\alpha_p = \beta_p = \alpha_q = \beta_q = 0.5$. These values correspond to the Jeffreys uninformative prior distribution in the standard Bernoulli model. We also carried out a sensitivity analysis, using informative priors with parameters $\alpha_p = 19$, $\beta_p = 3$; $\alpha_q = \beta_q = 11$. These values correspond to the idea that the chain has switched to state 1 in 2 out of 20 prior cases when it has been in state 0 and has switched to state 0 in 10 out of 20 prior cases when it has been in state 1. By this choice of priors we could increase the probability of state 1, so that it would correspond to our concept of depression.

For α_0, $\boldsymbol{\phi}$ and σ_ϵ^2 we gave improper, noninformative prior distributions. The prior of σ_ϵ^2 is equivalent to giving a uniform improper prior for $\log(\sigma_\epsilon^2)$ and is a common choice for positive-valued parameters. The prior distribution of α_1 prevents it from obtaining a positive value (that is, the state can then be interpreted as a depression state). Here, the notation $\mathrm{N}(\alpha_1|\mu_0, \sigma_0^2)$ refers to the Gaussian density with mean μ_0 and variance σ_0^2 and $I(\alpha_1 < -0.03)$ the indicator function obtaining the value 1, if $\alpha_1 < -0.03$, and 0, otherwise. The cut-off point -0.03 was chosen to draw a clear distinction between the states of the model and also to speed up posterior simulation. If the difference was allowed to be smaller, the iteration process was substantially slowed. We specified the values $\mu_0 = -0.1$, $\sigma_0^2 = 0.2^2$ as prior parameters, which results in a fairly noninformative prior distribution. We also experimented here with an informative alternative $\mu_0 = -0.05$, $\sigma_0^2 = 0.025^2$, which reduces the difference between the states and increases the probability of state 1. The results were similar to those obtained when informative priors were given to p and q.

All full conditional posterior distributions are needed to implement the Gibbs sampler. They can be found in Appendix A. The computations were performed and figures produced using the R computing environment (see http://www.r-project.org). The functions and data needed to replicate the results of this article can be found at http://mtl.uta.fi/codes/guarantee.

In Fig. 2, one simulated chain produced by the Gibbs sampler is shown. As will be seen, the chain converges rapidly to its stationary distribution, and the component series of the chain mix well, that is, they are not excessively autocorrelated. The summary of the estimation results, based on three simulated chains, as well as Gelman and Rubin's diagnostics (Gelman et al., 2004) are given in Appendix B. The values of the diagnostic are close to 1 and thus indicate good convergence.

4. Prediction of claim amounts

Our ultimate goal is to predict the pure premium and the required amount of risk capital needed for the claim deviation. The claim data were obtained from FCfP (see Section 2) and the years included in this study are 1966–2004. The claim amounts are predicted using the following simple regression model:

$$x_t = \beta_0 + \beta_1 x_{t-1} + \beta_2 y_t + u_t, \tag{1}$$

where $x_t = G^{-1}(x_t^*)$, x_t^* is the proportion of gross claim amount to technical provision at time t, G^{-1} is some strictly increasing and continuously differentiable link function transforming the open unit interval $(0, 1)$ to $(-\infty, \infty)$, y_t is the growth rate of the GNP and $u_t \sim \mathrm{N}(0, \sigma_u^2)$ is an i.i.d. Gaussian error process, independent of y_t. Note that G^{-1} can be interpreted as an inverse of some distribution function. The parameters β_0, β_1, β_2 and σ_u^2 are unknown and are estimated. Note that the assumptions on u_t imply that y_t is an exogenous process in the regression model, and, consequently, this model and the Hamilton model can be estimated separately. Moreover, the posterior predictive distribution of x_t, $t = T+1, T+2, \ldots$, can be easily simulated using the posterior simulations of β_0, β_1, β_2 and σ_u^2 and the posterior predictive simulations of y_t.

The model $x_t = \beta_0 + \beta_1 x_{t-1} + \beta_2 y_t + u_t$ may also be expressed in the form

$$\begin{aligned} x_t &= \frac{\beta_0}{1-\beta_1} + \frac{\beta_2}{1-\beta_1 B} y_t + \frac{1}{1-\beta_1 B} u_t \\ &= \frac{\beta_0}{1-\beta_1} + \beta_2 (y_t + \beta_1 y_{t-1} + \beta_1^2 y_{t-2} + \cdots) \\ &\quad + u_t + \beta_1 u_{t-1} + \beta_1^2 u_{t-2} + \cdots, \end{aligned}$$

from which one can see that it is a transfer function model (also called a dynamic regression model). We have not done general transfer function modelling here, but in more complicated situations it might be appropriate. A useful reference in this context is Pankratz (1991).

Since the density of x_t is normal, the density of $\mathbf{x}^* = (x_1^*, x_2^*, \ldots, x_T^*)$, conditional on $\mathbf{y} = (y_1, y_2, \ldots, y_T)$ and the parameters, is of the form

$$\begin{aligned} p(\mathbf{x}^*|\mathbf{y}, \boldsymbol{\beta}, \sigma_u^2) &= \prod_{t=1}^{T} \frac{1}{g(x_t)} \frac{1}{\sqrt{2\pi\sigma^2}} \\ &\quad \times \exp\left(-\frac{1}{2\sigma_u^2}(x_t - \beta_0 - \beta_1 x_{t-1} - \beta_2 y_t)^2\right), \end{aligned}$$

where $g(x) = \frac{\mathrm{d}}{\mathrm{d}x} G(x)$.

In the following, we will consider the cases where G is the distribution function of the standard normal distribution (probit link), the standard logistic distribution (logit link) or Student's t-distribution with ν degrees of freedom (t link). When using the t link we did not specify the degrees of freedom parameter ν, but estimated it from the data. We used two alternative prior distributions for ν. For other parameters we used uninformative prior distributions except that we restricted β_1 to be less than 1 in order to ensure that the estimated model for x_t is stationary. The prior distributions are as follows:

$$p_1(\nu) = \frac{1}{200}, \quad \nu = 1, 2, \ldots, 200,$$

$$p_2(\nu) \propto \frac{1}{\nu}, \quad \nu = 1, 2, \ldots, 200,$$

$$p(\boldsymbol{\beta}, \sigma_u^2|\nu) \propto \frac{1}{\sigma_u^2} \times I(\beta_1 < 1),$$

where $I(\beta_1 < 1)$ is an indicator function obtaining the value 1 when $\beta_1 < 1$, and 0 otherwise. The conditional posterior distributions needed in Gibbs sampling are presented in Appendix C.

For comparison, we also estimated the models using probit and logit links. The probit link is defined as $x_t = \Phi^{-1}(x_t^*)$, where $\Phi^{-1}(.)$ denotes the inverse function of the standard normal distribution function. The probit link can be regarded as a limiting case of the t link, as $\nu \to \infty$. The logit link is defined as $x_t = \mathrm{logit}(x_t^*) = \log(x_t^*/(1 - x_t^*))$. It does not correspond exactly to the t link with any ν, but in the value range of the original data set it approximates the t link with $\nu = 14$.

The premium P and the initial risk reserve are evaluated from the posterior predictive distribution of the proportions of claim amount to technical provision. For simplicity, the technical provision is set at 1 in the prediction. The predicted proportions x_t^* are then the same as the predicted claim amounts. We denote the predicted claim amount at time t on sample path i by x_{it}^*. The premium is set at an overall mean of x_{it}^* over all iterations $i = 1, 2, \ldots, n$ and predicted time periods $t = T+1, T+2, \ldots, T+h$, that is, $P = \sum_{i=1}^{n} \sum_{t=T+1}^{T+h} x_{it}^*/(nh)$. The balance at time t on sample path i is given by $b_{it} = P(t - T) - \sum_{j=T+1}^{t} x_{ij}^*$. For all simulations i we evaluate the minimum balance $b_i^{\min} = \min_t b_{it}$. These values

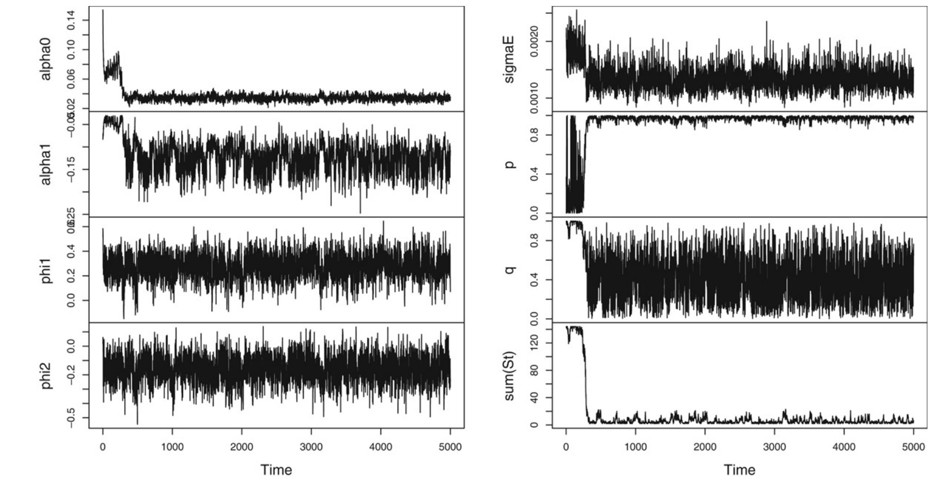


Fig. 2. Iterations of the Gibbs sampler.

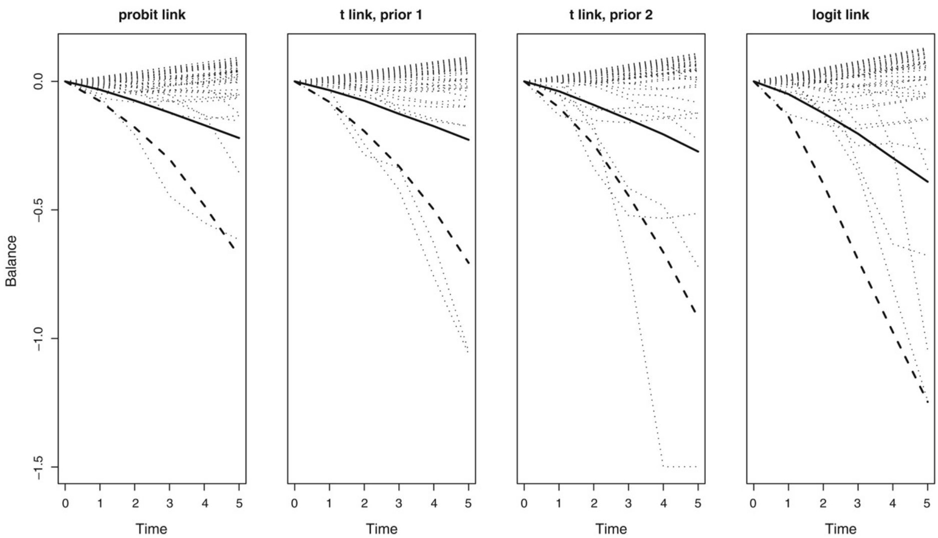


Fig. 3. Simulation results for the balance of guarantee insurance. The solid and dashed lines indicate the 95% and 99% values at risk, respectively, and the dotted lines 50 example paths. The simulation results based on the probit, t (with two different prior distributions) and logit links are shown.

constitute the simulated minimum balance distribution, which is used to evaluate 95% and 75% values at risk and to predict the required amount of risk capital.

The distribution of the minimum balance is extremely skewed, which can be explained by the rareness of depression and by the huge losses incurred for guarantee insurance once depression hits. This phenomenon can be seen from Fig. 3, which shows the 95% and 99% values at risk (VaR) evaluated from the five-year balance prediction for all the link functions. Noninformative prior distributions were used in estimating the Hamilton model. The solid and dashed lines indicate 95% and 99% VaRs, respectively. The curves indicating the 95% VaR differ from each other only moderately, while there is a major difference in the 99% VaR. The logit link gives the steepest slope, the probit link the most gentle, and the t link with the two different priors something between these two. These differences can be explained by the curvature of the distribution functions related to the links. These distributions differ considerably in the left tail area, where the original observations are being mapped.

Extensive simulations (10 000 000 iterations) were carried out to evaluate the pure premium and the 95% and 75% VaRs for the prediction period of five years. The results are presented in Table 1. The pure premium level ranges from 2.0% to 2.8%. The 95% VaR ranges from 2.3 to 2.9 times the five-year premium and the 75% VaR from 0.17 to 0.29 times the five-year premium. These results were obtained when noninformative prior distributions

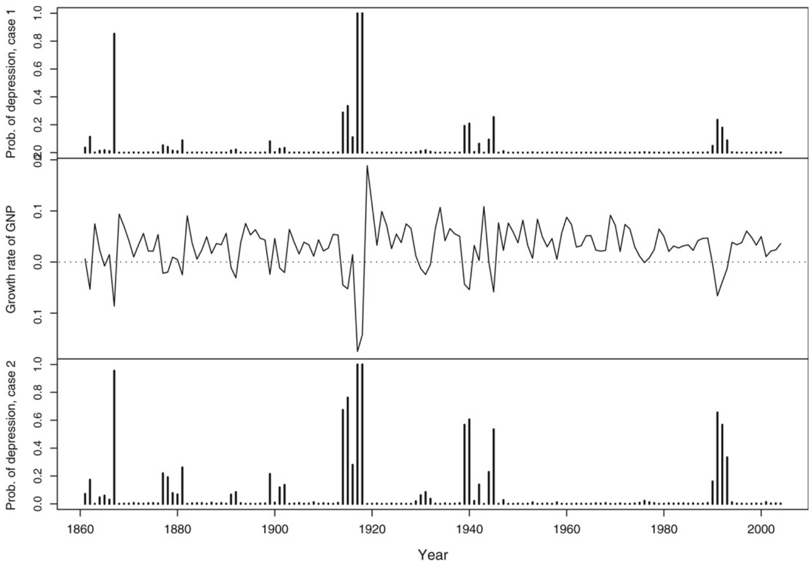


Fig. 4. The growth rate of the GNP and the probabilities of depression, estimated using two different kinds of prior information. The uppermost part of the figure corresponds to the case where noninformative prior distributions are used for all parameters and the undermost part that where informative prior distributions are used for p and q.

Table 1
Simulation results of the premium P, the 95% and 75% values at risk proportional to five-year premium and the average discrepancy for a transfer function model with different links

	P	95% VaR	75% VaR	$\hat{D}_{avg}(x^*)$
Probit link	1.96	2.28	0.288	−336.43
t link, prior 1	2.05	2.36	0.272	−335.98
t link, prior 2	2.17	2.45	0.250	−335.33
Logit link	2.76	2.91	0.173	−331.74

were used in estimating the Hamilton model. When informative prior distributions were used, the results did not substantially change.

5. Model checks for the Hamilton model and the transfer function model

We made some sensitivity analyses with respect to the prior distributions related to the Hamilton model. We found that by using informative prior distributions for the transition probabilities p and q we could increase the estimated probabilities of state 1, so as to correspond better to its interpretation as depression. This can be seen from Fig. 4, where the growth rate of the GNP is shown along with the probabilities of depression, estimated using two different kinds of prior information. The same goal was achieved by giving an informative prior for α_1. However, these adjustments did not markedly affect the estimated premium or values at risk. According to the posterior predictive checks, the model with noninformative prior distributions appeared to be somewhat better. However, the model with informative priors was also sufficiently good.

The residuals of the Hamilton model appeared to be normally or nearly normally distributed. In fact, our data set had one positive outlier which caused rejection of a normality test. This was due to the fact that the model does not include a regime for the strong boom periods of the economy. However, it is not necessary to make the model more complicated by introducing a third regime, since positive outliers are extremely rare and it was sufficient for our purpose to model the depression periods.

The fit of a model can be checked by producing replicated data sets by means of posterior predictive simulation. A replicated data set is produced by first generating the unknown parameters (and in the case of the Hamilton model also the states) from their posterior distribution and then, given these parameters, the new data values. One can simulate distributions of arbitrary test statistics under the checked model by calculating the test statistics from each replicated data set. Then one can compare these distributions with the statistics of the original data set. This approach to model checking is well explained in Chapter 6 of Gelman et al. (2004).

We generated 5000 replicates of the GNP data under the Hamilton model, and from them calculated some basic statistics. The resulting distributions were consistent with the observed statistics, as can be seen from Fig. 5. Only the maximum value of the original data set is extreme with respect to its simulated posterior distribution. This value is nonetheless plausible under the simulated model, that is, the Hamilton model. We also made a similar test for the simpler linear AR(2) model by producing 5000 replicates. The resulting distributions, as seen in Fig. 6, were not as consistent with the observed statistics as they were in the case of the Hamilton model. Specifically, the observed mean, skewness and kurtosis were more extreme than would be expected under a good model.

The discrepancy between the data and the model may be measured using several criteria (see Gelman et al. (2004)). We used the average discrepancy, defined as $D_{avg}(y) = \mathrm{E}(D(y,\theta)|y)$, the posterior mean of the deviance $D(y,\theta) = -2\log p(y|\theta)$. A smaller value of this criterion indicates a better model fit. The average discrepancy is estimated as $\hat{D}_{avg}(y) = \sum_{l=1}^{L} D(y,\theta_l)/L$, where the vectors θ_l are posterior simulations. The estimated average discrepancy for the Hamilton model with our noinformative prior distribution was $\hat{D}_{avg}(y) = -539.02$ and with our informative prior distribution $\hat{D}_{avg}(y) = -549.56$. The criterion value for the

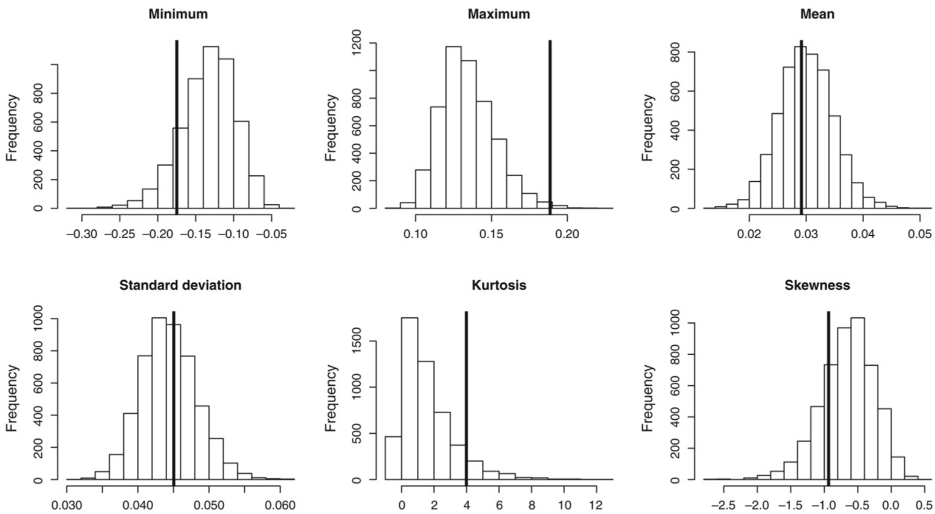


Fig. 5. Replication check for the Hamilton model with noninformative prior distributions.

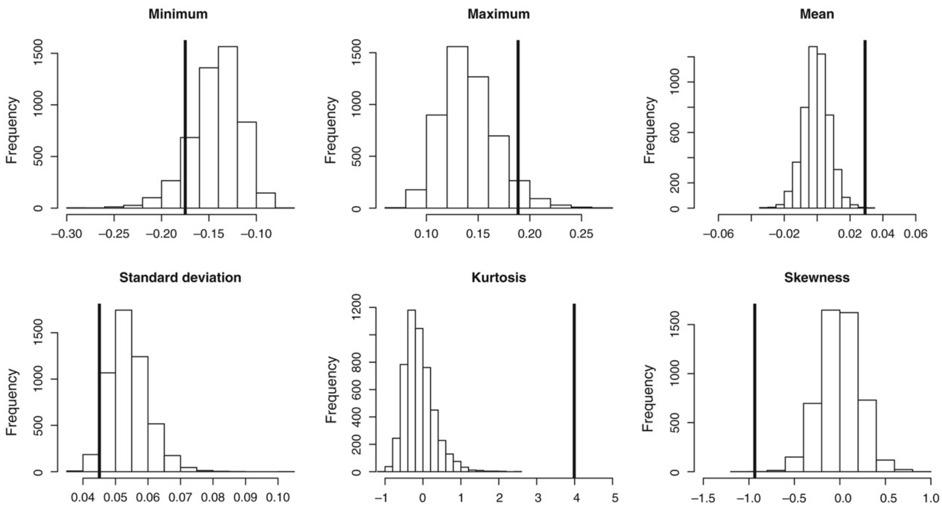


Fig. 6. Replication check for the AR(2) model.

AR(2) model was $\hat{D}_{\text{avg}}(y) = -274.19$, indicating that its model fit was considerably inferior to that of the Hamilton model.

We also made robustness checks using subsample data in estimation. The results did not markedly change when only the first half of the data set (years 1861–1932) was used. When the second half (years 1933–2004) was used the difference between the regimes became smaller and the probability of a depression regime increased. This is natural, since the second half does not contain the years when the GNP showed extreme drops, that is, the years 1867 (one of the great hunger years in Finland) and 1917-18 (when Finland became independent and had the Civil War).

We also made checks for our transfer function model, used in estimating the risk premium and the initial risk reserve. The predictive distributions of the basic statistics were consistent with their observed values with all link functions. The residuals, obtained after fitting the transformed data sets, appeared to be normally or nearly normally distributed. The average discrepancy of all models is presented in Table 1. It can be seen that with the probit link the model fit is best and with the logit link poorest. However, the difference between the probit link model and the t link models is small, and it might be advisable to use one of the t link models, since there is model uncertainty involved in the choice of the link, and it might be safer to average over several alternatives.

The observed proportions of claim amount to technical provision and the lines indicating predictive intervals are shown in Fig. 7.

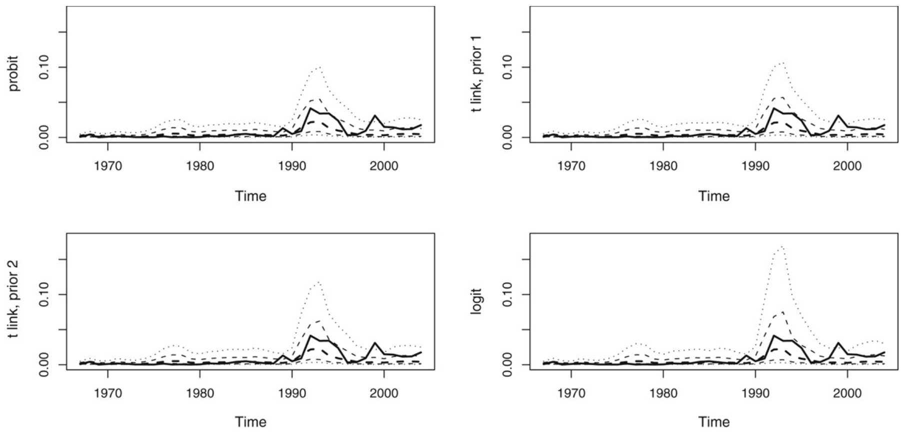


Fig. 7. Replication check for the transfer function models. The solid line indicates the observed data series, and the thick dashed line the medium of 5000 replications. At each time point, 50% of the simulated values lie between the thin dashed lines and 75% between the thin dotted lines.

These lines are based on 5000 replicated series. In all figures, the solid line is the observed data series, and the thick dashed line the medium of the predictive distribution. The thin dashed lines indicate the 50% predictive intervals, and the thin dotted lines the 75% intervals. On the basis of visual inspection, the median line and the 50% interval lines are almost identical in all models. The 75% predictive intervals differ from each other significantly during the depression period, the logit link being the extreme case. If the 90% lines were drawn, the upper line would go far beyond the range of the figure in the case of the logit link, which confirms our earlier observation that the logit link produces extreme simulation paths. This phenomenon was already noted at the end of Section 4. If the estimation period were longer, the variation between the models would probably be smaller.

A standard approach would be to use a compound Poisson process to model the numbers of claims and their sizes simultaneously. However, we found such an approach difficult, since both the claim size distribution and the intensity of claims turned out to be highly variable during our short estimation period.

6. Conclusions

In this paper we present an application of Bayesian modelling to financial guarantee insurance. Our goal was to model the claim process and to predict the premium and the required amount of risk capital needed for claim deviation. Even though the data used are from the Finnish economy and from the financial guarantee system of the Finnish statutory pension scheme, we would consider the model applicable in similar cases elsewhere. However, for the interpretation of the results it is important to note that the risks are probably smaller in conventional companies, which operate solely on a commercial basis, than in a statutory system.

The Markov regime-switching model was used to predict the frequency and severity of depressions in the future. We used real GNP data to measure economic growth. The claim amounts were predicted using a transfer function model where the predicted real GNP growth rate was an explanatory variable. We had no notable convergence problems when simulating the joint posterior distribution of the parameters, even though the prior distributions were noninformative or only mildly informative. The sensitivity to choice of link function (probit, logit and t link) in the context of the transfer function was much greater than that to the prior assumptions (informative or noninformative) in the growth rate model.

The simulation results can be summarized as follows. First, if the effects of economic depressions are not properly considered, there is a danger that the premiums of financial guarantee insurance will be set too low. The pure premium level based on the gross claim process is assessed to be at minimum 2.0% (range 2.0%–2.8%). Second, in order to get through a long-lasting depression, a financial insurer should have a fairly substantial risk reserve. The 95% value at risk for a five-year period is 2.3–2.9 times the five-year premium. The corresponding 75% value at risk is only 0.17–0.29 times the five-year premium. These figures illustrate the vital importance of reinsurance contracts in assessing the risk capital needed.

Some general observations may be made on the basis of this study:

- In order to understand the effects of business cycles on guarantee insurers' financial condition and better appreciate the risks, it is appropriate to extend the modelling horizon to cover a depression period;
- A guarantee insurance company may benefit from incorporating responses to credit cycle movements into its risk management policy;
- The use of Bayesian methods offers significant advantages for assessment of uncertainty;
- The present findings underline the observation that a niche insurance company may need special features (for example a transfer function model instead of the Poisson process approach) in its internal model when a specific product (for example pension guarantee insurance) is modelled.

We assume that the proposed method can also be applied to the financial guarantee and credit risks assessment of a narrow business sector whenever a suitable credit cycle model for the sector is found.

Acknowledgements

The authors of this study are grateful to Yrjö Romppainen, now deceased, for the data he provided us, and for most valuable comments and inspiring discussions. We would also like to thank Vesa Ronkainen and an anonymous referee for their useful comments.

Appendix A. The full conditional posterior distributions of the Hamilton model

We will use the following notation to simplify some of the expressions: $\mathbf{y} = (y_1, y_2, \ldots, y_T)'$, $\mathbf{s} = (s_1, s_2, \ldots, s_T)'$, $\mathbf{s}_{(-t)} = (s_1, s_2, \ldots, s_{t-1}, s_{t+1}, \ldots, s_T)'$ and $\mathbf{z} = (z_1, z_2, \ldots, z_T)'$. We will also need a matrix

$$\mathbf{Z} = \begin{pmatrix} \mathbf{z}_0' \\ \mathbf{z}_1' \\ \vdots \\ \mathbf{z}_{T-1}' \end{pmatrix},$$

whose rows are of the form $\mathbf{z_t} = (z_t, z_{t-1}, \ldots, z_{t-r+1})'$. Furthermore, we denote the vector of autoregressive coefficients by $\boldsymbol{\phi} = (\phi_1, \phi_2, \ldots, \phi_r)'$ and the vector of all parameters by $\boldsymbol{\eta} = (\alpha_0, \alpha_1, \boldsymbol{\phi}', \sigma_\epsilon^2, p, q)'$. In the following treatment we assume the pre-sample values $\mathbf{y}_0 = (y_0, \ldots, y_{1-r})'$ and $\mathbf{s}_0 = (s_0, \ldots, s_{1-r})'$ to be known. In fact, $\mathbf{s}_0$ is not known, but we will simulate its components in a similar way to that used for $\mathbf{s}$.

The full conditional posterior distributions of the Hamilton model are as follows:

$$\{p|\mathbf{s}\} \sim \text{Beta}\left(\sum_{t=1}^{T}[(1-s_t)(1-s_{t-1})] + \alpha_p, \sum_{t=1}^{T}[s_t(1-s_{t-1})] + \beta_p\right),$$

$$\{q|\mathbf{s}\} \sim \text{Beta}\left(\sum_{t=1}^{T}[s_t s_{t-1}] + \alpha_q, \sum_{t=1}^{T}[s_{t-1}(1-s_t)] + \beta_q\right),$$

$$\{s_t|\mathbf{s}_{(-t)}, \boldsymbol{\eta}, \mathbf{y}\} \sim \text{Bernoulli}\left(\frac{\Pr(s_t = 1|\mathbf{s}_{(-t)}, \boldsymbol{\eta}) p(\mathbf{y}|s_t = 1, \mathbf{s}_{(-t)}, \boldsymbol{\eta})}{\sum_{j=0}^{1}\Pr(s_t = j|\mathbf{s}_{(-t)}, \boldsymbol{\eta}) p(\mathbf{y}|s_t = j, \mathbf{s}_{(-t)}, \boldsymbol{\eta})}\right), \quad t = 1, \ldots, T,$$

$$\{\boldsymbol{\phi}|\mathbf{s}, \alpha_0, \alpha_1, \sigma_\epsilon^2, \mathbf{y}\} \sim \text{N}\left((\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{z}, \sigma_\epsilon^2(\mathbf{Z}'\mathbf{Z})^{-1}\right),$$

$$\{\sigma_\epsilon^2|\mathbf{s}, \alpha_0, \alpha_1, \boldsymbol{\phi}, \mathbf{y}\} \sim \text{Inv-}\chi^2\left(T, (\mathbf{z} - \mathbf{Z}\boldsymbol{\phi})'(\mathbf{z} - \mathbf{Z}\boldsymbol{\phi})/T\right),$$

$$\{\alpha_0|\mathbf{s}, \alpha_1, \boldsymbol{\phi}, \sigma_\epsilon, \mathbf{y}\} \sim \text{N}\left(\frac{\sum_{t=1}^{T} y_t^*}{T}, \frac{\sigma_\epsilon^2}{T}\right),$$

$$p(\alpha_1|\mathbf{s}, \alpha_0, \boldsymbol{\phi}, \sigma_\epsilon, \mathbf{y}) \propto \text{N}(\alpha_1|\hat{\alpha}_1, \hat{\sigma_1}^2) \times I(\alpha_1 < -0.03),$$

where we have denoted

$$\hat{\alpha}_1 = \frac{\frac{\sum_{t=1}^{T} s_t y_t^{**}}{\sigma_\epsilon^2} + \frac{1}{\sigma_0^2}\mu_0}{\frac{\sum_{t=1}^{T} s_t}{\sigma_\epsilon^2} + \frac{1}{\sigma_0^2}}, \qquad \hat{\sigma}_1^2 = \left(\frac{\sum_{t=1}^{T} s_t}{\sigma_\epsilon^2} + \frac{1}{\sigma_0^2}\right)^{-1},$$

and

$$y_t^* = y_t - \alpha_1 s_t - \boldsymbol{\phi}'\mathbf{z}_{t-1}, \qquad y_t^{**} = y_t - \alpha_0 - \boldsymbol{\phi}'\mathbf{z}_{t-1}.$$

The notation Inv-$\chi^2(m, s^2)$ means the scaled inverse-chi-square distribution, defined as $\frac{ms^2}{\chi_m^2}$, where χ_m^2 is a chi-square distributed random variable with m degrees of freedom.

When determining the distribution of $\{s_t|\mathbf{s}_{(-t)}, \mathbf{y}, \boldsymbol{\eta}\}$, one needs $\Pr(s_t = 1|\mathbf{s}_{(-t)}, \boldsymbol{\eta})$. This is easily calculated as

$$\Pr(s_t = 1|\mathbf{s}_{(-t)}, \boldsymbol{\eta}) = \frac{\Pr(s_t = 1|s_{t-1}, p, q)\Pr(s_{t+1}|s_t = 1, p, q)}{\Pr(s_{t+1}|s_{t-1}, p, q)}, \quad 0 < t < T.$$

Appendix B. Estimation results of the Hamilton model with noninformative prior distributions

The posterior simulations were performed using the R computing environment. The following output was obtained using the summary function of the add-on package MCMCpack:

```
Number of chains = 3
Sample size per chain = 2500

1. Empirical mean and standard deviation for each variable,
   plus standard error of the mean:

            Mean        SD  Naive SE Time-series SE
alpha0  0.034387 0.0037252 4.302e-05      0.0001311
alpha1 -0.127041 0.0300157 3.466e-04      0.0014282
phi1    0.272877 0.1075243 1.242e-03      0.0031602
phi2   -0.161943 0.0960893 1.110e-03      0.0027343
sigmaE  0.001325 0.0001898 2.192e-06      0.0000047
p       0.972081 0.0213633 2.467e-04      0.0010495
q       0.402894 0.2120644 2.449e-03      0.0037676
sum(St) 5.673333 3.8759548 4.476e-02      0.2744217

2. Quantiles for each variable:

            2.5\%       25\%      50\%      75\%    97.5\%
alpha0  0.0274775  0.031792  0.034295  0.036825  0.042004
alpha1 -0.1864689 -0.148860 -0.125884 -0.104208 -0.074866
phi1    0.0462000  0.204920  0.277852  0.345023  0.476358
phi2   -0.3557249 -0.225779 -0.158533 -0.096778  0.022228
sigmaE  0.0009913  0.001193  0.001312  0.001444  0.001734
p       0.9156113  0.962588  0.977417  0.987580  0.997095
q       0.0506376  0.236130  0.389157  0.554960  0.834136
sum(St) 2.0000000  3.000000  4.000000  7.000000 16.000000


Gelman and Rubin's diagnostics
(Potential scale reduction factors):

        Point est. 97.5\% quantile
alpha0        1.01           1.03
alpha1        1.03           1.11
phi1          1.00           1.01
phi2          1.00           1.00
sigmaE        1.01           1.02
p             1.03           1.09
q             1.00           1.01
sum(St)       1.05           1.16
```

Appendix C. The conditional posterior distributions of the transfer function models

The following notation is used to simplify some of the expressions: $\mathbf{x} = (x_1, x_2, \ldots, x_T)'$, $\mathbf{x}^* = (x_1^*, x_2^*, \ldots, x_T^*)'$, $\mathbf{y} = (y_1, y_2, \ldots, y_T)'$, $\mathbf{1} = (1, 1, \ldots, 1)'$, $\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2)'$ and

$$\mathbf{X} = \begin{pmatrix} 1 & x_0 & y_1 \\ 1 & x_1 & y_2 \\ \vdots & & \\ 1 & x_{T-1} & y_T \end{pmatrix}.$$

The conditional posterior distributions of the transfer function models are as follows:

$$p(\boldsymbol{\beta}|\sigma_u^2, \nu, \mathbf{x}^*, \mathbf{y}) \propto \mathrm{N}\left(\boldsymbol{\beta}|\hat{\boldsymbol{\beta}}, \sigma_u^2(\mathbf{X}'\mathbf{X})^{-1}\right) \times I(\beta_1 < 1),$$

$$p(\sigma_u^2|\nu, \mathbf{x}^*, \mathbf{y}) = \text{Inv-}\chi^2\left(\sigma_u^2|T-3, (\mathbf{x}-\mathbf{X}\hat{\boldsymbol{\beta}})'(\mathbf{x}-\mathbf{X}\hat{\boldsymbol{\beta}})/(T-3)\right),$$

$$p_1(\nu|\boldsymbol{\beta}, \sigma_u^2, \mathbf{x}^*, \mathbf{y}) \propto \frac{1}{\prod_{t=1}^{T} t_\nu(x_t)} \exp\left(-\frac{1}{2\sigma_u^2}SS\right),$$

$$\nu = 1, \ldots, 200,$$

$$p_2(\nu|\boldsymbol{\beta}, \sigma_u^2, \mathbf{x}^*, \mathbf{y}) \propto \frac{1}{\nu\prod_{t=1}^{T} t_\nu(x_t)} \exp\left(-\frac{1}{2\sigma_u^2}SS\right),$$

$$\nu = 1, \ldots, 200,$$

where $SS = \sum_{t=1}^{T}(x_t - \beta_0 - \beta_1 x_{t-1} - \beta_2 y_t)^2$ and $\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{x}$. The notation Inv-$\chi^2(x|m, s^2)$ means the density of the scaled inverse-chi-square distribution (see definition in Appendix A) and $t_\nu(x)$ the density of Student's t distribution. Note that ν is implicit in all the above formulas, since it is needed to transform $\mathbf{x}^*$ to $\mathbf{x}$.

The Gibbs sampler was implemented using two blocks, $\boldsymbol{\theta}_1 = (\boldsymbol{\beta}, \sigma^2)$ and $\boldsymbol{\theta}_2 = \nu$. The first block was simulated by at first generating σ^2 from $p(\sigma_u^2|\nu, \mathbf{x}^*, \mathbf{y})$ and then $\boldsymbol{\beta}$ from $p(\boldsymbol{\beta}|\sigma_u^2, \nu, \mathbf{x}^*, \mathbf{y})$. The second block (the scalar ν) was easy to simulate, since it has a one-dimensional discrete distribution with finite support.

References

Alexander, C., 2005. The present and future of financial risk management. Journal of Financial Econometrics 3, 3–25.

Best, N., Spiegelhalter, D., Thomas, A., Brayne, C., 1996. Bayesian analysis of realistically complex models. Journal of the Royal Statistical Society, Series A 159, 323–342.

Carlin, B.P., Polson, N.G., Stoffer, D.S., 1992. A Monte Carlo approach to nonnormal and nonlinear state-space modeling. Journal of the American Statistical Association 87, 493–500.

Coe, P., 2002. Financial crisis and the great depression — A regime switching approach. Journal of Money, Credit and Banking 34, 76–93.

Gelfand, A.E., Hills, S.E., Racine-Poon, A., Smith, A.F.M., 1990. Illustration of Bayesian inference in normal data models using Gibbs sampling. Journal of the American Statistical Association 85, 972–985.

Gelman, A., Carlin, J.B., Stern, H.S., Rubin, D.B., 2004. Bayesian Data Analysis, Second ed. Chapman & Hall/CRC.

Geman, S., Geman, D., 1984. Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. IEEE Transactions on Pattern Analysis and Machine Intelligence 6, 721–741.

Hamilton, J., 1989. A new approach to the economic analysis of nonstationary time series and the business cycle. Econometrica 57, 357–384.

Hardy, M., 2002. Bayesian risk management for equity-linked insurance. Scandinavian Actuarial Journal 3, 185–211.

Koskinen, L., Pukkila, T., 2002. Risk caused by the catastrophic downturns of the national economy, in: 27th Transactions of the International Congress of Actuaries. Cancun, Mexico.

McNeil, A., Frey, R., Embrechts, P., 2005. Quantitative Risk Management. Princeton Press.

Norberg, R., 2006. The Pension Crisis: Its causes, possibile remedies, and the role of the regulator. Erfaringer og utfordringer Kredittilsynet 1986–2006, Kredittilsynet.

Pankratz, A., 1991. Forecasting with Dynamic Regression Models. John Wiley & Sons, New York.

Pension Insurance Data Book, 2005. http://www.pbgc.gov/docs/2005databook.pdf.

Rantala, J., Hietikko, H., 1988. An application of time series methods to financial guarantee insurance. European Journal of Operational Research 37, 398–408.

Romppainen, Y., 1996 Credit insurance and the up and down turns of the national economy, XXVII Astin Colloquium, Copenhagen, Denmark.

Scollnik, D., 2001. Actuarial modelling with MCMC and Bugs. North American Actuarial Journal 5, 96–124.

Sigma, 2000. Credit insurance and surety; solidifying commitments, 6, Swiss Re.

Smith, R., Goodman, D., 2000. Bayesian risk analysis. In: Embrechts, P. (Ed.), Extremes and Integrated Risk Management. Risk Books.

II

Bayesian analysis of equity-linked savings contracts with American-style options

Arto Luoma

University of Tampere, Finland

Anne Puustelli

University of Tampere, Finland

Lasse Koskinen

The Finnish Financial Supervisory Authority
and
Helsinki School of Economics, Finland

Equity-linked components are common in many life insurance products. In this paper a full Bayesian procedure is developed for the market consistent valuation of a fairly general equity-linked savings contract. The return on the contract consists of a guaranteed interest rate and a bonus depending on the yield of a total return equity index. The contract includes an American-style path-dependent surrender option, and it is valued in a stochastic interest rate environment. From the insurance company's viewpoint this paper provides a realistic and flexible modeling tool for product design and risk analysis.

The underlying asset and interest rate processes are estimated using the Markov Chain Monte Carlo method, and their simulation is based on their posterior predictive distribution, which is, however, adjusted to give risk-neutral dynamics. We introduce a procedure to determine a point estimate and confidence interval for the fair bonus rate of the contract. The contract prices with given bonus rates are estimated using the regression method.

The focus is on a novel application of advanced theoretical and computational methods, which enable us to deal with a fairly realistic valuation framework and to address model and parameter error issues. Our empirical results support the use of elaborated instead of stylized models for asset dynamics in practical applications.

Key Words: Metropolis algorithm, model error, option pricing via simulaton, risk-neutral valuation, Solvency, stochastic interest rate

1. INTRODUCTION

The Solvency II Directive (SII) is an EU directive that codifies and harmonizes the EU insurance regulation. It will shape the financial world in the decades to come. A number of other country regulators are watching Solvency II with a view to introducing similar risk-based capital regulation locally. International Association of Insurance Supervisors is currently developing and introducing a new global solvency framework, which has many things in common with SII (IAIS, 2008).

Solvency II reflects modern risk management practices to define required capital and manage risk. The regime assumes a market-consistent valuation of the balance sheet. Under SII, undertakings will be able to calculate the solvency capital requirement using a 'standard formula' or their own 'internal model' as approved by the supervisory authority (see, e.g., European Commission, 2009; Gatzert and Schmeiser, 2006; Ronkainen et al., 2007)

Most equity-linked life insurance policies, for example variable annuities and equity-indexed annuities in the United States, unit-linked insurance in the United Kingdom and equity-linked insurance in Germany, include implicit options, which represent a significant risk to the company issuing these contracts. Some products are fairly simple; others are complex, with a wide choice of guarantees and options. Some products have well-established features, others are highly innovative. One can find a useful introduction to different types of equity-linked insurance contracts in Hardy (2003). SII will probably cause an increase in the solvency capital requirement for products including options or guarantees. This would result in a search for 'new traditional products' which fulfill the customer demands for traditional life contracts but in a capital-efficient manner (see Morgan Stanley and Oliver Wyman, 2010).

The European insurance regulator EIOPA (2011) emphasizes that the insurer should take into account both basis risk and market risk in their life products. Other risks that might also be relevant include path dependence risk, lapse risk and model risk. Further, the regulator insists that companies involved in complex, equity-linked or other, products should use their own internal models for the calculation of solvency capital requirement. Here, we address many of these risks in a way which is suitable for internal modeling: we provide a general procedure and R codes.

Market consistent valuation of life insurance contracts has become a popular research area among actuaries and financial mathematicians; see, for example, Briys and de Varenne (1997), Grosen and Jorgensen (2000), Tanskanen and Lukkarinen (2003), Ballotta et al. (2006) and Bauer et al. (2006). However, most valuation models allowing for sophisticated bonus distribution rules and the inclusion of frequently offered options assume a quite simplified set-up.

One of the aims of this paper is to present a more realistic framework in which equity-linked savings contracts including guarantees and options can be valuated and analyzed. Since the existing products vary considerably and new ones are developed in the future, our valuation framework is fairly flexible; it includes several financial components which are crucial for equity-linked life products' risk analysis. Many types of products can be covered just by excluding some components of the contract.

Assumptions on the price dynamics of underlying assets usually lead to a partial differential equation (PDE) characterizing the price of the option. However, a closed form solution of such a PDE exists only in simpliest cases, and several features may render its numerical solution impractical, or the PDE may even fail to exist. The approach based on solving PDEs is difficult when, for instance, the asset price dynamics are sufficiently complex, or the payoff of an option depends on the paths of the underlying assets, or the number of underlying assets required by the replicating strategy is large (greater than three). Instead, Monte Carlo methods are routinely used in pricing this kind of derivatives (Glasserman, 2004). Nonetheless, pricing American-style options via Monte Carlo simulation still remains a challenging task. The problem lies in the estimation of the early exercise decisions available.

Applications of Monte Carlo methods have become popular also in life insurance. For instance, Zaglauer and Bauer (2008) present a framework in which participating life insurance contracts can be valuated and analyzed in a stochastic interest rate environment using Monte Carlo and discretization methods. Bacinello et al. (2009) describe an algorithm based on the Least Squares Monte Carlo method to price American options. Their framework allows, for example, randomness in mortality. Hardy (2002) uses the Bayesian approach and Markov Chain Monte Carlo (MCMC) methods for the risk management of equity-linked insurance. Our study extends the previous work in that we deal with estimation and model uncertainty issues and a challenging valuation problem in the same context.

The price of an option depends on the model describing the behavior of the underlying instrument. Most approaches specify a particular stochastic process to represent the price dynamics of the underlying asset and then derive an explicit pricing model. However, neither the true model, nor its parameter values are known. A common practice is to assume a relatively simple model, and to use point estimates of the model parameters or to calibrate them using the prices of other options. Yet many options in practice require an elaborate time-series specification for the price dynamics of the underlying asset, since a too simple model might fail to explain the price of its derivative (see, for example, Brigo and Mercurio, 2001). Hence, it becomes difficult at best to derive explicit pricing formulae. Furthermore, with the additional complexity of a rich time-series specification, estimation uncertainty becomes a genuine concern.

In this article we utilize the constant elasticity of variance (CEV) model, introduced by Cox and Ross (1976), to model the equity index process. It generalizes the geometric Brownian (GBM) model, which underlies the Black-Scholes approach to option valuation. Although being a generalizition, the CEV process is still driven by one source of risk, so that option valuation and hedging remain straightforward. For stochastic interest rate, we assume the Chan-Karolyi-Longstaff-Sanders (CKLS) model (see Chan et al. (1992)), which generalizes several commonly used short-term interest rate models.

Equity-linked insurance savings contracts are characterized by an interest rate guarantee and some bonus distribution rules, which typically provide the policyholder participation at some specified rate in an underlying index. One of the most common options available is the possibility to exit (surrender) the contract before maturity and receive a lump sum reflecting the insurer's past contribution to the policy minus some charges. These American-style options are called surrender options. In the related research the emphasis has been on the mathematics of pricing these options and on Monte Carlo experiments.

Following the tradition of equivalence principle in the premium setting, one also determines the fair bonus rate, that is, the break-even participation rate for which the fair value of the contract equals the initial investment. In this article we describe in detail how to apply Bayesian statistics to value equity-linked savings contracts including surplus options using a fairly realistic model for assets and interest rates. One could also consider this type of contract as a life-insurance contract in a way that the model provides an upper bound for insurance liabilities. Then one can interpret the policyholder to survive until the maturity of the contract as Ballotta et al. (2006) note.

We follow Bunnin et al. (2002) who use Bayesian numerical techniques to price a European Call option on a share index. The two major benefits from using Bayesian techniques are that we can explicitly acknowledge the risks associated to model choice and parameter estimation. In order to value American-style options we use the Tsitsiklis and Van Roy (1999, 2001) regression approach, which approximates the value of the option against a set of basic functions.

From the methodological point of view we address questions about: (i) implementation of MCMC to estimate the underlying diffusion processes, (ii) implementation of the regression method to determine the fair price of the insurance contract and its confidence interval, and (iii) solution of an inverse problem to determine the fair bonus rate and its confidence interval. From a more applied point of view we investigate the effect of the contract conditions, such as contract length, guaranteed interest rate and penalty rate, on the fair bonus rate and its estimation accuracy. Furthermore, we compare the results obtained in constant and stochastic interest rate environments, and briefly note the differences in model fit and valuation results between the used CEV model and the simpler GBM model.

The paper is organized as follows. Section 2 introduces the framework and model, Section 3 presents the estimation and evaluation procedures and Section 4 the empirical results. The final Section 5 concludes.

2. THE FRAMEWORK

2.1. The equity-linked savings contract

Our goal is to price an equity-linked savings contract. The contract is not exactly any of the yet existing products, but it has many features covering a large scale of different types of policies.

The contract consists of two parts. The first part is a guaranteed interest and the second part a bonus depending on the yield of some total return equity index. Thus, our product resembles equity-indexed annuities in the United States and equity-linked insurance contracts in Germany. On the other hand, in some equity-linked contracts the bonus is linked to a fund or combinaton of funds, for example in variable annuities in the United States or segregated fund contracts in Canada (see Hardy, 2003).

We denote the amount of savings in the insurance contract at time t_i by $Y(t_i)$. Then its growth during a time interval of length $\delta = t_{i+1} - t_i$ is given by

$$\log \frac{Y(t_{i+1})}{Y(t_i)} = g\delta + b \max\left(0, \log \frac{X(t_{i+1})}{X(t_i)} - g\delta\right), \tag{1}$$

where $X(t_i) = \sum_{j=0}^{q} S(t_{i-j})/(q+1)$ is a moving average of a total return equity index $S(t_i)$, g is a guaranteed rate and b is a bonus rate, the proportion of the excessive equity index yield which is returned to the customer. In this study we use the time interval $\delta = 1/255$, where 255 is approximately the number of the days in a year on which the index is quoted, and the lag length of the moving average is chosen to be $q = 125$ (i.e., half a year). The use of a moving average decreases the volatility of the contract value, and thus facilitates hedging.

The model also incorporates a surrender (early exercise) option and possibility for a penalty p which occurs if the customer reclaims the contract before the final expiration date. If the penalty is set at a too high level, the contract becomes like a European-style option, which is exercised only at the end of the contract period. A further condition is that there will be a 1 % penalty if the contract is reclaimed during the first 10 working days. This condition essentially improves the estimation described in Section 3.3.2.

In the following, we will consider the two cases where (i) the riskless interest rate is fixed at a predetermined value r, or (ii) is assumed to be stochastic. For the constant interest rate r the guaranteed rate g is set at kr throughout the entire contract period for some constant $k < 1$. In the case of stochastic interest rate, the guaranteed rate is fixed for one year at a time. It is set annually at kr_t, where r_t is the riskless short-term interest rate at time t. By setting the guaranteed rate for one year at a time and not daily, the insurance company can better hedge its liabilities, and on the other hand, the customer will have a better idea of the guaranteed growth rate.

In this framework the penalty p for early exercise and the parameters k, g and b are predefined by the insurance company. However, in the case of a stochastic interest rate, g is reset annually. We will determine a fair bonus rate b so that the risk-neutral price of the contract is equal to the initial investment. This gives the contract a simple structure and makes its costs and returns visible and predictable for the insurer and the customer. Our main interest is to study the effects of the expiration date, guaranteed rate and penalty rate on the fair bonus rate in both constant and stochastic interest rate cases.

The equity-indexed annuity contract has a modification called an annual ratchet in which the index participation is evaluated year by year. Each year the amount of savings is increased by the greater of the floor rate, which is usually 0 percent, and the increase in the underlying index, multiplied by the participation rate. Our contract is similar to this apart from being evaluated on a daily basis. Our contract type is better linked to the dynamics of the financial markets, since the customer may follow the growth of the savings daily and also exercise the contract at market value.

2.2. Model with constant interest rate

The constant elasticity of variance (CEV) process introduced by Cox and Ross (1976) is used to model the equity index process. It is a nonnegative diffusion process, defined by the stochastic differential equation (SDE)

$$\mathrm{d}S_t = \mu S_t \mathrm{d}t + \nu S_t^{1-\alpha} \mathrm{d}W_t, \tag{2}$$

where μ, ν (> 0) and α are fixed parameters and W_t is a standard Brownian motion under a real-world probability measure $\mathcal{P}$. If $\alpha = 0$, the process (2) becomes a geometric Brownian motion. In the estimation, we assume that $\alpha > 0$, which means that the volatility is smaller for larger values of S_t. If $\alpha > \frac{1}{2}$, there is a positive probability that the process converges to zero. The model may also be written in the form

$$\mathrm{d}S_t = rS_t \mathrm{d}t + \nu S_t^{1-\alpha} \mathrm{d}Z_t, \tag{3}$$

where r is the riskless short-term interest rate and Z_t a standard Brownian motion under a risk-neutral probability measure $\mathcal{Q}$. The parameters μ, ν and α are unknown and will be estimated.

The transition densities of the process (2) have closed form solutions which use the modified Bessel function of the first kind (see Bunnin et al., 2002). However, we found the Euler discretization of the process to be accurate enough for estimation and simulation purposes, since our discretization interval is only one working day. The Euler scheme is the simplest standard method for approximate simulation of stochastic differential equations; for further details, see Iacus (2008) or Glasserman (2004).

Assuming that the discretized process (2) has been observed at equally-spaced time points $0, \delta, ...N\delta$, the likelihood function can be written in the form

$$p(y|\theta) = \prod_{i=1}^{N} \frac{1}{\sqrt{2\pi\nu^2 S_{(i-1)\delta}^{2(1-\alpha)}\delta}} \exp\left(-\frac{(\Delta S_{i\delta} - \mu S_{(i-1)\delta}\delta)^2}{2\nu^2 S_{(i-1)\delta}^{2(1-\alpha)}\delta}\right),$$

where y is data, $\theta = (\mu, \nu, \alpha)$ and $\Delta S_{i\delta} = S_{i\delta} - S_{(i-1)\delta}$.

2.3. Model with stochastic interest rate

In our second set-up, we assume that the dynamics of riskless short-term rate r_t and stock index S_t are described by the following system of SDEs:

$$\mathrm{d}r_t = \kappa(\xi - r_t)\mathrm{d}t + \sigma r_t^{\gamma}\mathrm{d}W_t^{(1)}, \tag{4a}$$

$$\mathrm{d}S_t = \mu S_t\mathrm{d}t + \nu S_t^{1-\alpha}\mathrm{d}W_t^{(2)}, \tag{4b}$$

with $W_t^{(1)}$ and $W_t^{(2)}$ two standard Brownian motions, correlated through $W_t^{(2)} = \rho W_t^{(1)} + \sqrt{1-\rho^2}W_t^{(3)}$, where $W_t^{(1)}$ and $W_t^{(3)}$ are independent standard Brownian motions under a real-world probability measure. Thus the correlation of $W_t^{(1)}$ and $W_t^{(2)}$ is ρ.

The short-term interest rate model (4a) was introduced by Chan et al. (1992), who provide a useful summary of short-term interest rate models in their paper. The two most commonly used models which may be derived from this model by parameter restriction are the following: If $\gamma = 0$, the model becomes the Ornstein-Uhlenbeck process, proposed by Vasiček (1977) as a model of the short rate, and, if $\gamma = \frac{1}{2}$, it becomes a square-root diffusion referred to as the Cox-Ingersoll-Ross (CIR) model (Cox et al., 1985). In estimation, the parameters κ, ξ, σ and γ are assumed to be positive.

Substituting $Z_t^{(1)} = W_t^{(1)}$ and $Z_t^{(3)} = W_t^{(3)} + (\mu - r_t)\nu^{-1}(1-\rho^2)^{-1/2}S_t^{\alpha}dt$, the system of SDEs (4a) and (4b) becomes

$$\mathrm{d}r_t = \kappa(\xi - r_t)\mathrm{d}t + \sigma r_t^{\gamma}\mathrm{d}Z_t^{(1)}, \tag{5a}$$

$$\mathrm{d}S_t = r_t S_t\mathrm{d}t + \nu S_t^{1-\alpha}\mathrm{d}Z_t^{(2)}, \tag{5b}$$

where $Z_t^{(2)} = \rho Z_t^{(1)} + \sqrt{1-\rho^2}Z_t^{(3)}$. Now a risk-neutral probability measure Q may be introduced by assuming that $Z_t^{(1)}$ and $Z_t^{(3)}$ are two independent standard Brownian motions under this measure. It can then be shown that the discounted price $\tilde{S}_t = S_t \exp(-\int_0^t r_s ds)$ is a martingale under Q.

The transition densities of the bivariate process described by (4a) and (4b) do not have a closed form solution, and we will use its Euler discretization to estimate the unknown parameters κ, ξ, σ, γ, μ, ν and α. Accordingly, we will simulate the risk-neutral process using the Euler discretization of (5a) and (5b).

In order to obtain numerical stability in estimation, we reparametrize model (4a) as

$$\mathrm{d}x_t = (\beta - \kappa x_t)\mathrm{d}t + \tau x_t^{\gamma}\mathrm{d}W_t^{(1)},$$

where $x_t = 100\, r_t$ (the interest rate given in percentages), $\beta = 100\,\kappa\xi$ and $\tau = (100)^{1-\gamma}\sigma$. Assuming that the bivariate process has been observed at equally-spaced time points $0, \delta, ...N\delta$, the likelihood function can be written in the form

$$p(y|\theta) = \prod_{i=1}^{N} \frac{1}{\sqrt{2\pi\tau^2 x_{(i-1)\delta}^{2\gamma}\delta}} \exp\left(-\frac{\left(\Delta x_{i\delta} - (\beta - \kappa x_{(i-1)\delta})\delta\right)^2}{2\tau^2 x_{(i-1)\delta}^{2\gamma}\delta}\right)$$

$$\times \prod_{i=1}^{N} \frac{1}{\sqrt{2\pi\nu^2 S_{(i-1)\delta}^{2(1-\alpha)}(1-\rho^2)\delta}}$$

$$\exp\left(-\frac{\left(\Delta S_{i\delta} - \mu S_{(i-1)\delta}\delta - \nu S_{(i-1)\delta}^{1-\alpha}\rho\Delta W_{i\delta}^{(1)}\right)^2}{2\nu^2 S_{(i-1)\delta}^{2(1-\alpha)}(1-\rho^2)\delta}\right), \tag{6}$$

where y is data, $\theta = (\mu, \nu, \alpha, \beta, \kappa, \tau, \gamma, \rho)$, $\Delta x_{i\delta} = x_{i\delta} - x_{(i-1)\delta}$, $\Delta S_{i\delta} = S_{i\delta} - S_{(i-1)\delta}$ and

$$\Delta W_{i\delta}^{(1)} = \frac{x_{i\delta} - x_{(i-1)\delta} - (\beta - \kappa x_{(i-1)\delta})\delta}{\tau x_{(i-1)\delta}^{\gamma}}.$$

3. ESTIMATION AND EVALUATION PROCEDURES

3.1. The Metropolis algorithm

The unknown parameters of the stock index and interest rate models are estimated using Bayesian methods. This makes it possible to take parameter uncertainty into account when evaluating the fair prices of derivatives. We follow Bunnin et al. (2002) who simulate the paths of the underlying asset using the posterior predictive distribution of the underlying asset. In their framework the posterior predictive distribution is constructed by averaging over alternative models and their parameters, thus taking into account the uncertainty related to them. However, we do not average over models, since we assume that model uncertainty can be taken into account by using a sufficiently general, continuously parametrized family of distributions. This approach is recommended in Section 6.7 of Gelman et al. (2004). We also use an estimation algorithm different from the SIR algorithm used by Bunnin et al. (2002). Even though this algorithm is fast and relatively simple, it requires an informative and carefully adjusted prior, which makes its use laborious when several parameters should be estimated. Instead, we use the Metropolis algorithm introduced by Metropolis et al. (1953) to simulate the joint posterior distribution of unknown parameters in both fixed and stochastic interest rate cases.

The Metropolis algorithm is a Markov Chain Monte Carlo (MCMC) method, and can be used to simulate Markov chains with given stationary distributions. The MCMC methods are especially useful when direct sampling from a probability distribution is difficult. The Metropolis algorithm is more flexible than the Gibbs sampler, which presumes the ability to generate random variates from the full conditional distributions of the target distribution. In order to implement the Metropolis algorithm, one only needs to know the joint density function of the target distribution up to a constant of proportionality.

Suppose that we wish to simulate a (multivariate) distribution with density $p(\theta)$. The algorithm works as follows: We first assign an initial value θ^0 such that $p(\theta^0) > 0$ from a starting distribution $p_0(\theta)$. Then, assuming that vectors $\theta^0, \theta^1, ..., \theta^{t-1}$ have been generated, we generate a proposal θ^* for θ^t from a jumping distribution $J(\theta^*|\theta^{t-1})$ which is symmetric in the sense that $J(\theta_a|\theta_b) = J(\theta_b|\theta_a)$ for all θ_a and θ_b. Finally, iteration t is completed by calculating the ratio

$$r = \frac{p(\theta^*)}{p(\theta^{t-1})}$$

and by setting the new value at

$$\theta^t = \begin{cases} \theta^* & \text{with probability } \min(r, 1) \\ \theta^{t-1} & \text{otherwise.} \end{cases}$$

It can be shown that, under mild conditions, the algorithm produces an ergodic Markov Chain whose stationary distribution is $p(\theta)$. We see that the transition kernel $T(\theta^t|\theta^{t-1})$ is a mixture of discrete probability at $\theta^t = \theta^{t-1}$ and the jumping density $J(\theta^*|\theta^{t-1})$.

As mentioned above, we use the Metropolis algorithm to simulate the posterior distribution. The posterior density is proportional to the product of the prior density and the likelihood,

$$p(\theta|y) \propto p(\theta)p(y|\theta).$$

We use an improper uniform prior distribution

$$p(\theta) \propto \begin{cases} 1 & \text{when } |\rho| < 1 \text{ and } \min(\beta, \kappa, \tau, \nu, \alpha) > 0 \\ 0 & \text{otherwise} \end{cases}$$

for the bivariate model of stock index and stochastic interest rate. The posterior function is thus proportional to the likelihood (6) in a feasible region of parameters. For the univariate stock index model (2) we use a uniform prior with the restriction $\min(\nu, \alpha) > 0$.

3.2. Pricing American options with regression methods

3.2.1. Point estimation of option prices

The equity-linked savings contract we want to price is in practice an American option with a path-dependent moving average feature. An American option gives the holder the right to exercise the option at any time up to the expiry date T. In pricing we adopt the least squares method introduced by Tsitsiklis and Van Roy (1999, 2001). It is a simple but powerful approximation method for American-style options. Longstaff and Schwartz (2001) provide a slightly different version of the method. In the following brief introduction we follow Glasserman (2004).

The pricing of an American option is based on an optimal exercising strategy. Let us assume that the relevant underlying security prices of the economy follow a d-dimensional Markov process $X(t)$ and that the payoff value of the option at time t is given by $\tilde{h}(X(t))$. The process $X(t)$ may be augmented to include a stochastic interest rate $r(t)$ and, in the case of path-dependent options, past values of the underlying processes as well.

Furthermore, let $\mathcal{T}$ denote a set of admissible stopping times with values in $[0, T]$. More specifically, we assume that the decision whether to stop at time t is a function of $X(t)$ and that the option can only be exercised at the m discrete times $0 < t_1 \le t_2 \le \cdots \le$

$t_m = T$. (If desirable, one can improve the approximation to continuously exercisable options by increasing m.) The goal in optimal exercising is to find a stopping time maximizing the expected discounted payoff of the option. The price of the option is given by

$$\sup_{\tau \in \mathcal{T}} \mathrm{E}\left[\exp\left(-\int_0^{\tau} r(s)ds\right)\tilde{h}(X(\tau))\right],$$

where $\{r(t), 0 \leq t \leq T\}$ is an instantaneous short rate process and the expectation is taken with respect to the risk-neutral probability measure.

To simplify notation we will write $X(t_i)$ as X_i. Let $\tilde{h}_i$ denote the payoff function for exercise at t_i and $\tilde{V}_i(x)$ the value of the option at t_i given $X_i = x$. One can then represent pricing algorithms recursively as follows:

$$\begin{aligned} \tilde{V}_m(x) &= \tilde{h}_m(x) \\ \tilde{V}_{i-1}(x) &= \max\{\tilde{h}_{i-1}(x), \mathrm{E}[D_{i-1,i}(X_i)\tilde{V}_i(X_i)|X_{i-1} = x]\}, \quad i = 1, \ldots, m, \end{aligned}$$

where $D_{i-1,i}(X_i)$ is the discount factor from t_{i-1} to t_i. We thus assume that the discount factor is a function of X_i, which may be achieved by augmenting X_i, if necessary. Usually, it is given by $D_{i-1,i}(X_i) = \exp\left(-\int_{t_{i-1}}^{t_i} r(u)du\right)$, but it is also possible to use a numeraire process other than the one based on riskless interest, provided that the expectation is taken with respect to a measure consistent with that numeraire. One can also show that equivalent to the procedure described above is to deal with time zero values $h_i(x) = D_{0,i}(x)\tilde{h}_i(x)$ and $V_i(x) = D_{0,i}(x)\tilde{V}_i(x)$, $i = 0, 1, ..., m$ (see Glasserman, 2004). Then, at time t_i, the decision to continue is based on comparing the discounted immediate exercise value $h_i(x)$ with the corresponding discounted continuation value $C_i(x) = \mathrm{E}[V_{i+1}(X_{i+1})|X_i = x]$. In the sequel, we will use these time zero values in order to simplify notation.

In regression methods it is assumed that the continuation value may be expressed as the linear regression

$$\mathrm{E}[V_{i+1}(X_{i+1})|X_i = x] = \sum_{r=1}^{M} \beta_{ir}\psi_r(x),$$

for some basis functions $\psi_r : \mathcal{R}^d \to \mathcal{R}$ and constants β_{ir}, $r = 1, ..., M$. In order to estimate the coefficients one first generates b independent paths $\{X_{1,j}, ..., X_{m,j}\}$, $j = 1, ..., b$, and sets $\hat{V}_{m,j} = h_m(X_{m,j})$, $j = 1, ..., b$, at terminal nodes. Then one proceeds backward in time and, using ordinary least squares, fits at time t_i the regression model

$$\hat{V}_{i+1,j}(X_{i+1,j}) = \sum_{r=1}^{M} \beta_{i,r}\psi_r(X_{i,j}) + \epsilon_{i,j}, \quad j = 1, ..., b, \tag{7}$$

where $\epsilon_{i,j}$ are residuals. The estimated value of the option for path j at time t_i is

$$\hat{V}_{i,j} = \max\{h_i(X_{i,j}), \hat{C}_i(X_{i,j})\},$$

where $\hat{C}_i(X_{i,j})$ is the fitted value from Equation 7. Finally, the estimate of the option price is given by $\hat{V}_0 = (\hat{V}_{1,1} + ... + \hat{V}_{1,b})/b$.

3.2.2. *Determining upper and lower bounds for option prices*

Glasserman (2004), Andersen and Broadie (2004) and Haugh and Kogan (2004) describe in detail methods to determine the upper and the lower bounds for the price of an American option. These bounds are important in assessing the reliability and accuracy of the price estimate. For the lower bound, one needs to simulate new paths $\{X_{1j}, \ldots, X_{mj}\}$, $j = 1, ..., b$, of the underlying process, independent of the paths used to estimate the regression coefficients. Define a stopping time as

$$\hat{\tau}_j = \min\{i : h_i(X_{ij}) \geq \hat{C}_i(X_{ij})\},$$

which is the first time when the immediate exercise value is greater than or equal to the estimated continuation value. The estimated continuation value $\hat{C}_i(X_{ij})$ may be computed using the estimated regression model or some other method. The low estimator for a single path j is

$$\hat{v}_j = h_{\hat{\tau}_j}(X_{\hat{\tau},j})$$

and the lower bound of the price is estimated as the mean of low estimators over all paths. Since no policy can be better than an optimal policy, this results in a low biased estimator.

The upper bounds are based on the inequality

$$V_0(X_0) = \sup_{\tau} \mathrm{E}[h_\tau(X_\tau)] \leq \mathrm{E}[\max_{i=1,\ldots,m} \{h_i(X_i) - M_i\}],$$

where $M = \{M_i, i = 0, ..., m\}$ is any martingale with $M_0 = 0$. It can be shown (see Glasserman, 2004) that this inequality holds with equality when M is defined as

$$M_i = \Delta_1 + ... + \Delta_i, \quad i = 1, ..., m,$$

where

$$\Delta_i = V_i(X_i) - \mathrm{E}[V_i(X_i)|X_{i-1}], \quad i = 1, ..., m. \tag{8}$$

There are two methods for constructing martingales $\hat{M}$ which approximate the optimal martingale M. The first method uses approximate value functions and the second approximate stopping times. We will use the first method, since it is less computationally intensive. For this method, one also needs to simulate new paths of the underlying process, independent of the paths used to estimate the regression coefficients. Suppose that we have simulated b additional paths $\{X_{1j}, ..., X_{mj}\}$, $j = 1, ..., b$. Then the option value at time $i + 1$ on path j is estimated as

$$\hat{V}_{i+1}(X_{i+1,j}) = \max\left(h_{i+1}\left(X_{i+1,j}\right), \hat{C}_{i+1}\left(X_{i+1,j}\right)\right),$$

where the continuation value, $\hat{C}_{i+1}\left(X_{i+1,j}\right)$, is computed using the estimated regression model.

Now we could use the martingale differences

$$\hat{\Delta}_{i+1,j} = \hat{V}_{i+1}(X_{i+1,j}) - \mathrm{E}\left(\hat{V}_{i+1}(X_{i+1,j})|X_{ij} = x\right)$$

to approximate the optimal differences (8) if we knew the expected value. However, the expectation can be replaced by a mean of nested simulations as follows: At each step X_{ij} of the Markov chain, new successors $X^{(k)}_{i+1,j}$, $k = 1, 2, \ldots, n$, are generated, and the corresponding option values,

$$\hat{V}_{i+1}\left(X^{(k)}_{i+1,j}\right) = \max\left(h_{i+1}\left(X^{(k)}_{i+1,j}\right), \hat{C}_{i+1}\left(X^{(k)}_{i+1,j}\right)\right),$$

are computed. Then we can define the martingale differences as follows:

$$\hat{\Delta}_{i+1,j} = \hat{V}_{i+1}(X_{i+1,j}) - \frac{1}{n}\sum_{k=1}^{n}\hat{V}_{i+1}\left(X^{(k)}_{i+1,j}\right),$$

and use them as an approximation to (8).

The corresponding martingale values are given by

$$\hat{M}_{i+1,j} = \hat{M}_{ij} + \hat{\Delta}_{i+1,j}$$

with $\hat{M}_{0j} = 0$. The upper bound of an option may now be estimated as

$$\mathrm{E}\left(\max_{i=1,\ldots,m}(h_i(X_{ij}) - \hat{M}_{ij})\right) \approx \frac{1}{b}\sum_{j=1}^{b}\max_{i=1,\ldots,m}(h_i(X_{ij}) - \hat{M}_{ij}).$$

3.3. Implementation

3.3.1. Choosing the regression variables

In our application, the continuation values of the option depend on the path of the underlying index value in a complicated way. Theoretically, we would need $q + 1$ state variables (or $q+2$ in the case of stochastic interest rate) to satisfy the Markovian assumption of the process. However, we consider that the current value of the index, its moving average, and the first index value appearing in the moving average may be used to predict the continuation value reasonably well. The use of the moving average may be motivated by observing that the growth of savings in the insurance contract depends on the path of the moving average (see Equation 1). The current index value and the first value appearing in the moving average help predict the future evolution of the moving average. The current amount of savings also helps predict the continuation value, but it is not included in the regression variables. Instead, it is subtracted from the regressed value before fitting the regression and subsequently added to the fitted value.

To avoid under- and overflows in the computations, the regression variables are scaled by the first index value. Thus, the following state variables are used:

$$X_1(t_i) = \frac{S(t_i)}{S(0)}$$

$$X_2(t_i) = \frac{\sum_{j=0}^{q} S(t_{i-j})/(q+1)}{S(0)}$$

$$X_3(t_i) = \frac{S(t_{i-q})}{S(0)}.$$

However, multicollinearity problems would occur if all the variables X_1, X_2 and X_3 were used at all time points. In fact, X_3 would be equal for all simulations paths for $i \le q$ and the moving averages X_2 would be very close to each other for small values of i. Therefore, we apply the following rule: The variable X_1 alone is used for $i < q/2$, X_1 and X_2 are used for $q/2 \le i < 3q/2$ and all variables are used for $i \ge 3q/2$.

We use Laguerre polynomials, suggested by Longstaff and Schwartz (2001), as basis functions. More specifically, we use the first two polynomials

$$L_0(X) = \exp(-X/2)$$
$$L_1(X) = \exp(-X/2)(1 - X)$$

for the variables X_1, X_2 and X_3. In addition, we use the cross-products $L_0(X_1)L_0(X_2)$, $L_0(X_1)L_1(X_2)$, $L_1(X_1)L_0(X_2)$, $L_0(X_1)L_0(X_3)$ and $L_0(X_2)L_0(X_3)$. We also tried adding L_2, and r_t in the case of stochastic interest rate, but these did not improve valuation accuracy. Thus, we have altogether 11 explanatory variables in the regression. At the time points where only X_1 is used we have only two explanatory variables, $L_0(X_1)$ and $L_1(X_1)$.

3.3.2. Inverse problem

Using the procedure described above we can determine the option price (i.e., the price of the insurance contract) when the bonus rate b and the guaranteed rate g have been given. However, we are interested to determine the bonus rate so that the price of the contract is equal to the initial savings. The problem of determining b is a kind of inverse prediction problem, and we need to estimate the option value for various values of b. Since there are several sources of uncertainty involved in the estimation, we also need to repeat it several times for fixed values of b. We end up estimating a regression model where the option price estimates are regressed on the corresponding bonus rates. We found the third degree polynomial curve to be flexible enough for this purpose. After fitting the curve, we solve the bonus rate b for which the option price is equal to 100, which we assume to be the initial amount of savings.

We repeat this procedure for the upper and lower bounds of the prices. Thus, we estimate altogether three regression models, represented by three cubic polynomial curves (see Figure 1). The intersections of the horizontal line at price level 100 with the upper and lower bound curves yield lower and upper bounds for the fair bonus rate, respectively. In order to facilitate the estimation of the lower bound of the fair bonus rate we set the further condition that there is a 1% penalty for reclaiming the contract during the first ten days.

Prior to fitting the polynomial, it is, however, necessary to determine an initial interval for the solution. For this purpose we developed a modified bisection method. In this method, one first specifies initial upper and lower limits for the bonus rate; we use the values $l = 0$ and $u = 1$. Then one estimates the option price as well as its upper and lower bounds at $(l + u)/2$. If the lower bound of the price is greater than 100, the upper limit of the bonus rate is set at $u - (u - l)/4$; if the upper bound of the price is smaller than 100, the lower limit of the bonus rate is set at $l + (u - l)/4$. In other cases the upper limit of the bonus rate is set at $u - (u - l)/8$ and the lower limit at $l + (u - l)/8$. This procedure is continued until $u - l = 0.25$. Note that the new limit is not set in the middle of the interval, as is done in the ordinary bisection method, since this might lead to missing the correct solution due to the randomness of price estimates.

Figure 1 illustrates the estimation procedure. The option price and its lower and upper bounds are estimated for 10 different bonus rates, and the estimation is repeated 10 times for each bonus rate, which produces 300 points to the scatter plot. Each estimation is based on 1000 simulated paths. The limits for the bonus rate were determined using the modified bisection method described above. When producing this figure, the time to maturity was set at 3 years, the guaranteed rate at 0, the penalty rate at 0 and the constant interest rate at 0.04. We can see that the fair bonus rate is approximately 0.33, the lower bound 0.31 and the upper bound 0.37.

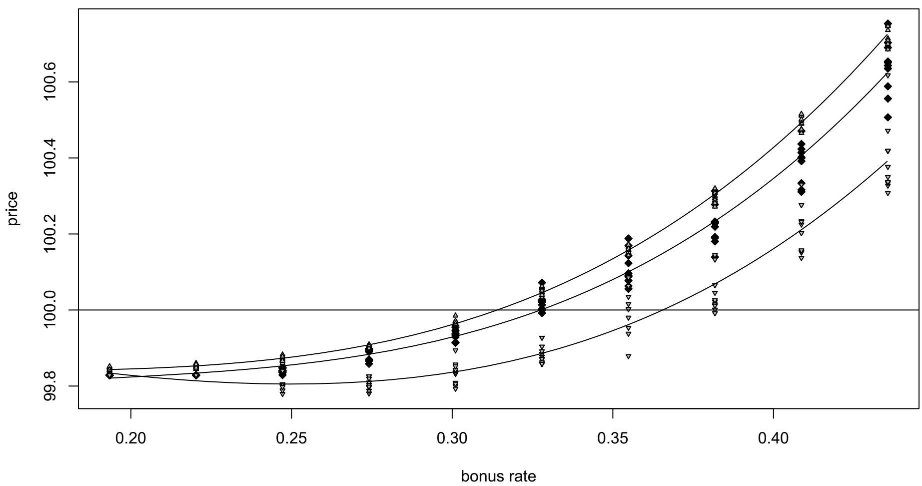


FIG. 1. Option price estimates (black diamonds), upper bounds (gray triangle point up) and lower bounds (gray triangle point down) vs. bonus rates.

As mentioned above, the bonus rate is solved from the equation $y = f(x)$, where y is the price of the contract and

$$f(x) = \hat{\beta}_0 + \hat{\beta}_1 x + \hat{\beta}_2 x^2 + \hat{\beta}_3 x^3 = \mathbf{x}'\hat{\boldsymbol{\beta}}, \tag{9}$$

with $\hat{\boldsymbol{\beta}} = (\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2, \hat{\beta}_3)'$ the ordinary least squares (OLS) estimate of the cubic regression model and $\mathbf{x} = (1, x, x^2, x^3)'$ a regression vector. The purpose of the initial penalty rate is to ensure that there is exactly one solution in the relevant interval.

Using the delta method, one also obtains an approximate variance for the estimate of x:

$$\mathrm{Var}(\hat{x}) \approx \frac{1}{[f'(x)]^2}\mathrm{Var}(f(x)) \approx \frac{1}{\left(\hat{\beta}_1 + 2\hat{\beta}_2\hat{x} + 3\hat{\beta}_3\hat{x}^2\right)^2}\hat{\mathbf{x}}'\mathrm{Cov}(\hat{\boldsymbol{\beta}})\hat{\mathbf{x}}.$$

4. EMPIRICAL RESULTS

4.1. Estimation of the parameters

In order to experiment with actual data and to estimate the unknown parameters of the models (2) and (4a,4b), we chose the following data sets: As an equity index we use the Total Return of Dow Jones EURO STOXX Total Market Index (TMI), which is a benchmark covering approximately 95 per cent of the free float market capitalization of Europe. The objective of the index is to provide a broad coverage of companies in the Euro zone including Austria, Belgium, Finland, France, Germany, Greece, Ireland, Italy, Luxembourg, the Netherlands, Portugal and Spain. The index is constructed by aggregating the stocks traded on the major exchanges of Euro zone. Only common stocks and those with similar characteristics are included, and any stocks that have had more than 10 non-trading days during the past three months are removed. In estimation, we use daily quotes from March 4th, 2002 until December 6th, 2007.

As a proxy for riskless short-term interest rate, we use Eurepo, which is the benchmark rate of the large Euro repo market. Eurepo is the rate at which one prime bank offers funds in euro to another prime bank if in exchange the former receives from the latter Eurepo GC as collateral. It is a good benchmark for secured money market transactions in the Euro zone. In the estimation of the interest rate model we use the 3 month Eurepo rate, since it behaves more regularly than the rates with shorter maturities. Both the index and interest series are presented in Figure 2.

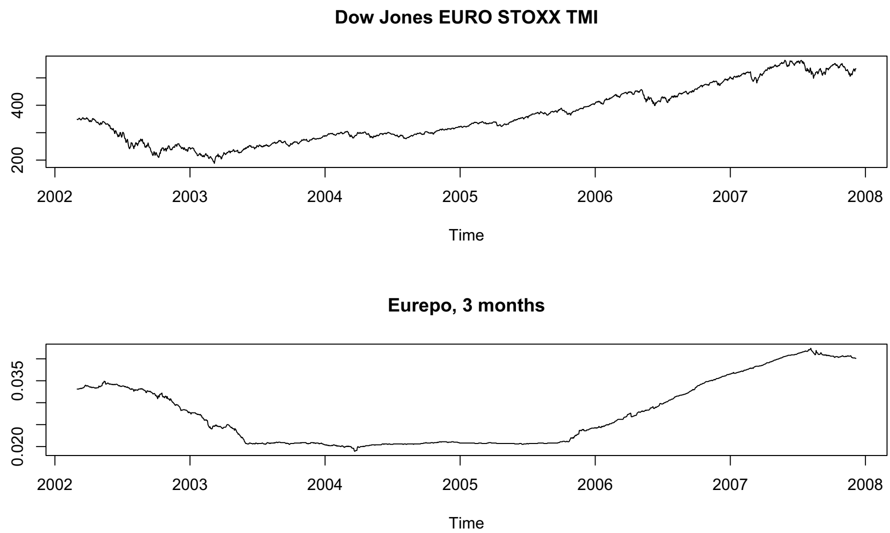


FIG. 2. The equity index and interest series.

We had no remarkable convergence problems when estimating the model parameters. We used three chains in MCMC simulation, and all of them converged rapidly to their stationary distributions. The summary of the estimation results, as well as Gelman and

Rubin's diagnostics (see Gelman et al., 2004), are given in Appendix A. The values of the diagnostic are close to 1 and thus indicate good convergence. All computations were made and figures produced using the R computing environment (R Development Core Team, 2010). To speed up computations we coded the most time consuming loops in C++. The code and data needed to replicate the results of this article are available at http://mtl.uta.fi/codes/savings.

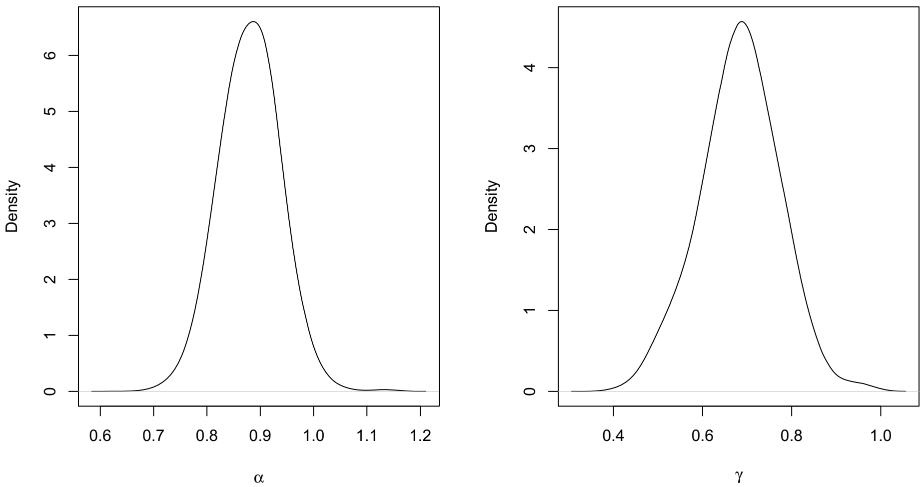


FIG. 3. Posterior distributions of the parameters α (index model) and γ (interest rate model).

The posterior distributions of the parameters α (Equation 2) and γ (Equation 4a) are shown in Figure 3. As already noted in Section 2.2, the CEV model becomes the geometric Brownian motion (GBM) when $\alpha = 0$. The figure reveals clearly that the posterior probability of α being around zero is vanishingly small, which makes the GBM highly improbable. We also tested its use in the pricing of the contract, and found that it gave substantially lower bonus rates than the more general alternative. Both models have, under the risk-neutral probability measure, equal expected yields for the underlying index, but the volatility will be greater with the GBM, since in the CEV model the volatility decreases as the level of the process increases. But greater volatility increases the probability of great profits, while not increasing the probability of great losses, since the accumulated capital is guaranteed to the customer. Consequently, the fair bonus rate is smaller in the case of the GBM.

This illustrates how our approach to use general models efficiently prevents the model error resulting from the use of a too simple model. On the other hand, we see that $\gamma = 1/2$ is not highly improbable in the interest rate model, so the model error would not be large if the CIR model were used instead of the more general CKLS model. The statistical inference of the parameters α and γ seems to be robust to the choice of data, since their posterior distributions remained similar when we included the data from year 2008, which was exceptional in that there was a collapse both in the stock markets and the interest rate.

Clearly, the fit of the index process could be further improved by modeling the volatility as a separate process. Such a stochastic volatility model would, however, be more difficult to estimate, and the valuation and hedging of the contract would be considerably more complicate.

4.2. Evaluation of the fair bonus rate

There are several parameters which may be varied in the equity-linked savings contract described by Equation 1. These include the duration of the contract, the lag length of the moving average, and the guaranteed rate. Furthermore, the number of simulated paths needs to be decided when the contract price is estimated, as well as the number of estimation repetitions when the fair bonus rate is determined. When the interest rate is assumed to be constant, it must be fixed at some predefined level, while when it is assumed to be stochastics, its starting level must be given. Our model also incorporates a possibility for a penalty rate which the customer has to pay if he reclaims the contract before the final expiration date. When the penalty rate is set at a high level, the price of the contract is determined like that of a European option, since then it is usually more profitable for the customer to keep the contract until the final expiration date.

We compared the accuracy of fair bonus rate estimation in the following two cases: first, we simulated 1000 paths to estimate the contract price and repeated the estimation 250 times to estimate the fair bonus price using the regression model (9), and, second, used 500 simulation paths and repeated it 500 times. We found that the standard error of the bonus rate estimate was in the second case almost twice as large as in the first case. This indicates that it is more important to increase the number of paths in the option price calculation than the number of repetitions in the bonus rate calculation. However, the differences in the bonus rate estimates resulting from the use of these two simulation schemes were very small; the maximum difference was 0.6 percentage units in our simulations.

The estimates of the fair bonus rate and the 95 % confidence intervals in the cases of constant interest rates 0.04 and 0.07 are shown in Tables 1 and 3, respectively. The confidence interval is calculated using the lower ($\hat{b}_l$) and upper ($\hat{b}_u$) estimates of the fair bonus rate and their standard errors:

$$\mathrm{CI} = \left(\hat{b}_l - 1.96\, s.e.(\hat{b}_l); \hat{b}_u + 1.96\, s.e.(\hat{b}_u)\right).$$

The guaranteed rate was set at 0, 1/3 and 2/3 of the interest rate, that is, 0, 0.013 and 0.027 for $r = 0.04$, and 0, 0.023 and 0.047 for $r = 0.07$. The corresponding results for the stochastic interest rate case with the starting interest rate levels 0.04 and 0.07 are shown in Tables 2 and 4, respectively. The guaranteed rate was not fixed at a constant value throughout the entire contract period but for one year at a time. More specifically, it was set at 0, 1/3 and 2/3 of the short-term rate at intervals of one year. In all the cases, the lag length of the moving average was 125 days, the number of simulated paths was 1000, the number of option price estimations used to determine the fair bonus rate with its upper and lower bounds was 300 (100 for each estimate), and the number of nested simulations used to determine the option price upper bounds was 40.

TABLE 1.

Fair bonus rates and their 95 % confidence intervals in the case of constant interest rate $r = 0.04$.

contract length	guarantee rate	penalty rate	CI lower bound	fair bonus rate	CI upper bound
3	0	0	0.308	0.327	0.387
3	1/3	0	0.217	0.234	0.299
3	2/3	0	0.120	0.133	0.167
3	0	0.02	0.485	0.496	0.509
3	1/3	0.02	0.364	0.368	0.378
3	2/3	0.02	0.201	0.205	0.211
10	0	0	0.295	0.326	0.415
10	1/3	0	0.221	0.259	0.310
10	2/3	0	0.109	0.14	0.177
10	0	0.02	0.452	0.475	0.499
10	1/3	0.02	0.343	0.357	0.374
10	2/3	0.02	0.199	0.206	0.216

TABLE 2.

Fair bonus rates and their 95 % confidence intervals in the case of stochastic interest rate with $r = 0.04$ as the starting level.

contract length	guarantee rate	penalty rate	CI lower bound	fair bonus rate	CI upper bound
3	0	0	0.304	0.322	0.390
3	1/3	0	0.213	0.236	0.302
3	2/3	0	0.115	0.135	0.169
3	0	0.02	0.484	0.496	0.509
3	1/3	0.02	0.363	0.368	0.379
3	2/3	0.02	0.204	0.206	0.213
10	0	0	0.255	0.334	0.443
10	1/3	0	0.197	0.258	0.322
10	2/3	0	0.102	0.141	0.177
10	0	0.02	0.440	0.477	0.503
10	1/3	0.02	0.334	0.361	0.378
10	2/3	0.02	0.198	0.210	0.219

By comparing Tables 1 and 3 (or Tables 2 and 4) one can see that the fair bonus rate is larger when the interest rate is larger. The reason is that the level of the index grows more rapidly when the interest rate is larger, since the 'percentage drift' equals the riskless interest rate under risk-neutral probability. This makes negative returns in the moving average of the stock index less probable, and the feature of the contract which protects

TABLE 3.

Fair bonus rates and their 95 % confidence intervals with constant interest rate $r = 0.07$.

contract length	guarantee rate	penalty rate	CI lower bound	fair bonus rate	CI upper bound
3	0	0	0.481	0.508	0.575
3	1/3	0	0.364	0.386	0.447
3	2/3	0	0.208	0.225	0.277
3	0	0.02	0.675	0.697	0.728
3	1/3	0.02	0.546	0.566	0.584
3	2/3	0.02	0.340	0.344	0.352
10	0	0	0.470	0.507	0.590
10	1/3	0	0.341	0.385	0.459
10	2/3	0	0.210	0.236	0.294
10	0	0.02	0.638	0.683	0.730
10	1/3	0.02	0.516	0.556	0.591
10	2/3	0.02	0.330	0.356	0.370

TABLE 4.

Fair bonus rates and their 95 % confidence intervals in the case of stochastic interest rate with $r = 0.07$ as the starting level.

contract length	guarantee rate	penalty rate	CI lower bound	fair bonus rate	CI upper bound
3	0	0	0.479	0.507	0.570
3	1/3	0	0.354	0.379	0.455
3	2/3	0	0.198	0.223	0.277
3	0	0.02	0.664	0.690	0.719
3	1/3	0.02	0.533	0.555	0.569
3	2/3	0.02	0.325	0.331	0.337
10	0	0	0.446	0.501	0.606
10	1/3	0	0.327	0.383	0.476
10	2/3	0	0.181	0.227	0.297
10	0	0.02	0.606	0.667	0.707
10	1/3	0.02	0.475	0.534	0.565
10	2/3	0.02	0.299	0.330	0.345

the accumulated capital against negative returns becomes less important. This, in turn, decreases the contract price, which is compensated by the increase in the fair bonus rate.

The results in Tables 1 and 2 show that the fixed and stochastic interest rate models with initial interest rate $r = 0.04$ produce similar estimates for the bonus rates. However, Tables 3 and 4 suggest that there is a a systematic difference between the constant and stochastic interest rate models when the initial interest rate is larger ($r = 0.07$). The estimated bonus rates tend to be smaller in the stochastic interest rate model. The reason is probably the mean-reverting property of the interest rate model, which causes the interest rate to decrease during the contract period. This phenomenon is emphasized when there is a penalty in the contract, which in most cases detains optimal stopping until the final expiration date. The increase in the guaranteed rate also magnifies the difference, which may also be explained by the drop in the interest rate.

The result tables indicate that the duration of the contract does not generally affect the bonus rate. However, when the contract includes a penalty and the guaranteed rate is 0 or 1/3 of the short-term interest rate, the fair bonus rate of the 10 years contract seems to be lower than that of the 3 years contract. If the discounted process of the payoff value (i.e., the immediate exercise value) were a martingale for some bonus rate, this would be the fair bonus rate for all maturities. However, this process is martingale only approximately, since the expectation of the future values depends on the path of the index, not only the current value of the savings. Furthermore, the payoff value process is discontinuous because of the penalty conditions.

The confidence intervals are largest with stochastic interest rate and long maturity, and shortest with fixed interest rate and short maturity. Moreover, the length of the confidence interval decreases as the guaranteed rate increases. This can be clearly seen from Figure 4. The figure also reveals that the estimated fair bonus rates are closer to their lower limits than their upper limits when the interest rate is fixed or when the interest rate is stochastic and the guaranteed rate is small. The insurance company probably wishes to set the bonus rate close to its lower limit, and it is good news to the customer that this lower limit is not far from the estimated fair value.

The result tables also show that the confidence intervals are shorter when the penalty is included in the contract. The reason is that the penalty changes the contract closer to a European-style option, which removes the uncertainty related to optimal stopping. Standard errors of the various estimates are shown in Tables 7, 8, 9 and 10 of Appendix B, and they indicate that the estimation errors related to Monte Carlo simulation are also smaller when the penalty is included. On the other hand, the standard errors are similar in the fixed and stochastic interest rate models.

There is an error related to the use of Euler discretization in estimation and simulation. However, the effect of discretization is vanishingly small, since our discretization interval is very short, one working day. If daily data were not available, one could use the high frequency augmentation technique described in Jones (1998) for estimation. On the other hand, it is important to select appropriate index and interest rate models. For example, a failure to choose a realistic model for the stock index might lead to over- or underestimation of volatility, which would make the price estimates biased. Finally, one should note that the regression methods used in determining prices of American options are approximative. In addition to Monte Carlo simulation errors, there is a modelling error related to the choice of regressors. These error sources are taken into account in the confidence intervals.

An important issue which we have not tackled here is the hedging of the contract liability. Our contract is so complicated that it would probably be infeasible to transfer the liability to a third party. This would also increase the overall costs. Instead, the insurer should manage the risk internally by constructing a replicating portfolio. This portfolio

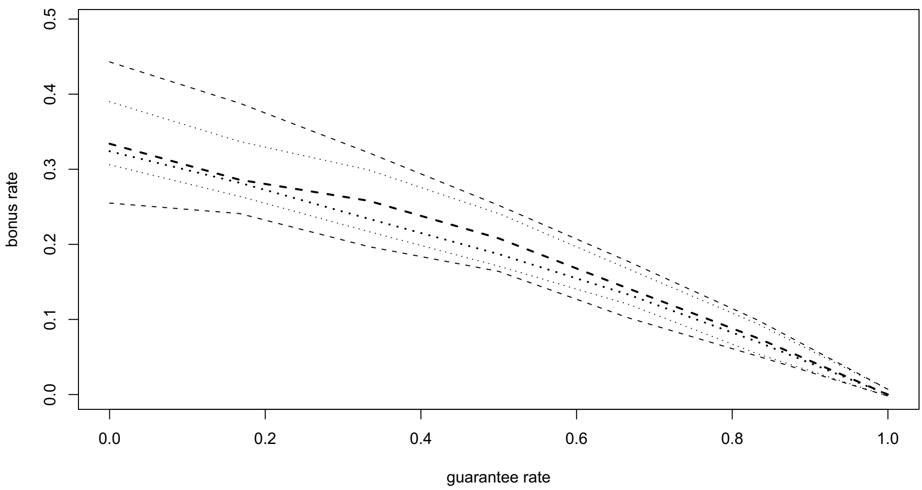


FIG. 4. Fair bonus rate is plotted vs. the proportion of guaranteed rate to riskless rate when $p = 0$ and $r = 0.04$. Dashed lines represent the case with stochastic interest rate and $T = 10$, and dotted lines the case with constant interest rate and $T = 3$. The thick lines represent estimates and the thin lines 95% confidence intervals.

would include investments in a reference fund tracking the stock index, and a money market account yielding the short-term interest rate. If necessary, a bond maturing at the end of the contract period could be included to hedge against the interest rate risk. In Luoma and Puustelli (2009), we have studied hedging our contract with a single-instrument hedge, which employs the underlying stock index and a money market account.

5. CONCLUSIONS

Without sound valuation, economic capital models give a false sense of security. Hence, valuation is the basis of financial risk management. This paper has attempted to provide a full Bayesian analysis of an equity-linked savings contract embedding an American-style path-dependent option in a way which leads to fair valuation. The introduced fairly realistic and flexible valuation framework suits for the design and risk analysis of new products. As a concrete problem we have quantified the effect of the discount rate, guaranteed rate and penalty rate on the fair bonus rate. We have shown how to determine the fair bonus rate and its confidence interval using the regression method. The code needed to utilize the introduced framework e.g. for an internal modeling, can be found at http://mtl.uta.fi/codes/savings.

The Bayesian approach enables us to analyze estimation and model errors, and to take estimation uncertainty into account in the valuation of the contract. Statistical methods, when used appropriately, help detect significant discrepancies between used models and empirical data. This in turn helps curb errors which stem from using inappropriate models, such as a too simple asset model. By studying posterior distributions we found clear evidence that the CEV model, which explicitly allows departures from the geometric Brownian motion, provides a better fit to data. This is an important finding, since we can avoid overestimating volatility, and thus the contract price, by using a more general model.

One of the major findings of our simulation experiments was that the duration of the contract did not generally affect the fair bonus rate. We also found that there were no significant differences in the fair bonus rate between the stochastic and fixed interest rate models when the initial interest rate was set at 4%. When the initial interest rate level was set at a higher level of 7%, the fair bonus rate was estimated to be lower in the stochastic interest rate case. This result suggests that it is more important to use a stochastic model for the interest rate when the initial interest rate is exceptionally low or high.

The accuracy of the fair bonus rate estimates was poorest when the stochastic interest rate model was used, the contract duration was long, and the penalty and guaranteed rates were zero. Although the confidence interval of the bonus rate was fairly long in some cases, the spread between the estimate and the lower confidence limit was reasonably small. This is a good result, since the insurance company would probably set the bonus rate close to its lower limit in order to hedge against the liability.

Since the model fit of the underlying financial time series could be further improved, it would be interesting to study in future research how the use of more sophisticated models, such as stochastic volatility models, would affect the valuation results. The impact of mortality on the results could also be analysed. Furthermore, it would be of practical importance to compare the perfomance of relevant hedging strategies.

ACKNOWLEDGMENTS

The authors of this study are grateful to Vesa Ronkainen for fruitful discussions and crucial suggestions during this project. We would also like to thank anonymous referees for their useful comments, which substantially helped in improving the manuscript.

REFERENCES

Andersen, L. and Broadie, M., 2004. Primal-dual simulation algorithm for pricing multidimensional American options. *Management Science*, 50, 1222–1234.

Bacinello, A.R., Biffis, E., Millossovich, P., 2009. Pricing life insurance contracts with early exercise features. *Journal of Computational and Applied Mathematics*, 233, 27–35.

Ballotta, L., Haberman, S., Wang, N., 2006. Guarantees in with-profit and unitized with-profit life insurance contracts: fair valuation problem in presence of the default option. *The Journal of Risk and Insurance*, 73, 97–121.

Bauer, D., Kiesel, R., Kling, A., Ruß, J., 2006. Risk-neutral valuation of participating life insurance contracts. *Insurance: Mathematics and Economics*, 39, 171–183.

Brigo, D. and Mercurio, F., 2001. *Interest Rate Models Theory and Practice.* Springer.

Briys, E. and de Varenne, F., 1997. On the risk of insurance liabilities: debunking some common pitfalls. *Journal of Risk and Insurance*, 64, 673–694.

Bunnin, F.O., Guo, Y., Ren, Y., 2002. Option pricing under model and parameter uncertainty using predictive densities. *Statistics and Computing*, 12, 37–44.

Chan, K.C., Karolyi, G.A., Longstaff, F.A., Sanders, A.B., 1992. An empirical comparison of alternative models of the short-term interest rate. *The Journal of Finance*, 47, 1209–1227.

Cox, J.C., Ingersoll, J.E., Ross, S.A., 1985. A theory of term structure of interest rates. *Econometrica*, 53, 385–407.

Cox, J.C. and Ross, S.A., 1976. The value of options for alternative stochastic processes. *Journal of Financial Economics*, 3, 145–166.

European Commission, 2009. Directive 2009/138/EC of the European Parliament and of the Council of 25 November 2009 on the taking-up and pursuit of the business of Insurance and Reinsurance (Solvency II).

European Insurance and Occupational Pensions Authority, 2011. Report on variable annuities. EIOPA, 11/031.

Gatzert, N. and Schmeiser, H., 2006. Implicit options in life insurance: Valuation and risk management. *Zeitschrift für die gesamte Versicherungswissenschaft, Ergänzungsband*, 111–128.

Gelman, A., Carlin, J.B., Stern, H.S., Rubin, D.B., 2004. *Bayesian Data Analysis*. Chapman & Hall/CRC.

Glasserman, P., 2004. *Monte Carlo Methods in Financial Engineering*. Springer.

Grosen, A. and Jorgensen, P., 2000. Fair valuation of life insurance liabilities: The impact of interest rate guarantees, surrender optios, and bonus policies. *Insurance: Mathematics and Economics*, 26, 37–57

Hardy, M.R., 2002. Bayesian risk management for equity-linked insurance. *Scandinavian Actuarial Journal*, 3, 185–211.

Hardy, M.R., 2003. *Investment Guaranteed – Modeling and Risk Management for Equity-Linked Life Insurance*. Wiley.

Haugh, M., and Kogan, L., 2004. Pricing American options: A duality approach. *Operations Research*, 52, 258–270.

Iacus, M.I., 2008. *Simulation and Inference for Stochastic Differential Equations with R Examples.* Springer.

International Association of Insurance Supervisors, 2008. Guidance paper on the structure of regulatory capital requirements, Guidance Paper No. 2.1.1.

Jones, C., 1998. Bayesian estimation of continuous-time finance models. Working paper, Simon School of Business, University of Rochester, New York.

Longstaff, F.A. and Schwartz, E.S., 2001. Valuing American options by simulation: A simple least-squares approach. *Review of Financial Studies*, 14, 113–148.

Luoma, A. and Puustelli, A., 2009. Hedging against volatility, jumps and longevity risk in participating life insurance contracts – a Bayesian analysis. In proc. of AFIR Colloquium, Munich, Germany, 8.-11.9.2009.

Metropolis, N., Rosenbluth, A.W., Rosenbluth, M.N., Teller, A.H., Teller, E., 1953. Equations of state calculations by fast computing machines. *Journal of Chemical Physics*, 21, 1087–1092.

Morgan Stanley and Oliver Wyman, 2010. Solvency 2: Quantitative & strategic impact, The tide is going out.

R Development Core Team, 2010. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org/.

Ronkainen, V., Koskinen L., Berglund R., 2007. Topical modelling issues in Solvency II. *Scandinavian Actuarial Journal*, 2, 135–146.

Tanskanen, A.J. and Lukkarinen, J., 2003. Fair valuation of path-dependent participating life insurance contracts. *Insurance: Mathematics and Economics*, 33, 595–609.

Tsitsiklis, J. and Van Roy, B., 1999. Optimal stopping of Markov processes: Hilbert space theory, approximation algorithms, and an application to pricing hig-dimensional financial derivatives. *IEEE Transactions on Automatic Control*, 44, 1840–1851.

Tsitsiklis, J., and Van Roy, B., 2001. Regression methods for pricing complex American-style options. *IEEE Transactions on Neural Networks*, 12, 694–703.

Vasiček, O., 1977. An equilibrium characterization of the term structure. *Journal of Financial Economics*, 5, 177–188.

Zaglauer, K. and Bauer, D., 2008. Risk-neutral valuation of participating life insurance contracts in a stochastic interest rate environment. *Insurance: Mathematics and Economics*, 43, 29–40.

APPENDICES

A. Estimation results

The posterior simulations were performed using the R computing environment. The following output was obtained using the summary function of the add-on package MCMC-pack:

TABLE 5.

Estimation results of the index model with constant interest rate.

```
Number of chains = 3
Sample size per chain = 5000

1. Empirical mean and standard deviation for each variable,
   plus standard error of the mean:

          Mean      SD  Naive SE Time-series SE
mu     0.08723 0.06768 0.0005526       0.001766
log nu 3.41621 0.32460 0.0026503       0.009327
alpha  0.88321 0.05544 0.0004526       0.001592


2. Quantiles for each variable:

          2.5%     25%     50%    75%  97.5%
mu     -0.0448 0.04134 0.08835 0.1316 0.2185
log nu  2.7876 3.20125 3.41310 3.6349 4.0524
alpha   0.7766 0.84630 0.88314 0.9205 0.9914


Gelman and Rubin’s diagnostics
(Potential scale reduction factors):

      Point est. 97.5% quantile
mu          1.00           1.01
log nu      1.01           1.03
alpha       1.01           1.03
```

TABLE 6.

Estimation results of the index model with stochastic interest rate.

```
Number of chains = 3
Sample size per chain = 5000

1. Empirical mean and standard deviation for each variable,
   plus standard error of the mean:

              Mean       SD  Naive SE Time-series SE
mu        0.079225 0.067939 5.547e-04      0.0042595
log nu    3.402534 0.327204 2.672e-03      0.0219129
alpha     0.880626 0.055834 4.559e-04      0.0037286
kappa     0.052439 0.045232 3.693e-04      0.0018596
beta      0.221869 0.132709 1.084e-03      0.0060462
tau^2     0.009487 0.001697 1.386e-05      0.0001046
gamma     0.683214 0.087154 7.116e-04      0.0051778
rho       0.091389 0.025618 2.092e-04      0.0016489

2. Quantiles for each variable:

               2.5%      25%      50%     75%   97.5%
mu        -0.048659 0.026263 0.079805 0.12508 0.21444
log nu     2.769493 3.176798 3.401768 3.61407 4.04341
alpha      0.772006 0.843013 0.881094 0.91714 0.98700
kappa      0.001606 0.018515 0.039505 0.07354 0.16534
beta       0.035210 0.126786 0.200871 0.29001 0.53552
tau^2      0.006695 0.008333 0.009257 0.01045 0.01355
gamma      0.504586 0.627500 0.687341 0.74042 0.84705
rho        0.041503 0.075450 0.090498 0.10661 0.14419


Gelman and Rubin's diagnostics
(Potential scale reduction factors):

           Point est. 97.5% quantile
mu               1.01           1.04
log nu           1.01           1.02
alpha            1.01           1.02
kappa            1.01           1.03
beta             1.01           1.03
tau^2            1.02           1.06
gamma            1.04           1.10
rho              1.01           1.03
```

B. Result tables

TABLE 7.
Fair bonus rate and its lower and upper bounds with standard errors in the case of constant interest rate $r = 0.04$.

length of the contract	guarantee rate	penalty rate	bonus rate	SE of bonus rate	lower bound	SE of lower bound	upper bound	SE of upper bound
3	0	0	0.327	0.009	0.314	0.003	0.365	0.011
3	1/3	0	0.234	0.023	0.227	0.005	0.275	0.012
3	2/3	0	0.133	0.008	0.126	0.003	0.151	0.008
3	0	0.01	0.466	0.006	0.451	0.002	0.484	0.006
3	1/3	0.01	0.357	0.005	0.348	0.002	0.364	0.004
3	2/3	0.01	0.206	0.003	0.204	0.001	0.206	0.003
3	0	0.02	0.496	0.005	0.487	0.001	0.499	0.005
3	1/3	0.02	0.368	0.004	0.366	0.001	0.370	0.004
3	2/3	0.02	0.205	0.003	0.205	0.002	0.205	0.003
10	0	0	0.326	0.009	0.313	0.009	0.372	0.022
10	1/3	0	0.259	0.015	0.245	0.012	0.283	0.014
10	2/3	0	0.14	0.014	0.131	0.011	0.157	0.01
10	0	0.01	0.442	0.006	0.423	0.005	0.468	0.006
10	1/3	0.01	0.333	0.004	0.319	0.003	0.35	0.005
10	2/3	0.01	0.192	0.004	0.187	0.003	0.198	0.005
10	0	0.02	0.475	0.004	0.456	0.002	0.491	0.004
10	1/3	0.02	0.357	0.003	0.347	0.002	0.366	0.004
10	2/3	0.02	0.206	0.003	0.203	0.002	0.208	0.004

TABLE 8.

Fair bonus rate and its lower and upper bounds with standard errors in the case of stochastic interest rate with $r = 0.04$ as the starting level.

length of the contract	guarantee rate	penalty rate	bonus rate	SE of bonus rate	lower bound	SE of lower bound	upper bound	SE of upper bound
3	0	0	0.322	0.011	0.312	0.004	0.366	0.012
3	1/3	0	0.236	0.02	0.225	0.006	0.277	0.013
3	2/3	0	0.135	0.008	0.121	0.003	0.155	0.007
3	0	0.01	0.468	0.007	0.45	0.002	0.484	0.007
3	1/3	0.01	0.356	0.004	0.346	0.001	0.364	0.005
3	2/3	0.01	0.207	0.003	0.203	0.001	0.208	0.003
3	0	0.02	0.496	0.006	0.486	0.001	0.499	0.005
3	1/3	0.02	0.368	0.004	0.365	0.001	0.369	0.005
3	2/3	0.02	0.206	0.003	0.206	0.001	0.207	0.003
10	0	0	0.334	0.09	0.306	0.026	0.4	0.022
10	1/3	0	0.258	0.022	0.228	0.016	0.298	0.012
10	2/3	0	0.141	0.014	0.124	0.011	0.157	0.01
10	0	0.01	0.445	0.005	0.411	0.004	0.47	0.006
10	1/3	0.01	0.334	0.005	0.31	0.002	0.355	0.005
10	2/3	0.01	0.196	0.004	0.18	0.003	0.204	0.005
10	0	0.02	0.477	0.004	0.446	0.003	0.495	0.004
10	1/3	0.02	0.361	0.004	0.338	0.002	0.37	0.004
10	2/3	0.02	0.21	0.003	0.2	0.001	0.213	0.003

TABLE 9.

Fair bonus rate and its lower and upper bounds with standard errors in the case of constant interest rate $r = 0.07$.

length of the contract	guarantee rate	penalty rate	bonus rate	SE of bonus rate	lower bound	SE of lower bound	upper bound	SE of upper bound
3	0	0	0.508	0.013	0.489	0.004	0.548	0.014
3	1/3	0	0.386	0.01	0.37	0.003	0.423	0.012
3	2/3	0	0.225	0.009	0.216	0.004	0.255	0.011
3	0	0.01	0.646	0.01	0.624	0.003	0.673	0.01
3	1/3	0.01	0.528	0.009	0.51	0.003	0.547	0.006
3	2/3	0.01	0.338	0.005	0.327	0.001	0.342	0.004
3	0	0.02	0.697	0.007	0.679	0.002	0.714	0.007
3	1/3	0.02	0.566	0.005	0.55	0.002	0.572	0.006
3	2/3	0.02	0.344	0.004	0.342	0.001	0.344	0.004
10	0	0	0.507	0.013	0.488	0.009	0.555	0.018
10	1/3	0	0.385	0.011	0.37	0.015	0.426	0.017
10	2/3	0	0.236	0.013	0.23	0.01	0.269	0.013
10	0	0.01	0.643	0.009	0.607	0.007	0.683	0.01
10	1/3	0.01	0.519	0.006	0.488	0.004	0.548	0.008
10	2/3	0.01	0.327	0.004	0.306	0.002	0.344	0.004
10	0	0.02	0.683	0.005	0.648	0.005	0.72	0.005
10	1/3	0.02	0.556	0.003	0.522	0.003	0.583	0.004
10	2/3	0.02	0.356	0.003	0.334	0.002	0.364	0.003

TABLE 10.

Fair bonus rate and its lower and upper bounds with standard errors in the case of stochastic interest rate with r = 0.07 as the starting level.

length of the contract	guarantee rate	penalty rate	bonus rate	SE of bonus rate	lower bound	SE of lower bound	upper bound	SE of upper bound
3	0	0	0.507	0.01	0.485	0.003	0.548	0.011
3	1/3	0	0.379	0.014	0.364	0.005	0.426	0.015
3	2/3	0	0.223	0.012	0.206	0.004	0.255	0.011
3	0	0.01	0.636	0.01	0.617	0.003	0.674	0.009
3	1/3	0.01	0.52	0.006	0.499	0.002	0.538	0.007
3	2/3	0.01	0.326	0.004	0.311	0.002	0.33	0.004
3	0	0.02	0.69	0.007	0.668	0.002	0.707	0.006
3	1/3	0.02	0.555	0.004	0.537	0.002	0.559	0.005
3	2/3	0.02	0.331	0.004	0.327	0.001	0.329	0.004
10	0	0	0.501	0.013	0.477	0.016	0.563	0.022
10	1/3	0	0.383	0.012	0.356	0.015	0.435	0.021
10	2/3	0	0.227	0.032	0.205	0.012	0.272	0.013
10	0	0.01	0.628	0.01	0.578	0.006	0.67	0.008
10	1/3	0.01	0.502	0.007	0.454	0.005	0.53	0.006
10	2/3	0.01	0.311	0.004	0.276	0.002	0.328	0.004
10	0	0.02	0.667	0.005	0.614	0.004	0.695	0.006
10	1/3	0.02	0.534	0.004	0.487	0.006	0.555	0.005
10	2/3	0.02	0.33	0.003	0.303	0.002	0.339	0.003

III

Hedging equity-linked life insurance contracts with American-style options in Bayesian framework

Arto Luoma
University of Tampere, Finland

Anne Puustelli
University of Tampere, Finland

We introduce a Bayesian approach to market consistent valuation and hedging of an equity-linked life insurance contract. This paper aims to provide a realistic and flexible modeling tool for product design and risk analysis for insurance companies.

The contract consists of a guaranteed interest rate and a bonus depending on the yield of a total return equity index. A fairly general and realistic framework is assumed, allowing interest rate, volatility and jumps in the asset dynamics to be stochastic, and incorporating stochastic mortality. We employ MCMC methods to estimate both the financial and mortality models, and show how parameter expansion can be effectively applied to estimate the financial model.

Our contract is an American-style path-dependent derivative, and we value it using the regression method. This is combined with an inverse prediction method to determine the fair bonus rate. As a hedging strategy we employ minimum variance hedging which relies on the underlying asset as a single hedging instrument. We compare its hedging effectiveness with a conventional delta-neutral hedge which uses a simpler model for asset dynamics. Parameter uncertainty is taken into account using posterior predictive simulations in valuation and hedging.

Key Words: Option pricing via simulaton, single-instrument hedging, stochastic interest rate, stochastic mortality, stochastic volatility

1. INTRODUCTION

Insurance markets around the world are changing as well as their supervision. In European Union a new risk-based insurance regulatory framework called Solvency II will soon be implemented. Solvency II is designed to reflect modern risk management practices, and the regime assumes a market-consistent valuation of the balance sheet.

In Solvency II insurers are encouraged to measure and manage their risks based on internal models (see, e.g., European Commission, 2009; Gatzert and Schmeiser, 2006; Ronkainen et al., 2007). Further, the European insurance regulator EIOPA (2011) insists that companies involved in complex equity-linked products should use their own internal models for the calculation of solvency capital requirement. On the other hand, Solvecy II will increase the price of more capital-intensive products such as life insurance contracts with capital guarantees. This would result in a search for 'new traditional products' which fulfill the customer demands for traditional life contracts but in a capital-efficient manner (Morgan Stanley and Oliver Wyman, 2010). This paper attempts to serve a need for creating internal models as well as a need to develop new types of products in a market-consistent framework.

There is a wide variety of life insurance products in the market. Most equity-linked life insurance policies, for example variable annuities and equity-indexed annuities in the United States, unit-linked insurance in the United Kingdom and equity-linked insurance in Germany, include implicit options, which represent a significant risk to the company issuing these contracts. One can find a useful introduction to different types of equity-linked insurance contracts in Hardy (2003). Here we do not assume any existing product in the market. The existing products vary considerably and new ones are developed in the future. Instead, our valuation framework is fairly flexible including several financial components which are crucial for the risk analysis of equity-linked life products. Many types of products can be covered from our procedure just by excluding some components of the contract.

Equity-linked and participating life insurance policies are characterized by a vast number of features including, for example, bonus and surrender options and interest rate guarantees. All these features have values which need to be priced. Pricing these policies in a market consistent framework was first studied by Briys and de Varenne (1997a,b). Since then a number of articles have appeared on the topic; see, for example, Grosen and Jorgensen (2000), Tanskanen and Lukkarinen (2003), Bernard et al. (2005), Ballotta et al. (2006), Bauer et al. (2006) and Zaglauer and Bauer (2008). However, most valuation models assume a simplified set-up. Our objective is to present a realistic valuation framework in which the guarantee and the bonus are priced in a stochastic framework, and a surrender option is included in the contract.

The price of an option depends on the assumption of the model describing the behaviour of the underlying instrument. Most approaches specify a particular stochastic process to represent the price dynamics of the underlying asset and then derive an explicit pricing formula. A traditional approach involves solving a partial differential equation. However, when the asset dynamics are assumed to follow a fairly complex model, a closed form solution of the partial differential equation may not exist or its numerical solution may become intractable. When the payoff of an option depends on the path of the underlying asset, the price cannot be evaluated in this manner. Instead, Monte Carlo simulation methods may be used (Glasserman, 2004). For example, Bacinello et al. (2009) apply the least squares Monte Carlo (or regression) method in pricing a participating life insurance with early exercise.

Most papers on pricing equity-linked life insurance contracts lack paying attention to parameter and model errors. Neither the true underlying model nor its parameter values are known. Typically, a relatively simple model is assumed and the point estimates of the parameters are used. This might lead to a crucial valuation error. In the Bayesian approach, parameter and model uncertainty plays a major role. While frequentist methods typically rely on large sample approximations, Bayesian inference is exact in finite samples. In derivative pricing an exact characterization of finite sample uncertainty is critical from the insurance company's risk management point of view. The Bayesian approach is particularly attractive, since it can link the uncertainty of parameters and latent variables to the predictive uncertainty of the process. Another advantage of Bayesian inference is its ability to incorporate prior information into the model.

We define an equity-linked life insurance contract with fairly general features. Specifically, it provides a participation in an equity index at a specified rate, option to surrender at any time, and a downside protection and guaranteed interest rate for the accumulated savings. These properties make the contract a path-dependent American-style derivative, whose pricing and hedging requires using advanced simulation techniques. In particular, we will use the regression approach in pricing (see, e.g., Tsitsiklis and Van Roy, 1999, 2001). The estimated regressions are also used to compute the deltas needed in hedging. The sample paths needed in the regression method are simulated using the posterior predictive distribution under risk-neutral dynamics, as suggested by Bunnin et al. (2002).

The used financial model is also fairly general and realistic, allowing the interest rate, volatility and jumps in the asset dynamics to be stochastic. In estimating this multivariate process we follow the guidelines provided by Jones (1998). However, we do not apply high frequency augmentation, since we are dealing with daily data and the discretization errors in the volatility and interest rate processes are small. Instead, we show how parameter expansion can be effectively used in parameter estimation under the general correlation structure.

Equity-linked life insurance policies involve not only risks arising from financial factors, but also risk related to mortality. Bacinello (2003) and Shen and Xu (2005) introduce mortality risk, but only in a simple set-up with deterministic or constant mortality rates. Biffis (2005) and Bacinello et al. (2009) incorporate stochastic mortality to the pricing framework. With a stochastic mortality model we do not need to make an assumption of a large insurance portfolio, and we avoid invoking to the law of large numbers. This again is significant from the risk management point of view. As a mortality model we will use a generalization of the Gompertz model which takes the cohort effect into account.

In Solvency II insurers are required to cover their risks appropriately, especially as guarantees offered to policyholders imply an increase of risks for insurers. Insurers can manage these risks either through a hedging programme or by additional tools, which include reinsurance and purchase of structured products. With hedging schemes insurers manage risks through financial instruments used for replicating cash flows representing the changes in the market value of liabilities. However, the hedging programme might not be able to properly take into account some of the risks, for example policyholder behaviour and other demographic risks, or the interactions amongst different risks. The insurer may transfer these risks to a reinsurer or to an investment bank. For more, see EIOPA (2011).

We also study dynamic hedging strategies to control for various risks by utilizing a replicating portfolio. As a hedging strategy we employ minimum-variance hedging which relies on the underlying asset as a single hedging instrument. We follow the work by Bakshi et al. (1997) when deriving the minimum-variance hedge. This type of hedge is needed, since a perfect delta-neutral hedge is not feasible due to untraded risks. However, a single-instrument hedge can only be partial, since in our set-up there is more than one source of risk. We also construct a conventional delta-neutral hedge which uses a simpler model for asset dynamics, and compare the performance of the hedges. We find that both methods produce fairly large hedging error standard deviations, which may result from using approximate delta values obtained with the regression method.

The paper is organized as follows. Section 2 introduces the framework and models for the asset dynamics and mortality, Section 3 presents the estimation and evaluation procedures and Section 4 hedging strategies. Section 5 presents the empirical results and the final Section 6 concludes. The full conditional distributions of the financial and mortality models as well as the estimation results are provided in the appendices.

2. THE FRAMEWORK

2.1. The equity-linked life insurance contract

We define the equity-linked life insurance contract as in Luoma et al. (2008). The contract consists of two parts, the first being a guaranteed interest and the second a bonus depending on the yield of some total return equity index. Thus, our product resembles equity-indexed annuities in the United States and equity-linked insurance contracts in Germany. On the other hand, in some equity-linked contracts the bonus is linked to a fund or combinaton of funds, for example in variable annuities in the United States or segregated fund contracts in Canada (see Hardy, 2003).

We denote the amount of savings in the insurance contract at time t_i by $A(t_i)$. Then its growth during a time interval of length $\delta = t_{i+1} - t_i$ is given by

$$\log \frac{A(t_{i+1})}{A(t_i)} = g\,\delta + b \max\left(0, \log \frac{X(t_{i+1})}{X(t_i)} - g\,\delta\right), \tag{1}$$

where $X(t_i) = \sum_{j=0}^{q} S(t_{i-j})/(q+1)$ is a moving average of a total return equity index $S(t_i)$, g is a guaranteed rate and b is a bonus rate, the proportion of the excessive equity index yield which is returned to the customer. In this study we use the time interval $\delta = 1/255$, where 255 is approximately the number of the days in a year on which the index is quoted and the lag length of the moving average is chosen to be $q = 125$ (i.e., half a year). The use of a moving average decreases the volatility of the contract value, and thus facilitates hedging.

The model also incorporates a surrender (early exercise) option and possibility for a penalty p which occurs if the customer reclaims the contract before the final expiration date. If the penalty is set too high, the contract is basically a European-style option and thus exercised only at the end of the contract period. A further condition is that there will be a 1 % penalty if the contract is reclaimed during the first 10 working days. This condition essentially improves the estimation described in Section 3.4. The penalty is not applied if the contract is reclaimed due to mortality.

In the following, we will consider the two cases where (i) the riskless interest rate is fixed at a predetermined value r, or (ii) is assumed to be stochastic. For the constant interest rate r the guaranteed rate g is set at kr throughout the entire contract period for some constant $k < 1$. In the case of stochastic interest rate, the guaranteed rate is fixed for one year at a time. It is set annually at kr_t, where r_t is the riskless short-term interest rate at time t. By setting the guaranteed rate for one year at a time and not daily, the insurance company can better hedge its liabilities, and on the other hand, the customer will have a better idea of the guaranteed growth rate.

In this framework the penalty p for early exercise and the parameters k, g and b are predefined by the insurance company. However, in the case of a stochastic interest rate, g is reset annually. In Section 3.4 we introduce a method to evaluate a fair bonus rate b so that the risk-neutral price of the contract is equal to the initial investment. This gives the contract a simple structure and makes its costs and returns visible and predictable for the insurer and the customer.

The equity-indexed annuity contract has a modification called an annual ratchet in which the index participation is evaluated year by year. Each year the amount of savings is increased by the greater of the floor rate, which is usually 0 percent, and the increase in the underlying index, multiplied by the participation rate. Our contract is similar to this apart from being evaluated on a daily basis. Our contract type is better linked to the dynamics of the financial markets, since the customer may follow the growth of the savings daily and also exercise the contract at market value.

2.2. Financial and mortality models

We assume that the dynamics of the stock index S_t, variance V_t and riskless short-term rate r_t are described by the following system of SDEs:

$$\mathrm{d}\log S_t = \mu \mathrm{d}t + \sqrt{V_t}\mathrm{d}B_t^{(1)} + U_t \mathrm{d}q_t \tag{2a}$$

$$\mathrm{d}V_t = (\alpha_1 + \beta_1 V_t)\mathrm{d}t + \sigma_V \sqrt{V_t}\mathrm{d}B_t^{(2)} \tag{2b}$$

$$\mathrm{d}r_t = (\alpha_2 + \beta_2 r_t)\mathrm{d}t + \sigma_r \sqrt{r_t}\mathrm{d}B_t^{(3)} \tag{2c}$$

where $B_t^{(1)}$, $B_t^{(2)}$ and $B_t^{(3)}$ are standard Brownian motions, q_t is a jump process, and U_t is the jump size. We further assume that these Brownian motions have the correlation structure

$$\mathrm{Cor}\left(B_t^{(1)}, B_t^{(2)}, B_t^{(3)}\right) = \begin{pmatrix} 1 & \rho_{12} & \rho_{13} \\ \rho_{12} & 1 & \rho_{23} \\ \rho_{13} & \rho_{23} & 1 \end{pmatrix}, \tag{3}$$

and q_t is a Poisson process with intensity λ, that is, $\Pr(\mathrm{d}q_t = 1) = \lambda \mathrm{d}t$ and $\Pr(\mathrm{d}q_t = 0) = 1 - \lambda \mathrm{d}t$. Conditional on a jump occurring, we assume that $U_t \sim \mathrm{N}(a, b^2)$. In addition, we assume that q_t is uncorrelated with U_t or with any other process. We abbreviate this model as SVJ-SI.

In order to facilitate estimation, we reparameterize the models (2a) -(2c) as follows:

$$\mathrm{d}\log S_t = \mu \mathrm{d}t + \sigma_1 \sqrt{Y_t}\mathrm{d}B_t^{(1)} + U_t \mathrm{d}q_t \tag{4a}$$

$$\mathrm{d}Y_t = (\alpha_1^* + \beta_1 Y_t)\mathrm{d}t + \sigma_2 \sqrt{Y_t}\mathrm{d}B_t^{(2)} \tag{4b}$$

$$\mathrm{d}R_t = (\alpha_2^* + \beta_2 R_t)\mathrm{d}t + \sigma_3 \sqrt{R_t}\mathrm{d}B_t^{(3)} \tag{4c}$$

where $Y_t = V_t/\sigma_1^2$ is rescaled variance and $R_t = 100\, r_t$ is the interest rate given in percentages. The new parameters are $\alpha_1^* = \alpha_1/\sigma_1^2$, $\sigma_2 = \sigma_V/\sigma_1$, $\alpha_2^* = 100\alpha_2$ and $\sigma_3 = 10\sigma_r$.

We introduce a risk-neutral probability measure $\mathbb{Q}$ under which the discounted index process $\tilde{S}_t = S_t \exp(-\int_0^t r_s ds)$ is a martingale. Specifically, we assume the risk neutral dynamics to be

$$\mathrm{d}S_t = (r_t - \lambda\mu_J)S_t \mathrm{d}t + \sqrt{V_t}S_t \mathrm{d}Z_t^{(1)} + J_t S_t \mathrm{d}q_t \tag{5a}$$

$$\mathrm{d}V_t = (\alpha_1 + \beta_1 V_t)\mathrm{d}t + \sigma_V \sqrt{V_t}\mathrm{d}Z_t^{(2)} \tag{5b}$$

$$\mathrm{d}r_t = (\alpha_2 + \beta_2 r_t)\mathrm{d}t + \sigma_r \sqrt{r_t}\mathrm{d}Z_t^{(3)} \tag{5c}$$

where $J_t = e^{U_t} - 1$, $\mu_J = \mathbb{E}(J_t) = \exp(a + \frac{1}{2}b^2) - 1$, and $Z_t^{(1)}$, $Z_t^{(2)}$ and $Z_t^{(3)}$ are three standard Brownian motions with correlation structure (3) under $\mathbb{Q}$. In logarithmic form, equation (5a) is given by

$$\mathrm{d}\log S_t = \left(r_t - \frac{1}{2}V_t - \lambda\mu_J\right)\mathrm{d}t + \sqrt{V_t}\mathrm{d}Z_t^{(1)} + U_t\mathrm{d}q_t.$$

For the intensity of mortality, we will use a generalization of the Gompertz model. The Gompertz model describes the age dynamics of human mortality fairly accurately in the middle span of ages, approximately between 30 and 80 years, which is enough for our purposes (see, e.g., Promislow, 2006). Specifically, we use a generalization of the form

$$\log(\mu_{ku}) = \beta_{00} + \beta_{01}u + \beta_{10}k + \beta_{11}ku + \epsilon_{ku}, \tag{6}$$

where μ_{ku} is the death rate for age k and for cohort u set by the year of birth. We assume the error term ϵ_{ku} follows an autoregressive process of order one: $\epsilon_{ku} = \phi\epsilon_{k-1,u} + a_{ku}$, where $a_{ku} \sim$ i.i.d. $\mathrm{N}(0, \sigma_m^2)$. The parameters $\beta_{00}, \beta_{01}, \beta_{10}, \beta_{11}$, ϕ and σ_m^2 are unobservable and must be estimated.

3. ESTIMATION AND EVALUATION PROCEDURES

3.1. Estimation of the financial model

We use Bayesian methods to estimate the unknown parameters of the stock index, volatility and interest rate models, as well as to estimate the latent volatility and jump processes. This makes it possible to take parameter uncertainty into account when the fair price of the contract is evaluated and a hedging strategy is employed. The major challenge in estimation is its high dimensionality, which results from the need to estimate latent processes.

We will use Euler discretization in the estimation of unknown parameters, since the transition density of the multivariate process described by (4a), (4b) and (4c) does not have a closed form solution. Accordingly, we will simulate the risk-neutral process using the Euler discretization of (5a), (5b) and (5c). The Euler scheme is the simplest standard method for approximate simulation of stochastic differential equations; for further details, see Iacus (2008) or Glasserman (2004).

A discrete version of (4a), (4b) and (4c) is given by

$$\begin{aligned} \log S_{k+1} &= \log S_k + \mu\delta + \sqrt{Y_k\,\delta}e_{k+1}^{(1)} + U_{k+1}l_{k+1} \\ Y_{k+1} &= Y_k + (\alpha_1^* + \beta_1 Y_k)\delta + \sqrt{Y_k\,\delta}e_{k+1}^{(2)} \\ R_{k+1} &= R_k + (\alpha_2^* + \beta_2 R_k)\delta + \sqrt{R_k\,\delta}e_{k+1}^{(3)} \end{aligned}$$

where δ denotes discretization interval length, $e_k^{(1)}$, $e_k^{(2)}$ and $e_k^{(3)}$ are three normal variables with zero means, variances σ_1^2, σ_2^2 and σ_3^2, and correlation structure (3), $U_k \sim \mathrm{N}(a, b^2)$ is jump size and $l_k \sim \mathrm{Ber}(\lambda\delta)$ an indicator variable of a jump.

Our estimation procedure is a single-component (or cyclic) Metropolis-Hastings algorithm (see, e.g., Gilks et al., 1996). The Metropolis-Hastings (M-H) algorithm is a general term for Markov Chain Monte Carlo (MCMC) methods which are used to simulate posterior distributions. The algorithm was introduced by Hastings (1970) as a generalization of the Metropolis algorithm (Metropolis et al., 1953). Also the Gibbs sampler (Geman and Geman, 1984) can be viewed as its special case.

The single-component M-H algorithm differs from the basic algorithm in that the simulated random vector is divided into components which are updated one by one. The purpose is to simulate the conditional distribution of each block given the current values of the other blocks. In the case of the Gibbs sampler, random variates from these distributions are drawn directly. In the more general case, a proposal from an approximative distribution is first generated, and it is accepted with certain probability, or otherwise the old value is retained.

In the case of our model, it is possible to divide the vector of all parameters into blocks which can be updated using Gibbs sampling, that is, the full conditionals of these blocks can be simulated directly. This is possible, since we have introduced a superfluous parameter σ_1 and we use the general correlation structure. Now posterior simulations of the dispersion matrix of the error vector $(e_k^{(1)}, e_k^{(2)}, e_k^{(3)})$ can be drawn from the Inverse-Wishart density. In order to stabilize simulation, only such draws are accepted for which $0.9 < \sigma_1^2 < 1.1$. Then the scaled variance vector Y is close to the true variance vector V. We estimate the GARCH(1,1) model to obtain an initial value for Y. Further details about the updating procedure are provided in Appendix A.

Note that the data do not contain enough information to estimate σ_1 and Y separately, but their joint posterior distribution determines the posterior of V, which is of interest. Adding a new parameter is called parameter expansion, which can be more generally used to improve the convergence of Markov chain simulation. This is discussed in Liu and Wu (1999), van Dyk and Meng (2001) and Liu (2003), and a simple example is provided by Gelman et al. (2004).

The volatility and jump processes cannot be updated using Gibbs sampling. Here we follow the guidelines provided by Jacquier et al. (1994) and Jones (1998). The scaled variances Y_k are updated one by one. Their full conditional distribution is $p(Y_k|Y_{-k}, H, \phi)$ where Y_{-k} comprises all of Y except Y_k, H comprises the index, interest rate and jump processes, and ϕ is a vector of all parameters. Since we are dealing with a Markov process,

$$p(Y_k|Y_{-k}, H, \phi) \propto p(Y_k|Y_{k-1}, H_{k-1}, H_k, \phi)p(Y_{k+1}, H_{k+1}|Y_k, H_k, \phi).$$

Now Y_k may be updated by first generating a proposal Y_k^* from $p(Y_k|Y_{k-1}, H_{k-1}, H_k, \phi)$ and accepting it with probability

$$\min\left(1, \frac{p(Y_{k+1}, H_{k+1}|Y_k^*, H_k, \phi)}{p(Y_{k+1}, H_{k+1}|Y_k, H_k, \phi)}\right).$$

A detailed description of this update can be found in Appendix A.

The jump process can be updated similarly. Let us denote the joint process of jumps and jump sizes as $I_k = (l_k, U_k)$ and the other processes as $L_k = (S_k, Y_k, R_k)$. Because the jumps are independent, their full conditional is given by $p(I_k|I_{-k}, L, \phi) = p(I_k|L_{k-1}, L_k, \phi)$, which is proportional to

$$p(I_k|\phi)p(L_k|L_{k-1}, I_k, \phi).$$

Now I_k is updated by first generating I_k^* from its marginal distribution $p(I_k|\phi)$ and accepting it with probability

$$\min\left(1, \frac{p(L_k|L_{k-1}, I_k^*, \phi)}{p(L_k|L_{k-1}, I_k, \phi)}\right).$$

The jumps l_k and their sizes U_k could also be updated separately. A detailed description of this update can be found in Appendix A.

3.2. Mortality estimation and prediction

To estimate the mortality model (6) we use Gibbs updates for $\beta = (\beta_{00}, \beta_{01}, \beta_{10}, \beta_{11})$ and σ_u^2, and a Metropolis update for ϕ. The needed conditional posterior distributions can be found in Appendix B. The data are imbalanced in the sense that later cohorts have fewer observations. The unobserved future death rates are considered as missing observations, and they are estimated similarly to the unknown parameters using Gibbs updates. Each missing value is initially given the corresponding death rate from the most recent cohort where it is available. Then the missing values of each cohort are updated by generating them from their multivariate normal conditional distribution.

When the mortality model is used to study hedging performance, we need to scale the estimated AR(1) model, which is based on yearly data, to daily observations. When the sampling frequency is changed from 1 to δ, the high frequency sampling parameters are given as $\phi_{hf} = \phi^\delta$ and $\sigma^2_{m.hf} = \sigma^2_m \frac{1-\phi^{2\delta}}{1-\phi^2}$ (see, e.g., Gourieroux and Jasiak, 2001).

When pricing and hedging the contract, we use the worst-case scenario of mortality from the insurance company's viewpoint. In practice, we simulate 1000 paths of death rates and choose the minimum rate for each time point. These minimum death rates are then used to generate the date of death for each simulation path.

3.3. Pricing American options with the regression method

Our equity-linked life insurance contract is an American-style option with a path-dependent moving average feature. An American option gives the holder the right to exercise the option at any time up to the expiry date. In pricing we adopt the regression (or least squares) method introduced by Tsitsiklis and Van Roy (1999, 2001). It is a simple but powerful approximation method for American-style options. Longstaff and Schwartz (2001) provide a slightly different version of the method.

In this method a number of paths from the underlying (multivariate) process are first generated under the $\mathbb{Q}$-measure. In our application we do not use fixed parameter values but generate the paths from an adjusted version of the posterior predictive distribution. The first adjustment is that μ is replaced with $r_t - V_t/2 - \lambda\mu_J$ in order to obtain risk-neutral dynamics, and the second that we use either 7% or 4% as an initial value of r_t in order to study its effect on the results.

Option pricing is based on an optimal exercising strategy in which the goal is to find a stopping time maximizing the expected discounted payoff of the option. The decision to continue is based on comparing the immediate exercise value with the corresponding continuation value. In the regression method it is assumed that the continuation value may be expressed as a linear regression of the discounted future value on known functions of the current state.

For simplicity, we assume that the value of the insurance contract at maturity, surrender or death is equal to its accumulated savings at the relevant time point. Special policies for these cases could be easily implemented. In the regression method mortality is taken into account so that at each time point only those simulation paths are used in regression for which the person is alive. Bacinello et al. (2009) provide a method to generate times of death, and also a valuation algorithm in which mortality is taken into account. Our approach is slightly different in that it is based on the Tsitsiklis and Van Roy (2001) version of the regression method, while theirs is based on that of Longstaff and Schwartz (2001). The deficit of the latter version is that it in practice produces low-biased price estimates (see Glasserman, 2004).

In our application, the continuation values of the option depend on the path of the underlying index value in a fairly complicated way. Theoretically, we would need $q+1$ state variables (or $q+2$ in the case of stochastic interest rate) to satisfy the Markovian assumption of the process. However, we consider that the current value of the index, its moving average, and the first index value appearing in the moving average may be used to predict the continuation value reasonably well. The use of the moving average may be motivated by observing that the growth of savings in the insurance contract depends on the path of the moving average (see equation 1). The current index value and the first value appearing in the moving average help predict the future evolution of the moving average. We also use the current values of interest rate and volatility to predict the continuation value. The current amount of savings also helps predict the continuation value, but it is not included in the regression variables. Instead, it is subtracted from the regressed value before fitting the regression and subsequently added to the fitted value.

To avoid under- and overflows in the computations, the regression variables related to the equity index are scaled by the first index value, and the current value of the interest rate is given in percentages. Thus, the following state variables are used: $X_1(t_i) = S(t_i)/S(0)$, $X_2(t_i) = \left[\sum_{j=0}^{q} S(t_{i-j})/(q+1)\right]/S(0)$, $X_3(t_i) = S(t_{i-q})/S(0)$, $X_4(t_i) = R(t_i)$ and $X_5(t_i) = V(t_i)$. However, multicollinearity problems would occur if the variables X_1, X_2 and X_3 were used at all time points. In fact, X_3 would be equal for all simulations paths for $i \le q$ and the moving averages X_2 would be very close to each other for small values of i. Therefore, we apply the following rule: The variables X_1, X_4 and X_5 are used for $1 \le i < q/2$, variables X_1, X_2, X_4 and X_5 are used for $q/2 \le i < 3q/2$, and all variables are used for $i \ge 3q/2$.

We use Laguerre polynomials, suggested by Longstaff and Schwartz (2001), as basis functions. More specifically, we use the first two polynomials

$$\begin{aligned} L_0(X) &= \exp(-X/2) \\ L_1(X) &= \exp(-X/2)(1-X) \end{aligned}$$

for all variables. In addition, we use the cross-products $L_0(X_1)L_0(X_4)$, $L_0(X_1)L_0(X_5)$, $L_0(X_1)L_0(X_2)$, $L_0(X_1)L_1(X_2)$, $L_1(X_1)L_0(X_2)$, $L_0(X_1)L_0(X_3)$ and $L_0(X_2)L_0(X_3)$. We also tried adding L_2, and r_t in the case of stochastic interest rate, but these did not improve valuation accuracy. Thus, we have altogether 17 explanatory variables in the regression.

3.4. Determining the fair bonus rate

Using the regression method we can determine the option price (that is, the price of the insurance contract) when the bonus rate b and the guarantee rate g have been given. However, we are interested to determine the bonus rate so that the price of the contract is equal to the initial investment. It makes the different hedging strategies comparable, since the bonus rate affects the optimal time to surrender, which is the most significant factor to produce large hedging errors. If the bonus rate is set at a high level, the contract is almost never reclaimed before the final expiration date, and, on the other hand, if the bonus rate is too low, early surrender is highly probable.

The problem of determining b is a kind of inverse prediction problem, and we need to estimate the option values for various values of b. Since we also wish to estimate the variance of the Monte Carlo error related to the regression method, we repeat the estimation several times for fixed values of b. We end up estimating a regression model where the option price estimates are regressed on the bonus rates. (This regression model should not be confused with the regression method used in the estimation of the option value for a fixed b). We found the third degree polynomial curve to be flexible enough for this purpose. After fitting the curve, we solve the bonus rate b for which the option price is equal to 100, which we assume to be the initial amount of savings. In order to facilitate the estimation of the fair bonus rate, we set the further condition that there is a 1% penalty for reclaiming the contract during the first ten days.

Prior to fitting the polynomial, it is, however, necessary to determine an initial interval for the solution. For this purpose we have developed a modified bisection method. In this method, one first specifies initial upper and lower limits for the bonus rate; we use the values $l = 0$ and $u = 1$. Then one estimates the option price at $(l + u)/2$. If the price is greater than 100, the upper limit of the bonus rate is set at $u - (u - l)/4$; if the price is smaller than 100, the lower limit of the bonus rate is set at $l + (u - l)/4$. This procedure is continued until $u - l = 0.25$. Note that the new limit is not set in the middle of the interval, as is done in the ordinary bisection method, since this might lead to missing the correct solution due to the randomness of the price estimates.

Figure 1 illustrates the estimation procedure. The option price is estimated for 10 different bonus rates, and the estimation is repeated 5 times for each bonus rate, which produces 50 points to the scatter plot. Each estimation is based on 1000 simulated paths. The initial limits of the bonus rate (0.14, 0.39) were determined using the modified bisection method described above. We can see that the fair bonus rate is approximately 28%.

As mentioned above, the bonus rate is solved from the equation $y = f(x)$, where y is the price of the contract and

$$f(x) = \hat{\beta}_0 + \hat{\beta}_1 x + \hat{\beta}_2 x^2 + \hat{\beta}_3 x^3 = \mathbf{x}'\hat{\boldsymbol{\beta}},$$

where $\hat{\boldsymbol{\beta}} = (\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2, \hat{\beta}_3)'$ is the ordinary least squares (OLS) estimate of the cubic regression model and $\mathbf{x} = (1, x, x^2, x^3)'$ a regression vector. The purpose of the initial penalty rate is to ensure that there is exactly one solution in the relevant interval.

Using the delta method, one also obtains an approximate variance for the estimate of x:

$$\mathrm{Var}(\hat{x}) \approx \frac{1}{[f'(x)]^2}\mathrm{Var}(f(x)) \approx \frac{1}{\left(\hat{\beta}_1 + 2\hat{\beta}_2\hat{x} + 3\hat{\beta}_3\hat{x}^2\right)^2}\hat{\mathbf{x}}'\mathrm{Cov}(\hat{\boldsymbol{\beta}})\hat{\mathbf{x}}.$$

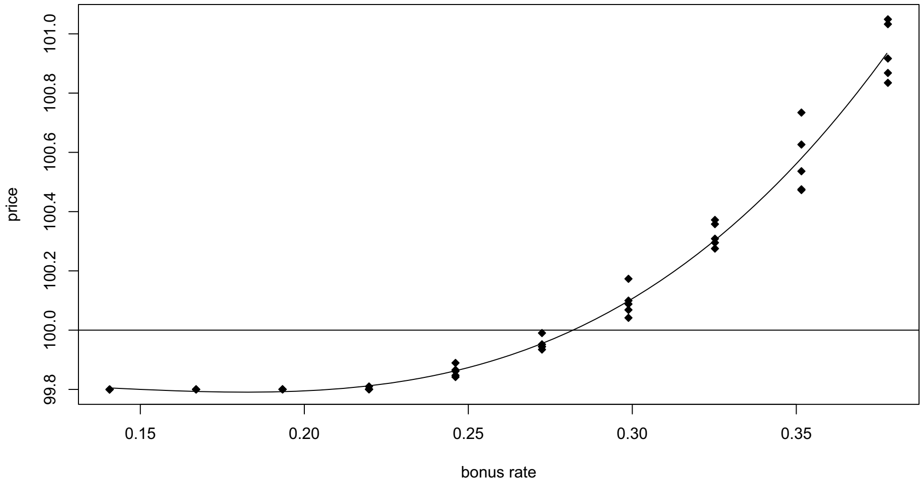


FIG. 1. Contract price estimates vs. bonus rates. The contract length is set at 3 years, guarantee rate at $r_t/3$, and starting level of r_t at 7%, and mortality is not included.

4. HEDGING

4.1. Minimum variance hedging

We construct a single instrument hedge, which employs only the underlying stock index. This hedge is only partial, since there are several sources of risks in our model. Uncontrolled risks are those which move the target option value but are uncorrelated with the underlying index. Such factors as model misspecification and transaction costs may render this type of hedge more practical to adopt than the conventional delta-neutral hedge. Besides a perfect delta-neutral hedge would be infeasible, since some of the risks are untraded.

We assume that there is a reference fund tracking the stock index, and a money market account (MMA) growing at the rate r_t. Let N_t^S be the number of shares to be purchased from the reference fund, and let N_t^0 be the residual position in the MMA. Then the time t value of the replicating portfolio is $N_t^0 + N_t^S S_t$. Furthermore, the hedging error $H_{t+\delta}$ at time $t + \delta$ is given by

$$H_{t+\delta} = N_t^S S_{t+\delta} + N_t^0 e^{r_t \delta} - C_{t+\delta}, \tag{7}$$

where δ is the updating interval of the replicating portfolio, and $C_{t+\delta}$ is the value of the contract at time $t + \delta$. In the limit when $\delta \to 0$, the mean squared hedging error is minimized by choosing

$$N_t^S = \frac{\mathrm{Cov}\,(\mathrm{d}S_t, \mathrm{d}C_t)}{\mathrm{Var}\,(\mathrm{d}S_t)}. \tag{8}$$

Let us denote the jump sizes of the index process S_t by $J_t \doteq e^{U_t} - 1$. The mean and variance of J_t are given by $\mu_J = \exp(a + \frac{1}{2}b^2) - 1$ and $\sigma_J^2 = \exp(2a + b^2)(\exp(b^2) - 1)$, respectively. Under our framework, the total return variance can be decomposed into two components

$$\frac{1}{\mathrm{d}t}\mathrm{Var}\left(\frac{\mathrm{d}S_t}{S_t}\right) = V_t + V_t^J, \tag{9}$$

where the instantaneous variance of the jump component is given by

$$\begin{aligned} V_t^J &= (1/\mathrm{d}t)\mathrm{Var}\,(J_t \mathrm{d}q_t) = (1/\mathrm{d}t)\left(\mathrm{E}(J_t \mathrm{d}q_t)^2 - [\mathrm{E}(J_t)\mathrm{E}(\mathrm{d}q_t)]^2\right) \\ &= (1/\mathrm{d}t)\left[\sigma_J^2 + \left(\mu^J\right)^2\right]\left[\lambda \mathrm{d}t + (\lambda \mathrm{d}t)^2\right] = \lambda\left[\sigma_J^2 + \left(\mu^J\right)^2\right]. \end{aligned}$$

Now let $C_t(S_t, V_t, r_t)$ denote the value of the contract at time t with index value S_t, variance V_t and interest rate r_t. The differential of $C_t(S_t, V_t, r_t)$ may be written as

$$\begin{aligned} \mathrm{d}C_t(S_t, V_t, r_t) = {} & \frac{\partial C_t(S_t, V_t, r_t)}{\partial S_t}\sqrt{V_t}\mathrm{d}Z_t^{(1)} + \frac{\partial C_t(S_t, V_t, r_t)}{\partial V_t}\mathrm{d}V_t + \frac{\partial C_t(S_t, V_t, r_t)}{\partial r_t}\mathrm{d}r_t \\ & + \left[C_t(S_t + J_t S_t, V_t, r_t) - C_t(S_t, V_t, r_t)\right]\mathrm{d}q_t. \end{aligned}$$

Using this and equations (8) and (9) we obtain that

$$\begin{aligned} N_t^S = {} & \Delta_t^{(S)}\frac{V_t}{(V_t + V_t^J)} + \Delta_t^{(V)}\frac{\rho_{12}\sigma_V V_t}{S_t(V_t + V_t^J)} + \Delta_t^{(r)}\frac{\rho_{13}\sigma_r\sqrt{V_t r_t}}{S_t(V_t + V_t^J)} \\ & + \frac{\lambda\left[\mathbb{E}_t\left(J_t C_t(S_t + J_t S_t, V_t, r_t)\right) - C_t(S_t, V_t, r_t)\mu_J\right]}{S_t(V_t + V_t^J)} \end{aligned} \tag{10}$$

where we have denoted the deltas as $\Delta_t^{(S)} = \frac{\partial C_t(S_t,V_t,r_t)}{\partial S_t}$, $\Delta_t^{(V)} = \frac{\partial C_t(S_t,V_t,r_t)}{\partial V_t}$ and $\Delta_t^{(r)} = \frac{\partial C_t(S_t,V_t,r_t)}{\partial r_t}$.

Equation (10) shows that the position to be taken in the shares of the reference fund must control not only for the direct impact of stock price changes on the target option, but also for the indirect impacts of those parts of volatility and interest rate changes which are correlated with index fluctuations. We can see that the additional number of shares needed besides $\Delta^{(S)}$ is increasing both in ρ_{12} and ρ_{13}. Furthermore, since the jump risk is present as well, the position to be taken in the reference fund must also hedge the impact of jump risk on the target option, which is reflected in the last term in (10). This term is increasing in λ and μ_J, meaning that the larger the random-jump risk, the more adjustment needs to be made in the hedging position.

We explained in Section 3.3 how to use the regression method to price our contract, which is an American-style derivative. The estimated regression model is used to determine the optimal stopping times needed in the simulation of hedging performance. For each simulation path the used stopping time is the first time when the estimated continuation value is smaller than the immediate exercise value.

We estimate the deltas $\Delta^{(S)}$, $\Delta^{(V)}$ and $\Delta^{(r)}$ by differentiating the estimated regression equation with respect to S, V and r. The computations are simplified by noting that the derivative of X_3 with respect to S is 0, and the corresponding derivative of X_2 very small, so that it can be ignored. Furthermore, we compute the expectation on the second line of (10) using numerical integration and the estimated regression model.

Comparison of the hedging schemes is based on 1000 simulation paths of the underlying process. For each path, the hedging error is computed as the difference of the replication portfolio and the balance at the estimated optimal stopping time. Then the mean difference, mean squared error and quantiles of the error distribution are computed over all simulation paths. Moreover, this procedure is repeated 100 times using different simulation paths and regression estimates, and the results of these repetitions are pooled. The real-world probability measure $\mathbb{P}$ is used in this simulation, since we are interested in the real-world outcome. The $\mathbb{Q}$-measure is only used for pricing and constructing the hedge portfolio.

Finally, we note that the hedging errors stem from three sources. First, perfect hedging is not possible even in continuous time, since we have more sources of randomness than hedging instruments. Second, since portfolio rebalancing takes place in discrete time, the discretization error given in (7) occurs. Third, since the exact values of the deltas are unknown, we have to use the regression method to estimate them, which adds to the hedging error.

The exact values of the parameters $\sigma_V, \sigma_\rho, \rho_{12}, \rho_{13}, \lambda, a$ and b, and the variance V_t, needed in (10), are also unknown, but this has probably only a minor effect on the hedging error. We use posterior median estimates for the parameters, and update the values of V_t using the GARCH(1,1) model estimated from the original data. In principle we could update the estimates of V_t using Bayesian simulation with the SVJ-SI model, but we found this to be too computationally-intensive for the hedging experiment.

4.2. Competing model and delta-neutral hedging

A similar approach is used when a delta-neutral hedge is constructed for a simpler model. The real-world asset dynamics of this model are described as

$$\mathrm{d}r_t = \kappa(\xi - r_t)\mathrm{d}t + \sigma r_t^\gamma \mathrm{d}W_t^{(1)} \tag{11a}$$

$$\mathrm{d}S_t = \mu S_t \mathrm{d}t + \nu S_t^{1-\alpha}\mathrm{d}W_t^{(2)}. \tag{11b}$$

Here $W_t^{(1)}$ and $W_t^{(2)}$ are two standard Brownian motions, correlated through $W_t^{(2)} = \rho W_t^{(1)} + \sqrt{1-\rho^2}W_t^{(3)}$, where $W_t^{(1)}$ and $W_t^{(3)}$ are two independent standard Brownian motions. The risk neutral dynamics are obtained by replacing the drift μ in (11b) with r_t. Details on estimation and pricing under this model may be found in Luoma et al. (2008). We abbreviate this model as CEV-SI.

In the delta-neutral hedge corresponding to this model, the number of shares in the replication portfolio is given by

$$N_t^S = \frac{\partial C_t(S_t, r_t)}{\partial S_t} = \Delta_t^{(S)} \geq 0.$$

Again we use the regression method to price the derivative and to compute $\Delta^{(S)}$. Here we do not include r_t in the regression variables and thus assume that $\Delta_t^{(r)} \approx 0$.

5. EMPIRICAL RESULTS

5.1. Estimation of the parameters

In order to experiment with actual data and to estimate the unknown parameters of the models, we chose the following data sets: As an equity index we use the Total Return of Dow Jones EURO STOXX Total Market Index (TMI), which is a benchmark covering approximately 95 per cent of the free float market capitalization of Europe. The objective of the index is to provide a broad coverage of companies in the Euro zone including Austria, Belgium, Finland, France, Germany, Greece, Ireland, Italy, Luxembourg, the Netherlands, Portugal and Spain. The index is constructed by aggregating the stocks traded on the major exchanges of Euro zone. Only common stocks and those with similar characteristics are included, and any stocks that have had more than 10 non-trading days during the past three months are removed. In estimation, we use daily quotes from March 4th, 2002 until December 6th, 2007.

As a proxy for riskless short-term interest rate, we use Eurepo, which is the benchmark rate of the large Euro repo market. Eurepo is the rate at which one prime bank offers funds in euro to another prime bank if in exchange the former receives from the latter Eurepo GC as collateral. It is a good benchmark for secured money market transactions in the Euro zone. In the estimation of the interest rate model we use the 3 month Eurepo rate, since it behaves more regularly than the rates with shorter maturities. Both the index and interest series are presented in Figure 2.

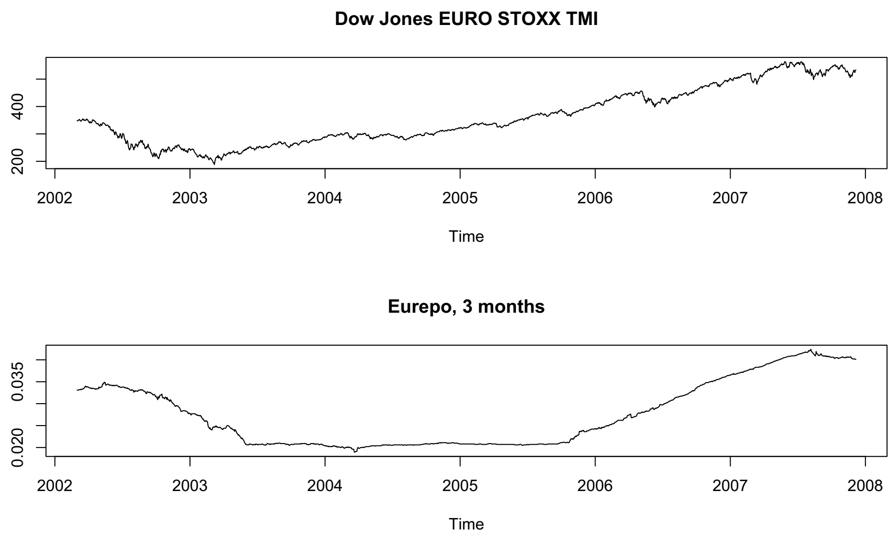


FIG. 2. The equity index and interest series.

In this study we use three alternative ways to model volatility. We apply the SVJ-SI and CEV-SI models for pricing and predictive simulation, while the GARCH(1,1) model is used for quick variance updating in hedging simulation. Figure 3 shows the returns of the used index series and curves indicating 95 % probability limits under these models. We see that the curve of the GARCH model reacts quickly to returns large in absolute value, after which it decays exponentially, while the peaks of the SVJ-SI curve are more symmetric. In the CEV-SI model the variance is a deterministic function of the index level, thus not reacting to the returns.

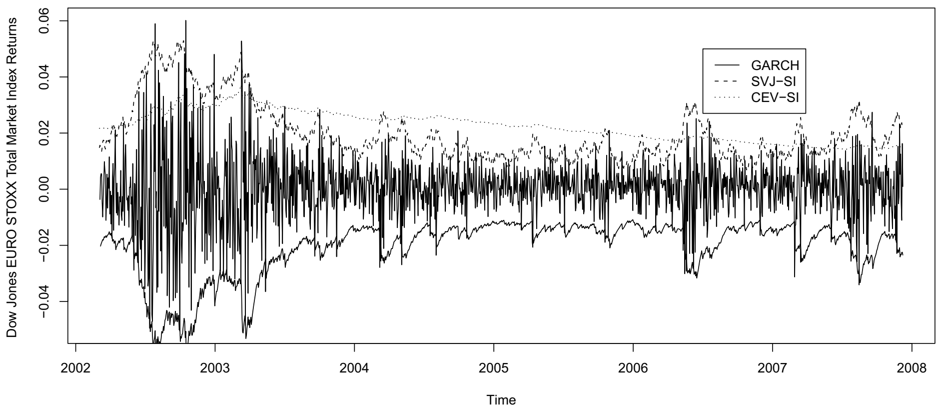


FIG. 3. Three volatility models fitted to the index returns. The lines indicate 2 times fitted volatility.

In mortality modelling we use mortality data provided by the Human Mortality Database (see http:// www.mortality.org). This was created to provide detailed mortality and population data to those interested in the history of human longevity. In our work we use Finnish mortality data for females between ages 30 and 80. More specifically, we use cohort death rates for cohorts born between 1926 and 1961.

All computations were made and figures produced using the R computing environment (R Development Core Team, 2010). To speed up computations we coded the most time consuming loops in C++. We had no remarkable convergence problems in the MCMC simulation used in estimation. Estimation of the finance model (2a)-(2c) was computationally more challenging, and we simulated three chains of length 200000 and picked every 10th simulation to obtain accurate results. In the estimation of the mortality model all chains converged rapidly to their stationary distributions. The summary of the estimation results, as well as Gelman and Rubin's diagnostics (see Gelman et al., 2004), are provided in Appendix C. The values of the diagnostic are close to 1 and thus indicate good convergence.

5.2. Hedging results

There are several parameters which may be varied in the equity-linked life insurance contract described by equation (1). We set the lag length of the moving average at 125 days, the number of simulated paths in contract price estimation at 1000 and the number of estimation repetitions at 100. The hedging results are based on 100 repetitions of 1000 simulation paths. In each case, the optimal stopping rule is based on the "correct" process, that is, the process used in simulating the paths. Furthermore, we set the duration of the contract at 3 or 10 years, and the starting level of interest rate at 4 or 7 percents. We do not fix the guarantee rate at a constant value throughout the entire contract period but set it at 0, 1/3 or 2/3 of the short-term rate at intervals of one year. In each case we assume that the initial investment is 100 (euros).

We calculate the results with and without mortality. When mortality is incorporated into the framework, the age of the insurant is assumed to be 80, and we use cohort data for those born in 1927. Due to simulation and other types of errors it is difficult to detect the effect of mortality, and it would be even more difficult if younger persons were used. Moreover, we consider only cases when the updating frequency of the replicating portfolio is either one day or 20 working days, since the results concerning daily and weekly updates do not differ considerably. Table 1 shows the fair bonus rates and hedging results when minimum variance hedging with SVJ-SI model is used, while Table 2 shows the results when delta-neutral hedging with CEV-SI model is used. Table 3 shows the results of delta-neutral hedging with CEV-SI model when the real-world predictive simulations are generated from the SVJ-SI model, while Table 4 presents the opposite situation. The results for the SVJ-SI model with mortality and 80 year old insurants may be found in Table 5.

TABLE 1.
Fair bonus rates and hedging errors when the SVJ-SI model is used for both hedging and predictive simulation (no mortality).

contract length	interest rate starting level	guarantee rate	fair bonus rate	0.5% VaR daily update	mean error daily update	0.5% VaR 20d update	mean error 20d update
3	4	0	24.1	-1.9	0.07	-2.2	0.08
3	4	1/3	17.3	-1.4	0.04	-1.6	0.04
3	4	2/3	8.8	-0.8	0.04	-0.8	0.04
3	7	0	37.9	-2.9	0.19	-3.4	0.2
3	7	1/3	28.2	-2.3	0.12	-2.7	0.13
3	7	2/3	15.3	-1.3	0.08	-1.5	0.09
10	4	0	26.8	-15.7	-0.09	-15.4	-0.07
10	4	1/3	19.1	-11.1	-0.08	-10.8	-0.06
10	4	2/3	9.2	-3.7	0.04	-3.5	0.05
10	7	0	38.1	-17.5	0.23	-16.6	0.26
10	7	1/3	29	-19.2	0.02	-18.9	0.05
10	7	2/3	15.4	-10.2	0.06	-10	0.07

From all these tables we may see that the estimated fair bonus rate increases as the guarantee rate decreases. This is logical but it is less obvious why the fair bonus rate also increases as the starting level of the interest rate increases. The probable explanation is as follows: When the interest rate is larger the level of the index grows more rapidly, since the 'percentage drift' equals the riskless interest rate under risk-neutral probability. This makes negative returns in the moving average of the stock index less probable, and the feature of the contract which protects the accumulated savings against negative returns becomes less important. This, in turn, decreases the contract price, which is compensated by the increase in the bonus rate.

From Tables 1 and 2 we may see that the mean hedging errors, that is, the mean differences between the replicating portfolio and pay-off values at the optimal stopping time, are very close to zero, as would be expected. Moreover, they are slightly positive in most

TABLE 2.
Fair bonus rates and hedging errors when the CEV-SI model is used for both hedging and predictive simulation (no mortality).

contract length	interest rate starting level	guarantee rate	fair bonus rate	0.5% VaR daily update	mean error daily update	0.5% VaR 20d update	mean error 20d update
3	4	0	32.7	-1.6	0.13	-1.8	0.13
3	4	1/3	23.8	-1.2	0.09	-1.3	0.09
3	4	2/3	13.1	-0.7	0.05	-0.7	0.05
3	7	0	50.2	-2.4	0.21	-2.7	0.21
3	7	1/3	38.3	-1.8	0.15	-2.1	0.15
3	7	2/3	22.6	-1.2	0.07	-1.3	0.07
10	4	0	32.6	-3.8	0.25	-3.7	0.26
10	4	1/3	26.7	-4.4	0	-4.2	0.01
10	4	2/3	14	-1.1	0.04	-1.1	0.04
10	7	0	50.6	-28.4	-0.12	-27.6	-0.11
10	7	1/3	38.3	-11.9	0.02	-11.7	0.03
10	7	2/3	24	-4.4	-0.01	-4.3	-0.01

TABLE 3.
Fair bonus rates and hedging errors when the CEV-SI model is used for hedging and the SVJ-SI model for predictive simulation (no mortality).

contract length	interest rate starting level	guarantee rate	fair bonus rate	0.5% VaR daily update	mean error daily update	0.5% VaR 20d update	mean error 20d update
3	4	0	32.7	-7	-1.01	-7.4	-0.95
3	4	1/3	23.8	-5.1	-0.76	-5.4	-0.72
3	4	2/3	13.1	-2.9	-0.46	-3.1	-0.43
3	7	0	50.2	-11.9	-1.78	-12.6	-1.66
3	7	1/3	38.3	-9.1	-1.35	-9.7	-1.27
3	7	2/3	22.6	-5.5	-1.01	-6	-0.97
10	4	0	32.6	-40.3	-5.09	-40	-4.9
10	4	1/3	26.7	-36.7	-9.34	-36.7	-9.1
10	4	2/3	14	-15.9	-3.73	-15.9	-3.63
10	7	0	50.6	-117.6	-19.44	-116.6	-18.97
10	7	1/3	38.3	-78.3	-14.78	-77.9	-14.44
10	7	2/3	24	-47.6	-12.82	-47.2	-12.62

cases, probably because the stopping rule based on the estimated regression model is slightly suboptimal. The exceptions are probably due to simulation errors.

The 0.5% values at risk (VaR) are substantially worse in the 10 years contracts than in the corresponding 3 years contracts. In the case of the SVJ-SI model the worst figure is around -20, which means that the hedging error is 20% of the initial investment. In most cases, when the guarantee rate increases, the VaR improves. This is understandable, since

TABLE 4.

Fair bonus rates and hedging errors when the SVJ-SI model is used for hedging and the CEV-SI model for predictive simulation (no mortality).

contract length	interest rate starting level	guarantee rate	fair bonus rate	0.5% VaR daily update	mean error daily update	0.5% VaR 20d update	mean error 20d update
3	4	0	24.1	-0.1	0.22	-0.4	0.22
3	4	1/3	17.3	-0.1	0.15	-0.3	0.15
3	4	2/3	8.8	0	0.06	0	0.06
3	7	0	37.9	-0.1	0.35	-0.4	0.35
3	7	1/3	28.2	0	0.23	-0.3	0.23
3	7	2/3	15.3	0	0.11	-0.1	0.11
10	4	0	26.8	-1.8	0.31	-2	0.31
10	4	1/3	19.1	-1	0.22	-1.1	0.21
10	4	2/3	9.2	-0.1	0.08	-0.1	0.08
10	7	0	38.1	-0.1	0.4	-0.5	0.4
10	7	1/3	29	-0.2	0.29	-0.5	0.28
10	7	2/3	15.4	0	0.11	-0.1	0.11

TABLE 5.

Fair bonus rates and hedging errors when the SVJ-SI model is used for both hedging and predictive simulation, and mortality is taken into account.

contract length	interest rate starting level	guarantee rate	fair bonus rate	0.5% VaR daily update	mean error daily update	0.5% VaR 20d update	mean error 20d update
3	4	0	24.5	-2.1	0.05	-2.3	0.06
3	4	1/3	17.3	-1.4	0.05	-1.6	0.05
3	4	2/3	8.9	-0.8	0.04	-0.9	0.04
3	7	0	38	-2.8	0.19	-3.4	0.2
3	7	1/3	28.3	-2.2	0.13	-2.6	0.14
3	7	2/3	15.4	-1.3	0.08	-1.5	0.09
10	4	0	23.6	-4.6	0.16	-4.4	0.17
10	4	1/3	18.7	-8.6	-0.01	-8.4	0
10	4	2/3	9.2	-3	0.04	-2.9	0.05
10	7	0	38.6	-17.4	0.17	-16.3	0.18
10	7	1/3	28.9	-17.2	0.04	-16.7	0.06
10	7	2/3	15.6	-9.6	0.05	-9.4	0.07

a larger guarantee reduces fluctuation in the value of the contract. In the 3 years contracts one can see that the VaRs are slightly poorer when the updating interval is 20 days, while in the 10 years contracts the opposite statement holds. In general, it is not easy to estimate extreme VaRs accurately using simulation; for example, the standard error for the VaR value -20 is around 0.6.

By comparing Tables 1 and 2 one can also see that the fair bonus rates for the CEV-SI model are substantially higher than for the SVJ-SI model. The reason here is that in the CEV model the variance of the return process is a decreasing function of the index level, implying lower predicted variances. This in turn decreases the value of the contract for a fixed bonus rate, which is compensated by an increase in the fair bonus rate. When the higher bonus rate is used and the true process is SVJ-SI, the result is unfavourable for the insurance company. One dramatic consequence of this are the negative mean hedging errors shown in Table 3, which are large in absolute value. The VaRs are also dramatically small, which may be explained by the fact that the optimal stopping time is at the end of the contract period or close to it.

Table 4 shows the results in the opposite situation where the SVJ-SI model is used for hedging and the true process is CEV-SI. In this case the bonus rate is too small, which implies that the hedging errors are positive on average. Moreover, the VaRs are very small in absolute value, since the optimal stopping times are close to the beginning of the contract period.

The effect of mortality can be studied by comparing Tables 1 and 5. In the 3 years contracts the fair bonus rate is usually slightly larger when mortality is taken into account, which is a kind of compensation for suboptimal stopping in the case of death. In the 10 years contracts no systematic difference can be observed, probably because of simulation errors. The hedging results also look similar in both cases.

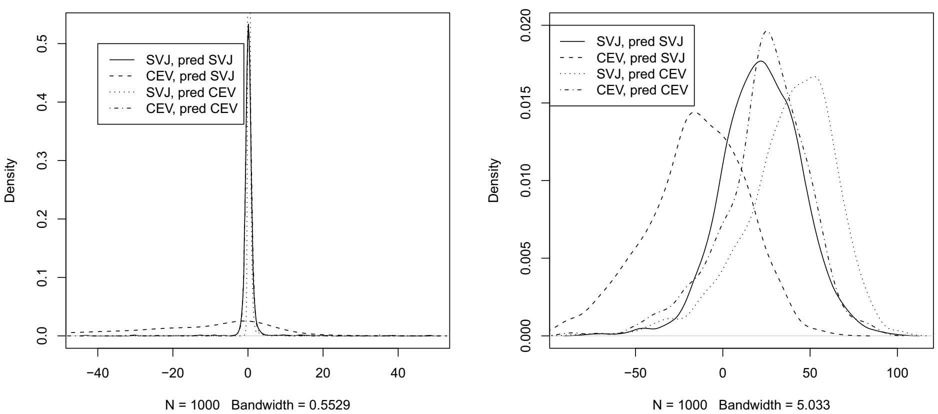


FIG. 4. Difference between hedging portfolio and pay-off value when both SVJ-SI and CEV-SI models are used in hedging and in predictive simulation. It is assumed that the initial interest rate is 7%, guarantee rate 0 and prediction length 10 years, and that there is no mortality. The bonus rate is 38% when the hedging model is SVJ-SI and 50.2% when it is CEV-SI. On the left the difference is calculated at the optimum stopping time, and on the right at the end of the contract period.

From Figure 4 one can see that the hedging error distributions are extremely peaked when the hedging model corresponds to the true process and the hedging error is defined as the difference between the replicating portfolio and the pay-off value at the optimal stopping time. The exceptionally small or large errors may occur when the stopping

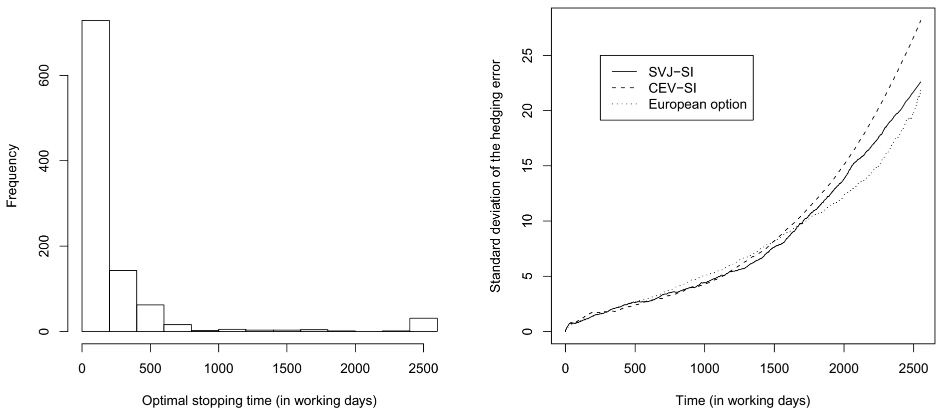


FIG. 5. (a) Distribution of the optimal stopping time in a 10 years contract. It is assumed that the initial interest rate is 7%, guarantee rate 0 and bonus rate 38.6%, and that there is no mortality. The SVJ-SI model is used both in predictive simulation and in determining the optimal stopping time. (b) Standard deviation of the hedging error when SVJ-SI and CEV-SI models are used for the insurance contract, and when the Black-Scholes model is used to hedge a 10 years European call option with initial price 1, strike price 4, interest rate 7% and volatility 20%. In each case, the initial investment is 100 (euros) and the replication portfolio is updated daily. These figures are based on 1000 simulation paths.

time is near the end of the contract period. The right hand side of the figure shows that the hedging error distributions have large standard deviations at the end of the contract period. When the correct model is used for hedging, the mean of the error distribution is slightly positive, since stopping at the end is suboptimal in most cases. When the SVJ-SI model is used for hedging and the true process is CEV-SI, the bonus is too small and the hedging error distribution is further on the positive side. In the opposite case, when the CEV-SI model is used for hedging and the true process is SVJ-SI, the bonus is too large and the hedging error distribution is concentrated on negative values.

From Figure 5 (a) one may see that the optimal stopping times are concentrated in the beginning part of the contract period, but in some cases the stopping is delayed until the end. Figure 5 (b) shows the standard deviation (SD) of the hedging error as a function of time. For comparison, the SD of the hedging error of a European call option with a suitably chosen strike price is also shown. Interestingly, it is close to the SDs of the insurance contract. This vanilla option is not easy to hedge because of its leverage, which makes its value very volatile, while the issue with the insurance contact is that the regression method does not provide delta values accurate enough.

These results indicate that model error might be crucial when hedging an equity-linked life insurance contract. In the worst scenarios the errors would mean huge losses to the insurance company. Small VaRs as such should not be an issue, since the hedging error distribution can be easily sifted to the positive side by decreasing the bonus rate.

We therefore suggest the following two-step approach to choose a sensible bonus rate. First, the theoretical fair bonus rate and the corresponding regression coefficient matrix is determined using the inverse prediction method described in Section 3.4. Second, the inverse prediction method is combined with hedging simulation by utilizing the regression coefficient matrix from the first step in order to determine a bonus rate such that the VaR of the hedging error at the maturity of the contract is acceptable for the insurance company. In order to eliminate the model risk, the SVJ-SI model should be used, since it implies lower bonus rates.

6. CONCLUSIONS

In this paper we present a full Bayesian analysis of valuation and hedging of an equity-linked life insurance contract. The Bayesian approach enables us to exploit MCMC methods and to take parameter uncertainty into account in both valuation and hedging. We value the contract with the regression method, since it embeds an American-style surrender option and no closed-form valuation formula is available. In the valuation we take both financial and mortality risks into account. Two alternative financial models, SVJ-SI and CEV-SI, are utilized. As a stochastic mortality model we use a generalization of the Gompertz model.

The main steps in this paper are the estimation of the financial and mortality models, generation of the posterior predictive distributions, pricing the American-style contract, evaluation of the fair bonus rate, and hedging simulation. Two alternative hedging strategies are employed: first, single-instrument hedging combined with the SVJ-SI model, and, second, conventional delta-neutral hedging combined with the CEV-SI model.

The hedging performances of these alternative strategies turn out to be similar. Probably the effect of the imperfectness of single-instrument hedging is vanishingly small compared to other sources of errors, such as discretization errors and estimation errors of the deltas. However, we find that correct model choice is crucial and that the use of an unrealistic model might lead to catastrophic losses for the insurance company. In particular, if the true data generating process were SVJ-SI, using the CEV-SI model in pricing might lead to losses, since the latter model implies significantly larger bonus rates.

We find that the duration of the contract is the most significant factor to produce large hedging errors. On the contrary, including mortality has only a slight effect on the estimated fair bonus rate and no observable effect on hedging performance. The effect of the updating interval is also small; rebalancing the hedging portfolio monthly produces in practice the same performance as daily updating.

Our results suggest the following two-step procedure to choose a sensible bonus rate: first, the theoretical fair bonus rate is determined, and second, it is adjusted so that the VaR of the hedging error becomes acceptable for the insurance company.

ACKNOWLEDGMENTS

The authors of this article are grateful to Vesa Ronkainen for introducing us to the challenges of life insurance contracts. They are also grateful to Lasse Koskinen for fruitful discussions and useful comments during this project. The possibility to use the computing resources of CSC, IT Center for Science Ltd, is gratefully acknowledged. The second author would like to thank the Finnish Academy of Science and Letters, Väisälä Fund, for the scholarship during which she could complete this project.

REFERENCES

Bacinello, A., 2003. Fair valuation of a guaranteed life insurance participating contract embedding a surrender option. *The Journal of Risk and Insurance*, 70, 461–487.

Bacinello, A.R., Biffis, E., Millossovich, P., 2009. Pricing life insurance contracts with early exercise features. *Journal of Computational and Applied Mathematics*, 233, 27–35.

Bakshi, G., Cao, C., Chen, Z., 1997. Empirical performance of alternative option pricing models. *Journal of Finance*, 52, 2003–2049.

Ballotta, L., Haberman, S., Wang, N., 2006. Guarantees in with-profit and unitized with-profit life insurance contracts: fair valuation problem in presence of the default option. *The Journal of Risk and Insurance*, 73, 97–121.

Bauer, D., Kiesel, R., Kling, A., Ruß, J., 2006. Risk-neutral valuation of participating life insurance contracts. *Insurance: Mathematics and Economics*, 39, 171–183.

Bernard, C., Le Courtois, O., Quittard-Pinon, F., 2005. Market value of life insurance contracts under stochastic interest rates and default risk. *Insurance: Mathematics and Economics*, 36, 499–516.

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

Briys, E. and de Varenne, F., 1997a. On the risk of previous termlife insurancenext term liabilities: debunking some common pitfalls. *Journal of Risk Insurance*, 64, 673–694.

Briys, E. and de Varenne, F., 1997b. Valuing risky fixed rate debt: an extension. *Journal of Financial and Quantitative Analysis*, 32, 239–248.

Bunnin, F.O., Guo, Y., Ren, Y., 2002. Option pricing under model and parameter uncertainty using predictive densities. *Statistics and Computing*, 12, 37–44.

European Commission, 2009. Directive 2009/138/EC of the European Parliament and of the Council of 25 November 2009 on the taking-up and pursuit of the business of Insurance and Reinsurance (Solvency II).

European Insurance and Occupational Pensions Authority, 2011. Report on variable annuities. EIOPA, 11/031.

Gatzert, N. and Schmeiser, H., 2006. Implicit options in life insurance: Valuation and risk management. *Zeitschrift für die gesamte Versicherungswissenschaft, Ergänzungsband*, 111–128.

Gelman, A., Carlin, J.B., Stern, H.S., Rubin, D.B., 2004. *Bayesian Data Analysis*. Chapman & Hall/CRC.

Geman, S. and Geman, D., 1984. Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 6, 721–741.

Gilks, W. R., Richardson, S., Spiegelhalter, D. (eds.), 1996. *Markov Chain Monte Carlo in Practice*. Chapman & Hall.

Glasserman, P., 2004. *Monte Carlo Methods in Financial Engineering*. Springer.

Gourieroux, C. and Jasiak, J., 2001. *Financial Econometrics*. Princeton Series in Finance.

Grosen, A. and Jorgensen, P., 2000. Fair valuation of life insurance liabilities: The impact of interest rate guarantees, surrender options, and bonus policies. *Insurance: Mathematics and Economics*, 26, 37–57

Hardy, M.R., 2003. *Investment Guaranteed – Modeling and Risk Management for Equity-Linked Life Insurance*. Wiley.

Hastings, W.K., 1970. Monte Carlo sampling methods using Markov chains and their applications. *Biometrica*, 57, 97–109.

Iacus, M.I., 2008. *Simulation and Inference for Stochastic Differential Equations with R Examples.* Springer.

Jacquier, E., Polson, N.G., Rossi, P.E., 1994. Bayesian analysis of stochastic volatility models. *Journal of Business and Economic Statistics*, 12, 371–389.

Jones, C., 1998. Bayesian estimation of continuous-time finance models. Working paper, Simon School of Business, University of Rochester, New York.

Liu, C., 2003. Alternating subspace-spanning resampling to accelerate Markov chain Monte Carlo simulation. *Journal of the American Statistical Association*, 98, 110–117.

Liu, J., and Wu, Y. N., 1999. Parameter expansion in for data augmentation. *Journal of the American Statistical Association*, 94, 1264–1274.

Longstaff, F.A. and Schwartz, E.S., 2001. Valuing American options by simulation: A simple least-squares approach. *Review of Financial Studies*, 14, 113–148.

Luoma, A., Puustelli, A., Koskinen, L., 2008. Bayesian analysis of participating life insurance contracts with American-style options. In Proc. of AFIR Colloquium, Rome, Italy, 31 September – 3 October, 2008.

Metropolis, N., Rosenbluth, A.W., Rosenbluth, M.N., Teller, A.H., Teller, E., 1953. Equations of state calculations by fast computing machines. *Journal of Chemical Physics*, 21, 1087–1092.

Morgan Stanley and Oliver Wyman, 2010. Solvency 2: Quantitative & strategic impact, The tide is going out.

R Development Core Team, 2010. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org/.

Ronkainen, V., Koskinen L., Berglund R., 2007. Topical modelling issues in Solvency II. *Scandinavian Actuarial Journal*, 2, 135–146.

Promislow, S.D., 2006. *Fundamentals of Actuarial Mathematics*. John Wiley & Sons, Ltd.

Shen, W. and Xu, H., 2005. The valuation of unit-linked policies with or without surrender options. *Insurance: Mathematics and Economics*, 36, 79–92.

Tanskanen, A.J. and Lukkarinen, J., 2003. Fair valuation of path-dependent participating life insurance contracts. *Insurance: Mathematics and Economics*, 33, 595–609.

Tsitsiklis, J. and Van Roy, B., 1999. Optimal stopping of Markov processes: Hilbert space theory, approximation algorithms, and an application to pricing hig-dimensional financial derivatives. *IEEE Transactions on Automatic Control*, 44, 1840–1851.

Tsitsiklis, J. and Van Roy, B., 2001. Regression methods for pricing complex American-style options. *IEEE Transactions on Neural Networks*, 12, 694–703.

van Dyk, D. A. and Meng, X. L., 2001. The art of data augmentation (with discussion). *Journal of Computational and Graphical Statistics*, 10, 1–111.

Zaglauer, K. and Bauer, D., 2008. Risk-neutral valuation of participating life insurance contracts in a stochastic interest rate environment. *Insurance: Mathematics and Economics*, 43, 29–40.

APPENDICES

A. Full conditional distributions of the option pricing model

Let us denote $\gamma_1 = (\alpha_1^*, \beta_1)$, $\gamma_2 = (\alpha_2^*, \beta_2)$, $\phi = (\mu, \gamma_1, \gamma_2, \sigma_1, \sigma_2, \sigma_3, \rho_{12}, \rho_{13}, \rho_{23})$, $I_k = (l_k, U_k)$, $Y = (Y_1, \dots Y_{K-1})$ and

$$e_k^{(1)} = \frac{\log S_k - \log S_{k-1} - \mu\delta - l_k U_k}{\sqrt{Y_{k-1}\delta}},$$

$$e_k^{(2)} = \frac{Y_k - Y_{k-1} - \left(\alpha_1^* + \beta_1 Y_{k-1}\right)\delta}{\sqrt{Y_{k-1}\delta}},$$

$$e_k^{(3)} = \frac{R_k - R_{k-1} - \left(\alpha_2^* + \beta_2 R_{k-1}\right)\delta}{\sqrt{R_{k-1}\delta}},$$

$$e_k = \begin{pmatrix} e_k^{(1)} \\ e_k^{(2)} \\ e_k^{(3)} \end{pmatrix},$$

and

$$\Sigma = \text{Cov}\left(e_k^{(1)}, e_k^{(2)}, e_k^{(3)}\right) = \begin{pmatrix} \sigma_1^2 & \sigma_{12} & \sigma_{13} \\ \sigma_{12} & \sigma_2^2 & \sigma_{23} \\ \sigma_{13} & \sigma_{23} & \sigma_3^2 \end{pmatrix}.$$

Updating μ with a Gibbs step

Prior: $p(\mu) \propto 1$
Conditional posterior:

$$\{\mu | z^{(1)}, Y, \sigma_{1.23}\} \sim \text{N}\left(\frac{\sum_{k=1}^{K-1} \frac{z_{k+1}^{(1)}}{Y_k}}{\sum_{k=1}^{K-1} \frac{1}{Y_k}}, \frac{\sigma_{1.23}}{\delta \sum_{k=1}^{K-1} \frac{1}{Y_k}}\right),$$

where

$$z^{(1)} = \left(z_2^{(1)}, z_3^{(1)}, \dots, z_K^{(1)}\right),$$

$$z_{k+1}^{(1)} = \frac{\log(S_{k+1}) - \log(S_k) - l_{k+1}U_{k+1}}{\delta} - \sqrt{\frac{Y_k}{\delta}}\mu_{k+1}^{(1.23)},$$

$$\mu_{k+1}^{(1.23)} = \begin{pmatrix} \sigma_{12} & \sigma_{13} \end{pmatrix} \begin{pmatrix} \sigma_2^2 & \sigma_{23} \\ \sigma_{23} & \sigma_3^2 \end{pmatrix}^{-1} \begin{pmatrix} e_{k+1}^{(2)} \\ e_{k+1}^{(3)} \end{pmatrix},$$

and

$$\sigma_{1.23} = \sigma_1 - \begin{pmatrix}\sigma_{12} & \sigma_{13}\end{pmatrix}\begin{pmatrix}\sigma_2^2 & \sigma_{23}\\ \sigma_{23} & \sigma_3^2\end{pmatrix}^{-1}\begin{pmatrix}\sigma_{12}\\ \sigma_{13}\end{pmatrix}.$$

Updating γ_1 with a Gibbs step

Prior: $p(\gamma_1) \propto 1$
Conditional posterior:

$$\{\gamma_1 | z^{(2)}, Y, \sigma_{2.13}\} \sim \mathrm{N}\left(\left(X'\Delta^{-1}X\right)^{-1} X'\Delta^{-1}z^{(2)}, \frac{\sigma_{2.13}}{\delta}\left(X'\Delta^{-1}X\right)^{-1}\right),$$

where

$$z^{(2)} = \left(z_2^{(2)}, z_3^{(2)}, \ldots, z_K^{(2)}\right),$$

$$z_{k+1}^{(2)} = \frac{Y_{k+1} - Y_k}{\delta} - \sqrt{\frac{Y_k}{\delta}}\mu_{k+1}^{(2.13)},$$

$$X = \begin{pmatrix}1 & Y_1\\ 1 & Y_2\\ \vdots & \vdots\\ 1 & Y_{K-1}\end{pmatrix},$$

$$\Delta^{-1} = \begin{pmatrix}\frac{1}{Y_1} & 0 & \ldots & 0\\ 0 & \ddots & & \vdots\\ \vdots & & \ddots & 0\\ 0 & \ldots & 0 & \frac{1}{Y_{K-1}}\end{pmatrix},$$

$$\mu_{k+1}^{(2.13)} = \begin{pmatrix}\sigma_{12} & \sigma_{23}\end{pmatrix}\begin{pmatrix}\sigma_1^2 & \sigma_{13}\\ \sigma_{13} & \sigma_3^2\end{pmatrix}^{-1}\begin{pmatrix}e_{k+1}^{(1)}\\ e_{k+1}^{(3)}\end{pmatrix},$$

and

$$\sigma_{2.13} = \sigma_2^2 - \begin{pmatrix}\sigma_{12} & \sigma_{23}\end{pmatrix}\begin{pmatrix}\sigma_1^2 & \sigma_{13}\\ \sigma_{13} & \sigma_3^2\end{pmatrix}^{-1}\begin{pmatrix}\sigma_{12}\\ \sigma_{23}\end{pmatrix}.$$

Updating γ_2 with a Gibbs step

Prior: $p(\gamma_2) \propto 1$
Conditional posterior:

$$\{\gamma_2 | z^{(3)}, X_*, \sigma_{3.12}\} \sim \mathrm{N}\left(\left(X_*'\Delta_*^{-1}X_*\right)^{-1} X_*'\Delta_*^{-1}z^{(3)}, \frac{\sigma_{3.12}}{\delta}\left(X_*'\Delta_*^{-1}X_*\right)^{-1}\right),$$

where

$$z^{(3)} = \left(z_2^{(3)}, z_3^{(3)}, \ldots, z_K^{(3)}\right),$$

$$z_{k+1}^{(3)} = \frac{R_{k+1} - R_k}{\delta} - \sqrt{\frac{R_k}{\delta}}\mu_{k+1}^{(3.12)},$$

$$X_* = \begin{pmatrix} 1 & R_1 \\ 1 & R_2 \\ \vdots & \vdots \\ 1 & R_{K-1} \end{pmatrix},$$

$$\Delta_*^{-1} = \begin{pmatrix} \frac{1}{R_1} & 0 & \ldots & 0 \\ 0 & \ddots & & \vdots \\ \vdots & & \ddots & 0 \\ 0 & \ldots & 0 & \frac{1}{R_{K-1}} \end{pmatrix},$$

$$\mu_{k+1}^{(3.12)} = \begin{pmatrix} \sigma_{13} & \sigma_{23} \end{pmatrix} \begin{pmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{pmatrix}^{-1} \begin{pmatrix} e_{k+1}^{(1)} \\ e_{k+1}^{(2)} \end{pmatrix},$$

and

$$\sigma_{3.12} = \sigma_3^2 - \begin{pmatrix} \sigma_{13} & \sigma_{23} \end{pmatrix} \begin{pmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{pmatrix}^{-1} \begin{pmatrix} \sigma_{13} \\ \sigma_{23} \end{pmatrix}.$$

Updating Σ with a Gibbs step

Prior: $p(\Sigma) \propto \text{Inv-Wishart}(\Sigma|\Psi, m) \times I_{(0.9,1.1)}(\sigma_1^2)$
Posterior: $p(\Sigma|\mu, \gamma_1, \gamma_2, S, Y, R, I) \propto \text{Inv-Wishart}(\Sigma|\Psi + A, m + K - 1) \times I_{(0.9,1.1)}(\sigma_1^2)$, where

$$A = \sum_{k=2}^{K} e_k e_k'.$$

Updating volatility with Metropolis-Hastings steps

Let us denote $H_k = (\log S_k, I_k, R_k)$. The conditional posterior of Y_k:

$$\begin{aligned} & p\left(Y_k|Y_{k-1}, Y_{k+1}, H_{k-1}, H_k, H_{k+1}, \phi\right) \\ & \propto p\left(Y_k|Y_{k-1}, H_{k-1}, H_k, \phi\right) p\left(Y_{k+1}, H_{k+1}|Y_k, H_k, \phi\right) \end{aligned}$$

Proposal Y_k^* is generated from $p\left(Y_k|Y_{k-1}, H_{k-1}, H_k, \phi\right)$:

$$Y_k^* = Y_{k-1} + (\alpha_1^* + \beta_1 Y_{k-1})\delta + \sqrt{Y_{k-1}\delta}\, e_k^{(2)*},$$

where $e_k^{(2)*} \sim \mathrm{N}\left(\mu_k^{(2.13)}, \sigma_{2.13}\right)$. For $k = 1$ the proposal is generated from unconditional distribution $\Pr(Y_1|\phi)$. Since Y is a CIR process, its stationary distribution is $\mathrm{Gamma}\left(-\frac{2\alpha_1^*}{\sigma_2^2}, -\frac{\sigma_2^2}{2\beta_1}\right)$. Acceptance probability:

$$\min\left(1, \frac{p(Y_{k+1}, H_{k+1}|Y_k^*, H_k, \phi)}{p(Y_{k+1}, H_{k+1}|Y_k, H_k, \phi)}\right)$$
$$= \min\left\{1,\ \exp\left[-\log(Y_k^*) + \log(Y_k) - \frac{1}{2}\left(e_{k+1}^{*\prime}\Sigma^{-1}e_{k+1}^* - e_{k+1}'\Sigma^{-1}e_{k+1}\right)\right]\right\},$$

where e_{k+1}^* is computed using Y_k^*. For $k = K$ the acceptance probability cannot be computed. The proposal is accepted with probability 1.

Updating the parameters of the jump process with Gibbs steps

Prior: $\lambda_0 \sim \mathrm{Beta}(p_1, p_2)$
Posterior: $\Pr(\lambda_0|I) \propto \mathrm{Beta}(p_1 + \sum l_i, p_2 + K - \sum l_i)$
Priors: $b^2 \sim \text{Inv-}\chi^2(df_0, \sigma_0^2)$, $\quad a|b^2 \sim \mathrm{N}(a_0, b^2/b_0)$
Posteriors:

$$b^2|l, U \sim \text{Inv-}\chi^2\left(df_0 + n, \frac{1}{df_0 + n}\left(df_0\sigma_0^2 + (n-1)s^2 + \frac{b_0 n}{b_0 + n}(\bar{U} - a_0)^2\right)\right)$$

$$a|b^2, l, U \sim \mathrm{N}\left(\frac{b_0 a_0 + n\bar{U}}{b_0 + n}, \frac{b^2}{b_0 + n}\right)$$

where $n = \sum l_i$, $\bar{U} = \frac{1}{n}\sum l_i U_i$, $s^2 = \frac{1}{n-1}\sum l_i(U_i - \bar{U})^2$.

Updating the jump process with Metropolis-Hastings steps

Let us denote and $L_k = (\log S_k, Y_k, R_k)$. Then the full conditional distribution of I_k is

$$p\left(I_k|I_{k-1}, I_{k+1}, L_k, L_{k-1}, L_{k+1}, \phi\right) = p\left(I_k|L_{k-1}, L_k, \phi\right) \propto$$
$$p\left(I_k|L_{k-1}, \phi\right) p\left(L_k|I_k, L_{k-1}, \phi\right) = p\left(I_k|\phi\right) p\left(L_k|I_k, L_{k-1}, \phi\right)$$

Proposal from distribution $p\left(I_k|\phi\right)$: $l_k^* \sim \mathrm{Ber}(\lambda_0)$, $U_k^* \sim \mathrm{N}(a, b^2)$.
Acceptance probability:

$$\min\left(1, \frac{p(L_k|L_{k-1}, I_k^*, \phi)}{p(L_k|L_{k-1}, I_k, \phi)}\right)$$
$$= \min\left\{1,\ \exp\left[-\frac{1}{2}\left(e_k^{*\prime}\Sigma^{-1}e_k^* - e_k'\Sigma^{-1}e_k\right)\right]\right\},$$

where e_k^* is computed using I_k^*.

B. Full conditional posterior distributions of the mortality model

Let us denote $y_{ku} = \log(\mu_{ku})$ for $k = 1, ..., K$ and $u = 1, ..., U$. Furthermore, $y_u = (y_{1u}, ..., y_{Ku})$, $y = (y_1, ..., y_U)$ and $X = (X_1', \ldots, X_U')'$, where

$$X_u = \begin{pmatrix} 1 & u & 1 & u \cdot 1 \\ 1 & u & 2 & u \cdot 2 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & u & K & u \cdot K \end{pmatrix},$$

and $\beta = (\beta_{00}, \beta_{01}, \beta_{10}, \beta_{11})$.

The inverse of $Cor(y|\beta, \phi)$ is

$$R_*^{-1} = \begin{pmatrix} R^{-1} & & 0 \\ & \ddots & \\ 0 & & R^{-1} \end{pmatrix} = I_U \otimes R^{-1}$$

where I_U is $U \times U$ identity matrix and

$$R = \begin{pmatrix} 1 & \phi & \phi^2 & \ldots & \phi^{K-1} \\ \phi & 1 & \phi & \ldots & \phi^{K-2} \\ \vdots & & \ddots & & \vdots \\ \phi^{K-1} & \phi^{K-2} & \ldots & \phi & 1 \end{pmatrix}.$$

Updating σ_m^2 with a Gibbs step

Prior: $p(\sigma_m^2) \propto \frac{1}{\sigma_m^2}$
Conditional posterior: $\sigma^2|y, \beta, \phi \sim \text{Inv-}\chi^2(KU, \text{SS}/KU)$, where $\text{SS} = (y - X\beta)' R_*^{-1} (y - X\beta)$.

Updating β with a Gibbs step

Prior: $p(\beta) \propto 1$
Conditional posterior: $\beta|y, \phi, \sigma_m^2 \sim \text{N}(\mu_\beta, \sigma_m^2 V_\beta)$, where $\mu_\beta = (X' R_*^{-1} X)^{-1} X' R_*^{-1} y$ and $V_\beta = (X' R_*^{-1} X)^{-1}$.

Updating ϕ with a Metropolis step

Prior: $p(\phi) = I_{(-1,1)}(\phi)$
Conditional posterior: $p(\phi|y, \beta, \sigma_m^2) \propto (1 - \phi^2)^{-\frac{1}{2}U(K-1)} \exp\left(-\frac{1}{2\sigma_m^2} SS\right) I_{(-1,1)}(\phi)$

III

C. Estimation results of the financial and mortality models

The posterior simulations were performed using the R computing environment. The following outputs were obtained using the summary function of the add-on package MCMCpack:

TABLE 6.

Estimation results of the financial model.

```
Number of chains = 3
Sample size per chain = 10000

1. Empirical mean and standard deviation for each variable,
   plus standard error of the mean:

               Mean        SD  Naive SE Time-series SE
mu        0.1157251 0.0365450 2.110e-04      7.554e-04
alpha1    0.1968175 0.0386987 2.234e-04      1.658e-03
beta1    -6.2996549 1.3530007 7.812e-03      5.483e-02
alpha2    0.2154899 0.1286814 7.429e-04      7.452e-04
beta2    -0.0495622 0.0429811 2.482e-04      2.270e-04
sigma22V  0.2218156 0.0444790 2.568e-04      2.137e-03
sigma33   0.0140068 0.0005198 3.001e-06      3.108e-06
rho12    -0.7681546 0.0523322 3.021e-04      2.070e-03
rho13     0.0792596 0.0264295 1.526e-04      2.204e-04
rho23    -0.1322315 0.0517667 2.989e-04      1.381e-03
a        -0.0060781 0.0068594 3.960e-05      1.366e-04
b2        0.0003484 0.0002063 1.191e-06      2.366e-06
lambda0   0.0090029 0.0027855 1.608e-05      6.318e-05
```

2. Quantiles for each variable:

	2.5%	25%	50%	75%	97.5%
mu	0.0426332	0.0911853	0.1162357	0.1406814	0.1859535
alpha1	0.1271985	0.1697973	0.1944505	0.2216304	0.2790440
beta1	-9.1239610	-7.1682486	-6.2259077	-5.3563520	-3.8502831
alpha2	0.0281787	0.1218338	0.1949668	0.2853363	0.5272006
beta2	-0.1612529	-0.0704730	-0.0382768	-0.0170129	-0.0015817
sigma22V	0.1457985	0.1902363	0.2177842	0.2496907	0.3184338
sigma33	0.0130214	0.0136508	0.0139906	0.0143487	0.0150585
rho12	-0.8556995	-0.8054466	-0.7729459	-0.7370593	-0.6501383
rho13	0.0266995	0.0615367	0.0792856	0.0971150	0.1311709
rho23	-0.2335275	-0.1670558	-0.1321860	-0.0979035	-0.0299741
a	-0.0193522	-0.0104156	-0.0062760	-0.0019092	0.0080387
b2	0.0001421	0.0002256	0.0002996	0.0004109	0.0008392
lambda0	0.0051479	0.0065661	0.0085743	0.0111619	0.0145301

Gelman and Rubin's diagnostics
(Potential scale reduction factors):

	Point est.	97.5% quantile
mu	1.00	1.00
alpha1	1.00	1.00
beta1	1.00	1.01
alpha2	1.00	1.00
beta2	1.00	1.00
sigma22V	1.00	1.00
sigma33	1.00	1.00
rho12	1.01	1.02
rho13	1.00	1.00
rho23	1.00	1.00
a	1.00	1.01
b2	1.01	1.01
lambda0	1.01	1.02

III

TABLE 7.

Estimation results of the mortality model.

```
Number of chains = 3
Sample size per chain = 2500

1. Empirical mean and standard deviation for each variable,
   plus standard error of the mean:

             Mean        SD  Naive SE Time-series SE
beta00 -7.218e+00 0.0230773 2.665e-04      3.524e-04
beta01 -1.391e-02 0.0011916 1.376e-05      6.582e-05
beta10  7.248e-02 0.0008204 9.473e-06      3.685e-05
beta11 -2.229e-05 0.0001162 1.341e-06      7.854e-06
sigma2m 2.941e-02 0.0015699 1.813e-05      4.937e-05
phi     2.486e-01 0.0366845 4.236e-04      1.603e-03

2. Quantiles for each variable:

             2.5%        25%        50%        75%      97.5%
beta00 -7.2635215 -7.2333503 -7.218e+00 -7.203e+00 -7.1731612
beta01 -0.0162149 -0.0146982 -1.392e-02 -1.315e-02 -0.0115150
beta10  0.0708985  0.0719250  7.247e-02  7.302e-02  0.0740813
beta11 -0.0002524 -0.0001005 -1.917e-05  5.772e-05  0.0002014
sigma2m 0.0266480  0.0283436  2.928e-02  3.038e-02  0.0327533
phi     0.1769527  0.2246326  2.472e-01  2.714e-01  0.3236979

Gelman and Rubin's diagnostics
(Potential scale reduction factors):

       Point est. 97.5% quantile
beta00       1.00           1.00
beta01       1.01           1.04
beta10       1.01           1.02
beta11       1.02           1.06
sigma2m      1.00           1.00
phi          1.00           1.01
```

A Bayesian smoothing spline method for mortality modeling

Arto Luoma
University of Tampere, Finland

Anne Puustelli
University of Tampere, Finland

Lasse Koskinen
Financial Supervisory Authority of Finland
and
Helsinki School of Economics, Finland

IV

We propose a new method for two-dimensional mortality modeling. Our approach smoothes the data set in the dimensions of cohort and age using Bayesian smoothing splines. The method allows the data set to be imbalanced, since more recent cohorts have fewer observations. We suggest an initial model for observed death rates, and an improved model which deals with the numbers of deaths directly. Unobserved death rates are estimated by smoothing the data with a suitable prior distribution. To assess the fit and plausibility of our models we perform model checks by introducing appropriate test quantities. We show that our final model fulfils nearly all of the model selection criteria proposed by Cairns et al. (2008).

Key Words: Bayesian analysis, cohort effect, forecasting, MCMC, model checking, parameter uncertainty, smoothing splines, stochastic mortality model

1. INTRODUCTION

Mortality forecasting is a problem of fundamental importance for the insurance and pensions industry. Due to the increasing focus on risk management and measurement for insurers and pension funds, stochastic mortality models have attracted considerable interest in recent years. A range of stochastic models for mortality have been proposed, for example the seminal models of Lee and Carter (1992), Renshaw and Haberman (2006) and Cairns et al. (2006b). Some models build on an assumption of smoothness in mortality rates between ages in any given year (e.g. Cairns et al., 2006b), while others allow for roughness, (e.g. Lee and Carter, 1992; Renshaw and Haberman, 2006).

In this paper we propose a new Bayesian method for two-dimensional mortality modeling. Our method is based on natural cubic smoothing splines, which are popular in statistical applications, since the smoothing problem can be solved using simple linear algebra. In this approach the distinct data values are taken as knots of the spline, and its smoothness is achieved by employing rouhgness penalty in a penalized likelihood function. In the Bayesian approach, the prior distribution takes the role of the roughness penalty term. A useful introduction to smoothing splines may be found, for example, in Green and Silverman (1994).

A more general penalized splines approach would employ a set of basis functions, such as B-splines. In the case that cubic B-splines are used, one may obtain the same solution as in the smoothing spline approach by using the same roughness penalty and by choosing the knots to be the distinct values of the data points. Compared to the general penalized splines approach our approach has the advantage that one does not need to optimize with respect to the number of knots and their locations. However, the drawback in our approach is that the matrices involved in computations become too large, unless one restricts the size of the estimation data set.

We use age-cohort data instead of age-period data, since we wish to preserve the sequential dependence of observations within each cohort. Therefore, we have to deal with imbalanced data, since more recent cohorts have fewer observations. We suggest an initial model for the observed death rates, and an improved model which deals with the numbers of deaths directly. We assume the number of deaths to follow a Poisson distribution, a common model for the number of deaths in a year in a particular cohort. Unobserved death rates are estimated by smoothing the data with one of our spline models. The proposed method is illustrated using Finnish mortality data for females, provided by the Human Mortality Database. We implement the Bayesian approach using the Markov chain Monte Carlo method (MCMC), or more specifically, the single-component Metropolis-Hastings algorithm.

The use of Bayesian methods is not new in this general context. Dellaportas et al. (2001) proposed a Bayesian mortality model in a parametric curve modeling context. Czado et al. (2005) and Pedroza (2006) provided Bayesian analyses for the Lee-Carter model using MCMC, with further work by Kogure and Kurache (2010). More recently, Reichmuth and Sarferaz (2008) have applied MCMC to a version of the Renshaw and Haberman (2006) model. Schmid and Held (2007) present software which allows analysis of incidence count data with a Bayesian age-period-cohort model. Cairns et al. (2011) use the same model to compare results based on a two-population approach with single-population results. Currie et al. (2004) and Richards et al. (2006) assume smoothness in both age and cohort dimensions through the use of P-splines in a non-Bayesian set-up. Lang and Brezger (2004) introduce two-dimensional P-splines in a Bayesian set-up but in a different context.

Cairns et al. (2008) evaluated several types of stochastic mortality models using a checklist of criteria. These criteria are based on general characteristics and the ability of the model to explain historical patterns of mortality. None of the existing models met all of the criteria. However, Plat (2009) later proposed a model which apart from partly meeting the parsimony criteria meets all of the criteria. We also follow the same list in assessing the fit and plausibility of our model.

The plan of the paper is as follows. In the next section we describe the data and its use in estimation. In Section 3 we explain the smoothing problem and present the Bayesian formulation of the preliminary model, and in Section 4 we describe our final model. In Section 5 we introduce the estimation method and provide some convergence results. The model checks are described in Section 6, after which we conclude with a brief discussion.

2. DATA

We use mortality data provided by the Human Mortality Database (see Human Mortality Database, 2009). This was created to provide detailed mortality and population data to those interested in the history of human longevity. In our work we use Finnish cohort mortality data for females. We use age-cohort data instead of age-period data, since we wish to take into account the dependence of consequtive observations within each cohort. In the complete data matrix the years of birth included are between 1807 and 1977; hence there are 171 different cohorts. The most recent data are from 2006. When the age group of persons 110 years and older is excluded, the dimensions of the data matrix become 110×171. These data are illustrated in Figure 1, in which the observed area is denoted by vertical lines and the unobserved by two white triangles in the upper left and lower right corners.

Our estimation method would produce huge matrices if all these data were used simultaneously. Therefore, we define estimation areas which are parts of the complete data set. A rectangular estimation area shown in Figure 1 indicates the cohorts and ages for which a smooth spline surface is fitted. The mortality rates are known for part of this area, and they are predicted for the unknown part. More specifically, an estimation area is defined by minimum age x_1, maximum age x_K, minimum cohort t_1 and the maximum cohort t_T. The maximum age for which data are available in cohort t_T is denoted as x_*. Thus, the number of ages included is $K = x_K - x_1 + 1$ and the number of cohorts $T = t_T - t_1 + 1$.

Since some readers might be more familiar with age-period data, we have also plotted the data set in the dimensions of age and year in Figure 2. One should, however, remember that the figures in these two types of mortality tables are not computed in the same way. One figure in an age-period table is based on persons who have a certain (discrete) age during one calendar year and are born during two consecutive years, while each figure in an age-cohort table is based on data from two consecutive calendar years about persons born in a certain year (for details, see Wilmoth et al., 2007).

3. PRELIMINARY MODEL

We start building our model in a simplified set-up. Let us denote the logarithms of observed death rates as $y_{xt} = \log(m_{xt})$ for ages $x = x_1, x_2, ..., x_K$ and cohorts (years of birth) $t = t_1, t_2, ..., t_T$. The observed death rates are defined as

$$m_{xt} = \frac{d_{xt}}{e_{xt}},$$

where d_{xt} is the number of deaths and e_{xt} the person years of exposure. In our preliminary set-up we model the observed death rates directly, while in our final set-up we model the theoretical, unobserved death rates μ_{xt}.

3.1. The smoothing problem

Our goal is to smooth and predict logarithms of observed death rates. We fit a smooth two-dimensional curve $\theta(x, t)$, and denote its values at discrete points as θ_{xt}. In matrix form we may write

$$\mathbf{Y} = \mathbf{\Theta} + \mathbf{E},$$

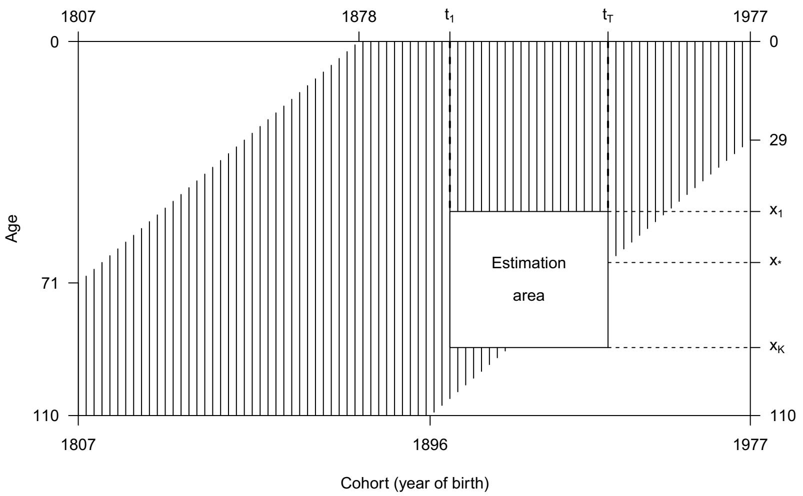


FIG. 1. Age-cohort representation of the data set. The complete data set is indicated by the streaked area, and the imbalanced estimation set by the white rectangle.

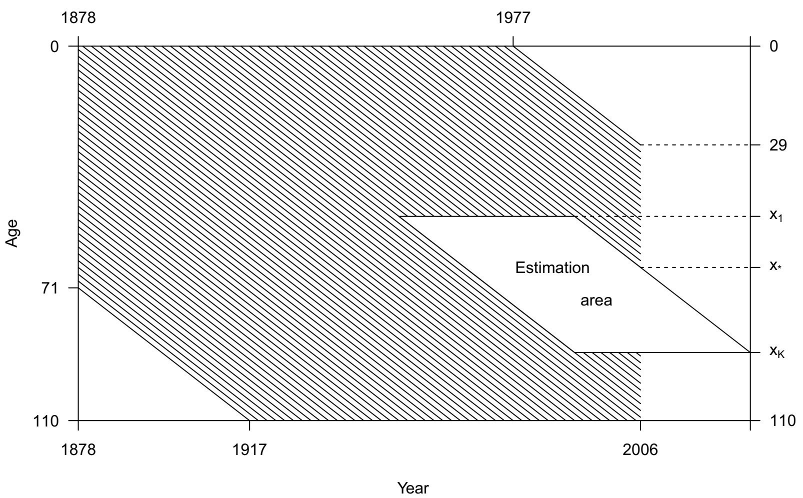


FIG. 2. Age-period representation of the data set. The complete data set is indicated by the streaked area, and the imbalanced estimation set by the white parallelogram.

where $\mathbf{Y}$ is a $K \times T$ matrix of observations, $\mathbf{\Theta}$ is a matrix of smoothed values, and $\mathbf{E}$ is a matrix of errors. We denote the columns of $\mathbf{Y}$, $\mathbf{\Theta}$ and $\mathbf{E}$ by $\mathbf{y}_j$, $\boldsymbol{\theta}_j$ and $\boldsymbol{\varepsilon}_j$, respectively. Concatenating the columns we obtain $\mathbf{y} = \mathrm{vec}(\mathbf{Y})$, $\boldsymbol{\theta} = \mathrm{vec}(\mathbf{\Theta})$ and $\boldsymbol{\varepsilon} = \mathrm{vec}(\mathbf{E})$.

We further assume that the death rates within a cohort follow a multivariate normal distribution having an AR(1) correlation structure with autocorrelation coefficient ϕ. Thus,

$$\boldsymbol{\epsilon}_j \sim N(\mathbf{0}, \sigma^2 \mathbf{P}), \quad j = 1, 2, ..., T,$$

where $\mathbf{P}$ is a correlation matrix with elements $\rho_{rs} = \phi^{|r-s|}$. The observations in different cohorts are assumed to be independent.

In general, all observations are not available for all cohorts. For each j, we may partition $\mathbf{y}_j$ into observed $\mathbf{y}_{j1}$ and unobserved $\mathbf{y}_{j2}$, and $\boldsymbol{\theta}_j$ correspondingly to $\boldsymbol{\theta}_{j1}$ and $\boldsymbol{\theta}_{j2}$, and $\mathbf{P}$ to $\mathbf{P}_{j,rs}$, $r, s = 1, 2$. The unobserved part of the data can be predicted using the result about the conditional distribution of the multivariate normal distribution:

$$\{\mathbf{y}_{j2} | \mathbf{y}_{j1}, \sigma^2, \phi\} \sim N(\boldsymbol{\theta}_{j2.1}, \sigma^2 \mathbf{P}_{j,22.1}),$$

where $\boldsymbol{\theta}_{j2.1} = \boldsymbol{\theta}_{j2} + \mathbf{P}_{j,21} \mathbf{P}_{j,11}^{-1} (\mathbf{y}_{j1} - \boldsymbol{\theta}_{j1})$ and $\mathbf{P}_{j,22.1} = \mathbf{P}_{j,22} - \mathbf{P}_{j,21} \mathbf{P}_{j,11}^{-1} \mathbf{P}_{j,12}$.

When estimating $\boldsymbol{\theta}$ we wish to minimize the generalized sum of squares

$$SS_1 = \sum_{j=1}^{T} (\mathbf{y}_{j1} - \boldsymbol{\theta}_{j1})' \mathbf{P}_{j,11}^{-1} (\mathbf{y}_{j1} - \boldsymbol{\theta}_{j1}). \tag{1}$$

The vector of all observed mortality rates is $\mathbf{y}^{obs} = \mathbf{Sy}$, where $\mathbf{S}$ is a selection matrix selecting the known values from the complete data vector $\mathbf{y}$. The matrix $\mathbf{S}$ can be constructed from the identity matrix of size KT by including the ith row ($i = 1, 2, ..., KT$) if the ith element of $\mathbf{y}$ is known. Now we can write (1) as

$$SS_1 = (\mathbf{y}^{obs} - \mathbf{S}\boldsymbol{\theta})' (\mathbf{S}\mathbf{P}_* \mathbf{S}')^{-1} (\mathbf{y}^{obs} - \mathbf{S}\boldsymbol{\theta}), \tag{2}$$

where $\mathbf{P}_* = \mathbf{I}_T \otimes \mathbf{P}$.

In addition to maximizing fit, we wish to smooth $\boldsymbol{\Theta}$ in the dimensions of cohort and age. Specifically, we minimize the roughness functional

$$\int_{x_1}^{x_K} \left[\frac{\partial^2}{\partial x^2} \theta(x, t_j) \right]^2 dx \tag{3}$$

for each $j = 1, 2, ..., T$ and

$$\int_{t_1}^{t_T} \left[\frac{\partial^2}{\partial t^2} \theta(x_k, t) \right]^2 dt \tag{4}$$

for each $k = 1, 2, ..., K$.

If $\theta(x, t_j)$ is considered a smooth function of x obtaining fixed values at points $x_1, x_2, ..., x_K$, then using variational calculus it can be shown that the integral in (3) is minimized by choosing $\theta(x, t_j)$ to be a cubic splines curve with knots at $x_1, x_2, ..., x_K$. Furthermore, this integral can be expressed as a squared form $\boldsymbol{\theta}_j' \mathbf{G}_K \boldsymbol{\theta}_j$, where $\mathbf{G}_K$ is a so-called roughness matrix with dimensions $K \times K$ (for proof, see Green and Silverman, 1994). Similarly, if $\theta(x_k, t)$ is a cubic splines curve with knots at $t_1, ..., t_T$, the integral in (4) equals $\boldsymbol{\theta}_{(k)}' \mathbf{G}_T \boldsymbol{\theta}_{(k)}$, where $\boldsymbol{\theta}_{(k)}$ denotes the kth row of $\boldsymbol{\Theta}$ and $\mathbf{G}_T$ is a $T \times T$ roughness matrix. Thus, we wish to minimize

$$SS_2 = \sum_{j=1}^{T} \boldsymbol{\theta}_j' \mathbf{G}_K \boldsymbol{\theta}_j = \boldsymbol{\theta}' (\mathbf{I}_T \otimes \mathbf{G}_K) \boldsymbol{\theta} \tag{5}$$

and

$$SS_3 = \sum_{k=1}^{K} \boldsymbol{\theta}'_{(k)} \mathbf{G}_T \boldsymbol{\theta}_{(k)} = \boldsymbol{\theta}'(\mathbf{G}_T \otimes \mathbf{I}_K)\boldsymbol{\theta}. \tag{6}$$

An $N \times N$ roughness matrix is defined as $\mathbf{G}_N = \boldsymbol{\nabla}_N \boldsymbol{\Delta}_N^{-1} \boldsymbol{\nabla}'_N$ where the non-zero elements of banded $N \times (N-2)$ and $(N-2) \times (N-2)$ matrices $\boldsymbol{\nabla}_N$ and $\boldsymbol{\Delta}_N$, respectively, are defined as follows:

$$\nabla_{i,i} = \frac{1}{x_{i+1} - x_i}, \quad \nabla_{i+1,i} = -\left(\frac{1}{x_{i+1} - x_i} + \frac{1}{x_{i+2} - x_{i+1}}\right), \quad \nabla_{i+2,i} = \frac{1}{x_{i+2} - x_{i+1}}$$

and

$$\Delta_{i,i+1} = \Delta_{i+1,i} = \frac{x_{i+2} - x_{i+1}}{6}, \quad \Delta_{i,i} = \frac{x_{i+2} - x_i}{3},$$

with data points x_i, $i = 1, \ldots, n$. In our case the data are given at equal intervals, implying that

$$\nabla_{i,i} = 1, \quad \nabla_{i+1,i} = -2, \quad \nabla_{i+2,i} = 1$$

and

$$\Delta_{i,i+1} = \Delta_{i+1,i} = \frac{1}{6}, \quad \Delta_{i,i} = \frac{2}{3}.$$

Combining the previous results, we obtain the bivariate smoothing splines solution for $\boldsymbol{\theta}$ by minimizing the exression $SS_1 + \lambda_1 SS_2 + \lambda_2 SS_3$, where SS_1, SS_2 and SS_3 are given in the equations (2), (5) and (6), respectively, and the parameters λ_1 and λ_2 control smoothing in the dimensions of cohort and age, respectively. Using matrix differentiation and the properties of Kronecker's product, it is easy to show that for fixed values of λ_1 and λ_2 the minimal solution is given by

$$\hat{\boldsymbol{\theta}} = \left[\mathbf{S}'(\mathbf{S}\mathbf{P}_*\mathbf{S}')^{-1}\mathbf{S} + \mathbf{A}\right]^{-1} \mathbf{S}'(\mathbf{S}\mathbf{P}_*\mathbf{S}')^{-1}\mathbf{y}^{obs}, \tag{7}$$

where

$$\mathbf{A} = \lambda_1(\mathbf{I}_T \otimes \mathbf{G}_K) + \lambda_2(\mathbf{G}_T \otimes \mathbf{I}_K). \tag{8}$$

In the special case that the data set is balanced ($\mathbf{S}$ is an identity matrix), the solution is simplified to $\hat{\boldsymbol{\theta}} = (\mathbf{I} + \mathbf{P}_*\mathbf{A})^{-1}\mathbf{y}$.

3.2. Bayesian formulation

Bayesian statistical inference is based on the posterior distribution, which is the conditional distribution of unknown parameters given the data. In order to compute the posterior distribution one needs to define the prior distribution, which is the unconditional distribution of parameters, and the likelihood function, which is the probability density of observations given the parameters. Bayes' theorem implies that the posterior distribution is proportional to the product of the prior distribution and the likelihood:

$$p(\boldsymbol{\eta}|\mathbf{y}) \propto p(\boldsymbol{\eta})p(\mathbf{y}|\boldsymbol{\eta}),$$

where $\mathbf{y}$ is the data vector and $\boldsymbol{\eta}$ the vector of all unknown parameters.

In our case, the likelihood is given by

$$p(\mathbf{y}^{obs}|\boldsymbol{\eta}) = (2\pi\sigma^2)^{-\frac{K_*}{2}} |\mathbf{SP}_*\mathbf{S}'|^{-\frac{1}{2}} \mathrm{e}^{-\frac{1}{2\sigma^2}(\mathbf{y}^{obs}-\mathbf{S}\boldsymbol{\theta})'(\mathbf{SP}_*\mathbf{S}')^{-1}(\mathbf{y}^{obs}-\mathbf{S}\boldsymbol{\theta})}, \tag{9}$$

where K_* is the length of $\mathbf{y}^{obs}$.

In order to facilitate estimation we reparametrize the smoothing paramaters as follows: $\lambda = \lambda_1$ and $\omega = \lambda_2/\lambda_1$, where λ_1 and λ_2 control the smoothing in the dimensions of age and cohort, respectively. Furthermore, we use the following hierarchical prior for $\boldsymbol{\eta}$:

$$p(\boldsymbol{\eta}) = p(\sigma^2)p(\lambda)p(\omega)p(\phi)p(\boldsymbol{\theta}|\sigma^2, \lambda, \omega, \phi),$$

where

$$\begin{aligned} p(\sigma^2) &\propto \frac{1}{\sigma^2} \\ p(\lambda) &\propto \lambda^{\alpha_1 - 1} e^{-\beta_1 \lambda} \\ p(\omega) &\propto \omega^{\alpha_2 - 1} e^{-\beta_2 \omega} \\ p(\phi) &\propto 1, \quad -1 < \phi < 1. \end{aligned} \tag{10}$$

As hyperparameters we set $\alpha_1 = \beta_1 = 0.001$ and $\alpha_2 = \beta_2 = 10$. Thus, the prior of σ^2 is the standard uninformative improper prior used for positive parameters, and the priors of λ and ϕ are also fairly uninformative. The prior of ω is instead more informative, having mean 1 and variance 0.1, since we found that the data do not contain enough information about ω, and with a looser prior we would face convergence problems in estimation. We made sensitivity analysis with respect to the prior of λ and found that increasing or decreasing the order of magnitude of α_1 and β_1 did not essentially affect the results.

The smoothing effect can now be obtained by choosing a conditional prior for $\boldsymbol{\theta}$ which is consistent with the smoothing splines solution. Such a prior contains information only on the curvature, or roughness, of the spline surface, not on its position or gradient. Thus, we assume that $\{\boldsymbol{\theta}|\sigma^2, \lambda, \omega, \phi\}$ is multivarite normal with density

$$p(\boldsymbol{\theta} \mid \sigma^2, \lambda, \omega, \phi) = (2\pi\sigma^2)^{-\frac{KT}{2}} \left|\lambda\left[(\mathbf{I}_T \otimes \mathbf{G}_{K,\gamma} + \omega(\mathbf{G}_T \otimes \mathbf{I}_K)\right]\right|^{\frac{1}{2}} \mathrm{e}^{-\frac{\lambda}{2\sigma^2}\boldsymbol{\theta}'\left[(\mathbf{I}_T\otimes\mathbf{G}_{K,\gamma})+\omega(\mathbf{G}_T\otimes\mathbf{I}_K)\right]\boldsymbol{\theta}}, \tag{11}$$

where $\mathbf{G}_{K,\gamma}$ is a positive definite matrix approximating $\mathbf{G}_K$. More specifically, we define $\mathbf{G}_{K,\gamma} = \mathbf{G}_K + \gamma\mathbf{XX}'$, where $\gamma > 0$ can be chosen to be arbitrarily small, and $\mathbf{X} = (\mathbf{1}\ \mathbf{x})$ with $\mathbf{1} = (1, ..., 1)'$ and $\mathbf{x} = (x_1, ..., x_K)'$. Initially, we use $\mathbf{G}_{K,\gamma}$ instead of $\mathbf{G}_K$, since otherwise $p(\boldsymbol{\theta}|\sigma^2, \lambda, \omega, \phi)$ would be improper, which would lead to difficulties when deriving the conditional posteriors for λ and ω.

Multiplying the densities in (11) and (9) and picking the factors which include $\boldsymbol{\theta}$ we obtain the full conditional posterior for $\boldsymbol{\theta}$ up to a constant of proportionality:

$$p(\boldsymbol{\theta}|\mathbf{y}, \sigma^2, \lambda, \omega, \phi) \propto \mathrm{e}^{-\frac{1}{2\sigma^2}\left\{(\mathbf{y}^{obs}-\mathbf{S}\boldsymbol{\theta})'(\mathbf{SP}_*\mathbf{S}')^{-1}(\mathbf{y}^{obs}-\mathbf{S}\boldsymbol{\theta})+\lambda\boldsymbol{\theta}'\left[(\mathbf{I}_T\otimes\mathbf{G}_{K,\gamma})+\omega(\mathbf{G}_T\otimes\mathbf{I}_K)\right]\boldsymbol{\theta}\right\}}. \tag{12}$$

Manipulating this expression and replacing $\mathbf{G}_{K,\gamma}$ with $\mathbf{G}_K$ we obtain

$$p(\boldsymbol{\theta}|\mathbf{y}, \sigma^2, \lambda, \omega, \phi) \propto \mathrm{e}^{-\frac{1}{2\sigma^2}(\boldsymbol{\theta}-\hat{\boldsymbol{\theta}})'\mathbf{B}(\boldsymbol{\theta}-\hat{\boldsymbol{\theta}})},$$

where $\hat{\boldsymbol{\theta}}$ is given in (7) and $\mathbf{B} = \mathbf{A} + \mathbf{S}'(\mathbf{SP}_*\mathbf{S}')^{-1}\mathbf{S}$. From this we see that the conditional posterior distribution of $\boldsymbol{\theta}$ is multivariate normal with mean $\hat{\boldsymbol{\theta}}$ and covariance matrix $\sigma^2\mathbf{B}^{-1}$ in the limiting case when $\mathbf{G}_{K,\gamma} \to \mathbf{G}_K$. This implies that the conditional posterior mode for $\boldsymbol{\theta}$ is equal to the smoothing splines solution provided in the previous section. Thus, using the multivariate prior described above, we can implement the roughness penalty of smoothing splines in the Bayesian framework.

In order to implement estimation using the Gibbs sampler, the full conditional posterior distributions of the parameters are needed. In the following, we will provide these for σ^2, λ, ω and ϕ in the limiting case when $\mathbf{G}_{K,\gamma} \to \mathbf{G}_K$.

The conditional posterior of σ^2 is

$$p(\sigma^2|\mathbf{y}, \lambda, \omega, \phi) \propto (\sigma^2)^{-\left(\frac{K_*+KT}{2}+1\right)} \mathrm{e}^{-\frac{1}{2\sigma^2}\left[(\mathbf{y}^{obs}-\mathbf{S}\boldsymbol{\theta})'(\mathbf{SP}_*\mathbf{S}')^{-1}(\mathbf{y}^{obs}-\mathbf{S}\boldsymbol{\theta})+\boldsymbol{\theta}'\mathbf{A}\boldsymbol{\theta}\right]},$$

which is the density of the scaled inverted χ^2-distribution Inv-$\chi^2(\nu, b)$[1], where $\nu = K_* + KT$ and $b = (SS_1 + \lambda SS_2 + \lambda\omega SS_3)/\nu$ with SS_1, SS_2 and SS_3 given in (2), (5) and (6).

The conditional posterior of λ is

$$p(\lambda|\mathbf{y}, \boldsymbol{\theta}, \sigma^2, \omega, \phi) \propto \lambda^{\alpha_1 - 1 + \frac{KT}{2}} \mathrm{e}^{-\lambda\left[\beta_1 + \frac{1}{2\sigma^2}\boldsymbol{\theta}'(\mathbf{I}_T\otimes\mathbf{G}_K + \omega\mathbf{G}_T\otimes\mathbf{I}_K)\boldsymbol{\theta}\right]}, \tag{13}$$

which is the density of Gamma$(\alpha_1 + KT/2,\ \beta_1 + (SS_2 + \omega SS_3)/(2\sigma^2))$.

The conditional posterior of ω is

$$p(\omega|\mathbf{y}, \boldsymbol{\theta}, \sigma^2, \lambda, \phi) \propto \omega^{\alpha_2 + T - 2} \left[\prod_{k=1}^{K-2}\prod_{j=1}^{T-2}\left(1 + \omega\frac{\mu_j}{\nu_k}\right)\right]^{\frac{1}{2}} \mathrm{e}^{-\omega\left[\beta_2 + \frac{\lambda}{2\sigma^2}\boldsymbol{\theta}'(\mathbf{G}_T\otimes\mathbf{I}_K)\boldsymbol{\theta}\right]}, \tag{14}$$

where $\mu_j,\ j = 1, ..., T-2$, and $\nu_k,\ k = 1, ..., K-2$, are the nonzero eigenvalues of $\mathbf{G}_T$ and $\mathbf{G}_K$, respectively. This is not a standard distribution, but since it is log-concave, it is possible to generate values from it using adaptive rejection sampling, introduced by Gilks and Wild (1992).

Finally, the conditional posterior of ϕ, given by

$$p(\phi|\mathbf{y}, \boldsymbol{\theta}, \sigma^2, \lambda, \omega) \propto (1 - \phi^2)^{-\frac{1}{2}(K_* - T)} \mathrm{e}^{-\frac{1}{2\sigma^2}SS_1},$$

is not of standard form, and it is therefore difficult to generate random variates from it directly. Instead, we may employ a Metropolis step within the Gibbs sampler.

Now, the estimation algorithm is implemented so that the parameters $\lambda, \omega, \sigma^2$ and $\boldsymbol{\theta}$ are updated one by one using Gibbs steps, and ϕ is updated using a Metropolis step. Further details will be given in Section 5.

4. THE FINAL MODEL

In our final set-up we are able to control for unsystematic mortality risk in addition to systematic risk. Unsystematic risk means that even if the true mortality rate were known, the numbers of deaths would remain unpredictable. When the population becomes larger, the unsystematic mortality risk becomes smaller due to diversification.

[1]Notation $X \sim$ Inv-$\chi^2(\nu, b)$ means that $\nu b/X \sim \chi^2_\nu$.

4.1. Formulation and estimation

In the final model the inference is rendered more accurate by modeling the observed numbers of deaths directly. Specifically, we assume that

$$d_{xt} \sim \text{Poisson}(\mu_{xt} e_{xt}),$$

where d_{xt} is the number of deaths at age x and cohort t, μ_{xt} is the theoretical death rate (also called intensity of mortality or force of mortality) and e_{xt} is the person years of exposure. This is an approximation, since neither the death rate nor the exposure is constant during any given year. Our purpose is to model $\theta_{xt} = \log(\mu_{xt})$ with a smooth spline surface. Compared to the preliminary model we have replaced m_{xt} by μ_{xt} and removed the error term and its autocorrelation structure.

Similarly to the preliminary model, we obtain the smoothing effect by using a suitable conditional prior distribution for $\boldsymbol{\theta}$. Specifically, we obtain $p(\boldsymbol{\theta}|\lambda, \omega)$ by replacing σ^2 with 1 in equation (11). For λ and ω we use the same prior distributions as earlier, given by (10), and their conditional posteriors are obtained from (13) and (14) when σ^2 is set at 1. However, here we use hyperparameters $\alpha_1 = \beta_1 = 10^{-6}$, since removing σ^2 changes the scale of λ several orders of magnitude.

Now the full conditional posterior distribution of $\boldsymbol{\theta}$ may be written as

$$p(\boldsymbol{\theta}|\mathbf{d}^{obs}, \lambda, \omega) \propto \exp\left\{\sum_{t=t_1}^{t_T}\sum_{x=x_1}^{x_{K_t}} [d_{xt}\theta_{xt} - e_{xt}\exp(\theta_{xt})] - \frac{1}{2}\boldsymbol{\theta}'\mathbf{A}\boldsymbol{\theta}\right\}, \tag{15}$$

where $\mathbf{d}^{obs}$ is a vector of observed death numbers, and K_t the number of ages for which data are available in cohort t. The double sum in this expression comes from the likelihood function and the squared form from the prior distribution.

This model can be estimated similarly to the preliminary model, using Gibbs sampling. However, since the conditional distribution in (15) is non-standard, it is difficult to sample from it directly. Here we may use a Metropolis-Hastings step within the Gibbs sampler. As a proposal distribution we may use a multivariate normal approximation to (15), given by

$$J(\boldsymbol{\theta}) \propto \exp\left\{-\frac{1}{2}\left(\mathbf{y}^{obs} - \mathbf{S}\boldsymbol{\theta}\right)'(\mathbf{S}\boldsymbol{\Sigma}\mathbf{S}')^{-1}\left(\mathbf{y}^{obs} - \mathbf{S}\boldsymbol{\theta}\right) - \frac{1}{2}\boldsymbol{\theta}'\mathbf{A}\boldsymbol{\theta}\right\},$$

where $\mathbf{y}^{obs}$ is a vector of observed log death rates, $\boldsymbol{\Sigma}$ is a diagonal matrix with $1/d_{xt},\ x = x_1, ..., x_K,\ t = t_1, ..., t_T$, as its diagonal elements, and $\mathbf{S}$ is a selection matrix defined in Section 3.1. The approximate variance $1/d_{xt}$ of the log death rate is obtained by applying the delta method to the relevant transformation of the underlying Poisson variable.

Thus, the proposal $\boldsymbol{\theta}^*$ is distributed as

$$\boldsymbol{\theta}^* \sim \text{MVN}(\mathbf{C}^{-1}\mathbf{S}'(\mathbf{S}\boldsymbol{\Sigma}\mathbf{S}')^{-1}\mathbf{y}^{obs}, \mathbf{C}^{-1}),$$

where $\mathbf{C} = \mathbf{A} + \mathbf{S}'(\mathbf{S}\boldsymbol{\Sigma}\mathbf{S}')^{-1}\mathbf{S}$, and is accepted with probability

$$\min\left(1, \frac{p(\boldsymbol{\theta}^*|\mathbf{d}^{obs}, \lambda, \omega)/J(\boldsymbol{\theta}^*)}{p(\boldsymbol{\theta}|\mathbf{d}^{obs}, \lambda, \omega)/J(\boldsymbol{\theta})}\right).$$

This update is an independence sampler, since the proposal distribution of $\boldsymbol{\theta}^*$ does not depend on the current value $\boldsymbol{\theta}$. The whole algorithm is once more a special case of the single-component Metropolis-Hastings. Further details on this algorithm will be provided in the next section.

5. ESTIMATION

5.1. Estimation procedure

Our estimation procedure is a single-component (or cyclic) Metropolis-Hastings algorithm. This is one of the Markov Chain Monte Carlo (MCMC) methods, which are useful in drawing samples from posterior distributions. Generally, MCMC methods are based on drawing values from approximate distributions and then correcting these draws to better approximate the target distribution, and hence they are used when direct sampling from a target distribution is difficult. A useful reference for different versions of MCMC is Gilks et al. (1996).

The Metropolis-Hastings algorithm was introduced by Hastings (1970) as a generalization of the Metropolis algorithm (Metropolis et al., 1953). Also the Gibbs sampler proposed by Geman and Geman (1984) is its special case. The Gibbs sampler assumes the full conditional distributions of the target distribution to be such that one is able to generate random numbers or vectors from them. The Metropolis and Metropolis-Hastings algorithms are more flexible than the Gibbs sampler; with them one o nly needs to know the joint density function of the target distribution with density $p(\theta)$ up to a constant of proportionality.

With the Metropolis algorithm the target distribution is generated as follows: first a starting distribution $p_0(\theta)$ is assigned, and from it a starting-point θ^0 is drawn such that $p(\theta^0) > 0$. For iterations $t = 1, 2, \ldots$, a proposal θ^* is generated from a jumping distribution $J(\theta^*|\theta^{t-1})$, which is symmetric in the sense that $J(\theta_a|\theta_b) = J(\theta_b|\theta_a)$ for all θ_a and θ_b. Finally, iteration t is completed by calculating the ratio

$$r = \frac{p(\theta^*)}{p(\theta^{t-1})} \tag{16}$$

and by setting the new value at

$$\theta^t = \begin{cases} \theta^* & \text{with probability } \min(r, 1) \\ \theta^{t-1} & \text{otherwise.} \end{cases}$$

It can be shown that, under mild conditions, the algorithm produces an ergodic Markov Chain whose stationary distribution is the target distribution.

Metropolis-Hastings algorithm generalizes the Metropolis algorithm by removing the assumption of symmetric jumping distribution. The ratio r in (16) is replaced by

$$r = \frac{p(\theta^*)/J(\theta^*|\theta^{t-1})}{p(\theta^{t-1})/J(\theta^{t-1}|\theta^*)}$$

to correct for the asymmetry in the jumping rule.

In the single-component Metropolis-Hastings algorithm the simulated random vector is divided into components or subvectors which are updated one by one. If the jumping distribution for a component is its full conditional posterior distribution, the proposals are accepted with probability one. In the case that all the components are simulated in this way, the algorithm is called a Gibbs sampler. As stated above, in the case of our preliminary model we can simulate all parameters except ϕ directly, and may therefore use a Gibbs sampler with one Metropolis step. As the jumping distribution of ϕ we use the normal distribution $N(\phi^{t-1}, 0.05^2)$. For the final model we use a Gibbs sampler with one Metropolis-Hastings step for $\boldsymbol{\theta}$. The proposal distribution and its acceptance probability were already given in Section 4.

5.2. Empirical results

All the computations in this article were performed and figures produced using the R computing environment (R Development Core Team, 2010). The functions and data needed to replicate the results can be found at http://mtl.uta.fi/codes/mortality. A minor drawback is that we cannot use all available data in estimation but must restrict ourselves to a relevant subset. This is due to the huge matrices involved in computations if many ages and cohorts are included in the data set. For example, if we used our complete data set, whose dimensions are $T = 110$ and $K = 171$, we would have to deal with Kronecker product matrices of dimension 18810×18810. This would require 5 GB of memory for storing one matrix and much more for computations. Although we can alleviate the storage problem and also speed up the computations using sparse matrix methods, we still cannot use the complete data set. In our implementation we use the R package SparseM.

To assess the convergence of the simulated Markov chain to its stationary distribution we used 5 representative values of $\boldsymbol{\theta}$, denoted as $\theta_1, ..., \theta_5$, from each corner and the middle of the data matrix, in addition to the upper level parameters. The value θ_5 is in the lower right corner of the matrix and corresponds to an unobserved data item.

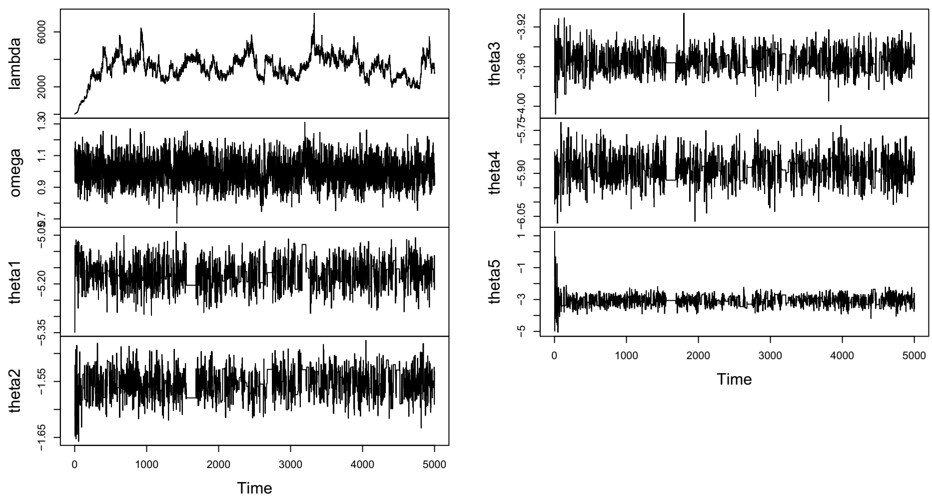


FIG. 3. Posterior simulations of the final model.

For each data set and both models we assessed the convergence of iterative simulation using three simulated sequences with 5000 iterations. In the case of the final model we discarded 1500 first iterations of each chain as a burn-in period, while in the case of the preliminary model the convergence was more rapid and we discarded only 200 iterations.

Figure 3 shows one simulated chain corresponding to the final model and the data set with ages 50–90 and cohorts 1901–1941. It seems that the chain converges to its stationary distribution after about 300 iterations, and the component series of the chain mix well, that is, they are not excessively autocorrelated, except for the parameter λ. Summaries of the estimation results for both preliminary and final model as well as the diagnostics of Gelman and Rubin (1992) are given in Appendix A. The values of the diagnostic are close to 1 and thus indicate good convergence.

6. MODEL CHECKING

Cairns et al. (2008) provide a checklist of criteria against which a stochastic mortality model can be assessed. We will follow this list as we assess the fit and plausibility of our two models. The list is as follows:

1. Mortality rates should be positive.
2. The model should be consistent with historical data.
3. Long-term dynamics under the model should be biologically reasonable.
4. Parameter estimates should be robust relative to the period of data and range of ages employed.
5. Model forecasts should be robust relative to the period of data and range of ages employed.
6. Forecast levels of uncertainty and central trajectories should be plausible and consistent with historical trends and variability in mortality data.
7. The model should be straightforward to implement using analytical methods or fast numerical algorithms.
8. The model should be relatively parsimonious.
9. It should be possible to use the model to generate sample paths and calculate prediction intervals.
10. The structure of the model should make it possible to incorporate parameter uncertainty in simulations.
11. At least for some countries, the model should incorporate a stochastic cohort effect.
12. The model should have a non-trivial correlation structure.

Both of our models fulfil the first item in the list, since we model log death rates. To assess the consistency of the models with historical data we will introduce three Bayesian test quantities in Section 6.1.

A model is defined by Cairns et al. (2006a) to be biologically reasonable if the mortality rates are increasing with age and if there is no long-run mean reversion around a deterministic trend. Our spline approach implies that the log death rate increases linearly beyond the estimable region. The preliminary model allows for short-term mean reversion (or autocorrelation) for the observed death rate, while there is no mean reversion at all in the final model.

The fourth and fifth points in the list, that is, the robustness of parameter estimates and model forecasts, will be studied in Sections 6.2 and 6.3. The figures of posterior predictions in Section 6.3 help assess the plausibility and uncertainty of forecasts and their consistency with historical trends and variability.

Implementing the models is fairly straightforward but involves several algorithms. Basically, we use the Gibbs sampler, and supplement it with rejection sampling and Metropolis and Metropolis-Hastings steps, which are needed to update certain parameters or parameter blocks. A further complication is that we have to use sparse matrix methods to increase the maximum size of the data set.

In the Bayesian approach one typically uses posterior predictive simulation, in which parameter uncertainty is taken into account, to generate sample paths and calculate prediction intervals. This will be explained in detail in Section 6.3.

The hierarchical structure of the spline models makes them parsimonious: on the upper level the preliminary model has 4 parameters, the final model only 2. Both models also incorporate a stochastic cohort effect. The preliminary model incorporates an AR(1) structure for observed mortality, while the final model has no correlation structure for deviations from the spline surface.

One should note, however, that the spline model in itself implies a covariance structure. In a one-dimensional case the Bayesian smoothing spline model can be interpreted as a sum of a linear trend and integrated Brownian motion (Wahba, 1978). The prior distribution does not contain information on the intercept or slope of the trend but implies the covariance structure of the integrated Brownian motion. Similarily, in our two-dimensional case, the spline surface can be interpreted as a sum of a plane and deviations from this plane. The conditional prior of $\boldsymbol{\theta}$, given the smoothing parameters, does not include information on the plane but implies a specific spatial covariance structure for the deviations.

6.1. Tests for the consistency of the model

In the Bayesian framework, posterior predictive simulations of replicated data sets may be used to check the model fit (see Gelman et al., 2004). Once several replicated data sets y^{rep} have been produced, they may be compared with the original data set y. If they look similar to y, the model fits.

The discrepancy between data and model can be measured using arbitrarily defined test quantities. A test quantity $T(y, \theta)$ is a scalar summary of parameters and data which is used to compare data with predictive simulations. If the test quantity depends only on data and not on parameters, then it is said to be a test statistic. If we already have N posterior simulations $\theta_i,\ i = 1, ..., N$, we can generate one replication y_i^{rep} using each θ_i, and compute the test quantities $T(y, \theta_i)$ and $T(y_i^{rep}, \theta_i)$. The Bayesian p-value is defined to be the posterior probability that the test quantity computed from a replication, $T(y^{rep}, \theta)$, will exceed that computed from the original data, $T(y, \theta)$. This test may be illustrated by a scatter plot of $(T(y, \theta_i), T(y_i^{rep}, \theta_i)),\ i = 1, ..., N$, where the same scale is used for both coordinates. Further details on this approach can be found in Chapter 6 of Gelman et al. (2004) or Chapter 11 of Gilks et al. (1996).

In the case of our preliminary model, a replication of data is generated as follows: First, $\boldsymbol{\theta}$, σ^2 and ϕ are generated from their joint posterior distribution. Then, using these parameter values, a replicated data vector $\mathbf{y}^{rep}$ is generated from the multivariate normal distribution $N(\boldsymbol{\theta}, \mathbf{I} \otimes \sigma^2 \mathbf{P})$. Finally, the elements of $\mathbf{y}^{rep}$ which correspond to the observed values in $\mathbf{y}^{obs}$ are selected. In the case of the final model, $\boldsymbol{\theta}$ is first generated and then the numbers of deaths d_{xt} and exposures e_{xt} are generated recursively by starting from the smallest age included in the estimation data set. The numbers for the smallest age are not generated but they are taken to be the same as in the estimation set. Finally, the replicated death rates are computed as $y_{xt} = \log(d_{xt}/e_{xt})$, and the values corresponding to the observed values in $\mathbf{y}^{obs}$ are selected. Further details about this procedure are provided in Appendix B.

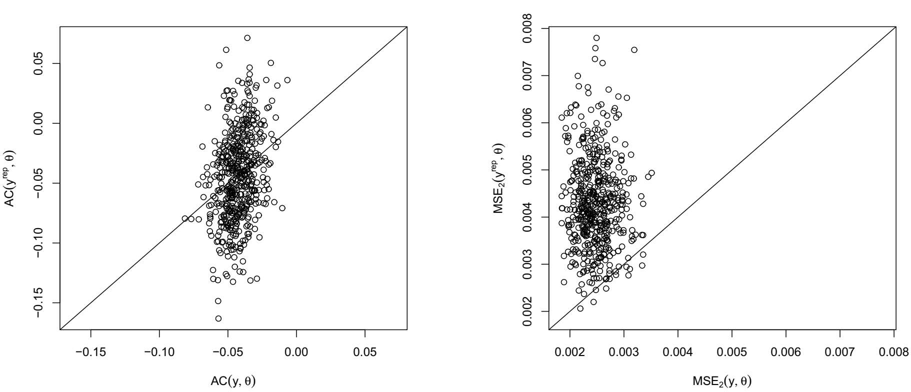
We introduce three test quantities to check the model fit. The first measures the autocorrelation of the observed log death rate and the second and third its mean square error:

$$AC(y, \theta) = \frac{\sum_{t=t_1}^{T} \sum_{x=x_1}^{x_K - 1} (y_{x+1,t} - \theta_{x+1,t})(y_{xt} - \theta_{xt})}{\sum_{t=1}^{T} K_t},$$

where K_t is the number of observations in cohort t, and

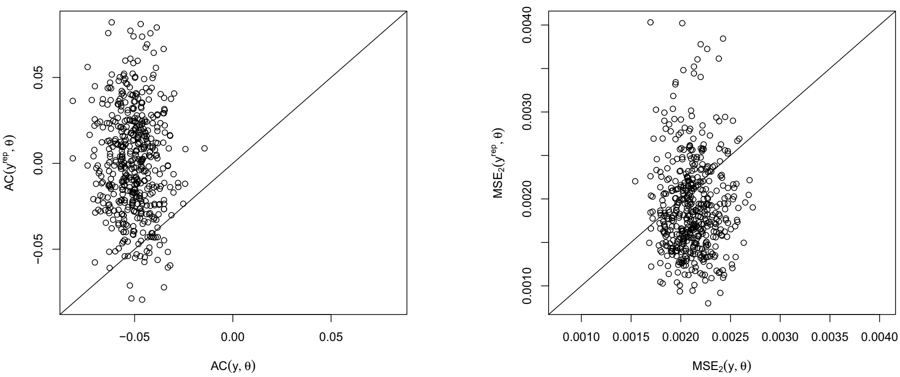
$$MSE_1(y, \theta) = \frac{\sum_{t=t_1}^{t_T} \sum_{x=x_1}^{x_{K_t}} (y_{xt} - \theta_{xt})^2}{\sum_{t=1}^{T} K_t}, \quad MSE_2(y, \theta) = \frac{\sum_{t=t_1}^{t_T} \left(y_{x_{K_t} t} - \theta_{x_{K_t} t} \right)^2}{T}.$$

Figures 4 and 5 show the results when using the data set with ages 50–90 and cohorts 1901–1941. Each figure is based on 500 simulations. If the original data and replicated data were consistent, about half the points in the scatter plot would fall above the 45° line and half below. Figure 4a indicates that the preliminary model adequately explains the autocorrelation observed in the original data set, while Figure 5a suggests that there might be slight negative autocorrelation in the residuals not explained by the model. However, since the Bayesian p-value, which is the proportion of points above the line, is approximately 0.95, there is no sufficient evidence to reject the assumption of independent Poisson observations.



(a) Autocorrelation test (b) MSE test

FIG. 4. Goodness-of-fit testing for the preliminary model.



(a) Autocorrelation test

(b) MSE test

FIG. 5. Goodness-of-fit testing for the final model.

The test statistic MSE_1 measures the overall fit of the models, and both models pass it (figures not shown). The test statistic MSE_2 measures the fit at the largest ages of the cohorts. From Figure 5b we see that that the final model passes this test. However, Figure 4b suggests that under the preliminary model the MSE_2 simulations based on the original data are smaller than those based on replicated data sets ($p_B = 0.98$). The reason here is that the homoscedasticity assumption of logarithmic mortality data is not valid. The validity of the homoscedasticity and independence assumptions could be further assessed by plotting the standardized residuals (not shown here).

6.2. Robustness of the parameter estimates

The robustness of the parameters may be studied by comparing the posterior distributions when two different but equally sized data sets are used. Here we used two data sets with ages 40–70 and 60–90, and cohorts 1917–1947 and 1886–1916, respectively. We refer to these as the younger and older age groups, respectively. Figure 6 (c) indicates that the variance parameter σ^2 of the preliminary model is clearly higher for the younger age group. This results from the fact that the variance of log mortality data becomes smaller when the age grows. This also causes a robustness problem for λ, since its posterior distribution is dependent with that of σ^2. Also ϕ seems to have a slight robustness problem, suggested by Figure 6 (d). On the contrary, ω does not suffer from robustness problems, but the reason is that the data do not contain enough information for its estimation, which implies that the posterior is practically the same as the prior. Since ω does not significantly deviate from unity, one could consider using only one smoothing parameter instead of two.

Figure 7 (a) indicates that under the final model the posterior of λ is more concentrated on small values for the younger age group. However, the difference between the age groups is not as clear as in the case of the preliminary model. Besides, the range of the distribution is fairly large in both cases.

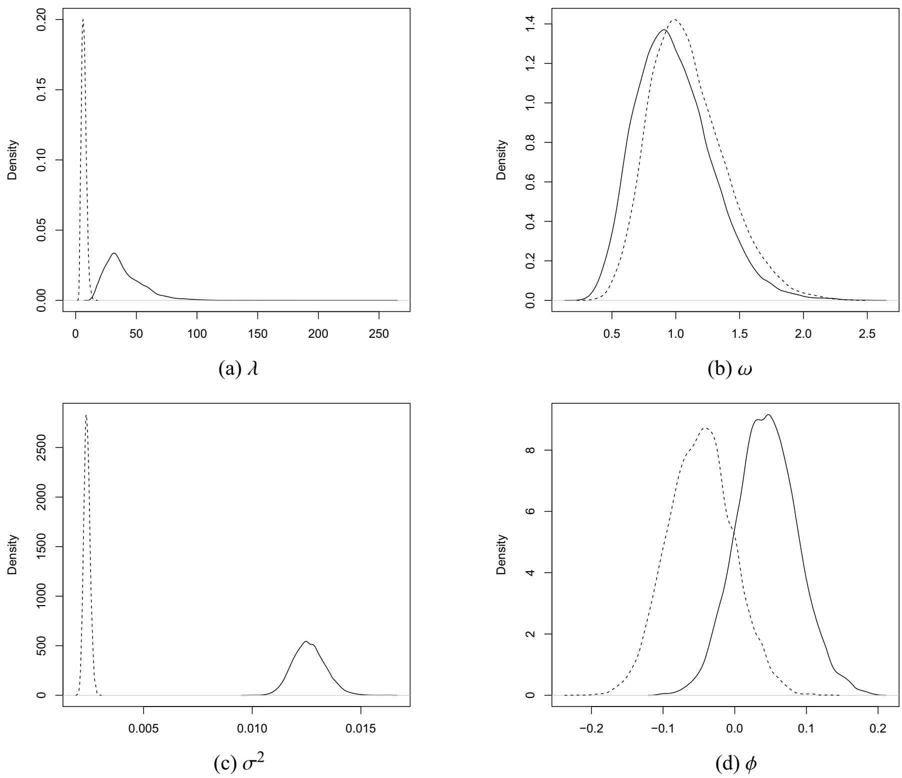


FIG. 6. Distributions of λ, ω, σ^2 and ϕ for the preliminary model. The solid line corresponds to the younger (ages 40–70) and the dashed line the older (ages 60–90) age group.

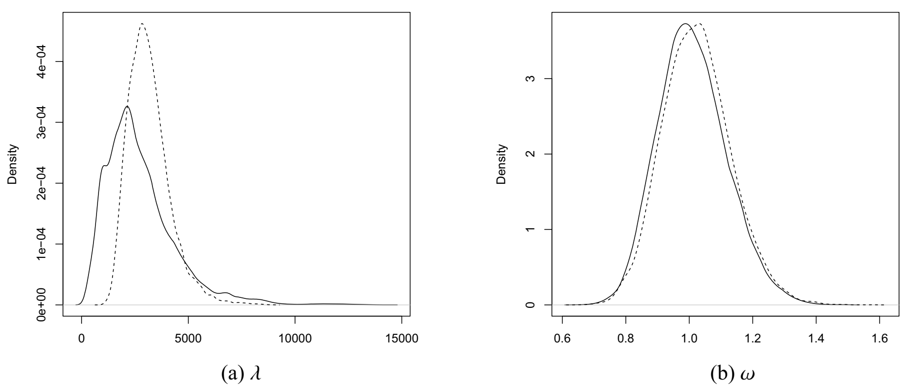


FIG. 7. Distributions of λ and ω for the final model. The solid line corresponds to the younger (ages 40–70) and the dashed line the older (ages 60–90) age group.

6.3. Forecasting

Our procedure for forecasting mortality is as follows. We first select a rectangular estimation area which includes in its lower right corner the ages and cohorts for which the death rates are to be predicted. Thus we have in our estimation set earlier observations from the same age as the predicted age and from the same cohort as the predicted cohort. An example of an estimation area is shown in Figure 1.

In the Bayesian approach, forecasting is based on the posterior predictive distribution. In the case of our preliminary model, a simulation from this distribution is drawn as follows: First, $\boldsymbol{\theta}$, σ^2 and ϕ are generated from their joint posterior distribution. Then the unobserved data vectors $\mathbf{y}_{j2}$, $j = 1, 2, ..., T$, (which are to be predicted) are generated from their conditional multivariate normal distributions, given the observed data vectors $\mathbf{y}_{j1}$ and the parameters $\boldsymbol{\theta}$, σ^2 and ϕ. These distributions were provided in Section 3. In the case of our final model, $\boldsymbol{\theta}$ is first generated. Then the numbers of deaths d_{xt} and the exposures e_{xt} are generated recursively starting from the most recent observed values within each cohort. In this way we obtain simulation paths for each cohort and a predictive distribution for each missing value in the mortality table. Further details are provided in Appendix B.

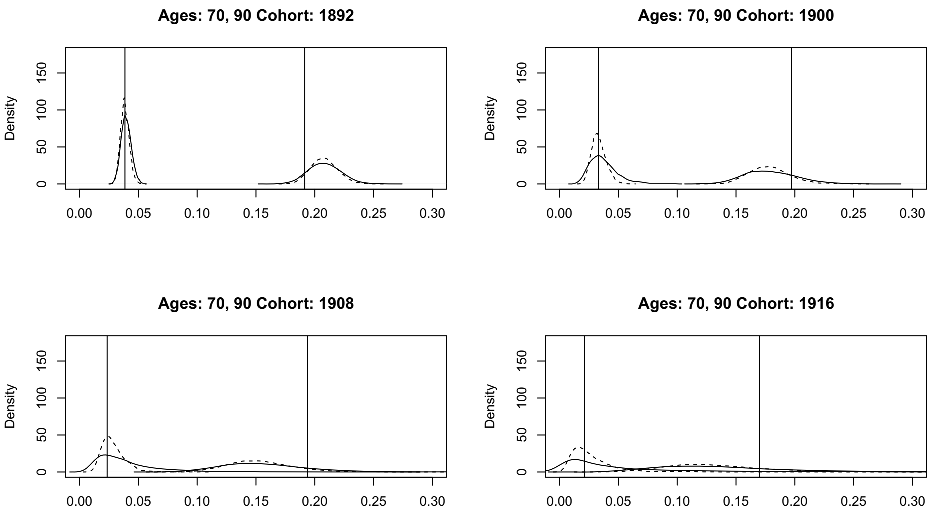


FIG. 8. Posterior predictive distributions of the death rates at ages 70 and 90, based on the preliminary model. The solid curves correspond to the larger data set (cohorts 1876 – 1916, and ages 30–70 when the the death rate at age 70 is predicted, and ages 50–90 when the death rate at age 90 is predicted) and the dashed curves the smaller (cohorts 1886 – 1916, and ages 40–70 when the the death rate at age 70 is predicted, and ages 60–90 when the death rate at age 90 is predicted). The vertical lines indicate the realized death rates.

In studying the accuracy and robustness of forecasts, we use estimation areas similar to those used earlier. However, we choose them so that we can compare the predictive distribution of the death rate with its realized value. The estimation is done as if the triangular area in the right lower corner of the estimation area, indicated in Figure 1, were not known. The posterior predictive distributions shown in Figure 8 are based on

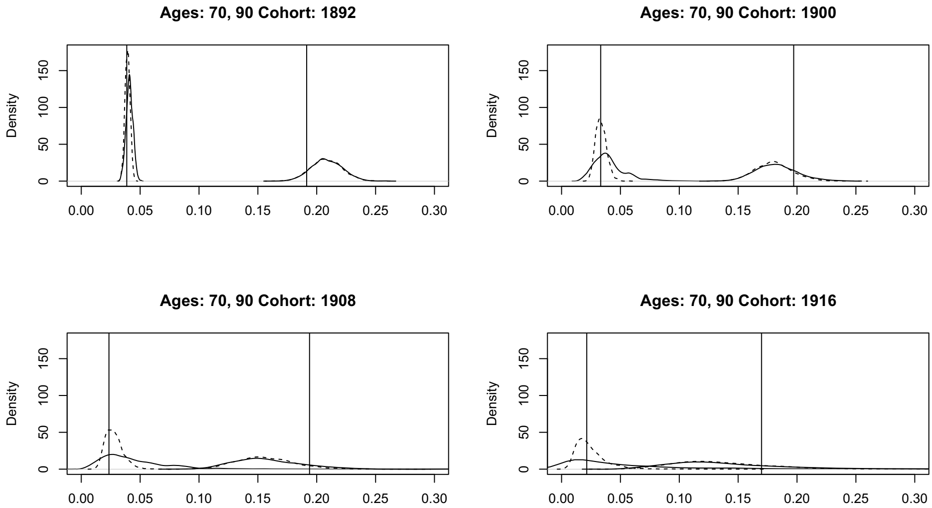


FIG. 9. Posterior predictive distributions of the death rates at ages 70 and 90, based on the final model. The solid curves correspond to the larger data set (cohorts 1876 – 1916; ages 30–70 when the the death rate at age 70 is predicted, and ages 50–90 when the death rate at age 90 is predicted) and the dashed curves the smaller (cohorts 1886 – 1916; ages 40–70 when the the death rate at age 70 is predicted, and ages 60–90 when the death rate at age 90 is predicted). The vertical lines indicate the realized death rates.

the preliminary model, while those in Figure 9 are based on the final model. The four cases in both figures correspond to forecasts 1, 9, 17 and 25 years ahead, for cohorts 1892, 1900, 1908 and 1916, respectively, when the death rate at ages 70 and 90 are forecast. The distributions indicated by solid lines are based on larger estimation sets than those indicated by dashed lines.

It may be seen that increasing uncertainty is reflected by the growing width of the distributions. Furthermore, the size of the estimation set does not considerably affect the distributions when the death rate at age 90 is predicted, while when it is predicted at age 70, the smaller data sets produce more accurate distributions. The obvious reason is that in the latter case the larger estimation set contains observations from the age interval 30–40 in which the growth of mortality is less regular than at larger ages, inducing more variability in the estimated model. In all cases, the realized values lie within the 90% prediction intervals.

Figures 10 and 11 show posterior predictive simulations when the preliminary and the final model is used, respectively. In each case, the results are shown for the cohorts 1891, 1904 and 1916. Three predictive simulation paths (gray lines) are shown for the cohorts 1904 and 1916. Their starting points indicate the beginning of the forecast region. As may be seen, their variability resembles that of the observed paths (thin black lines). The variability of observed death rates around the central trajectories (thick black lines) reflects sample variability around the expected values.

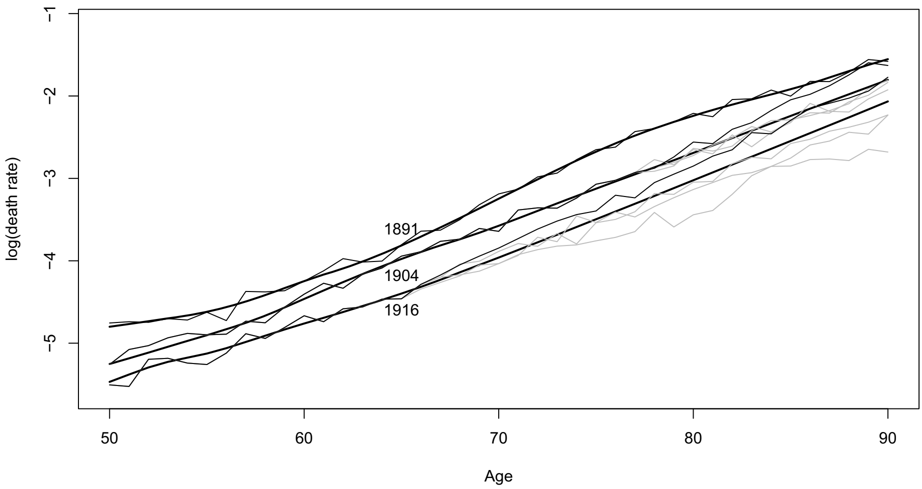


FIG. 10. Posterior predictions with the preliminary model for ages 50 − 90 and cohorts 1876 − 1916. The gray lines represent posterior predictions of log death rates, thin black lines their observed values, and thick black lines the averages of the posterior simulations of θ. The forecast region starts at ages 78 and 66 for cohorts 1904 and 1916, respectively.

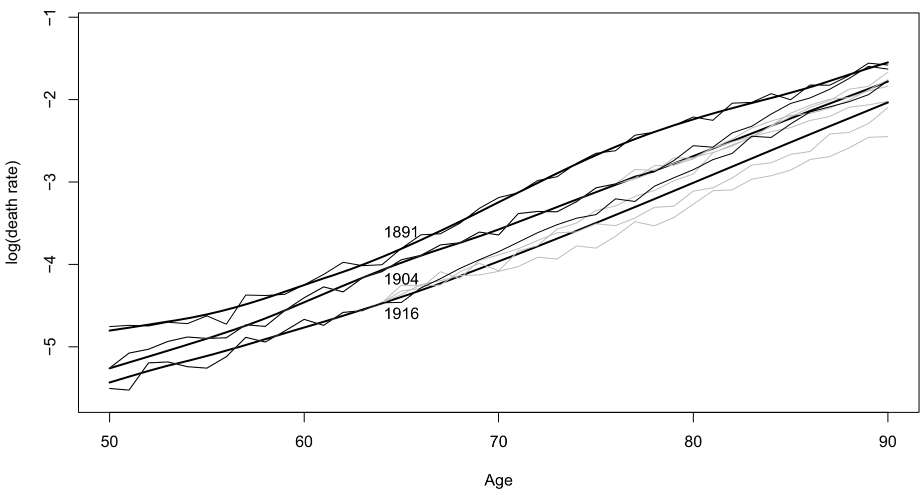


FIG. 11. Posterior predictions with the final model for ages 50 − 90 and cohorts 1876 − 1916. The gray lines represent posterior predictions of log death rates, thin black lines their observed values, and thick black lines the averages of the posterior simulations of θ. The forecast region starts at ages 78 and 66 for cohorts 1904 and 1916, respectively.

7. CONCLUSIONS

In this article we have introduced a new method to model mortality data in both age and cohort dimensions with Bayesian smoothing splines. The smoothing effect is obtained by means of a suitable prior distribution. The advantage in this approach compared to other splines approaches is that we do not need to optimize with respect to the number of knots and their locations. In order to take into account the serial dependence of observations within cohorts, we use cohort data sets, which are imbalanced in the sense that they contain fewer observations for more recent cohorts. We consider two versions of modeling: first, we model the observed death rates, and second, the numbers of deaths directly.

To assess the fit and plausibility of our models we follow the checklist provided by Cairns et al. (2008). The Bayesian framework allows us to easily assess parameter and prediction uncertainty using the posterior and posterior predictive distributions, respectively. In order to assess the consistency of the models with historical data we introduce test quantities. We find that our models are biologically reasonable, have non-trivial correlation structures, fit the historical data well, capture the stochastic cohort effect, and are parsimonious and relatively simple. Our final model has the further advantages that it has less robustness problems with respect to parameters, and avoids the heteroscedasticity of standardized residuals.

A minor drawback is that we cannot use all available data in estimation but must restrict ourselves to a relevant subset. This is due to the huge matrices involved in computations if many ages and cohorts are included in the data set. However, this problem can be alleviated using sparse matrix computations. Besides, for practical applications using "local" data sets should be sufficient.

In both models we have two smoothing parameters, controlling smoothing in the dimensions of cohort and age. Since it turned out that the data do not contain information to distinguish between these parameters, we might consider simplifying the model and using only one smoothing parameter. On the other hand, the model might be generalized by allowing for dependence between the smoothing parameter and age.

In conclusion, we may say that our final model meets well the mortality model selection criteria proposed by Cairns et al. (2008) except that it has a somewhat local character. This locality is partly due to limitations on the size of the estimation set and partly due to slight robustness problems related to the smoothing parameter and forecasting uncertainty.

ACKNOWLEDGMENTS

The authors are grateful to referees for their insightful comments and suggestions, which substantially helped in improving the manuscript. The second author of the article would like to thank the Finnish Academy of Science and Letters, Väisälä Fund, for the scholarship during which she could complete this project.

REFERENCES

Cairns, A.J.G., Blake, D., Dowd, K., 2006a. Pricing death: Frameworks for the valuation and securitization of mortality risk. *ASTIN Bulletin*, 36, 79–120.

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

Cairns, A.J.G., Blake, D., Dowd, K., 2008. Modelling and management of mortality risk: a review. *Scandinavian Actuarial Journal*, 2, 79–113.

Cairns, A.J.G., Blake, D., Dowd, K., Coughlan, G.D., Khalaf-Allah M., 2011. Bayesian stochastic mortality modelling for two populations. *ASTIN Bulletin*, 41, 29–59.

Currie, I.D., Durban, M., Eilers, P.H.C., 2004. Smoothing and forecasting mortality rates. *Statistical Modelling*, 4, 279–298.

Czado, C., Delwarde, A., Denuit, M., 2005. Bayesian Poisson log-linear mortality projections. *Insurance: Mathematics and Economics*, 36, 260–284.

Dellaportas, P., Smith, A.F.M., Stavropoulos, P., 2001. Bayesian analysis of mortality data. *Journal of the Royal Statistical Society*. Series A, 164, 275–291.

Gelman, A., Carlin, J.B., Stern, H.S., Rubin, D.B., 2004. *Bayesian Data Analysis*. Chapman & Hall/CRC.

Gelman, A. and Rubin, D.B., 1992. Inference from iterative simulation using multiple sequences. *Statistical Science*, 7, 457–511.

Geman, S. and Geman, D., 1984. Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 6, 721–741.

Gilks, W. R. and Wild, P., 1992. Adaptive rejection sampling for Gibbs sampling. *Applied Statistics*, 41, 337–348.

Gilks, W. R., Richardson, S., Spiegelhalter, D. (eds.), 1996. *Markov Chain Monte Carlo in Practice*. Chapman & Hall.

Green, P. and Silverman, B.W., 1994. *Nonparametric Regression and Generalized Linear Models*. Chapman & Hall, London.

Hastings, W.K., 1970. Monte Carlo sampling methods using Markov chains and their applications. *Biometrika*, 57, 97–109.

Human Mortality Database. University of California, Berkeley (USA), and Max Planck Institute for Demographic Research (Germany). Available at www.mortality.org or www.humanmortality.de (data downloaded on April 8, 2009).

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

Lang, S. and Brezger, A., 2004. Bayesian P-Splines. *Journal of Computational and Graphical Statistics*, 13, 183–212.

Lee, R.D. and Carter, L.R., 1992. Modeling and forecasting U.S. mortality. *Journal of the American Statistical Association*, 87, 659–675.

Metropolis, N., Rosenbluth, A.W., Rosenbluth, M.N., Teller, A.H., Teller, E., 1953. Equations of state calculations by fast computing machines. *Journal of Chemical Physics*, 21, 1087–1092.

Pedroza, C., 2006. A Bayesian forecasting model: predicting U.S. male mortality. *Biostatistics*, 7, 530–550.

Plat, R., 2009. On stochastic mortality modeling. *Insurance: Mathematics and Economics*, 45, 393–404.

R Development Core Team, 2010. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org/.

Reichmuth, W. and Sarferaz, S., 2008. Bayesian demographic modelling and forecasting: An application to US mortality. SFB 649 Discussion paper 2008-052.

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

Richards, S.J., Kirkby, J.G., Currie, I.D., 2006. The importance of year of birth in two-dimensional mortality data. *British Actuarial Journal*, 12, 5–38.

Schmid, V.J. and Held, L., 2007. BAMP - Bayesian age-period-cohort modeling and prediction. *Journal of Statistical Software*, 21, 8, Available at: http://www.jstatsoft.org.

Wahba, G., 1978. Improper priors, spline smoothing, and the problem of guarding against model errors in regression. *Journal of the Royal Statistical Society B*, 40, 364–372.

Wilmoth, J.R., Andreev, K., Jdanov, D., Glei, D.A., 2007. Methods Protocol for the Human Mortality Database. Available at: http://www.mortality.org.

APPENDIX A

The posterior simulations were performed using the R computing environment. The following outputs were obtained using the summary function of the add-on package MCMCpack:

TABLE 1.

Estimation results of the preliminary mortality model.

```
Number of chains = 3
Sample size per chain = 4800

1. Empirical mean and standard deviation for each variable,
   plus standard error of the mean:

            Mean        SD  Naive SE Time-series SE
lambda 16.940473 5.0795384 4.233e-02      4.001e-01
omega   1.010386 0.2781601 2.318e-03      1.765e-02
sigma2  0.004243 0.0001720 1.433e-06      3.552e-06
phi    -0.047612 0.0295280 2.461e-04      6.206e-04
theta1 -5.163545 0.0335889 2.799e-04      4.310e-04
theta2 -1.552719 0.0331673 2.764e-04      2.790e-04
theta3 -3.958345 0.0121454 1.012e-04      1.156e-04
theta4 -5.903402 0.0333465 2.779e-04      2.878e-04
theta5 -3.099490 0.2857720 2.381e-03      3.118e-03

2. Quantiles for each variable:

            2.5%       25%       50%       75%     97.5%
lambda  8.643654 13.198089 16.351764 20.291032 28.209809
omega   0.551153  0.814618  0.979278  1.175660  1.634720
sigma2  0.003923  0.004125  0.004237  0.004355  0.004597
phi    -0.104964 -0.067280 -0.047333 -0.027680  0.011365
theta1 -5.229142 -5.186360 -5.163524 -5.141287 -5.097599
theta2 -1.617886 -1.574978 -1.552572 -1.530468 -1.488354
theta3 -3.982101 -3.966441 -3.958372 -3.950303 -3.934482
theta4 -5.969645 -5.925643 -5.903172 -5.880918 -5.837735
theta5 -3.659055 -3.288343 -3.100254 -2.910501 -2.526937
```

```
Potential scale reduction factors:

       Point est. 97.5% quantile
lambda       1.02           1.08
omega        1.03           1.08
sigma2       1.00           1.01
phi          1.00           1.01
theta1       1.00           1.00
theta2       1.00           1.00
theta3       1.00           1.00
theta4       1.00           1.00
theta5       1.00           1.00

Multivariate psrf
1.02
```

TABLE 2.

Estimation results of the final mortality model.

```
Number of chains = 3
Sample size per chain = 3500

1. Empirical mean and standard deviation for each variable,
   plus standard error of the mean:

           Mean        SD  Naive SE Time-series SE
lambda 3635.488 921.74852 8.9953444      7.067e+01
omega     1.001   0.08031 0.0007838      1.945e-03
theta1   -5.169   0.04009 0.0003913      1.320e-03
theta2   -1.553   0.02386 0.0002329      7.500e-04
theta3   -3.955   0.01060 0.0001035      3.319e-04
theta4   -5.886   0.04394 0.0004288      1.288e-03
theta5   -3.064   0.28431 0.0027745      8.642e-03

2. Quantiles for each variable:

            2.5%      25%       50%      75%    97.5%
lambda 2173.3167 2952.634 3483.3201 4195.765 5756.134
omega     0.8506    0.946    0.9989    1.054    1.162
theta1   -5.2484   -5.198   -5.1685   -5.142   -5.091
theta2   -1.5998   -1.569   -1.5536   -1.536   -1.506
theta3   -3.9765   -3.962   -3.9548   -3.948   -3.934
theta4   -5.9737   -5.917   -5.8853   -5.856   -5.801
theta5   -3.6252   -3.264   -3.0631   -2.875   -2.515


Potential scale reduction factors:

       Point est. 97.5% quantile
lambda       1.06           1.19
omega        1.00           1.00
theta1       1.00           1.01
theta2       1.00           1.01
theta3       1.01           1.01
theta4       1.00           1.02
theta5       1.00           1.01

Multivariate psrf
1.04
```

APPENDIX B

In the case of the final model, the numbers of deaths d_{xt} and the exposures e_{xt} should be forecast for the ages and cohorts for which they are unknown. Furthermore, these values should be generated when replications of the original estimation data set are produced.

In the case of forecasting, we use an iterative procedure to generate d_{xt} and e_{xt}, starting from the most recent observation of death rate within each cohort. In the case of data replication, we start from the smallest age available in the data set. In each case, the initial cohort size is estimated on the basis of the relationship

$$q_{xt} = 1 - \exp(-\mu_{xt}),$$

where q_{xt} is the probability that a person in cohort t dies at age x. The same equality applies for the maximum likelihood estimates of q_{xt} and μ_{xt}, given by $\hat{q}_{xt} = d_{xt}/n_{xt}$ and $m_{xt} = d_{xt}/e_{xt}$, where n_{xt} is the number of persons reaching age x in cohort t. Thus, we obtain the formula

$$\frac{d_{xt}}{n_{xt}} = 1 - \exp\left(-\frac{d_{xt}}{e_{xt}}\right), \tag{17}$$

from which we may solve n_{xt} when d_{xt} and e_{xt} are known.

Further, the number of persons alive is updated recursively as $n_{x+1,t} = n_{xt} - d_{xt}$, and the number of deaths is generated from the binomial distribution:

$$d_{x+1,t} \sim \text{Bin}\left(n_{x+1,t}, q_{x+1,t}\right).$$

Then $e_{x+1,t}$ is solved using (17).