Netspar THESES

Richard Plat

Essays on Valuation and Risk Management for Insurers

Essays on Valuation and Risk Management for Insurers

Essays on Valuation and Risk Management for Insurers

ACADEMISCH PROEFSCHRIFT

ter verkrijging van de graad van doctor
aan de Universiteit van Amsterdam
op gezag van de Rector Magnificus
Prof. Dr. D.C. van den Boom
ten overstaan van een door het college voor promoties
ingestelde commissie,
in het openbaar te verdedigen in de Agnietenkapel
op dinsdag 1 Februari 2011, te 10.00 uur

door

Hendrikus Jozef Plat

geboren te Edam-Volendam

Promotiecommissie

Promotor:
Prof. dr. A.A.J. Pelsser

Overige leden:
Prof. dr. ir. M.H. Vellekoop
Prof. dr. R. Kaas
Prof. dr. A.J.G. Cairns
Prof. dr. A.M.B. De Wagenaere
Dr. K. Antonio

Faculteit der Economie en Bedrijfskunde

Preface

This thesis is the result of three years of (part-time) research at the Quantitative Economics department of the University of Amsterdam. The combination of doing research at the university and my job at insurer Eureko has been enjoyable, valuable and fruitful. One of the reasons for this is that the link between academia and the insurance industry has become stronger the last few years, which gives the opportunity to perform research that is directly applicable in the day-to-day business of insurance companies. During these years of research I have received support, in one form or the other, from a number of people.

First of all I would like to thank my supervisor Antoon Pelsser for his guidance, enthusiasm and ideas. Also, his early work on insurance contracts was an inspiration for me to start a PhD. Next, I would like to thank my co-authors Alexander van Haastrecht and Katrien Antonio. Alexander has always been very open and helpful, which provided a basis for having many interesting discussions about valuation and risk management, as well as discussions about topics that were not related to work at all. Katrien has a modest personality, but this cannot hide the fact that she is very good in her work. Next to this, it was very pleasant to work together on a paper. I would also like to thank the people from the actuarial department Rob Kaas, Angela van Heerwaarden, Michiel Janssen, Jan Kuné, Willem-Jan Willemse, Michel Vellekoop, Marc Goovaerts, Agnes Joseph and Julien Tomas for providing a pleasant and inspiring atmosphere at the university.

Furthermore, I am grateful to Eureko and Netspar for their financial support. At Eureko, I would like to thank my manager Martin Sandford for giving me the opportunity to perform a PhD and my colleagues of the Group Risk Management department for the excellent atmosphere and for taking into account the fact that I was only part-time available for Eureko.

Of course I would like to thank my friends, family and family-in-law for their interest and for providing the necessary distractions. Above all, I would like to thank Anne-Marie and my daughters Noa and Mila for being such a good reason to go home on time and not to think about valuation and risk management at all while being there.

Contents

Chapter 1

Introduction and Outline

Individual persons, companies and other entities are exposed to several risks that potentially can lead to undesirable financial consequences. For example, for an individual person it could be damage to a car, property damage, living longer or shorter than expected, expenses related to health and several other risks. Companies could be exposed to, amongst others, a liability claim, a company building on fire, damage to the products and disabled employees. These risks can be transferred by buying an insurance policy at an insurance company. In exchange for this the insurance company receives a premium from the policyholder. The insurance company pools the risks so that the results on the individual policies compensate each other.

As a result of writing insurance business for decennia, most insurers have to pay considerable amounts in the future to their policyholders. The company holds a reserve to cover for this, which is based on a valuation of these future insurance liabilities. Besides this, the insurance company is exposed to several risks, for which it holds additional capital. As such, valuation of insurance liabilities and measuring and managing the risks are two major building blocks for running an insurance company successfully. This thesis is a combination of papers on several issues related to valuation and risk management for insurers.

In the remainder of this chapter some more background is given on valuation and risk management for insurers, followed by an outline and discussion of the research presented in this thesis.

1.1 Valuation and Risk Management for Insurers

At this moment, most insurers are reporting their liabilities on a 'book value' basis, where the economic assumptions are often not directly linked to the financial market. Furthermore, regulators require additional (solvency) capital to be held by insurers which is a fixed percentage of the reserve, premiums or claims and thus not based on the actual risks of the insurer. However, in recent years there has been an increasing amount of attention of the insurance industry for market valuation of insurance liabilities and the quantification of insurance risks. Important drivers of this development are the introduction of IFRS 4 Phase 2 and Solvency 2.

With the introduction of Solvency 2 and IFRS 4 Phase 2 (both expected in 2013) insurers face major challenges. IFRS 4 Phase 2 will define a new accounting model for insurance contracts, based on market values of liabilities. In the document 'Preliminary Views on Insurance

Contracts' (May 2007, discussion paper) the International Accounting Standards Board (IASB) states that an insurer should base the measurement of all its insurance liabilities (for reserving) on best estimates of the contractual cash flows, discounted with current market discount rates. On top of this, margins that market participants are expected to require for bearing risk should be added to this. The IASB is currently further developing the standards, of which a consultation paper will appear in 2010.

Solvency 2 will lead to a change in the regulatory required solvency capital for insurers. Under Solvency 2 the so-called Solvency Capital Requirement (SCR) will be risk-based, and market values of assets and liabilities will be the basis for these calculations. The directive[1] of Solvency 2 prescribes that the reserve "... *shall be equal to the sum of the best estimate and a risk margin…*" and that "*the best estimate will correspond to the probability-weighted average of future cash-flows, taking account of the time value of money, using the relevant risk-free interest rate term structure*". Furthermore, it states that "*the calculation of the best estimate shall be based upon up-to-date and credible information and realistic assumptions, and be performed using adequate, applicable and relevant actuarial and statistical methods*".

The SCR aims to reflect all of the risks an insurance company is exposed to: market risk, operational risk, life underwriting risk, health underwriting risk, non-life underwriting risk, counterparty default risk and intangible asset risk. CEIOPS[2], the advising committee of the European Commission on Solvency 2, has developed a standard formula that leads to a required solvency margin that is aimed at covering all risks over a one-year horizon with a probability of 99,5%. However, insurance companies are encouraged to develop their own internal models to reflect the specific risks of the company more accurately.

Given the above, it is clear that the measurement of future cash flows and its uncertainty thus becomes more and more important.

1.2 Outline

This thesis consists of a collection of papers that each tackle a specific issue in valuation or risk management for insurers. First chapter 2 will cover some general concepts that are used throughout the thesis, mainly relating to stochastic processes of some kind.

Life insurance products often have profit sharing features in combination with guarantees. Valuation of these so-called embedded options is one of the key challenges in market valuation of the insurance liabilities. Chapter 3 and 4 are both covering the valuation of specific embedded options. In chapter 3 analytical approximations for prices of swap rate dependent embedded options are developed. These options are very common in products of European insurers. Chapter 4 covers the valuation of Guaranteed Annuity Options, which have been written by U.K. insurance companies for many years. The valuation of embedded options is not only a valuation issue, it is also an important aspect in risk management. After all, the risk of variations in the

---

[1] See 'Directive of the European parliament and of the council on the taking-up and pursuit of the business of insurance and re-insurance (Solvency 2)' of the European parliament.
[2] Committee of European Insurance and Occupational Pensions Supervisors

prices of embedded options is a risk element that has to be managed by the insurance company, for example by hedging this risk exposure.

Important risks to be quantified for Life insurers (and pension funds) are mortality and longevity risk. Chapter 5 and 6 will both cover different aspect in quantifying these risks. Chapter 5 will introduce a new stochastic mortality model for the population of a country. Chapter 6 will focus on another stochastic model that is the missing link to come to a full stochastic mortality model for specific insurance portfolios. The latter also gives the opportunity to quantify the basis risk that is involved when insurance portfolios are hedged with instruments of which the payoff depends on country population mortality rates.

The other underwriting risks, related to the health and non-life business, are treated in chapter 7. Usually, reserving and risk management for this business is based on actuarial techniques that are applied to aggregated data. This chapter describes a new stochastic reserving technique on the level of individual claims (micro-level).

The remainder of this chapter contains a short introduction on the subjects covered in the different chapters.

1.2.1   Chapter 3: Valuation of Swap Rate Dependent Embedded Options

Many life insurance products have profit sharing features in combination with guarantees. These so-called embedded options are often dependent on or approximated by forward swap rates. In practice, these kinds of options are mostly valued by Monte Carlo simulation, a computer intensive calculation technique. However, for risk management calculations and reporting processes, lots of valuations are needed. Therefore a more efficient method to value these options would be helpful.

In this chapter analytical approximations are derived for these kinds of options. The analytical approximation for options where profit sharing is paid directly is almost exact while the approximation for compounding profit sharing options is also satisfactory. In addition, the proposed analytical approximation can be used as a control variate in Monte Carlo valuation of options for which no analytical approximation is available, such as similar options with management actions. This considerately speeds up the calculation process for these options. Furthermore, it's also possible to construct analytical approximations when returns on additional assets (such as equities) are part of the profit sharing rate.

1.2.2   Chapter 4: Valuation of Guaranteed Annuity Options using a Stochastic Volatility Model for Equity Prices

Guaranteed Annuity Options are options providing the right to convert a policyholder's accumulated funds to a life annuity at a fixed rate when the policy matures. These options were a common feature in UK retirement savings contracts issued in the 1970's and 1980's when interest rates were high, but caused problems for insurers as the interest rates began to fall in the 1990's. Currently, these options are frequently sold in the U.S. and Japan as part of variable annuity products.

The last decade the literature on pricing and risk management of these options evolved. Until now, for pricing these options generally a process for equity prices is assumed where volatility is constant. However, given the long maturities of the insurance contracts a stochastic volatility model for equity prices would be more suitable. In this chapter explicit expressions are derived for prices of guaranteed annuity options assuming stochastic volatility for equity prices and either a 1-factor or 2-factor Gaussian interest rate model. The results indicate that the impact of ignoring stochastic volatility can be significant.

1.2.3 Chapter 5: On stochastic mortality modeling

The last decennium a vast literature on stochastic mortality models has been developed, mainly for use in risk management. All well known models have nice features but also disadvantages. In this chapter a stochastic mortality model is proposed that aims at combining the nice features from existing models, while eliminating the disadvantages. More specifically, the model fits historical data very well, is applicable to a full age range, captures the cohort effect, has a non-trivial (but not too complex) correlation structure and has no robustness problems, while the structure of the model remains relatively simple. Also, the chapter describes how to incorporate parameter uncertainty in the model. Furthermore, a version of the model is given that can be used for pricing.

1.2.4 Chapter 6: Stochastic portfolio specific mortality and the quantification of mortality basis risk

Chapter 5 will describe several stochastic mortality models that have been developed over time, usually applied to mortality rates of a country population. However, these models are often not directly applicable to insurance portfolios because:
a)   For insurers and pension funds it is more relevant to model mortality rates measured in insured amounts instead of measured in number of policies.
b)   Often there is not enough insurance portfolio specific mortality data available to fit such stochastic mortality models reliably.

Therefore, in this chapter a stochastic model is proposed for portfolio specific mortality experience. Combining this stochastic process with a stochastic country population mortality process leads to stochastic portfolio specific mortality rates, measured in insured amounts. The proposed stochastic process is applied to two insurance portfolios, and the impact on the height of the longevity risk is quantified. Furthermore, the model can be used to quantify the basis risk that remains when hedging portfolio specific mortality risk with instruments of which the payoff depends on population mortality rates.

1.2.5 Chapter 7: Micro-level stochastic loss reserving

The last decennium also a substantial literature about stochastic loss reserving for the non-life insurance business has been developed. Apart from few exceptions, all of these papers are based on data aggregated in run-off triangles. However, such an aggregate data set is a summary of an underlying, much more detailed data based that is available to the insurance company. This data set at individual claim level as will be referred to as 'micro-level data'. In this chapter it is investigated whether the use of such micro-level claim data can improve the reserving process. A realistic micro-level data set on general liability claims (material and injury) from a European insurance company is modeled. Stochastic processes are specified for the various aspects involved in the development of a claim: the time of occurrence, the delay between occurrence

and the time of reporting to the company, the occurrence of payments and their size and the final settlement of the claim. These processes are calibrated to the historical individual data of the portfolio and used for the projection of future claims. Through an out-of-sample prediction exercise it is shown that the micro-level approach provides the actuary with detailed and valuable reserve calculations. A comparison with results from traditional actuarial reserving techniques is included. For our case-study reserve calculations based on the micro-level model are preferable: compared to traditional methods, they reflect real outcomes in a more realistic way.

Chapter 2

Stochastic processes

At the heart of most valuation and all risk management calculations are assumptions about the stochastic processes of the relevant variables. Stochastic processes required for valuation are often of a different nature than the stochastic processes required for risk management.

For the valuation of embedded options it is important that the underlying stochastic model is arbitrage free. Arbitrage free means that it is not possible to generate a non-zero payoff without any initial investment. A convenient way to accomplish this is the use of a so-called 'risk-neutral' model. The risk-neutral stochastic processes used in this thesis are described in section 2.1.

For risk management it is more important that the stochastic processes are as realistic as possible reflecting the dynamics of the underlying stochastic variable. This means that a 'real-world' model is required. The real-world stochastic processes used in this thesis are described in section 2.2.

2.1 Risk Neutral Stochastic Processes for Valuation

In this thesis the topics regarding valuation of embedded options require arbitrage free stochastic processes for interest rates and equity prices. The stochastic processes used are members of a more general class of models, the affine jump-diffusions. This section describes this general class of models and the specific interest rate and equity model used in this thesis. This will be preceded by a short introduction in the notion of martingales and measures. The section ends with a short discussion about stochastic processes for valuation of unhedgeable insurance risks.

2.1.1 Martingales and Measures

The foundation of option pricing theory is the assumption that arbitrage opportunities do not exist. Another important underlying concept is *completeness* of the economy. If in an economy the payoffs of all derivative securities can be replicated by a self-financing trading strategy, the economy is called complete. If no arbitrage opportunities and no transaction costs exist in an economy, the value of a self-financing trading strategy should be equal to the value of the corresponding derivative. If this would not be the case, arbitrage opportunities exist.

Harrison and Kreps (1979) and Harrison and Pliska (1981) brought the concepts of arbitrage free and completeness together in what is called 'The Fundamental Theorem of Asset Pricing'. Any

asset which has strictly positive prices for all future times is called a numéraire. Numéraires can be used to denominate all prices in an economy (instead of Euro's or Dollars). A martingale is a stochastic process with a zero drift. Harrison and Kreps (1979) and Harrison and Pliska (1981) proved that a continuous economy is complete and arbitrage free if for every choice of numéraire there exists a unique equivalent martingale measure. In other words, given a choice of numéraire, there is a unique probability measure such that the relative price processes are martingales. This important result is very useful for option valuation.

For example, say that price at time t of an option H maturing at time T relative to the price of security M is defined as V. Then under the relevant measure Q^M the process V is a martingale. This means that:

$$(2.1) \quad V(t) = E^M \left[V(T) \right] \Rightarrow \frac{H(t)}{M(t)} = E^M \left[\frac{H(T)}{M(T)} \right] \Rightarrow H(t) = M(t) \, E^M \left[\frac{H(T)}{M(T)} \right]$$

where $E^M[x]$ is the expectation under the relevant measure. By choosing a convenient numéraire the option price calculation can be simplified considerably in some cases.

Usually as a starting point the riskless money-market account is used as the numéraire. Under the unique probability measure corresponding to this numéraire the expected return on all assets is equal to the risk-free rate. Therefore, this measure is called the risk-neutral measure, usually denoted as Q. Often stochastic processes intended to be used for valuation are defined in the risk-neutral measure. However, sometimes it is more convenient to change to another measure.

Consider two numéraires N and M with the martingale measures Q^N and Q^M. Geman et al (1995) proved that the Radon-Nikodym derivative that changes the equivalent martingale measure Q^M into Q^N is given by:

$$(2.2) \quad \frac{dQ^N}{dQ^M} = \frac{N(T)/N(t)}{M(T)/M(t)} = \boldsymbol{r}(t)$$

Girsanov's Theorem states that if this Radon-Nikodym derivative can be written as:

$$(2.3) \quad \boldsymbol{r}(t) = \exp \left[\int_0^t \boldsymbol{k}(s) \, dW^M(s) - 0.5 \int_0^t \boldsymbol{k}(s)^2 \, ds \right]$$

where W^M is a Brownian motion under the measure Q^M. This leads to:

$$(2.4) \quad W^N(t) = W^M(t) - \int_0^t \boldsymbol{k}(s) \, ds \quad or \quad dW^M = dW^N + \boldsymbol{k}(t) \, dt$$

So in order to use Girsanov's Theorem the process $\boldsymbol{k}(t)$ has to be found that yields (2.3). An application of Ito's Lemma shows that $d\boldsymbol{r}(t) = \boldsymbol{r}(t)\boldsymbol{k}(t)dW^M$, showing that $\boldsymbol{r}(t)$ is a martingale

under the measure Q^M under the condition $\mathrm{E}\left[\exp\left\{\dfrac{1}{2}\displaystyle\int_0^t |\boldsymbol{k}(s)|^2\,ds\right\}\right] < \infty$. Now applying Ito's Lemma to the ratio (2.2) will give $\boldsymbol{k}(t)$.

2.1.2 Affine Jump-Diffusions

The stochastic processes used in this thesis for interest rates and equity prices are part of a broader class of models, called the affine jump-diffusions. A class of affine models was introduced first in the context of interest rates by Duffie and Kan (1996). Later this is generalized by Duffie et al (2000) and Duffie et al (2003). The class of affine jump-diffusions provides a flexible and general model structure combined with analytical tractability. The latter feature facilitates the calibration and simulation of such models. Well known term structure models that are members of this class are, amongst others, the models of Hull and White (1993), Cox et al (1985) and Longstaff and Schwartz (1992). Next to the equity price model of Black and Scholes (1973) also the stochastic volatility models of Heston (1993), Schöbel and Zhu (1999) and the stochastic volatility with jumps model of Bates (1996) are members of this class.

The class of affine jump-diffusions can be defined as follows. Let X be a real-valued n-dimensional Markov process satisfying:

$$(2.5) \quad dX(t) \;=\; \boldsymbol{m}\big(X(t)\big)\,dt + \boldsymbol{s}\big(X(t)\big)\,dW(t) + dZ(t)$$

Where $W(t)$ is a standard Brownian motion in $\Re^n$, $\boldsymbol{m}(\cdot) \in \Re^n$, $\boldsymbol{s}(\cdot) \in \Re^{n \times n}$, and Z is a pure jump process whose jumps have a fixed probability distribution v and arrive with intensity $\boldsymbol{l}(X(t))$. The jump times of Z are the jump times of a Poisson process with time-inhomogeneous intensity. Poisson processes are further highlighted in section 2.2. The process X is affine if and only if the diffusion coefficients are of the following form:

$$(2.6) \quad \boldsymbol{m}(x) \;=\; K_0 + K_1 x \qquad\qquad \text{for } K=(K_0,K_1) \in \Re^n \times \Re^{n \times n}$$

$$(2.7) \quad \big(\boldsymbol{s}(x)\boldsymbol{s}(x)^T\big)_{ij} \;=\; (H_0)_{ij} + (H_1)_{ij}\,x \qquad\qquad \text{for } H=(H_0,H_1) \in \Re^{n \times n} \times \Re^{n \times n \times n}$$

$$(2.8) \quad \boldsymbol{l}(x) \;=\; l_0 + l_1 x \qquad\qquad \text{for } l=(l_0,l_1) \in \Re \times \Re^n$$

$$(2.9) \quad r(x) \;=\; \boldsymbol{r}_0 + \boldsymbol{r}_1 x \qquad\qquad \text{for } \boldsymbol{r}=(\boldsymbol{r}_0,\boldsymbol{r}_1) \in \Re^n \times \Re^{n \times n}$$

where $r(\mathrm{x})$ is the short term interest rate. Now it can be proved that the characteristic function of $X(t)$, including the effects of any discounting, is known in closed form up to the solution of a system of Ordinary Differential Equations. Duffie et al (2000) show that for $u \in C^n$ the Fourier transform $\boldsymbol{f}(u,X(t),t,T)$ of $X(t)$, conditional on filtration F_t, is given by:

$$(2.10) \quad \boldsymbol{f}(u, X(t), t, T) \;=\; \mathrm{E}\left[e^{-\int_t^T r(X(s))ds} \, e^{uX(t)} \mid F_t \right] \;=\; e^{A(t)+B(t)X(t)}$$

where $A(\cdot)$ and $B(\cdot)$ satisfy the following system of Ricatti equations:

$$(2.11) \quad \frac{dA(t)}{dt} \;=\; \boldsymbol{r}_0 - K_0 B(t) - \frac{1}{2} B(t)^T H_0 B(t) - l_0 \big[\boldsymbol{q}(B(t)) - 1 \big]$$

$$(2.12) \quad \frac{dB(t)}{dt} \;=\; \boldsymbol{r}_1 - K_1^T B(t) - \frac{1}{2} B(t)^T H_1 B(t) - l_1 \big[\boldsymbol{q}(B(t)) - 1 \big]$$

with boundary conditions $A(T) = 0$ and $B(T) = u$. The 'jump transform' $\boldsymbol{q}(\cdot)$ is given by:

$$(2.13) \quad \boldsymbol{q}(c) \;=\; \int_{\Re^n} e^{cz}\, dv(z)$$

In general the solutions of $A(\cdot)$ and $B(\cdot)$ have to be computed numerically, although the well known models mentioned above result in explicit expressions for $A(\cdot)$ and $B(\cdot)$.

2.1.3 Gaussian interest rate models

In this thesis the underlying interest rate model for the valuation is the class of multi-factor Gaussian models. Special cases of this class of models are the 1-factor and 2-factor Hull-White model, which are often used in practice. These models are appealing because of their analytical tractability.

The Gaussian interest rate models are also a special case of the affine term structure models introduced by Duffie and Kan (1996). The *m*-factor Gaussian model describes the stochastic process for the instantaneous short rate as follows[3]:

$$(2.14) \quad r(t) = \boldsymbol{1}' Y(t) + \boldsymbol{a}(t)$$

$$(2.15) \quad dY(t) = -CY(t)dt + \Sigma\, dW^Q(t)$$

where $W^Q(t)$ is a *m*-dimensional Brownian motion under the risk-neutral measure and *C* and **S** are *m* x *m* matrices. *C* is a diagonal matrix.

The function $\boldsymbol{a}(t)$ is chosen in such a way that the fit of the model to the initial term structure is perfect. The covariance matrix of the *Y*-variables is equal to **SS**'.

The analytical tractability of this model makes it possible to obtain bond prices analytically, from which swap and zero rates can be derived. The price at time *t* of a zero bond maturing at time *T* is given by:

---

[3] See Brigo & Mercurio (2006) for an extensive explanation of and pricing formulas for the 2-factor Gaussian model.

$$(2.16) \quad D(t,T) = A(t,T) \exp\left(-\sum_{i=1}^{m} B^{(i)}(t,T) Y^{(i)}(t)\right)$$

where $B^{(i)}(t,T) = 1 / A_{(ii)}\left(1 - \exp(-A_{(ii)}(T-t))\right)$

The expression for $A(t,T)$ is further specified for the 1-factor and 2-factor case in chapter 4.

2.1.4 Stochastic volatility model for equity prices

In a seminal paper Black & Scholes (1973) made a major breakthrough in the pricing of equity options. The underlying stochastic model for equity prices has become known as the Black-Scholes model. The Black-Scholes model assumes the volatility to be constant. However, in practice the volatility varies through time. For this reason a significant literature has evolved on alternative models that incorporate stochastic volatility. Next to leading to more realistic dynamics of the stochastic process for equity prices, these models have the advantage that they provide a better fit of the model to actual market (option) data. This is an important feature for being able to adequately price more exotic options such as embedded options in insurance products. Well known stochastic volatility models are the models of Hull and White (1987), Stein and Stein (1991), Heston (1993) and Schöbel and Zhu (1999).

The aim in chapter 4 is to combine a stochastic volatility model for equity prices with a stochastic interest rate model. Van Haastrecht et al (2009) show that it is possible to obtain an explicit expression for the price of European equity options when the Schöbel and Zhu (1999) model is combined with a stochastic Gaussian model for interest rates, explicitly taking into account the correlation between those processes. That makes this combined model suitable for valuation of the Guaranteed Annuity Options in chapter 4.

In the Schöbel and Zhu (1999) model, the process for equity price $S(t)$ under the risk-neutral measure Q is:

$$(2.17) \quad \frac{dS(t)}{S(t)} = r(t)dt + v(t)dW_s^Q(t) \qquad S(0) = S_0$$

$$(2.18) \quad dv(t) = \mathbf{k}\left(\mathbf{y} - v(t)\right)dt + \mathbf{t}\,dW_v^Q(t) \qquad v(0) = v_0$$

Here $v(t)$, which follows an Ornstein-Uhlenbeck process, is the (instantaneous) stochastic volatility of the equity $S(t)$. The parameters of the volatility process are the positive constants *?* (mean reversion), v_0 (short-term mean), *?* (long-term mean) and t (volatility of the volatility).

2.1.5 Stochastic processes for valuation of unhedgeable insurance risks

The valuation of insurance liabilities also requires the valuation of (unhedgeable) insurance risks. For example, mortality models for the valuation of mortality or longevity liabilities (or derivatives) are given by Dahl (2004), Schrager (2006), Cairns et al (2006b) and Bauer et al (2008). The models of Dahl (2004) and Schrager (2006) belong to the general class of affine

jump-diffusions defined in paragraph 2.1.2 and as a result allow for closed form expressions of the survival rate.

Usually insurance risk models are calibrated to historical data and are therefore defined in the real world measure, denoted by *P*. Given the techniques mentioned in paragraph 2.1.1, one could apply a change of measure to risk neutral measure *Q*, under which the insurance liability can be valued. However, in this case one crucial condition is not satisfied, being the completeness of the economy. As explained in paragraph 2.1.1, the completeness of the economy forces the risk neutral measure *Q* to be unique. The market for insurance risks is far from complete, meaning that the insurance risks are unhedgeable and therefore a range of possibilities for *Q* exist. As mentioned by Cairns et al (2006a) the choice of *Q* needs to be consistent with the limited market information, but beyond this restriction the choice of *Q* becomes a modeling assumption.

An alternative method for valuation in incomplete markets is the use of utility functions and the *principle of equivalent utility*, see Young and Zariphopoulou (2002), Young and Moore (2003) and Young (2004). This principle implies that the maximal expected utility with and without the specific insurance risk are examined. The compensation at which the insurer is indifferent between the two alternative alternatives yields the value of the unhedgeable insurance risk. However, this approach is currently only feasible for relatively simple products.

2.2 Real World Stochastic Processes for Risk Management

As mention above, for risk management it is particularly important that the stochastic processes used realistically reflect the observed characteristics of the underlying stochastic variable. In chapter 5 and 6 parametric models are fit to yearly observations, leading to time series of fitted variables. Stochastic processes have to be fit to these time series, for which the Autoregressive Integrated Moving Average (ARIMA) models can be used. These are described in paragraph 2.2.1. The stochastic processes needed in chapter 7 are of a different nature and are described in paragraph 2.2.2.

2.2.1 ARIMA Time Series Models

A seminal work on the estimation and identification of ARIMA models is the monograph by Box and Jenkins (1976). Additional details and discussion of more recent topics can be found in for example Mills (1990), Enders (2004) and Hamilton (1994). An important issue is whether a time series process is stationary, meaning that the distribution of the variable of interest does not depend on time. If this is not case, the first step would be to difference the time series until the differenced time series is stationary. Box and Jenkins found that usually only one or two differencing operations are required.

The general ARIMA(p, d, q) model for a time series of a variable y_t can be written as:

(2.10) $\quad \mathrm{V}^d\, y_t = y_t^*$

$$y_t^* = a_0 + \sum_{i=1}^{p} a_i y_{t-i}^* + e_t + \sum_{j=1}^{q} b_i e_{t-i}$$

where the **a**'s and **b**'s are the unknown parameters, the **e**'s are independent and identically distributed normal errors and $\boldsymbol{D}^d$ represents the differencing, meaning $\boldsymbol{D}^0 y_t = y_t$, $\boldsymbol{D}^1 y_t = y_t - y_{t-1}$, $\boldsymbol{D}^2 y_t = (y_t - y_{t-1}) - (y_{t-1} - y_{t-2})$, etc. The parameter p is the number of lagged values of y_t, representing the order of the autoregressive (AR) dimension of the model, and q is the number of lagged values of the error term, representing the order of the moving average (MA) dimension of the model.

Box and Jenkins define three steps for the development of an ARIMA model:

1) *Model identification and model selection*: determining the values for p, d, q.
2) *Parameter estimation*: either by using Maximum Likelihood or (non-linear) Least Squares estimation.
3) *Diagnostic checking*: testing whether the estimated model meets the specifications of a stationary univariate process.

Often an extension is needed to allow the modeling of multivariate time series. This requires a multivariate generalization of the ARIMA process, see for example Verbeek (2008).

2.2.2 Poisson processes and renewal processes

The required stochastic processes in chapter 7 are of a different nature than those described above. Poisson processes and the related renewal processes are convenient concepts for modeling the development process of individual claims. For an extensive overview of these techniques, see Cook and Lawless (2007).

Poisson Processes   A Poisson process describes situations where events occur randomly in such a way that the numbers of events in non-overlapping time intervals are independent. Poisson processes are therefore *Markov*, with an intensity function:

$$(2.11) \quad \boldsymbol{I}\left(t \mid H(t)\right) \;=\; \lim_{\Delta t \downarrow 0} \frac{\Pr\left(N(t+\Delta t) - N(t) = 1\right)}{\Delta t} \;=\; \boldsymbol{r}(t)$$

Where $N(t)$ is the cumulative number of events occurring over the time interval $[0,t]$ and $H(t)$ is the process history. In the case where $\boldsymbol{r}(t)$ is constant, $\boldsymbol{r}(t) = \boldsymbol{r}$, the process is called homogeneous. Otherwise, it is inhomogeneous. The above specification implies:

$$(2.12) \quad N(t) - N(s) \;\sim\; Poisson\left(\int_s^t \boldsymbol{r}(u)\,du\right)$$

Position Dependent Marked Poisson Process (PDMPP)   In chapter 7 the individual claims process is modeled as a PDMPP. A marked Poisson process with intensity $\boldsymbol{r}(t)$ and position-dependent marks is a process

$$(2.13) \quad \left((T_i, Z_i)\right)_{i=1,\dots,N}$$

where the claims counting process $N(t)$ is an inhomogeneous Poisson point process with intensity $r(t)$, points T_i and marks Z_i. The $(Z_t)_{t>0}$ are mutually independent, are independent of the Poisson point process $N(\cdot)$ and have time-dependent probability assumptions.

Renewal processes   Related to the Poisson process is the renewal process, in which the waiting (gap) times between successive events are statistically independent: that is, an individual is 'renewed' after each event occurrence. Renewal models for waiting times are defined as processes for which

$$(2.14) \quad \lambda\left(t \mid H(t)\right) \;=\; h\!\left(t - T_{N(t^-)}\right)$$

where $h(\cdot)$ is the hazard rate and $t - T_{N(t^-)}$ is the time since the most recent event before t.

Often used models for the time to an event, say T, are the Exponential, Weibull and the Gompertz distribution. These distributions have the convenient property that the hazard function has a simple form. The following hazard functions $g(u)$ are implied by these distributions:
- $T \sim \text{Exponential}(\lambda) \Rightarrow h(u) = \lambda$ (constant hazard)
- $T \sim \text{Weibull}(\alpha, \gamma) \quad \Rightarrow h(u) = \alpha\gamma u^{\gamma-1}$
- $T \sim \text{Gompertz}(\alpha, \gamma) \Rightarrow h(u) = \gamma e^{\alpha u}$

Other possibilities are a piecewise constant specification for the hazard rate or the Cox proportional hazard model (see Cox (1972)).

Chapter 3

Valuation of swap rate dependent embedded options*

* This chapter has appeared as:

PLAT, R. AND A.A.J. PELSSER (2008): Analytical approximation for prices of swap rate dependent embedded options in insurance products, *Insurance: Mathematics and Economics 44*, pp. 124-134

3.1 Introduction

An important part of the market valuation of liabilities is the valuation of embedded options. Embedded options are options that have been sold to the policyholders and are often the more complex features in insurance products. An embedded option that is very common in insurance products in Europe, is a profit sharing rule based on a (moving average) fixed income rate, in combination with a minimum guarantee. This fixed income rate is either from an external source or could be the book value return on a fixed income portfolio. For example, in the Netherlands the profit sharing is often based on the so-called u-yield, which is more or less an average return of several treasury rates. In other parts of Europe, the book value return on the fixed income portfolio is often the basis for the profit sharing. In practice the exact rates are difficult to determine and to project forward, and implied volatilities from the market are not available. Therefore, often the euro swap rate is used as a proxy. So what remains is the valuation of an option on a moving or weighted average of forward and historic swap rates.

Most insurers use Monte Carlo simulations for the valuation of their embedded options. The advantage of this is that many kinds of options can be valued with it (including the more complex ones) and that it gives one uniform simulation framework that is applicable for various embedded options. However, an important disadvantage is the computational time it requires. Embedded option calculations are required for Fair Value reporting, Market Consistent Embedded Value, Asset Liability Management, product development and pricing, Economic Capital calculations and Mergers & Acquisitions. For most of these purposes several calculations are required. For the calculation of Economic Capital for example 20.000 or more simulations

are used and in each of these scenario's the market value of liabilities (and thus the value of embedded options) has to be calculated. Also for other purposes, often sensitivities and analysis of changes are necessary. If an insurer then also exists of several business units or legal entities, the total computational time can be significant. Therefore, analytical solutions for the valuation of embedded options would be very helpful.

In this chapter analytical approximations are derived for the above mentioned swap rate dependent embedded options. The underlying interest rate model is a multi-factor Gaussian model. This model is very appealing because of its analytical tractability. Also, the model implicitly accounts for the volatility skew to some extent, what is important for these kind of options because those are in most cases not at-the-money. Because of this the model is often used in practice (in most cases the 1-factor or 2-factor Hull-White variant). Analytical approximations are derived for the case of direct payment of profit sharing, as well as for the case of compounding profit sharing. In case of (very) complex options with management actions, the analytical approximation for the direct payment case can be used as a control variate in combination with Monte Carlo simulation, reducing the computational time to a great extent.

It could well be that an insurance company has other kinds of embedded options for which no analytical approximations are available. These embedded options probably have to be valued using Monte Carlo simulation. Since the multi-factor Gaussian models are often used in practice, the analytical approximation for the swap rate dependent options can in that case be used in conjunction with the simulation model that may be required for the valuation of other embedded options. This results in a consistent underlying interest rate model for the valuation of embedded options, despite the fact that perhaps some of the options are valued with Monte Carlo simulations and others with analytical formulas.

The basis for the analytical approximation is the result of Schrager and Pelsser (2006), who have developed an approximation for swaption prices for affine term structure models (of which the multi-factor Gaussian models are a subset). They determine the dynamics of the swap rate under the relevant swap measure and these dynamics are approximated by replacing some low-variance martingales by their time zero values. This technique is already used extensively in the context of Libor Market Models and given the results of Schrager and Pelsser, it also proves to work well in an affine setting. By use of the Change of Numéraire techniques developed by Geman et al (1995), the result of Schrager and Pelsser can be used to derive analytical approximations for swap rate dependent options.

Most of the existing literature on valuation of embedded options in insurance products focuses on Unit Linked products, with-profits products or Guaranteed Annuity Options. For example, Grosen and Jorgensen (2000), Schrager and Pelsser (2004) and Castellani et al (2007) developed analytical approximations for Unit Linked type products with guarantees. Wilkie et al (2003) use numerical techniques to value Guaranteed Annuity Options, while Sheldon and Smith (2004) developed analytical formulas for these products. Nielsen and Sandmann (2002) and Prieul et al (2001) use numerical techniques for valuation of With-Profits contracts.

However, to our knowledge there has been little focus on profit sharing based on (moving average) fixed income rates, despite this being one of the most common types of profit sharing in

Europe. Our contribution to the existing literature is that we provide analytical approximations for these kinds of profit sharing. Analytical approximations for direct payment of profit sharing and for compounding profit sharing are given, while a combination with returns on other assets (such as equities) is also possible. In addition, the proposed analytical approximation can be used as a control variate in Monte Carlo valuation of options for which no analytical approximation is available, such as similar options with management actions. This potentially reduces the number of simulations required to a great extent.

Some of the techniques proposed in this chapter can also be used for financial products, such as options on an average of Constant Maturity Swap (CMS) rates, (callable) CMS accrual swaps and (callable) CMS range notes.

The remainder of the chapter is organized as follows. First, in section 3.2 the characteristics of the swap rate dependent embedded options are described. In section 3.3 the underlying Gaussian interest model is given. In section 3.4 the Schrager-Pelsser result for swaptions is repeated and this is applied to the direct payment case in section 3.5. In section 3.6 possibilities are given for more complex embedded options. Then numerical examples are worked out in section 3.7 and conclusions are given in section 3.8.

3.2 Swap rate dependent embedded options

Traditional non-linked life insurance products often guarantee a certain insured amount. Common practice was (and often still is) to calculate the price of this insurance by discounting the expected cash flows with a relatively low interest rate, called the technical interest rate. Often this is combined with profit sharing, where some reference return is paid out to the policyholder if this exceeds the technical interest rate, possibly under subtraction of a margin. There exist various types of profit sharing, such as:

- Profit sharing based on an external reference index
- Profit sharing based on the (book or market value) return on the underlying investment portfolio
- Profit sharing based on the performance and profits of the insurance company
- Profit sharing of the so-called with-profits products, where regular and terminal bonuses are given though the life of the product, based on the return of the underlying investment portfolios. The terms of these policies often contain management actions that allow the insurance companies to reduce the risks of these products.

In most cases where the profit sharing rate depends on a certain fixed income rate, the exact profit sharing rate is either very complex or not fully known, or implied volatilities from the market are not available. In practice, these kinds of options are often valued using an (average) forward swap rate as an approximation for the profit sharing rate. The profit sharing payoff *PS(t)* in year *t* is in that case:

(3.1) $PS(t) = L(t) \, Max\{c\,(R(t) - K(t)), 0\}$

where $L(t)$ is the profit sharing basis, c is the percentage that is distributed to the policyholder and $K(t)$ is the strike of the option. The strike equals the sum of the technical interest rate $TR(t)$ and a margin. In most cases, either the margin or the c is used for the benefits of the insurer. $R(t)$ is the profit sharing rate and is a (weighted) average of historic and forward swap rates.

The profit sharing as described in (3.1) is a call option on a rate $R(t)$ and has to be valued using option valuation techniques. The profit sharing is either paid directly or is being compounded and paid at the end of the contract.

Note that it depends on the specific profit sharing rules whether the swap rate is a good approximation for the profit sharing rate. This has to be verified for each specific profit sharing arrangement. Below two examples are given of profit sharing arrangements where the swap rate is often used as approximation in practice.

Example 1 – book value return on underlying portfolio
One of the most common forms of profit sharing across the European life insurance business is the one where the profit sharing rate is based on the book value return of the underlying fixed income portfolio[4]. To be able to value this option, assumptions have to be made about the reinvestment strategy. An example of how this problem is often tackled in practice is to assume:
  - a certain average turnover rate d
  - a reinvestment strategy favoring m-year maturity assets.
  - the m-year swap rate being an approximation for the yield on the m-year maturity assets

Given these assumptions the book value return of the portfolio can be modeled as follows:

$$(3.2) \quad R(t) = (1-d)\,R(t-1) + d\,y_{t,t+m}(t)$$

where $y_{t,t+m}(t)$ is the m-year swap rate at time t. The book value return on time t can also be expressed in terms of the current book value return $R(0)$, leading to an exponentially weighted moving average:

$$(3.3) \quad R(t) = (1-d)^t\,R(0) + \sum_{i=0}^{t} y_{i,i+m}(i)(1-d)^{t-i}\,d$$

being a weighted combination of forward swap rates and the current book value return.

Another approach that is often used is approximating the book value return by a moving average of swap rates:

$$(3.4) \quad R(t) = \frac{1}{n}\sum_{i=t-n+1}^{t} y_{i,i+m}(i)$$

where $n\ (=1/d)$ is the number of fixings of the moving average.

---

[4] This is common practice in for example France, Germany, Italy, Czech Republic, Switzerland and Norway.

Example 2 – "u-rate" profit sharing in the Netherlands
In the Netherlands the most common form of profit sharing is based on a moving average of the so-called u-rate. The u-rate is the 3-months average of u-rate-parts, where the subsequent u-rate-parts are weighted averages of an effective return on a basket of government bonds. This leads to a complicated expression, and no implied volatilities are available for government bonds. Therefore, it is common practice in the Netherlands to approximate the u-rate or the u-yield parts by a swap-rate[5]. That means that the profit sharing rate is approximated by a moving average of swap rates, as in (3.4).

Besides the direct payment and compounding versions of (3.1), other variants of this profit sharing exist, such as:

1) Profit sharing including the return on an additional asset
2) (Compounding) profit sharing with additional management actions or other complex features.

In case of 1), the underlying investment portfolio also contains additional non-fixed income assets. This means that the profit sharing rate is a combination of a (weighted) moving average of swap rates and the return on additional assets. The profit sharing rate could then be expressed as:

$$(3.5) \quad R^*(T_i) = \sum_{k=T_{i-s}}^{T_i} w_k^{FI} y_{k,k+m}(k) + \sum_j w_j^S r_{S_j}$$

where w_j^S is the weight in additional asset S_j, r_{S_j} is the return on that asset and

$$\sum w_k^{FI} + \sum w_l^S = 1.$$

In case of 2), the insurer has added management actions or other complexities to the profit sharing rules, mainly to lower the risk exposure for the insurer.

In the following sections analytical approximations are developed for prices of embedded options where the profit sharing rate depends on or is approximated by forward swap rates. Note that the developed formulas are approximating swap rate dependent embedded options. When considering the results or using the formulas one always has to be aware of the fact that the first error is introduced when the swap rate is being used as a proxy for the profit sharing rate.

3.3 The underlying interest rate model

The analytical approximations in this chapter are based on an underlying multi-factor Gaussian interest rate model. This model is described in paragraph 3.3.1. Paragraph 3.3.2 gives a

---

[5] Historical data that show that u-rate parts have behaved similarly as swap rates in the past, is available upon request.

discussion whether similar techniques as developed in this chapter can be used for analytical valuation of the options described in section 3.2 given other underlying interest rate models.

3.3.1 Multi-factor Gaussian models

As mentioned in paragraph 2.1.3, the underlying interest rate model for the valuation is the class of multi-factor Gaussian models. These models are very appealing because of their analytical tractability. This makes the model easy to implement, while there are also more possibilities for analytical approximations (or solutions) for embedded options

In the swaption market, the observed implied Black volatility is varying for different strike levels, leading to the so-called volatility skew. This volatility skew exists because the market apparently does not believe in lognormally distributed swap rates. Instead, the volatility skew seems to indicate a distribution that is closer to the normal distribution[6]. Therefore, the Gaussian models implicitly account for the volatility skew to a certain extent. This is also an appealing property of these models in the context of embedded options in insurance products, since these options are in most cases not at-the-money.

3.3.2 Valuation for other interest rate models

This paragraph gives a discussion whether similar techniques as developed in this chapter can be used for analytical valuation of the options described in section 3.2 given other underlying interest rate models.

General affine models
Schrager and Pelsser (2006) developed approximations for swaption prices for general affine interest rate models. For non-Gaussian affine models they come to an approximate solution for swaption prices for which only a numerical integration is necessary. An approximation for the characteristic function of the swap rate under the swap measure and the method of Carr and Madan (1999) is used for this. As a first step in this process they derive approximate dynamics for the swap rate in similar fashion as described in section 3.4. With an additional approximation a square-root process for the swap rate results.

Dassios and Nagaradjasarma (2006) develop explicit prices for Asian options, given an underlying square root process. They also obtain distributional results concerning the square-root process and its average over time, including analytic formulae for their joint density and moments.

For the embedded options discussed in this chapter a suggested approach would be to use the approximate dynamics for the swap rate from Schrager and Pelsser (2006) and combine this with the techniques in Dassios and Nagaradjasarma (2006).

Libor Market Model (LMM)
As mentioned in section 3.4, the approximation technique used in this chapter is already used extensively in the context of Libor Market Models. For example, Brigo and Mercurio (2006) use

---

[6] See Levin (2004) for a discussion on this issue.

the technique for approximation of swaption prices in the LMM model. Gatarek (2003) uses it to approximate prices of Constant Maturity Swaps.

Now when using this technique, the resulting distribution of the approximate swap rate in the LMM model is lognormal. However, for the valuation of the embedded options in this chapter the distribution of the average swap rate is needed. In case the swap rate is lognormally distributed, the distribution of the average swap rate is unknown. This is a well known problem in the context of valuation of Asian options. Methods for approximate analytical valuation of options on the average of lognormally distributed variables are proposed in, amongst others, Levy (1992), Curran (1994) and Rogers and Shi (1995). Lord (2006) gives an overview of existing methods, compares the quality of those numerically and develops approximations that outperform the other methods.

Swap Market Model (SMM)
In a standard SMM as proposed by Jamshidian (1997) each swap rate is modeled in its own swap measure, making it hard to apply for pricing of most exotic interest rate products. This could be one of the reasons that the SMM has not been discussed extensively in financial literature. The co-sliding SMM proposed by, amongst others, Pietersz and Van Regenmortel (2006) seems promising though and is applicable especially for Constant Maturity Swap (CMS) and swap rate products.

In the SMM the swap rate is modeled directly in a lognormal setting, so no approximation of the distribution of the swap rate in the swap measure is necessary. A price for the profit sharing options discussed in this chapter can be obtained by applying the relevant convexity and timing adjustments and using one of the above mentioned techniques for approximate analytical valuation of Asian options.

3.4 The Schrager-Pelsser result for swaptions

Schrager and Pelsser (2006) developed an approximation for swaption prices for affine interest rate models. In this section their main result for the Gaussian models is repeated.

The swap rate $y_{n,N}$ is the par swap rate at which a person would like to enter into a swap contract with a value of 0, starting at time T_n (first payout at time T_{n+1}) and lasting until T_N. The swap rate at time t is given by:

$$(3.6) \quad y_{n,N}(t) = \frac{D(t,T_n) - D(t,T_N)}{\sum_{k=n+1}^{N} \Delta_{k-1}^{Y} D(t,T_k)} = \frac{D(t,T_n) - D(t,T_N)}{P_{n+1,N}(t)}$$

where Δ_{k-1}^{Y} is the market convention for the calculation of the daycount fraction for the swap payment at T_k. When using $P_{n+1,N}(t)$ as a numéraire, all $P_{n+1,N}(t)$ rebased values must be martingales under the measure $Q^{n+1,N}$, associated with this numéraire. That means that $y_{n,N}$ is a martingale under this so-called swap measure, which is introduced by Jamshidian (1998). When

applying Ito's Lemma to the model defined in (2.14) and (2.15) the following dynamics for the swap rate $y_{n,N}(t)$ under the swap measure result:

$$(3.7) \quad dy_{n,N}(t) = \frac{\partial y_{n,N}(t)}{\partial Y(t)} \Sigma \, dW^{n+1,N}(t)$$

Where $dW^{n+1,N}$ is a m-dimensional Brownian motion under the swap measure $Q^{n+1,N}$ corresponding to the numéraire $P_{n+1,N}(t)$. Schrager and Pelsser (2006) determine the partial derivatives in (3.7), which are stochastic, and approximate these by replacing low-variance martingales by their time zero values. This technique is already used extensively in the context of Libor Market Models[7] and given the results of Schrager and Pelsser, it also proves to work well in an affine setting. This approximation makes the swap rate volatility deterministic and thus leads to a normally distributed forward swap rate. The approach described leads to an analytical approximation for the integrated variance of $y_{n,N}$ (associated with a $T_n \times T_N$ swaption) over the interval $[0,T_n]$ (for the proof, see appendix 3a):

$$(3.8) \quad s_{n,N}^2 \approx \sum_{i=1}^{m} \sum_{j=1}^{m} \hat{\Sigma}_{(ij)} \tilde{C}_{n,N}^{(i)} \tilde{C}_{n,N}^{(j)} \left[\frac{e^{\left[A_{(ii)}+A_{(jj)}\right]T_n} - 1}{A_{(ii)} + A_{(jj)}} \right]$$

where $\hat{\Sigma}_{(ij)}$ is the element (i,j) of $\dot{a} \, \dot{a}'$ and

$$(3.9) \quad \tilde{C}_{n,N}^{(i)} = \frac{1}{A_{(ii)}} \left[e^{-A_{(ii)}T_n} D^P(0,T_n) - e^{-A_{(ii)}T_N} D^P(0,T_N) - y_{n,N}(0) \sum_{k=n+1}^{N} \Delta_{k-1}^{Y} e^{-A_{(ii)}T_k} D^P(0,T_k) \right]$$

where $D^P(t,T_n) = D(t,T_n) / P_{n+1,N}(t)$, the bond price normalized by the numéraire.

The result is an easy to implement analytical approach to calibrate Gaussian models to the swaption market. A nice by-product of the approach (as opposed to other approaches for approximating swaption prices) is that the dynamics of the swap rates are approximated. These approximate dynamics can be used for approximating prices of other swap-rate dependent options.

3.5 Analytical approximation – direct payment

Assume that the profit sharing rate at time T_i is a weighted average of t-year maturity swap rates with weights w_k and the averaging period is from time T_{i-s} to time T_i:

$$(3.10) \quad R(T_i) = \sum_{k=T_{i-s}}^{T_i} w_k \, y_{k,k+t}(k)$$

where $S \, w_k = 1$.

---

[7] See Andersen and Andreasen (1998), Gatarek (2003) and Brigo and Mercurio (2006).

In case of direct payment of profit sharing, the embedded option is in fact a strip of options that mature at time T_i (i=1,2,…) and lead to a direct payment of an option payoff on $R(T_i)$ on these dates. Since the individual $y_{k,k+t}(k)$'s are approximately normally distributed (see section 3.4), $R(T_i)$ is also approximately normally distributed. So to value the option the expectation and the variance of $R(T_i)$ have to be approximated under the T_i-forward measure and feed into a Gaussian option formula for each time T_i. For determining the variance of $R(T_i)$ the covariance's of the $y_{k,k+t}(k)$'s with the $y_{l,l+t}(k)$'s have to be specified.

3.5.1 Determining the expectation of R(Tᵢ)

The above means that each individual option has to be priced in the T_i-forward measure. To come to the expectations of $R(Ti)$ under the right measure the following steps are necessary:

a) For each (forward) swap rate $y_{n,N}$ a change of measure has to be done from the swap measure $Q^{n+1,N}$ to the T_n-forward measure Q^{Tn}.

b) If the payoff of the option on the average of the swap rates is at time T_i, for each of the individual swap rates observed at time (T_{i-s}), a change of measure has to be done from the (T_{i-s})-forward measure to the T_i-forward measure.

The corrections mentioned above can be interpreted as convexity corrections (a) and timing corrections (b). The formulas for these corrections are given in (3.11) and (3.12), of which the proofs are given in appendix 3b. Note that due to the changes of measure it's not guaranteed that the quality of the approximation will remain. Therefore, this will be tested in section 3.7.

The convexity correction $CC_{n,N}(T_n)$ for time $T_n > 0$ for the swap rate $y_{n,N}$ is:

$$(3.11) \quad CC_{n,N}(T_n) \approx \sum_{i=1}^{m}\sum_{j=1}^{m} \hat{\Sigma}_{(ij)} \tilde{C}_{n,N}^{(i)} \tilde{G}_{n,N}^{(j)} \left[\frac{e^{\left[A_{(ii)}+A_{(jj)}\right]T_n}-1}{A_{(ii)} + A_{(jj)}} \right]$$

$$\text{where} \quad \tilde{G}_{n,N}^{(j)} = \frac{1}{A_{(jj)}}\left[e^{-A_{(jj)}T_n} - \sum_{k=n+1}^{N} \Delta_{k-1}^{Y} e^{-A_{(jj)}T_k} D^P(0,T_k) \right]$$

The timing correction $TC_{n,N}(T_n,T_{n+u})$ representing a change of measure from time $T_n > 0$ to T_{n+u} is:

$$(3.12) \quad TC_{n,N}(T_n,T_{n+u}) \approx \sum_{i=1}^{m}\sum_{j=1}^{m} \hat{\Sigma}_{(ij)} \tilde{C}_{n,N}^{(i)} \tilde{H}_{T_n,T_{n+u}}^{(j)} \left[\frac{e^{\left[A_{(ii)}+A_{(jj)}\right]T_n}-1}{A_{(ii)} + A_{(jj)}} \right]$$

$$\text{where} \quad \tilde{H}_{T_n,T_{n+u}}^{(j)} = \frac{1}{A_{(jj)}}\left[e^{-A_{(jj)}(T_{n+u})} - e^{-A_{(jj)}T_n} \right]$$

For $T_n < 0$, the convexity corrections and the timing corrections are 0. Note that in the derivation of (3.11) also stochastic terms are replaced by their time zero values, leading to a deterministic convexity correction.

The expectation $\boldsymbol{m}_{R(T_i)}$ of $R(T_i)$ becomes:

$$(3.13) \quad \boldsymbol{m}_{R(T_i)} \approx \sum_{k=T_{i-s}}^{T_i} w_k \left[y_{k,k+\boldsymbol{t}}(0) + CC_{k,k+\boldsymbol{t}}(k) + TC_{k,k+\boldsymbol{t}}(k,T_i) \right]$$

The convexity correction is positive and the timing correction is negative, so they are partly offsetting each other. The formulas (3.11) and (3.12) have the same structure as in case of the swaptions in section 3.4, so the implementation is not much more complicated than that.

3.5.2 Determining the variance of $R(T_i)$

Given that the drift term is deterministic, the change of measure has no impact on the volatility, so expression (3.8) can be used to determine the variance of $R(T_i)$. The variance $\boldsymbol{s}_{R(T_i)}^2$ of $R(T_i)$ is:

$$(3.14) \quad \boldsymbol{s}_{R(T_i)}^2 = \sum_{k=T_{i-s}}^{T_i} \sum_{l=T_{i-s}}^{T_i} w_k w_l \, Cov[y_{k,k+\boldsymbol{t}}(k), y_{l,l+\boldsymbol{t}}(l)]$$

where $Cov(.)$ is the covariance between the swap rates. From stochastic calculus we know (for $s \le t$):

$$(3.15) \quad Cov\left[\int_0^t f(u)dW_u, \int_0^s g(u)dW_u \right] = \int_0^s f(u)g(u)du$$

Using this and expression (3.8) the covariance between swap rates is

$$Cov[y_{k,k+\boldsymbol{t}}(k), y_{l,l+\boldsymbol{t}}(l)] \approx \int_0^{k \wedge l} e^{As'} diag(\tilde{C}_{k,k+\boldsymbol{t}}) \Sigma \Sigma' diag(\tilde{C}_{l,l+\boldsymbol{t}}) e^{As} \, ds$$

$$(3.16)$$

$$= \sum_{i=1}^m \sum_{j=1}^m \hat{\Sigma}_{(ij)} \tilde{C}_{k,k+\boldsymbol{t}}^{(i)} \tilde{C}_{l,l+\boldsymbol{t}}^{(j)} \left[\frac{e^{\left[A_{(ii)} + A_{(jj)}\right](k \wedge l)} - 1}{A_{(ii)} + A_{(jj)}} \right]$$

where $k \wedge l = \min(k,l)$.

3.5.3 Pricing formulas

The total value of the embedded option is the sum of the values of the strip of options that mature at time T_i (i=1,2,…). The profit sharing specified in (3.1) is in fact a call option on the normally distributed rate $R(T_i)$ with expectation (3.13) and variance (3.14) under the T_i-forward measure.

Let $\boldsymbol{j}_{\boldsymbol{ms}}(\cdot)$ be the density of a Gaussian random variable with mean $\boldsymbol{m}$ and standard deviation $\boldsymbol{s}$, $\boldsymbol{F}_{\boldsymbol{ms}}$ the corresponding distribution function and $\boldsymbol{F} = \boldsymbol{F}_{0,1}$.

The value at time 0 of the profit sharing payoff $PS(T_i)$ at time T_i is[8]:

$$V[PS(T_i)] = D(0,T_i)\,L(T_i)\,c\,E^{Q^{T_i}}\left[Max\{R(T_i) - K(T_i),0\}\right]$$

(3.17)
$$= D(0,T_i)\,L(T_i)\,c\int_{K(T_i)}^{\infty}(x - K(T_i))\,\boldsymbol{f}_{\boldsymbol{m}_{R(Ti)},\boldsymbol{s}_{R(Ti)}}(x)\,dx$$

$$= D(0,T_i)\,L(T_i)\,c\left[(\boldsymbol{m}_{R(T_i)} - K(T_i))\,\Phi\left(\frac{\boldsymbol{m}_{R(T_i)} - K(T_i)}{\boldsymbol{s}_{R(T_i)}}\right) + \boldsymbol{s}_{R(T_i)}\,\boldsymbol{f}\left(\frac{K(T_i) - \boldsymbol{m}_{R(T_i)}}{\boldsymbol{s}_{R(T_i)}}\right)\right]$$

The total value of the profit sharing at time 0 is then:

(3.18) $$V[PS] = \sum_i V[PS(T_i)]$$

When the profit sharing payoff at a time > 0 is dependent on observations at a time < 0, a slight adjustment has to be done. In that case the expectation to be valued is:

$$V[PS(T_i)] = D(0,T_i)\,L(T_i)\,E^{Q^{T_i}}\left[Max\{R(T_i) - K(t),0\}\right]$$

(3.19)
$$= D(0,T_i)\,L(T_i)\,E^{Q^{T_i}}\left[Max\{R(T_i)_{t>0} + R(T_i)_{t\leq0} - K(t),0\}\right]$$

$$= D(0,T_i)\,L(T_i)\,E^{Q^{T_i}}\left[Max\{R(T_i)_{t>0} - K^*(t),0\}\right]$$

$$where\ R(T_i)_{t\leq0} = \sum_{k=T_{i-s}}^{T_j=0} w_k\,y_{k,k+t}(k),\ R(T_i)_{t>0} = \sum_{k=T_j}^{T_i} w_k\,y_{k,k+t}(k)\ and\ K^*(t) = K(t) - R(T_i)_{t\leq0}$$

So these profit sharing options can be priced with a relatively simple and relatively easy to implement Gaussian option formula.

3.6 Valuation for more complex profit sharing rules

In section 3.5 an analytical approximation is derived for the case of direct payment of the profit sharing payoff specified in (3.1). However, in practice other variants of this profit sharing exist, such as:

1) Compounding variant of the profit sharing in (3.1)
2) Profit sharing including the return on an additional asset
3) (Compounding) profit sharing with additional management actions or other complex features

---

[8] These results can be derived in a similar fashion in case of a put-option on rate $R(T_i)$.

For 1) and 2), an analytical approximation can be derived in line with the approximation developed in section 3.5. For 3), either volatility scaling or Monte Carlo simulation will be necessary. In case of Monte Carlo simulation, the approximation in (3.17) can be used as a control variate, potentially reducing the amount of simulations necessary to a great extent.

3.6.1 Compounding profit sharing

It is also common that profit sharing is not paid directly, but is compounded and paid out at the end of the contract term. Valuation of this option with Monte Carlo simulation often takes a significant amount of time. The reason for this is the dependency of the profit sharing rates on the future cash flows, resulting in the need to use the original liability cash flow model in a stochastic way. An analytical approximation would significantly (even more than in the direct payment case) reduce computational time, since these formulas can be used as input for the liability cash flow model without the need to run these stochastically.

Let the maturity of the product be T_n and total payoff $L(T_n)$ be of the form:

$$(3.20) \quad L(T_n) = L(0) \prod_{i=0}^{n} s(T_i)\left[1 + TR(T_i) + Max\{c\left(R(T_i) - K(T_i)\right), 0\}\right]$$

where the definition of the variables is as in (3.1) and $s(T_i)$ is the probability that the policyholder stays in the portfolio.

The distribution of the right term of (3.20) is unknown so there is no analytical expression for this payoff. However, if we assume that the $R(T_i)$'s are independent (which is obviously a crude assumption in this case), the expectation of $L(T_n)$ under the T_n-forward measure is:

$$\begin{aligned}
(3.21) \quad E^{T_n}\left[L(T_n)\right] &= E^{T_n}\left[L(0)\prod_{i=0}^{n} s(T_i)\left[1 + TR(t) + Max\{c\left(R(T_i) - K(T_i)\right), 0\}\right]\right] \\
&\approx L(0)\prod_{i=0}^{n} s(T_i)\left[1 + TR(T_i) + E^{T_n}\left(Max\{c\left(R(T_i) - K(T_i)\right), 0\}\right)\right]
\end{aligned}$$

where the latter expectations can be calculated with (3.17), excluding the term $D(0,T_i) \, L(T_i)$. Note that this expectation has to be determined under the T_n-forward measure by making a timing correction to time T_n using formula (3.12).

The value of the compounding profit sharing option would then be:

$$\begin{aligned}
(3.22) \quad V[PS] &= D(0,T_n)\left[E^{T_n}\left[L(T_n)\right] - K\right] \\
where \ K &= \prod_{i=1}^{n}\left[1 + TR(T_i)\right]
\end{aligned}$$

Despite the crude assumption on independence, the analytical approximation could still work well. When the expected $R(T_i)$'s are low, the impact of the compounding effect is relatively low, resulting in a relatively good approximation of the time value of the option. When the expected

$R(T_i)$'s are high, the impact of the compounding effect is relatively high and the quality of the approximation will be less (in terms of time value). However, in this case the total value of the option will also be high and the impact of approximation errors in the time value on the total value will be less. This reasoning is being tested in section 3.7.

Instead of using this analytical approximation, it is also possible to use Monte Carlo simulation with the analytical approximation of (3.17) as a control variate, reducing the amount of simulations needed significantly. This technique is further described in paragraph 3.6.3.

3.6.2 Profit sharing including the return on an additional asset

In some cases the underlying investment portfolio also contains additional non-fixed income assets. The profit sharing rate could then be expressed as in (3.5).

Assume that the additional asset class S_j follows a standard geometric Brownian motion under the risk neutral measure Q:

$$(3.23) \quad dS_j(t) = S_j(t)\left[r(t)\,dt + \boldsymbol{s}_{S_j}\,dW_{S_j}^{Q}(t)\right]$$

In this case there is an analytical expression for the distribution of return r_{Sj} and the covariances with $y_{k,k+m}$, under normally distributed stochastic interest rates in a T-forward measure. The analytical expression for the distribution of r_{Sj} is worked out in Brigo and Mercurio (2006) for the 1-factor model and the result is similar for multi-factor models. The covariance's with $y_{k,k+t}$ can be determined using (3.15) and the formulas in Brigo and Mercurio (2006).

In practice, often r_{Sj} is a book value return. The specification of this book value return can be complex and possibly differs for every insurance company. Often, Monte Carlo simulations are necessary. However, an alternative is the approach described above, where the volatility parameters $\boldsymbol{s}_{Sj}$ can be calibrated to results of Monte Carlo simulation or derived from historical patterns of book value returns relative to total returns.

3.6.3 Additional management actions or other complex features

In some cases the insurer has added management actions or other complexities to the profit sharing rules, mainly to lower the risk exposure for the insurer. In most cases, it's not possible to properly value these options analytically. Other possibilities would then be:

a) Use a volatility scaling factor that is calibrated to results obtained with Monte Carlo simulation and use this as input for the analytical approximation in (3.17) and (3.22).
b) Value the option with Monte Carlo simulation, using the analytical approximation in (3.17) as a control variate.

Both possibilities are described below.

a) Volatility scaling factor

When the impact of the management actions or complexities is expected to be low or in cases where it is sufficient to use an approximation, one could use a volatility scaling factor $f(T_i)$, such that:

(3.24) $\quad \mathbf{s}_{R(T_i)}^{Adj} = [1 + f(T_i)] \mathbf{s}_{R(T_i)}$

The factor $f(T_i)$ can be calibrated for each time T_i to output from Monte Carlo simulation. This approach can be useful when lots of valuations are needed, for example for Economic Capital or Asset Liability Management calculations.

b) Control Variate technique

When the impact of the management actions or complexities is significant and exact valuation is necessary, Monte Carlo simulation can be used in conjunction with a control variate algorithm. For a thorough description of the control variate technique, see for example Glasserman (2004). When using the control variate algorithm, the value of the profit sharing is:

(3.25) $\quad V[PS] = V[PS]^{sim} - b\left(X^{sim} - E[X]\right)$

where $V[PS]^{sim}$ is the simulated value of the profit sharing option, X^{sim} is the simulated value of another asset *X* and *E[X]* is the expected value of *X*, which is assumed to be known. When choosing the proper control variate, the standard error of the Monte Carlo estimate can be reduced significantly. This means that significantly less simulations are needed to come to an estimate with the same quality as an ordinary Monte Carlo estimation.

The deterministic coefficient b that minimizes the standard error of the Monte Carlo estimation is given by:

(3.26) $\quad b = \dfrac{Cov(PS, X)}{Var(X)}$

The control variate algorithm is most effective when the correlation between *PS* and *X* is high. Therefore a suitable choice for the control variate would be a carefully selected combination of payer swaptions or CMS caplets[9].

An alternative can be the use of the direct payment option of section 3.5 as control variate. Since the management actions or complexities are added to a profit sharing as in (3.1), the correlation between this profit sharing and the direct payment variant of (3.1) is probably very high. Therefore, using the direct payment option of section 3.5 as a control variate would significantly reduce the number of simulations necessary. This can be implemented by adding the approximate dynamics (A.4) to the simulations to determine X^{sim} and using (3.17) to determine *E[X]*.

An example of the benefits of this technique is the following. In section 3.7 the quality of the approximation (3.17) is assessed. For testing this quality, the option values coming from (3.17) were in first instance compared with the result of 1.000.000 Monte Carlo simulations. The result from the simulations is seen as the "true" value, since the standard error of the estimation is

---

[9] The authors thank the anonymous referee for this suggestion.

sufficiently low for this number of simulations. Now when we use the same option (valued under the approximate dynamics) as a control variate and (3.17) as its expected value, only 1.000 simulations are needed to come to the same standard error. Of course in this case the correlation between the option to be valued and the control variate is almost maximal, but one could imagine that in case of more complex options the reduction of the number of simulations needed would still be substantial.

Whether the carefully selected combination of payer swaption / CMS caplets or the direct payment option of section 3.5 performs better as a control variate, will be subject for future research. An advantage of the selection of simpler instruments is that the market price of these instruments is usually available, so no model assumption has to be used for the valuation of this part.

3.7 Numerical examples

In this section the results of the approximation formulas will be shown for 2 example products and compared with the "true" values resulting from Monte Carlo simulation. When considering the results one has to be aware of the fact that before using the approximation already "errors" are introduced in the valuation, for example in the calibration of the interest rate model to market prices and by using the swap rate as a proxy for the profit sharing rate.

7.1 Example 1: 10-year average of 7-year swap rate, direct payment

This example is a specification of (3.1) and (3.4) with direct payment. This specification is for example commonly applied in pricing the u-rate profit sharing in the Netherlands, where the 7-year swap rate is often used as a proxy for the u-rate. Also, as in (3.4) it can be interpreted as a proxy for profit sharing based on the book value return on a underlying fixed income portfolio with an assumed turnover rate of 10% and a reinvestment strategy favoring 7-year maturity assets (on average). The underlying interest rate model used is a 2-factor Gaussian interest rate model.

The data used for the profit sharing basis and the technical interest rates are based on an example portfolio of a long term pension insurance portfolio, with cash flows up to 50 years ahead. This data is given in appendix 3c, along with the yield curve, implied volatility matrix and the specific parameter setting of the 2-factor Gaussian interest rate model. A margin of 0.5% is applied and c is assumed to be 1.

The analytical approximation described in section 3.5 is tested with Monte Carlo simulation, where 5000 (antithetic) simulations are used in combination with the control variate technique described in paragraph 3.6.3[10]. The results are given in table 3.1, where the total value of the option is given for both approaches and for different yield curve, volatility, mean reversion and strike sensitivities.

---

[10] Note, as described in paragraph 6.3, that 1.000 simulations in combination with the described control variate technique leads to a similar standard error as 1.000.000 simulation without the control variate technique. For this example $b = 1$ is used.

Table 3.1: comparison analytical / Monte Carlo approach, example 1

Total option value	Analytical	Monte Carlo	error	%
Base scenario	103.32	103.19	0.12	0.12%
Interest rates: + 1.5%	207.41	207.20	0.21	0.10%
Interest rates: - 1.5%	37.10	36.88	0.22	0.59%
Volatilities: + 0.15%	131.36	130.96	0.40	0.31%
Volatilities: - 0.15%	76.08	75.99	0.08	0.11%
Mean reversion: + 1.5%	93.16	93.03	0.13	0.14%
Mean reversion: -1.5%	116.26	116.00	0.26	0.22%
Strike: +1%	35.82	36.00	-0.18	-0.49%
Strike: -1%	238.18	237.73	0.46	0.19%

The table shows that the quality of the analytical approximation is excellent for all calculated scenarios. Note that the error as a percentage of the total value of the insurance liabilities would be around 0.01% in most cases.

The analytical approximation is potentially more exact than Monte Carlo simulation (without using a control variate algorithm), since the number of simulations used in practice is usually less than 1.000.000.

In table 3.2 a comparison between the analytical approximation and Monte Carlo simulation is given for different swap rate maturities and averaging periods. The table shows that the quality of the analytical approximation is also excellent for these product variants.

Table 3.2: comparison analytical / Monte Carlo – sensitivities*

Error of approximation	Averaging period		
Swap rate maturity	5	10	15
5	112,48 112,33 -0,13%	102,50 102,39 -0,11%	86,26 86,14 -0,14%
10	114,22 113,99 -0,20%	104,65 104,38 -0,27%	92,10 91,89 -0,23%
15	117,81 117,35 -0,39%	106,89 106,52 -0,35%	95,18 94,83 -0,38%

 * In each cell, top left: analytical price, top right: Monte Carlo, bottom: percentage error

3.7.2 Example 2: 10-year average of 7-year swap rate, compounding option

In this example the value of compounded profit sharing is calculated for a savings product with maturity 20. The compounding profit sharing is of form (3.20), where again the 10-year average of 7-year swap rates is used as the profit sharing rate. The assumed technical interest rate is 3.5%, $s(T_i)$ is assumed to be 1 and a margin of 0.5% is applied. The fund value at the start of the projection is 1.000.

The analytical approximation described in paragraph 3.6.1 is tested with Monte Carlo simulation, where 100.000 (antithetic) simulations are used. The results are given in table 3.3, where again the total value of the option is given for both approaches and for different yield curve, volatility, mean reversion and strike sensitivities. Also results are included for different maturities of the insurance product.

Table 3.3: comparison analytical / Monte Carlo approach, example 2

Total option value	Analytical	Monte Carlo	error	%
Base scenario	115,37	117,47	-2,10	-1,78%
Interest rates: + 1.5%	228,08	234,12	-6,05	-2,58%
Interest rates: - 1.5%	38,95	39,57	-0,62	-1,56%
Volatilities: + 0.15%	136,24	140,19	-3,96	-2,82%
Volatilities: - 0.15%	94,83	96,19	-1,36	-1,41%
Mean reversion: + 1.5%	109,01	110,83	-1,82	-1,65%
Mean reversion: -1.5%	122,81	125,80	-2,99	-2,38%
Strike: +1%	32,03	32,21	-0,17	-0,54%
Strike: -1%	276,14	282,00	-5,86	-2,08%
Maturity product: 15	80,69	80,78	-0,09	-0,11%
Maturity product: 25	147,38	154,10	-6,73	-4,37%

The table shows that the quality of the analytical approximation is reasonable for all calculated scenarios. Note that the error as a percentage of the initial fund value is less than 0.5% in most cases. The assumption of independent profit sharing rates over time introduces an additional error. However, considering the "errors" made earlier in the process (calibration of interest rate model, approximation with swap rate) and the quality of the assumptions usually made for non-economic parameters (mortality, lapses), the error could still be considered as being acceptable. The results for different maturities indicate that the quality of the approximation decreases when the maturity of the product exceeds 20 years.

In table 3.4 a comparison for different swap rate maturities and averaging periods is given. The table shows that the quality of the analytical approximation is increasing (decreasing) when the averaging period is longer (shorter).

Table 3.4: comparison analytical / Monte Carlo – sensitivities*

Error of approximation	Averaging period		
Swap rate maturity	5	10	15
5	123,80 128,20 -3,43%	113,52 115,91 -2,06%	126,43 126,12 0,25%
10	125,88 129,88 -3,08%	117,31 119,86 -2,13%	131,15 130,42 0,56%
15	124,55 127,68 -2,45%	117,59 119,50 -1,60%	132,52 131,98 0,41%

* In each cell, top left: analytical price, top right: Monte Carlo, bottom: percentage error

As mentioned in 3.6.1 the quality of the approximation (in terms of time value of the option) is less when the impact of the compounding is relatively high. However, since the total value of the option is higher in this case, the error will still be reasonable in terms of the total value of the option (as shown in the table above). This effect is also shown in figure 3.1, where the results of the analytical and the Monte Carlo approach are given for different yield curve sensitivities for example 2.

Figure 3.1: comparison analytical / Monte Carlo approach, example 2
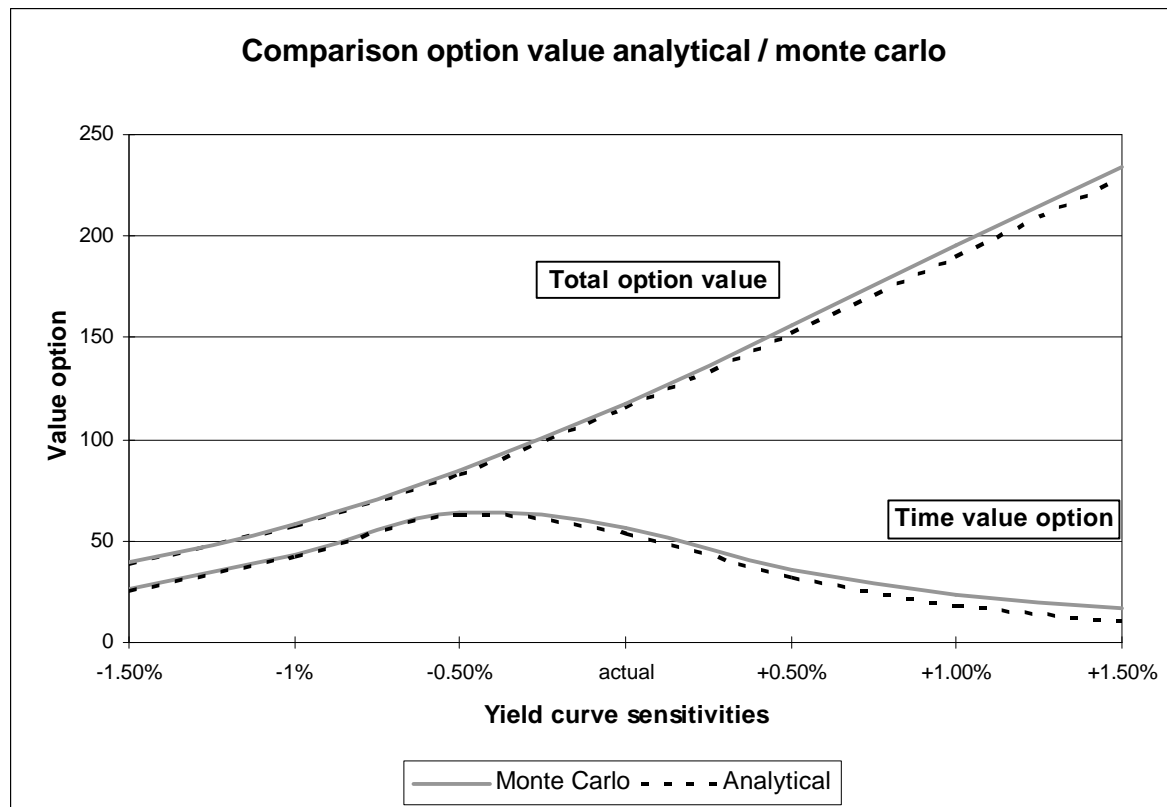
3.8 Conclusions

In this chapter analytical approximations are derived for prices of swap rate dependent embedded options in insurance products. In practice these options are often valued using Monte Carlo simulations. However, for risk management calculations and reporting processes, lots of valuations are needed and therefore a more efficient method to value these options would be helpful. The basis for the approximations is the result of Schrager and Pelsser (2006), who derived an approximate distribution for the forward swap rates under the relevant swap measure. After some changes of measure, this result is used to derive analytical approximations for swap rate dependent embedded options, given an underlying multi-factor Gaussian interest rate model.

The analytical approximation for options with direct payment is almost exact while the approximation for compounding options is also satisfactory. For similar options with additional management actions that significantly impact the option value, no analytical approximation is possible. However, using the analytical approximation for an option with direct payment as a control variate, the number of Monte Carlo simulations can be reduced significantly for these kinds of options. Furthermore, it's also possible to construct analytical approximations when returns on additional assets (such as equities) are part of the profit sharing rate.

Appendix 3a: proof of (3.8)

Each element of the vector of derivatives of (3.7) can be written as:

(3.27)
$$\frac{\partial y_{n,N}(t)}{\partial Y^{(i)}(t)} = -B^{(i)}(t,T_n)D^P(t,T_n) + B^{(i)}(t,T_N)D^P(t,T_N)$$

$$+ y_{n,N}(t)\sum_{k=n+1}^{N}\Delta^Y_{k-1}B^{(i)}(t,T_k)D^P(t,T_k)$$

where $D^P(t,T_n) = D(t,T_n)/P_{n+1,N}(t)$, the bond price normalized by the numéraire.

Note that since bond prices in this model are stochastic, the volatility of the swap rate is stochastic as well. The approximation of Schrager and Pelsser consists of replacing the stochastic terms $D^P(t,T_i)$ by their time zero values $D^P(0,T_i)$. This results in:

(3.28)
$$\frac{\partial y_{n,N}(t)}{\partial Y^{(i)}(t)} \approx -B^{(i)}(t,T_n)D^P(0,T_n) + B^{(i)}(t,T_N)D^P(0,T_N)$$

$$+ y_{n,N}(0)\sum_{k=n+1}^{N}\Delta^Y_{k-1}B^{(i)}(t,T_k)D^P(0,T_k) = \overline{\frac{\partial y_{n,N}(t)}{\partial Y^{(i)}(t)}}$$

This approximation makes the swap rate volatility deterministic and thus leads to a normally distributed forward swap rate. Furthermore, we can rewrite

(3.29) $\quad B^{(i)}(t,T) = \dfrac{1}{A_{(ii)}} - \dfrac{e^{-A_{(ii)}T}}{A_{(ii)}}e^{A_{(ii)}t}$

Using this, (3.28) can be split in a time dependent part and a constant part:

(3.30)
$$\overline{\frac{\partial y_{n,N}(t)}{\partial Y^{(i)}(t)}} = \frac{1}{A_{(ii)}}e^{A_{(ii)}t}\left[e^{-A_{(ii)}T_n}D^P(0,T_n) - e^{-A_{(ii)}T_N}D^P(0,T_N) \right.$$

$$\left. - y_{n,N}(0)\sum_{k=n+1}^{N}\Delta^Y_{k-1}e^{-A_{(ii)}T_k}D^P(0,T_k) \right] = e^{A_{(ii)}t}\,\tilde{C}^{(i)}_{n,N}$$

So in the approximate model, the swap rate at time T_n is given by:

(3.31)
$$\int_0^{T_n} dy_{n,N}(s) = \int_0^{T_n}\frac{\partial y_{n,N}(t)}{\partial Y(t)}\Sigma\, dW^{n+1,N}(t) \approx \int_0^{T_n}\overline{\frac{\partial y_{n,N}(t)}{\partial Y(t)}}\Sigma\, dW^{n+1,N}(t)$$

$$= \int_0^{T_n} e^{As'}\,diag(\tilde{C}_{n,N})\Sigma\, dW^{n+1,N}(t)$$

$$\text{where } e^{As} = \begin{bmatrix} e^{A_{(11)}s} \\ \vdots \\ e^{A_{(mm)}s} \end{bmatrix} \text{ and } diag(\tilde{C}_{n,N}) = \begin{bmatrix} \tilde{C}_{n,N}^{(1)} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \tilde{C}_{n,N}^{(m)} \end{bmatrix}$$

By using Ito's Isometry, this leads to an analytical expression for the integrated variance of $y_{n,N}$ (associated with a T_n x T_N swaption) over the interval $[0,T_n]$:

$$\mathbf{s}_{n,N}^2 \approx \int_0^{T_n} e^{As'} diag(\tilde{C}_{n,N}) \Sigma \Sigma' diag(\tilde{C}_{n,N}) e^{As} \, ds$$

(3.32)

$$= \sum_{i=1}^m \sum_{j=1}^m \hat{\Sigma}_{(ij)} \tilde{C}_{n,N}^{(i)} \tilde{C}_{n,N}^{(j)} \left[\frac{e^{[A_{(ii)}+A_{(jj)}]T_n} - 1}{A_{(ii)} + A_{(jj)}} \right]$$

Appendix 3b: proofs of (3.11) and (3.12)

Proof of (3.11)
A change of measure has to be done from the swap measure $Q^{n+1,N}$ to the T_n-forward measure Q^{T_n}. In this case the Radon-Nikodym derivative is:

(3.33) $\quad \dfrac{dQ^{T_n}}{dQ^{n+1,N}} = \mathbf{r}(t) = \dfrac{D(t,T_n)\big/D(0,T_n)}{\displaystyle\sum_{k=n+1}^N \Delta_{k-1}^Y D(t,T_k) \Big/ \displaystyle\sum_{k=n+1}^N \Delta_{k-1}^Y D(0,T_k)}$

Then using Ito's Lemma leads to:

(3.34) $\quad d\mathbf{r}(t) = \mathbf{k}(t)\,\mathbf{r}(t)\,dW^{T_n}$

where $\mathbf{k}(t)$ is an 1 x m vector with for each element $\mathbf{k}^{(i)}(t)$:

(3.35) $\quad \mathbf{k}^{(i)}(t) = -B^{(i)}(t,T_n) + \displaystyle\sum_{k=n+1}^N \Delta_{k-1}^Y B^{(i)}(t,T_k) D^P(t,T_k)$

Now like in appendix 3a replacing the stochastic terms $D^P(t,T_i)$ by their time zero values $D^P(0,T_i)$ and using (3.29) results in:

(3.36) $\quad \mathbf{k}^{(i)}(t) \approx \dfrac{1}{A_{(ii)}} e^{A_{(ii)}t} \left[e^{-A_{(ii)}T_n} - \displaystyle\sum_{k=n+1}^N \Delta_{k-1}^Y e^{-A_{(ii)}T_k} D^P(0,T_k) \right] = e^{A_{(ii)}t} \tilde{G}_{n,N}^{(i)}$

Using (3.30) and integrating $dy_{n,N}$ leads to the following formula for the convexity correction $CC_{n,N}(T_n)$ for time $T_n > 0$ for the swap rate $y_{n,N}$:

$$
CC_{n,N}(T_n) \approx \int_0^{T_n} e^{As'} \, diag(\tilde{C}_{n,N}) \, \Sigma\Sigma' \, diag(\tilde{G}_{n,N}) \, e^{As} \, ds
$$

(3.37)

$$
= \sum_{i=1}^{m}\sum_{j=1}^{m} \hat{\Sigma}_{(ij)} \, \tilde{C}_{n,N}^{(i)} \tilde{G}_{n,N}^{(j)} \left[\frac{e^{\left[A_{(ii)}+A_{(jj)}\right]T_n} - 1}{A_{(ii)} + A_{(jj)}} \right]
$$

$$
where \quad \tilde{G}_{n,N}^{(j)} = \frac{1}{A_{(jj)}} \left[e^{-A_{(jj)}T_n} - \sum_{k=n+1}^{N} \Delta_{k-1}^{Y} \, e^{-A_{(jj)}T_k} \, D^P(0,T_k) \right]
$$

Proof of (3.12)
In this case the Radon-Nikodym derivative is:

$$
\frac{dQ^{T_{n+u}}}{dQ^{T_n}} = r(t) = \frac{D(t,T_{n+u})/D(0,T_{n+u})}{D(t,T_n)/D(0,T_n)}
$$

(3.38)

$$
= \frac{D(0,T_n)}{D(0,T_{n+u})} \, \exp\left[A(t,T_{n+u}) - A(t,T_n) - \left(\sum_{i=1}^{m} B^{(i)}(t,T_{n+u}) - \sum_{i=1}^{m} B^{(i)}(t,T_n) \right) Y^{(i)}(t) \right]
$$

Then using the same procedure as above:

$$
(3.39) \quad k^{(i)}(t) = B^{(i)}(t,T_n) - B^{(i)}(t,T_{n+u}) = \frac{1}{A_{(ii)}} e^{A_{(ii)}t} \left[e^{-A_{(ii)}T_{n+u}} - e^{-A_{(ii)}T_n} \right] = e^{A_{(ii)}t} \tilde{H}_{n,n+u}^{(i)}
$$

Using (3.30) and integrating $dy_{n,N}$ leads the following formula for the timing correction $TC_{n,N}(T_n,T_{n+u})$ representing a change of measure from time $T_n > 0$ to T_{n+u}:

$$
TC_{n,N}(T_n,T_{n+u}) \approx \int_0^{T_n} e^{As'} \, diag(\tilde{C}_{n,N}) \, \Sigma\Sigma' \, diag(\tilde{H}_{T_n,T_{n+u}}) \, e^{As} \, ds
$$

(3.40)

$$
= \sum_{i=1}^{m}\sum_{j=1}^{m} \hat{\Sigma}_{(ij)} \, \tilde{C}_{n,N}^{(i)} \tilde{H}_{T_n,T_{n+u}}^{(j)} \left[\frac{e^{\left[A_{(ii)}+A_{(jj)}\right]T_n} - 1}{A_{(ii)} + A_{(jj)}} \right]
$$

$$
where \quad \tilde{H}_{T_n,T_{n+u}}^{(j)} = \frac{1}{A_{(jj)}} \left[e^{-A_{(jj)}(T_{n+u})} - e^{-A_{(jj)}T_n} \right]
$$

Appendix 3c: input example 1

In this appendix the data and assumptions are given that are used for example 1. The data used for the profit sharing basis *L(t)* and the technical interest rates *TR(t)* are based on an example portfolio of a long term pension insurance portfolio and are given in table 3.5.

Table 3.5: used data for profit sharing basis and technical interest rate

Time	TR(t)	L(t)	Time	TR(t)	L(t)
0	3.8%	1,000	25	3.6%	655
1	3.7%	1,043	26	3.6%	625
2	3.7%	1,066	27	3.5%	594
3	3.7%	1,060	28	3.5%	563
4	3.7%	1,054	29	3.5%	532
5	3.7%	1,046	30	3.5%	501
6	3.7%	1,038	31	3.5%	470
7	3.7%	1,028	32	3.5%	440
8	3.7%	1,016	33	3.5%	410
9	3.7%	1,004	34	3.5%	381
10	3.7%	991	35	3.5%	353
11	3.7%	976	36	3.5%	326
12	3.7%	961	37	3.5%	300
13	3.7%	944	38	3.5%	275
14	3.6%	926	39	3.5%	251
15	3.6%	907	40	3.4%	228
16	3.6%	887	41	3.4%	206
17	3.6%	865	42	3.4%	186
18	3.6%	842	43	3.4%	167
19	3.6%	819	44	3.4%	149
20	3.6%	794	45	3.4%	132
21	3.6%	768	46	3.4%	116
22	3.6%	741	47	3.4%	102
23	3.6%	713	48	3.4%	89
24	3.6%	684	49	3.4%	77

The swap curve used is from ultimo 2006 and the parameters of the 2 factor Gaussian interest rate model are calibrated to the swaption implied volatility surface at the same date. This information is given in table 3.6 (where *s* and *a* belong to factor 1 and *h* and *b* to factor 2).

Table 3.6: swap curve, implied volatility surface and parameters 2F Gaussian model

Swaption ATMF Volatility Surface											
Expiry/Tenor	1Y	2Y	3Y	4Y	5Y	7Y	10Y	15Y	20Y	25Y	30Y
1Y	14,3%	14,0%	14,4%	14,5%	14,5%	14,3%	14,0%	13,5%	13,0%	13,0%	12,9%
2Y	14,6%	14,9%	15,0%	14,9%	14,8%	14,6%	14,2%	13,6%	13,2%	13,0%	12,9%
3Y	15,1%	15,0%	15,1%	15,0%	14,8%	14,5%	14,1%	13,6%	13,2%	13,1%	12,9%
4Y	15,1%	15,0%	15,0%	14,8%	14,6%	14,2%	13,9%	13,4%	13,1%	13,0%	12,8%
5Y	14,8%	14,8%	14,7%	14,5%	14,3%	14,0%	13,6%	13,2%	12,9%	12,8%	12,6%
7Y	14,0%	14,0%	14,0%	13,9%	13,6%	13,4%	13,1%	12,9%	12,4%	12,4%	12,2%
10Y	13,0%	13,1%	13,1%	13,0%	12,8%	12,7%	12,5%	12,1%	11,8%	11,6%	11,4%
15Y	12,0%	12,0%	12,0%	12,0%	12,0%	12,0%	12,0%	11,6%	11,2%	11,0%	10,9%
20Y	11,6%	11,6%	11,6%	11,7%	11,8%	11,8%	11,8%	11,2%	10,8%	10,5%	10,5%
30Y	11,1%	11,1%	11,2%	11,3%	11,3%	11,3%	11,3%	10,7%	10,6%	10,7%	10,7%

Time	Swap Rate
1	4,08%
2	4,14%
3	4,12%
4	4,12%
5	4,13%
6	4,14%
7	4,15%
8	4,16%
9	4,18%
10	4,20%
15	4,28%
20	4,31%
30	4,29%
40	4,25%
50	4,20%

Parameters	
s	0,51%
a	2,75%
h	0,28%
b	2,75%
rho	0,497

Chapter 4

Valuation of Guaranteed Annuity Options using a Stochastic Volatility Model for Equity Prices*

4.1 Introduction

As mentioned in chapter 3, life insurers often include embedded options in the terms of their products. One of the most familiar embedded options is the Guaranteed Annuity Option (GAO). A GAO provides the right to convert a policyholder's accumulated funds to a life annuity at a fixed rate when the policy matures. These options were a common feature in retirement savings contracts issued in the 1970's and 1980's in the United Kingdom (UK). According to Bolton et al. (1997) the most popular guaranteed conversion rate was about 11%. Due to the high interest rates at that time, the GAOs were far out of the money. However, as the interest rate levels decreased in the 1990's and the (expected) mortality rates improved, the value of the GAOs increased rapidly and amongst others led to the downfall of Equitable Life in 2000. Currently, similar options are frequently sold under the name Guaranteed Minimum Income Benefit (GMIB) in the U.S. and Japan as part of variable annuity products. The markets for variable annuities in the U.S. and Japan have grown explosively over the past years, and a growth in Europe is also expected, see Wyman (2007).

The last decade the literature on pricing and risk management of these options evolved. Approaches for risk management and hedging of GAOs were described in Dunbar (1999), Yang (2001), Wilkie et al. (2003) and Pelsser (2003). The pricing of GAOs and GMIBs has been described by several authors, for example van Bezooyen et al. (1998), Boyle and Hardy (2001), Ballotta and Haberman (2003), Boyle and Hardy (2003), Biffis and Millossovich (2006), Chu

and Kwok (2007), Bauer et al. (2008) and Marshall et al. (2009). In most of these papers, the focus is on unit linked deferred annuity contracts purchased originally by a single premium. Generally a standard geometric Brownian motion is assumed for equity prices. However, Ballotta and Haberman (2003) and Chu and Kwok (2007) noted that, given the long maturities of the insurance contracts, a stochastic volatility model for equity prices would be more suitable.

In this chapter explicit expressions are derived for prices of GAOs, assuming stochastic volatility for equity prices and (of course) stochastic interest rates. The model used for this is the Schöbel-Zhu Hull-White (SZHW) model, introduced in van Haastrecht et al. (2009). The model combines the stochastic volatility model of Schöbel and Zhu (1999) with the 1-factor Gaussian interest rate model of Hull and White (1993), taking the correlation structure between those processes explicitly into account. Furthermore, this is extended to the case of a 2-factor Gaussian interest rate model.

The remainder of the chapter is organized as follows. First, in Section 4.2 the characteristics of the GAO are given. Section 4.3 describes the SZHW model to be used for the pricing of the GAO and section 4.4 discusses its calibration. In Section 4.5 explicit pricing formulas are derived for the GAOs given an underlying SZHW model. These results are extended to a 2-factor Hull-White model in Section 4.6. In Section 4.7 two numerical examples are worked out: the first shows the impact of stochastic volatility on the pricing of the GAO, whilst the second example deals with a comparison of the efficiency of our explicit formula for the 2-factor model with existing methods in the literature. Conclusions are given in Section 4.8.

4.2 Guaranteed Annuity Contract

A GAO gives the holder the right to receive at the retirement date T either a cash payment equal to the investment in the equity fund $S(T)$ or a life annuity of this investment against the guaranteed rate g. A rational policy holder would choose the greater of the two assets. In other words, if at inception, the policy holder is aged x and the normal retirement date is at time T, then the annuity value at maturity is $S(T) + H(T)$ with GAO payoff $H(T)$ equal to

$$(4.1) \quad H(T) \;=\; \left(gS(T)\sum_{i=0}^{n} c_i P(T,t_i) - S(T) \right)^{+}$$

provided that the policy holder is still alive at that time. Here $(x)^{+} = \max(x,0)$, g is the guaranteed rate, $P(T, t_i)$ the zero-coupon bond at time T maturing at t_i and c_i the insurance amounts for time i multiplied by the probability of survival from time T until time t_i for the policyholder. Without loss of generality, we will use unit insured amounts in the remainder of this chapter. Furthermore, we assume that the survival probabilities are independent of the equity prices and interest rates. Note that

$$(4.2) \quad H(T) \;=\; gS(T) \left(\sum_{i=0}^{n} c_i P(T,t_i) - K \right)^{+},$$

where $K = 1/g$. This last equality shows that one can interpret the GAO as a quanto call option with strike K on the zero-coupon bond portfolio $\sum_{i=0}^{n} c_i P(T, t_i)$ which is paid out using the exchange rate/currency $S(T)$, see Boyle and Hardy (2003). Under the risk-neutral measure Q, which uses the money market account $B(T)$,

$$(4.3) \quad B(T) \ = \ \exp\left(\int_0^T r(u)du\right)$$

as numéraire, the price of this option can be expressed as

$$(4.4) \quad C(T) \ = \ {}_{r-x}p_x \mathrm{E}^Q\left[\exp\left(-\int_0^T r(u)du\right)gS(T)\left(\sum_{i=0}^{n} c_i P(T, t_i) - K\right)^+\right]$$

where ${}_{r-x}p_x$ denotes the probability that the policy holder aged x survives r - x years, i.e. until the retirement age r at time T. To derive an explicit expression for the GAO of (4.4), it is more convenient to measure payments in terms of units of stock instead of money market values. Mathematically, we can establish this by using the equity price $S(T)$ as numéraire and changing from the risk-neutral measure to the equity-price measure Q^S, see Geman et al. (1995). Under the equity-price measure Q^S, the GAO price is then given by

$$(4.5) \quad C(T) \ = \ {}_{r-x}p_x gS(0)\mathrm{E}^{Q^S}\left[\left(\sum_{i=0}^{n} c_i P(T, t_i) - K\right)^+\right].$$

To evaluate this expectation we need to take into account the dynamics of the zero-coupon bonds prices $P(T, t_i)$ under the equity price measure.

Apart from the guaranteed rate, the drivers of the GAO price are the interest rates, the equity prices, the correlation between those, and the survival probabilities. The combined model for interest rates and equity prices is explained in Section 4.3. This model needs an assumption for the correlation, which could be derived from historical data. Note that if it is assumed that equity prices and interest rates are independent, it does not matter which model is assumed for equity prices. Both from historical data as well from market quotes, one however rarely finds that the equity prices and interest rates behave in an independent fashion. As this dependency structure is one of the main driver for the GAO price and its sensitivities, a non-trivial structure therefore has to be taken into account for a proper pricing and risk management of these derivatives, see Boyle and Hardy (2003), Ballotta and Haberman (2003) or Bauer (2009).

4.3 The Schöbel-Zhu-Hull-White model

The model used in this chapter is the Schöbel-Zhu Hull-White (SZHW) model, introduced in van Haastrecht et al. (2009). The model combines the stochastic volatility model of Schöbel and Zhu

(1999) with the 1-factor Gaussian interest rate model of Hull and White (1993), taking explicitly into account the correlation between these processes. In the SZHW model, the process for equity price $S(t)$ under the risk-neutral measure is the Schöbel and Zhu (1999) model described in paragraph 2.1.4.:

(4.6) $\dfrac{dS(t)}{S(t)} = r(t)dt + v(t)dW_S^Q(t)$ $S(0) = S_0$

(4.7) $dv(t) = \boldsymbol{k}\big(\boldsymbol{y} - v(t)\big)dt + \boldsymbol{t}\,dW_v^Q(t)$ $v(0) = v_0$

Here $v(t)$, which follows an Ornstein-Uhlenbeck process, is the (instantaneous) stochastic volatility of the equity $S(t)$. The parameters of the volatility process are the positive constants *?* (mean reversion), v_0 (short-term mean), *?* (long-term mean) and *t* (volatility of the volatility). We assume the interest rates are given by a one-factor Hull and White (1993) process, whose dynamics under Q can be parameterized by

(4.8) $r(t) = \boldsymbol{a}(t) + x(t)$ $r(0) = r_0$

(4.9) $dx(t) = -ax(t)dt + \boldsymbol{s}\,dW_x^Q(t)$ $x(0) = 0$

Here *a* (mean reversion) and $\boldsymbol{s}$ (volatility) are the positive parameters of the model. The function $\boldsymbol{a}(t)$ can be used to fit the current term structure of interest rates exactly, see Pelsser (2000) or Brigo and Mercurio (2006). Under the above dynamics for the equity, volatility and interest rates explicit expressions for the prices of European equity options exist, see van Haastrecht et al. (2009). Moreover the model allows for a general correlation structure, i.e.

(4.10) $dW_v^Q(t)dW_S^Q(t) = \boldsymbol{r}_{vS}dt, \quad dW_x^Q(t)dW_S^Q(t) = \boldsymbol{r}_{xS}dt, \quad dW_x^Q(t)dW_v^Q(t) = \boldsymbol{r}_{xv}dt$

where $\boldsymbol{r}_{vS}$, $\boldsymbol{r}_{xS}$ and $\boldsymbol{r}_{xv}$ are the instantaneous correlation parameters between the Brownian motions of the equity price, the stochastic volatility and the interest rate. Having the flexibility to correlate the equity price with both stochastic volatility and stochastic interest rates yields a realistic model, which is of practical importance for the pricing and hedging of options with long-term exposures such as guaranteed annuities, see Boyle and Hardy (2003).

It is hardly necessary to motivate the inclusion of stochastic volatility in a long-term derivative pricing model. First, compared to constant volatility models, stochastic volatility models are significantly better able to fit the market's option data, see Andreasen (2006) or Andersen and Brotherton-Ratcliffe (2001). Second, as stochastic interest rates and stochastic volatility are empirical phenomena, the addition of these factors yields a more realistic model, which becomes important for the pricing and especially the hedging of long-term derivatives. The addition of stochastic volatility and stochastic interest rates as stochastic factors is important when considering long-maturity equity derivatives and has been the subject of empirical investigations most notably by Bakshi et al. (2000). These authors show that the hedging performance of delta hedging strategies of long-maturity options improves when stochastic volatility and stochastic interest rates are taken into account.

Stochastic volatility models have been described by several others, for example Stein and Stein (1991), Heston (1993), Schöbel and Zhu (1999), Duffie et al. (2000), Duffie et al. (2003), van der Ploeg (2006) and van Haastrecht et al. (2009). Also regime-switching models are suggested in the literature for the pricing of equity-linked insurance policies, see Hardy (2001) and Brigo and Mercurio (2006). In the limit of an infinite number of regimes these models again converge to a continuous-time stochastic volatility model, however in discrete time they can benefit from a greater analytical tractability. A proper model assessment greatly depends on the properties of the embedded options in the insurance contract.

To investigate the impact of using a stochastic volatility model on the pricing of GAOs, note that the GAO directly depends on the stochastic interest rates, the underlying equity fund and the correlation between the rates and the equity. For the pricing of GAOs we therefore choose to use the SZHW model over other stochastic volatility models, as this model distinguishes itself models by an explicit incorporation of the correlation between the underlying equity fund and the term structure of interest rates, whilst maintaining a high degree of analytical tractability.

In Section 4.7 the impact of stochastic volatility on the pricing of GAOs is investigated. That is, we compare the pricing of GAOs in the SZHW stochastic volatility model with the Black-Scholes Hull-White (BSHW) constant volatility model. The BSHW process for equity prices $S(t)$ under the risk neutral measure Q is:

$$(4.11) \quad \frac{dS(t)}{S(t)} \;=\; r(t)dt + \boldsymbol{s}_S dW_S^Q(t) \qquad S(0) = S_0$$

where the interest rate process $r(t)$ follows Hull and White (1993) dynamics as in (4.8) and with the instantaneous correlation between Brownian motions of the interest rate and the equity price equal to

$$(4.12) \quad dW_S^Q(t)dW_x^Q(t) = \boldsymbol{r}_{xS}dt$$

In the following section both the SZHW and BSHW model are calibrated to market data.

4.4 Calibration of the SZHW and BSHW model

To come up with a fair analysis of the impact of stochastic volatility on the pricing of GAOs, we first calibrate the BSHW and SZHW model to the same market's option data per end July 2007. This is done by first calibrating the interest rate parameters, than estimating the effective correlation between the interest rates and equity, and finally we specify the equity components of the BSHW/SZHW model. We detail the calibration approach below.

Interest Rates First we calibrate the Hull and White (1993) interest rate models to EU and U.S. swaption markets. The option prices and corresponding swap curves are obtained from Bloomberg. Here a total of 151 swaption prices, which are contributed by various issuers and

maintained by Bloomberg, can be found for different tenors and maturities ranging from 1 to 30 years. For the calibration of the interest rate model we used close (mid) swaption prices per 31st of July 2007. We calibrate the Hull and White (1993) models to these prices by minimizing the sum of the squared differences between the model's and the market's swaption implied volatilities. For the U.S. market, the mean average price error is 1.88% and for the EU market 1.34% which is very good given the large set of option prices that is fitted using only 2 interest rate parameters.

Terminal Correlation   After calibrating the interest rate component, we need to calibrate the equity and correlation parameters. For the equity component of the GAO we assume a large stock index, for which the Eurostoxx50 index (EU) and the S&P500 (U.S.) are used. The Eurostoxx50 consists of 50 large European companies and is traded on the Dow-Jones exchange, whilst the S&P500 is maintained by Standard & Poors and consists of NASDAQ and NYSE denoted shares. The effective 10 years correlation between the log equity returns and the interest rates is determined by time series analysis of the 10-year swap rate and the log returns of the EuroStoxx50 (EU) and S&P500 (U.S.) index over the period from February 2002 to July 2007. For the EU and the U.S. this resulted in correlation coefficients of 34.65% and 14.64% between the interest rates and the log equity returns.

It is well known that it is hard to calibrate the correlation coefficient. Furthermore large bid-ask spreads have to be paid to hedge this risk, which shows that the markets for correlation risks are unfortunately not very liquid. As a result, additional capital needs to be reserved in order to protect against this unhedgeable risk.

Equity   Using the interest rate and correlation parameters determined in the previous steps, the equity specific parameters are calibrated to option prices on the EuroStoxx50 (EU) and S&P500 (U.S.) index. These option prices are obtained from the implied volatility service of MarkIT, a financial data provider, which provides (mid) implied volatility quotes by averaging quotes from a large number of issuers. For large indices a total of 94 liquid quotes are available for 10 maturities ranging from 1 month up to 15 years, and 10 strikes ranging from 60% to 200%.

To aid a fair comparison between the models, the SZHW model is calibrated in such a way that the effective correlation between interest rates and equity prices is equal to that of the BSHW process. Finally, as the considered GAO in Section 4.7 only depends on terminal asset price distribution after 10 years, we have calibrated the equity model to market option prices maturing in 10 years time. This estimation is performed by minimizing the sum of absolute differences between market's and model's implied volatilities. The calibration results to the Eurostoxx50 and S&P500 are shown in table 4.1 below.

Table 4.1: calibration results for the SZHW and BSHW models, for EU and U.S.

EU				US			
strike	**Market**	**SZHW**	**BSHW**	**strike**	**Market**	**SZHW**	**BSHW**
80	27,8%	27,9%	26,4%	**80**	27,5%	27,5%	25,8%
90	27,1%	27,1%	26,4%	**90**	26,6%	26,6%	25,8%
95	26,7%	26,7%	26,4%	**95**	26,2%	26,2%	25,8%
100	26,4%	26,4%	26,4%	**100**	25,8%	25,8%	25,8%
105	26,0%	26,0%	26,4%	**105**	25,4%	25,4%	25,8%
110	25,7%	25,7%	26,4%	**110**	25,0%	25,0%	25,8%
120	25,1%	25,1%	26,4%	**120**	24,3%	24,4%	25,8%

The tables show that SZHW is significantly better in capturing the market's implied volatility structure and provides an extremely good fit. The fit of the BSHW model is relatively poor. Furthermore, a direct consequence of the log-normal distribution of the BSHW model, it that the asset returns have thin tails, which does not correspond to historical data nor to the market's view on long-term asset returns. The SZHW model provides a more realistic picture on the market's view on long-term asset returns as it can incorporate heavy-tailed returns. The latter can be made especially clear by looking at the risk-neutral densities of the log-asset price of the SZHW and BSHW model. These are plotted in figure 4.1 below for the BSHW and SZHW model, calibrated to EU option prices.

Figure 4.1: risk-neutral density of the log-asset price for the SZHW and BSHW model
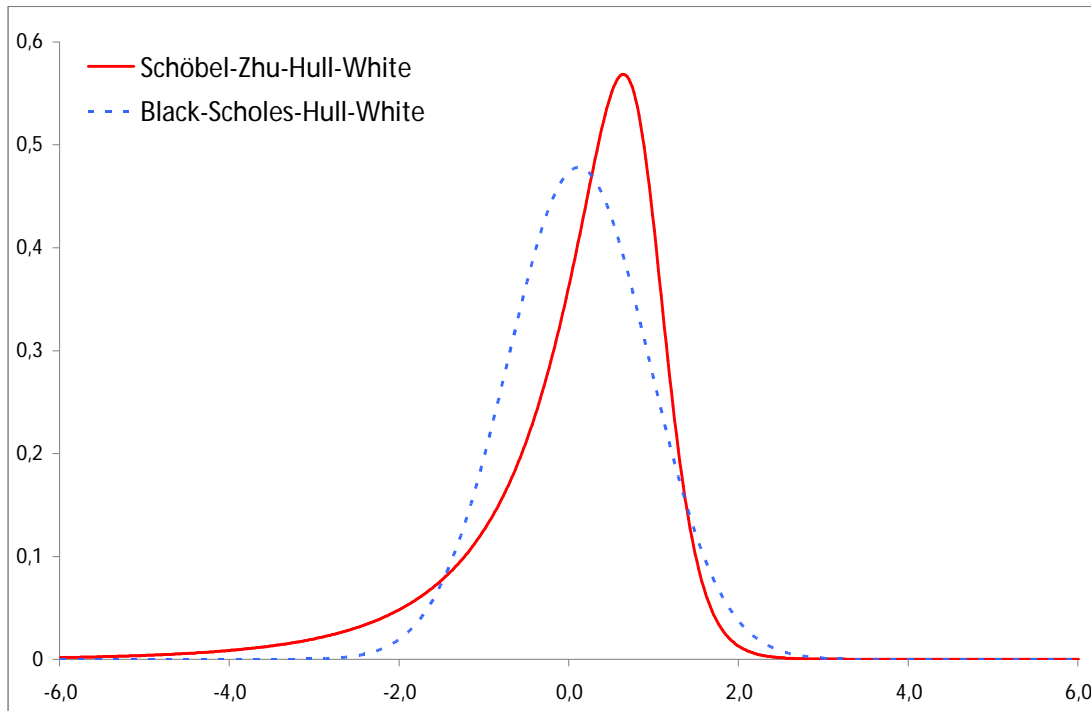
Clearly, the SZHW model incorporates the skewness and heavy-tails seen in option markets (see Bakshi et al. (1997)) a lot more realistically than the BSHW model. The effects of these log-asset price distributions on the pricing of GAOs, combined with correlated interest rates, are analyzed in Section 4.7.

4.5 Pricing the Guaranteed Annuity Option under stochastic volatility and stochastic interest rates

For the pricing of the GAO in the SZHW model, i.e. the evaluation of (4.5), we need to consider the pricing of zero-coupon bonds in the Gaussian interest rate model. In the Hull and White (1993) model, one has the following expression for the time-T price of a zero-coupon bond $P(T,t_i)$ maturing at time t_i:

$$(4.13) \quad P(T,t_i) \;=\; A(T,t_i)e^{-B(T,t_i)x(T)}$$

where

$$(4.14) \quad A(T,t_i) \;=\; \frac{P^M(0,t_i)}{P^M(0,T)}e^{\left[\frac{1}{2}(V(T,t_i)-V(0,t_i)+V(0,T))\right]}$$

$$(4.15) \quad B(T,t_i) \;=\; \frac{1-e^{-a(t_i-T)}}{a}$$

$$(4.16) \quad V(T,t_i) \;=\; \frac{s^2}{a^2}\left((t_i-T)+\frac{2}{a}e^{-a(t_i-T)}-\frac{1}{2a}e^{-2a(t_i-T)}-\frac{3}{2a}\right)$$

and with $P^M(0, s)$ denoting the market's time zero discount factor maturing at time s. Using (4.13), we have for the GAO price (4.5) under the equity price measure Q^S:

$$(4.17) \quad C(T) \;=\; {}_{r-x}p_x\, gS(0)\mathrm{E}^{Q^S}\left[\left(\sum_{i=0}^{n}c_i A(T,t_i)e^{-B(T,t_i)x(T)}-K\right)^{+}\right]$$

To further evaluate this expression, we first have to consider the dynamics of $x(T)$ under the equity price measure Q^S in the SZHW model.

4.5.1 Taking the equity price as numéraire

To change the money market account numéraire into the equity price numéraire, we need to calculate the corresponding Radon-Nikodym derivative (see Geman et al. (1995)), which is given by

$$(4.18) \quad \frac{dQ^S}{dQ} \;=\; \frac{S(T)B(0)}{S(0)B(T)} \;=\; \exp\left[-\frac{1}{2}\int_0^T v^2(u)du+\int_0^T v(u)dW_S^Q(u)\right]$$

The multi-dimensional version of Girsanov's theorem (see Oksendal (2005)) hence implies that

$$(4.19) \quad dW_S^{Q^S}(t) \;\mapsto\; dW_S^Q(t)-v(t)dt$$

$$(4.20) \quad dW_x^{Q^S}(t) \;\mapsto\; dW_x^{Q}(t) - \boldsymbol{r}_{xS}\boldsymbol{v}(t)dt$$

$$(4.21) \quad dW_v^{Q^S}(t) \;\mapsto\; dW_v^{Q}(t) - \boldsymbol{r}_{vS}\boldsymbol{v}(t)dt$$

are Brownian motions under Q^S. Hence under Q^S one has the following model dynamics for the volatility and interest rate process

$$(4.22) \quad dx(t) \;=\; -ax(t)dt + \boldsymbol{r}_{xS}\boldsymbol{s}\,v(t)dt + \boldsymbol{s}\,dW_x^{Q^S}(t) \qquad x(0)=0$$

$$(4.23) \quad \begin{aligned} dv(t) &= \boldsymbol{k}\big(\mathbf{y}-v(t)\big)dt + \boldsymbol{r}_{vS}\boldsymbol{t}v(t)dt + \boldsymbol{t}\,dW_v^{Q^S}(t) \\ &= \boldsymbol{\breve{k}}\big(\boldsymbol{\tilde{y}}-v(t)\big)dt + \boldsymbol{t}\,dW_v^{Q^S}(t) \qquad\qquad v(0)=v_0 \end{aligned}$$

where $\boldsymbol{\breve{k}}=\boldsymbol{k}-\boldsymbol{r}_{vS}\boldsymbol{t}$ and $\boldsymbol{\tilde{y}}=\dfrac{\boldsymbol{k}\mathbf{y}}{\boldsymbol{\breve{k}}}$. After some calculations one can show that:

$$(4.24) \quad v(T) \;=\; \boldsymbol{\tilde{y}} + \big(v(0)-\boldsymbol{\tilde{y}}\big)e^{-\boldsymbol{\breve{k}}T} + \boldsymbol{t}\int_0^T e^{-\boldsymbol{\breve{k}}(T-u)}dW_v^{Q^S}(u)$$

$$(4.25) \quad \begin{aligned} x(T) \;=\;& \boldsymbol{r}_{xS}\boldsymbol{s}\left(\frac{\boldsymbol{\tilde{y}}}{a}\big[1-e^{-aT}\big]+\frac{v(0)-\boldsymbol{\tilde{y}}}{(a-\boldsymbol{\breve{k}})}\big[e^{-\boldsymbol{\breve{k}}T}-e^{-aT}\big]\right) \\ &+\frac{\boldsymbol{r}_{xS}\boldsymbol{s}\boldsymbol{t}}{(a-\boldsymbol{\breve{k}})}\int_0^T\big[e^{-\boldsymbol{\breve{k}}(T-u)}-e^{-a(T-u)}\big]dW_v^{Q^S}(u)+\boldsymbol{s}\int_0^T e^{-a(T-u)}dW_x^{Q^S}(u) \end{aligned}$$

Using Ito's isometry and Fubini's theorem, we have that $x(T)$ is normally distributed with mean μ_x and variance $\boldsymbol{s}_x^2$ given by

$$(4.26) \quad \boldsymbol{m}_x \;=\; \boldsymbol{r}_{xS}\boldsymbol{s}\left(\frac{\boldsymbol{\tilde{y}}}{a}\big[1-e^{-aT}\big]+\frac{v(0)-\boldsymbol{\tilde{y}}}{(a-\boldsymbol{\breve{k}})}\big[e^{-\boldsymbol{\breve{k}}T}-e^{-aT}\big]\right)$$

$$(4.27) \quad \boldsymbol{s}_x^2 \;=\; \boldsymbol{s}_1^2 + \boldsymbol{s}_2^2 + 2\boldsymbol{r}_{12}\boldsymbol{s}_1\boldsymbol{s}_2$$

where

$$(4.28) \quad \boldsymbol{s}_1 \;=\; \boldsymbol{s}\sqrt{\frac{1-e^{-2aT}}{2a}}$$

$$(4.29) \quad \boldsymbol{s}_2 \;=\; \frac{\boldsymbol{r}_{xS}\boldsymbol{s}\boldsymbol{t}}{a-\boldsymbol{\breve{k}}}\sqrt{\frac{1}{2\boldsymbol{\breve{k}}}+\frac{1}{2a}-\frac{2}{(\boldsymbol{\breve{k}}+a)}-\frac{e^{-2\boldsymbol{\breve{k}}T}}{2\boldsymbol{\breve{k}}}-\frac{e^{-2aT}}{2a}+\frac{2e^{-(\boldsymbol{\breve{k}}+a)T}}{(\boldsymbol{\breve{k}}+a)}}$$

$$(4.30) \quad r_{12} \quad = \quad r_{xv} \frac{s^2 r_{xS} t}{s_1 s_2 (a - \bar{k})} \left(\frac{1 - e^{-(a+\bar{k})T}}{(a + \bar{k})} - \frac{1 - e^{-2aT}}{2a} \right)$$

4.5.2 Explicit formula for the GAO price

Using the results from the previous paragraph, we can now further evaluate the expression (4.17) for the GAO price in the SZHW model: as the zero-coupon bond price is a monotone function of one state variable, $x(T)$, one can use the Jamshidian (1989) result and write the call option (4.17) on the sum of zero-coupon bonds as a sum of zero-coupon bond call options: let x^* solve

$$(4.31) \quad \sum_{i=0}^{n} c_i A(T,t_i) e^{-B(T,t_i)x^*} \quad = \quad K$$

and let

$$(4.32) \quad K_i \quad = \quad A(T,t_i) e^{-B(T,t_i)x^*}$$

Using Jamshidian (1989), we have that the price of GAO is equal to the price of a sum of zero-coupon bond options, i.e.

$$(4.33) \quad C(T) \quad = \quad _{r-x} p_x g S(0) \mathrm{E}^{Q^S} \left[\sum_{i=0}^{n} c_i \left(A(T,t_i) e^{-B(T,t_i)x(T)} - K_i \right)^+ \right]$$

As $x(T)$ is normally distributed, we have that $P(T, t_i) = A(T,t_i)e^{-B(T,ti)x(T)}$ is log-normally distributed. Provided that we know the mean M_i and variance V_i of $\ln P(T, t_i)$ under Q^S, one can directly express the above expectation in terms of the Black and Scholes (1973) formula:

$$(4.34) \quad C(T) \quad = \quad _{r-x} p_x g S(0) \sum_{i=0}^{n} c_i \left(F_i N(d_1^i) - K_i N(d_2^i) \right)$$

$$(4.35) \quad F_i \quad = \quad e^{M_i + \frac{1}{2} V_i}$$

$$(4.36) \quad d_1^i \quad = \quad \frac{\ln(F_i / K_i) + \frac{1}{2} V_i}{\sqrt{V_i}}$$

$$(4.37) \quad d_2^i \quad = \quad d_1^i - \sqrt{V_i}$$

where N denotes the cumulative standard normal distribution function. To determine M_i and V_i, recall from (4.26) and (4.27) that $x(T)$ is normally distributed with mean μ_x and variance s_x^2. Hence with $P(T, t_i) = A(T, t_i)e^{-B(T,ti)x(T)}$, one can directly obtain that the mean M_i and variance V_i of $\ln P(T, t_i)$ are given by

$$(4.38) \quad M_i \quad = \quad \ln A(T,t_i) - B(T,t_i)\boldsymbol{m}_x$$

$$(4.39) \quad V_i \quad = \quad B^2(T,t_i)\boldsymbol{s}_x^2$$

Hence under the SZHW dynamics (4.6)-(4.9), we have derived the explicit formula (4.34) for the price of a GAO under stochastic volatility and correlated stochastic interest rates. With this result, we are able to investigate the impact of stochastic volatility on the pricing of GAOs, which will be the subject of paragraph 4.7.1.

4.6 Extension to two-factor interest rate model

A one-factor assumption for the short interest rate unfortunately means that all future interest rates are driven by one factor. As reported in Brigo and Mercurio (2006), principal components analysis shows that the full interest rate curve is (depending on the currency) typically driven by two or more factors. When calibrating to European swaption prices, it is demonstrated that a two-factor Gaussian model gives significantly better fits and produces more realistic future interest rate curves. Furthermore, as noted in Chu and Kwok (2007), the one-factor assumption typically leads to a full correlation of all future interest rates. In particular these authors recommend using a two-factor interest rate model for the pricing of long-term derivatives and GAO contracts in particular. In this section, we therefore generalize the setting of the previous section from one to two-factor Gaussian interest rates. That is under the risk-neutral measure Q, we assume the following dynamics for the short interest rate process:

$$(4.40) \quad r(t) \quad = \quad \boldsymbol{j}(t) + x(t) + y(t) \qquad\qquad r(0) = r_0$$

$$(4.41) \quad dx(t) \quad = \quad -ax(t)dt + \boldsymbol{s}\,dW_x^Q(t) \qquad x(0) = 0$$

$$(4.42) \quad dy(t) \quad = \quad -by(t)dt + \boldsymbol{h}\,dW_y^Q(t) \qquad y(0) = 0$$

$$(4.43) \quad dW_x^Q(t)dW_y^Q(t) = \boldsymbol{r}_{xy}dt$$

Here a, b (mean reversion) and $\boldsymbol{s}$, $\boldsymbol{h}$ (volatility) are the positive parameters of the model and $|\boldsymbol{r}_{xy}|$ < 1. The deterministic function $\boldsymbol{j}(t)$ can be used to exactly fit the current term structure of interest rates, see Brigo and Mercurio (2006) who name this model the 'G2++' model. Much of the analytical structure of the one-factor Gaussian is preserved in this two-factor setting. For example prices of time T zero-coupon bonds maturing at time t_i are given by

$$(4.44) \quad P(T,t_i) \quad = \quad A(T,t_i)e^{-B(a,T,t_i)x(T) - B(b,T,t_i)y(T)}$$

where

$$(4.45) \quad A(T,t_i) \;=\; \frac{P^M(0,t_i)}{P^M(0,T)}\, e^{\left[\frac{1}{2}\left(V(T,t_i)-V(0,t_i)+V(0,T)\right)\right]}$$

$$(4.46) \quad B(z,T,t_i) \;=\; \frac{1-e^{-z(t_i-T)}}{z}$$

$$(4.47) \quad V(T,t_i) \;=\; \frac{\boldsymbol{s}^2}{a^2}\left((t_i-T)+\frac{2}{a}e^{-a(t_i-T)}-\frac{1}{2a}e^{-2a(t_i-T)}-\frac{3}{2a}\right)$$
$$+\frac{\boldsymbol{h}^2}{b^2}\left((t_i-T)+\frac{2}{b}e^{-b(t_i-T)}-\frac{1}{2b}e^{-2b(t_i-T)}-\frac{3}{2b}\right)$$
$$+2\boldsymbol{r}_{xy}\frac{\boldsymbol{sh}}{ab}\left((t_i-T)+\frac{e^{-a(t_i-T)}-1}{a}+\frac{e^{-b(t_i-T)}-1}{b}-\frac{e^{-(a+b)(t_i-T)}-1}{a+b}\right)$$

Substituting the zero-coupon bond expression (4.44) into the pricing equation (4.5) and evaluating this expectation, results in the following explicit expression for the GAO price:

$$(4.48) \quad C(T) \;=\; _{r-x}P_x\,gS(0)\int_{-\infty}^{\infty}\frac{e^{-\frac{1}{2}\left(\frac{x-\boldsymbol{m}_x}{\boldsymbol{s}_x}\right)^2}}{\boldsymbol{s}_x\sqrt{2\boldsymbol{p}}}\Big[F_i(x)N\big(h_2(x)\big)-KN\big(h_1(x)\big)\Big]dx$$

where

$$(4.49) \quad h_1(x) \;=\; \frac{y^*-\boldsymbol{m}_y}{\boldsymbol{s}_y\sqrt{1-\boldsymbol{r}_{xy}^2}}-\frac{\boldsymbol{r}_{xy}(x-\boldsymbol{m}_x)}{\boldsymbol{s}_x\sqrt{1-\boldsymbol{r}_{xy}^2}}$$

$$(4.50) \quad h_2(x) \;=\; h_1(x)+B(b,T,t_i)\boldsymbol{s}_y\sqrt{1-\boldsymbol{r}_{xy}^2}$$

$$(4.51) \quad l_i(x) \;=\; c_i A(T,t_i)e^{-B(a,T,t_i)x}$$

$$(4.52) \quad k_i(x) \;=\; -B(b,T,t_i)\left[\boldsymbol{m}_y-\frac{1}{2}\boldsymbol{s}_y^2\left(1-\boldsymbol{r}_{xy}^2\right)B(b,T,t_i)+\boldsymbol{r}_{xy}\boldsymbol{s}_y\frac{(x-\boldsymbol{m}_x)}{\boldsymbol{s}_x}\right]$$

$$(4.53) \quad F_i(x) \;=\; \sum_{i=0}^{n}l_i(x)e^{k_i(x)}$$

and where y^* is the unique solution of

$$(4.54) \quad \sum_{i=0}^{n}l_i(x)e^{-B(b,T,t_i)y^*} \;=\; K$$

The proof of (4.48) is given in appendix 4a.

In the pricing formula (4.48) it remains to determine the first two moments of $x(T)$ and $y(T)$ and the (terminal) correlation between $x(T)$ and $y(T)$, under the equity price measure Q^S. These are given in appendix 4b. Note that in the pricing formula (4.48), one is integrating a Gaussian probability density function against a bounded function. Because the Gaussian density functions decays very rapidly[11], one can therefore truncate the integration domain in an implementation of (4.48) to a suitable number of standard deviations s_x around the mean μ_x.

4.7 Numerical examples

In this section two numerical examples are given. In paragraph 4.7.1 the values of the GAO using the stochastic volatility model described in Section 4.3 are compared with values that result when a geometric Brownian motion is assumed for equity prices. Paragraph 4.7.2 deals with sensitivity analyses of different risk drivers. In paragraph 4.7.3 our approach for two-factor interest rate models is compared with the methods described in Chu and Kwok (2007).

4.7.1 Comparison results SZHW model and Black-Scholes Hull-White model

In this section the impact of stochastic volatility of equity prices is shown for an example policy. The results for the SZHW model given in (4.6)-(4.9) are compared with a model that combines a Black-Scholes process for equity prices with a one-factor Hull White model for interest rates, the so-called Black-Scholes-Hull-White (BSHW) model given in (4.11)-(4.12). The SZHW and BSHW models are both calibrated to market information (implied volatilities and interest rates) per end July 2007, see Section 4.4.

In the example, the policyholder is 55 years old, the retirement age is 65, giving the maturity T of the GAO option of 10 years. Furthermore, $S(0)$ is assumed to be 100. The survival rates are based on the PNMA00 table of the Continuous Mortality Investigation (CMI) for male pensioners[12].

In table 4.2 the prices for the GAO are given for different guaranteed rates g for both models. The results for the SZHW model are obtained using the explicit expressions given in (4.33) - (4.39). The pricing formula for the BSHW is a special case of this, and is also derived in Ballotta and Haberman (2003). The results are determined for EU data and U.S. data with an equity-interest rate correlation of respectively 0.347 and 0.146 (see Section 4.4). The table presents the total value of the GAO as well as the time value. The time value is determined as the difference between the total value and the (forward) intrinsic value. The latter is determined by setting all volatilities to zero. While the total value gives the impact on the total prices, the time value gives more insight in the relative impact of the models (since those only have impact on the time value). Also, the time value of the GAO is often reported separately, for example within Embedded Value reporting of insurers. The at-the-money guaranteed rate g is 8,44% for the U.S. and 8,23% for the EU.

---

[11] For instance, 99,9999% of the probability mass of a Gaussian density function lies within five standard deviations around its mean.

[12] This table is available at: http://www.actuaries.org.uk/knowledge/cmi/cmi_tables/00_ series_tables

Table 4.2: comparison of GAO values for the SZHW and BSHW model

EU, Total Value				EU, Time Value			
g	**SZHW**	**BSHW**	**Rel. Diff**	**g**	**SZHW**	**BSHW**	**Rel. Diff**
8,23%	3,82	3,07	+ 24,5%	**8,23%**	3,82	3,07	+ 24,5%
7%	0,59	0,39	+ 50,7%	**7%**	0,59	0,39	+ 50,7%
8%	2,89	2,26	+ 28,0%	**8%**	2,89	2,26	+ 28,0%
9%	8,40	7,25	+ 15,8%	**9%**	2,43	1,29	+ 88,9%
10%	17,02	15,53	+ 9,6%	**10%**	-0,11	-1,60	-93,0%
11%	27,37	25,69	+ 6,5%	**11%**	-0,93	-2,60	-64,4%
12%	38,30	36,47	+ 5,0%	**12%**	-1,17	-2,99	-61,0%
13%	49,35	47,37	+ 4,2%	**13%**	-1,28	-3,26	-60,7%

U.S., Total Value				U.S., Time Value			
g	**SZHW**	**BSHW**	**Rel. Diff**	**g**	**SZHW**	**BSHW**	**Rel. Diff**
8,44%	5,43	4,84	+ 12,0%	**8,44%**	5,43	4,84	+ 12,0%
7%	1,04	0,88	+ 18,0%	**7%**	1,04	0,88	+ 18,0%
8%	3,54	3,11	+ 13,6%	**8%**	3,54	3,11	+ 13,6%
9%	8,53	7,74	+ 10,3%	**9%**	7,27	6,47	+ 12,3%
10%	16,06	14,90	+ 7,8%	**10%**	4,15	2,99	+ 39,1%
11%	25,42	23,96	+ 6,1%	**11%**	2,86	1,40	+ 104,2%
12%	35,73	34,06	+ 4,9%	**12%**	2,53	0,86	+ 195,5%
13%	46,43	44,58	+ 4,1%	**13%**	2,58	0,74	+ 250,1%

The table shows that the use of a stochastic volatility model such as the SZHW model has a significant impact on the total value of the GAO. The value increases with 4% -50% for a EU data and 4% -18% for a U.S. data, depending on the level of the guarantee.

These price differences are not caused by a volatility effect as both models are calibrated to the same market data in Section 4.4. Figure 4.1 however showed that the distribution of equity prices under a SZHW process has a heavy left tail, but also relatively more mass on the right of the distribution compared to the BSHW process. Given a positive correlation between equity prices and interest rates, and the fact that the GAO pays off when interest rates are low, this means that for the SZHW model there will be some very low payoffs for equity prices in the left tail, but relatively higher payoffs for the remaining scenarios. This is illustrated in table 4.3. For the EU data and $g = 8,23\%$, 50.000 Monte Carlo simulations are generated for both models and the discounted payoffs are segmented in specific intervals.

Table 4.3: classification in intervals of payoffs of SZHW and BSHW model

Payoff	SZHW	BSHW	Diff
0	58,3%	58,5%	-0,2%
(0,1]	7,5%	5,2%	2,2%
(1,10]	22,0%	26,3%	-4,3%
(10,20]	7,2%	6,8%	0,4%
(20,30]	2,7%	1,9%	0,8%
(30,40]	1,2%	0,7%	0,4%
(40,50]	0,5%	0,3%	0,2%
(50,60]	0,3%	0,1%	0,1%
(60,70]	0,2%	0,1%	0,1%
(70,80]	0,1%	0,1%	0,0%
(80,90]	0,1%	0,0%	0,0%
(90,100]	0,0%	0,0%	0,0%
(100,110]	0,0%	0,0%	0,0%
> 110	0,1%	0,0%	0,1%

The table shows that indeed:
- SZHW has relatively much payoffs in the interval (0,1) due to the heavy left tail.
- For the remaining intervals, SZHW has more mass to the right, illustrated by the less frequent payoffs in the interval (1,10) and more frequent payoffs in the intervals greater than 10.

Since the models only have an impact on the time value, the relative changes in time value for in-the-money GAOs are higher, which is also illustrated in table 4.2. One might wonder why the time values for the EU data as negative for high levels of g. The reason for this is that due to the positive correlation between interest rates and equity prices, higher equity volatility means that there is a higher chance of lower payoffs, leading to a lower total option value compared to the intrinsic value. For the U.S. data no negative time values are reported. Reason for this is that due to the lower correlation between interest rates and equity prices, the effect described above is less significant than the positive impact of interest rates on the time value.

4.7.2 Impact of different risk drivers

As noted in Section 4.2, we assume that the survival probabilities are independent of the equity prices and interest rates. It is interesting though to see the impact of significant changes in those survival probabilities on the GAO price and to compare it with the impact of changes in equity prices and interest rates. To get a feeling about this, we apply shocks for these risk drivers as defined in the technical specifications of the Quantitative Impact Study 5 (QIS5) of CEIOPS. QIS5 is the basis for the Solvency 2, a new risk-based framework for regulatory required solvency capital. The shocks are aimed to represent the 99.5% percentile on a 1 year horizon.

Table 4.4 shows the impact of 2 shifts in survival probabilities. The shifts are based on a 25% reduction of mortality rates (longevity risk) and a 15% increase in mortality rates (mortality risk). Table 4.5 shows the impact of 2 shifts in the yield curve. The shifts are given in Appendix 4d. Table 4.6 shows the impact of shocks of +39% and -/- 39% on equity prices.

Table 4.4: impact of survival probabilities on GAO total value

EU, Total value				U.S., Total value			
g	SZHW	qx * 75%	qx * 115%	g	SZHW	qx * 75%	qx * 115%
8,23%	3,82	7,28	2,53	8,23%	5,43	9,08	3,88
7%	0,59	1,61	0,31	7%	1,04	2,88	0,85
8%	2,89	5,82	1,85	8%	3,54	7,61	3,07
9%	8,40	13,63	6,17	9%	8,53	15,22	7,71
10%	17,02	24,01	13,72	10%	16,06	25,05	14,90
11%	27,37	35,49	23,38	11%	25,42	36,13	24,01
12%	38,30	47,25	33,86	12%	35,73	47,76	34,14
13%	49,35	59,05	44,52	13%	46,43	59,56	44,68

Table 4.5: impact of changes in yield curve on GAO total value

EU, Total value				U.S., Total value			
g	SZHW	rates down	rates up	g	SZHW	rates down	rates up
8,23%	3,82	9,91	1,11	8,23%	5,43	10,48	1,54
7%	0,59	2,49	0,10	7%	1,04	3,42	0,25
8%	2,89	8,10	0,77	8%	3,54	8,83	1,15
9%	8,40	17,38	3,24	9%	8,53	17,20	3,58
10%	17,02	28,80	8,65	10%	16,06	27,69	8,21
11%	27,37	40,94	16,71	11%	25,42	39,27	15,05
12%	38,30	53,23	26,19	12%	35,73	51,26	23,50
13%	49,35	65,54	36,13	13%	46,43	63,37	32,79

Table 4.6: impact of shocks on equity prices on GAO total value

EU, Total value				U.S., Total value			
g	SZHW	equity up	equity down	g	SZHW	equity up	equity down
8,23%	3,82	5,31	2,33	8,23%	5,43	7,55	3,31
7%	0,59	0,81	0,36	7%	1,04	1,45	0,63
8%	2,89	4,01	1,76	8%	3,54	4,92	2,16
9%	8,40	11,67	5,12	9%	8,53	11,86	5,20
10%	17,02	23,66	10,38	10%	16,06	22,32	9,80
11%	27,37	38,04	16,70	11%	25,42	35,33	15,51
12%	38,30	53,24	23,36	12%	35,73	49,66	21,80
13%	49,35	68,60	30,10	13%	46,43	64,54	28,32

Although the impact differs for different strike levels, the tables show that the impact of the different risk drivers is reasonably similar for this particular example. Table 4.4 shows that indeed the GAO value increases significantly when a shift down is applied to the mortality rates. A shift up in mortality rate has the opposite effect on the value of the GAO. Similar effects can be seen in table 4.5 for the yield curve shifts. Note that the impact of the yield curve shifts is (coincidently) approximately equal to a 1% shift in the strike level. Of course, higher (lower) equity prices will lead to higher (lower) payments, as shown in table 4.6. But for low strike prices, the impact of changes in equity prices is less than the impact of interest rates and longevity

4.7.3 Comparison results of the two-factor model with Chu and Kwok (2007)

A special case of our modeling framework is considered in Chu and Kwok (2007), namely an equity model with constant volatility. Chu and Kwok (2007) argue that for a two-factor Gaussian interest rate model no analytical pricing formulas exist. Therefore they propose three approximation methods for the valuation of GAOs:

Method of minimum variance duration: This method approximates the annuity with a single zero-coupon bond and minimizes the approximation error by choosing the maturity of the zero-coupon bond to be equal to the stochastic duration.

Edgeworth expansion: This method makes use of the Edgeworth approximation of the probability distribution of the value of the annuity (see Collin-Dufresne and Goldstein (2002)).

Affine approximation: This method approximates the conditional distributions of the risk factors in affine diffusions.

In the paper the runtimes and approximation errors are compared with benchmark results using Monte Carlo simulations and the method of minimum variance duration comes out most favorably. The other approximation methods do have very long runtime, the Edgeworth expansion method requires even more time then a Monte Carlo simulation.

However, as we have shown in section 4.6, it is possible to derive an explicit expression where only a single numerical integration is needed for the case of a two-factor Gaussian interest rate model. It takes hardly any runtime (a couple of hundreds of seconds) to do this numerical integration, whilst it provides exact results. Table 4.7 shows a comparison of the results for the different methods and a Monte Carlo simulation with 1.000.000 sample paths, which are compared to the exact GAO prices obtained by the quasi closed-form expression in (4.48). The parameter setting used is the same as in Chu and Kwok (2007) and is given in Appendix 4e. Numbers in parentheses are relative differences compared to the exact formula for the GAO price. Values that are within the 95% confidence interval of the Monte Carlo estimates are starred (*) and made bold.

Table 4.7: comparison GAO prices for different methods

r_0	Strike	Exact	Min. Var. Duration		Edgeworth Expansion		Affine Approximation		Monte Carlo
0,5%	127%	**11,800***	**11,810***	(-0,1%)	11,816	(-0,1%)	**11,791***	(-0,1%)	11,792
1,0%	123%	**9,756***	**9,771***	(-0,2%)	**9,750***	(-0,1%)	**9,741***	(-0,1%)	9,749
1,5%	118%	**7,874***	**7,896***	(-0,3%)	**7,848***	(-0,3%)	**7,853***	(-0,3%)	7,868
2,0%	114%	**6,169***	6,195	(-0,4%)	6,129	(-0,6%)	**6,142***	(-0,4%)	6,163
2,5%	110%	**4,661***	4,686	(-0,5%)	4,620	(-0,9%)	4,631	(-0,6%)	4,656
3,0%	106%	**3,373***	3,391	(-0,5%)	3,341	(-1,0%)	3,346	(-0,8%)	3,368
3,5%	103%	**2,322***	**2,327***	(-0,2%)	2,300	(-0,9%)	**2,304***	(-0,7%)	2,317
4,0%	99%	**1,510***	**1,501***	(-0,6%)	1,490	(-1,3%)	**1,506***	(-0,3%)	1,507
4,5%	96%	**0,921***	0,901	(-2,2%)	0,894	(-2,9%)	0,931	(-1%)	0,92
5,0%	93%	**0,525***	0,498	(-5,1%)	0,492	(-6,2%)	0,544	(-3,6%)	0,524
5,5%	90%	**0,278***	0,252	(-9,4%)		-		-	0,278
6,0%	88%	**0,136***	0,115	(-15,4%)		-		-	0,135
6,5%	85%	**0,061***	0,047	(-23,3%)		-		-	0,061
7,0%	83%	**0,025***	0,017	(-32,8%)		-		-	0,025

The table shows that the approximation methods considered by Chu and Kwok (2007) break down for higher interest rates, where the guarantee is out-the-money. Note hereby that the first moment of the underlying distribution is the main driving factor for the option price, while for the price of out-of-the-money options the higher moments play a more important role, see Brigo and Mercurio (2006). Taking into account that the mean of the underlying annuity is determined exactly in the approximations, this implies that these methods have severe difficulties in estimating the higher moments of the underlying distribution, resulting in poor an approximation quality of the out-of-money GAOs, see Table 4.7.

The explicit (quasi closed-form) exact formula (4.48) does give highly accurate prices for GAOs across for all strike levels. Both the Monte Carlo method as the explicit formula are unbiased. Differences between the Monte Carlo method and the exact formula are sampling errors as we can see that the 95% confidence interval of the Monte Carlo method for all cases is overlapping with the price of the explicit exact formula. Typically such Monte Carlo noise increases for out-of-the-money options (see Glasserman (2003)) as can also be seen from table 4.7 for the considered GAOs. The careful reader may notice that in the above example the sign of the difference between the Monte Carlo price and the exact formula is always negative, which is due to the fact that the same set of Monte Carlo paths is used for all strikes.

Where the Affine approximation method and the Edgeworth expansion method require a very long runtime (according to Chu and Kwok (2007), the runtime of the Edgeworth expansion is even larger than of the Monte Carlo method with 100.000 sample paths), the runtime of the explicit expression derived in this chapter is comparable to the method of minimum variance duration and takes only a few hundreds of a second. The closed-form exact approach proposed in Section 4.6 is preferable compared to the approaches described in Chu and Kwok (2007), as it gives exact GAO prices over all strike levels whilst being computationally very efficient.

4.8 Conclusions

In this chapter explicit expressions are given for the pricing of GAOs using a stochastic volatility model for equity prices. The considered framework further allows for one-factor and two-factor Gaussian interest rate models, whilst taking the correlation between the equity, the stochastic volatility and the stochastic interest rates explicitly into account. The basis for the explicit formulas for GAOs lies in the fact that under the equity price measure, the GAO can be written in terms of an option on a sum of coupon bearing bonds: after some calculations the Jamshidian (1989) result can be used that expresses the latter option on a sum into a sum of options which can be priced in closed-form. For one-factor interest rate models the price of a GAO can be expressed as a sum of Black and Scholes (1973) options, whereas an explicit expression using a single integral can be established for the case of a two-factor Gaussian interest rate model.

The results in this chapter indicate that the impact of using a stochastic volatility model is significant. In the considered empirical test cases we found that the prices for the GAOs using a stochastic volatility model for equity prices are considerably higher in comparison to the constant volatility model, especially for GAOs with out-of-the-money strikes.

A special case of our modeling framework, that is an equity model with constant volatility, is considered in Chu and Kwok (2007). These authors argue that for a two-factor Gaussian interest rate model no analytical pricing formulas exist and propose several approximation methods for the valuation of GAOs. In this chapter we did derive an explicit expression for the price of a GAO in terms of a single numerical integral, which called for a comparison between these valuation methods. The numerical results show that the use of the quasi closed-form exact approach is preferable compared to the approaches described in Chu and Kwok (2007), as it

gives exact GAO prices over all strike levels whilst being computational very efficient to compute.

Appendix 4a: Pricing of a coupon bearing option under a two-factor interest rate model

Let the pair $x(T)$, $y(T)$ follow a bivariate normal distribution with means μ_x, μ_y, variances $\mathbf{s}_x^2$, σ_y^2 and correlation $\mathbf{r}_{xy}$. The probability density function $f(x, y)$ of $(x(T), y(T))$ is hence given by

$$(4.55) \quad f(x,y) \;=\; \frac{\exp\left\{-\dfrac{1}{2\left(1-\mathbf{r}_{xy}^2\right)}\left[\left(\dfrac{x-\mathbf{m}_x}{\mathbf{s}_x}\right)^2 - 2\mathbf{r}_{xy}\dfrac{(x-\mathbf{m}_x)(y-\mathbf{m}_y)}{\mathbf{s}_x\mathbf{s}_y} + \left(\dfrac{y-\mathbf{m}_y}{\mathbf{s}_y}\right)^2\right]\right\}}{2\mathbf{p}\mathbf{s}_x\mathbf{s}_y\sqrt{1-\mathbf{r}_{xy}^2}}$$

Furthermore, let the time T price of the zero-coupon bond $P(T, t_i)$ maturing at time ti be given by

$$(4.56) \quad P(T,t_i) \;=\; A(T,t_i)e^{-B(a,T,t_i)x(T)-B(b,T,t_i)y(T)}$$

We then come to the following proposition.

Proposition A.1 The expected value of the coupon-bearing option maturing at time T, paying coupons c_i at times $i = 0,..., n$ and with strike K is given by a one-dimensional integral:

$$\mathrm{E}\left\{\left(\sum_{i=0}^{n}c_i P(T,t_i)-K\right)^{+}\right\}$$

$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\left(\sum_{i=0}^{n}c_i A(T,t_i)e^{-B(a,T,t_i)x(T)-B(b,T,t_i)y(T)}-K\right)^{+}f(x,y)dydx$$

$$(4.57) \quad = \int_{-\infty}^{\infty}\frac{e^{-\frac{1}{2}\left(\frac{x-\mathbf{m}_x}{\mathbf{s}_x}\right)^2}}{\mathbf{s}_x\sqrt{2\mathbf{p}}}\left[F_i(x)N\left(h_2(x)\right)-KN\left(h_1(x)\right)\right]dx \;=\; G\left(\mathbf{m}_x,\mathbf{m}_y,\mathbf{s}_x,\mathbf{s}_y,\mathbf{r}_{xy}\right)$$

Proof The result is analogous to the derivation of the swaption price under the two-factor Gaussian interest rate model, we therefore refer to equation (4.31) in Brigo and Mercurio (2006) on pp. 158-159 and the corresponding proof on pp. 173-175.

Appendix 4b: Moments and terminal correlation of the two-factor Gaussian interest rate model

To determine the moments of $x(T)$ and $y(T)$ under the equity price measure, we need to consider the dynamics of (4.40), given under the risk-neutral measure Q, under the equity price measure Q^S. To change the underlying numéraire (see Geman et al. (1996)), we calculate the corresponding Radon-Nikodym derivative which is given by

$$(4.58) \quad \frac{dQ^S}{dQ} = \frac{S(T)B(0)}{S(0)B(T)} = \exp\left[-\frac{1}{2}\int_0^T v^2(u)du + \int_0^T v(u)dW_S^Q(u)\right]$$

The multi-dimensional version of Girsanov's theorem (see Oksendal (2005)) hence implies that

$$(4.59) \quad dW_S^{Q^S}(t) \mapsto dW_S^Q(t) - v(t)dt$$

$$(4.60) \quad dW_x^{Q^S}(t) \mapsto dW_x^Q(t) - \boldsymbol{r}_{xS}v(t)dt$$

$$(4.61) \quad dW_y^{Q^S}(t) \mapsto dW_y^Q(t) - \boldsymbol{r}_{yS}v(t)dt$$

$$(4.62) \quad dW_v^{Q^S}(t) \mapsto dW_v^Q(t) - \boldsymbol{r}_{vS}v(t)dt$$

are Brownian motions under Q^S. Hence under Q^S one has the following model dynamics for the volatility and interest rate process

$$(4.63) \quad dx(t) = -ax(t)dt + \boldsymbol{r}_{xS}\boldsymbol{s}v(t)dt + \boldsymbol{s}\,dW_x^{Q^S}(t) \qquad x(0)=0$$

$$(4.64) \quad dy(t) = -by(t)dt + \boldsymbol{r}_{yS}\boldsymbol{h}v(t)dt + \boldsymbol{h}dW_y^{Q^S}(t) \qquad y(0)=0$$

$$(4.65) \quad dv(t) = \boldsymbol{\check{k}}\left(\boldsymbol{\tilde{y}}-v(t)\right)dt + \boldsymbol{t}\,dW_v^{Q^S}(t) \qquad v(0)=v_0$$

where $\boldsymbol{\check{k}} = \boldsymbol{k} - \boldsymbol{r}_{vS}\boldsymbol{t}$ and $\boldsymbol{\tilde{y}} = \dfrac{\boldsymbol{ky}}{\boldsymbol{\check{k}}}$. Integrating the latter dynamics yields the following explicit expressions:

$$(4.66) \quad v(T) = \boldsymbol{\tilde{y}} + \left(v(0)-\boldsymbol{\tilde{y}}\right)e^{-\boldsymbol{\check{k}}T} + \boldsymbol{t}\int_0^T e^{-\boldsymbol{\check{k}}(T-u)}dW_v^{Q^S}(u)$$

$$(4.67) \quad x(T) = \boldsymbol{r}_{xS}\boldsymbol{s}\left(\frac{\boldsymbol{\tilde{y}}}{a}\left[1-e^{-aT}\right]+\frac{v(0)-\boldsymbol{\tilde{y}}}{(a-\boldsymbol{\check{k}})}\left[e^{-\boldsymbol{\check{k}}T}-e^{-aT}\right]\right)$$

$$+ \frac{\boldsymbol{r}_{xS}\boldsymbol{st}}{(a-\boldsymbol{\check{k}})}\int_0^T\left[e^{-\boldsymbol{\check{k}}(T-u)}-e^{-a(T-u)}\right]dW_v^{Q^S}(u) + \boldsymbol{s}\int_0^T e^{-a(T-u)}dW_x^{Q^S}(u)$$

$$(4.68) \quad y(T) = r_{xS}h\left(\frac{\tilde{y}}{b}\left[1-e^{-bT}\right]+\frac{v(0)-\tilde{y}}{(b-\bar{k})}\left[e^{-\bar{k}T}-e^{-bT}\right]\right)$$

$$+\frac{r_{yS}ht}{(b-\bar{k})}\int_0^T\left[e^{-\bar{k}(T-u)}-e^{-b(T-u)}\right]dW_v^{Q^S}(u)+h\int_0^T e^{-b(T-u)}dW_y^{Q^S}(u)$$

Using Ito's isometry, we have that $(x(T), y(T))$ is normally distributed with means μ_x, μ_y, variances $\mathbf{s}_x^2$, σ_y^2 and correlation $\mathbf{r}_{xy}(T)$ given by

$$(4.69) \quad \mathbf{m}_x = r_{xS}\mathbf{s}\left(\frac{\tilde{y}}{a}\left[1-e^{-aT}\right]+\frac{v(0)-\tilde{y}}{(a-\bar{k})}\left[e^{-\bar{k}T}-e^{-aT}\right]\right)$$

$$(4.70) \quad \mathbf{m}_y = r_{yS}h\left(\frac{\tilde{y}}{b}\left[1-e^{-bT}\right]+\frac{v(0)-\tilde{y}}{(b-\bar{k})}\left[e^{-\bar{k}T}-e^{-bT}\right]\right)$$

$$(4.71) \quad \mathbf{s}_x^2 = \mathbf{s}_1^2(\mathbf{s},a)+\mathbf{s}_2^2(\mathbf{s},a,\mathbf{r}_{xS})+2\mathbf{r}_{12}(\mathbf{s},a,\mathbf{r}_{xv},\mathbf{r}_{xS})\mathbf{s}_1(\mathbf{s},a)\mathbf{s}_2(\mathbf{s},a,\mathbf{r}_{xS})$$

$$(4.72) \quad \mathbf{s}_y^2 = \mathbf{s}_1^2(\mathbf{h},b)+\mathbf{s}_2^2(\mathbf{h},b,\mathbf{r}_{yS})+2\mathbf{r}_{12}(\mathbf{h},b,\mathbf{r}_{yv},\mathbf{r}_{yS})\mathbf{s}_1(\mathbf{h},b)\mathbf{s}_2(\mathbf{h},b,\mathbf{r}_{yS})$$

$$(4.73) \quad \mathbf{r}_{xy} = \frac{Cov\left(x(T),y(T)\right)}{\mathbf{s}_x\mathbf{s}_y}$$

where

$$(4.74) \quad \mathbf{s}_1(l,z) = l\sqrt{\frac{1-e^{-2zT}}{2z}}$$

$$(4.75) \quad \mathbf{s}_2(l,z,r) = \frac{rlt}{z-\bar{k}}\sqrt{\frac{1}{2\bar{k}}+\frac{1}{2z}-\frac{2}{(\bar{k}+z)}-\frac{e^{-2\bar{k}T}}{2\bar{k}}-\frac{e^{-2zT}}{2z}+\frac{2e^{-(\bar{k}+z)T}}{(\bar{k}+z)}}$$

$$(4.76) \quad \mathbf{r}_{12}(l,z,r_1,r_2) = r_1\frac{l^2r_2t}{\mathbf{s}_1(l,z)\mathbf{s}_2(l,z,r_2)(z-\bar{k})}\left(\frac{1-e^{-(z+\bar{k})T}}{(z+\bar{k})}-\frac{1-e^{-2zT}}{2z}\right)$$

$$(4.77) \quad Cov\big(x(T), y(T)\big) \;=\; \boldsymbol{r}_{xy}\boldsymbol{sh}\left(\frac{1-e^{-(a+b)T}}{(a+b)}\right) + \boldsymbol{r}_{xv}\boldsymbol{s}\,\frac{\boldsymbol{r}_{yS}\boldsymbol{ht}}{(b-\bar{\boldsymbol{k}})}\left(\frac{1-e^{-(a+\bar{k})T}}{(a+\bar{\boldsymbol{k}})} - \frac{1-e^{-(a+b)T}}{(a+b)}\right)$$

$$+\, \boldsymbol{r}_{yv}\boldsymbol{h}\,\frac{\boldsymbol{r}_{xS}\boldsymbol{st}}{(a-\bar{\boldsymbol{k}})}\left(\frac{1-e^{-(b+\bar{k})T}}{(b+\bar{\boldsymbol{k}})} - \frac{1-e^{-(a+b)T}}{(a+b)}\right) + \frac{\boldsymbol{r}_{xS}\boldsymbol{st}}{(a-\bar{\boldsymbol{k}})}\frac{\boldsymbol{r}_{yS}\boldsymbol{ht}}{(b-\bar{\boldsymbol{k}})}$$

$$\left(\frac{1-e^{-2\bar{k}T}}{2\bar{\boldsymbol{k}}} + \frac{1-e^{-(a+b)T}}{(a+b)} - \frac{1-e^{-(a+\bar{k})T}}{(a+\bar{\boldsymbol{k}})} - \frac{1-e^{-(b+\bar{k})T}}{(b+\bar{\boldsymbol{k}})}\right)$$

Appendix 4c: Special case: independent equity price process or pure interest rate guaranteed annuities

If one does not link the GAO to the performance of the equity (e.g. seen in the Netherlands) the expression (4.4) for the GAO price can be simplified significantly. One then has that the GAO price is given by

$$(4.78) \quad C(T) \;=\; {}_{r-x}p_x \mathrm{E}^{Q}\left[\exp\left(-\int_0^T r(u)du\right)g\left(\sum_{i=0}^{n} c_i P(T,t_i)-K\right)^{+}\right]$$

$$(4.79) \qquad =\; {}_{r-x}p_x\, g\, P(0,T)\mathrm{E}^{Q^T}\left[\left(\sum_{i=0}^{n} c_i P(T,t_i)-K\right)^{+}\right]$$

where the above expectation is taken with respect to the T-forward measure Q^T, which uses the zero-coupon bond price maturing at time T as numéraire. Moreover, also in case one assumes the equity price is independent from the annuity, e.g. according to Boyle and Hardy (2003) and Pelsser (2003), one ends up with the same expectation as (4.78); one only has to multiply the currency $P(0, T)$ with the expectation future equity price, i.e. in (4.79) one only has to multiply this formula with the expected future equity price. In the following paragraphs we will derive explicit expressions for the GAO price under both one-factor and two-factor Gaussian interest rates.

C.1 Hull-White model

Under Q^T, one has the following expression for the stochastic process $x(T)$, driving the short interest rate (see Brigo and Mercurio (2006), Pelsser (2000)):

$$(4.80) \quad x(T) \;=\; \boldsymbol{m}_x^T + \boldsymbol{s}\int_0^T e^{-a(T-u)}dW_x^{Q^T}(u)$$

hence from Ito's isometry, we have $x(T)$ is normally distributed with mean μ_x and variance $\boldsymbol{s}_x^2$ given by

$$(4.81) \quad m_x^T = -\frac{s^2}{a^2}\left(1-e^{-aT}\right)+\frac{s^2}{2a^2}\left(1-e^{-2aT}\right)$$

$$(4.82) \quad s_x^T = s\sqrt{\frac{1-e^{-2aT}}{2a}}$$

Just as in Section 4.5, we have that $x(T)$ is normally distributed with the same variance s_x^2, but with a different mean μ_x^T. Hence completely analogous to Section 4.5, one can use the Jamshidian (1989) result and write the call option on the sum of zero-coupon bonds as a sum of zero-coupon bond call options: let x^* solve

$$(4.83) \quad \sum_{i=0}^{n} c_i A(T,t_i)e^{-B(T,t_i)x^*} = K$$

and let

$$(4.84) \quad K_i = A(T,t_i)e^{-B(T,t_i)x^*}$$

Using Jamshidian (1989), we have that the price of GAO is equal to the price of a sum of zero-coupon bond options, i.e.

$$(4.85) \quad C(T) = {}_{r-x}p_x g P(0,T)\mathrm{E}^{Q^T}\left[\sum_{i=0}^{n} c_i \left(A(T,t_i)e^{-B(T,t_i)x(T)} - K_i\right)^+\right]$$

As the bond price again follows a log-normal distribution, one can express the GAO price in terms of the Black and Scholes (1973) formula:

$$(4.86) \quad C(T) = {}_{r-x}p_x g P(0,T)\sum_{i=0}^{n} c_i \left(F_i N(d_1^i) - K_i N(d_2^i)\right)$$

$$(4.87) \quad F_i = e^{M_i+\frac{1}{2}V_i}$$

$$(4.88) \quad d_1^i = \frac{\ln(F_i / K_i)+\frac{1}{2}V_i}{\sqrt{V_i}}$$

$$(4.89) \quad d_2^i = d_1^i - \sqrt{V_i}$$

where

(4.90) $M_i = \ln A(T, t_i) - B(T, t_i) \boldsymbol{m}_x$

(4.91) $V_i = B^2(T, t_i) \left(\boldsymbol{s}_x^T \right)^2$

and note that the above expression only deviates from (4.33) in the different means and variances for the $x(T)$ process.

C.2 Gaussian Two-factor model

Under Q^T, one has the following expression for the stochastic processes $x(T)$, $y(T)$ that drive the short interest rate (see Brigo and Mercurio (2006)):

(4.92) $x(T) = \boldsymbol{m}_x^T + \boldsymbol{s} \int_0^T e^{-a(T-u)} dW_x^{Q^T}(u)$ $y(T) = \boldsymbol{m}_y^T + \boldsymbol{h} \int_0^T e^{-b(T-u)} dW_y^{Q^T}(u)$

hence $x(T)$, $y(T)$ are normally distributed with means μ_x^T, μ_y^T, variances $\boldsymbol{s}_x^2$, $\boldsymbol{s}_y^2$ and correlation $\boldsymbol{r}_{xy}(T)$ given by:

(4.93) $\boldsymbol{m}_x^T = -\left(\dfrac{\boldsymbol{s}^2}{a^2} + \boldsymbol{r}_{xy} \dfrac{\boldsymbol{sh}}{ab} \right)\left[1 - e^{-aT}\right] + \dfrac{\boldsymbol{s}^2}{2a^2}\left[1 - e^{-2aT}\right] + \dfrac{\boldsymbol{r}_{xy}\boldsymbol{sh}}{b(a+b)}\left[1 - e^{-(a+b)T}\right]$

(4.94) $\boldsymbol{m}_y^T = -\left(\dfrac{\boldsymbol{h}^2}{b^2} + \boldsymbol{r}_{xy} \dfrac{\boldsymbol{sh}}{ab} \right)\left[1 - e^{-bT}\right] + \dfrac{\boldsymbol{h}^2}{2b^2}\left[1 - e^{-2bT}\right] + \dfrac{\boldsymbol{r}_{xy}\boldsymbol{sh}}{a(a+b)}\left[1 - e^{-(a+b)T}\right]$

(4.95) $\boldsymbol{s}_x^T = \boldsymbol{s}\sqrt{\dfrac{1 - e^{-2aT}}{2a}}$

(4.96) $\boldsymbol{s}_y^T = \boldsymbol{h}\sqrt{\dfrac{1 - e^{-2bT}}{2b}}$

Hence analogously to Section 4.6, one has that the GAO price is given by

(4.97) $C(T) = {}_{r-x}p_x\, gP(0,T) G\left(\boldsymbol{m}_x^T, \boldsymbol{m}_y^T, \boldsymbol{s}_x, \boldsymbol{s}_y, \boldsymbol{r}_{xy}(T) \right)$

where G is given by an explicit expression, defined by equation (4.57) of appendix 4a.

Appendix 4d: Yield curve shocks

In paragraph 4.7.2 the 2 shocks given in table 4.8 are applied to the yield curves. These shocks are aimed to represent the 99.5% percentile on a 1 year horizon in the Quantitative Impact Study 5 (QIS 5) of CEIOPS.

Table 4.8: yield curve changes for the up and down shock in QIS5

Maturity	up	down
1	75%	-75%
2	65%	-65%
3	56%	-56%
4	50%	-50%
5	46%	-46%
6	42%	-42%
7	39%	-39%
8	36%	-36%
9	33%	-33%
10	31%	-31%
11	30%	-30%
12	29%	-29%
13	28%	-28%
14	28%	-28%
15	27%	-27%
16	28%	-28%
17	28%	-28%
18	28%	-28%
19	29%	-29%
20	29%	-29%
21	29%	-29%
22	30%	-30%
23	30%	-30%
24	30%	-30%
25	30%	-30%
26	30%	-30%
27	30%	-30%
28	30%	-30%
29	30%	-30%
30	30%	-30%

Appendix 4e: Model setup of the Chu and Kwok (2007) case

In this appendix we describe the numerical input of the example being used in Chu and Kwok (2007). We also report the relative differences between the GAO price obtained by their methods and the explicit expression in (4.48) for that example. Note that as the Black-Scholes-G2++ model, used in Chu and Kwok (2007), is special case of the Schöbel-Zhu-G2++ considered in 4.6, we can one on one translate their parameters into our modeling framework. As the notation is slightly different, we explicitly provide this translation into our modeling framework.

The GAO is specified using the guaranteed rate $g = 9\%$, the current age of the policy holder $x = 50$ and his retirement age $r = 65$, with corresponding probability of survival $_{r-x}p_x = 0.9091$ and time to expiry for the GAO equal to $T = 15$ years. The equity price is modeled by the Black and Scholes (1973) model with parameters $q = 5\%$, $S(0) = 100 \exp(-q\,T) = 47,24$ and $s_S = 10\%$, where q denotes the continuous dividend rate and s_S the constant volatility of the equity price. The current (continuous) yield curve is given by (4.98) and for the G2++ interest rate model (e.g.

see Brigo and Mercurio (2006)) the following parameters are used: $a = 0{,}77$, $b = 0{,}08$, $s = 2\%$, $h = 1\%$, $r_{xy} = -0.7$, where the correlations between equity and interest rate drivers given by $r_{xS} = 0{,}5$ and $r_{yS} = 0{,}0071$. Finally, the i-year survival probabilities c_i from policy holder's retirement age 65 are provided in the following table:

Table 4.9: i-year survival probabilities from policy holder's retirement age 65

c0	1,000	c9	0,830	c18	0,489	c27	0,100
c1	0,987	c10	0,802	c19	0,441	c28	0,073
c2	0,973	c11	0,771	c20	0,393	c29	0,050
c3	0,958	c12	0,737	c21	0,345	c30	0,033
c4	0,941	c13	0,702	c22	0,298	c31	0,020
c5	0,923	c14	0,663	c23	0,252	c32	0,012
c6	0,903	c15	0,623	c24	0,209	c33	0,006
c7	0,881	c16	0,580	c25	0,168	c34	0,003
c8	0,857	c17	0,535	c26	0,132	c35	0,001

In paragraph 4.7.3 we compared the prices of the explicit solution (4.48) and estimates obtained using 1.000.000 Monte Carlo simulations with the Minimum Variance, the Edgeworth and Affine Approximation method which are used in Chu and Kwok (2007). These results can be found in table 7.6, where a comparison is given for different levels r_0 of the yield curve provided by the (continuous) yields

$$(4.98)\quad Y(T) = r_0 + 0{,}04(1 - e^{-0{,}2T})$$

Chapter 5

On Stochastic Mortality Modeling*

5.1 Introduction

As mentioned in chapter 1, important risks to be quantified are mortality and longevity risk. Not only is this an important risk for most (life) insurers and pension funds, the resulting solvency margin will also be part of the fair value reserve. Reason for this is that it is becoming best practice for the quantification of the risk margin to apply a Cost of Capital rate to the solvency capital necessary to cover for unhedgeable risks, such as mortality and longevity risks.

There is a vast literature on stochastic modeling of mortality rates. Often used models are for example those of Lee and Carter (1992), Renshaw and Haberman (2006), Cairns et al (2006a), Currie et al (2004) and Currie (2006). For an extensive review we refer to section 5.2.

All well known models have nice features but also disadvantages. In this chapter a mortality model is proposed that aims at combining the nice features from existing models, while eliminating the disadvantages of existing models. More specifically, the model fits historical data very well, is applicable to a full age range, captures the cohort effect, has a non-trivial (but not too complex) correlation structure, has no robustness problems and can take into account parameter risk, while the structure of the model remains relatively simple.

The remainder of the chapter is organized as follows. First, in section 5.2 the existing literature is extensively reviewed, focusing on stochastic mortality models and the criteria for them. In section 5.3 a new mortality model is proposed. Section 5.4 describes the fitting procedure of the model and gives results of the fitting process for mortality of different countries. Section 5.5 shows simulation results of mortality rates and the results of a robustness test. In section 5.6 a risk neutral version of the model is given, which can be used for pricing. Section 5.7 describes a possible method to account for parameter risk for the proposed mortality model. Conclusions are given in section 5.8.

5.2 Literature review: criteria and models

Due to the increasing focus on risk management and measurement for insurers and pension funds, the literature on stochastic mortality models has developed rapidly during the last decennium. In this section an overview of current literature on stochastic mortality models and criteria for them is given.

5.2.1 Criteria for stochastic mortality models

It is important to consider whether a specific stochastic mortality model is a good model or not. Therefore, Cairns et al (2008a) defined criteria against which a model can be assessed:

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 and model forecasts should be robust relative to the period of data and range of ages employed.
5) Forecast levels of uncertainty and central trajectories should be plausible and consistent with historical trends and variability in mortality data.
6) The model should be straightforward to implement using analytical methods or fast numerical algorithms.
7) The model should be relatively parsimonious.
8) It should be possible to use the model to generate sample paths and calculate prediction intervals.
9) The structure of the model should make it possible to incorporate parameter uncertainty in simulations.
10) At least for some countries, the model should incorporate a stochastic cohort effect.
11) The model should have a non-trivial correlation structure.

An important additional criterion is that the model is applicable for a full age range. Some models are designed for higher ages only (say 60 years or older). However, the portfolios of insurers and pension funds usually exist of policyholders from age 20 and older. One would want to model the mortality rates and their dependencies for the whole portfolio consistently, therefore the model should be applicable for the whole age range.

The existing models meet most of the above criteria. However, as far as we know, none of the existing models meet all of the above criteria (although some are close), see section 5.8 and Cairns et al (2007).

5.2.2 Stochastic mortality models

Stochastic mortality models either model the central mortality rate or the initial mortality rate (see Coughlan et al (2007)). The central mortality rate $m_{x,t}$ is defined as:

$$(5.1) \quad m_{x,t} = \frac{D_{x,t}}{E_{x,t}} = \frac{\#\, deaths\ during\ calendar\ year\ t\ aged\ x\ last\ birthday}{average\ population\ during\ calendar\ year\ t\ aged\ x\ last\ birthday}$$

The initial mortality rate q_x is the probability that a person aged x dies within the next year. The different mortality measures are linked by the following approximation:

$$(5.2) \quad q_x \approx 1 - e^{-m_x}$$

One of the most well known stochastic mortality models is the model of Lee and Carter (1992):

$$(5.3) \quad \ln(m_{x,t}) = a_x + b_x \boldsymbol{k}_t$$

where a_x and b_x are age effects and $\boldsymbol{k}_t$ is a random period effect. Cairns et al (2007, 2008a and 2008b) noted several disadvantages of the Lee-Carter model:

- It is a 1-factor model, resulting in mortality improvements at all ages being perfectly correlated (trivial correlation structure).
- For countries where a cohort effect is observed in the past, the model gives a poor fit to historical data.
- The uncertainty in future death rates is proportional to the average improvement rate b_x. For high ages this can lead to this uncertainty being too low, since historical improvement rates have often been lower at high ages.
- The basic version of the model can result in a lack of smoothness in the estimated age effect b_x.

There is whole strand of literature on additions or modifications of the Lee-Carter model, for example Brouhns et al (2002), Lee and Miller (2001), Booth et al (2002), Girosi and King (2005), De Jong and Tickle (2006), Delwarde et al (2007) and Renshaw and Haberman (2003). Most of these models tackle one of the problems of the Lee-Carter model, but the other disadvantages still remain.

The first model that incorporated the cohort effect was proposed in Renshaw and Haberman (2006):

$$(5.4) \quad \ln(m_{x,t}) = a_x + b_x^1 \boldsymbol{k}_t + b_x^2 \boldsymbol{g}_{t-x}$$

where $\boldsymbol{g}_{t-x}$ is a random cohort effect that is a function of the year of birth (t-x).
For countries where a cohort effect is observed in the past, this model provides a significant better fit to the historical data. However, CMI (2007) and Cairns et al (2007, 2008b) find that the Renshaw-Haberman suffers from a lack of robustness. Furthermore, although the model has an additional stochastic factor for the cohort effect, for most of the simulated mortality rates the correlation structure is still trivial. Especially when using a wide age range, the simulated cohort parameters are only relevant for the higher ages in the far end of the projection.

Currie (2006) introduced a simplification of the Renshaw-Haberman model that removes the robustness problem:

$$(5.5) \quad \ln(m_{x,t}) = a_x + \boldsymbol{k}_t + \boldsymbol{g}_{t-x}$$

However, the fit quality is less good compared to the Renshaw-Haberman model, and the problem with the trivial correlation structure still remains.

When fitting models (5.4) and (5.5) to an age range of say 20-85, the modeled cohort effect can result in odd looking humps in the projected mortality rates over time. This problem will be further highlighted in the next paragraph.

Furthermore, Cairns et al (2008b) observe that for England & Wales and United States data, the fitted cohort effect appears to have a trend in the year of birth. This suggests that the cohort effect compensates the lack of a second age-period effect, as well as trying to capture the cohort effect in the data.

Cairns et al (2006a) introduced the following model:

$$(5.6) \quad \text{logit}(q_{x,t}) = \ln\left(\frac{q_{x,t}}{1-q_{x,t}}\right) = \boldsymbol{k}_t^1 + \boldsymbol{k}_t^2 (x-\bar{x})$$

where $\bar{x}$ is the mean age in the sample range and $(\boldsymbol{k}_t^1, \boldsymbol{k}_t^2)$ is assumed to be a bivariate random walk with drift. Cairns et al (2007) also introduced some additions on model (5.6), amongst others capturing the cohort effect. The models have multiple factors that result in a (desired) non-trivial correlation structure, while the structure of the model is relatively simple. However, those models are all designed for higher ages only. When using these models for full age ranges, the fit quality will be relatively poor and the projections are likely to be biologically unreasonable.

### 5.2.3 Problems with modeling cohort effect

Various explanations have been put forward for cohort effects that have been identified in the past. For example, for the United Kingdom Willets (2004) mentions historical patterns of smoking behavior and the impact of early life experience on health in later life. He states that there are a number of reasons to believe that this cohort effect will have an enduring impact on rates of mortality improvement in future decades.

The investigations on cohort effects often concentrate on birth years until about 1945. This is natural, since in most cases the cohort effect is an effect on health in later life, so one needs observations of mortality rates for middle age and higher ages to verify the existence of the cohort effect. When applying models (5.4) and (5.5) to a full age range, say 20-85, cohort parameters are also fitted for birth years 1945-1980. This means that for these birth years, (cohort) movements for young ages (which can be volatile) are projected into the future. This affects the mortality rates for higher ages in a similar degree, since the cohort effect is usually modeled in a multiplicative way. However, given the possible nature of the movements for these specific birth years (for example AIDS, drug and alcohol abuse and violence) it is unclear

whether these effects do have a persistent effect on the future mortality rates for these cohorts. And if so, it is questionable whether a high relative cohort effect for young ages will have a similar relative effect on mortality of higher ages, given the nature of the cohort effect for young ages[13].

Figure 5.1 shows a best estimate and percentiles of mortality rates for 75 years old males, using the Currie (2006) model applied to United States mortality for an age range of 20-84.

Figure 5.1: projected mortality rates, 75 years old male – Currie (2006) model
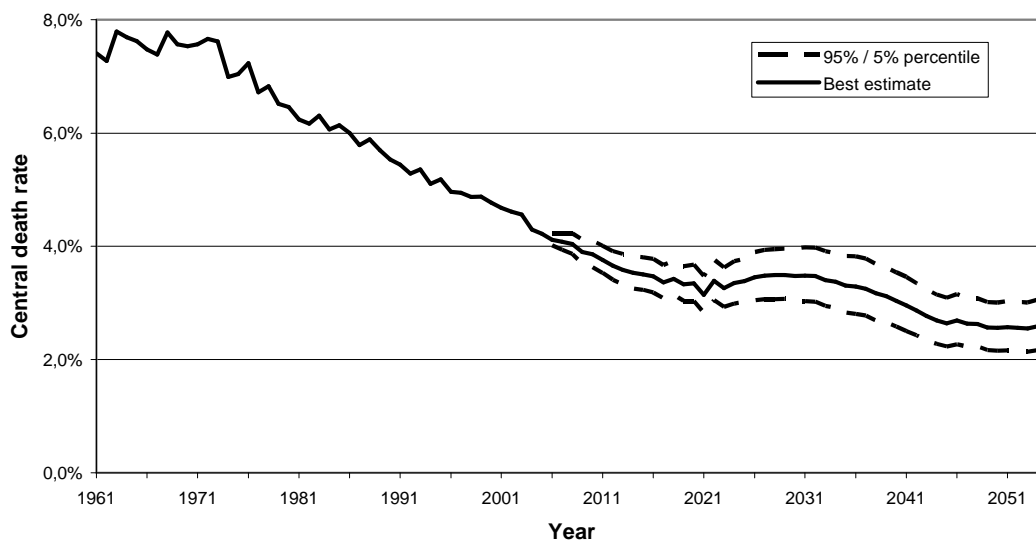
The figure shows an odd-looking hump around 2020-2040 and flattening of projected mortality thereafter, corresponding with patterns in the fitted cohort parameters for birth years 1945-1980.

Given the considerations above and the odd-looking results when taking into account cohort effects of recent birth years, it might be wise to only include the cohort effect for early birth years (say until year 1945) in the fitting of the model. The cohort effect for later birth years (so > 1945) can be simulated from the fitted distributions. An additional advantage of this is that the simulation of the cohort effect becomes relevant for higher ages already in the beginning of the projection, leading to a non-trivial correlation structure.

5.3 A new stochastic mortality model

The models mentioned in the previous section all have some nice features:

-   the a_x term of the Lee-Carter model makes it suitable for full age ranges

---

[13] Note that the Renshaw-Haberman tries to capture this in the parameter b_x^2. However, this is based on the cohort effects for earlier birth years, which could have a significantly different nature.

- the Renshaw-Haberman model addresses the cohort effect and fits well to historical data
- the Currie model has a simpler structure then the Renshaw-Haberman model, making it more robust
- the models of Cairns et al (2006a, 2007) have multiple factors, resulting in a non-trivial correlation structure, while the structure of the model is relatively simple

In this section a mortality model is proposed that combines those nice features, while eliminating the disadvantages mentioned in the previous section.

5.3.1 The proposed model

As for most other stochastic mortality models, the quantity of interest is the central mortality rate $m_{x,t}$. The proposed model for $m_{x,t}$ is:

$$(5.7) \quad \ln(m_{x,t}) = a_x + k_t^1 + k_t^2 \left(\bar{x} - x \right) + k_t^3 \left(\bar{x} - x \right)^+ + g_{t-x}$$

where $\left(\bar{x} - x \right)^+ = \max \left(\bar{x} - x, 0 \right)$. The model has 4 stochastic factors, but has a similar relatively simple structure as the Currie (2006) and the Cairns et al (2006a, 2007) models.

The a_x is similar as in the Lee-Carter model and makes sure that the basic shape of the mortality curve over ages is in line with historical observations. Next to the a_x, the model has 4 stochastic factors $(k_t^1, k_t^2, k_t^3, g_{t-x})$. The parameters of the model can be fitted using the methodology described in section 5.4, after which suitable ARIMA-processes (see paragraph 2.2.1) are selected for the various factors.

The factor k^1 represents changes in the level of mortality for all ages. Following the reasoning in Cairns et al (2006b), the (long-term) stochastic process for this factor should not be mean reverting. Reason for this is that it is not expected that higher mortality improvements in some years will surely be compensated by lower mortality improvements in later years.

The factor k^2 allows changes in mortality to vary between ages, to reflect the historical observation that improvement rates can differ for different age classes.

Furthermore, historical data seems to indicate that the dynamics of mortality rates at lower ages (up to age 40 / 50) can be (significantly) different at some times. For example, think of developments like AIDS, drugs and alcohol abuse, and violence. The factor k^3 is added to capture these dynamics.

The factor g_{-x} is capturing the cohort effect, in the same way as the models of Currie (2006) and Cairns et al (2007). As mentioned in paragraph 5.2.2, the process for this factor should not have a trend. Therefore, a trendless mean reverting process will be assumed for g_{-x}.

Next to g_{-x}, the factors k^2 and k^3 allow the model to have a non-trivial correlation structure between ages. Fitting non-stationary ARIMA-process for factors k^2 and k^3 could result (in some scenarios) in projected scenarios where the shape of the mortality curve over ages is not

biologically reasonable. Therefore, a stationary (mean reverting) process will be assumed for these factors.

In most cases mortality projections for a wide age range are needed. However, if one is only interested in higher ages (say age 60 and older), the factor $\boldsymbol{k}^3$ is not needed and can be left out. This reduces the model to:

$$(5.8) \quad \ln(m_{x,t}) = a_x + \boldsymbol{k}_t^1 + \boldsymbol{k}_t^2 (\overline{x} - x) + \boldsymbol{g}_{t-x}$$

This reduced model still has all the favorable characteristics of model (5.7), but is more suitable for high ages only.

5.3.2 Identifiability constraints

Just like most stochastic mortality models, the proposed mortality model has an identifiability problem, meaning that different parameterizations could lead to identical values for $\ln(m_{x,t})$. Note that the following parameterization leads to similar values for $\ln(m_{x,t})$:

$$(5.9) \quad \begin{aligned} \tilde{\boldsymbol{g}}_{t-x} &= \boldsymbol{g}_{t-x} + \boldsymbol{y}_1 + \boldsymbol{y}_2 (t - x) \\ \boldsymbol{k}_t^1 &= \boldsymbol{k}_t^1 - \boldsymbol{y}_1 - d\overline{x}\boldsymbol{y}_2 - \boldsymbol{y}_2 t \\ \boldsymbol{k}_t^2 &= \boldsymbol{k}_t^2 + d\boldsymbol{y}_2 \\ \tilde{a}_x &= a_x + (1+d) x\boldsymbol{y}_2 \end{aligned}$$

where $\boldsymbol{y}_1$, $\boldsymbol{y}_2$ and d are constants.

This can be resolved by setting identifiability constraints. We use the approach of Cairns et al (2007, model M6) for this, leading to the following constraints:

$$\sum_{c=c_0}^{c_1} \boldsymbol{g}_c = 0$$

$$(5.10) \quad \sum_{c=c_0}^{c_1} c \boldsymbol{g}_c = 0$$

$$\sum_t \boldsymbol{k}_t^3 = 0$$

where c_0 and c_1 are the earliest and latest year of birth to which a cohort effect is fitted, and $c = t-x$. The rationale behind the choice of first two constraints is that if the function $\boldsymbol{y}_1 + \boldsymbol{y}_2 (t\text{-}x)$ is fitted to $\gamma_{t\text{-}x}$ the constraints ensure that $\hat{\boldsymbol{y}}_1 = \hat{\boldsymbol{y}}_2 = 0$. This results in a fitted process for $\gamma_{t\text{-}x}$ that will fluctuate around 0 and there will be no constant trend up or down. This means that the constraints in (5.10) force the process of $\boldsymbol{g}_{t\text{-}x}$ only to be used to capture the cohort effect and not to compensate lack of age-period effects. The third constraint is only used to normalize the estimates for $\boldsymbol{k}^3$.

Other approaches for setting the identifiability constraints are also possible, see for example Cairns et al (2007) and Renshaw and Haberman (2006).

5.4 Fitting the model

An important aspect of stochastic mortality models is the quality of the fit of the model to historical mortality data. In this section the methodology for fitting the model is described, and a comparison of fit quality with other models is made for mortality rates of the United States (US), England & Wales (E&W) and The Netherlands (NL).

5.4.1 Fitting methodology

Brouhns et al (2002) described a fitting methodology for the Lee-Carter model based on a Poisson model. The main advantage of this is that it accounts for heteroskedasticity of the mortality data for different ages. This method has been used more commonly after that, also for other models, see for example Renshaw and Haberman (2003, 2006) and Cairns et al (2007).

This fitting methodology will be applied to the model proposed in section 5.3. Therefore, the number of deaths is modeled using the Poisson model, implying:

(5.11) $D_{x,t} \sim \text{Poisson}(E_{x,t}\, m_{x,t})$

where $D_{x,t}$ is the number of deaths, $E_{x,t}$ is the exposure (see (5.1)) and $m_{x,t}$ is modeled as in (5.7). The parameter set $\boldsymbol{f}$ is fitted with maximum likelihood estimation, where the log-likelihood function of model (5.11) is given by:

(5.12) $L(\boldsymbol{f};D,E) = \sum_{x,t}\left\{ D_{x,t}\ln\left[E_{x,t}m_{x,t}(\boldsymbol{f})\right] - E_{x,t}m_{x,t}(\boldsymbol{f}) - \ln\left(D_{x,t}!\right) \right\}$

Because of the specific nature of the problem, there are (as far as we know) no commercial statistical packages that implement this Poisson regression with constraints. Therefore we have used the R-code of the (free) software package "Lifemetrics" as a basis for fitting (5.12)[14]. Another reason for using this is to make an honest comparison of the fit quality of the model proposed in this chapter and existing models (also modeled in Lifemetrics), which is the topic of the next paragraph.

Besides estimates for a_x, the fitting procedure described above leads to time series of estimations of $\boldsymbol{k}^1$, $\boldsymbol{k}^2$, $\boldsymbol{k}^3$ and $\boldsymbol{g}_{t-x}$. The next step in fitting the model is selecting and fitting a suitable ARIMA-process to these time series (see paragraph 5.4.3).

---

[14] See www.lifemetrics.com and http://www.r-project.org/. Lifemetrics is an (open source) toolkit for measuring and managing longevity and mortality risk, designed by J.P. Morgan.

5.4.2 Comparison of fit quality with existing models

To evaluate whether the proposed model fits historical data well, we have fitted the model to three different data sets and compared the fitting results with those of models from the Lifemetrics toolkit. The three used data sets are:
- United States, Males, 1961-2005, ages 20-84
- England & Wales, Males, 1961-2005, ages 20-89
- The Netherlands, Males, 1951-2005, ages 20-90

The data consists of numbers of deaths $D_{x,t}$ and the corresponding exposures $E_{x,t}$ and is extracted from www.mortality.org[15].

As in Cairns et al (2007), the models are compared using the Bayes Information Criterion (*BIC*). The measure *BIC* provides a trade-off between fit quality and parsimony of the model. The *BIC* is defined as:

$$(5.13) \quad BIC = L(\hat{f}) - \frac{1}{2} K \ln N$$

where $\hat{f}$ is the maximum likelihood estimate of the parameter vector, N is the number of observations and K is the number of parameters being estimated.

Table 5.1 gives a comparison of the fitting results (in terms of *BIC*) of the model proposed in section 5.3 and existing models (fitted with the Lifemetrics toolkit).

Table 5.1: comparison BIC for proposed model and existing models

BIC mortality models *	U.S.	E&W	NL
Plat	-24.506	-18.151	-18.425
Renshaw-Haberman (2006)	-25.971	-18.062	-18.632
Currie (2006)	-37.489	-19.805	-18.597
Lee-Carter (1992)	-47.542	-22.949	-20.353
Cairns et al (2007, model M7)	-56.571	-27.730	-21.055
Cairns et al (2006)	-294.928	-66.744	-31.511

* higher BIC is more favorable

The table shows that for these specific data sets the proposed model gives the best fitting results, closely followed by the Renshaw-Haberman model. The *BIC* for the other models is (sometimes significantly) less. The models of Cairns et al (2006a, 2007) do not perform very well for this age range, since they are designed for higher ages only.

In the fitting process of the models above the cohort effect was taken into account for all birth years of the dataset. However, given the reasons mentioned in paragraph 5.2.3, for the remaining of this chapter we will exclude the cohort effect for birth years later than 1945 in the fitting of the model. In general this will reduce the quality of the fit somewhat, as is shown in table 5.2.

---

[15] Note that a longer history is available. We used these historic periods (for the U.S. and E&W) to be able to roughly compare results with Cairns et al (2007) and Cairns et al (2008).

BIC mortality models *	U.S.	E&W	NL
Plat	-24.506	-18.151	-18.425
Plat (exluding cohorts > 1945)	-32.392	-18.927	-18.378

* higher BIC is more favorable

The fitting results of the model are still good when excluding the cohort effect for birth years > 1945, certainly considering the fact that the *BIC* of the other models which include a cohort effect (Renshaw-Haberman (2006), Currie (2006) and Cairns et al (2007)) would also be less favorable when excluding these birth years[16].

Note that the proposed model nests the model of Currie (2006). For nested models, the use of a likelihood ratio test is more appropriate than the use of the *BIC* measure. The likelihood ratio (*LR*) test can be used to test the null hypothesis that the nested model (in this case, the Currie (2006) model) is the correct model against the alternative that the more general model (the model proposed in this chapter) is correct. The likelihood ratio test statistic is:

$$(5.14) \quad \mathbf{x}_{LR} = 2\left[L(\hat{\mathbf{f}}) - L(\tilde{\mathbf{f}}) \right]$$

where $L(\hat{\mathbf{f}})$ is the log-likelihood of the general model and $L(\tilde{\mathbf{f}})$ of the nested model.

Under the null hypothesis, $\mathbf{x}_{LR}$ has a Chi-squared distribution with *J* degrees of freedom, *J* being the additional parameters being estimated in the general model compared to the nested model. Therefore, the null hypothesis can be rejected if:

$$(5.15) \quad \mathbf{x}_{LR} > \mathbf{c}^2_{J,\mathbf{a}}$$

where **a** is the significance level. Alternatively, the *p*-value can be determined for this test:

$$(5.16) \quad p = 1 - \mathbf{c}^{2\ -1}_{J}\left(2\left[L(\hat{\mathbf{f}}) - L(\tilde{\mathbf{f}}) \right] \right)$$

The *p*-value is the probability of obtaining the observed value, assuming that the null hypothesis is true. If the *p*-value is lower than **a**, the null hypothesis is rejected. Table 5.3 shows the results of the likelihood ratio test for the three data sets.

---

[16] An alternative way of presentation could be to exclude the birth years > 1945 for all models that include the cohort effect and compare the *BIC* on that basis. However, the other models are all fitted with the Lifemetrics tool that includes all birth years.

Table 5.3: LR test, null hypothesis Currie (2006) model against proposed model

|	Likelihood Ratio test statistic	Degrees of freedom	p-value
U.S.	26.677	89	< 0,0001
E&W	4.023	89	< 0,0001
Holland	1.245	109	< 0,0001

The table shows that for each data set the null hypothesis is rejected overwhelmingly. Therefore, the conclusion above (based on *BIC* results) that the model proposed in this chapter is preferable to the nested Currie (2006) model is supported by the results from the likelihood ratio test.

5.4.3 Fitting the ARIMA processes – U.S. Males

In the remainder of this chapter, we will focus on the population of U.S. males[17]. The next step in the process is selecting and fitting a suitable ARIMA-process to the time series of k^1, k^2, k^3 and g_{t-x}. The fitted parameters k^1, k^2, k^3 and g_{t-x} for U.S. males are given in figure 5.2. The figure shows that the pattern of the important parameter k^1 is well-behaved. The patterns of the other parameters all reveal some autoregressive behavior.

Since the factor k^1 drives a significant part of the uncertainty in mortality rates, its relatively regular behavior (for this particular dataset) will also show in the projected uncertainty (in other words, the confidence intervals will be relatively narrow).

Figure 5.2: estimated values of k^1, k^2, k^3 and g_{t-x}

---
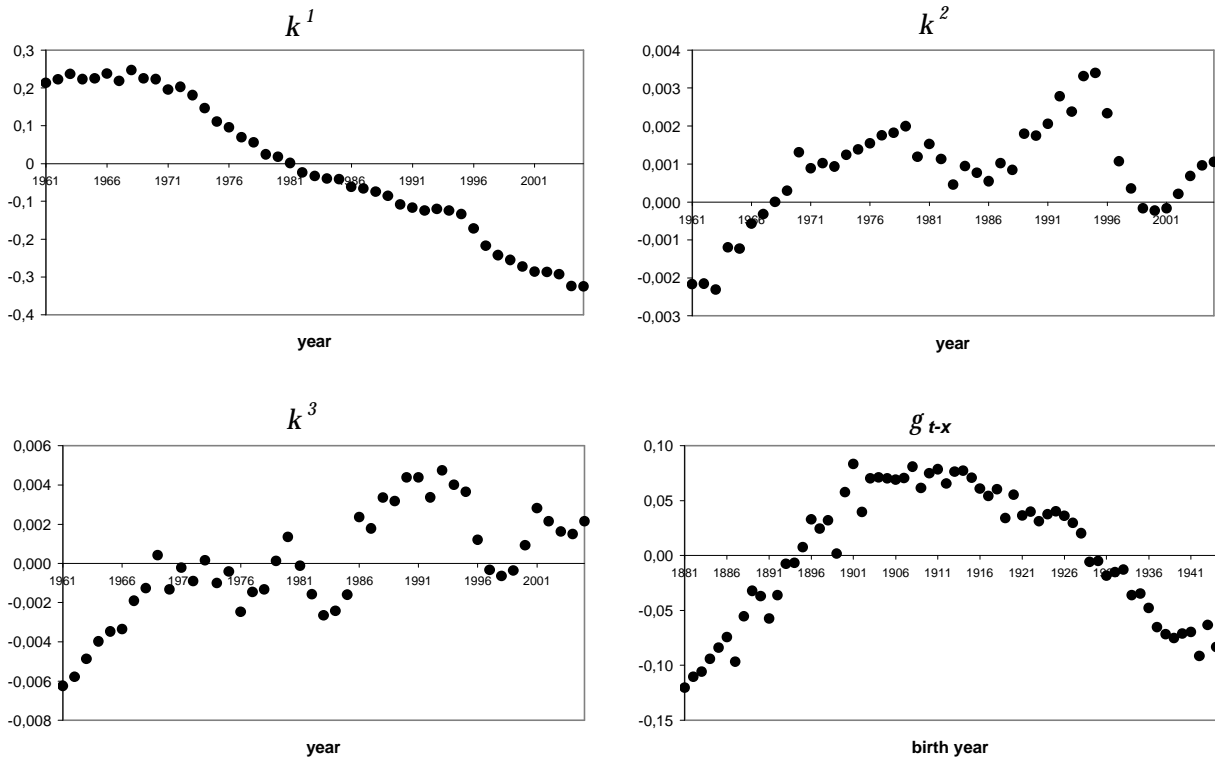
[17] To be able to compare simulation results with Cairns et al (2007), we can either use US males or E&W males. The choice for U.S. males is more or less arbitrary.

Now for each of these time series all relevant ARIMA(p,d,q) processes for the range p, d, q = 0, 1, 2, 3 are fitted and the most favorable process in terms of *BIC* is selected. The selected ARIMA processes are:

-     k^1: ARIMA(0,1,0)
-     k^2, k^3 and g_{t-x}: ARIMA(1,0,0), no constant

It is commonly assumed (see Renshaw and Haberman (2006), CMI (2007) and Cairns et al (2008b), that the process for g_{t-x} is independent of the other processes, so the parameters of this process can be fitted independently using Ordinary Least Squares (OLS). The other processes can be fitted simultaneously using Seemingly Unrelated Regression (SUR, see Zellner (1963)).

Table 5.4 gives the fitted parameters, standard errors, t-ratios and *BIC*'s and table 5.5 shows the fitted standard deviations (on the diagonal) and correlations.

Table 5. 4: fitted parameters for the processes $y_{t+1} - y_t = q_1$ and $y_{t+1} = q_2\, y_t$ *

Fit results	k^1	k^2	k^3	g_{t-x}
q_1	-0,0131 0,0022 -5,925			
q_2		0,9539 0,0495 19,280	0,8440 0,0656 12,871	0,9366 0,0361 25,941
BIC	120,83	267,65	229,21	163,55

* In each cell for $?_1$ and $?_2$, top: fitted parameter, bottom left: standard error, bottom right: t-ratio

Table 5.5: fitted standard deviations (on the diagonal) and correlations

	k^1	k^2	k^3	g_{t-x}
k^1	0,0150	0,2539	0,0274	0
k^2	0,2539	0,0005	0,0144	0
k^3	0,0274	0,0144	0,0012	0
g_{t-x}	0	0	0	0,0175

5.5 Mortality projections – U.S. Males

Cairns et al (2008b) performed an extensive assessment of the out-of-sample performance of several stochastic mortality models, focusing on England & Wales and U.S. Males between 60 and 90 years old. The main criteria used in this assessment were biological reasonableness and robustness of the (stochastic) forecasts. Based on these criteria and specifically for these datasets, they concluded that the models of Lee and Carter (1992), Renshaw and Haberman (2006) and Cairns et al (2007, model M8) did not perform in a satisfactory way. Furthermore, they concluded that the models of Currie (2006) and Cairns et al (2006, 2007 model M7) did produce plausible results and seem robust.

This section shows the simulation results and results of robustness tests for the proposed mortality model.

5.5.1 Simulation results – U.S. Males

Using the fitted ARIMA processes and the fitted values for a_x and $\boldsymbol{g}_{t-x}$ (see appendix 5a), future mortality rate scenarios for U.S. males can be constructed using Monte Carlo simulation. Figure 5.3 shows simulation results for ages 25, 45, 65 and 84 for U.S. males[18]. The best estimate projection is given and the 5% and 95% percentiles.

The results are biologically plausible. For higher ages, the widths of the confidence intervals are broadly similar as the models of which Cairns et al (2008b) concluded that they produced biologically plausible results. The results for younger ages (25 and 45) also seem plausible, where the observed historical variability is reflected in the confidence intervals.

Figure 5.3: simulation results for U.S. Males
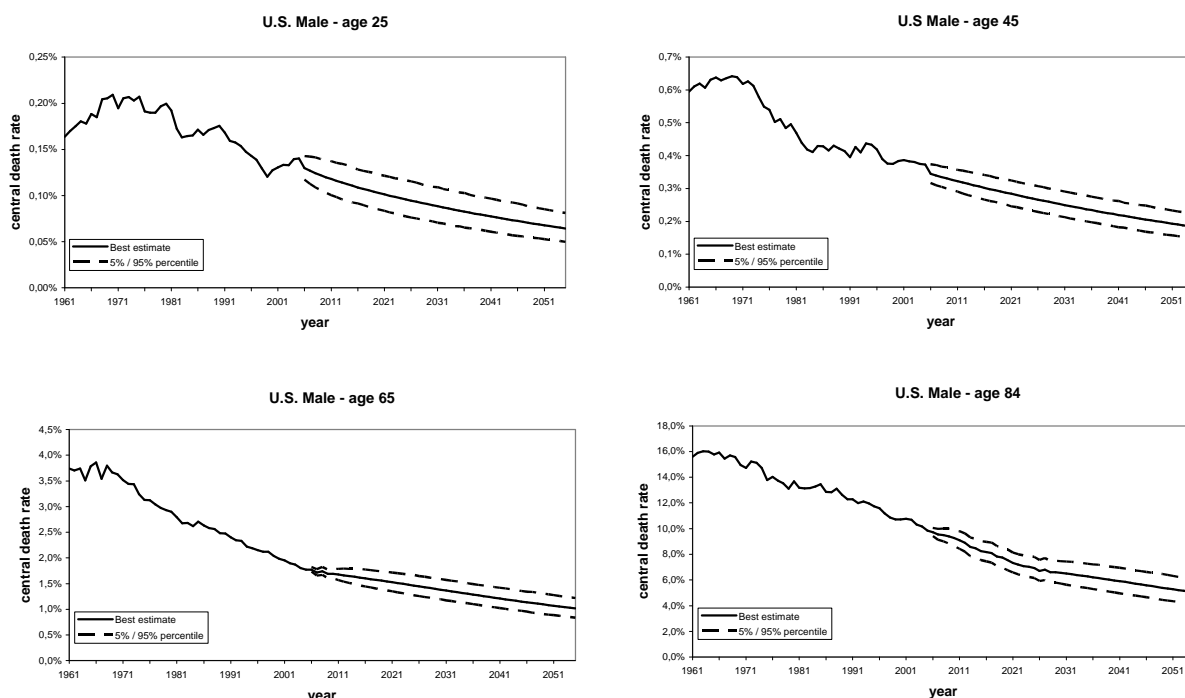
5.5.2 Robustness of simulation results

Some models suffer from a lack of robustness. For example, Cairns et al (2007, 2008b) find that the Renshaw-Haberman model is not robust for changes in range of years. They link this to the shape of the likelihood function. Robust models probably have a unique maximum that remains broadly unchanged when the range of years or ages is changed. Models that lack robustness possibly have more than one maximum, so when changing the range of years or ages the optimizer can jump from one local maximum to another, yielding different parameter estimates.

---

[18] Simulation results for England & Wales and the Netherlands are given in appendix 5b.

The model proposed in this chapter is tested for robustness using the same test as in Cairns et al (2008b). This means that the simulation results above are compared with those of two sensitivities. These sensitivities are:
1) The model is fitted only to historical data from 1981-2005 (instead of 1961-2005)
2) The model (5.7) is fitted to historical data from 1961-2005, but the stochastic models for k^1, k^2, k^3 and g_{t-x} are only fitted to a restricted set of parameter estimates (being only the final 24 $k^{(i)}$'s and the final 45 g_{t-x}'s)

Of course, if there is a change in trend or variability in the period 1981-2005 compared to 1961-2005, it is inevitable, for all models, that the simulation results will be somewhat different.

The results are given in appendix 5c and are not significantly different as the results shown in paragraph 5.5.1. The confidence intervals for age 25 are wider, due to the higher variability for younger ages in the past 25 years. Conclusion is that the proposed model is robust for the sensitivities given above.

Furthermore, a backtest is carried out, meaning that the model is fitted to historical data from 1961-1986 and the forecast results are compared with the actual observations for the period 1987-2005. Also for this backtest, the proposed model performs adequately (see the results in appendix 5c).

5.5.3 Comparison with other models
Paragraph 5.5.1 and 5.5.2 showed that the proposed model produces plausible results and seems robust. Cairns et al (2008b) came to the same conclusion for the models of Currie (2006) and Cairns et al (2006, 2007 model M7).

The models of Cairns et al (2006, 2007 model M7) are designed for higher ages, so will not produce plausible results for lower ages. Compared to those models the proposed model has the advantage that it does produce plausible results for a full age range.

Compared to the model of Currie (2006) the proposed model has the advantage that it has a non-trivial correlation structure. This is important because often insurers and pension funds have different type of exposures for younger or middle ages (term insurance, pre-retirement spouse option) than for higher ages (pensions, annuities). Aggregating these different types of exposures can only be done sufficiently if the model has a non-trivial correlation structure. Assuming an almost perfect correlation between ages, as in the Currie (2006) model, will possibly lead to an overstatement of the diversification benefits that arise when aggregating these exposures.

5.6 Risk neutral specification of the model

The model proposed in section 5.3 is set up in the so-called real-world measure, suitable for assessing risks for example in the context of Solvency 2. For pricing instruments of which the payoff depends on future mortality rates, a risk adjusted pricing measure has to be defined. A common approach is to specify a risk neutral measure Q that is a suitable basis for pricing, see for example Milevsky and Promislow (2001), Dahl (2004), Schrager (2006), Cairns et al (2006a,

2006b) and Biffis et al (2006). The risk neutral specification proposed below is in line with the approach of Cairns et al (2006a).

Note that the market for longevity or mortality instruments is currently (very) far from complete. Consequence of this is that the risk neutral measure Q is not unique. Given the absence of any market price data, it seems wise to keep the specification of the risk neutral process relatively simple. For the same reason, it is difficult to judge whether one risk neutral mortality model is better than another.

5.6.1 Risk neutral dynamics

The stochastic process for the factors $\boldsymbol{k}^1$, $\boldsymbol{k}^2$, $\boldsymbol{k}^3$ and $\boldsymbol{g}_{-x}$ in the real world measure P can generally be specified as:

$$(5.17) \quad \mathrm{K}_t = \Theta\left(\mathrm{K}_{t-1}, \boldsymbol{e}_{t-1}\right) + \Sigma \, \mathrm{Z}_t^P$$

Where K_t is the vector with factors $\boldsymbol{k}^1$, $\boldsymbol{k}^2$, $\boldsymbol{k}^3$ and $\boldsymbol{g}_{-x}$, $\Theta(\mathrm{K}_{t-1}, \boldsymbol{e}_{t-1})$ is the drift of the process, $\boldsymbol{SS}$' is the covariance matrix and Z^P is a 4 x 1 vector with standard normal random variables under measure P.

Now the proposed dynamics under the risk neutral measure Q are:

$$(5.18) \quad \begin{aligned} \mathrm{K}_t &= \Theta\left(\mathrm{K}_{t-1}, \boldsymbol{e}_{t-1}\right) + \Sigma\left[\mathrm{Z}_t^Q - \boldsymbol{1}\right] \\ &= \Theta\left(\mathrm{K}_{t-1}, \boldsymbol{e}_{t-1}\right) - \Sigma \boldsymbol{1} + \Sigma \, \mathrm{Z}_t^Q \end{aligned}$$

where the vector $\boldsymbol{1}$ represents the "market price of risk" associated with the process K_t. Like Cairns et al (2006a), we assume that the market price of risk is constant over time. When market prices for longevity or mortality derivatives are available, the vector $\boldsymbol{1}$ should be calibrated in such a way that the theoretical prices under the measure Q are approximately equal to market prices.

5.6.2 Calibration of the market price of risk

Currently, there is no developed market for longevity derivatives. However, Loeys et al (2007) have the opinion that q-forwards are most likely to become the basis of a longevity market. Therefore, in this paragraph the risk neutral model (5.18) is calibrated to q-forward prices. Of course, calibration to other instruments such as longevity bonds or survivor swaps, would also be possible.

A q-forward is a simple capital market instrument with similar characteristics as an interest rate swap. The instrument exchanges a realized mortality rate in a future period for a pre-agreed fixed mortality rate. This is shown in figure 5.4. The pre-agreed fixed mortality rate is based on a projection of mortality rates, coming from the Lifemetrics toolkit.

Figure 5.4: working of a q-forward
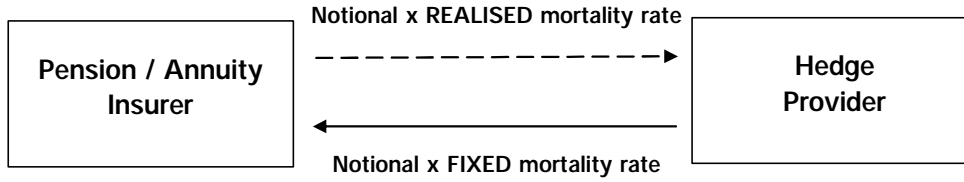
For example, when the realized mortality rate is lower than expected, the pension / annuity insurer will receive a payment which (partly) compensates the increase of the expected value of the insurance liabilities (caused by the decreasing mortality rates).

The basis for the instrument is the (projected) mortality of a country population, not the mortality of a specific company or portfolio. This makes the product and the pricing very transparent compared to traditional reinsurance.

Although there have been some transactions involving q-forwards, currently no market quotes for *q*-forwards are publicly available. However, Loeys et al (2007) give an indication and examples on how such an instrument would be priced in practice. In absence of real market data, we calibrate the model to *q*-forward prices resulting from these examples.

Loeys et al (2007) give the following formula for setting the fixed q-forward rate:

(5.19) $q_{forward} = (1 - horizon * Sharpe\ ratio * q_{vol}) * q_{expected}$

where they have used 10 years for the horizon of the derivative, 25% for the Sharpe ratio and the volatility q_{vol} based on historical data. Table 5.6 shows the results for q-forwards with a horizon of 10 and starting ages of 35, 45, 55 and 65, where $q_{expected}$ is based on model (5.7). Since in this chapter the central mortality rate $m_{x,t}$ is modeled, the results are also translated into these terms, which makes the calibration easier.

Table 5.6: indication q-forward rate for horizon 10 and translation to m-forward

Age start	Age end	q_{vol}	$q_{expected}$	$q_{forward}$	$m_{expected}$	$m_{forward}$	h
35	45	2,31%	0,306%	0,288%	0,307%	0,289%	0,060
45	55	1,53%	0,709%	0,682%	0,712%	0,685%	0,039
55	65	1,01%	1,618%	1,578%	1,632%	1,590%	0,026
65	75	1,47%	3,542%	3,412%	3,606%	3,471%	0,038

Now interpreting $m_{forward}$ as the expectation under the risk neutral measure and $m_{expected}$ as the expectation under real world measure leads to:

(5.20) $E^Q\left(m_{x_{end},t_{end}}\right) = g\,E^P\left(m_{x_{end},t_{end}}\right)$

where x_{end} and t_{end} are age and year at the end of the contract and g can be extracted from the market (or in this case, from table 6). Taking logarithms leads to:

$$(5.21) \quad \ln\left[E^Q\left(m_{x_{end},t_{end}}\right)\right] = \ln\left[g\right] + \ln\left[E^P\left(m_{x_{end},t_{end}}\right)\right]$$

Because the only difference between the processes under Q and P is in the drift term, we can assume that:

$$(5.22) \quad \ln\left[E^Q\left(m_{x_{end},t_{end}}\right)\right] - \ln\left[E^P\left(m_{x_{end},t_{end}}\right)\right] = E^Q\left(\ln\left[m_{x_{end},t_{end}}\right]\right) - E^P\left(\ln\left[m_{x_{end},t_{end}}\right]\right) = \ln\left[g\right]$$

Now since this difference in the drift term is the matrix $-\Sigma l$, for a horizon k the following holds:

$$(5.23) \quad E^Q\left(\ln\left[m_{x_{end},t_{end}}\right]\right) - E^P\left(\ln\left[m_{x_{end},t_{end}}\right]\right) = -\sum_{t=1}^{k} W_t \Sigma l$$

where W_t is the matrix of weights that are used to translate the values for $\boldsymbol{k}^1$, $\boldsymbol{k}^2$, $\boldsymbol{k}^3$ and $\boldsymbol{g}_{-x}$ into values for $\ln(m_{x,t})$. This leads to:

$$(5.24) \quad h = \sum_{t=1}^{k} W_t \Sigma l$$

where $h = -\ln[g]$, of which the values are given in the table above. From (5.24) the market prices of risk can be solved:

$$(5.25) \quad \hat{l} = \left(\sum_{t=1}^{k} W_t \Sigma\right)^{-1} h$$

Now we use this formula to calibrate the market prices of risk to the q-forwards specified above. The weights matrices W_t vary slightly for each year t depending on the development of the age:

$$(5.26) \quad W_t = \begin{pmatrix} 1 & (\bar{x}-45) & (\bar{x}-45)^+ & 1 \\ 1 & (\bar{x}-55) & (\bar{x}-55)^+ & 1 \\ 1 & (\bar{x}-65) & (\bar{x}-65)^+ & c_t \\ 1 & (\bar{x}-75) & (\bar{x}-75)^+ & 0 \end{pmatrix}$$

where $c_t = 0$ for $t < 5$ and $c_t = 1$ otherwise. Reason for this time dependence is that the simulated cohort effect gradually comes into the projections for age 65. The right bottom item of W_t is 0 because for age 75 the cohort effect does not play a role within the horizon of 10 year for this age.

Applying formula (5.25) using (5.26), the results in table 5 and the vector h from table 5.6 leads to the calibrated market prices of risk given in table 5.7.

	I
k^1	1,2430
k^2	0,9793
k^3	-0,5756
g_{t-x}	-0,7338

When the market develops and a number of q-forward prices are available, the market prices of risk can be calibrated by minimizing the squared errors between the theoretical prices and the market prices:

$$(5.27) \quad \hat{I} = \underset{I}{Min} \sum_{i=1}^{p} \left(h_i - \sum_{t=1}^{k} W_t^i \Sigma I \right)^2$$

where p is the number of q-forwards the model is calibrated to.

5.7 Parameter Uncertainty

As mentioned in the criteria for stochastic mortality models in paragraph 5.2.1, the structure of the model should make it possible to incorporate parameter uncertainty in simulations. There are three possible approaches for including this parameter uncertainty:

1) Using a formal Bayesian framework, see Cairns (2000) and Cairns et al (2006a)
2) Simulate the parameter values using the estimates and the standard errors obtained in the estimation process
3) Applying a bootstrap procedure such as described in Brouhns et al (2005) and Renshaw and Haberman (2008)

In a Bayesian framework a prior, possibly non-informative, distribution is assumed for the parameters. Combining this prior distribution with the sample data and the assumed density function of a particular stochastic process leads to a posterior distribution. This posterior distribution can be used to assess the parameter uncertainty.

Approach 2) uses the standard errors of the fitted parameters to incorporate the parameter uncertainty. When least squares or maximum likelihood estimation is used the estimators are either normally or asymptotically normally distributed.

Approach 3) uses bootstrapping techniques, either applied to $D_{x,t}$ (semi-parametric bootstrap) or to the residuals $D_{x,t} - \tilde{D}_{x,t}$ (residual bootstrap),

While a formal Bayesian approach is more elegant than approach 2) and 3), it generally leads to significantly more complexity. Market Chain Monte Carlo (MCMC) methods or importance sampling might be necessary, because the posterior distribution often does not belong to a known

class of probability density functions (see for example Kleibergen and Hoek (1996)). Since the approaches should not lead to significantly different parameter uncertainty, it is questionable whether it is worth increasing the complexity of the model significantly for slightly more elegance. Therefore, approach 1) does not have our preference.

By using approach 2), parameter uncertainty can be incorporated in the model proposed in this chapter. For the stochastic processes of $\boldsymbol{k}^1$, $\boldsymbol{k}^2$, $\boldsymbol{k}^3$ and $\boldsymbol{g}_{-x}$ the estimates and standard errors given in table 4 can be used as the moments of the normal distributions of the parameters. For the parameter estimates of the $\boldsymbol{g}_{-x}$'s (until birth year 1945) and the a_x's the standard errors have to be calculated separately. Starting point for this (see for example Verbeek (2008)) is the vector of first derivatives of the log-likelihood function, the so-called score vector $s(\boldsymbol{f})$:

$$(5.28) \quad s(\boldsymbol{f}) = \frac{\partial L(\boldsymbol{f})}{\partial \boldsymbol{f}} = \sum_{i=1}^{N} \frac{\partial L_i(\boldsymbol{f})}{\partial \boldsymbol{f}} = \sum_{i=1}^{N} s_i(\boldsymbol{f})$$

where $\boldsymbol{f}$ is the parameter set, $L_i(\boldsymbol{f})$ is de log-likelihood function for observation i and N is the number of observations. Now the covariance matrix V_{par} to be used can be estimated with:

$$(5.29) \quad V_{par} = \left(\sum_{i=1}^{N} s_i(\hat{\boldsymbol{f}}) s_i(\hat{\boldsymbol{f}})' \right)^{-1}$$

The standard errors for the $\boldsymbol{g}_{-x}$'s (until birth year 1945) and the a_x's are the square roots of the relevant diagonal elements.

Approach 2) is the most practical method. However, Renshaw and Haberman (2008) noted that the confidence intervals for the Lee-Carter and Renshaw-Haberman models vary for different versions of identifiability constraints when using this method. This phenomenon was not seen when using approach 3). Although the question remains whether their conclusion still holds for other models (such as the one proposed in this chapter) and different sort of constraints (such as the ones used in this chapter and in Cairns et al (2007, model M6), approach 3) seems the most appropriate method for addressing parameter uncertainty in the model proposed in this chapter.

5.8 Conclusions

All well known stochastic mortality models have nice features but also disadvantages. In this chapter a stochastic mortality model is proposed that aims at combining the nice features from existing models, while eliminating the disadvantages.

The chapter shows that the fit of the model to historical data is better than the well-known mortality models. Also, the model has 4 stochastic factors, leading to a (desired) non-trivial (but not too complex) correlation structure between ages. Due to a (Lee-Carter type) variable that describes the shape of the mortality curve over ages and the inclusion of a separate stochastic factor for young ages, the model is applicable to a full age range. Furthermore, the model captures the cohort effect and has no robustness problems. The chapter also describes how to

incorporate parameter uncertainty into the model.

In paragraph 5.2.1, a list of criteria for stochastic mortality models is given. Table 5.8 shows whether the existing models and the proposed model meet those criteria. A large part of the table is based on Cairns et al (2007) and the conclusions in Cairns et al (2008b).

Table 5.8: comparison of stochastic mortality models – satisfaction of criteria *

Satisfaction of criteria mortality models	Renshaw & Haberman	Currie	Lee & Carter	Cairns et al (2006)	Cairns et al (2007, M7)	Plat
1) Positive mortality rates	+	+	+	+	+	+
2) Consistency historical data	+	+	+ / -	+	+	+
3) Long-term biological reasonableness	+	+	+	+	+	+
4) Robustness	-	+	+	+	+	+
5) Forecasts biological reasonable	+	+	+/-	+	+	+
6) Ease of implementation	+	+	+	+	+	+
7) Parsimony	+/-	+/-	+	+	+/-	+/-
8) Possibility generating sample paths	+	+	+	+	+	+
9) Allowance for parameter uncertainty	+	+	+	+	+	+
10) Incorporation cohort effects	+	+	-	-	+	+
11) Non-trivial correlation structure	+/-	+/-	-	+	+	+
12) Applicable for full age range	+/-	+/-	+/-	-	-	+

* +: meets criterion, +/-: partly meets criterion, -: does not meet criterion

The table shows that, apart from partly meeting the parsimony criteria, the proposed model meets all of the criteria. None of the existing models meet all of the criteria. Of the existing models, the model of Currie (2006) is most close to meeting all criteria. However, the advantages of the proposed model compared to the model of Currie (2006) are:
- Better fit to historical data
- Non-trivial correlation structure, which is important in solvency calculations
- Better applicable to a full age range, amongst others due to the inclusion of a separate factor for younger ages

So by combining the nice features of the existing models, the proposed model has eliminated the disadvantages of those models, and as a result the model meets all of the criteria set for stochastic mortality models.

For pricing purposes, a risk neutral version of the model is given, that can be used for pricing. This model is calibrated to some indicative prices for longevity derivatives.

Appendix 5a: U.S. Male - estimates for a_x and g_{t-x}

age	a_x	age	a_x	birth year	g_{t-x}	birth year	g_{t-x}
20	-6,3983	53	-4,6403	1881	-0,1202	1914	0,0773
21	-6,3328	54	-4,5587	1882	-0,1103	1915	0,0708
22	-6,3382	55	-4,4660	1883	-0,1057	1916	0,0609
23	-6,3483	56	-4,3843	1884	-0,0940	1917	0,0544
24	-6,3826	57	-4,3008	1885	-0,0839	1918	0,0603
25	-6,3812	58	-4,1935	1886	-0,0743	1919	0,0340
26	-6,3777	59	-4,1249	1887	-0,0967	1920	0,0554
27	-6,3645	60	-4,0197	1888	-0,0553	1921	0,0366
28	-6,3301	61	-3,9471	1889	-0,0321	1922	0,0400
29	-6,3426	62	-3,8376	1890	-0,0370	1923	0,0313
30	-6,3080	63	-3,7746	1891	-0,0574	1924	0,0376
31	-6,2548	64	-3,7040	1892	-0,0361	1925	0,0402
32	-6,2078	65	-3,6100	1893	-0,0074	1926	0,0361
33	-6,1571	66	-3,5483	1894	-0,0067	1927	0,0297
34	-6,1303	67	-3,4648	1895	0,0075	1928	0,0203
35	-6,0643	68	-3,3801	1896	0,0329	1929	-0,0057
36	-6,0026	69	-3,3068	1897	0,0244	1930	-0,0050
37	-5,9383	70	-3,2088	1898	0,0321	1931	-0,0184
38	-5,8493	71	-3,1444	1899	0,0016	1932	-0,0152
39	-5,8183	72	-3,0406	1900	0,0576	1933	-0,0128
40	-5,7386	73	-2,9656	1901	0,0834	1934	-0,0360
41	-5,6532	74	-2,8864	1902	0,0396	1935	-0,0347
42	-5,5658	75	-2,7994	1903	0,0703	1936	-0,0477
43	-5,4921	76	-2,7193	1904	0,0711	1937	-0,0651
44	-5,4221	77	-2,6375	1905	0,0702	1938	-0,0717
45	-5,3288	78	-2,5568	1906	0,0690	1939	-0,0751
46	-5,2423	79	-2,4671	1907	0,0705	1940	-0,0709
47	-5,1539	80	-2,3649	1908	0,0809	1941	-0,0697
48	-5,0575	81	-2,2824	1909	0,0615	1942	-0,0914
49	-4,9965	82	-2,1863	1910	0,0749	1943	-0,0632
50	-4,8950	83	-2,0906	1911	0,0785	1944	-0,0831
51	-4,8109	84	-1,9992	1912	0,0655	1945	-0,0702
52	-4,7155			1913	0,0765		

Appendix 5b: simulation results England & Wales and the Netherlands

In this appendix the simulation results for England & Wales (E&W) and the Netherlands are given. The best estimate projection is given and the 5% and 95% percentiles. Information about the fitted parameters and underlying ARIMA processes is available upon request.
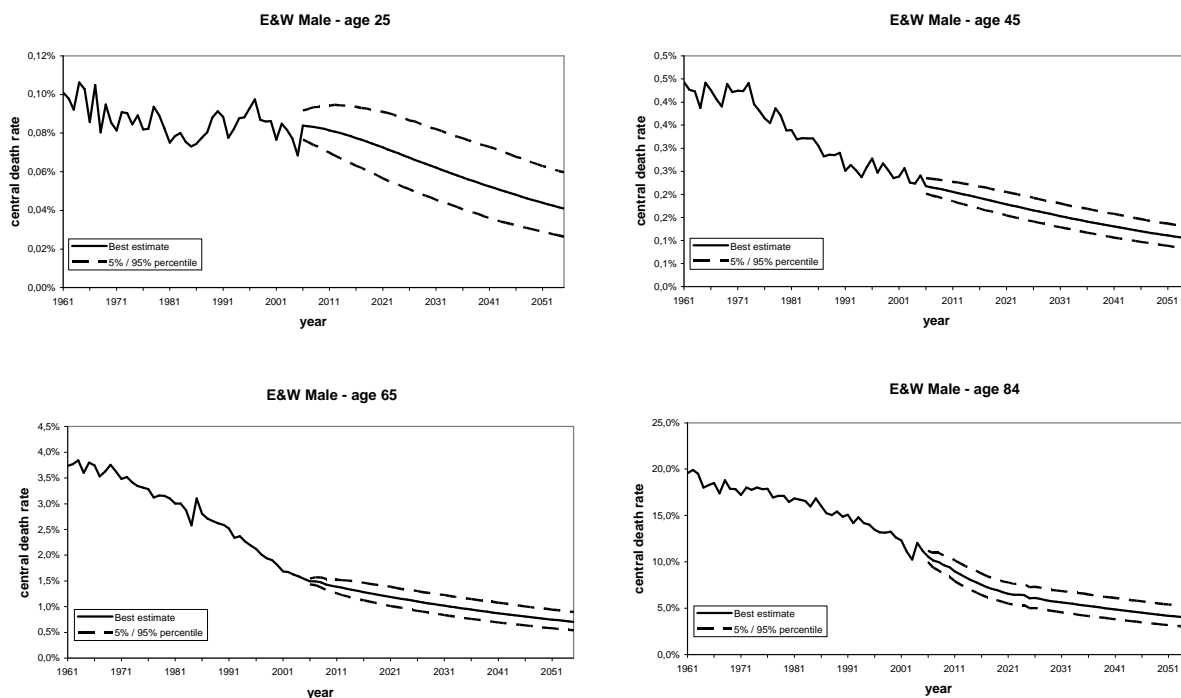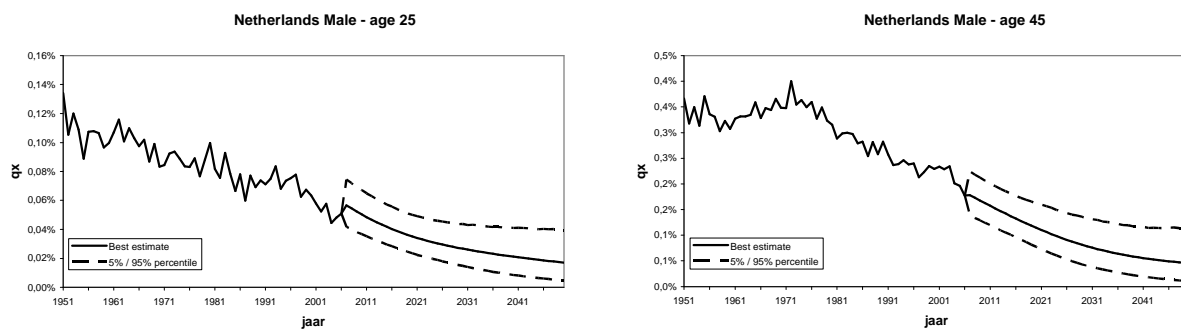
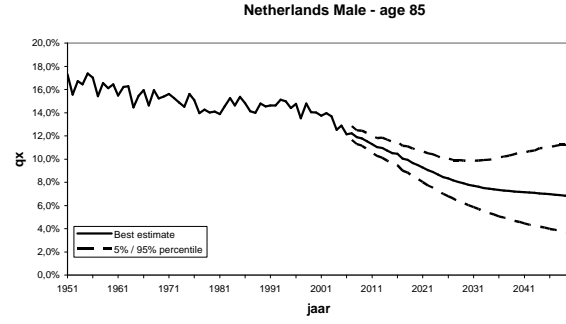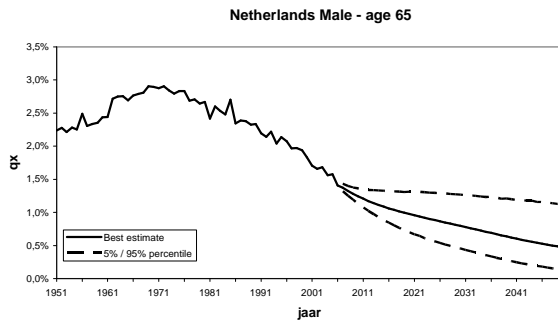*Figure 5.5: simulation results for England & Wales males*

Figure 5.6: simulation results the Netherlands

Netherlands Male - age 65

Netherlands Male - age 85

Appendix 5c: simulation results robustness tests

In this appendix the simulation results are given for the sensitivities that have been specified to test the robustness of the model:

1) The model is fitted only to historical data from 1981-2005 (instead of 1961-2005)
2) The model (5.7) is fitted to historical data from 1961-2005, but the stochastic models for k^1, k^2, k^3 and g_{t-x} are only fitted to a restricted set of parameter estimates (being only the final 24 $k^{(i)}$'s and the final 45 g_{t-x}'s)

Figure 5.7: simulation results sensitivity 1

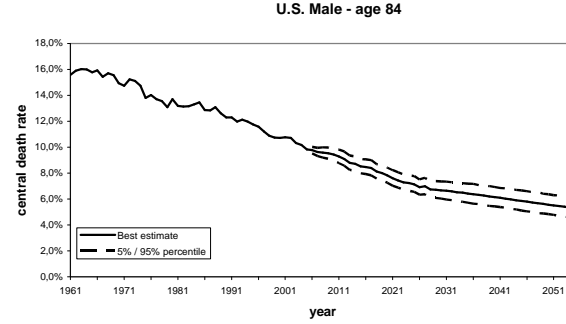**U.S. Male - age 25**

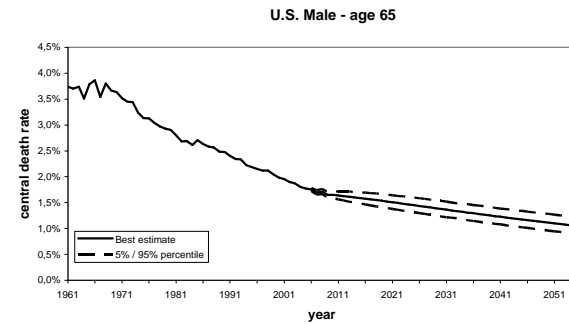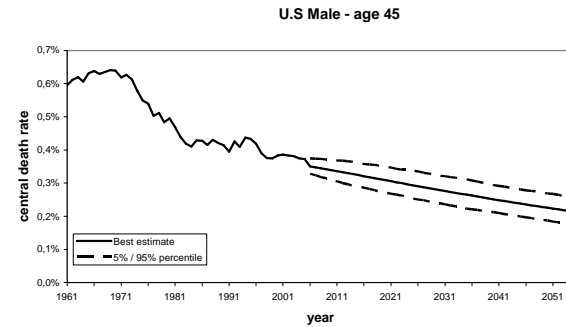**U.S Male - age 45**

U.S. Male - age 65

U.S. Male - age 84

85
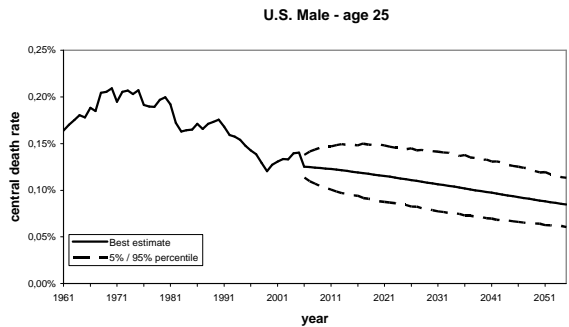
Figure 5.8: simulation results sensitivity 2)

U.S. Male - age 25

U.S Male - age 45

U.S. Male - age 65

U.S. Male - age 84
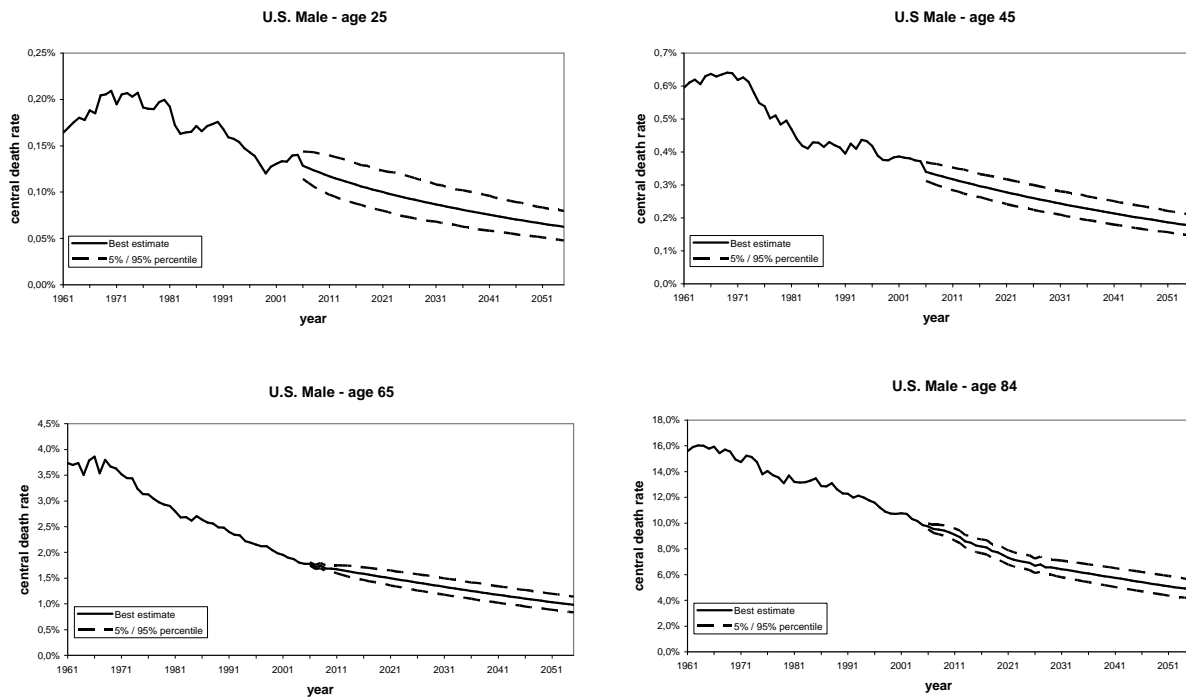
Figure 5.9: simulation results backtest
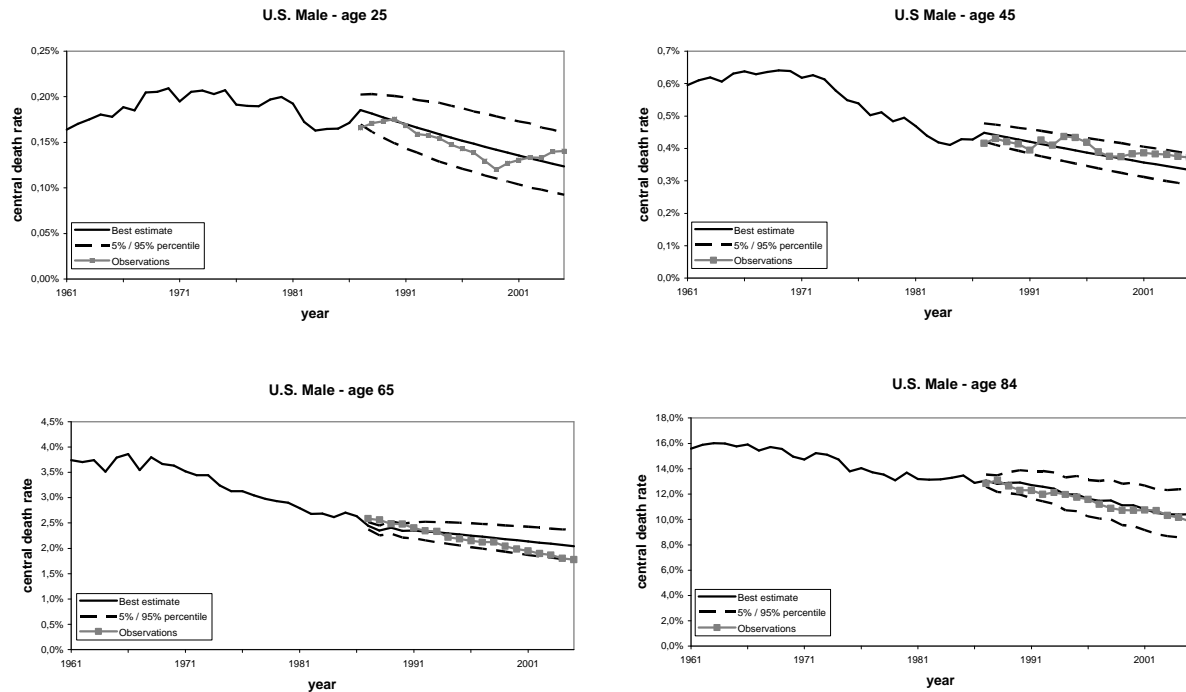
U.S. Male - age 25

U.S Male - age 45

U.S. Male - age 65

U.S. Male - age 84

Chapter 6

Stochastic Portfolio Specific Mortality and the Quantification of Mortality Basis Risk*

* This chapter has appeared as:

6.1 Introduction

As noted in chapter 5, there exists a vast literature on stochastic modeling of mortality rates. Frequently used models are for example those of Lee and Carter (1992), Brouhns et al (2002), Renshaw and Haberman (2006), Cairns et al (2006a), Currie et al (2004) and Currie (2006). These models are generally tested on a long history of mortality rates for large country populations, such as the United Kingdom or the United States. However, the ultimate application is to quantify the risks for specific insurance portfolios. And in practice there is often not enough insurance portfolio specific mortality data to fit such stochastic mortality models reliably, since:

- The historical period for which observed mortality rates for the insurance portfolio are available is usually limited, often in a range of only 5 to 15 years.
- The number of people in an insurance portfolio is much less than the country's population.

Also, for insurers it is more relevant to model mortality rates measured in insured amounts instead of measured by the number of people, because in the end the insured amounts have to be paid by the insurer. Measuring mortality rates in insured amounts has two effects:

- Policyholders with higher insured amounts tend to have lower mortality rates[19]. So measuring mortality rates in insured amounts will generally lead to lower mortality rates.

---

[19] See for example CMI (2004).

- The standard deviation of the observations will increase. For example, the risk of an insurance portfolio with 100 males with average salaries will be lower then that of a portfolio with 99 males with average salaries and 1 billionaire.

So fitting the before mentioned stochastic mortality models to the limited mortality data of insurers, measured in insured amounts, will in many cases not lead to results that are sufficiently reliable. In practice, this issue is often solved by applying a (deterministic) portfolio experience factor to projected (stochastic) mortality rates of the whole country population. However, it is reasonable to assume that this portfolio experience factor is a stochastic variable.

In this chapter a stochastic model is proposed for portfolio specific mortality experience. This stochastic process can be combined with the stochastic country population mortality process, leading to stochastic portfolio specific mortality rates. The proposed model is, amongst others, based on historical mortality rates measured in insured amounts, but can also be used when only historical mortality rates measured in number of policies are available.

The model can be used to quantify portfolio specific mortality or longevity risks for the purpose of determining the Value at Risk (VaR) or SCR, which could also be the basis for the quantification of the Market Value Margin. Also, it gives more insight into the basis risk when hedging portfolio mortality or longevity risks with hedge instruments of which the payoff depends on country population mortality. The market for mortality or longevity derivatives is emerging (see Loeys et al (2007)) and one of the characteristics of these derivatives is that the payoff depends on country population mortality. While this certainly has advantages regarding transparency and market efficiency, the impact of the basis risk is unclear. This basis risk is the result of differences between country population mortality and portfolio specific mortality, which is exactly what the proposed model is able to quantify.

Measurement of (portfolio specific) mortality rates in insured amounts has already been used for a long time, starting with CMI (1962) and more recently for example in Verbond van Verzekeraars[20] (2008) and CMI (2008). In these papers portfolio experience factors, measured in amounts, are determined based on portfolio data that is collected from a representative part of the insurance market. The results of this are frequently used by the market participants as part of an estimate of future mortality rates. Furthermore, Brouhns et al (2002) also determine deterministic portfolio experience factors for the Belgian annuity policyholders, based on 3 years of historical data.

The literature on stochastic modeling of portfolio specific experience and mortality basis risk is less developed, possibly because of a lack of historical insurance portfolio data. Van Broekhoven (2002) determines a Market Value Margin for portfolio specific mortality risk. However, the model is not set up to be easily combined with existing country population models and the structure of the model over ages is very restrictive. Since the pattern of the portfolio experience factor over ages can vary for different portfolios, there has to be enough flexibility in the assumed structure over ages.
A related paper is the one of Jarner and Kryger (2009) who set up a model for mortality in small (country) populations, using the concept of frailty. The model seems to be too complex though to

---

[20] Dutch Association of Insurers

be calibrated to the limited data of insurance portfolios. Sweeting (2007) focuses in a more qualitative way on basis risk in survivor swaps. More generally, Dahl and Møller (2006) look at hedge strategies for mortality risk in life insurance liabilities.

So the model proposed in this chapter is the first stochastic model for portfolio specific mortality that:
- can be combined easily with any stochastic country mortality process
- has enough flexibility in the assumed structure over ages
- has a structure that is simple enough to be able to calibrate it to limited historical data of life insurance portfolios

The remainder of the chapter is organized as follows. First, in Section 6.2 the general model for stochastic portfolio specific experience mortality is defined. In Section 6.3 a 1-factor version of this model is applied to two insurance portfolios. Then in Section 6.4 and 6.5 the impact on the VaR and on the hedge effectiveness is quantified. Section 6.6 gives conclusions.

6.2 General model for stochastic portfolio specific mortality experience

The first step in stochastic modeling of portfolio specific mortality rates is determining the historical portfolio mortality rates, measured by insured amounts. There are different kinds of definitions for mortality rates which are calculated in a slightly different manner (see Coughlan et al (2007)), for example the initial mortality rate or the central mortality rate. Regardless which definition is used, it is important that the same mortality rate definition is used for setting the country population mortality rates and the portfolio specific mortality rates. In the remaining part of this chapter, we use the following definition for the initial mortality rate (see for example Namboodiri and Suchindran (1987):

$$(6.1) \quad q_{x,t} = \frac{D_{x,t}}{\frac{1}{2}(N_{x,t}^{P} + N_{x,t}^{U} + D_{x,t})}$$

where $D_{x,t}$ is the number of deaths and $N_{x,t}^{P}$ and $N_{x,t}^{U}$ are the primo and ultimo total populations. The related portfolio mortality rate, measured by insured amounts, is:

$$(6.2) \quad q_{x,t}^{A} = \frac{A_{x,t}^{D}}{\frac{1}{2}(A_{x,t}^{P} + A_{x,t}^{U} + A_{x,t}^{D})}$$

where $A_{x,t}^{P}$ and $A_{x,t}^{U}$ are the insured amounts primo and ultimo for the total portfolio and $A_{x,t}^{D}$ the insured amount of the deaths, for age x and year t.

Now the aim is to define a stochastic mortality model for the so-called portfolio experience mortality factor $P_{x,t}$ for age x and year t:

$$(6.3) \quad P_{x,t} = \frac{q_{x,t}^A}{q_{x,t}^{Pop}}$$

where $q_{x,t}^{Pop}$ is the specific country population mortality rate for age x and year t, determined using (6.1). So $P_{x,t}$ represents the relation between a portfolio specific mortality rate (measured by insured amounts) and a country population mortality rate. Multiplying stochastic country mortality rates with stochastic $P_{x,t}$'s will give stochastic portfolio specific mortality rates. In this context a portfolio is seen as a group of homogenous risks, or a product group. $P_{x,t}$ is specific for each product group, it behaves differently for annuities than it does for term insurance. For reasons of convenience, the product specific nature is left out of the notation in the remaining of the chapter, but the reader should be aware that all of the following is product (group) specific.

6.2.1 The basic model

Given that the model will often be based on a limited amount of data, it is desirable that the model for $P_{x,t}$ is as parsimonious as possible. Furthermore, the conjecture is that the difference between portfolio mortality and country population mortality is expected to be less at the highest ages, since the remaining country population at the highest ages is expected to have a relatively high percentage of people that are insured and have relatively high salaries. This is corroborated by the results in CMI (2004), where the difference between portfolio mortality and country population mortality is decreasing with age. Therefore, the proposed model leads to an expectation of $P_{x,t}$ that approaches 1 for the highest ages.

Given the above, we propose to model the mortality experience factor $P_{x,t}$ as:

$$(6.4) \quad P_{x,t} = 1 + \sum_{i=1}^{n} X^i(x) b_t^i + x_{x,t}$$

where n is the number of factors of the model, $X^i(x)$ is the element for age x in the ith column of design matrix X, b_t^i is the ith element of a vector with factors for year t and $x_{x,t}$ the error term. Another way to define the model is in matrix notation:

$$(6.5) \quad P_t = \mathbf{1} + X b_t + x_t$$

where P_t is the vector of mortality experience factors, b_t the vector with factors and x_t the vector of error terms for time t. Furthermore, to ensure that $P_{x,t}$ approaches 1, we require:

$$(6.6) \quad \sum_{i=1}^{n} X_t^i(w) b_t^i = 0$$

where w is the closing age of the mortality table (usually 120).

Now given a design matrix X, the vector $\boldsymbol{b}_t$ has to be estimated for each year. The structure of X (and the corresponding $\boldsymbol{b}$'s) can be set in different ways, depending on what fits best with the data and the problem at hand. One could use for example:

a) principal components analysis to derive the preferred shape of the columns X^i.
b) a similar structure as the multi-factor model proposed by Nelson and Siegel (1987) for modeling of yield curve dynamics.
c) a more simple structure, for example using 1 factor where the vector X is a linear function in age.

a) Principal Components Analysis (PCA)
Principal components analysis is a statistical technique that linearly transforms an original set of variables into a substantially smaller set of uncorrelated variables that represents most of the information in the original set of variables. Its goal is to reduce the dimensionality of the original data set.

The (m x k) matrix P contains historical observations of $P_{x,t}$ for m years and k ages. Instead of assessing the $P_{x,t}$ process for each age individually, the goal of PCA is to derive r linear combinations (where $r < k$) that capture most of the information in the original variables:

$$
\begin{aligned}
Z_1 &= v_{11}P_1 + v_{21}P_2 + ... + v_{k1}P_k \\
Z_2 &= v_{12}P_1 + v_{22}P_2 + ... + v_{k2}P_k \\
&\ \vdots \\
Z_r &= v_{1r}P_1 + v_{2r}P_2 + ... + v_{kr}P_k
\end{aligned}
$$

(6.7)

where P_j is the vector of observations for age class j.

Or, in matrix notation:

(6.8) $Z = PV$

It can be shown (see for example Jolliffe (1986)) that the difference between the original data set and the set of linear combinations can be minimized by taking the eigenvectors of the covariance matrix $\boldsymbol{S}_{P*}$ of the de-meaned historical observation matrix $P*$ as the columns of matrix V. The corresponding eigenvalues $\boldsymbol{l}_j$ indicate the proportion of variance that each eigenvector (principal component) accounts for. By ordering the eigenvalues in such a way that $\boldsymbol{l}_1 \geq \boldsymbol{l}_2 \geq ... \geq \boldsymbol{l}_k \geq 0$, the dimensionality of the problem can be reduced by selecting the r eigenvalues (and the corresponding eigenvectors) that explain most of the variance of the original data set. The selected eigenvectors can be used as the columns X^i in (6.4).

b) Similar structure as Nelson and Siegel (1987)
Nelson and Siegel (1987) developed a parsimonious multi-factor model for yield curves that has the ability to represent shapes generally associated with yield curves. They model the instantaneous forward curve as:

(6.9) $f_t(\boldsymbol{t}) = \boldsymbol{b}_{1t} + \boldsymbol{b}_{2t}e^{-l_t t} + \boldsymbol{b}_{3t}\boldsymbol{l}_t e^{-l_t t}$

where the parameters $\boldsymbol{b}_{1t}$, $\boldsymbol{b}_{2t}$, $\boldsymbol{b}_{3t}$ and $\boldsymbol{l}_t$ have to be estimated from the observed yield curves. In practice $\boldsymbol{l}_t$ is often fixed at a pre-specified value, simplifying the estimation procedure.

We are interested in the curve of $P_{x,t}$ over the ages. As mentioned above, the observed shapes of this curve are often upward (and sometimes downward) sloping towards 1 for higher ages. So although much more erratic, these historical shapes roughly resemble possible shapes (not levels) of yield curves. Therefore a structure similar as (6.9) could be used for modeling the $P_{x,t}$'s. An example of a possible 2-factor structure is given in appendix 6a.

c) A more simple structure
An alternative for structure a) and b) is a more simple structure, for example one where it is assumed that $P_{x,t}$ is linear in age for each t. It depends on the size of the insurance portfolio and the historical period whether structure a) leads to usable results and structure b) leads to a better fit to the data than this simple structure. For very large portfolios, structure a) and b) could be the most appropriate solutions. However, for the insurance portfolios considered in this chapter, with 14 years of history and respectively about 100.000 policies and about 45.000 policies, principle components analysis did not lead to usable results, and structure b) did not fit the data better than a simple linear structure (see section 6.3).

6.2.2 Fitting the basic model
The structure of the model is such that it could be fitted with Ordinary Least Squares (OLS). However, the observations $P_{x,t}$ are all based on different exposures to death and observed deaths, so there is generally significant heteroscedasticity. Therefore Generalized Least Squares (GLS) should be used (Verbeek (2008)). When applying GLS in case of heteroscedasticity, each observation is weighted by (a factor proportional to) the inverse of the error standard deviation. Fitting this transformed model with OLS gives the GLS estimator, which accounts for the heteroscedasticity in the data.

When the available data are a cross-section of group averages with different group sizes and the observations are homoscedastic at individual level, the variance of the error term of the group averages is inversely related to the number of observations per group. In that case the square root of the number of observations in the group can be used as weights (Verbeek (2008)). For the problem in this chapter this means that the square root of the number of deaths can be used as weights. So using a diagonal weight matrix W_t with these weights and applying it to (6.5) leads to a transformed model:

(6.10) $W_t\left(P_t - \mathbf{1}\right) = W_t X \boldsymbol{b}_t + W_t \boldsymbol{x}_t \quad or \quad \left(P_t - \mathbf{1}\right)^* = X^* \boldsymbol{b}_t + \boldsymbol{x}_t^*$

Where the vectors or matrices labeled with an * are weighted with W_t. Now applying OLS to (6.10) gives the GLS estimator for $\boldsymbol{b}_t$:

$$(6.11) \quad \hat{\boldsymbol{b}}_t = \left(X^{*\prime} X^* \right)^{-1} X^{*\prime} \left(P_t - \mathbf{1} \right)^* = \left(X' W_t' W_t X \right)^{-1} X' W_t' W_t \left(P_t - \mathbf{1} \right)$$

This procedure can be repeated for each historical observation year, leading to a time series of vector $\hat{\boldsymbol{b}}_t$.

6.2.3 Adding stochastic behavior

Now using the time series of the fitted $\boldsymbol{b}_t$'s, a Box-Jenkins analysis can be performed to determine which stochastic process fits the historical data best [21]. However, an important requirement in this case is biological reasonableness. For example, when assuming a non-stationary process such as a Random Walk for the $\boldsymbol{b}_t$'s, in certain scenarios the $P_{x,t}$'s could be 0 for all ages for some time, which is not biologically reasonable. Since the difference between country population mortality and portfolio mortality is dependent on factors that in our experience are normally relatively stable (size, composition and relative welfare of the portfolio), it does not seem reasonable to assume that this difference can increase without limit. Therefore, a stationary process seems the most appropriate in this case. Given the often limited historical period of observations and the requirement of parsimoniousness, in most cases the most appropriate model will then be a set of correlated first order autoregressive ($AR(1)$) processes or equivalently, a restricted first order Vector Autoregressive (VAR) model:

$$(6.12) \quad \boldsymbol{b}_t = \boldsymbol{d} + \Theta_1 \boldsymbol{b}_{t-1} + \boldsymbol{e}_t$$

where $\boldsymbol{Q}_1$ is a $n \times n$ diagonal matrix, $\boldsymbol{d}$ is a n-dimensional vector and $\boldsymbol{e}_t$ is a n-dimensional vector of white noise processes with covariance matrix $\boldsymbol{S}$.

Possible alternatives are an unrestricted $VAR(1)$ model or a first order (restricted) Vector Moving Average (VMA) model. In some cases an even simpler process than (6.12) is possible, being the so-called $ARIMA(0,0,0)$[22] process:

$$(6.13) \quad \boldsymbol{b}_t = \boldsymbol{d} + \boldsymbol{e}_t$$

Model (6.12) and (6.13) can be fitted using OLS equation by equation. From the residuals e of the n equations the elements (i,j) of $\boldsymbol{S}$ can be estimated as:

$$(6.14) \quad \hat{\boldsymbol{s}}_{ij} = \frac{1}{(T-K)} \sum_{t=1}^{T} e_{it} e_{jt}$$

where K is the maximum number of parameters used in either equations i or j (that is 2 when both processes are $AR(1)$ processes).

---

[21] This is possible under the assumption that the historical fitted parameters are certain. Another possible approach would be to fit the parameters and the stochastic process at once, for example using a state space method combined with the Kalman filtering technique.

[22] Note that various names are used for this process in literature. Since the name ARIMA(0,0,0) seems to be the most widely used, we have adopted this name in this chapter.

An alternative is to estimate this simultaneously with the stochastic processes of the country population mortality model, which is the subject of the next paragraph.

When the insurance portfolio has developed significantly over the years, the fitted parameters over time are subject to heteroscedasticity. In this case GLS could be used, using either the results from table 6.5 in appendix 6b or the square root of the number of deaths (see paragraph 6.2.2) as weights. When the portfolio has grown significantly and the current size of the portfolio is believed to be more representative for the future, the *relative* weights can also be applied to the residuals, weighting the earlier residuals less than the more recent ones.

6.2.4 Combine the process with the stochastic country population model

To end up with a stochastic process for portfolio specific mortality rates, the correlation between country population mortality rates and the portfolio mortality experience factors has to be taken into account. Therefore, the processes of the drivers of these have to be estimated simultaneously. Let us assume that the country population mortality is driven by m factors of which the processes $\boldsymbol{a}_t$ can be written as:

$$(6.15) \quad \boldsymbol{a}_t^k = X_k^a \boldsymbol{h}_k^a + \boldsymbol{e}_k^a \qquad k = 1, ..., m$$

Now when the historical observation period is equal for the country mortality rates and the portfolio mortality experience factors, Seemingly Unrelated Regression (SUR, see Zellner (1963)) can be applied to fit all processes simultaneously. The processes do not have to be similar, so $AR(1)$, Random Walk or other $ARIMA$ models can be combined.

Re-writing (6.12) for each element i in a more general form as $\boldsymbol{b}_t^i = X_i^b \boldsymbol{h}_i^b + \boldsymbol{e}_i^b$ and combining all processes gives:

$$(6.16) \quad \begin{bmatrix} \boldsymbol{b}^1 \\ \vdots \\ \boldsymbol{b}^n \\ \boldsymbol{a}^1 \\ \vdots \\ \boldsymbol{a}^m \end{bmatrix} = \begin{bmatrix} X_1^b & 0 & \cdots & \cdots & \cdots & 0 \\ 0 & \ddots & & & & 0 \\ \vdots & 0 & X_n^b & & & \vdots \\ \vdots & \vdots & & X_1^a & & \vdots \\ \vdots & \vdots & & & \ddots & 0 \\ 0 & 0 & \cdots & \cdots & 0 & X_m^a \end{bmatrix} \begin{bmatrix} \boldsymbol{h}_1^b \\ \vdots \\ \boldsymbol{h}_n^b \\ \boldsymbol{h}_1^a \\ \vdots \\ \boldsymbol{h}_m^a \end{bmatrix} + \begin{bmatrix} \boldsymbol{e}_1 \\ \vdots \\ \vdots \\ \vdots \\ \vdots \\ \boldsymbol{e}_{n+m} \end{bmatrix}$$

which can be written more compactly as:

$$(6.17) \quad Y = X^{a,b} \boldsymbol{h}^{a,b} + \boldsymbol{e}^{a,b}$$

Now these processes can be fitted with SUR using the following steps:

1) Fit equation by equation using OLS
2) Use the residuals to estimate the total covariance matrix $\hat{\Omega}$ with (6.14)

3) Estimate $\hbar$ using GLS

To be more specific, the resulting estimator in step 3) is determined as:

$$(6.18) \quad \hbar = \left(X^{a,b\prime} \hat{\Omega}^{-1} X^{a,b} \right)^{-1} \left(X^{a,b\prime} \hat{\Omega}^{-1} Y \right)$$

As mentioned earlier, in most cases the historical data period for portfolio mortality will be shorter than of country population mortality. In this case an alternative is only to do steps 1) and 2). In step 1) all available historical observations can be used for the different processes. In step 2) for the country population mortality the same historical data period should be used as is available for the portfolio mortality.

6.3 Application to example insurance portfolios

As mentioned in section 6.2, $P_{x,t}$ is specific for each product group or portfolio of homogeneous risks. In this section the general model described in section 6.2 is applied to two insurance portfolios[23]. The portfolios are respectively large and medium sized, and only data for males from age 65 on is taken into account. The large portfolio is a collection of collective pension portfolios of the Dutch insurers and contains about 100.000 male policyholders aged 65 or older. The medium portfolio is an annuity portfolio with about 45.000 male policyholders aged 65 or older. Note that this medium portfolio has developed significantly over time, so had fewer policyholders in earlier years. For both portfolios 14 years of historical mortality data is available.

Due to the relatively short historical period an the erratic pattern over the ages of the observed $P_{x,t}$'s, principal components analysis does not give usable or interpretable results for these portfolios. To be specific, the resulting shapes of the columns X^i are very erratic and do not have a clear interpretation (such as for example level, slope or curvature).

For both portfolios, we examined a collection of 1-, 2-, and 3-factor models and concluded that the 2- and 3- factor models did not fit the data much better than a 1-factor model[24]. Since the 1-factor model uses fewer parameters, the Bayesian Information Criterion (BIC)[25] is more favourable for this structure. Therefore, the model we use is model (6.4) with $n = 1$ and:

$$(6.19) \quad X^1(x) = 1 - \frac{x-d}{w-d} \qquad\qquad d \leq x \leq w$$

where d is the start age (in this case 65) and w is the end age (120). So in this formulation of model (6.4), the vector X is a linear function in age and, as required, $X^1(w) = 0$.

[23] The author thanks the Centrum voor Verzekeringsstatistiek (CVS) and Erik Tornij for the data of the large portfolio, and Femke Nawijn and Christel Donkers for the data of the medium portfolio.
[24] The fitting results for the 2-factor and 3-factor models are available upon request.
[25] BIC is a criterion that provides a trade-off between goodness-of-fit and the parsimony of the model.

The reason why the 1-factor model fits the data as well as 2- or 3-factor models is that the data shows an upward slope for increasing ages, but the pattern along the ages is very volatile. For example, figure 6.1 shows two fits for the years 2006 and 2000. Fitting a more complex model through this data will not reduce the residuals significantly. Of course, this observation depends on the characteristics of the specific portfolio to which the model is fitted. For larger portfolios a 2- or 3-factor could give better results, since such a model is able to capture more shapes of the portfolio experience mortality factor curve.

Figure 6.1: example fit of model to actual observations for years 2006 and 2000
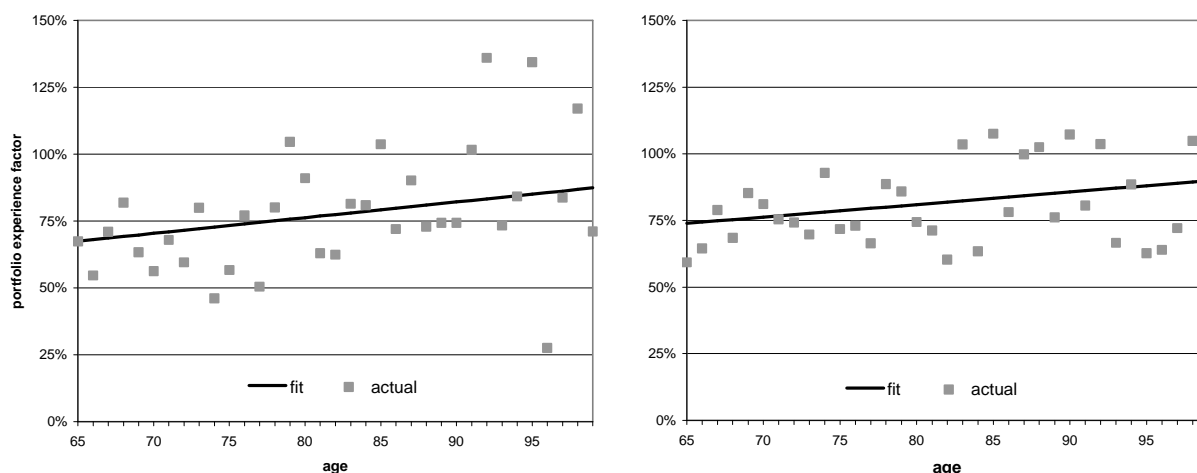
The model is fitted using the procedure described in paragraph 6.2.2, where we have used the square root of the number of deaths as weights. The fitted b's are shown in figure 6.2[26]. Further results are given in table 6.5 in appendix 6b.

Figure 6.2: fitted b's for historical years 1993-2007

---
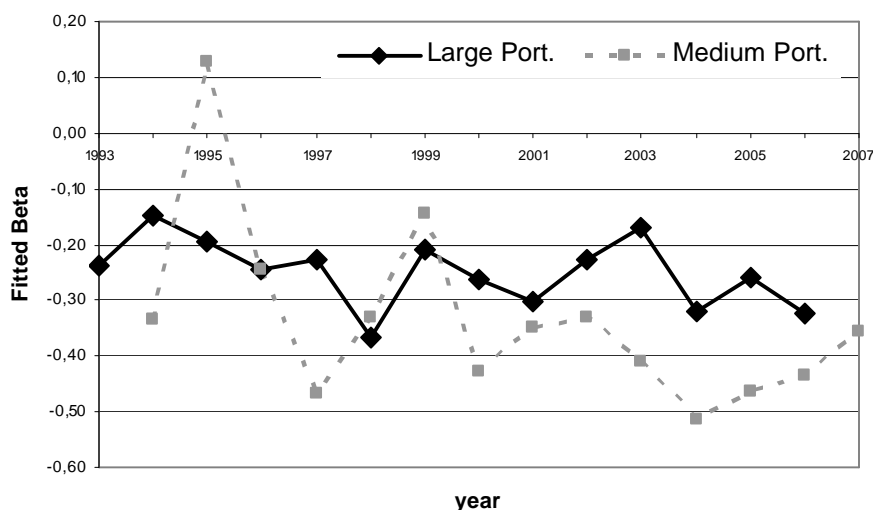
[26] Note that although we have 14 years of data for both portfolios, the periods are slightly different, having data from 1993-2006 for the large portfolio and from 1994-2007 for the medium portfolio.

For both portfolios the results show an autoregressive pattern for the **b**'s. Now a stochastic process for the future **b**'s has to be selected. As mentioned in paragraph 6.2.3, a stationary process will be most appropriate. Also, since the historical data period is limited, the model should be as parsimonious as possible. We have fitted an *AR*(1), *AR*(2) and *ARIMA*(0,0,0) process to the data shown in figure 2. For both portfolios the *ARIMA*(0,0,0) process led to a more favorable *BIC* compared to the other processes.

Because of the significant development of the medium sized portfolio over the historical years, GLS is used for fitting the *ARIMA*(0,0,0) process. The square roots of the *relative* number of deaths in a year are used as weights. Relative means relative to the average number of deaths. These weights are also applied to the residuals, giving less weight to years where the portfolio was relatively small. Since the large portfolio was relatively stable over time, OLS is used for fitting the *ARIMA*(0,0,0) process for this portfolio.

The fitted processes for the portfolios are:

(6.20) Large portfolio: $\boldsymbol{b}_t = -0,2497 + \boldsymbol{e}_t$, $\hat{\boldsymbol{s}} = 0,0625$

(6.21) Medium portfolio: $\boldsymbol{b}_t = -0,3798 + \boldsymbol{e}_t$, $\hat{\boldsymbol{s}} = 0,1130$

The estimated error standard deviation $\hat{\boldsymbol{s}}$ is significantly larger for the medium sized portfolio, which is mainly the result of having fewer policyholders. The result of this is shown in figure 6.3, where the best estimates and the 99,5% / 0,5% percentiles are given for the portfolio experience mortality factors in the year 2016[27]. These specific percentiles are shown because the SCR of Solvency 2 is based on a 99,5% percentile.

Figure 6.3: best estimates and 99,5% / 0,5% percentiles for both portfolios - 2016
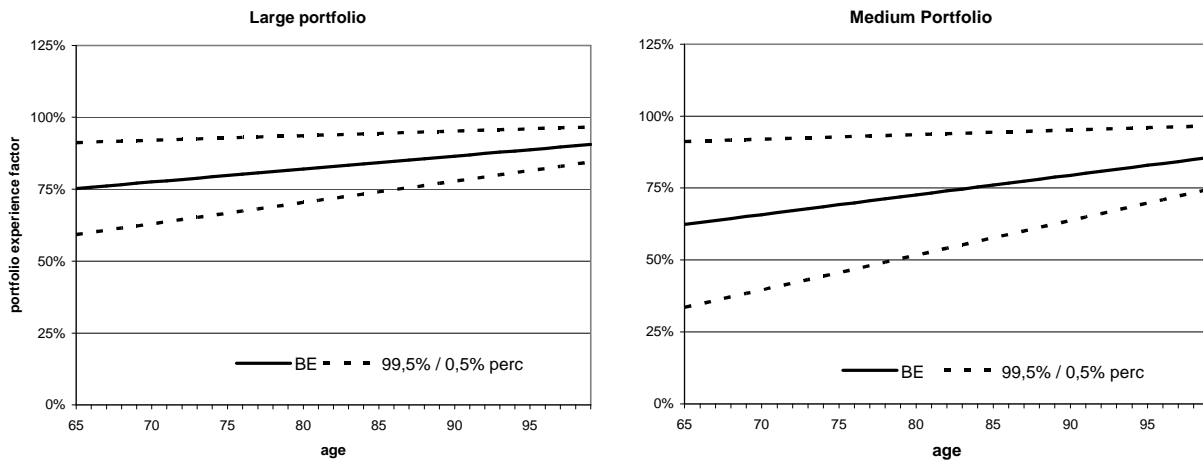
The figure shows that for the large portfolio the difference between the best estimate and the percentile(s) is in the range 10-15 %-point for ages 65-80. So taking this stochastic behaviour of the portfolio experience mortality factor into account can have a reasonable impact on for

---

[27] Since a stationary process is assumed, the figure will be similar for other projection years.

example the Value at Risk. As expected, the impact is larger for the medium portfolio, where the difference between the best estimate and the percentile(s) is almost 30 %-point at its maximum.

6.4 Numerical example 1: Value at Risk

An important application of the presented model is the quantification of the Value at Risk (VaR) or SCR for longevity or mortality risk. In this paragraph the VaR is determined for the two portfolios, for different definitions / horizons of the VaR. First the model has to be combined with a model for country population mortality risk.

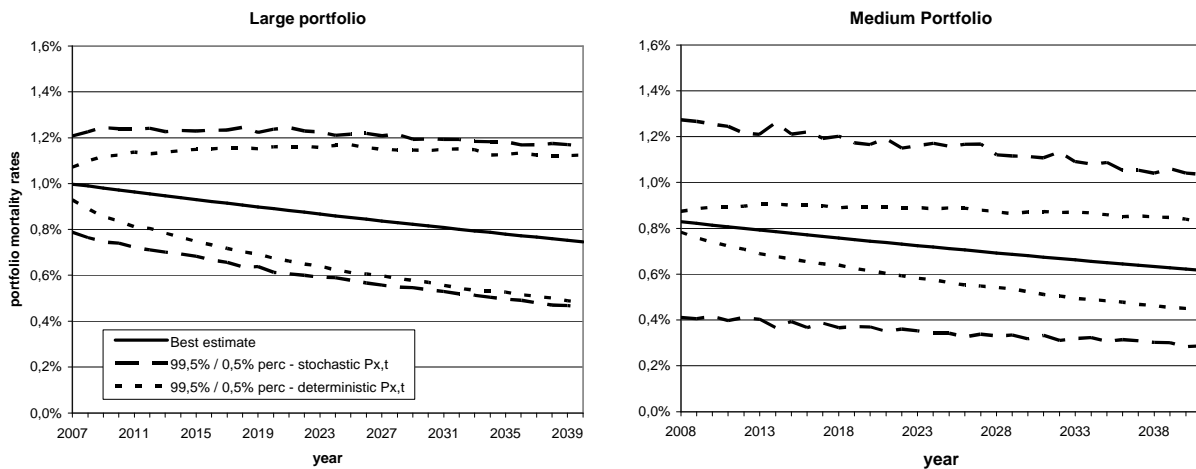### *6.4.1 Stochastic country population mortality model*
For the stochastic country population model we use the model of Cairns et al (2006a):

$$(6.22) \quad \text{logit } q_{x,t}^{Pop} = \boldsymbol{k}_t^1 + \boldsymbol{k}_t^2 \left(x - \overline{x} \right)$$

Where $\overline{x}$ is the mean age in the sample range and $\boldsymbol{k}_t^1$ and $\boldsymbol{k}_t^2$ the two (stochastic) factors. We fitted this model to data of the Dutch population for the years 1950 – 2007. Using the resulting time series of parameter estimates, a 2-dimensional random walk process is fitted for the factors. The fitted parameters and the covariance matrices, including the covariances with the portfolio experience mortality process of both portfolios, are given in appendix 6b.

Now combining the stochastic process above and the process described in section 6.3 leads to stochastic portfolio specific mortality rates. Figure 6.4 gives the best estimate mortality rates and percentiles for age 65. The percentiles are based on respectively deterministic and stochastic $P_{x,t}$'s.

Figure 6.4: best estimates and percentiles, with stochastic or deterministic $P_{x,t}$


The figure shows that the additional risk of including stochastic $P_{x,t}$'s is highest at the start of the projection and decreases slowly in time. The reason for this is that the country population

mortality rate risk is gradually increasing over time, resulting in a higher diversification effect between country population mortality rates and the $P_{x,t}$'s over time.

The percentiles for the medium portfolio seem quite dramatic. However, note that the shown percentiles are a result of picking the particular percentile every year, and not picking 1 scenario that represents the *x*%-percentile for the whole projection. Because of the assumed *ARIMA*(0,0,0) process the extremely low outliers will normally be (partially) compensated somewhere in time by high outliers. This is shown in figure 6.5, where two random (simulated) scenarios of the **b**'s are given as an example.

Figure 6.5: two random (simulated) scenarios for **b** *- medium portfolio*
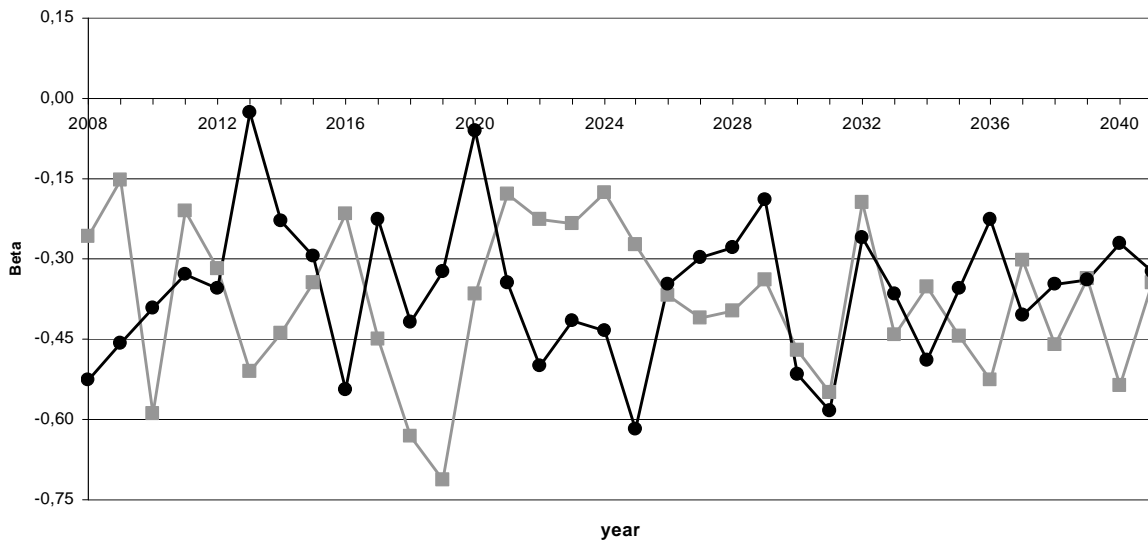
6.4.2 Impact on Value at Risk

Now using the described stochastic processes the impact on the VaR of stochastic (instead of deterministic) $P_{x,t}$'s is determined for both portfolios. The (present) value of liabilities is calculated for all simulated mortality rate scenarios[28]. The VaR is then defined as the difference between the *x*%-percentile and the average value of the liabilities. The impact is determined for three different definitions / horizons, which are all being used in practice:

1) 1-year horizon, 99,5% percentile, including effect on best estimate after 1 year
2) 10-year horizon, 95% percentile, including effect on best estimate after 10 years
3) Run-off of the liabilities, 90% percentile

So for definitions 1) and 2), at the 1-year or 10-year horizon all parameters are re-estimated using the (simulated) observations in the first 1 or 10 years, for each simulated scenario. The impact of the new parameterization on the best estimate of liabilities (for each scenario) is taken into account in the VaR. The results for the large and medium portfolio are given in respectively table 6.1 and table 6.2.

---

[28] For convenience we assumed that the portfolios only contain pension or annuity payments, so no spouse pension or annuities on a second life.

Table 6.1: impact of stochastic $P_{x,t}$ on VaR – large portfolio (in millions of Euros)

VAR definition	Deterministic $P_{x,t}$	Stochastic $P_{x,t}$	% difference
1-year, 99,5%	126,4	138,2	+ 9,3%
10-year, 95%	182,3	194,3	+ 6,6%
Run off, 90%	136,5	145,9	+ 6,8%

Table 6.2: impact of stochastic $P_{x,t}$ on VaR – medium portfolio (in millions of Euros)

VAR definition	Deterministic $P_{x,t}$	Stochastic $P_{x,t}$	% difference
1-year, 99,5%	45,1	73,0	+ 61,8%
10-year, 95%	69,2	95,1	+ 37,4%
Run off, 90%	54,1	75,1	+ 38,8%

Table 6.1 shows that for the large portfolio stochastic $P_{x,t}$'s lead to a VaR that is about 7%-9% higher compared to the VaR calculated with deterministic $P_{x,t}$'s. Table 6.2 shows that the impact for the medium portfolio is very high. The increase in VaR is between 37% and 68%, depending on the definition for VaR used. The reason for this is the large increase in volatility due to the addition of the stochastic $P_{x,t}$'s, which is mainly related to the size of the portfolio. Since a large part of the insurance portfolios in practice are of this size or smaller, this should be a point of attention when developing or reviewing internal models for mortality and longevity.

6.5 Numerical example 2: hedge effectiveness / basis risk

Because of the increasing external requirements and focus on risk measurement and risk management, the interest in hedging mortality or longevity risk is also increasing. A result of this is that a market for mortality and longevity derivatives is emerging (see Loeys et al (2007)). One of the main characteristics of these derivatives is that the payoff depends on country population mortality. While this certainly has advantages regarding transparency and market efficiency, the impact of the basis risk is unclear. Basis risk is the risk arising from a difference between the underlying of the derivative and the actual risk in the liability portfolio. The model presented in this chapter can be used to quantify this basis risk. In the example below the basis risk will be quantified for the two portfolios, where the longevity risk is (partly) hedged with the so-called q-forwards.

A *q*-forward is a simple capital market instrument with similar characteristics as an interest rate swap. The instrument exchanges a realized mortality rate in a future period for a pre-agreed fixed mortality rate. This is shown in figure 6.6. The pre-agreed fixed mortality rate is based on a projection of mortality rates, using a freely available and well documented projection tool[29].

---

[29] For more information, see http://www.jpmorgan.com/pages/jpmorgan/investbk/solutions/lifemetrics

Figure 6.6: mechanics of a q-forward
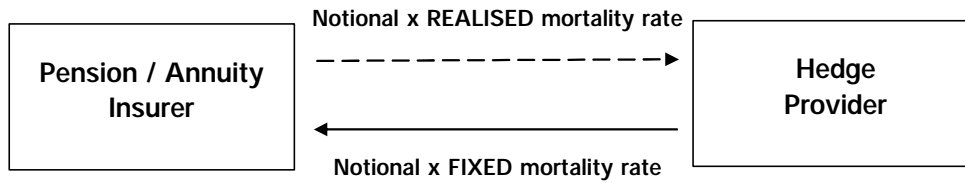
For example, when the realized mortality rate is lower than expected, the pension / annuity insurer will receive a payment which (partly) compensates for the increase of the expected value of the insurance liabilities (caused by the decreasing mortality rates).

The basis for the instrument is the (projected) mortality of a country population, not the mortality of a specific company or portfolio. This makes the product and the pricing very transparent compared to traditional reinsurance.

For both insurance portfolios we determined a minimum variance hedge, based on deterministic $P_{x,t}$'s. The hedge is determined for a horizon of 10 years, but including the effect on the best estimate after 10 years (conform definition 2 of VaR in paragraph 6.4.2). The hedge is determined for age-buckets of 5 years. For every bucket i, the impact of small shocks of the two factors of the country population model on the value of the liabilities and the value of an appropriate q-forward contract are calculated. The required nominal a_i^* for the q-forward of bucket i is then determined as:

$$(6.23) \quad a_i^* = \frac{l_1 h_1 + l_2 h_2}{h_1^2 + h_2^2}$$

where l_i and h_i are the impact of the shock of the ith factor on respectively the liabilities (l) and the hedge instrument (h). This expression is obtained by solving the required nominal from the equation that results when minimising the variance of the hedge result.

The resulting hedge portfolio consists of 5 q-forwards for age-buckets of 5, from age 65 until age 89. The payoff of such a q-forward depends on the average mortality rate for the 5 ages in the bucket. The exact composition of both the hedge portfolios is given in appendix 6c.

Tables 6.3 and 6.4 show the impact on the hedge effectiveness when the $P_{x,t}$'s are assumed to follow the stochastic process described in section 6.3.

Table 6.3: impact of stochastic $P_{x,t}$ on hedge effectiveness – large portfolio

	VAR unhedged	VAR hedged	% reduction
Deterministic $P_{x,t}$	182,3	64,0	64,9%
Stochastic $P_{x,t}$	194,3	81,9	57,8%

Table 6.4: impact of stochastic P*_{x,t} *on hedge effectiveness – medium portfolio

	VAR unhedged	VAR hedged	% reduction
Deterministic $P_{x,t}$	69,2	23,6	65,8%
Stochastic $P_{x,t}$	95,1	47,9	49,7%

The tables show that given deterministic $P_{x,t}$'s, the hedge reduces the VaR with about 65%. The risk is not fully hedged, because the hedge is based on small shocks of the two country population factors, while the factors in the tails of the distributions (which are relevant for VaR) are often more extreme.

For the large portfolio, table 6.3 shows that the hedge quality is decreasing, but is still reasonable. The basis risk for this portfolio is therefore limited. The reason for this is that on a longer horizon the impact of stochastic $P_{x,t}$'s levels out because of the assumed autoregressive process.

For the medium portfolio the hedge effectiveness is reduced to a larger extent. The effectiveness of the hedge can be improved by periodically adjusting the hedge portfolio. For smaller portfolios than this, it is probably questionable whether it is sensible to set up such hedge constructions.

6.6 Conclusions

In this chapter a stochastic model is proposed for stochastic portfolio experience. Adding this stochastic process to a stochastic country population mortality model leads to stochastic portfolio specific mortality rates, measured in insured amounts. The proposed stochastic process is applied to two insurance portfolios. The results show that the uncertainty for the portfolio experience factor $P_{x,t}$ can be significant, mostly depending on the size of the portfolio.

The impact of the VaR for longevity risk is quantified. Depending on the definition used, the VaR increases by about 7%-9% for the large portfolio. The impact for the medium portfolio is very high, with an increase in VaR of 37%-68%. The reason for this is the high increase in volatility due to the addition of the stochastic $P_{x,t}$'s. Since a large part of the insurance portfolios in practice are of this size or smaller, this should be a point of attention when developing or reviewing internal models for mortality and longevity.

Furthermore, the basis risk is quantified when hedging portfolio specific mortality risk with q-forwards, of which the payoff depends on country population mortality rates. For the large portfolio the hedge quality is decreasing, but is still reasonable. The reason for this is that on a longer horizon the impact of stochastic $P_{x,t}$'s levels out because of the assumed autoregressive process. For the medium portfolio hedge effectiveness is reduced to a larger extent. For smaller portfolios than this, it is probably questionable whether it is sensible to set up such hedge constructions.

Appendix 6a: example 2-factor model based on Nelson & Siegel

Nelson and Siegel (1987) proposed a parsimonious model for yield curves, which allows for different shapes of the curve. The Nelson-Siegel forward curve can be viewed as a constant plus a Laguerre function, which is a polynomial times an exponential decay term. It has three elements, respectively for the short, medium and long term. The model is very often used for yield curves and could serve as a basis for thinking for the P_t curves that are the subject of this chapter. However, the Nelson-Siegel curve cannot directly be used for the P_t curves because $P_{x,t}$ should approach 1 near the closing age. Also, another requirement mentioned in section 6.2 is that the model is as parsimonious as possible, so a 2-factor model might be more appropriate in most cases.

Many variations on the Nelson-Siegel curve are possible. An example of such a model is the following model:

$$(6.24) \quad P_t(t) = 1 + \boldsymbol{b}_{1t}\, e^{-l_1 t} + \boldsymbol{b}_{2t} w_{\boldsymbol{t}} \left(e^{-l_1 t} - e^{-l_2 t} \right)$$

$$where\ w_{\boldsymbol{t}} = \boldsymbol{j}\left(\boldsymbol{a}\left[\frac{\boldsymbol{t} - \boldsymbol{t}_m}{\boldsymbol{t}_m} \right] \right) \boldsymbol{f}$$

The variable $\boldsymbol{t}$ is 0 for the starting age of the data (in this case 65 years), $\boldsymbol{t}_m$ is a strategically set middle point of the age interval (in this case 20, representing age 85), $\boldsymbol{j}$ is the density of a standard normal distributed variable, $\boldsymbol{a}$ is a variable that arranges the shape of $w_{\boldsymbol{t}}$ and can be set at 2 for example, and $\boldsymbol{f}$ is a scale variable. The variable $\boldsymbol{l}_1$ can be solved in such a way that the second term of (6.24) approaches 0 for the closing age. The variable $\boldsymbol{l}_2$ can be solved in such a way that the third term of (6.24) is at its maximum somewhere between $\boldsymbol{t} = 0$ and $\boldsymbol{t}_m$ (in this case 75 years). The factors are shown in figure 6.7, where x_1 represents the second term and x_2 the third term of (6.24).

As can be seen from the figure and (6.24), the curve starts at age 65 at $1 + \boldsymbol{b}_{1t}$ (where $\boldsymbol{b}_{1t}$ will be negative in general) and ends at 1 at higher ages. With the model (6.24) different shapes of the curve can be fitted, and the requirements in section 6.2 are met. A disadvantage of the model is the large number of parameters, of which some are set more or less arbitrarily.

Figure 6.7: factors for model (A.1)
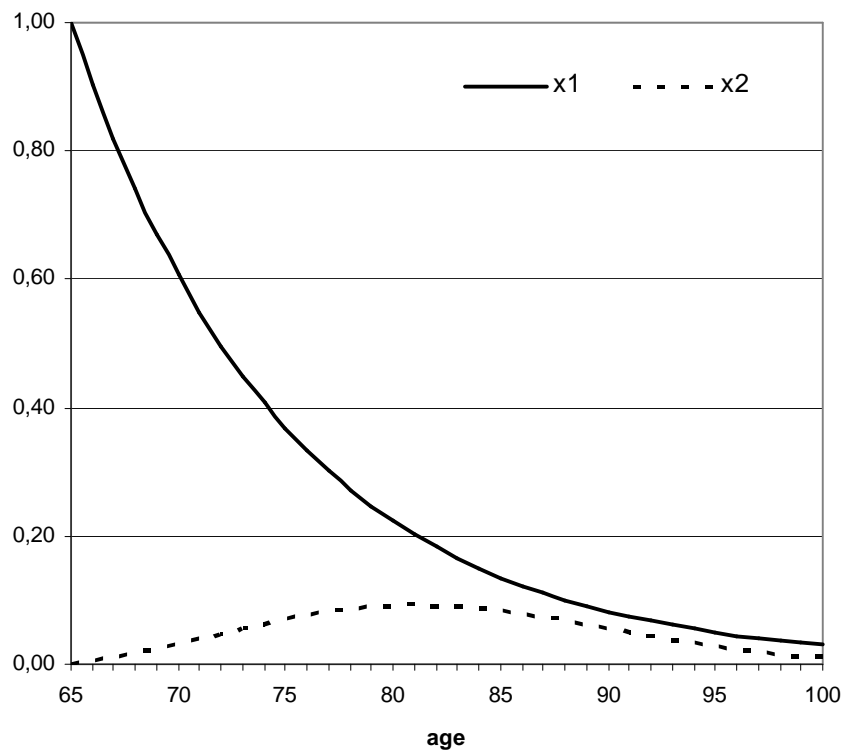
Appendix 6b: further results

Table 6.5 shows the fitting results for the **b**'s in each year, for the large and medium sized portfolio.

*Table 6.5: yearly fitting results for **b**'s*

Results large portfolio					Results medium portfolio			
Year	**b**	s.e.	t-ratio		Year	**b**	s.e.	t-ratio
1993	-0,239	0,036	-6,55		1994	-0,333	0,103	-3,23
1994	-0,149	0,041	-3,67		1995	0,127	0,201	0,63
1995	-0,194	0,030	-6,55		1996	-0,243	0,127	-1,92
1996	-0,246	0,033	-7,43		1997	-0,467	0,091	-5,14
1997	-0,228	0,032	-7,20		1998	-0,330	0,056	-5,92
1998	-0,368	0,023	-16,12		1999	-0,143	0,065	-2,21
1999	-0,208	0,036	-5,77		2000	-0,427	0,057	-7,56
2000	-0,261	0,029	-8,91		2001	-0,349	0,089	-3,94
2001	-0,304	0,032	-9,46		2002	-0,331	0,052	-6,32
2002	-0,226	0,033	-6,88		2003	-0,408	0,050	-8,11
2003	-0,168	0,046	-3,62		2004	-0,515	0,035	-14,77
2004	-0,321	0,048	-6,71		2005	-0,462	0,047	-9,81
2005	-0,259	0,042	-6,11		2006	-0,434	0,046	-9,52
2006	-0,325	0,040	-8,18		2007	-0,355	0,079	-4,52

Table 6.6 shows the fitted parameters for the 2-dimensional random walk model of section 6.4, and the covariance matrix including the covariances with the process of section 6.3. Note that the country population parameter estimates slightly differ for the large and medium portfolio, because for the medium portfolio the year 2007 is also taken into account.

Table 6.6: fit of country population model and covariance matrices

Fit - large portfolio		
	n	**s**
k^1	-0,006206	0,0312395
k^2	0,000182	0,00140485

Fit - medium portfolio		
	n	**s**
k^1	-0,006722	0,022663
k^2	0,000175	0,001436

Covariance matrix - large portfolio			
	b	k^1	k^2
b	0,003911	0,000445	0,000041
k^1	0,000445	0,000976	0,000022
k^2	0,000041	0,000022	0,000002

Covariance matrix - large portfolio			
	b	k^1	k^2
b	0,013808	0,002174	0,000053
k^1	0,002174	0,000514	0,000020
k^2	0,000053	0,000020	0,000002

Appendix 6c: hedge portfolios

Table 6.7: hedge portfolios for large and medium insurance portfolio

Characteristics hedge portfolio - large portfolio				
q-forward	Start age	End age	Nominal	Tick Size
1	65	69	117.865.528	100
2	70	74	34.221.141	100
3	75	79	10.420.145	100
4	80	84	2.213.481	100
5	85	89	315.640	100

Characteristics hedge portfolio - medium portfolio				
q-forward	Start age	End age	Nominal	Tick Size
1	65	69	74.346.033	100
2	70	74	24.226.106	100
3	75	79	3.203.979	100
4	80	84	143.135	100
5	85	89	8.224	100

Chapter 7

# Micro-Level Stochastic Loss Reserving*

* This chapter is based on:

ANTONIO, K. AND R. PLAT (2010): Micro-Level Stochastic Loss Reserving, *Working Paper*

7.1 Introduction

In this chapter a micro-level stochastic model for the run-off of general insurance[30] claims is developed. Figure 7.1 illustrates the run-off (or development) process of a general insurance claim. It shows that a claim occurs at a certain point in time (t_1), consequently it is declared to the insurer (t_2) (possibly after a period of delay) and one or several payments follow until the settlement (or closing) of the claim. Depending on the nature of the business and the claim, the claim can re-open and payments can follow until the claim finally settles.

Figure 7.1: run-off process of an individual general insurance claim

---
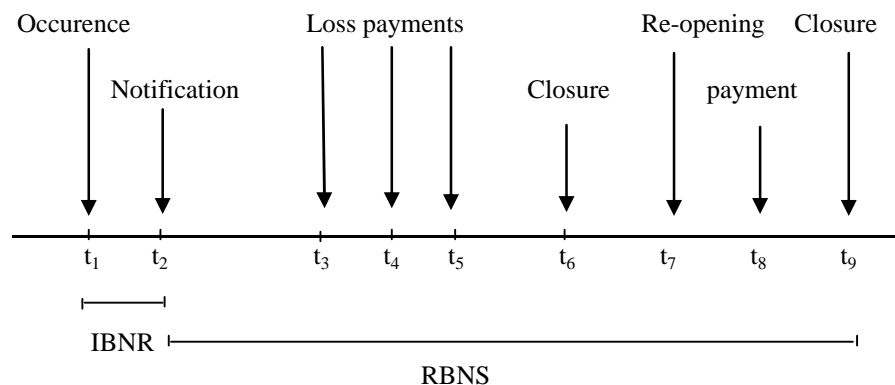
[30] General insurance is also often referred to as 'Non-Life' or 'Property and Casualty' insurance.

At the present moment (say t) the insurer needs to put reserves aside to fulfill its liabilities in the future. This actuarial exercise will be denoted as 'loss-' or 'claims reserving'. Insurers, shareholders, regulators and tax authorities are interested in a rigorous picture of the distribution of future payments corresponding with the open (i.e. not settled) claims in a loss reserving exercise. General insurers distinguish between RBNS and IBNR reserves. 'RBNS' claims are claims that are Reported to the insurer But Not Settled, whereas 'IBNR' claims Incurred But are Not Reported to the company. For an RBNS claim occurrence and declaration take place before the present moment and settlement occurs afterwards (i.e. $t \geq t_2$ and $t < t_6$ (or $\boldsymbol{t} < t_9$) in figure 7.1). An IBNR claim has occurred before the present moment, but its declaration and settlement follow afterwards (i.e. $t \in [t_1, t_2)$ in figure 7.1). The interval $[t_1, t_2]$ represents the so-called reporting delay. The interval $[t_2, t_6]$ (or $[t_2, t_9]$) is often referred to as the settlement delay. Data bases within general insurers typically contain detailed information about the run-off process of historical and current claims. The structure in figure 7.1 is generic for the kind of information that is available. In the remaining of this chapter we will use the label 'micro-level' data to denote this sort of data structures.

The measurement of future cash flows and its uncertainty becomes more and more important. That also gives rise to the question whether the currently used techniques can be improved. In this chapter we will address that question for general insurance. Currently reserving for general insurance is based on aggregated data in run-off triangles. In a run-off triangle observable variables are summarized per arrival year and development year combination. An arrival year is the year in which the claim occurred, while the development year refers to the delay in payment relative to the origin year. Examples of run-off triangles are given in section 7.6.

There exists a vast literature about techniques for claims reserving, largely designed for application to loss triangles. An overview of these techniques is given in England and Verrall (2002), Wüthrich and Merz (2008) or Kaas et al (2008). These techniques can be applied to run-off triangles containing either paid losses or incurred losses (i.e. the sum of paid losses and case reserves). The most popular approach is the Chain Ladder approach, largely because of is practicality. However, the use of aggregated data in combination with (stochastic variants of) the Chain Ladder approach (or similar techniques) gives rise to several issues. A whole literature on itself has evolved to solve these issues, which are (in random order):

1) Different results between projections based on paid losses or incurred losses, addressed by Quarg and Mack (2008), Posthuma et al (2008) and Halliwell (2009).
2) Lack of robustness and the treatment of outliers, see Verdonck et al (2009).
3) The existence of the Chain Ladder bias, see Halliwell (2007) and Taylor (2003).
4) Instability in ultimate claims for recent arrival years, see Bornhuetter and Ferguson (1972).
5) Modeling negative or zero cells in a stochastic setting, see Kunkler (2004).
6) The inclusion of calendar year effects, see Verbeek (1972) and Zehnwirth (1994).
7) The possibly different treatment of small and large claims, see Alai and Wüthrich (2009).
8) The need for including a tail factor, see for example Mack (1999).
9) Over parametrization of the Chain Ladder method, see Wright (1990) and Renshaw (1994).

10) Separate assessment of IBNR and RBNS claims, see Schieper (1991) and Liu and Verrall (2009).
11) The realism of the Poisson distribution underlying the Chain Ladder method.
12) Not using lots of useful information about the individual claims data, as noted by England and Verrall (2002) and Taylor and Campbell (2002).

Most references above present useful additions to the Chain Ladder method, but these additions cannot all be applied simultaneously. More importantly, the existence of these issues and the substantial literature about it indicate that the use of aggregate data in combination with (stochastic variants of) the Chain Ladder approach (or similar techniques) is not fully adequate for capturing the complexities of stochastic reserving for general insurance.

England and Verrall (2002) and Taylor and Campbell (2002) questioned the use of aggregate loss data when the underlying extensive micro-level data base is available as well. With aggregate data, lots of useful information about the claims data remains unused. Covariate information from policy, policy holder or the past development process cannot be used in the traditional stochastic model, since each cell of the run-off triangle is an aggregate figure. Quoting England and Verrall (2002, page 507) "*[...] it has to be borne in mind that traditional techniques were developed before the advent of desktop computers, using methods which could be evaluated using pencil and paper. With the continuing increase in computer power, it has to be questioned whether it would not be better to examine individual claims rather than use aggregate data*".

As a result of the observations mentioned above, a small stream of literature has emerged about stochastic loss reserving on an individual claim level. Arjas (1989), Norberg (1993) and Norberg (1999) formulated a mathematical framework for the development of individual claims. Using ideas from martingale theory and point processes, these authors present a probabilistic, rather than statistical, framework for individual claims reserving. Haastrup and Arjas (1996) continue the work by Norberg and present a first detailed implementation of a micro-level stochastic model for loss reserving. They use non-parametric Bayesian statistics which may complicate the accessibility of the paper. Furthermore, the case study is based on a small data set with fixed claim amounts. Recently, Larsen (2007) revisited the work of Norberg, Haastrup and Arjas with a small case-study. However, detailed information about his modeling choices is not available in the paper. Zhao et al (2009) and Zhao and Zhou (2009) present a model for individual claims development using (semi-parametric) techniques from survival analysis and copula methods. However, a case study is lacking in their work.

In this chapter a micro-level stochastic model is used to quantify the reserve and its uncertainty for a realistic general liability insurance portfolio. Stochastic processes for the occurrence times, the reporting delay, the development process and the payments are fitted to the historical individual data of the portfolio and used for projection of future claims and its (estimation and process) uncertainty. Both the Incurred But Not Reported (IBNR) reserve as well as the Reported But Not Settled (RBNS) reserve are quantified and the results are compared with those of traditional actuarial techniques.

We investigate whether the quality of reserves and their uncertainty can be improved by using more detailed claims data in this way. A micro-level approach allows much closer modeling of

the claims process. Lots of issues mentioned above will not exist when using a micro-level approach, because of the availability of lots of data and the potential flexibility in modeling the future claims process. For example, covariate information (deductibles, policy limits, calendar year) can be included in the projection of the cash flows when claims are modeled at an individual level. The use of lots of (individual) data avoids robustness problems and over parametrization. Also the problems with negative or zero cells and setting the tail factor are circumvented, and small and large claims can be handled simultaneously. Furthermore, individual claim modeling can provide a natural solution for the dilemma within the traditional literature whether to use triangles with paid claims or incurred claims. This dilemma is important because practicing actuaries put high value to their companies' expert opinion which is expressed by setting an initial case reserve. Using micro-level data we use the initial case reserve as a covariate in the projection process of future cash flows.
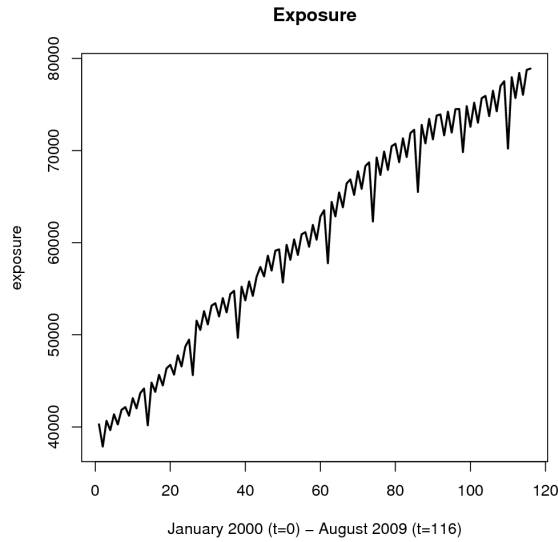
The remainder of the chapter is organized as follows. First, the dataset is introduced in section 7.2. In section 7.3 the statistical model is described. Results from estimating all components of the model are in section 7.4. Section 7.5 presents the prediction routine and section 7.6 shows results and a comparison with traditional actuarial techniques. Section 7.7 gives conclusions.

7.2 Data

The data set used in this chapter contains information about a general liability insurance portfolio (for private individuals) of a European insurance company. The data available consists of the exposure per month from January 2000 till August 2009, as well as a claim file that provides a record of each claim filed with the insurer from January 1997 till August 2009. Note that we are missing exposure information for the period January 1997 till December 1999, but the impact of this lack on our reserve calculations will be very small.

Exposure   The exposure is not the number of policies, but the "earned" exposure. That implies that 2 policies which are both only insured for half of the period are counted as 1. Figure 7.2 shows the exposure per month. Note that the downward spikes correspond to the month February.

Exposure

Random development processes The claim file consists of 1.525.376 records corresponding with 491.912 claims. Figure 7.3 shows the development of 3 claims, taken at random from our data set. It shows the timing of events as well as the cost of the corresponding payments (if any). These are indicated as jumps in the figure. Starting point of the development process is the accident date. This is indicated with a sub–title in each of the plots and corresponds with the point $x = 0$. The x–axis is in months since the accident date. The y–axis represents the cumulative amount paid for the claim.

Figure 7.3: development of 3 claims from the data set
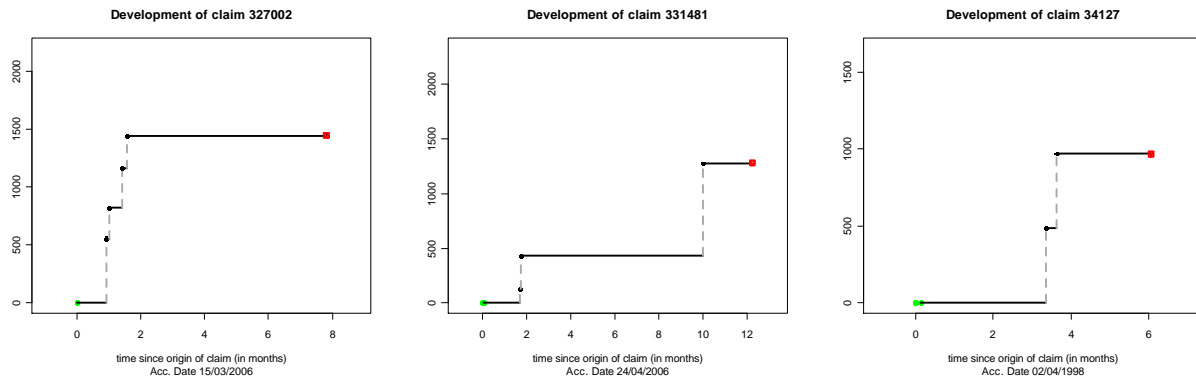
Type and number of claims In this general liability portfolio, there are 2 types of claims: material damage ('material') and bodily injury ('injury'). Figure 7.4 shows the number of open and closed claims per arrival year, and whether they are closed or still open.

110

Figure 7.4: number of open and closed claims, material and injury
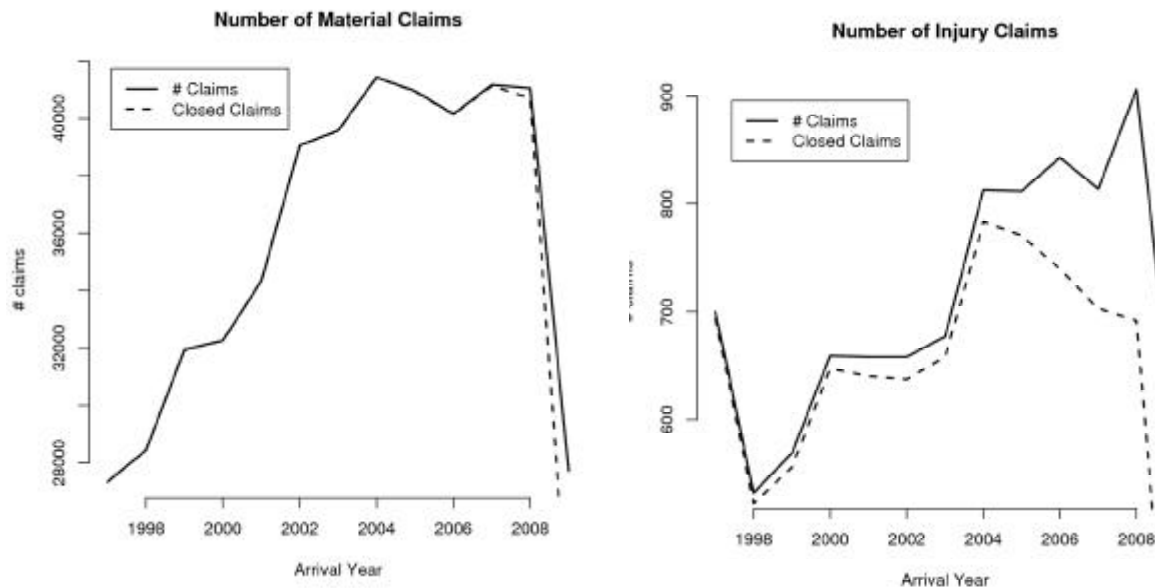
The development pattern and loss distributions of those claim types are usually very different. In practice they are therefore treated separately in separate run-off triangles. Following this approach we will treat them separately too.

Reporting and Settlement delay   Important drivers of the IBNR and RBNS reserves are the reporting delays and settlement delays. Figure 7.5 and 7.6 show the reporting delays and settlement delays separately for material and injury losses. The reporting delay is the time that passes between the occurrence date of the accident and the date it was reported to the insurance company. It is measured in month since the occurrence of the claim. The settlement delay is the time elapsed between the reporting data of the claim and the date of final settlement by the company. It is measured in months and only available for closed claims.

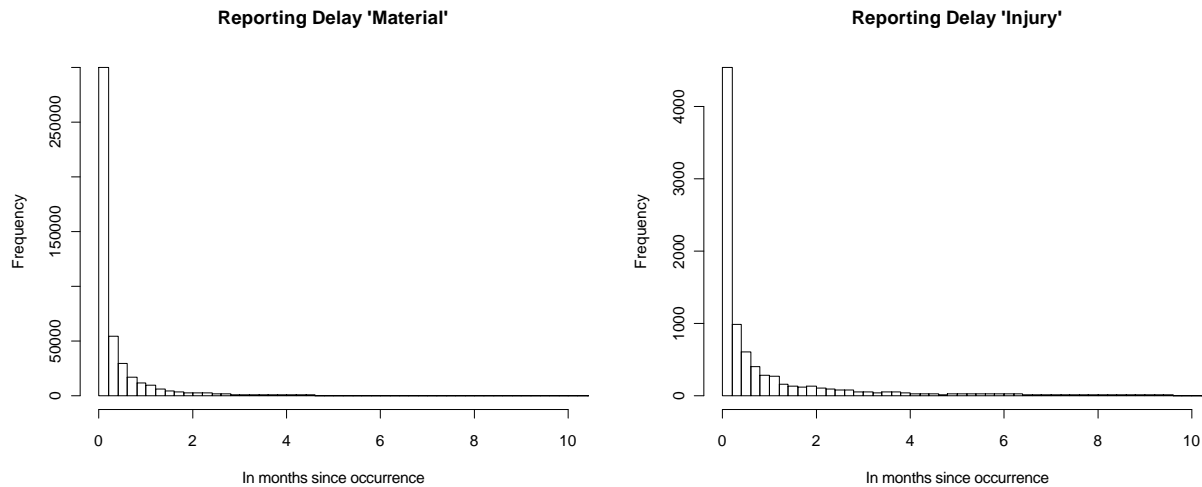*Figure 7.5: histogram of reporting delay for material claims and injury claims*

Reporting Delay 'Material'

Reporting Delay 'Injury'

Figure 7.6: histogram of settlement delay for material claims and injury claims

Settlement Delay 'Material'

Settlement Delay 'Injury'
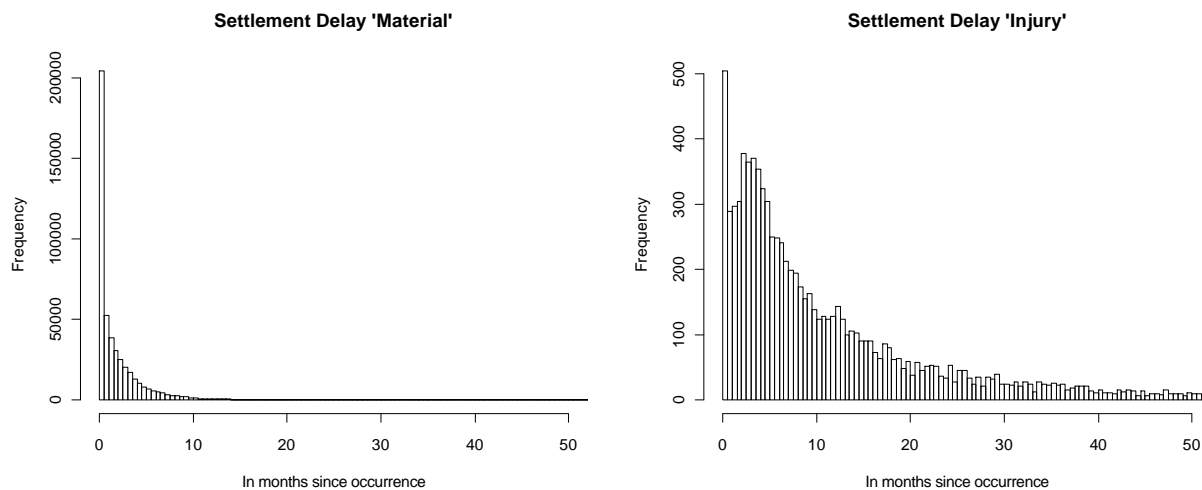
The figures above show that the observed reporting delays are of similar length for material and injury losses. However, the settlement delay is very different. The settlement delay is far more skewed to the right for the injury claims than for the material claims.

Events in the development The settlement delay is the result of the development process of the claim. During the development process, different types of events are possible. In this chapter we will distinguish three types of events that can occur during the development of a claim. "Type 1" events imply settlement of the claim without payment. With a "type 2" event we will refer to a payment with settlement at the same time. Intermediate payments (without settlement) are "type 3" events.

Figure 7.7 gives the relative frequency of the different types of events over development quarters. With micro-level data the first development quarter is the period of 3 months following the reporting of the claim, the second quarter the period of 3 months following the first development quarter, et cetera.

Figure 7.7: number of each event type as percentage of total number of events
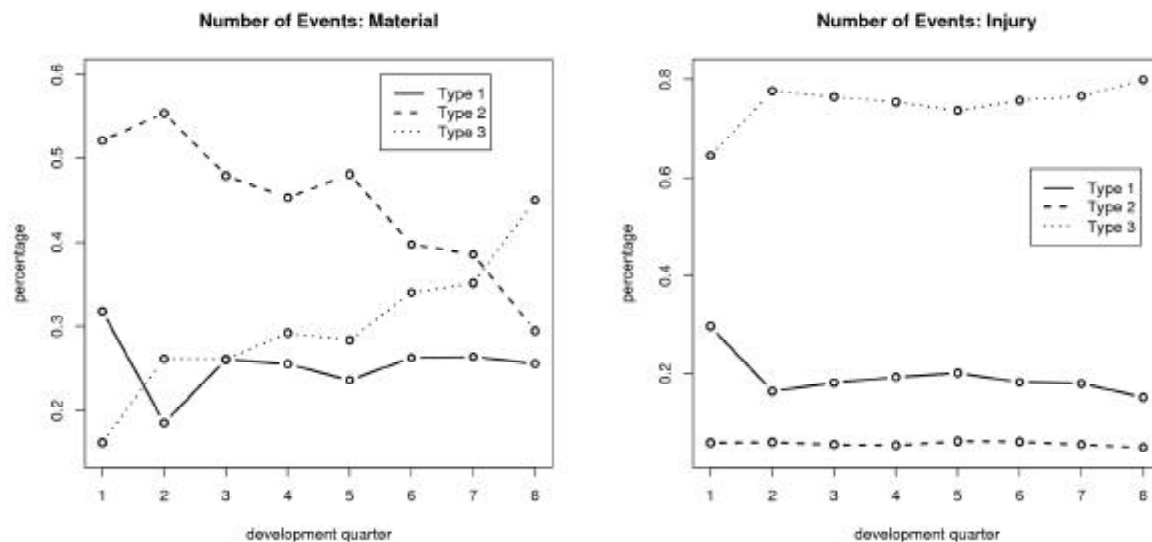
The figure shows that the proportions of each event type are stable over the development quarters for injury claims. For material claims, the proportion of event type 2 decreases for later development quarters, while the proportion of event type 3 increases.

Payments Events of type 2 and type 3 come with a payment. The distribution of these payments differs materially for the different type of claims. Figure 7.8 shows the distribution of the log payments, separate for material and injury claims. The payments are discounted to 1-1-1997 with the Dutch consumer price inflation, to exclude the impact of inflation on the distribution of the payments.

Figure 7.8: distribution of payments for material claims and injury claims
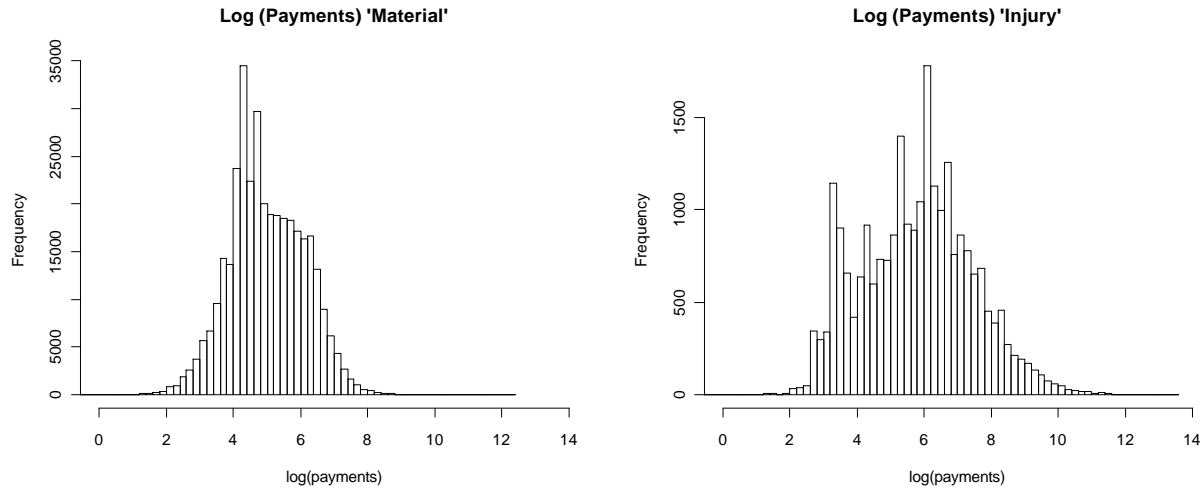
The figures above suggest that a lognormal distribution would probably be reasonable for describing the distribution of the payments. This will be discussed further in section 7.4.

Table 7.1 gives characteristics of the observed (discounted) payments for both material and injury losses.

Table 7.1: characteristics observed payments

Measure	Material	Injury
Mean	277	1.395
Median	129	361
Minimum	0,0008	0,4875
Maximum	198.931	779.398
1% perc.	12	16
5% perc.	25	25
25% perc.	69	89
75% perc.	334	967
95% perc.	890	4.927
99% perc.	1.768	16.664

Initial case estimates   As noted in section 7.1, often the problem arises that the projection based on paid losses is far different than the projection based on incurred losses. This problem is addressed recently by Quarg and Mack (2008), Posthuma et al (2008) and Halliwell (2009), who simultaneously model paid and incurred losses. Disadvantage of those methods is that models based on incurred losses can be instable because the methods for setting the case reserves are sometimes changed (for example, as a result of adequacy test results or profit policy of the company). Reserving models that are directly based on these case reserves (as part of the incurred losses) can therefore be instable. However, the case reserves can have added value as an explaining variable when projecting future payments. We have defined different categories of initial case reserves (separately for material claims and injury claims) that can be used as

explaining variables. Table 7.2 and 7.3 shows the number of claims, the average settlement delay (in months) and the average cumulative payment for these categories.

Table 7.2: output for initial reserve categories, material claims

Initial case reserve	# claims	average settl. delay	average cum. payments
= 10.000	465.015	1,87	252
> 10.000	385	10,88	7.950

Table 7.3: output for initial reserve categories, injury claims

Initial case reserve	# claims	average settl. delay	average cum. payments
= 1.000	3.709	9,87	2.570
(1.000 - 15.000]	5.165	15,17	3.872
> 15.000	360	35,20	33.840

The tables clearly show the differences in settlement delay and cumulative payments for the different initial reserve categories. Therefore, it might be worthwhile to include these categories as explaining variables into the projection routine.

7.3 The statistical model

By a claim i is understood a combination of an occurrence time T_i, a reporting delay U_i and a development process X_i. Hereby X_i is short for $(E_i(v),\ P_i(v))_{v \in [0,Vi]}$. $E_i(v_{ij}) := E_{ij}$ is the type of the jth event in the development of claim i. This event occurs at time v_{ij}, expressed in months after notification of the claim. V_i is the total waiting time from notification to settlement for claim i. If the event includes a payment, the corresponding severity is given by $P_i(v_{ij}) := P_{ij}$. The different types of events are specified in section 7.2. The development process X_i is a jump process. It is modeled here with two separate building blocks: the timing and type of events and their corresponding severities. The complete description of a claim is given by:

(7.1) $(T_i,\ U_i,\ X_i)$ with $X_i = (E_i(v),\ P_i(v))_{v \in [0,Vi]}$

Assume that outstanding liabilities are to be predicted at calendar time t. We distinguish IBNR, RBNS and settled claims.

- For an IBNR claim: $T_i + U_i > t$ and $T_i < t$
- For an RBNS claim: $T_i + U_i \pounds\ t$ and the development of the claim is censored at $(t - T_i - U_i)$, i.e. only $(E_i(v),\ P_i(v))_{v \in [0, t - Ti - Ui]}$ is observed.
- For a settled claim: $T_i + U_i \pounds\ t$ and $(E_i(v),\ P_i(v))_{v \in [0,Vi]}$ is observed.

7.3.1 Position Dependent Marked Poisson Process

Following the approach in Haastrup and Arjas (1996) and Norberg (1993) we treat the claims process as a Position Dependent Marked Poisson Process (PDMPP), see Karr (1991). In this

application, a point is an occurrence time and the associated mark is the combined reporting delay and development of the claim. We denote the intensity measure of this Poisson Process with $\mathbf{1}$ and the associated mark distribution with $(P_{Z|t})_{t \geq 0}$. In the claims development framework the distribution $P_{Z|t}$ is given by the distribution $P_{U|t}$ of the reporting delay, given occurrence time t, and the distribution $P_{X|t,u}$ of the development, given occurrence time t and reporting delay u. The complete development process then is a Poisson Process on claim space $? = [0,\infty) \times [0,\infty) \times \mathbf{c}$ with intensity measure:

(7.2) $\quad \mathbf{1}(dt) \times P_{U|t}(du) \times P_{X|t,u}(dx)$ with $(t,u,x) \in ?$

The reported claims (which are not necessarily settled) belong to the set:

$$?^r = \{ (t,u,x) \in ? \mid t+u \leq \mathbf{t} \}$$

whereas the IBNR claims belong to:

$$?^i = \{ (t,u,x) \in ? \mid t \leq \mathbf{t}, t+u > \mathbf{t} \}$$

Since both sets are disjoint, both processes are independent (see Karr (1991)). The process of reported claims is a Poisson Process with measure

$$\mathbf{1}(dt) \times P_{U|t}(du) \times P_{X|t,u}(dx) \times 1_{[(t,u,x) \in C^r]}$$

which equals

(7.3) $\quad \underbrace{\mathbf{1}(dt) P_{U|t}(\mathbf{t}-t) 1_{(t \in [0,\mathbf{t}])}}_{(a)} \times \underbrace{\frac{P_{U|t}(du) 1_{(u \leq \mathbf{t}-t)}}{P_{U|t}(\mathbf{t}-t)}}_{(b)} \times \underbrace{P_{X|t,u}(dx)}_{(c)}$

Part (a) is the occurrence measure. The mark of this claim is composed by a reporting delay, given the occurrence time (its conditional distribution is given by (b)) and the conditional distribution (c) of the development, given the occurrence time and reporting delay.

Similarly, the process of IBNR claims is a Poisson process with measure:

(7.4) $\quad \underbrace{\mathbf{1}(dt)\left(1 - P_{U|t}(\mathbf{t}-t) 1_{(t \in [0,\mathbf{t}])}\right)}_{(a)} \times \underbrace{\frac{P_{U|t}(du) 1_{(u > \mathbf{t}-t)}}{1 - P_{U|t}(\mathbf{t}-t)}}_{(b)} \times \underbrace{P_{X|t,u}(dx)}_{(c)}$

where similar components can be indentified as in (7.3).

7.3.2 The Likelihood

The approach followed in this chapter is parametric. Therefore, we will optimize the likelihood expression for observed data over the unknown parameters used in this expression.

The observed part of the claims process consists of the development up to time t of claims reported before t. We denote these observed claims as follows:

$$\left(T_i^0, U_i^0, X_i^0\right)_{i \geq 1}$$

where the development of claim i is censored $t - T_i^0 - U_i^0$ time units after notification.

The likelihood of the observed claim development process can be written as (see Cook and Lawless (2007)):

$$(7.5) \quad \Lambda(obs) \propto \left(\prod_{i \geq 1} \boldsymbol{1}\left(T_i^0\right) P_{U|t}\left(\boldsymbol{t} - T_i^0\right)\right) \exp\left(-\int_0^t w(t) \boldsymbol{1}(t) P_{U|t}(\boldsymbol{t} - t)dt\right)$$

$$\times \left(\prod_{i \geq 1} \frac{P_{U|t}\left(dU_i^0\right)}{P_{U|t}\left(\boldsymbol{t} - T_i^0\right)}\right) \times \prod_{i \geq 1} P_{X|t,u}^{\boldsymbol{t} - T_i^0 - U_i^0}\left(dX_i^0\right)$$

The superscript $\boldsymbol{t} - T_i^0 - U_i^0$ in the last term of this likelihood indicates the censoring of the development of this claim $\boldsymbol{t} - T_i^0 - U_i^0$ time units after notification. The function $w(t)$ gives the exposure at time t.

For the reporting delay and the development process we will use techniques from survival analysis. The reporting delay is a one-time single type event that can be modeled using standard distributions from survival analysis. For the development process the statistical framework of recurrent events will be used. Cook and Lawless (2007) provide a recent overview of statistical techniques for the analysis of recurrent events. These techniques primarily address the modeling of an event intensity (or hazard rate).

As mentioned in (7.1) for each claim i its development process consists of $X_i = (E_i(v), P_i(v))_{v \hat{\boldsymbol{1}} [0, Vi]}$. Hereby $E_i(v_{ij}) := E_{ij}$ is the type of the jth event in development of claim i, occurring at time v_{ij}. V_i is the total waiting time from notification to settlement for claim i. If the event includes a payment, the corresponding severity is given by $P_i(v_{ij}) := P_{ij}$. To model the occurrence of the different events a hazard rate is specified for each type. The hazard rates h_{se}, h_{sep} and h_p correspond to respectively type 1 (settlement without payment), type 2 (settlement with a payment at the same time) and type 3 (payment without settlement) events.

Events of type 2 and 3 come with a payment. We denote the density of a severity payment with P_p. Using this notation the likelihood of the development process of claim i is given by:

$$(7.6) \quad \left(\prod_{j=1}^{N_i}\left(h_{se}^{d_{ij1}}(V_{ij}) \times h_{sep}^{d_{ij2}}(V_{ij}) \times h_p^{d_{ij3}}(V_{ij})\right)\right) \exp\left(-\int_0^{t_i}(h_{se}(u) + h_{sep}(u) + h_p(u))du\right) \prod_j P_p(dV_{ij})$$

Here $\boldsymbol{d}_{ijk}$ is an indicator variable that is 1 if the jth event in the development of claim i is of type k. N_i is the total number of events, registered in the observation period for claim i. This observation period is $[0, \boldsymbol{t}_i]$ with $\boldsymbol{t}_i = \min(\boldsymbol{t} - T_i - U_i, V_i)$.

Combining (7.5) and (7.6) gives the likelihood for the observed data:

$$(7.7) \quad \Lambda(obs) \; \propto \; \left(\prod_{i \geq 1} \boldsymbol{1}\left(T_i^0\right) P_{U|t}\left(\boldsymbol{t} - T_i^0\right) \right) \exp\left(-\int_0^t w(t)\boldsymbol{1}(t)P_{U|t}(\boldsymbol{t} - t)dt \right) \times \left(\prod_{i \geq 1} \frac{P_{U|t}\left(dU_i^0\right)}{P_{U|t}\left(\boldsymbol{t} - T_i^0\right)} \right)$$

$$\times \; \prod_{i \geq 1}\left(\prod_{j=1}^{N_i}\left(h_{se}^{\boldsymbol{d}_{ij1}}(V_{ij}) \times h_{sep}^{\boldsymbol{d}_{ij2}}(V_{ij}) \times h_p^{\boldsymbol{d}_{ij3}}(V_{ij}) \right) \right) \exp\left(-\int_0^{t_i}(h_{se}(u) + h_{sep}(u) + h_p(u))du \right)$$

$$\times \; \prod_{i \geq 1}\prod_{j=1}^{N_i} P_p(dVij)$$

7.3.3 Distributional assumptions

In this paragraph we discuss the likelihood (7.7) in more detail. Distribution assumptions for the various building blocks, being the reporting delay, the occurrence times – given the reporting delay – and the development process, are presented. At each stage it is possible to include covariate information such as the initial case reserve categories. Our final choices and estimation results will be covered in section 7.4.

Reporting delay The notification of a claim is a one-time single type event that can be modeled using standard distributions from survival analysis (such as the Exponential, Weibull or Gompertz distribution). Figure 7.5 indicates that for a large part of the claims the claim will be reported in the first few days after the occurrence. Therefore we will use a mixture of one of the above mentioned distributions with one or more degenerate distributions for notification during the first few days. For example, for a mixture of a survival distribution f_U with n degenerate components the density is given by:

$$(7.8) \quad \sum_{k=0}^{n-1} p_k I_{\{k\}}(u) \; + \; \left(1 - \sum_{k=0}^{n-1} p_k \right) f_{U|U>n-1}(u)$$

where $I_{\{k\}} = 1$ for the kth day after occurrence time t and $I_{\{k\}} = 0$ otherwise.

Occurrence process When optimizing the likelihood for the occurrence process the reporting delay distribution and its parameters (as obtained in the previous step) are used. The likelihood

$$(7.9) \quad L \; \propto \; \prod_{i \geq 1} \boldsymbol{1}\left(T_i^0\right) P_{U|t}\left(\boldsymbol{t} - T_i^0\right) \exp\left(-\int_0^t w(t)\boldsymbol{1}(t)P_{U|t}(\boldsymbol{t} - t)dt \right)$$

needs to be optimized over $\boldsymbol{1}(t)$. A piecewise constant specification is used for this occurrence rate:

$$(7.10) \quad \boldsymbol{I}(t) = \begin{cases} \boldsymbol{I}_1 & 0 \le t < d_1 \\ \boldsymbol{I}_2 & d_1 \le t < d_2 \\ \vdots & \vdots \\ \boldsymbol{I}_m & d_{m-1} \le t < d_m \end{cases}$$

where the intervals are chosen in such a way that $\boldsymbol{t} \in [d_{m-1}, d_m)$ and the exposure $w(t) := w_l$ for $d_{l-1} \le t < d_l$.

Let the indicator variable $\boldsymbol{d}_l(l, t_i)$ be 1 if $d_{l-1} \le t_i < d_l$, with t_i the occurrence time of claim i. The number of claims in interval $[d_{l-1}, d_l)$ can be expressed as:

$$(7.11) \quad N_{oc}(l) = \sum_i \boldsymbol{d}_1(l, t_i)$$

The likelihood corresponding to the occurrence times is given by:

$$(7.12) \quad L \propto \boldsymbol{I}_1^{N_{oc}(1)} \boldsymbol{I}_2^{N_{oc}(2)} \dots \boldsymbol{I}_m^{N_{oc}(m)} \prod_{i \ge 1} P_{U|t}(\boldsymbol{t} - t_i)$$

$$\times \exp\left(-\boldsymbol{I}_1 w_1 \int_0^{d_1} P_{U|t}(\boldsymbol{t} - t)dt\right) \exp\left(-\boldsymbol{I}_2 w_2 \int_{d_1}^{d_2} P_{U|t}(\boldsymbol{t} - t)dt\right)$$

$$\dots \exp\left(-\boldsymbol{I}_m w_m \int_{d_{m-1}}^{d_m} P_{U|t}(\boldsymbol{t} - t)dt\right)$$

Optimizing this expression over $\boldsymbol{I}_l$ (with $l = 1, \dots, m$) leads to:

$$(7.13) \quad \hat{\boldsymbol{I}}_l = \frac{N_{oc}(l)}{w_l \int_{d_{l-1}}^{d_l} P_{U|t}(\boldsymbol{t} - t)dt}$$

Development process   Similar distributions as given for the reporting delay can be used for each type of event in the development process. Another alternative is a piecewise constant specification of the hazard rates. This implies:

$$(7.14) \quad h_{\{se,sep,p\}}(t) = \begin{cases} h_{\{se,sep,p\},1} & for\ 0 \le t < a_1 \\ h_{\{se,sep,p\},2} & for\ a_1 \le t < a_2 \\ \vdots & \vdots \\ h_{\{se,sep,p\},q} & for\ a_{q-1} \le t < a_q \end{cases} = \prod_{l=1}^{q} h_{\{se,sep,p\},l}^{d_1(l,t)}$$

where $\boldsymbol{d}_l(l,t)$ is 1 if $a_{l-1} \le t < a_l$ and 0 otherwise. This piecewise specification can be integrated in a straightforward way in likelihood specification (7.6) and (7.7), although the resulting

expression is complex in notation. The optimization of the likelihood expression can be done analytically or numerically. It might be worthwhile to fit the distribution separately for 'first events' and 'later events'. This will be investigated in section 7.4.

Payments   Event type 2 and type 3 come with a payment. Section 7.2 showed that the observed distribution of the payments has similarities with a lognormal distribution, but there might be more flexible distributions that fit the historical payment data better. Therefore, next to the lognormal distribution, we experimented as well with a generalized beta of the second kind (GB2), Burr and Gamma distribution. Also covariate information such as the initial reserve category and the development year can be taken into account.

7.4 Estimation results

In this paragraph the results of the calibration of the model to the historical data are given.
Given the very different characteristics of material claims and injury claims, the processes described in section 7.3 are fitted (and projected) separately for those types of claims. This is in line with actuarial practice, where usually separate run-off triangles are constructed for material and injury claims. Optimization of all likelihood specifications was done with the Proc NLMixed routine in SAS.

Reporting delay   In paragraph 7.3.3 we specified the possible models for the reporting delay. In this chapter we will use a mixture of a Weibull distribution and 9 degenerate distributions. Figure 7.9 shows the fit of this mixture with the observed reporting delays.

Figure 7.9: estimate of reporting delay
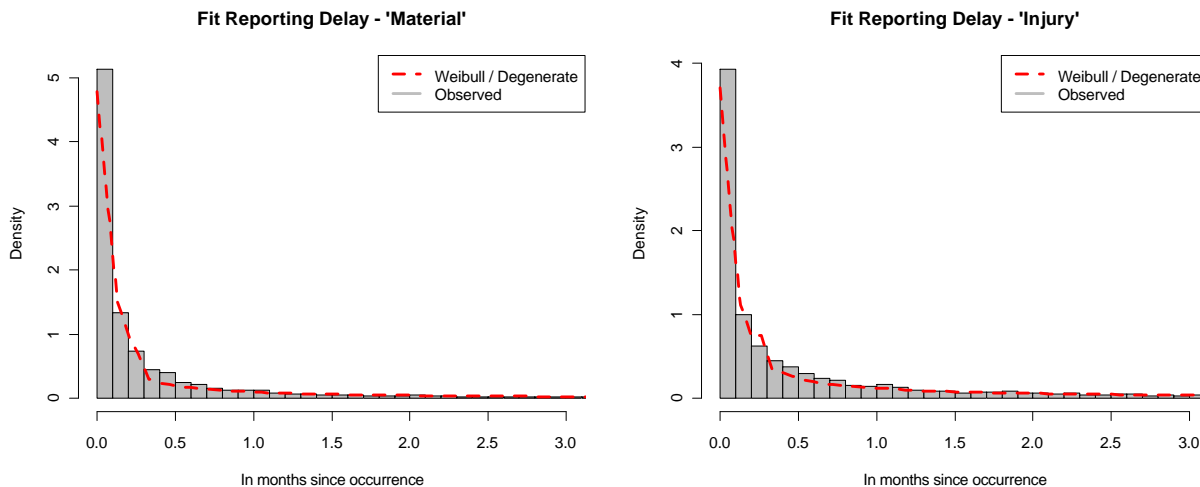
**Occurrence process**   Given the above specified distribution for the reporting delay, the likelihood (7.12) for the occurrence times can be optimized[31]. Monthly intervals are used for this,

---

[31] This is done numerically with Proc NLMixed instead of using (3.13), in order to obtain the standard errors of the parameter estimates. These standard errors will also be used in the prediction process.

ranging from January 2000 till August 2009. The estimated l_i's (black line) and their 95% confidence intervals (grey area) are given in figure 7.10.

Figure 7.10: estimate lambda's and their uncertainty
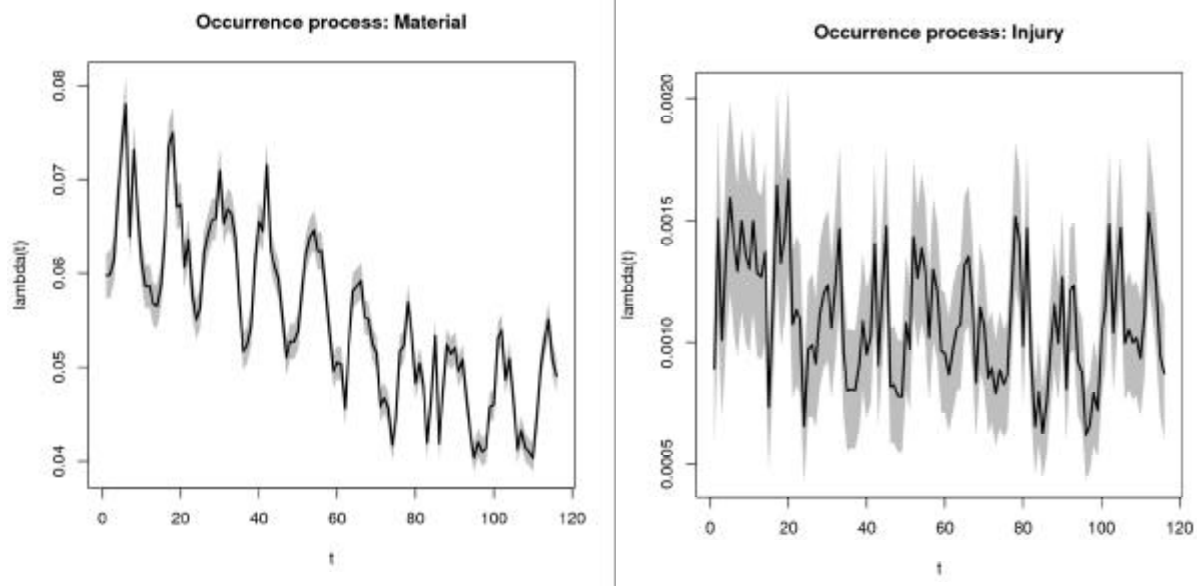
Development process   For the different event types in the development process delay the use of constant, Weibull and piecewise constant hazard rates are investigated. In the piecewise constant hazard rate specification for the development of the material claims, the hazard rate is assumed to be continuous on four month intervals: [0 – 4) months, [4 – 8) months, [8 – 12) months and ≥ 12 months. For injury claims, the hazard rate is assumed continuous on intervals of six months: [0 – 6) months, [6 – 12) months, …, [36 – 42) months and ≥ 42 months.

Figure 7.11 shows the estimates for the Weibull and piecewise constant hazard rates. All models are estimated separately for 'first events' and 'later events'.

Figure 7.11: estimates for Weibull and piecewise constant hazard rates
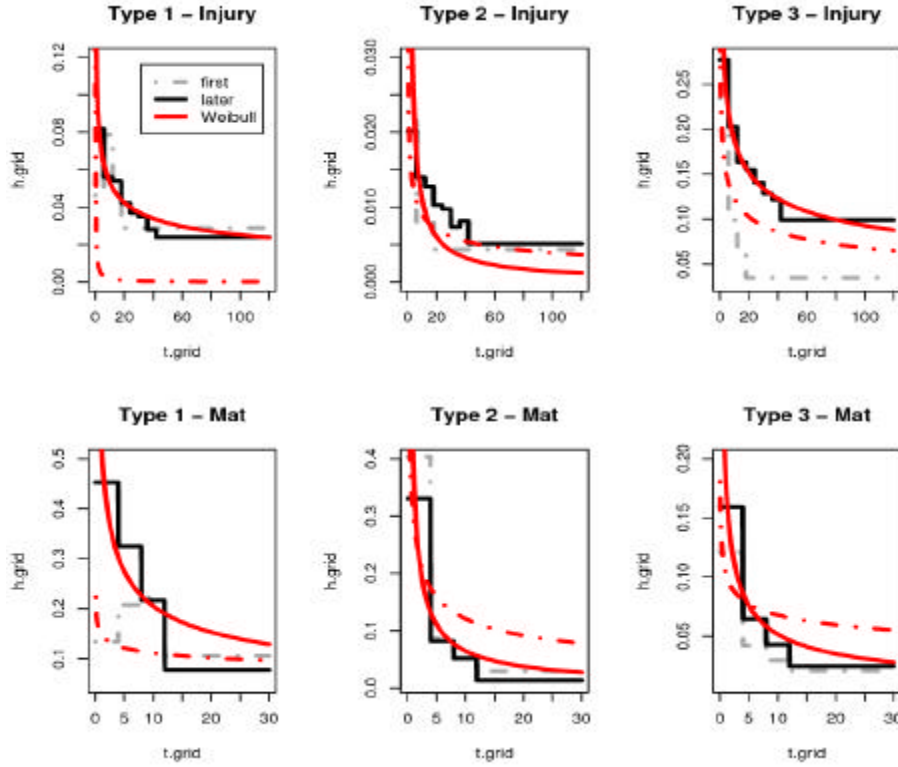
The piecewise constant specification reflects the actual data. The figure shows that the Weibull distribution is reasonably close to the piecewise constant specification. In the rest of this chapter we will use the piecewise constant specification. Because the Weibull distribution is a good alternative, we explain how to use both specifications in the prediction routine (see section 7.5).

Payments   Several distributions have been fitted to the historical payments (that are discounted to 1-1-1997 with Dutch price inflation). We examined the fit of the Burr, Gamma and Lognormal distribution, combined with covariate information. Distributions for the payments are truncated at the coverage limit of € 2,5 million per claim. A comparison based on Bayes Information Criterium (*BIC*) showed that the lognormal distribution achieves a better fit than the Burr and Gamma distributions. When including the initial reserve category as covariate or both the initial reserve category and the development year, the fit further improves. Given these results, the lognormal distribution with the initial reserve category and the development year as covariates will be used in the prediction. The covariate information is included in both the mean m_i and standard deviation s_i of the lognormal distribution for observation i as follows:

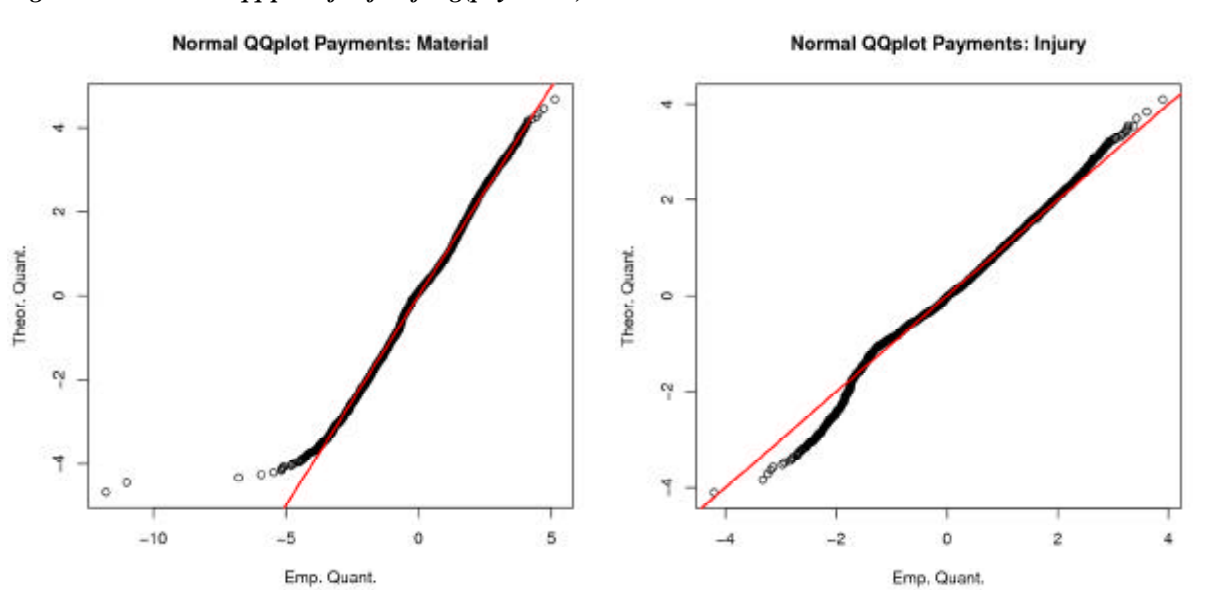$$(7.15) \quad m_i \quad = \quad \sum_r \sum_s m_{r,s} I_{DY_i=s} I_{i \in r}$$

$$(7.16) \quad s_i \quad = \quad \sum_r \sum_s s_{r,s} I_{DY_i=s} I_{i \in r}$$

122

where r is the initial reserve category and DY_i is the development year. $I_{DYi=s}$ and $I_{i\in r}$ are indicator variables denoting whether observation i corresponds with development year s and reserve category r.

Figure 7.12 shows the corresponding qq-plots.

Figure 7.12: normal qq-plots for fit of log(payments)

The figures show that the fit to the data is good. Note that the fit in the left tail seems to be less good, but this is corresponding to payments of about 0 (so not important in this case).

7.5 Predicting future cash flows

To predict the outstanding liabilities with respect to this portfolio of liability claims, we distinguish between IBNR and RBNS claims. The following step by step approach allows to obtain random draws from the distribution of both IBNR and RBNS claims.

7.5.1 Predicting IBNR claims

As noted in section 7.3, an IBNR claim occurred already but is not reported to the insurer. Therefore, $T_i + U_i > \tau$ where T_i is the occurrence time of the claim and U_i is its reporting delay. The T_i's are missing data: they are determined in the development process but unknown to the actuary at time t.

The prediction process for the IBNR claims requires the following steps:

a) Simulate the number of IBNR claims in [0,t] and their occurrence times
According to the discussion in section 7.3 the IBNR claims are governed by a Poisson process with non-homogeneous intensity or occurrence rate:

(7.17) $w(t)\mathbf{l}(t)(1-P_{U|t}(\mathbf{t}-t))$

were $\mathbf{l}(t)$ is piecewise constant according to specification (7.10). The following property follows from the definition of non-homogeneous Poisson processes:

$$(7.18) \quad N_{IBNR}(l) \sim Poisson\left(\mathbf{l}_l w_l \int_{d_{l-1}}^{d_l} (1-P_{U|t}(\mathbf{t}-t))dt\right)$$

were $N_{IBNR}(l)$ is the number of IBNR claims in time interval $[d_{l-1},d_l)$. Note that the integral expression has already been evaluated (numerically) in the fitting procedure.

Given the simulated number of IBNR claims $n_{IBNR}(l)$ for each interval $[d_{l-1},d_l)$, the occurrence times of the claims are uniformly distributed in $[d_{l-1},d_l)$.

b) Simulate the reporting delay for each IBNR claim
Given the simulated occurrence time t_i of an IBRN claim, its reporting delay is simulated by inverting the distribution:

$$(7.19) \quad P(U \leq u \,|\, U > \mathbf{t}-t_i) = \frac{P(\mathbf{t}-t_i < U \leq u)}{1-P(U \leq \mathbf{t}-t_i)}$$

In case of our assumed mixture of a Weibull distribution and 9 degenerate distributions this expression has to be evaluated numerically.

c) Simulate the initial reserve category
For each IBNR claim an initial reserve category has to be simulated for use in the development and payment process. Given m initial reserve categories, the probability density for initial reserve category c is:

$$(7.20) \quad f(c) = \begin{cases} p_c & \text{for } c = 1, 2, ..m-1 \\ 1 - \sum_{k=1}^{m-1} p_k & \text{for } c = m \end{cases}$$

The probabilities used in (7.20) are the empirically observed percentages of policies in a particular initial reserve category.

d) Simulate the payment process for each IBNR claim
This step is common with the procedure for RBNS claims and will be explained in the next paragraph.

7.5.2 Predicting RBNS claims

Given the RBNS claims and the simulated IBNR claims, the process proceeds as below. Note that we use the piecewise hazard specification for the development process. As an alternative for

the analytical specifications given below, numerical routines could be used. Using the alternative Weibull specification would require numerical operations as well.

e) Simulate the next event's exact time
In case of RBNS claims, the time of censoring c_i of claim i is known. For IBNR claims this censoring time $c_i = 0$. The next event at time $v_{i,next}$ can take place at any time $v_{i,next} > c_i$. To simulate its exact time we need to invert (with p randomly drawn from a Uniform(0,1) distribution):

$$(7.21) \quad P\left(V < v_{i,next} \mid V > c_i\right) = \frac{P\left(c_i < V \le v_{i,next}\right)}{1 - P\left(V \le c_i\right)} = p$$

From the relation between a hazard rate and the cdf, we know:

$$(7.22) \quad P\left(V \le v_{i,next}\right) = 1 - \exp\left(-\int_0^{v_{i,next}} \sum_e h_e(t)dt\right)$$

with $e \in \{se, sep, p\}$. For instance with a Weibull specification for the hazard rates this equation will be inverted numerically. With a piecewise constant specification for the hazard rates numerical routines can be used. Alternatively analytical expressions can be derived. In that case, step (e) should then be replaced by (e1) – (e2):

e1) Simulate the next event's time interval
In case of RBNS claims, the time of censoring c_i of claim i belongs to a certain interval $[a_{l-1}, a_l)$. The next event – at time $v_{i,\text{next}} > c_i$ – can take place in any interval from $[a_{l-1}, a_l)$ on. The probability that $v_{i,\text{next}}$ belongs to a certain interval $[a_{l-1}, a_l)$ is given by:

$$(7.23) \quad P\left(a_{k-1} \le V < a_k\right) = \begin{cases} \dfrac{P(c_i < V < a_k)}{1 - P(V \le c_i)} & \text{if } c_i \in [a_{k-1}, a_k) \\[4mm] \dfrac{P(a_{k-1} < V < a_k)}{1 - P(V \le c_i)} & \text{if } c_i \notin [a_{k-1}, a_k) \end{cases}$$

Using the notation introduced above the involved probabilities can be expressed as (for instance):

$$\frac{P(c_i < V < a_k)}{1 - P(V \leq c_i)} = \frac{P(V < a_k) - P(V < c_i)}{1 - P(V \leq c_i)}$$

(7.24)
$$= \frac{1 - \exp\left\{-\int_0^{a_k} \sum_e h_e^f(t)dt\right\} - 1 + \exp\left\{-\int_0^{c_i} \sum_e h_e^f(t)dt\right\}}{\exp\left\{-\int_0^{c_i} \sum_e h_e^f(t)dt\right\}}$$

$$where \int_0^z \sum_e h_e^f(t)dt = \sum_e \sum_{l=1}^d h_{el}^f\left[(a_l - a_{l-1})\mathbf{d}_2(l,z) + (z - a_{l-1})\mathbf{d}_1(l,z)\right] \ for \ z = c_i, a_k$$

with $e \in \{se, sep, p\}$ and $f \in \{\text{'first event', 'later events'}\}$.

In case of IBNR claims, there is no censoring so the probability that $v_{i,\text{next}}$ belongs to a certain interval $[a_{l-1}, a_l)$ simplifies to:

(7.25) $$P(a_{k-1} \leq V < a_k) = \exp\left\{-\int_0^{a_{k-1}} \sum_e h_e^f(t)dt\right\} - \exp\left\{-\int_0^{a_k} \sum_e h_e^f(t)dt\right\}$$

e2) Simulate the exact time of the next event
Given the time interval of the next event, $[a_{l-1}, a_l)$, its exact time is simulated by inverting the following equation for $v_{i,\text{next}}$:

(7.26)
$$P(V < v_{i,next} \mid c_i < V < a_k) = p \quad if \ c_i \in [a_{k-1}, a_k)$$
$$P(V < v_{i,next} \mid a_{k-1} < V < a_k) = p \quad otherwise$$

where p is randomly drawn from a Uniform(0,1) distribution. For example, for P($V < v_{i,\text{next}}|a_{k-1} \leq V < a_k$) this inverting operation goes as follows:

$$P(V < v_{i,next}) = p \ P(a_{k-1} < V < a_k) + P(V < a_{k-1}) \Rightarrow$$

$$1 - \exp\left\{-\int_0^{v_{i,next}} \sum_e h_e^f(t)dt\right\} = p \ P(a_{k-1} < V < a_k) + P(V < a_{k-1}) \Rightarrow$$

$$-\log\left[p \ P(a_{k-1} < V < a_k) + P(V < a_{k-1})\right] = \sum_e \sum_{l=1}^{k-1} h_{el}^f(a_l - a_{l-1}) + \sum_e h_{ek}^f(v_{i,next} - a_{k-1}) \Rightarrow$$

(7.27) $$v_{i,next} = a_{k-1} + \frac{-\log\left[p \ P(a_{k-1} < V < a_k) + P(V < a_{k-1})\right] - \sum_e \sum_{l=1}^{k-1} h_{el}^f(a_l - a_{l-1})}{\sum_e h_{ek}}$$

f) Simulate the event type
Given the exact time of the next event, its type is simulated using the following argument:

$$(7.28) \quad \lim_{\Delta v \to 0} P\left(E = e \mid v \leq V < v + \Delta v\right) = \frac{P(v \leq V < v + \Delta v \cap E = e)/\Delta v}{P(v \leq V < v + \Delta v)/\Delta v} = \frac{h_e(v)}{\sum_e h_e(v)}$$

where e $\in$ {*se, sep, p*}.

g) Simulate the corresponding payment
Given the covariate information for claim *i*, the payment can be drawn from the appropriate lognormal distribution. Note that the cumulative payment cannot exceed the coverage limit of € 2,5 million per claim.

h) Stop or continue
Depending on the simulated event type in step f), the prediction stops (in case of settlement) or continues.

In the next section, this prediction process will be applied separately for the material claims and the injury claims.

7.5.3 Comment on estimation uncertainty
With regards of the uncertainty of predictions a distinction can be made between process uncertainty and estimation uncertainty (see England and Verrall (2002)). The process uncertainty wil be taken care of by sampling from the distributions proposed in section 7.3. To include parameter uncertainty the bootstrap technique or concepts from Bayesian statistics can be used. While a formal Bayesian approach is very elegant, it generally leads to significantly more complexity, which is not contributing to the accessibility and transparency of the techniques towards practicing actuaries. Applying a bootstrap procedure would be possible, but is very computer intensive, since our sample size is very large and several stochastic processes are used. To avoid computational problems when dealing with parameter uncertainty, we will use the asymptotic normal distribution of our maximum likelihood estimators. At each iteration of the prediction routine we sample each parameter from its corresponding asymptotic normal distribution. Note that – due to our large sample size – confidence intervals are narrow. This is in contrast with run-off triangles where sample sizes are typically very small and estimation uncertainty is an important point of concern.

7.6 Numerical results

The prediction process described in Section 7.5 is applied separately for the material and injury claims. In this section results obtained with the micro–level reserving model are shown. Our results are compared with those from traditional techniques based on aggregate data. We show results for an out–of–sample exercise, so that the estimated reserves can be compared with actual payments. This out–of–sample test is done by estimating the reserves per 1-1-2005. The data set that is available at 1-1-2005 can be summarized using run-off triangles, displaying data from arrival years 1999 –2004. Table 7.4 (material) and 7.5 (injury) show the run–off triangles that are the basis for this out–of–sample exercise. The lower triangle is known up to 3 cells. The actual

observations are given in bold. Of course, these were not known at 1-1-2005 so cannot be used as input for calibration of the models.

Table 7.4: run-off triangle 'Material' claims, arrival year 1997-2004

arrival year	development year							
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**
1997	4.379.653	971.591	81.875	9.264	35.942	26.720	34.277	10.750
1998	4.333.968	975.501	55.978	35.004	75.768	23.769	572	**16.481**
1999	5.225.441	1.218.325	58.894	107.716	107.832	11.751	**390**	**0**
2000	5.365.758	1.119.476	161.148	14.451	5.927	**4.253**	**36**	**10.014**
2001	5.535.075	1.619.956	118.336	119.202	**12.711**	**2.988**	**350**	**2.184**
2002	6.538.549	1.547.253	67.331	**65.414**	**16.509**	**5.256**	**9.120**	**8.847**
2003	6.535.125	1.601.255	**90.721**	**20.505**	**30.838**	**7.424**	**1.685**	
2004	7.109.492	**1.347.123**	**98.695**	**76.384**	**19.926**	**12.896**		

Table 7.5: run-off triangle 'Injury' claims, arrival year 1997-2004

arrival year	development year							
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**
1997	307.166	635.084	366.324	530.201	548.906	137.401	132.076	338.865
1998	256.758	481.893	311.525	336.221	268.519	56.043	178.618	**78.124**
1999	291.719	589.928	410.442	272.972	254.240	285.602	**132.109**	**96.813**
2000	315.509	601.364	439.408	498.131	406.642	**371.131**	**247.141**	**275.271**
2001	464.813	846.150	566.122	566.855	**445.835**	**375.499**	**146.507**	**239.922**
2002	314.422	614.945	540.023	**449.435**	**132.515**	**131.172**	**332.044**	**1.081.869**
2003	302.699	801.452	**617.225**	**268.342**	**222.621**	**215.501**	**172.566**	
2004	333.075	**864.120**	**411.705**	**245.176**	**272.621**	**100.128**		

Output from the micro–level model The distribution of the reserve per 1-1-2005 is determined for the individual (micro–level) model proposed in this chapter. We will first look at the output that becomes available when using the micro–level model. Figure 7.13 shows results for injury payments done in calendar year 2006, based on 10.000 simulations. In table 7.5 this is the diagonal going from 412, 268, ..., up to 97. The first row in figure 7.13 shows (from left to right): the number of IBNR claims reported in 2006, the total amount of payments done in this calendar year and the total number of events occurring in 2006. The IBNR claims are claims that occurred before 1-1-2005, but were reported to the insurer during calendar year 2006. The total amount paid in 2006 is the sum of payment for RBNS and IBNR claims, which are separately available from the micro-model. In the second row of plots we take a closer look at the events registered in 2006 by splitting into type 1 – type 3 events. In each of the plots the black solid line indicates what was actually observed.

Figure 7.13: results for injury payments, calendar year 2006
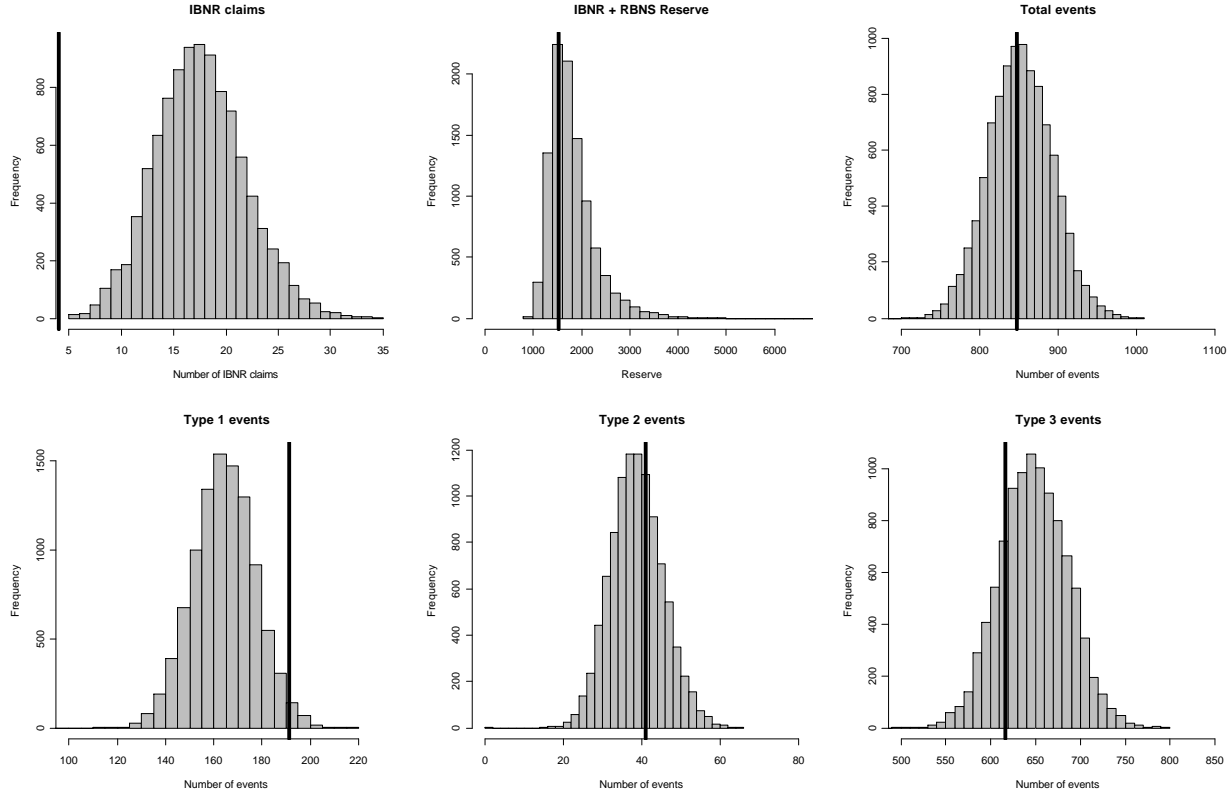
The figure shows that the resulting distributions of the micro-level model are realistic, given the actual observations. Only the actual number of IBNR claims is far in the tail of the distribution. However, note that this relates to a relatively low number of IBNR claims.

Comparing reserves   The results from the micro-level model are now compared with results from two standard actuarial models developed for aggregate data. To the data in tables 7.4 and 7.5, a stochastic Chain-Ladder model is applied which is based on the Overdispersed Poisson distribution and the Lognormal distribution, respectively. With Y_{ij} denoting cell (i,j) from a run-off triangle, corresponding with arrival year i and development year j, the model specifications are:

(7.29)   *Overdispersed Poisson* : $Y_{ij} = fM_{ij}$ $M_{ij} \sim Poi\left(m_{ij} / f\right)$ $m_{ij} = a_i + b_j$

(7.30)   *Lognormal* : $\log(Y_{ij}) \sim m_{ij} + e_{ij}$ $m_{ij} = a_i + b_j$ $e_{ij} \sim N\left(0, s^2\right)$

Both aggregate models are implemented in a Bayesian framework[32].

---

[32] The implementation of the Overdispersed Poisson is in fact empirically Bayesian. f is estimated on beforehand and held fixed. We use vague normal priors for the regression parameters in both models and a gamma prior for s^{-1} in the Lognormal model.

Figure 7.14 shows the distributions of the total payments (in thousands Euro) for material claims, as obtained with the different methods. The results are shown for calendar years 2005 – 2009 separately and for the total. The total reserve predicts the complete lower triangle (all bold numbers + three missing cells in tables 7.4 and 7.5). The solid black line in each plot indicates what has really been observed. In the plot of the total reserve the dashed line is the sum of all observed payments in the lower triangle. This is – up to three unknown cells – the total reserve. Corresponding numerical results are in table 7.6.

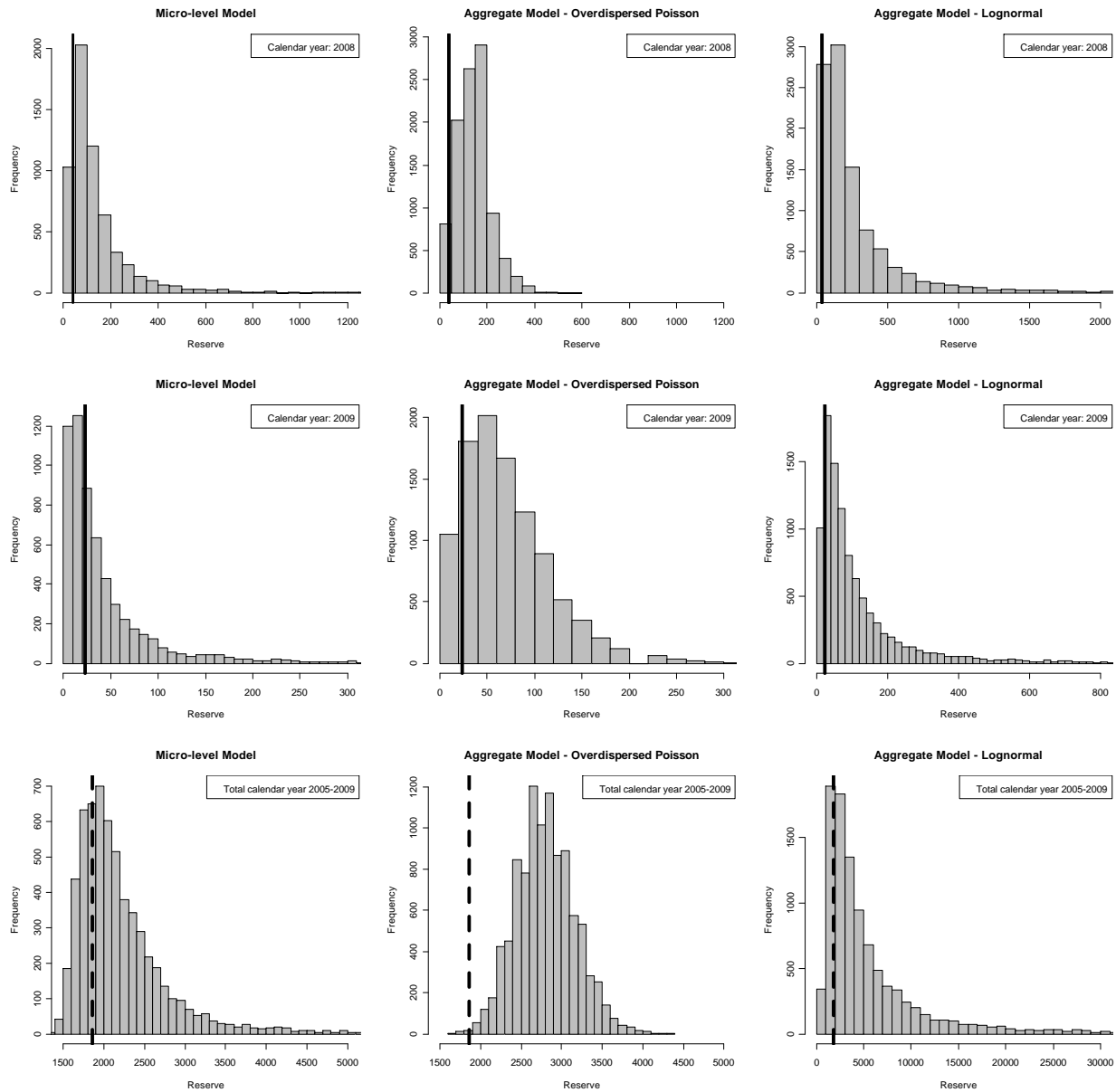*Figure 7.14: out-of-sample results – Material claims*

131
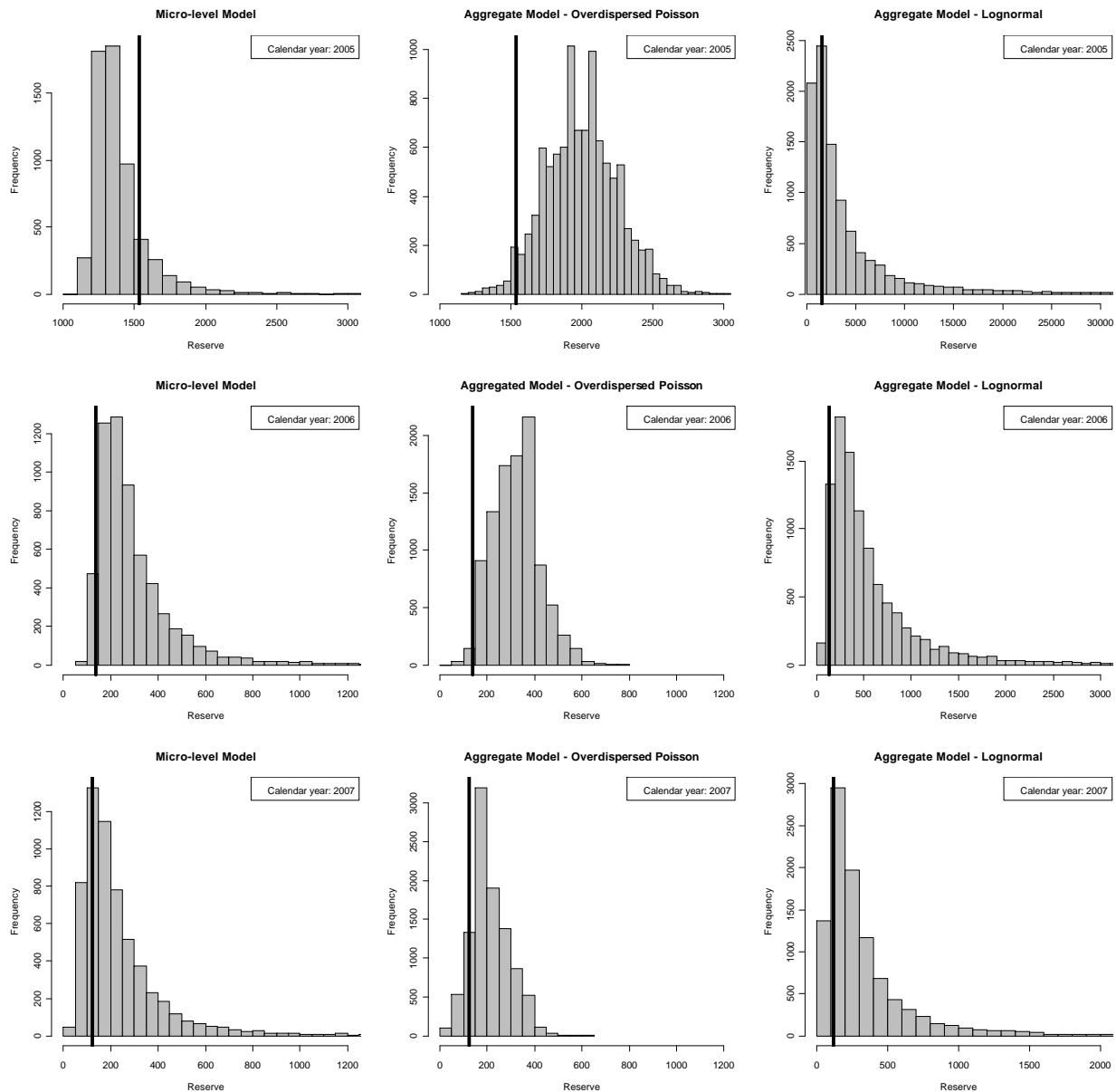
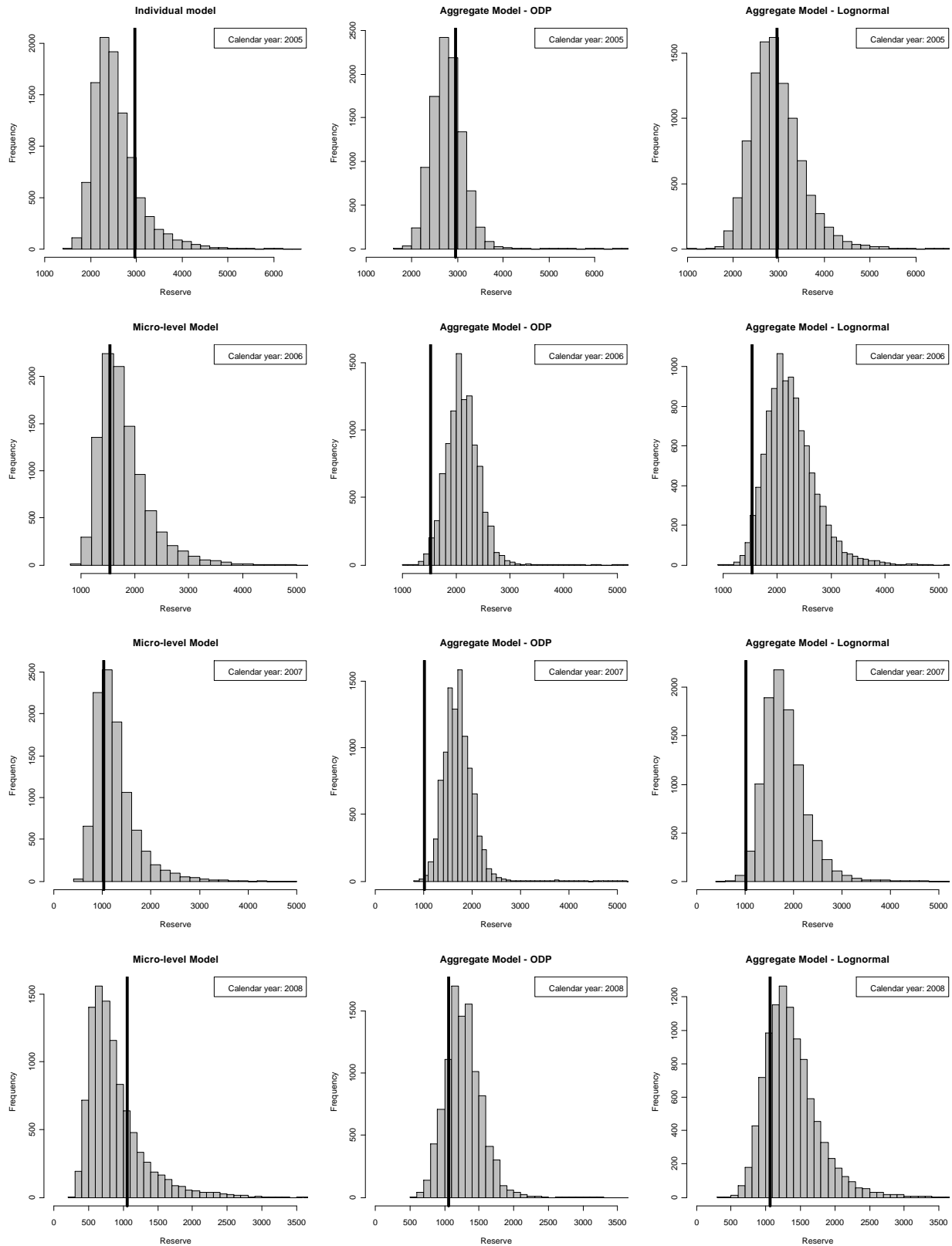*Table 7.6: out-of-sample exercise per 1-1-2005: numerical results for material claims (in thousands Euro)*

Method	Observation	Cal. Year	Mean	Median	Min.	Max.	5%	25%	75%	90%	95%	99.5%
Micro--level	1.537	2005	1.404	1.342	1.093	5.574	1.204	1.272	1.449	1.627	1.783	3.143
	139	2006	307	248	76	2.738	138	191	346	498	630	1.779
	123	2007	246	183	30	2.74	72	123	286	444	618	1.688
	39	2008	146	98	7	2.426	30	61	164	283	402	1.225
	23	2009	52	26	0	2.216	4	12	53	104	167	639
	> 1861	*Total*	*2.208*	*2.054*	*1.374*	*7.875*	*1.622*	*1.831*	*2.401*	*2.871*	*3.305*	*5.074*
Aggregate ODP	1.537	2005	2	1.989	1.194	3.028	1.591	1.834	2.166	2.321	2.431	2.674
	139	2006	324	309	44	774	177	265	376	442	486	597
	123	2007	214	199	0	619	88	155	265	332	354	464
	39	2008	144	133	0	553	44	88	177	243	265	354
	23	2009	66	66	0	376	0	22	88	133	155	243
	> 1861	*Total*	*2.803*	*2.785*	*1.613*	*4.354*	*2.232*	*2.564*	*3.028*	*3.271*	*3.426*	*3.846*
Aggregate LogN.	1.537	2005	5.340	2.253	70	587.5	497	1.146	4.896	10.79	17.985	77.671
	139	2006	699	410	32	164.2	135	254	710	1.231	1.818	6.522
	123	2007	380	228	8	23.72	67	137	403	734	1.11	3.731
	39	2008	326	167	2	48.85	41	93	317	627	998	4.053
	23	2009	163	71	1	33.66	14	36	146	304	499	2.051
	> 1861	*Total*	*7.071*	*3.645*	*201*	*645.5*	*1.11*	*2.135*	*6.936*	*13.692*	*21.931*	*84.712*

In figure 7.14 we use the same scale for plots showing reserves obtained with the micro–level and the Overdispersed Poisson model. However, for the Lognormal model a different scale on the *x*–axis is necessary because of the long right tail of the frequency histogram obtained for this model. These unrealistically high reserves (see also table 7.6) are a disadvantage of the lognormal model for the portfolio of material claims. Concerning the Poisson model for aggregate data, we conclude from figure 7.14 that the overdispersed Poisson model overstates the reserve: the actually observed amount is always in the left tail of the histogram. For instance, in the plots with the total reserve, the median of the simulations from overdispersed Poisson is at 2,785,000 euro, the median of the simulations from the micro–level model is 2,054,430 euro, whereas the total amount registered for the lower triangle is 1,861,000 euro. Recall that the latter is the total reserve up to the three unknown cells in table 7.4.

The best estimates (see the 'Mean' and 'Median' columns) obtained with the micro-level model are realistic and closer to the true realizations than the best estimates from aggregate techniques.

Figure 7.15 shows the total payments (in thousands Euro) for the different methods for injury claims. Once again the actual payments are indicated with a solid black line. The results of the log-linear model are now presented on a similar scale as the other two models. Corresponding numerical results are in table 7.7.

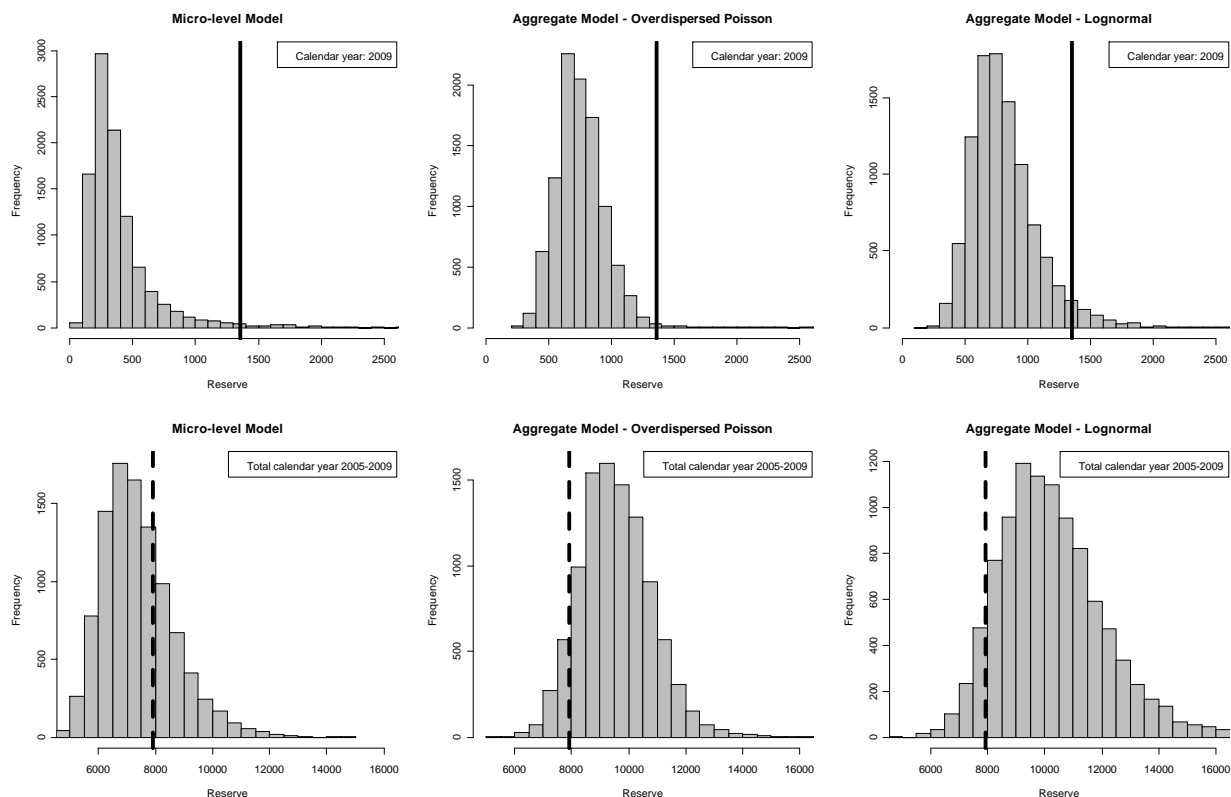*Figure 7.15: out-of-sample results – Injury claims*

Micro-level Model — Calendar year: 2009

Aggregate Model - Overdispersed Poisson — Calendar year: 2009

Aggregate Model - Lognormal — Calendar year: 2009

Micro-level Model — Total calendar year 2005-2009

Aggregate Model - Overdispersed Poisson — Total calendar year 2005-2009

Aggregate Model - Lognormal — Total calendar year 2005-2009

Table 7.7: out-of-sample exercise per 1-1-2005: numerical results for injury claims (in thousands Euro)

Method	Outcome	Cal. Year	Mean	Median	Min.	Max.	5%	25%	75%	90%	95%	99.5%
Micro--level	2.957	2005	2.548	2.453	1.569	6.587	1.951	2.212	2.764	3.154	3.499	4.567
	1532	2006	1798	1699	909	6.79	1246	1477	2001	2393	2703	3.752
	1.02	2007	1254	1159	453	4945	774	968	1.42	1778	2088	3.125
	1.06	2008	884	776	267	4.381	458	613	1024	1393	1694	2.743
	1354	2009	390	313	63	3.745	149	226	448	678	908	1875
	> 7923	*Total*	*7.386*	*7.209*	*4.209*	*14.85*	*5.666*	*6.489*	*8.092*	*9.035*	*9.721*	*11.725*
Aggregate ODP	2.957	2005	2798	2.774	1.727	8.247	2.259	2.553	2.994	3.233	3.38	4.298
	1532	2006	2134	2112	1065	6723	1.67	1929	2314	2498	2627	3472
	1.02	2007	1721	1708	845	6172	1286	1525	1892	2076	2186	3049
	1.06	2008	1286	1249	551	5933	882	1102	1433	1616	1727	2627
	1354	2009	759	735	220	4114	478	625	863	992	1084	1543
	> 7.923	*Total*	*9.639*	*9.478*	*5.474*	*40.67*	*7.66*	*8.688*	*10.36*	*11.2*	*11.77*	*17.36*
Aggregate LogN.	2.957	2005	2.948	2.882	1175	6729	2181	2.57	3.254	3648	3.944	4.944
	1532	2006	2251	2196	957	6898	1623	1.94	2.5	2.825	3.05	3.934
	1.02	2007	1817	1759	567	5313	1244	1526	2.04	2355	2583	3.426
	1.06	2008	1377	1315	374	5768	864	1.11	1571	1861	2087	2.944
	1354	2009	815	768	195	4054	472	632	941	1151	1313	1.867
	> 7.923	*Total*	*10.277*	*10.04*	*4459*	*26.01*	*7661*	*8.954*	*11.31*	*12.68*	*13.73*	*17.59*

The figure shows that for the total reserve, the distribution obtained with micro-level model seem to be more realistic than the other two models, given the actual observed realisations. All models do well for calendar year 2005, while the individual model does the best job for calendar years 2006 and 2007. For these calendar years the actual amount paid is – again – in the very left tail of the distribution obtained with aggregate techniques. The overdispersed Poisson and the Lognormal distribution perform better in calendar years 2008 and 2009. Note however that the year 2008 and 2009 were extraordinary years, when looking at injury payments. In 2009 the two highest claims of the whole data set settled with a payment in 2009. The highest (the €779.383 payment shown in table 7.1) is extremely far from all other payments in the data set. The

134

observed outcome from calendar year 2009 should be considered as a very pessimistic scenario. Indeed, this realized outcome is in the very right tail of the distribution obtained with the micro-level model. The year 2008 was less extreme, but had an unusual number of very large claims (of the 15 highest claims in the data set, 4 of them occurred in 2008).

Conclusion of the out-of-sample test is that for these case studies the reserve calculations based on the micro-level model are preferable above the traditional methods applied to aggregate data.

Note that although we only present the results obtained for the out-of-sample test that calculates the reserve per 1-1-2005, we also calculated reserves per 1-1-2006/2007/2008/2009. Our conclusions for these tests were similar to those reported above. Full details are available on the home page of the first author.

7.7 Conclusions

The measurement of future cash flows and its uncertainty becomes more and more important, also for general insurance portfolios. Currently, reserving for general insurance is based on aggregated data in run-off triangles. A vast literature about techniques for claims reserving exists, largely designed for application to run-off triangles. The most popular approach is the Chain Ladder approach, largely because of is practicality. However, the use of aggregated data in combination with the Chain Ladder approach gives rise to several issues, implying that the use of aggregate data in combination with the Chain Ladder technique (or similar techniques) is not fully adequate for capturing the complexities of stochastic reserving for general insurance.

In this chapter micro-level stochastic modeling is used to quantify the reserve and its uncertainty for a realistic general liability insurance portfolio. Stochastic processes for the occurrence times, the reporting delay, the development process and the payments are fit to the historical individual data of the portfolio and used for projection of future claims and its (estimation and process) uncertainty. A micro-level approach allows much closer modeling of the claims process. Lots of issues mentioned in our discussion of the Chain Ladder approach will not exist when using a micro-level approach, because of the availability of lots of data and the potential flexibility in modeling the future claims process.

The chapter shows that micro-level stochastic modeling is feasible for real life portfolios with over a million data records, and that it gives the flexibility to model the future payments realistically, not restricted by limitations that exist when using aggregated data. The prediction results of the micro-level model are compared with models applied to aggregate data, being an Overdispersed Poisson and a Lognormal model. We present our results through an out-of-sample, so that the estimated reserves can also be compared with actual payments. Conclusion of the out-of-sample test is that – for the case-study under consideration – traditional techniques tend to overestimate the real payments. Predictive distributions obtained with the micro-level model reflect reality in a more realistic way: 'regular' outcomes are close to the median of the predictive distribution whereas pessimistic outcomes are in the very right tail. As such, reserve calculations based on the micro-level are preferable: they reflect real outcomes in a more realistic way.

The results obtained in this chapter make it worthwhile to further investigate the use of micro–level data for reserving purposes. Several directions for future research can be mentioned. One could try to refine the performance of the individual model with respect to very pessimistic scenarios by using a combination of a lognormal distribution for losses below and a generalized Pareto distribution for losses above a certain threshold. Analyzing the performance of both the micro–level model and techniques for aggregate data on simulated data sets will bring more insight in their performance. In that respect it is our intention to collect and study new case–studies.

References

ALAI, D.H. AND M.V. WÜTHRICH (2009): Modelling small and large claims in a chain ladder framework, *Working paper*

ANDERSEN, L. AND J. ANDREASEN (1998): Volatility Skews and Extensions of the LIBOR Market Model, *Applied Mathematical Finance 7*, 1-32

ANDERSEN, L. AND R. BROTHERTON-RATCLIFFE (2001): Extended libor market models with stochastic volatility, *Working paper, Gen Re Securities*

ANDREASEN, J. (2006) Closed form pricing of FX options under stochastic rates and volatility, *Global Derivatives Conference, Paris*

ARJAS, E. (1989): The claims reserving problem in non-life insurance: some structural ideas, *ASTIN Bulletin 19,* 139-152

BAKSHI, S., C. CAO, AND Z. CHEN (1997): Empirical performance of alternative option pricing models, *Journal of Finance 52*, 2003–2049

BAKSHI, S., C. CAO, AND Z. CHEN (2000): Pricing and hedging long-term options, *Journal of Econometrics 94*, 277–318

BALLOTTA L. AND S. HABERMAN (2003): Valuation of guaranteed annuity conversion options, *Insurance: Mathematics and Economics 33*, 87–108

BATES, D.S. (1996): Jumps and stochastic volatility: exchange rate processes implicit in deutsche mark options, *The Review of Financial Studies 9*, 69-107

BAUER, D., A. KLING AND J. RUSS (2008): A universal pricing framework for guaranteed minimum benefits in variable annuities, *paper presented at AFIR Colloquium in Stockholm*

BAUER, D., M. BÖRGER, J. RUSS AND H. ZWIESLER (2008): The volatility of mortality, *Asia-Pacific Journal of Risk and Insurance 3*

BAUR, D.G (2009): Stock-bond co-movements and cross-country linkages, *Working paper*

BIFFIS, E. AND P. MILLOSSOVICH (2006): The fair value of guaranteed annuity options, *Scandinavian Actuarial Journal 1*, 23–41

BIFFIS, E., M. DENUIT, AND P. DEVOLDER (2006): Stochastic mortality under measure changes, *Pensions Institute Discussion Paper PI-0512*

BLACK, F. AND M. SCHOLES (1973): The pricing of options and corporate liabilities, *Journal of Political Economy 81(3)*

BOLTON, M.J. AND D.H. CARR, P.A. COLLIS, C.M. GEORGE, V.P. KNOWLES, AND A.J. WHITEHOUSE (1997): Reserving for annuity guarantees, *The Report of the Annuity Guarantees Working Party*

BOOTH, H., J. MAINDONALD AND L. SMITH (2002): Applying Lee-Carter under conditions of variable mortality decline, *Population Studies 56*, 325-336

BORNHUETTER, R.L. AND R.E. FERGUSON (1972): The actuary and IBNR, *Proc. CAS LIX*, 181-195

BOX, G.E.P. AND G.M. JENKINS (1976): Time series analysis: forecasting and control, revised

edition, *Holden-Day San Fransico*

BOYLE, P.P. AND M. HARDY (2001): Mortality derivatives and the option to annuitize, *Insurance: Mathematics and Economics 29(3)*

BOYLE, P.P. AND M. HARDY (2003): Guaranteed annuity options, *Astin Bulletin 33(2)*, 125–152

BRIGO, D. AND F. MERCURIO (2006): Interest Rate Models – Theory and Practice, 2nd edition, *Springer-Verlag*

BROUHNS, N., M. DENUIT AND J.K. VERMUNT (2002): Measuring the longevity risk in mortality projections, *Bulletin of the Swiss Association of Actuaries 2*, 105-130

BROUHNS, N., M. DENUIT, AND I. VAN KEILEGOM (2005): Bootstrapping the Poisson log-bilinear model for mortality forecasting, *Scandinavian Actuarial Journal 3*, 212-224

BROUHNS, N., M. DENUIT, AND J.K. VERMUNT (2002): A Poisson log-bilinear regression approach to the construction of projected life tables, *Insurance: Mathematics and Economics 31,* 373-393

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., D. BLAKE, AND K. DOWD (2006a): 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., D. BLAKE, AND K. DOWD (2006b): Pricing death: Frameworks for the valuation and securitization of mortality risk, *ASTIN Bulletin 36*, 79-120

CAIRNS, A.J.G., D. BLAKE, AND K. DOWD (2008a): Modelling and Management of Mortality Risk: A Review, *Working paper, Heriot-Watt University, and Pensions Institute Discussion Paper PI-0814*

CAIRNS, A.J.G., D. BLAKE, K. DOWD, G. D. COUGHLAN, D. EPSTEIN AND M. KHALLAF-ALLAH (2008b): The plausibility of mortality density forecasts: an analysis of six stochastic mortality models, *Working paper, Heriot-Watt University, and Pensions Institute Discussion Paper PI-0801*

CAIRNS, A.J.G., D. BLAKE, K. DOWD, G. D. COUGHLAN, D. EPSTEIN, A. ONG AND I. BALEVICH (2007): A quantitative comparison of stochastic mortality models using data from England & Wales and the United States, *Working paper, Heriot-Watt University, and Pensions Institute Discussion Paper PI-0701*

CARR, P. AND D. B. MADAN (1999): Option Valuation Using the Fast Fourier Transform, *Journal of Computational Finance 2*, 61-73

CASTELLANI, G ET AL (2007): Pricing Formulae for Financial Options and Guarantees Embedded in Profit Sharing Life Insurance Policies, *Working Paper*

CEIOPS (2007): QIS 4 Technical Specifications

CEIOPS (2010): QIS 5 Technical Specifications

CHU, C.C. AND Y. K. KWOK (2007): Valuation of guaranteed annuity options in affine term structure models, *International Journal of Theoretical and Applied Finance, 10(2)*, 363–387

COLLIN-DUFRESNE, P. AND R.S. GOLDSTEIN (2002): Pricing swaptions within an affine framework, *Journal of Derivatives, Fall issue*, 1–18

CONTINUOUS MORTALITY INVESTIGATION (1962): Continuous investigation into the mortality of pensioners under life office pension schemes, *available at:*
http://www.actuaries.org.uk/knowledge/cmi

CONTINUOUS MORTALITY INVESTIGATION (2004): *Working paper 9, available at:*
http://www.actuaries.org.uk/knowledge/cmi

CONTINUOUS MORTALITY INVESTIGATION (2007): Stochastic projection methodologies: Lee-

Carter model features, example results and implications, *Working paper 25*

CONTINUOUS MORTALITY INVESTIGATION (2008): *Working paper 31, available at: http://www.actuaries.org.uk/knowledge/cmi*

COOK, R. AND J. LAWLESS (2007): The statistical analyses of recurrent events, *Springer New York*

COUGHLAN, G., D. EPSTEIN, A. ONG, A. SINHA, I. BALEVICH, J. HEVIA-PORTOCARRERO, E. GINGRICH, M. KHALAF-ALLAH AND P. JOSEPH (2007): Lifemetrics Technical Document, *available at: http://www.jpmorgan.com/pages/jpmorgan/investbk/solutions/lifemetrics*

COX, D.R. (1972): Regression models and life tables (with discussion), *Journal of the Royal Statistical Society 34*, 187-220

COX, J.C., J.E. INGERSOLL AND S.A. ROSS (1985): A theory of the term structure of interest rates, *Econometrica 53*, 385-407

CURRAN, M (1994): Valuing Asian and Portfolio Options by Conditioning on the Geometric Mean Price, *Management Science 40*, 1705-1711

CURRIE, I.D. (2006): Smoothing and forecasting mortality rates with P-splines, *Talk given at the Institute of Actuaries, June 2006, available at: http:www.ma.hw.ac.uk/~iain/research/talks.html*

CURRIE, I.D., M. DURBAN, AND P.H.C. EILERS (2004): Smoothing and forecasting mortality rates, *Statistical Modelling 4*, 279-298

DAHL, M. (2004): Stochastic mortality in life insurance: Market reserves and mortality-linked insurance contracts, *Insurance: Mathematics and Economics 35*, 113-136

DAHL, M., AND T. MØLLER (2006): Valuation and hedging of life insurance liabilities with systematic mortality risk, *Insurance: Mathematics and Economics 39*, 193-217

DASSIOS, A. AND J. NAGARADJASARMA (2006): The square-root process and Asian options, *Quantitative Finance 6*, 337-347

DE JONG, P., AND L. TICKLE (2006): Extending the Lee-Carter model of mortality projection, *Mathematical Population Studies 13*, 1-18

DELWARDE, A., M. DENUIT AND P. EILERS (2007): Smoothing the Lee-Carter and Poisson log-bilinear models for mortality forecasting: A penalized log-likelihood approach, *Statistical Modelling 7*, 29-48

DIEBOLD, F.X., AND C. LI (2006): Forecasting the term structure of government bond yields, *Journal of Economics 130*, 337-364

DUFFIE, D. AND R. KAN (1996): A Yield-Factor Model of Interest Rates, *Mathematical Finance 6*, 379-406

DUFFIE, D., D. FILIPOVIC, AND W. SCHACHERMAYER (2003): Affine processes and applications in finance, *Annals of Applied Probability 13*

DUFFIE, D., J. PAN, AND K. SINGLETON (2000): Transform analysis for affine jump diffusions, *Econometrica 68*, 1343-1376

DUNBAR, N.(1999): Sterling swaptions under new scrutiny, *Risk (12)*

ENDERS, W. (2004): Applied econometric time series, 2nd edition, *John Wiley and Sons*

ENGLAND, P.D. AND R.J. VERRALL (1999): Analytic and bootstrap estimates of prediction errors in claims reserving, *Insurance: Mathematics and Economics 25,* 281-293

ENGLAND, P.D. AND R.J. VERRALL (2002): Stochastic claims reserving in general insurance, *Britisch Actuarial Journal 8*, 443-518

GATAREK, D (2003).: Constant maturity swaps, forward measure and LIBOR market model, *Available at SSRN: http://ssrn.com/abstract=394201*

GEMAN, H., E. KAROUI, AND J.C. ROCHET (1995): Changes of numéraire, changes of probability measures and pricing of options, *Journal of Applied Probability 32*, 443–548

GLASSERMAN, P. (2004): Monte Carlo Methods in Financial Engineering, *Springer-Verlag*

GROSEN, A. AND P. JORGENSEN (2000), Fair Valuation of Life Insurance Liabilities: The Impact of Guarantees, Surrender Options and Bonus Policies, *Insurance, Mathematics and Economics 26*, 37-57

HAASTRUP, S. AND E. ARJAS (1996): Claims reserving in continuous time: a nonparametric Bayesian approach, *ASTIN Bulletin 26,* 139-164

HALLIWELL, L.J. (2007): Chain-ladder bias: its reason and meaning, *Variance 1*, 214-247

HALLIWELL, L.J. (2009): Modeling paid and incurred losses together, *Casualty Actuarial Society E-forum spring 2009*

HAMILTON, J.D. (1994): time series analysis, *Princeton University Press*

HANDEL, R. (2007): Stochastic Calculus, Filtering, and Stochastic Control, *Lecture Notes*

HARDY, M.R. (2001): A regime-switching model of long-term stock returns, *North American Actuarial Journal*

HARDY, M.R.(2004): Ratchet Equity Indexed Annuities, *AFIR Colloquium, Boston*

HARRISON, J.M. AND D. KREPS (1979): Martingale and arbitrage in multiperiod securities markets, *Journal of Economic Theory 20*, 381-408

HARRISON, J.M. AND S. PLISKA (1981): Martingales and stochastic integrals in the theory of continuous trading*, Stochastic processes and their applications 11*, 215-260

HESTON, S.L. (1993): A closed-form solution for options with stochastic volatility with applications to bond and currency options, *Review of Financial Studies 6*, 327–343

HULL, J. AND A. WHITE (1993): One factor interest rate models and the valuation of interest rate derivative securities, *Journal of Financial and Quantitative Analysis 28(2)*

HULL, J.C. (2006): Options, Futures, and Other Derivatives, 6th edition, *Prentice Hall*

JAMSHIDIAN, F. (1989): An exact bond option pricing formula, *Journal of Finance 44*, 205-209

JAMSHIDIAN, F. (1998): LIBOR and Swap Market Models and Measures, *Finance and Stochastics 1*, 293-330

JOLLIFFE, I.T. (2002): Principal Component Analysis, *Springer-Verlag New York, Inc.*

KAAS, R., M. GOOVAERTS, J. DHAENE AND M. DENUIT (2008): Modern actuarial risk theory, *Springer-Verlag*

KARR, A.F. (1991): Point processes and their statistical inference, 2nd edition, *Marcel Dekker INC*

KLEIBERGEN, F., AND H. HOEK (1996): Bayesian Analysis of ARMA Models using Noninformative Priors, *Tinbergen Institute discussion paper*

KUNKLER, M. (2004): Modelling zeros in stochastic reserving models, *Insurance: Mathematics and Economics 34,* 23-35

LARSEN, C.R. (2007): An individual claims reserving model, *ASTIN Bulletin 37,* 113-132

LEE, R., AND T. MILLER (2001): Evaluating the performance of the Lee-Carter model for forecasting mortality, *Demography 38*, 537-549

LEE, R.D., AND L.R. CARTER (1992): Modelling and forecasting U.S. mortality, *Journal of the American Statistical Association 87*, 659-675

LEVIN, A. (2004): Interest Rate Model Selection, *The Journal of Portfolio Management*, 74-86

LEVY, E. (1992): Pricing European average rate currency options, *Journal of International Money and Finance 11*, 474-491

LIU, H. AND R. VERRALL (2009): Predictive distributions for reserves which separate true IBNR and IBNER claims, *Working paper*

LOEYS, J., N. PANIGIRTZOGLOU AND R.M. RIBEIRO (2007): Longevity: a market in the making,

available at: http://www.jpmorgan.com/pages/jpmorgan/investbk/solutions/lifemetrics

LONGSTAFF, F.A. AND E.S. SCHWARTZ (1992): Interest rate volatility and the term structure: a two-factor general equilibrium model, *Journal of finance XLVII (4)*, 1259-1282

LORD, R (2006): Partially exact and bounded approximations for arithmetic Asian options, *Journal of Computational Finance 10*, 1-52

MACK, T. (1999): The standard error of chain ladder reserve estimates: recursive calculation and the inclusion of a tail factor, *ASTIN Bulletin 29,* 361-366

MARSHALL, C., M.R. HARDY, AND D. SAUNDERS (2009): Static hedging strategies for guaranteed minimum income benefits, *North American Actuarial Journal 14*, 38-58

MILEVSKY, M.A., AND S.D. PROMISLOW (2001): Mortality derivatives and the option to annuitize, *Insurance: Mathematics and Economics 29*, 299-318

MILLS, T.C. (1990): Time series techniques for economists, *Cambridge University Press*

MOORE, K.S. AND V.R. YOUNG (2003): Pricing equity-linked pure endowments via the principle of equivalent utility, *Insurance: Mathematics and Economics 33*, 497-516

NAMBOODIRI, K. AND C.M. SUCHINDRAN (1987): Life Table Techniques and Their Applications, *Academic Press, Inc.*

NELSON, C.R., AND A.F. SIEGEL (1987): Parsimonious modeling of yield curve, *Journal of Business 60*, 473-489

NIELSEN, J.A. AND K. SANDMANN (2002): The fair premium of an equity-linked life and pension insurance, *In: Schönbucher, P. and K. Sandmann (Eds.). Advances in Finance and Stochastics: essays in Honor of Dieter Sondermann, Springer-Verlag*

NORBERG, R. (1993): Prediction of outstanding liabilities in non-life insurance, *ASTIN Bulletin 23,* 95-115

NORBERG, R. (1999): Prediction of outstanding liabilities ii. Model variations and extensions, *ASTIN Bulletin 29,* 5-25

OKSENDAL, B. (2005): Differential Equations: An Introduction with Applications. *Springer-Verlag*

PELSSER, A.A.J. (2003): Pricing and hedging guaranteed annuity options via static option replication, *Insurance, Mathematics and Economics 33,* 283-296

PELSSER, A.A.J. (2004): Efficient Methods for Valuing Interest Rate Derivatives, 2nd printing, *Springer-Verlag*

PIETERSZ, R. AND M. VAN REGENMORTEL (2006): Generic market models, *Finance and Stochastics 10*, 507-528

POSTHUMA, B., E.A. CATOR, W. VEERKAMP AND E.W. VAN ZWET (2008): Combined analysis of paid and incurred losses, *Casualty Actuarial Society E-forum fall 2008*

PRIEUL, D., ET AL (2001): On pricing and reserving with-profits life insurance contracts, *Applied Mathematical Finance 8*, 145-166

QUARG, G. AND T. MACK (2008): Munich Chain Ladder: a reserving method that reduces the gap between IBNR projections based on paid losses and IBNR projections based on incurred losses, *Variance 2,* 266-299

RENSHAW, A.E. (1994): Claims reserving by joint modelling, *Actuarial Research Paper 72, Department of Actuarial Sciences and Statistics, City University, London*

RENSHAW, A.E., AND S. HABERMAN (2003): Lee-Carter mortality forecasting with age-specific enhancement, *Insurance: Mathematics and Economics 33*, 255-272

RENSHAW, A.E., AND S. HABERMAN (2006): A cohort-based extension to the Lee-Carter model for mortality reduction factors, *Insurance: Mathematics and Economics 38,* 556-570

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

ROGERS, L.C.G. AND Z. SHI (1995): The value of an Asian option, *Journal of Applied Probability 32*, 1077-1088

SCHNIEPER, R. (1991): Separating true IBNR and IBNER claims, *ASTIN Bulletin 21,* 111-127

SCHÖBEL AND J. ZHU (1999): Stochastic volatility with an Ornstein-Uhlenbeck process: An extension, *European Finance Review 4*, 23–46

SCHRAGER D.F. AND A.A.J. PELSSER (2004): Pricing Rate of Return Guarantees in Regular Premium Unit Linked Insurance, *Insurance: Mathematics and Economics 35*, 369-398

SCHRAGER D.F. AND A.A.J. PELSSER (2006): Pricing swaptions and coupon bond options in affine term structure models, *Mathematical Finance 16*, 673-694

SCHRAGER, D.F. (2006): Affine stochastic mortality, *Insurance: Mathematics and Economics 38*, 81-97

SHELDON, T.J. AND A.D. SMITH (2004): Market consistent valuation of life assurance business, *Britisch Actuarial Journal 10*, 543-605

STEIN, E.M. AND J.C. STEIN (1991): Stock-price distributions with stochastic volatility: an analytic approach, *Review of Financial Studies 4*, 727–752

TAYLOR, G. (2003): Chain ladder bias, *ASTIN Bulletin 33*, 313-330

TAYLOR, G. AND M. CAMPBELL (2002): Statistical case estimation, *Research paper 104, The University of Melbourne, Australia 1*

VAN BEZOOYEN, J.T.S., C.J.E. EXLEY, AND S.J.B. MEHTA (1998): Valuing and hedging guaranteed annuity options

VAN BROEKHOVEN, H. (2002): Market Value of Liabilities Mortality Risk: A Practical Model, *North American Actuarial Journal 6,* 95-106

VAN DER PLOEG, A.P.C. (2006): Stochastic Volatility and the Pricing of Financial Derivatives, PhD thesis, *Tinbergen institute/ University of Amsterdam*

VAN HAASTRECHT, A., R. LORD, A.A.J. PELSSER, AND D. SCHRAGER (2009): Pricing long-dated insurance contracts with stochastic volatility and stochastic interest rates, *Insurance: Mathematics and Economics*

VERBEEK, H.G. (1972): An approach to the analysis of claims experience in motor liability excess of loss reinsurance, *ASTIN Bulletin 6*, 195-202

VERBEEK, M. (2008): Modern Econometrics, 3th edition, *John Wiley & Sons, Ltd*

VERBOND VAN VERZEKERAARS (2008): Generatietafels Pensioenen 2008, *publication in dutch*

VERDONCK, T., M. VAN WOUWE AND J. DHAENE (2009): A robustification of the chain-ladder method, *North American Actuarial Journal 13,* 280-298

WILKIE, A.D., H.R, WATERS AND S. YANG (2003): Reserving, pricing and hedging for policies with guaranteed annuity options, *Britisch Actuarial Journal 9*, 263-391

WILLETS, R.C. (2004): The cohort effect: Insights and explanations, *British Actuarial Journal 10*, 833-877

WRIGHT, T.S. (1990): A stochastic method for claims reserving in general insurance, *J. Institute Actuaries 117*, 677-731

WÜTHRICH, M.V. AND M. MERZ (2008): Stochastic claims reserving methods in insurance, *Wiley Finance*

WYMAN, O. (2007): Va va voom, *http://www.mmc.com/knowledgecenter/ OliverWymanVariableAnnuities.pdf*

YANG, S. (2001): Reserving, Pricing and Hedging for Guaranteed Annuity Options, PhD thesis, *Department of Actuarial Mathematics and Statistics, Heriot Watt University*

YOUNG, V.R. (2004): Pricing in an incomple market with an affine term structure, *Mathematical Finance 14 (3)*, 359-381

YOUNG, V.R. AND T. ZARIPHOPOULOU (2002): Pricing dynamic insurance risks using the principle of equivalent utility, *Scandinavian Actuarial Journal 4*, 246-279

ZEHNWIRTH, B. (1994): Probabilistic development factor models with application to loss reserve variability, prediction intervals, and risk based capital, *Casualty Actuarial Society*

ZELLNER, A. (1963): Estimators of Seemingly Unrelated Regressions: Some Exact Finite Sample Results, *Journal of the American Statistical Association 58*, 977-992

ZHAO, X.B. AND X. ZHOU (2010): Applying copula models to individual claim loss reserving methods, *Insurance: Mathematics and Economics 46(2),* 290-299

ZHAO, X.B., X. ZHOU AND J.L. WANG (2009): Semiparametric model for prediction of individual claim loss reserving, *Insurance: Mathematics and Economics 45(1),* 1-8

Samenvatting (Summary in Dutch)

Individuen, bedrijven en andere entiteiten staan bloot aan verschillende soorten risico's die kunnen leiden tot ongewenste financiële consequenties. Een individu kan bijvoorbeeld schade hebben aan zijn of haar auto, huis of inboedel, kan langer of korter leven dan verwacht, of onverwacht hoge kosten maken in verband met de gezondheid. Bedrijven kunnen blootstaan aan schades veroorzaakt door bijvoorbeeld een schadeclaim, verbranding van een bedrijfsgebouw, schade aan goederen en arbeidsongeschikte werknemers. Deze risico's kunnen worden overgedragen door het aangaan van een verzekeringscontract bij een verzekeringsmaatschappij. In ruil hiervoor vraagt de verzekeringsmaatschappij een premie van de 'polishouder'. De verzekeringsmaatschappij brengt alle individuele risico's samen waardoor de resultaten op individuele polissen elkaar compenseren.

Het resultaat van het vele jaren verkopen van verzekeringen is dat verzekeraars in de toekomst nog aanzienlijke bedragen moeten betalen aan hun polishouders (bijvoorbeeld hun pensioen). Verzekeraars houden hiervoor een reserve aan, welke is gebaseerd op de waardering van deze toekomstige verzekeringsverplichtingen. Daarnaast staat de verzekeraar bloot aan verschillende risico's, waarvoor het additioneel kapitaal aanhoudt. Derhalve zijn correcte waardering van verzekeringsverplichtingen en het meten en managen van risico's twee belangrijke voorwaarden voor het succesvol runnen van een verzekeringsbedrijf. Dit proefschrift is een bundeling van artikelen over verschillende kwesties gerelateerd aan waardering en risicomanagement voor verzekeraars.

In het vervolg van deze samenvatting wordt meer context gegeven over waardering en risicomanagement voor verzekeraars, gevolgd door een korte behandeling van de verschillende artikelen.

Context

Momenteel stellen verzekeraars de waarde van hun verplichtingen vast op basis van 'boekwaarde', wat inhoudt dat de economische aannames meestal niet direct afgeleid zijn uit de financiële markten. Daarnaast verplicht de toezichthouder verzekeraars additioneel (solvabiliteits)kapitaal aan te houden. Dit kapitaal wordt bepaald als vast percentage van de reserve, premies of schades en is dus niet direct gebaseerd op het risicoprofiel van de verzekeraar. Echter, de laatste jaren is er een toenemende aandacht van de verzekeringsindustrie voor marktwaardering van de verzekeringsverplichtingen en het kwantificeren van verzekeringsrisico's. Belangrijke redenen hiervoor zijn de komende introducties van IFRS 4 Fase 2 en Solvency 2.

De introductie van Solvency 2 en IFRS 4 Fase 2 (beiden in 2013) stelt verzekeraars voor een grote uitdaging. IFRS 4 fase 2 zal een nieuw accounting model voor verzekeringscontracten definiëren, gebaseerd op marktwaardering van de verplichtingen. In het document 'Preliminary Views on Insurance Contracts' (Mei 2007) stelt de 'International Accounting Standards Board' (IASB) dat verzekeraars de waardering van hun verplichtingen moeten baseren op zo actueel en juist mogelijke inschattingen van toekomstige kasstromen, gedisconteerd met de actuele rentes uit de markt. Verder wordt verwacht dat verzekeraars additioneel een risicomarge opnemen. De IASB is de principes momenteel verder aan het uitwerken.

Solvency 2 zal leiden tot een verandering in de eisen van de toezichthouder wat betreft het additioneel aan te houden solvabiliteitskapitaal. Onder Solvency 2 zal de kapitaalseis risico-gebaseerd zijn, en marktwaardering van beleggingen en verplichtingen vormt de basis hiervoor. De kapitaalseis zal alle risico's moeten dekken waaraan een verzekeraar blootstaat: marktrisico, operationeel risico, risico's van leven- en pensioen producten, risico's van schade en zorg producten, tegenpartij risico en het risico van overige bezittingen. Binnen Solvency 2 is een standaard formule ontwikkeld die leidt tot een kapitaalseis die erop gericht is om de risico's voor 1 jaar te dekken met een 99,5% betrouwbaarheid. Echter, verzekeraars worden gestimuleerd om hun eigen interne modellen te ontwikkelen om zodoende de specifieke risico's van de verzekeraar beter in te kunnen schatten.

Gegeven bovenstaande is de conclusie dat het meten van toekomstige kasstromen en de onzekerheid hiervan steeds belangrijker wordt voor de verzekeringsindustrie.

Indeling proefschrift

In dit proefschrift zijn enkele artikelen gebundeld op het gebied van waardering van verzekeringsverplichtingen en risicomanagement voor verzekeraars. Eerst worden in hoofdstuk 2 de algemene concepten toegelicht die gebruikt worden in dit proefschrift, met name gerelateerd aan stochastische processen.

Leven- en pensioen producten bevatten vaak een vorm van winstdeling in combinatie met een garantie. Waardering van deze zogenoemde 'embedded opties' is een van de grootste uitdagingen bij marktwaardering voor verzekeraars. Hoofdstuk 3 en 4 behandelen beiden de waardering van specifieke embedded opties.

Belangrijke risico's bij leven- en pensioenverzekeraars zijn het 'langlevenrisico' (het risico dat mensen langer leven dan verwacht) en het 'kortlevenrisico' (het risico dat mensen korter leven dan verwacht). Hoofdstuk 5 en 6 behandelen verschillende aspecten in het kwantificeren van deze risico's.

Hoofdstuk 7 behandelt de risico's van schadeproducten. In dit hoofdstuk wordt een nieuwe techniek gepresenteerd om de waarde van de verplichtingen (en de onzekerheid daarvan) te kwantificeren.

In de volgende secties worden de hoofdstukken afzonderlijk toegelicht.

Hoofdstuk 3: Waardering van swap-afhankelijke embedded opties

Veel verzekeringsproducten kennen een vorm van winstdeling in combinatie met een garantie. Deze zogenoemde embedded opties zijn vaak afhankelijk van of worden geschat middels (forward) swaprentes. De swaprente is het tarief waartegen op de geld- en kapitaalmarkt leningen met verschillende looptijden worden geruild. Normaal gesproken worden deze opties berekend door middel van Monte Carlo simulatie, een computerintensieve berekeningstechniek. Echter, voor risicomanagement en rapportage processen zijn vele waarderingen benodigd. Daarom zou een meer efficiënte berekeningstechniek welkom zijn.

In dit hoofdstuk worden (benaderende) analytische formules ontwikkeld voor deze klasse van embedded opties. De analytische formule voor directe betaling van winstdeling is vrijwel exact en de benadering voor cumulatieve winstdelingsbetalingen is ook voldoende. Daarnaast kunnen de formules gebruikt worden als 'control variate' bij Monte Carlo simulatie, wat de berekeningstijden van Monte Carlo simulatie significant verlaagd. Dit kan van pas komen bij meer complexe embedded opties waarvoor geen analytische formules bestaan. Tot slot kan de formule ook uitgebreid worden voor het geval waar de winstdeling mede afhankelijk is van het rendement op aandelen.

Hoofdstuk 4: Waardering van Gegarandeerde Annuïteit Opties gebruik makend van een model met stochastische volatiliteit voor aandelenprijzen

Een Gegarandeerde Annuiteit Optie (GAO) is een optie die een polishouder het recht biedt om het op de pensioendatum opgebouwde kapitaal om te zetten naar een levenslange lijfrente tegen een vaste rente. Deze embedded optie was een standaard onderdeel van pensioencontracten in het Verenigd Koninkrijk in de jaren '70 en '80 toen het renteniveau hoog lag. Echter, deze opties zorgden voor problemen toen de rente begon te dalen in de jaren '90. Momenteel worden deze opties nog veelvuldig verkocht in de Verenigde Staten en Japan.

Het laatste decennium is de literatuur over waardering en risicomanagement voor deze opties sterk uitgebreid. Tot op dit moment is er bij de waardering vooral uitgegaan van een proces voor aandelenprijzen waarbij de volatiliteit constant is. Echter, gegeven de lange looptijd van deze contracten en de observatie uit het verleden dat de volatiliteit niet constant is, is een model met stochastische volatiliteit te prefereren. In dit hoofdstuk zijn expliciete formules bepaald voor prijzen van GAO's, gebruik makend van een model met stochastische volatiliteit voor aandelenprijzen en een stochastisch model voor rentes. De resultaten wijzen uit het meenemen van stochastische volatiliteit een grote impact heeft op de prijsstelling en het risicomanagement van deze opties.

Hoofdstuk 5: Over stochastische modellering van sterftekansen

Het laatste decennium heeft er een grote toename van de literatuur over stochastische modellen voor sterftekansen plaatsgevonden, met name voor gebruik in risico management. Alle bekende modellen hebben voordelen en nadelen. In dit hoofdstuk wordt een nieuw stochastisch sterftemodel voorgesteld die de goede eigenschappen van bestaande modellen combineert, en waarbij de nadelen van bestaande modellen niet meer voorkomen. Meer concreet, het model sluit goed aan bij de historische waarnemingen van sterftekansen, is bruikbaar voor alle leeftijden, adresseert ook effecten die specifiek voor geboortejaren gelden, modelleert de samenhang tussen leeftijden adequaat en heeft geen robuustheid problemen. Ook is beschreven hoe parameteronzekerheid kan worden meegenomen. Tot slot is ook een versie van het model gegeven die gebruikt kan worden voor waardering.

Hoofdstuk 6: Stochastische portefeuille specifieke sterfte en het kwantificeren van sterfte basis risico

In hoofdstuk 5 zijn een aantal stochastische sterftemodellen beschreven, veelal toegepast op bevolkingssterfte. Echter, deze modellen zijn meestal niet direct toe te passen op verzekeringsportefeuilles, omdat:
   a) het voor verzekeraars en pensioenfondsen relevanter is om sterftekansen te meten in bedragen in plaats van aantallen.
   b) er vaak niet voldoende data beschikbaar is van de historische sterftekansen van de specifieke portefeuille van verzekeraars.
Om deze reden wordt in dit hoofdstuk een stochastisch model voorgesteld voor portefeuille specifieke ervaringssterfte. Combinatie van dit stochastische proces met een stochastisch model voor bevolkingsterfte resulteert in stochastische portefeuillespecifieke sterftekansen, gemeten in bedragen. Het stochastische proces is getest op twee voorbeeld portefeuilles, en de impact op de hoogte van het langlevenrisico is gekwantificeerd. Daarnaast kan het model ook gebruikt worden voor het kwantificeren van sterfte basis risico. Dit is het risico dat overblijft als portefeuille specifieke sterfte door een verzekeraar afgedekt wordt met instrumenten waarvan de betalingen afhangen van bevolkingssterfte.

Hoofdstuk 7: Stochastische schadereservering op micro-niveau

Er heeft zich een substantiële literatuur ontwikkeld over stochastische schadereservering. Echter, vrijwel alle literatuur is gebaseerd op technieken die toegepast worden op een zogenaamde 'schadedriehoek' met geaggregeerde data. Echter, deze geaggregeerde data is een samenvatting van een onderliggende, veel gedetailleerdere database die beschikbaar is binnen verzekeraars. Deze data op het niveau van individuele schades wordt micro-niveau data genoemd. In dit hoofdstuk is onderzocht of het gebruik van data op micro-niveau de kwaliteit van schadereservering kan verbeteren. Een realistische dataset op micro-niveau van een aansprakelijkheidsportefeuille van een Europese verzekeraar is daarvoor gebruikt. Stochastische processen zijn gespecificeerd voor de verschillende onderdelen in de ontwikkeling van een schade: de tijd van plaatsvinden van de schade, de vertraging tussen het plaatsvinden van de

schade en het op de hoogte stellen van de verzekeraar, eventuele betalingen en de hoogte ervan en de afsluiting van de schade. De parameters behorende bij deze processen worden geschat op basis van de historische data van de portefeuille en worden gebruikt voor de projectie van toekomstige betalingen. Een 'out-of-sample' exercitie toont aan dat de voorgestelde aanpak de actuaris voorziet van gedetailleerde en waardevolle berekeningen van de reserve. Een vergelijking met traditionele reservering technieken is ook gemaakt. Voor het voorbeeld gebruikt in dit hoofdstuk is het voorgestelde model te prefereren: de resultaten zijn realistischer en sluiten beter aan bij de historische observaties.