Risk Decomposition for Annuity Portfolios

Chengrong Xie

A Thesis

for The Department of

Mathematics and Statistics

Presented in Partial Fulfillment of the Requirements

for the Degree of Master of Science (Mathematics) at

Concordia University

Montréal, Québec, Canada

July 2017

CONCORDIA UNIVERSITY

School of Graduate Studies

This is to certify that the thesis prepared

By: **Chengrong Xie**

Entitled: **Risk Decomposition for Annuity Portfolios**

and submitted in partial fulfillment of the requirements for the degree of

Master of Science (Actuarial Mathematics)

complies with the regulations of the University and meets the accepted standards with respect to originality and quality.

Signed by the final examining committee:

_____ Examiner

Dr. M. Mailhot

_____ Examiner

Dr. F. Godin

_____ Thesis Supervisor

Dr. P. Gaillardetz

Approved by _____

Chair of Department or Graduate Program Director

_____

Dean of Faculty

Date _____

ABSTRACT

Risk Decomposition for Annuity Portfolios

by Chengrong Xie

A life annuity is a series of payments made at fixed intervals while the annuitant is alive. It has been a major part of actuarial science for a long time and it plays an important role in life insurance operations. In order to explore the interaction of various risks in an annuity portfolio, we decompose the liabilities by using the so called Martingale Representation Theorem (MRT) decomposition. The MRT decomposition satisfies all 6 meaningful properties proposed by Schilling et al. (2015).

Before presenting some numerical examples to illustrate its applicability, several stochastic mortality models are compared and the Renshaw–Haberman (RH) model is chosen as our projection model. Then we compare two one-factor short rate models and estimate the parameters of CIR model to construct the stochastic interest rate setting. Finally, we allocate risk capitals to risk factors obtained from the MRT decomposition according to the Euler principle and analyze them when the age of cohort and the deferred term change.

Acknowledgments

First of all, I would like to express my sincere gratitude to my supervisor, Dr. Patrice Gaillardetz, for his support and insightful guidance. His patience and expertise have been playing an important role in my graduate studies.

I would also like to take this opportunity express a deep gratitude to my family, especially my parents, for their nurturing, support and encouragement in my life. Further, I would like to thank my thesis committee; Professor Mélina Mailhot and Professor Frédéric Godin, for taking the time to read and providing valuable comments. I am also grateful to the financial and academic support provided by the Math and Statistics department, with which I was able to initiate my study in Canada.

Last but not least, I am very grateful to my friends for their help and encouragement throughout the entire process both in academy and life.

Contents

List of Figures

List of Tables

Introduction

A life annuity is a series of payments made at fixed intervals while the annuitant is alive. There are many forms of life annuities. It may be limited to a given term of years, or it may be payable for the whole life. An annuity can have a deferral phase and an annuitization phase in which the insurance company actually makes payments. Otherwise, the payment intervals may commence immediately. There also exists fixed and variable annuities. Annuities that make payments in fixed amounts or in amounts that increase by a fixed percentage are called fixed annuities. Alternatively, the payments of variable annuities are based upon the performance of some specified portfolio of securities, such as bond or equity mutual funds.

Life annuity has been a major part of actuarial science for a long time and it plays a major role in life insurance operations. For example, defined benefit pension plans are a form of life annuity typically provided by employers or governments as long as the annuitants are alive. The size of payouts is usually determined based on the employee's years of service, age, and salary. As a result, life annuity actually transfers longevity risk from the annuitant to the issuer. In fact, life annuity pricing elements include not only the mortality of the insured, but also the time value of money, the benefits promised and loadings to cover expenses, taxes, profits, and contingencies (Black and Skipper (2000)). That is being said, a life annuity has at least two major sources of risk, namely, mortality risk and interest rates risk.

The interaction of various risks can be quite complex, so that identification and quantification of each individual risk are of practical importance in view of risk management. For example, the determination of the most relevant risk drivers helps insurance companies to develop adequate risk management strategies such as prod-

uct modifications and hedging. To this end, different methodologies for deriving risk factors have already been proposed. The most common approach is a conditional expectation approach. Bühlmann (1995) use this approach to decompose the annual loss into financial and technical losses. A desirable property of this decomposition is that the sum of variances of risk factors amounts to the variance of the total loss. As a result, this approach is referred to as the variance decomposition. Schilling et al. (2015) propose the martingale representation theorem (MRT) decomposition which provides a way to allocate the randomness of liabilities to different sources of risk. They also introduce properties for meaningful risk decompositions and show that the MRT decomposition satisfies all of the properties while the variance decomposition violates some of the properties.

The first risk we analyze is the mortality risk. To better understand the variations in human mortality and to have a reasonable way to forecast mortality rates, it is natural to incorporate a time variable (year) and some uncertainty into models. Lee and Carter (1992) pioneered the research of stochastic mortality models. They introduced an age-independent time index to capture changes in the general level of mortality. This period term can then be modeled and forecast as a random walk with drift.

Lee-Carter model is a simple but useful model for capturing the behavior of human mortality. However, it does not include the cohort effect, which has a significant impact on mortality. Following the proposal of Lee-Carter model, there are a number of extensions or modifications of it, for example Brouhns et al. (2002), Booth et al. (2002) and Renshaw and Haberman (2003). Among all of these generalizations of the Lee-Carter model, Renshaw and Haberman (2006) proposed a model that incorporates a cohort effect. In the model, they add a random cohort term that is a function of the year of birth.

In lieu of modifying Lee-Carter model, Cairns et al. (2006) introduced a logit model regards to mortality rates $q_{x,t}$. In fact, it can be seen as a stochastic generalization of the Perks model (Cairns et al. (2006)). In their model (CBD model), random effects are captured by a bivariate random walk with drift $(\kappa_t^{(1)}, \kappa_t^{(2)})$, where $\kappa_t^{(1)}$ reflects

general mortality decline at all ages and $\kappa_t^{(2)}$ reflects the rate of decline with respect to different ages.

Cairns et al. (2009) compared eight stochastic mortality models quantitatively when fitting mortality rates in England and Wales and in the United States. They proposed criteria when assessing models. More specifically, a nice model would have following properties: be relatively parsimonious, provide a good fit to the historical data, able to generate sample paths, incorporate the cohort effect, have nontrivial correlation structure, and remain relatively simple. All three models have some nice features, however none of them meet all of the criteria. For example, Lee–Carter model does not capture the cohort effect and has a trivial correlation structure. Renshaw–Haberman also has trivial correlation structure and suffers from a lack of robustness (CMI (2007)). CBD model fits good for higher ages (60 years or older) while the performance deteriorates when fitting to the whole age range. Lastly, we point out that these three models are all extrapolative models. We choose those models since we believe that historical patterns would continue in the future.

Life insurance products often span from years to decades. This characteristic of life insurance makes the assumption of flat term structure of interest rates unrealistic. Thus we resort to the so called stochastic interest models. Single-factor models are a popular class of interest rate models. The pioneering one of this class is proposed by Vasicek (1977). It models the short term interest rate as an Ornstein-Uhlenbeck process. The Vasicek model is popular due to its analytical formulas for bonds pricing. Cox et al. (1985) develop an equilibrium model where the diffusion coefficient of the dynamics is a square-root term. Its analytical tractability and positivity make it stand out from other models.

The thesis is organized as follows. Chapter 1 presents three different stochastic mortality models and we choose the most suitable one to be the mortality model in the risk decomposition setting. Chapter 2 introduces the Vasicek and CIR models and illustrate the estimation of the parameters of the CIR model using maximum likelihood estimation (MLE) method. Chapter 3 lays out the considered life insurance modeling framework and introduces the MRT decomposition. Finally, Chapter 4

presents numerical examples using whole life and deferred life annuities.

Chapter 1

Stochastic Mortality Models

Mortality rates play a central role in life insurance. In fact the use of life tables, some of which had been derived from observed mortality rates, can be traced back to the birth of actuarial science. Thus constructing or choosing good models for mortality rates is one of the first tasks confronted by actuaries. The past 100 years have seen many improvements in life expectancy due to improvements in medical care, to the establishment of global health systems, etc. Demographers also reached a conclusion that mortality rates have improved during the last century. Thus deterministic mortality laws, which only describe a mortality schedule in analytical terms, cannot give an accurate prediction for mortality improvements. Therefore, we focus on the class of stochastic mortality models.

In this chapter, three stochastic mortality models will be fitted to Canadian population data and compared, based on AIC and BIC criteria. Then we derive a continuous time stochastic mortality model driven by a Wiener process from the chosen discrete time model as a part of the risk decomposition framework.

1.1 Notations and Data

1.1.1 Notations

Before introducing stochastic mortality models, it is important that we introduce notations that will be used to describe different models. We adopt the notations introduced by Cairns et al. (2009).

First, we define $m_c(t,x)$ to be the crude (observed) death rate for age x in calendar year t. Let $D(t,x)$ be the number of deaths during year t at age x and $E(t,x)$ be the exposure which represents the average number of people alive who were aged x at time t. More specifically,

$$m_c(t,x) = \frac{\text{Number of deaths}}{\text{Average population}} = \frac{D(t,x)}{E(t,x)}.$$

The average population is usually approximated by an estimate of the population aged x in the middle of the year, that is $E(t,x)$.

The underlying (central) death rate is then $m(t,x)$, which is equal to the expected number of deaths divided by the exposure, i.e.

$$m(t,x) = \frac{\mathrm{E}\left[D(t,x)\right]}{E(t,x)},$$

where $\mathrm{E}[\cdot]$ denotes expectation.

On the one hand, we can assume that $D(t,x)$ is a Poisson random variable. Specifically, we consider that

$$D(t,x) \sim Poisson\left(E(t,x)m(t,x)\right).$$

This is the so called Poisson assumption. If we adopt this assumption, which is commonly used in the literature (Brouhns et al. (2002); Renshaw and Haberman (2006); Cairns et al. (2009)), then the stochastic mortality models considered in this thesis can be estimated using Poisson maximum likelihood approach. The parameters in different models are determined in a similar manner by maximizing the log-likelihood based on this Poisson assumption. More details will be given in Section 1.2.1.

Mortality rate $q(t,x)$ is the probability that an individual aged exactly x at time t will die between t and $t+1$.

The force of mortality $\mu(t, x)$ is a general version of the usual $\mu(x)$ which allows for a stochastic force. For small dt, the rate of death between time t and $t + dt$ is approximately $\mu(t, x) \times dt$. So it can interpreted as the instantaneous death rate at time t for individuals aged x at that time. We assume that the force of mortality remains constant within each year of age and within each calendar year, i.e. for all $0 \leq s, u < 1, \mu(t + s, x + u) = \mu(t, x)$, where t and x are integers.

The assumption can be best understood with a Lexis diagram which implies the following:

$$m(t, x) = \mu(t, x).$$

This makes statistical inference much easier since death rates are estimated by the number of deaths, $D(t, x)$, and $E(t, x)$. They are contained in a typical dataset. Since $m(t, x) = \mu(t, x)$, we immediately have

$$q(t, x) = 1 - \exp(-\mu(t, x)) = 1 - \exp(-m(t, x)).$$

1.1.2 Data

In this thesis we use crude mortality rates for Canadian males between years 1960 and 2011. Since we are more concerned about longevity risk, the risk that realized mortality rates might be lower than anticipated, to which annuity providers are exposed, data at medium to higher ages (40 - 99 inclusive) will be used when we make comparisons of the different models. A typical dataset consists of numbers of deaths $D(t, x)$, and corresponding exposures, $E(t, x)$. Data were taken from the Human Mortality Database (www.mortality.org). As we can see in Figure 1.1 , crude death rates have been declining over the selected period.

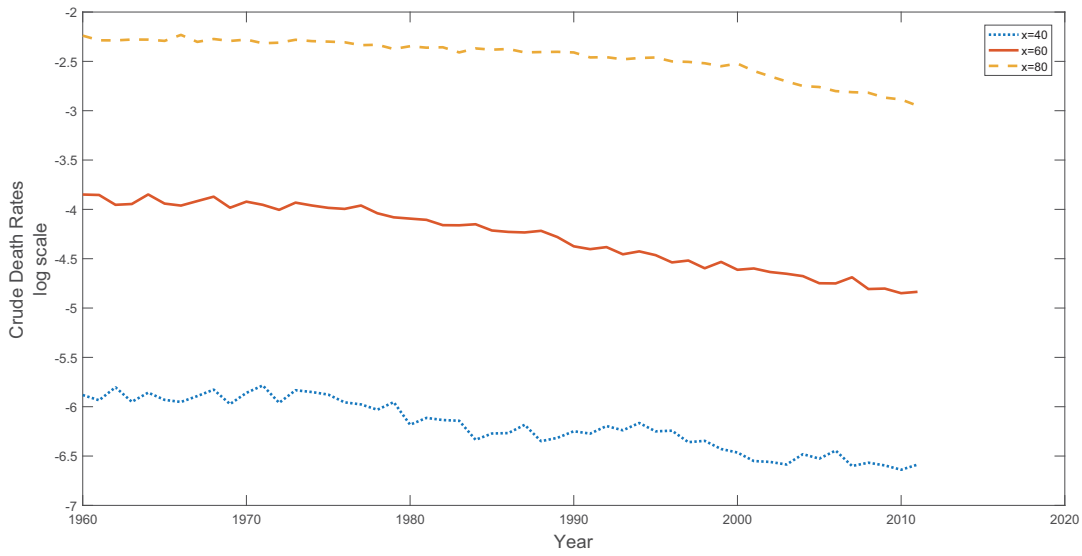

Figure 1.1: $\log m_c(t,x)$ at age 40, 60 and 80.

1.2 Lee–Carter(LC) Model

We start from the single factor model proposed by Lee and Carter (1992). It is a log-bilinear projection model with age and period effect, where

$$\ln \mu(t,x) = \beta_x^{(1)} + \beta_x^{(2)}\kappa_t^{(2)}, \tag{1.1}$$

the $\beta_x^{(1)}$ coefficients describe the average shape of the age profile, then $\exp(\beta_x^{(1)})$ is the general shape of the mortality schedule (age effect). The $\beta_x^{(2)}$ coefficients describe the pattern of deviations from this age profile when the parameter $\kappa_t^{(2)}$ varies. It indicates the sensitivity of the $\ln \mu(t,x)$ at age x to variations in the time index κ_t^2. The time-varying parameters κ_t^2 reflect the general level of mortality (period effect). The $\kappa_t^{(2)}$ series characterizes the general speed of mortality improvement.

Assuming we have data covering a set of consecutive calendar years $t = t_1, t_2, ..., t_n$ and a set of consecutive ages $x = x_1, x_2, ..., x_m$. We are trying to estimate the parameters appeared in this model. But before that, there is an identifiability problem to be addressed. To see this, note that

$$\ln \mu(t,x) = \tilde{\beta}_x^{(1)} + \tilde{\beta}_x^{(2)}\tilde{\kappa}_t^{(2)},$$

where $\tilde{\beta}_x^{(1)} = \beta_x^{(1)} + b\beta_x^{(2)}$, $\tilde{\beta}_x^{(2)} = \beta_x^{(2)}/a$ and $\tilde{\kappa}_t^{(2)} = a(\kappa_t^{(2)} - b)$. We can see that arbitrary selection of a and b leads to an arbitrary parameterization.

Therefore, we need to impose two constraints on the parameters to circumvent this problem. Lee and Carter (1992) propose the following constraints:

$$\sum_t \kappa_t^{(2)} = 0,$$
$$\sum_x \beta_x^{(2)} = 1. \tag{1.2}$$

Actually, there are no rules for choosing the constraints. The likelihood linked with the model has an infinite number of equivalent maximums. The important point is that they will lead to the same quality of fit or forecast no matter which constraints we choose. However, the first constraint here simplifies the estimation. For each x, the estimate for $\beta_x^{(1)}$ is the mean over t of the $\ln m_c(t,x)$, specifically,

$$\beta_x^{(1)} = \frac{\sum_{t=t_1}^{t_n} \ln m_c(t,x)}{t_n - t_1 + 1}.$$

Now we can fit the model to the data set. We first need a statistical model. Lee and Carter (1992) use the model

$$\ln m_c(t,x) = \beta_x^{(1)} + \beta_x^{(2)}\kappa_t^{(2)} + \epsilon(t,x),$$

where the error term $\epsilon(t,x)$ has mean 0 and variance σ_ϵ^2. It reflects particular age-specific historical influence that are not captured by the model. There are two ways to estimate the parameters in the model, namely ordinary least square (OLS) estimation and (Poisson) maximum likelihood estimation. We will discuss in details the OLS estimation in the following subsection. The maximum likelihood estimation will be addressed after introducing all the mortality models as a generalized estimation method.

1.2.1 OLS Estimation

It is worth mentioning that this model is not a simple regression model, since there are no observed independent variables on the right-hand side of (1.1). Specifically,

age x and time t are treated as factors and effect on death rates are captured by $\beta_x^{(1)}$'s and $\beta_x^{(2)}$'s for age, and by $\kappa_t^{(2)}$'s for time. Essentially this is a relational model.

As said before, $\beta_x^{(1)}$ can be calculated as long as we adopt normal constraints (1.2). In mathematical terms, we have

$$\hat{\beta}_x^{(1)} = \frac{\sum_{t=t_1}^{t_n} \ln m_c(t,x)}{t_n - t_1 + 1}.$$

Secondly, if we still want to use the regression method, the $\kappa_t^{(2)}$ values must be calculated. Since the sum of $\beta_x^{(2)}$ has been fixed to 1, we have

$$\kappa_t^{(2)} = \sum_{x=x_1}^{x_m} \left(\ln m_c(t,x) - \beta_x^{(1)} \right).$$

Now, the parameter $\beta_x^{(2)}$ can be estimated using the linear regressions over time, performed separately for each age x_i. In our linear regression, the dependent variable is the difference between the natural logarithms over time of the crude death rates at age x and $\beta_x^{(1)}$. It means that we regress $y = \ln m_c(t,x) - \beta_x^{(1)}$ on $x = \kappa_t^{(2)}$. Then $\hat{\beta}_x^{(2)}$'s are the estimated coefficients.

Beside this cumbersome estimation method, we can employ the singular value decomposition (SVD) to find a least-square solution to an objective function. Specifically, we need to find values of $\beta_x^{(1)}, \beta_x^{(2)}$ and $\kappa_t^{(2)}$ which they minimize the objective function

$$OLS(\boldsymbol{\beta^{(1)}}, \boldsymbol{\beta^{(2)}}, \boldsymbol{\kappa^{(2)}}) = \sum_{x=x_1}^{x_m} \sum_{t=t_1}^{t_n} \left(\ln m_c(t,x) - \beta_x^{(1)} - \beta_x^{(2)} \kappa_t^{(2)} \right)^2.$$

In fact, this is equivalent to maximum likelihood estimation provided that $\epsilon(t,x)$'s obey the Normal distribution. As usual, the $\beta_x^{(1)}$ coefficients are estimated as the average values over time of the $\ln m_c(t,x)$ values for each x. That is

$$\hat{\beta}_x^{(1)} = \frac{1}{t_n - t_1 + 1} \sum_{t=t_1}^{t_n} \ln m_c(t,x).$$

To see this, set $\frac{\partial}{\partial \beta_x^{(1)}} OLS$ equal to 0 yields to

$$\sum_{t=t_1}^{t_n} \ln m_c(t,x) = (t_n - t_1 + 1)\beta_x^{(1)} + \beta_x^{(2)} \sum_{t=t_1}^{t_n} \kappa_t^{(2)}.$$

From the constraints, we know that $\sum_{t=t_1}^{t_n} \kappa_t^{(2)} = 0$, thus we get the $\beta_x^{(1)}$'s. We then obtain the $\beta_x^{(2)}$'s and $\kappa_t^{(2)}$'s from the first term of a singular value decomposition of the matrix $\ln m_c(t,x) - \beta_x^{(1)}$.

Before proceeding to the application of the SVD, we brief review this concept. SVD is based on a theorem from linear algebra which says that a rectangular matrix A can be broken down into a product of three matrices.

Theorem 1.2.1 *Suppose M is a $m \times n$ real matrix. Then there exists a factorization, called a singular value decomposition of M, of the form*

$$M = U\Sigma V^T$$

where Σ is a $m \times n$ diagonal matrix with the square roots of eigenvalues of $M^T M$ in descending order on the diagonal and $U_{m\times m}$, $V_{n\times n}$ are orthogonal matrices.

Proof. A proof of the theorem can be found in Friedberg et al. (2003). □

We list two important properties regarding U and V here:

- $U^T U = I$, the columns of U are orthonormal eigenvectors of MM^T.
- $V^T V = I$, the columns of V are orthonormal eigenvectors of $M^T M$.

SVD is a method for identifying and ordering the dimensions where data points exhibit the most variation. Once we have identified where the most variation is, it is possible to find the best approximation of the original data points using fewer dimensions. Hence, SVD can be seen as a method for data reduction.

Now turn back to our estimation problem. Let us create the matrix

$$M = \begin{bmatrix} \ln m_c(t_1,x_1) - \beta_{x_1}^{(1)} & \dots & \ln m_c(t_n,x_1) - \beta_{x_1}^{(1)} \\ \vdots & \ddots & \vdots \\ \ln m_c(t_1,x_m) - \beta_{x_m}^{(1)} & \dots & \ln m_c(t_n,x_m) - \beta_{x_m}^{(1)} \end{bmatrix}.$$

M has m rows and n columns. Now, the $\beta_x^{(2)}$'s and $\kappa_t^{(2)}$'s are such that they minimize

$$OLS(\boldsymbol{\beta^{(2)}}, \boldsymbol{\kappa^{(2)}}) = \sum_{i=1}^{m}\sum_{j=1}^{n}(M_{ij} - \beta_{x_i}^{(2)}\kappa_{t_j}^{(2)})^2,$$

where M_{ij} are entries of matrix M. The solution is given by the singular value decomposition of M.

$M = U\Sigma V^T$ using the singular value decomposition. The best approximation of M in the least-squares sense is known to be

$$M \approx M^\star = \sqrt{\lambda_1}\mathbf{u}_1\mathbf{v}_1^T$$

where $\mathbf{u}_1$ is the first column of U, $\mathbf{v}_1$ is the first column of V, and $\sqrt{\lambda_1}$ is the first element of the diagonal of Σ.

Then we obtain

$$\beta_{x_j}^{(2)} = \frac{U_{j,1}}{\sum_{j=1}^m U_{j,1}},$$

$$\kappa_{t_i}^{(2)} = \sqrt{\lambda_1}\left(\sum_{j=1}^m U_{j,1}\right) V_{i,1}.$$

The constraints (1.2) are then satisfied by the $\hat{\beta}_x$'s and $\hat{\kappa}_t$'s.

If we use the $\kappa_t^{(2)}$'s obtained from SVD, there will be a discrepancy between the observed number of deaths $D(t) = \sum_x D(t,x)$ (in year) and the fitted one. It is mainly because we are modeling on the logarithmic scale. To solve this problem, we adjust the $\kappa_t^{(2)}$ by solving

$$\sum_x D(t,x) = \sum_x E(t,x)\exp\left(\beta_x^{(1)} + \beta_x^{(2)}\kappa_t^{(2)}\right).$$

Now with the re-estimated $\kappa_t^{(2)}$ we can produce exactly the same number of deaths actually observed in the data. We note that no explicit solution is available for the equation, which has to be solved numerically.

1.2.2 Mortality Projection

We could forecast mortality rates using the estimated parameters. In the LC model, the time index $\kappa_t^{(2)}$ can be seen as a time series when we need to forecast mortality rates. It is the only source of uncertainty involved in the projection process. As in their original paper (Lee and Carter (1992)), we first use an ARIMA(0, 1, 0) process

to fit the estimated $\kappa_t^{(2)}$ values, then the derived slope and drift values can be used in the extrapolation process. This ARIMA process is tested against by using standard Box-Jenkins procedures in many of the empirical studies in the literature. In most applications so far, $\kappa_t^{(2)}$ is well-modeled as a random walk with drift:

$$\kappa_{t_j}^{(2)} = \kappa_{t_{j-1}}^{(2)} + d + \sigma Z_j$$

where the Z_j's are i.i.d. standard Normal random variables.

From the above $\kappa_t^{(2)}$ modeling dynamic, we know that $(\kappa_{t_j}^{(2)} - \kappa_{t_{j-1}}^{(2)}), j = 2, 3, ..., n$ are independent random variables and normally distributed with mean d and standard derivation σ. Thus the maximum likelihood estimators of d and σ^2 are given by the sample mean and variance of the $(\kappa_{t_j}^{(2)} - \kappa_{t_{j-1}}^{(2)})$'s. They are

$$\begin{aligned} \hat{d} &= \frac{1}{t_n - t_1} \sum_{t=t_2}^{t_n} (\kappa_t^{(2)} - \kappa_{t-1}^{(2)}) = \frac{\kappa_{t_n}^{(2)} - \kappa_{t_1}^{(2)}}{t_n - t_1}, \\ \hat{\sigma}^2 &= \frac{1}{t_n - t_1} \sum_{t=t_2}^{t_n} (\kappa_t^{(2)} - \kappa_{t-1}^{(2)} - \hat{d})^2. \end{aligned} \tag{1.3}$$

Then for the projection period starting from t_{n+1}, we use $\beta_x^{(1)}, \beta_x^{(2)}$ estimated from the model together with $\kappa_t^{(2)}$'s calculated from the random walk with drift projection model. The forecast central death rate is

$$\ln m(t_j, x) = \beta_x^{(1)} + \beta_x^{(2)} \kappa_{t_j}^{(2)},$$

$j = n + 1, n + 2, \ldots$.

1.3 Renshaw–Haberman(RH) Model

Renshaw and Haberman (2006) propose the following model for population mortality with a cohort effect:

$$\ln \mu(t, x) = \beta_x^{(1)} + \beta_x^{(2)} \kappa_t^{(2)} + \beta_x^{(3)} \gamma_{t-x}^{(3)}. \tag{1.4}$$

This is an extension of the Lee-Carter model with an extra cohort effect term $\beta_x^{(3)} \gamma_{t-x}^{(3)}$. To see this, notice that $t - x$ is the year of birth of a cohort.

This model has similar identifiability problems as the LC model (Cairns et al. (2009)). They use the following constraints to ensure that we are able to perform the estimation:

$$\begin{aligned}
\sum_t \kappa_t^{(2)} &= 0, \\
\sum_x \beta_x^{(2)} &= 1, \\
\sum_{x,t} \gamma_{t-x}^{(3)} &= 0, \\
\sum_x \beta_x^{(3)} &= 1.
\end{aligned} \tag{1.5}$$

The first and third equations allow us to estimate $\beta_x^{(1)}$ as mean over time of the $\ln m(t,x)$. The second and fourth equations are not natural choices, but it has no effect on the quality of fit as in the Lee–Cater model.

1.3.1 Poisson Maximum Likelihood Estimation

As we mentioned earlier, $D(t,x)$ is a Poisson random variable with mean $E(t,x)m(t,x)$. Thus we can estimate parameters by maximizing the Poisson log-likelihood over all age, period, and cohort parameters. To make it more general, we denote by ϕ the full set of parameters. In RH model, there are five sets of parameters, i.e., the $\beta_x^{(1)}, \beta_x^{(2)}, \beta_x^{(3)}$, the $\kappa_t^{(2)}$ and the $\gamma_{t-x}^{(3)}$ terms. The log-likelihood is

$$L(\phi) = \sum_{t,x} \Big(D(t,x) \ln \left[E(t,x) m(t,x;\phi) \right] - E(t,x) m(t,x;\phi) - \ln \left[D(t,x)! \right] \Big).$$

Brouhns et al. (2002) give an iterative algorithm using a uni-dimensional or elementary Newton method to estimate log-linear models with bilinear terms. In iteration step $\nu+1$, a single set of parameters is updated fixing other parameters at their current estimates using the following updating scheme

$$\hat{\theta}^{(\nu+1)} = \hat{\theta}^{(\nu)} - \frac{\partial L^{(\nu)} / \partial \theta}{\partial^2 L^{(\nu)} / \partial \theta^2},$$

where $L^{(\nu)} = L^{(\nu)}(\hat{\theta}^{(\nu)})$.

With respect to RH model, we can proceed as follows. First specify the starting values, we set $\beta_x^{(1)} = \frac{\sum_{t=t_1}^{t_n} \ln m_c(t,x)}{t_n - t_1 + 1}, \beta_x^{(2)} = 0, \beta_x^{(3)} = 1/m$ for each age x. Also, set $\kappa_{t_j}^{(2)} = \frac{n+1}{2} - j$, for $j = 1, 2, ..., n$, (arbitrary values can also be used) and each calendar year t. Next, we set $\gamma_{t-x}^{(3)} = 0$ for each cohort year $t - x$. Within each iteration:

- Update each of the $\beta_x^{(1)}$'s in turn, using the presented algorithm. For a given x, this amounts to increasing the likelihood over age x cells only. The likelihood for all other ages is unaffected.
- Update each of the $\gamma_{t-x}^{(3)}$'s in turn. For a given $t - x$, this amounts to increasing the likelihood over cells, (t, x), that have a common year of birth only. The likelihood for all other cohort years of birth is unaffected.
- Update each of the $\beta_x^{(2)}$'s in the same manner.
- Update each of the $\kappa_t^{(2)}$'s in turn. For a given t, this amounts to increasing the likelihood over calendar years t cells only. The likelihood for all other calendar years is unaffected.
- Update each of the $\beta_x^{(3)}$'s in the same manner.
- Apply the identifiability constraints.

We specify a criterion to stop the procedure, usually a very small increase of the log-likelihood function. This ensures that the log-likelihood converges within a specified degree of tolerance. This method can easily be adapted to fit the Lee–Carter model. In fact, we use Poisson maximum likelihood estimation to fit all the mortality models in this thesis.

1.3.2 Mortality Projection

For the period effect, Cairns et al. (2011) argue that random walk processes have been widely used to drive the dynamics of the period effect since the introduction of the original Lee and Carter (1992) model. In their original paper, Renshaw and Haberman (2006) also project two time series $\kappa_t^{(2)}$ and $\gamma_{t-x}^{(3)}$ by using univariate ARIMA

models. Thus we fit and forecast $\kappa_t^{(2)}$ and $\gamma_{t-x}^{(3)}$ by two independent random walks with drift (or ARIMA(0,1,0) processes) in this thesis.

1.4 Cairns–Blake–Dowd(CBD) Model

Cairns et al. (2006) propose the following model for mortality rates $q(t,x)$

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

where $\bar{x}$ is the mean of ages in the sample range and logit $(x) = \ln \frac{x}{1-x}, 0 < x < 1$.

Note that we have two stochastic time series here, which means the period effect enters the model in two different ways. The intercept period term $\kappa_t^{(1)}$ affects mortality at different ages in the same way, which corresponds to the feature that mortality rates are decreasing over time at all ages. On the other hand, the slope period term $\kappa_t^{(2)}$ affects mortality proportionally to age. It uses two period effect parameters to capture the trend improvement in mortality rates (the intercept or level term) and the differential higher age dynamics (slope term). This is a fundamental difference when compared with Lee–Carter model since it has only one period term $\kappa_t^{(2)}$ that affects all ages at the same time. Moreover, this model specification does not suffer from any identifiability problems as in the Lee–Carter model class.

1.4.1 Estimation

Again, assuming we have data covering a set of consecutive calendar years $t = t_1, t_2, ..., t_n$ and a set of consecutive ages $x = x_1, x_2, ..., x_m$. Now we want to estimate $\kappa_t^{(1)}$ and $\kappa_t^{(2)}$ parameters. This can be done by least squares or simple linear regression. For each year t, we have m dependent variables logit $q(t,x_1)$, logit $q(t,x_2)$, ..., logit $q(t,x_m)$ and corresponding explanatory variables $x_1 - \bar{x}, x_2 - \bar{x}, ..., x_m - \bar{x}$. Now suppose a simple regression model $y = \beta_0 + \beta_1 x$, then we can estimate $\kappa_t^{(1)}$ and $\kappa_t^{(2)}$ as using the linear regression setting. Note that CBD model implicitly assumes that

$$\text{logit } q(t,x) = \ln \frac{q(t,x)}{1 - q(t,x)} = \kappa_t^{(1)} + (x - \bar{x})\kappa_t^{(2)} + \epsilon(t,x)$$

where the error terms $\epsilon(t,x)$ are i.i.d. normal random variables with mean 0 and constant variance σ^2. In the OLS estimation process, period parameters $\kappa_t^{(1)}$ and $\kappa_t^{(2)}$ are estimated separately for each calendar year t.

There are alternative ways to fit CBD model, such as Poisson maximum likelihood method. We only need to transform $q(t,x)$ to $m(t,x)$ using the relationship $m(t,x) = -\ln(1-q(t,x))$. Then we can maximize the log likelihood introduced in Section 1.2.1 using the elementary Newton method to get the estimated values for $\kappa_t^{(1)}$ and $\kappa_t^{(2)}$.

1.4.2 Mortality Projection

In order to forecast mortality rates improvements, we need to fit a two-dimensional random walk with drift $\kappa^{(1)}(t), \kappa^{(2)}(t)$. We suggest to adopt symbols used in their original model. Specifically, let $K(t) = \left(\kappa^{(1)}(t), \kappa^{(2)}(t)\right)'$ with dynamic

$$K(t+1) = K(t) + \mu + CZ(t+1),$$

where μ is a constant 2×1 vector of drift parameters, C is a constant 2×2 lower triangular Cholesky square root matrix of the covariance matrix V (that is $V = CC^T$), and $Z(t)$ is a two-dimensional standard normal random variable. As pointed by Cairns et al. (2006), the choice of C makes no difference in the analysis. The restriction of C to a lower triangular form means that C is straightforward from V and that this (Cholesky) decomposition is unique.

1.5 Models Comparison

In this section we evaluate and compare the mortality models and determine which one produce the best fit to data for Canadian males aged between 40 and 99 over the period 1960 to 2011.

When fitting models it is possible to increase log-likelihood by adding parameters. As pointed by Cairns et al. (2009), such improvements are almost guaranteed if models

Model	L	AIC (Rank)	BIC (Rank)	ν
LC	-18,787	37,914 (2)	38,941 (2)	170
RH	-15,585	31,934 (1)	33,983 (1)	339
CBD	-20,821	41,851 (3)	42,480 (3)	104

Table 1.1: Comparison results

are nested: if one model is a special case of another, then the model with more parameters will typically have a higher maximum likelihood, even if the true model is the one with fewer parameters. This may result in overfitting. A overfit model has poor predictive performance, as it overreacts to minor fluctuations in the dataset, or we can say it is more 'sensitive' to the noise. To resolve this problem, we use AIC and BIC as criteria of the selection.

They have some advantages when dealing with model selection problem. First, their use allows to strike a balance between quality of fit and parsimony by introducing a penalty term for the number of parameters in the model. Moreover, BIC makes no assumptions about "prior" model rankings: that is, all models have equal status in terms of how we rank them. In a word, the two information criteria deal with the trade-off between the goodness of fit of the model and the complexity of the model.

The AIC and BIC for model i are specified respectively as

$$AIC_i = -2 \ln L(\hat{\phi}_i) + 2\nu_i. \tag{1.6}$$

$$BIC_i = -2 \ln L(\hat{\phi}_i) + \nu_i \ln N, \tag{1.7}$$

where ϕ_i is the parameter vector for model i, $\hat{\phi}_i$ is its maximum likelihood estimate, $\ln L(\hat{\phi}_i)$ is the maximum log likelihood, N is the number of observations and ν_i is the effective number of parameters being estimated. For our models, ν_i equals the number of constraints subtracted from the actual number of estimated parameters. For instance, effective number of parameters in Lee–Carter model $\nu_1 = 2 \times 60$ (ages) + 52 (years) $-$ 2 (number of constraints) $= 170$. Then the models can be ordered, with the top model having the lowest AIC or BIC.

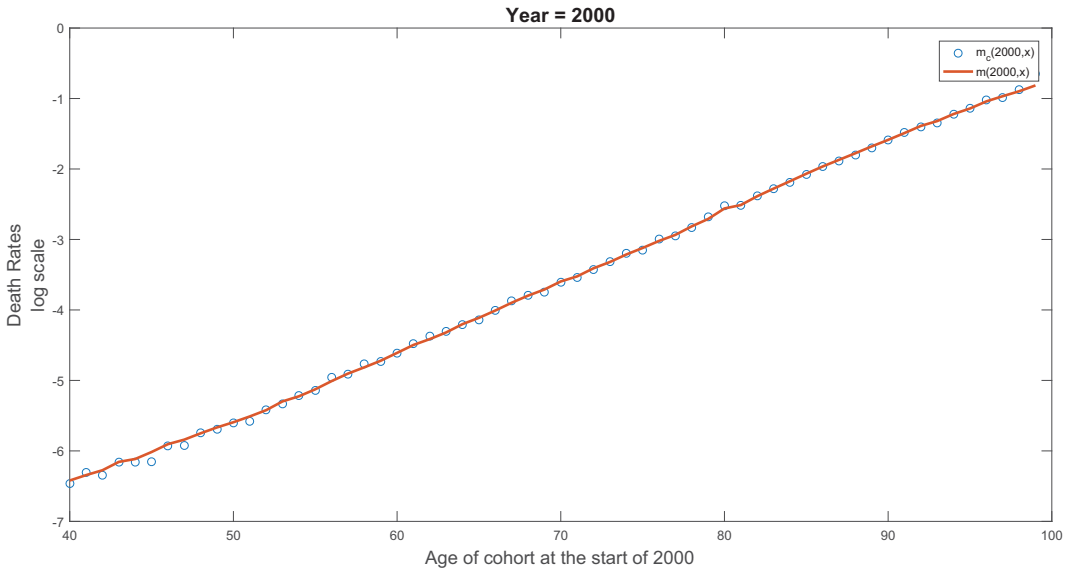

Figure 1.2: Mortality rates for Canadian males for the year 2000 (dots) and fitted RH model.

Values for the AIC and BIC, the maximum log likelihood and corresponding effective number of parameters are presented in Table 1.1. We conclude that Renshaw–Haberman model stands out on top with the lowest AIC and BIC values. Also we note here that RH model parameters converge very slowly to their maximum likelihood estimates when compared to other two models which is observed also by other authors (Renshaw and Haberman (2006); Cairns et al. (2009)).

Next we present the fitted mortality curve using RH model and mortality rates for the year 2000 in our dataset in Figure 1.2 and 1.3. The fit is clearly very good, especially for old ages. Estimated values are $\beta_x^{(1)}$, $\beta_x^{(2)}$ and $\beta_x^{(3)}$ for the ages 40-99, $\kappa_t^{(2)}$ for the years 2000 and 2010 and $\gamma_{t-x}^{(3)}$ for corresponding cohort years.

1.6 Continuous Time Stochastic Mortality Model

In this section, a continuous Renshaw–Haberman stochastic mortality model $\mu(t)$ will be derived as a corner stone of the life insurance modeling framework.

When we move to the continuous time model world, the Wiener process, sometimes

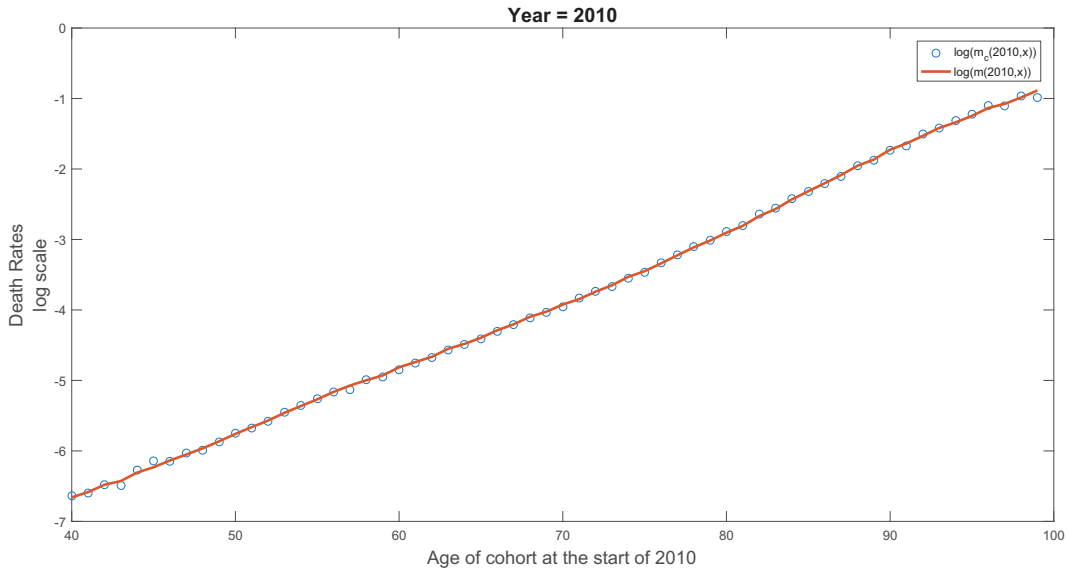

Figure 1.3: Mortality rates for Canadian males for the year 2010 (dots) and fitted RH model.

called the Brownian motion, is the fundamental building block in modern theory of random process.

Definition 1.6.1. A stochastic process $\{W(t), t \geq 0\}$ is said to be a standard Wiener process if

i) $W(0) = 0$;

ii) $\{W(t), t \geq 0\}$ has stationary and independent increments;

iii) for every $t > 0$, $W(t)$ is normally distributed with mean 0 and variance t.

The name of Brownian motion goes back to the English botanist Robert Brown who discovered a motion exhibited by a small particle that is immersed in a liquid. The standard Wiener process can be generalized to allow $W(t)$ has the normal distribution with mean 0 and variance $\sigma^2 t$ for some constants σ^2. From the definition of $W(t)$, we expect $W(t)$ to be a continuous function of t. It turns out to be the case; with probability 1, $W(t)$ is indeed a continuous function of t.

To derive the dynamics of the mortality intensity process, it is inevitable to use stochastic calculus. First we need the definition of Itô process.

Definition 1.6.2. Let $W(t)$, $t \geq 0$, be a Wiener process. An Itô process is a stochastic process of the form

$$X(t) = X(0) + \int_0^t a(s)ds + \int_0^t b(s)dW(s)$$

where $X(0)$ is a scalar starting point and $a(t)$ and $b(t)$ are adapted stochastic processes.

$a(t)$ and $b(t)$ are respectively the drift and the diffusion terms. A shorthand notation is the following stochastic differential equation for $dX(t)$,

$$dX(t) = a(t)dt + b(t)dW(t).$$

This can be seen as a Brownian motion with an instantaneous drift $a(t)$ and an instantaneous variance rate $b^2(t)$.

Theorem 1.6.1 (Itô formula) *Let $X(t)$ be an Itô process and let $f(t,x)$ be a function for which the partial derivatives $f_t(t,x), f_x(t,x)$, and $f_{xx}(t,x)$ are defined and continuous, and let $W(t)$ be a Wiener process. Then, for every $T \geq 0$,*

$$f(T, X(T)) = f(0, X(0)) + \int_0^T f_t(t, X(t))dt + \int_0^T f_x(t, X(t))dX(t) + \frac{1}{2}\int_0^T f_{xx}(t, X(t))(dX(t))^2.$$

Proof. The proof can be found in Shreve (2004). □

Remark. Again we have the differential notation for this formula

$$df(t, X(t)) = f_t(t, X(t))dt + f_x(t, X(t))dX(t) + \frac{1}{2}f_{xx}(t, X(t))dX(t)dX(t).$$

If we apply the rules $dW(t)dW(t) = dt$, $dtdW(t) = dW(t)dt = 0$, $dtdt = 0$, then the Ito formula in differential form simplifies further to

$$\begin{aligned} df(t, X(t)) &= f_t(t, X(t))dt + f_x(t, X(t))(a(t)dt + b(t)dW(t)) + \frac{1}{2}f_{xx}(t, X(t))(b^2(t)dt) \\ &= \left[f_t(t, X(t)) + a(t)f_x(t, X(t)) + \frac{1}{2}b^2(t)f_{xx}(t, X(t))\right]dt + b(t)f_x(t, X(t))dW(t). \end{aligned}$$

where $dX(t)dX(t) = (a(t)dt + b(t)dW(t))(a(t)dt + b(t)dW(t)) = b^2(t)dt$.

1.6.1 A Generalized Renshaw–Haberman Model

First, we remove the assumption that $\mu(t,x)$ remains constant over each calendar year. Now we are ready to extend the discrete time RH model to continuous time. In light of (1.4), a general continuous-time intensity can be written as

$$\ln \mu(t,x) = \beta_1(x+t) + \beta_2(x+t)\kappa(t) + \beta_3(x+t)\gamma, \tag{1.8}$$

for some functions β_1, β_2 and β_3 and a constant γ. Meanwhile, $\kappa(t)$ is a stochastic process.

In (1.8), we have assumed that the person is of age x at time 0 (which can be an arbitrary calendar time). The parameter t then describes the time that has passed since time 0. This explains the dependence of functions β_1, β_2 and β_3 on age $x+t$ since the process $\mu(t,x)$ describes the evolution of the intensity of mortality of an individual aged $x+t$ at each time t. Thus, the $\beta_x^{(1)}$s, $\beta_x^{(2)}$s and $\beta_x^{(3)}$ of expression (1.4) must be seen as the point-wise estimates of the functions β_1's, β_2's and β_3's at each age $x+t$. Then we apply some interpolation methods to get the values of three functions.

As we mentioned in Section 1.3.2, $\kappa_t^{(2)}$ is modeled as a random walk with drift $\kappa_t^{(2)} = \kappa_{t-1}^{(2)} + \theta + \sigma Z$. So in the continuous case, it is natural to assume $\kappa(t)$ has the dynamic

$$d\kappa(t) = \theta dt + \sigma dW(t),$$

where θ and σ are to be estimated from the $\kappa_t^{(2)}$'s as stated in (1.3).

The dynamic of $\mu(t,x)$ can be found by using Ito formula. To see this, let $X(t,x) = \ln \mu(t,x)$ and $f(x) = e^x$, then we have $\mu(t,x) = f(X(t,x))$, the differential of $X(t,x)$ is

$$\begin{aligned}
dX(t,x) &= \beta_1'(x+t)dt + d(\beta_2(x+t)\kappa(t)) + \gamma\beta_3'(x+t)dt \\
&= (\beta_1'(x+t) + \gamma\beta_3'(x+t))dt + \kappa(t)\beta_2'(x+t)dt + \beta_2(x+t)d\kappa(t) \\
&= \big[\beta_1'(x+t) + \gamma\beta_3'(x+t) + \kappa(t)\beta_2'(x+t)\big]dt + \beta_2(x+t)(\theta dt + \sigma dW(t)) \\
&= \big[\theta\beta_2(x+t) + \beta_1'(x+t) + \kappa(t)\beta_2'(x+t) + \gamma\beta_3'(x+t)\big]dt + \sigma\beta_2(x+t)dW(t).
\end{aligned}$$

Then by the Ito formula we have

$$
\begin{aligned}
d\mu(t,x) &= f_x(X(t,x))dX(t,x) + \frac{1}{2}f_{xx}(X(t,x))dX(t,x)dX(t,x) \\
&= e^{X(t,x)}dX(t,x) + \frac{1}{2}e^{X(t,x)}\beta_2^2(x+t)\sigma^2 dt \\
&= \mu(t,x)\Big(\big[\theta\beta_2(x+t) + \beta_1^{'}(x+t) + \kappa(t)\beta_2^{'}(x+t) + \gamma\beta_3^{'}(x+t)\big]dt + \sigma\beta_2(x+t)dW(t)\Big) \\
&+ \frac{1}{2}\sigma^2\beta_2^2(x+t)\mu(t,x)dt \\
&= \Big(\theta\beta_2(x+t) + \frac{1}{2}\sigma^2\beta_2^2(x+t) + \beta_1^{'}(x+t) + \kappa(t)\beta_2^{'}(x+t) + \gamma\beta_3^{'}(x+t)\Big)\mu(t,x)dt \\
&+ \sigma\beta_2(x+t)\mu(t,x)dW(t).
\end{aligned}
\tag{1.9}
$$

As for the initial value, $\mu(0,x) = \exp\left(X(0,x)\right) = \exp\left(\beta_1(x) + \beta_3(x)\gamma + \beta_2(x)\kappa(0)\right)$ for a given cohort age x. We choose as $\kappa(0)$ the last estimate of $\kappa_t^{(2)}$ in the Poisson maximum likelihood estimation process.

Chapter 2

Interest Rate Models

The concept of interest rates plays a very important role in modern finance. Interest rate derivatives, i.e. derivatives with payoffs depending on interest rates or bond prices, are also traded in large volume. From this point of view, an annuity is closely related to a fixed income derivative since an annuity often exists over decades with steady payments. Thus constant interest rates assumption used in textbooks when introducing annuities is not suitable for our analysis. The main goal here is to introduce the basic concepts and terminologies involved with stochastic interest rate modeling. We begin with the definition of a bank account. Let $B(t)$ be the value of a bank account at time $t \geq 0$. We assume that we started our bank account with initial deposit of 1 ($B(0) = 1$). The value of the bank account evolves according to the following differential equation:

$$dB(t) = r(t)B(t)dt.$$

where $r(t)$ is a positive function of time.

Remark. Integrating over time together with initial condition $B(0) = 1$ leads to

$$B(t) = \exp\left(\int_0^t r(s)ds\right).$$

The above equation tells us that investing a unit amount at time 0 yields at time t the value $B(t)$. $r(t)$ is the instantaneous rate at which the bank account accrues. This instantaneous rate is referred to as short rate.

The discount factor $D(t,T)$ is the amount at time t that is "equivalent" to one unit of currency payable at time T, and is given by

$$D(t,T) = \frac{B(t)}{B(T)} = \exp\left(-\int_t^T r(s)ds\right).$$

Since we are going to model the evolution of r in time through a stochastic process, the discount factor D is also a stochastic process.

A Bond is a another fundamental instrument used in financial theory. A bond is a debt investment in which an investor loans money to an entity which borrows the funds for a defined period of time at a variable or fixed interest rate. Thus a bond can be viewed as a loan. A T-maturity zero-coupon bond is a contract that guarantees its holder the payment of one unit of currency at time T, with no intermediate payments. The contract value at time $t \leq T$ is denoted by $P(t,T)$.

Zero-coupon bond prices play a very important role in interest-rate theory. In fact, if we want to know the present value of a derivative or a financial product, we have to deal with zero-coupon bond prices first. Moreover, most interest rate quantities can be obtained from the zero-coupon bond price, and vice versa.

The continuously-compounded spot interest rate prevailing at time t for the maturity T is denoted by $R(t,T)$ and is equivalent to the constant rate at which an investment of $P(t,T)$ units of currency at time t accrues continuously to yield a unit amount of currency at maturity T, i.e.

$$P(t,T)e^{R(t,T)(T-t)} = 1.$$

Thus, the bond price $P(t,T)$ can be expressed in terms of R:

$$P(t,T) = e^{-R(t,T)(T-t)}.$$

$R(t,T)$ is closely related to the short rate $r(t)$. Indeed, time to maturity $T-t$ can be partitioned into infinitesimal time intervals. From this point of view, modeling $R(t,T)$ is equivalent to modeling $r(t)$. More specifically, the short rate is a limit of the spot rate, that is, for each t,

$$r(t) = \lim_{T \to t^+} R(t,T).$$

2.1 Interest Rate Models

In order to price interest rate derivatives we need a model which can describe the evolution of interest rate reasonably. In this section, we present two classical time-homogeneous short-rate models, namely the Vasicek and the Cox, Ingersoll and Ross models.

2.1.1 The Vasicek Model

Vasicek (1977) assumes that the instantaneous spot interest rate is driven by a one-dimensional diffusion process, namely an Ornstein–Uhlenbeck process with constant coefficients. That is why this model belongs to the family of one-factor short-rate models. In his original paper, $r(t)$ evolves under the real world measure P. But for a suitable choice of the market price of risk, we can write the dynamics of r under the risk-neutral measure Q as

$$dr(t) = \kappa[\theta - r(t)]dt + \sigma dW(t), \tag{2.1}$$

where $r(0), \kappa, \theta$ and σ are positive constants. The parameter θ represents long-term mean. If the interest rate r is larger than mean θ, then the drift term is negative since $\kappa > 0$, so that r will be pulled down in the direction of θ. Similarly, the drift of r is positive whenever the interest rate is below θ, hence $r(t)$ is pushed to be closer to θ. We call this behavior mean reversion. Therefore, κ is the speed of reversion. In fact, possessing the mean reversion characteristic is a compelling feature of the interest rate model. It prevents the interest rate from rising indefinitely.

Integrating (2.1), we obtain,

$$\begin{aligned} r(t) &= e^{-\kappa t}\left[r_0 + \int_0^t e^{\kappa s}\kappa\theta ds + \int_0^t e^{\kappa s}\sigma dW(s)\right] \\ &= r_0 e^{-\kappa t} + e^{-\kappa t}\left[\kappa\theta(\frac{1}{\kappa}e^{\kappa t} - \frac{1}{\kappa}) + \sigma\int_0^t e^{\kappa s}dW(s)\right] \\ &= r_0 e^{-\kappa t} + e^{-\kappa t}\theta(e^{\kappa t} - 1) + \sigma\int_0^t e^{-\kappa(t-s)}dW(s) \\ &= r_0 e^{-\kappa t} + \theta(1 - e^{-\kappa t}) + \sigma\int_0^t e^{-\kappa(t-s)}dW(s). \end{aligned}$$

Similarly, for each $s \leq t$, we have

$$r(t) = r(s)e^{-\kappa(t-s)} + \theta(1 - e^{-\kappa(t-s)}) + \sigma \int_s^t e^{-\kappa(t-u)} dW(u).$$

Hence, the conditional distribution of $r(t)$ is a Normal distribution with the mean and variance given by

$$\begin{aligned} \mathrm{E}[r(t)|\mathcal{F}_s] &= r(s)e^{-\kappa(t-s)} + \theta(1 - e^{-\kappa(t-s)}), \\ \mathrm{Var}(r(t)|\mathcal{F}_s) &= \frac{\sigma^2}{2\kappa}[1 - e^{-2\kappa(t-s)}], \end{aligned}$$

where $\{\mathcal{F}_t\}_{t\geq 0}$ is the filtration generated by $\{W(t)\}_{t\geq 0}$. It contains all the information about $W(t)$ up to time t. Consequently, we have

$$\begin{aligned} \lim_{t\to\infty} \mathrm{E}[r(t)] &= \theta, \\ \lim_{t\to\infty} \mathrm{Var}(r(t)) &= \frac{\sigma^2}{2\kappa}. \end{aligned}$$

Again, we justify the name long-term mean θ from the asymptotic expectation. Normally distributed interest rate implies that there is a positive possibility for $r(t)$ to be negative, which indeed is a main drawback of the Vasicek model. However, we value more the analytical tractability implied by the Gaussian density compared to the shortcoming.

As stated in Brigo and Mercurio (2007), the Vasicek model possesses an affine term structure, i.e. the continuously compounded spot rate $R(t,T)$ is an affine function in the short rate $r(t)$,

$$R(t,T) = \alpha(t,T) + \beta(t,T)r(t),$$

where α and β are deterministic functions of time. This relationship is always satisfied when the zero–coupon bond price can be written in the form

$$P(t,T) = A(t,T)e^{-B(t,T)r(t)}.$$

Indeed, in the Vasicek model

$$\begin{aligned} A(t,T) &= \exp\left[\big(\theta - \frac{\sigma^2}{2\kappa^2}\big)(B(t,T) - T + t) - \frac{\sigma^2}{4\kappa}B(t,T)^2\right], \\ B(t,T) &= \frac{1}{\kappa}\big[1 - e^{-\kappa(T-t)}\big]. \end{aligned}$$

The existence of affine term structure is a very convenient property from a computational and analytical point of view, because the explicit formula for the term structure makes the model easier to calibrate.

2.1.2 The Cox-Ingersoll-Ross(CIR) Model

Cox, Ingersoll and Ross (1985) introduce an extension of the Vasicek model which includes a "square-root" term in the diffusion coefficient of the instantaneous short-rate dynamics. Under the risk-neutral measure Q, the dynamics follows the stochastic differential equation

$$dr(t) = \kappa(\theta - r(t))dt + \sigma\sqrt{r(t)}dW(t), \tag{2.2}$$

where $r(0), \kappa, \theta$ and σ are positive constants.

The drift term $\kappa(\theta - r(t))$ is the same as in the Vasicek model. Therefore, the short rate r is mean reverting with long-term mean θ. If the parameters satisfy the following condition (known as the Feller condition)

$$2\kappa\theta > \sigma^2,$$

then the process $r(t)$ would never reach 0, so that we can guarantee that $r(t)$ remains positive according to dynamics (2.2). When the interest rate approaches zero then the volatility term $\sigma\sqrt{r(t)}$ becomes very small, which decreases randomness. Consequently, the evolution of short rate $r(t)$ is dominated by the drift factor, which pushes the rate upwards.

The process r admits a non-central chi-squared distribution rather than a Normal distribution, see Brigo and Mercurio (2007). The density function for the interest rate process is

$$\begin{aligned} f_{r(t)}(x) &= f_{\frac{\chi^2(\nu,\lambda_t)}{c_t}}(x) = c_t f_{\chi^2(\nu,\lambda_t)}(c_t x), \\ c_t &= \frac{4\kappa}{\sigma^2(1-\exp(-\kappa t))}, \\ \nu &= 4\kappa\theta/\sigma^2, \\ \lambda_t &= c_t r_0 \exp(-\kappa t). \end{aligned}$$

where $\chi^2(\nu, \lambda)$ is a non-central chi-squared distribution with ν degrees of freedom and non-centrality parameter λ.

The mean and variance of $r(t)$ conditional on $\mathcal{F}_s$ are given by

$$\begin{aligned}
\mathrm{E}[r(t)|\mathcal{F}_s] &= r(s)e^{-\kappa(t-s)} + \theta(1 - e^{-\kappa(t-s)}), \\
\mathrm{Var}(r(t)|\mathcal{F}_s) &= r(s)\frac{\sigma^2}{\kappa}\big(e^{-\kappa(t-s)-e^{-2\kappa(t-s)}}\big) + \theta\frac{\sigma^2}{2\kappa}\big(1 - e^{-\kappa(t-s)}\big)^2.
\end{aligned}$$

The CIR model is tractable because its transition density has a closed-form expression in which maximum likelihood estimation (MLE) can be undertaken and there also exists a closed formula for bond price. In fact, the CIR model also belongs to the family of affine term-structure models. We have explicit solutions for the bond prices. The price at time t of a zero-coupon bond with maturity T is

$$P(t, T) = A(t, T)e^{-B(t,T)r(t)}, \tag{2.3}$$

where

$$\begin{aligned}
A(t, T) &= \left[\frac{2h\exp\{(\kappa + h)(T - t)/2\}}{2h + (\kappa + h)(\exp\{(T - t)h\} - 1)}\right]^{2\kappa\theta/\sigma^2}, \\
B(t, T) &= \frac{2(\exp\{(T - t)h\} - 1)}{2h + (\kappa + h)(\exp\{(T - t)h\} - 1)}, \\
h &= \sqrt{\kappa^2 + 2\sigma^2}.
\end{aligned}$$

2.2 Estimation

It is natural to perform the historical estimation under the real-world measure P. In this section, we estimate the parameters of the CIR model using historical one-month yields of the Canadian treasury bills. To preserve the same square–root structure in (2.2) under the real-world measure P, we adopt the following formulation (Brigo and Mercurio (2007)):

$$\begin{aligned}
dr(t) &= [\kappa\theta - (\kappa + \lambda\sigma)r(t)]dt + \sigma\sqrt{r(t)}dW^0(t) \\
&= (\kappa + \lambda\sigma)\left[\frac{\kappa\theta}{\kappa + \lambda\sigma} - r(t)\right]dt + \sigma dW^0(t) \\
&= \kappa^*(\theta^* - r(t)) + \sigma^* dW^0(t),
\end{aligned} \tag{2.4}$$

where λ is a new parameter, contributing to the market price of risk. The real-world dynamics (2.4) now has the same form as in (2.2) and we only need to estimate κ^*, θ^*, and σ^*. Therefore, formulas and results presented above are also available under the real-world measure P.

For maximum likelihood estimation transition densities are required. Cox et al. (1985) obtain the result that given r_t at time t the distribution of $r_{t+\Delta t}$ at time $t+\Delta t$ is a non-central chi-square distribution and give the formulation:

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

where

$$\begin{aligned} c &= \frac{2\kappa}{\sigma^2(1-e^{-\kappa\Delta t})}, \\ u &= cr_t e^{-\kappa\Delta t}, \\ v &= cr_{t+\Delta t}, \\ q &= \frac{2\kappa\theta}{\sigma^2} - 1, \end{aligned}$$

and $I_q(2\sqrt{uv})$ is a modified Bessel function of the first kind of order q.

The likelihood function for interest rate time series with N observations is

$$L(\kappa,\theta,\sigma) = \prod_{i=1}^{N-1} p(r_{t_{i+1}}|r_{t_i}).$$

It is computationally convenient to work with the log-likelihood function

$$\ln L(\kappa,\theta,\sigma) = \sum_{i=1}^{N-1} \ln p(r_{t_{i+1}}|r_{t_i}),$$

from which we derive the log-likelihood function of the CIR process given by

$$\ln L = (N-1)\ln c + \sum_{i=1}^{N-1}\left(-u_{t_i} - v_{t_{i+1}} + \frac{1}{2}q\ln\frac{v_{t_{i+1}}}{u_{t_i}} + \ln I_q(2\sqrt{u_{t_i}v_{t_{i+1}}})\right), \quad (2.5)$$

where $u_{t_i} = cr_{t_i}e^{-\kappa\Delta t}$ and $v_{t_{i+1}} = cr_{t_{i+1}}$.

The monthly dataset is obtained from the CANSIM database (v122529) and covers the period from January 1980 to February 2017, providing 446 observations. We set the time step $\Delta t = 1/12$ here since we are using monthly data. We then find maximum likelihood estimates of parameters κ^*, θ^* and σ^* by maximizing the log-likelihood function (2.5).

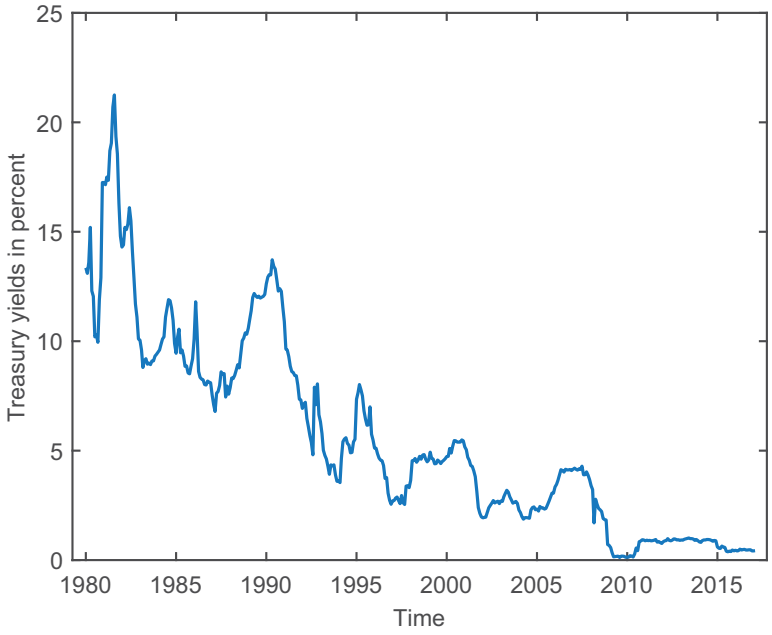

Figure 2.1: Canadian one-month treasury yields (in percents) from January 1980 to February 2017

	$\hat{\kappa}$	$\hat{\theta}$	$\hat{\sigma}$	$\ln L/N$
MLE	0.1090	0.0236	0.0681	-0.4076

Table 2.1: Maximum likelihood estimation results for CIR model. The $\ln L$ refers to the maximized value of the log-likelihood divided by the number of observations.

2.2.1 Empirical results

The time series of Canadian treasury yields is plotted in Figure 2.1. Table 2.1 reports the estimates and the log-likelihood of the MLE for the CIR model. Note that the parameters satisfy the Feller condition $2\kappa^*\theta^* > (\sigma^*)^2$.

Chapter 3

Martingale Representation Theorem Decomposition

In this chapter, we first present the properties that define a meaningful risk decomposition as proposed by Schilling et al. (2015). The life insurance modeling framework and the martingale representation theorem (MRT) decomposition are then introduced. We finally conclude with the discussion of the calculation of this decomposition.

3.1 Risk Decomposition

In general, an insurance contract is a transfer of risk from a policyholder to the insurer. Life insurance liabilities are influenced by various sources of risk such as financial factors, aggregate demographic trends, and actual deaths observed in the portfolio of insured. The latter two cause uncertain timings of cash flows in a long time horizon. In many cases, insurance companies need to assess the relative size of each source of risk in order to be able to allocate resources and propose efficient risk management strategies. Nevertheless, the interaction of various sources can be quite complex, so that the identification and quantification of each individual risk is not trivial. Therefore, an effective risk decomposition methodology is very important and has great practical value from an actuarial perspective.

Suppose an insurer's total risk is given by the (normalized) loss random variable

L, where $\mathrm{E}[L] = 0$. Assume there are k sources of risk, where $Z_i = Z_i(t)$ denotes the i-th source of risk and $Z = (Z_1, \ldots, Z_k)$. A decomposition methodology is a method that assigns each source of risk a corresponding risk factor.

Schilling et al. (2015) propose 6 properties that a *meaningful risk decomposition* should satisfy.

P1 *Randomness*

Individual risk factors are given by *random variables* $R_1, R_2, \ldots, R_k$, where random variable R_i corresponds to risk factor $i \in \{1, 2, ..., k\}$. We introduce the relation $\leftrightarrow$ for a decomposition methodology and write $(L, Z_1, \ldots, Z_k) \leftrightarrow (R_1, R_2, \ldots, R_k)$ to indicate that the loss L depending on $(Z_1, \ldots, Z_k)$ corresponds to the decomposition $R_1, R_2, \ldots, R_k$.

P2 *Attribution*

R_i represents the risk factor related to risk i. Formally, we require that whenever the loss L is $\sigma(Z_i)$-measurable and Z_i is independent of $(Z_1, \ldots, Z_{i-1}, Z_{i+1}, \ldots, Z_k)$, then $R_j = 0$ for all $j \neq i$.

P3 *Uniqueness*

The decomposition methodology yields a unique decomposition. Formally, we require that $(L, Z_1, \ldots, Z_k) \leftrightarrow (R_1, R_2, \ldots, R_k)$ and $(L, Z_1, \ldots, Z_k) \leftrightarrow (\tilde{R}_1, \tilde{R}_2, \ldots, \tilde{R}_k)$ implies $R_i = \tilde{R}_i, i \in \{1, 2, ..., k\}$.

P4 *Order invariance*

The decomposition is invariant to the order of the risks $1, 2, \ldots, k$. Formally, consider a permutation $\pi : \{1, 2, \ldots, k\} \to \{1, 2, \ldots, k\}$ and assume $(L, Z_1, \ldots, Z_k) \leftrightarrow (R_1, R_2, \ldots, R_k)$. Then we require:

$$(L, Z_{\pi(1)}, \ldots, Z_{\pi(k)}) \leftrightarrow (R_{\pi(1)}, R_{\pi(2)}, \ldots, R_{\pi(k)}).$$

P5 *Scale invariance*

The decomposition is invariant to changes in the scale of sources of risk. Formally, assume $(L, Z_1, \ldots, Z_k) \leftrightarrow (R_1, R_2, \ldots, R_k)$, and let $\tilde{Z}_i(t) := f_i(Z_i(t))$ for

all $i = 1, \ldots, k, 0 \leq t \leq T^*$, where, for each i, $f_i : \mathbb{R} \to \mathbb{R}$ is a smooth, invertible function. If $(L, \tilde{Z}_1, \ldots, \tilde{Z}_k) \leftrightarrow (\tilde{R}_1, \tilde{R}_2, \ldots, \tilde{R}_k)$, then we require that $R_i = \tilde{R}_i$ for all $i \in \{1, 2, ..., k\}$.

P6 *Aggregation*

The decomposition aggregates to the total risk faced by the company. Formally, we require that for each loss L and risks$Z$ with $(L, Z_1, \ldots, Z_k) \leftrightarrow (R_1, R_2, \ldots, R_k)$, there exists a function $A_{(L,Z)} : \mathbb{R}^k \to \mathbb{R}$ such that

$$L = A_{(L,Z)}(R_1, R_2, \ldots, R_k).$$

They also mention a special case of P6 called additive aggregation. It is a desirable property since it allows for the natural interpretation that the risk factors sum up to the total risk.

P6* *Additive aggregation*

A special case of P6 is an additive aggregation funtion, i.e. the case where L is given as the sum of the individual risk factors:

$$L = \sum_{i=1}^{k} R_i.$$

As the chapter title suggests, we introduce a decomposition approach named MRT decomposition in Section 4.3 that satisfies all these six meaningful risk decomposition properties.

3.2 Life Insurance Modeling Framework

First, we fix a finite time horizon T^* and a filtered probability space $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P})$ with $\mathbb{F} = (\mathcal{F}_t)_{0 \leq t \leq T^*}$. Throughout the chapter, $\mathcal{F}_t$ describes all the information available at time t and $\mathcal{F} = \mathcal{F}_{T^*}$. The main factors that affect annuity values are mortality rates and interest rates. More formally, we assume that the uncertainty arises from the evolution of the short rate $r(t)$ and the mortality intensity $\mu(t)$ as well as the observed number of deaths.

For the former two factors, we introduce an 2-dimensional state process $X(t) = (X_1(t), X_2(t))'$, $0 \leq t \leq T^*$, where the symbol $'$ denotes transpose. In our modeling framework, the state process is a 2-dimensional diffusion process satisfying

$$dX(t) = \theta(t)dt + \sigma(t)dW(t), \tag{3.1}$$

with deterministic initial value $X(0) = (X_1(0), X_2(0))' \in \mathbb{R}^2$. For simplicity, we assume that the state process is driven by a 2-dimensional standard Brownian motion $W = (W_1(t), W_2(t))'$, where $W_1(t)$ and $W_2(t)$ are independent. Therefore, $\theta(t)$ is a two dimensional drift vector and $\sigma(t)$ is a 2×2 volatility matrix here. Since $r(t)$ and $\mu(t)$ are risk sources in an annuity, we choose them as components of the state process $X(t)$. Thus we write the state process as $X(t) = (r(t), \mu(t))'$.

For the latter source of risk, i.e. the actual occurrence of deaths, we consider m homogeneous policyholders aged x at time 0. The remaining lifetime τ_x^i of policyholder i as seen from $0, i = 1, \ldots, m$, is defined as the first jump time of a doubly stochastic or Cox process with intensity $(\mu(t))_{0 \leq t \leq T^*}$, i.e.

$$\tau_x^i = \inf \left\{ t \in [0, T^*] : \int_0^t \mu(s)ds \geq E_i \right\}, i = 1, \ldots, m, \tag{3.2}$$

where $E_i, i = 1, \ldots, m$, are i.i.d. exponential random variables with mean 1 which are all independent of the filtration $\mathcal{G}$ generated by intensity process $\mu(t)$. We use the convention $\inf \emptyset = \infty$. Therefore the first jump time can be interpreted as the time required to consume an exponential random variable if it is consumed at rate $\mu(s)$ at time s. Thus this definition also gives us a way to simulate the values of τ_x^i.

Note that the residual lifetimes $\tau_x^i, i = 1, \ldots, m$, of the homogeneous policyholders are by construction conditionally i.i.d. random variables given the σ-algebra $\mathcal{G}_{T^*}$. Finally, we denote the number of deaths up to time t by the counting process

$$N(t) = \sum_{i=1}^{m} \mathbb{I}\{\tau_x^i \leq t\}, \tag{3.3}$$

where $\mathbb{I}$ denotes indicator function, having the value of 1 when the ith policyholder dies before time t.

In order to keep the presentation concise, we consider the time-0 present value L_0 of an insurer's future losses. It is the sum of discounted future cash flows as from

time 0 and is given by:

$$L_0 = C_0 + \sum_{k=0}^{n}(m - N(t_k))C_{a,k} + \sum_{k=1}^{n}(N(t_k) - N(t_{k-1}))C_{ad,k}, \tag{3.4}$$

where $0 = t_0 < t_1 < \cdots < t_n = T^*, n \in \mathbb{N}$, are discrete points in time. There are three different cash flows in (3.4):

C_0 the discounted payments independent of the remaining lifetimes.

$C_{a,k}$ the sum of all discounted payments at or after time t_k that are conditional on survival until time t_k, $k = 0, \ldots, n$. Discrete annuity payments fall into this category.

$C_{ad,k}$ the sum of all discounted payments at or after time t_k that are conditional on death within time interval $(t_{k-1}, t_k]$, $k = 1, \ldots, n$.

Thus L_0 represents the insurer's total net liability at time 0. To clarify expressions, we remind that positive payments are the amounts made by the insurance company while payments received are negative.

For annuity portfolios, all the payments made by insurance companies are conditional on the survival of policyholders. Therefore, we have $C_0 = 0$, $C_{a,k} = e^{-\int_0^{t_k} r(s)ds}$, and $C_{ad,k} = 0$. Thus the total discount loss (3.4) can be written as

$$L_0 = \sum_{k=0}^{n}(m - N(t_k))e^{-\int_0^{t_k} r(s)ds}. \tag{3.5}$$

For the determination of superscript n, we introduce the limiting age ω, means that for a life table $l_n = 0$ for all $n \geq \omega$. Then index n can be determined by the initial cohort age x and the limiting age ω as $\omega - x - 1$.

Moreover, we need a total risk random variable when dealing with risk decompositions. Thus, the insurer's risk at time 0 is identified with $L := L_0 - \mathrm{E}[L_0]$, which has mean 0. We remark here that the net liability L_0 is the (stochastic) amount of money the insurance company needs at time 0 to pay off future contract obligations. Since the company should at least prepare for the expected value $\mathrm{E}[L_0]$ (equivalence principle), risk is interpreted as the excessive part.

3.3 MRT Decomposition

This section is mainly an introduction to the propositions and results obtained by Schilling et al. (2015). They propose a decomposition into stochastic integrals with respect to the compensated sources of risk. These integrals represent the risk factors of the corresponding sources of risk.

In an annuity portfolio, we have three sources of risk, namely $r(t), \mu(t)$, and the actual number of deaths $N(t)$. Therefore, we have three different risk factors in the MRT decomposition according to property P1 randomness. They are denoted by interest risk R_1, systematic mortality risk R_2, and unsystematic mortality risk R_3, respectively. To see where the names come from, we note that the first two risk factors R_1 and R_2 are connected with the state process $X(t) = (r(t), \mu(t))'$. In fact, they are associated with the changes in the underlying stochastic processes $r(t)$ and $\mu(t)$, whereas the third risk factor R_3 is identified with the randomness of deaths $N(t)$ in a portfolio with fixed mortality intensity. That is because the process $N(t)$ is binomial conditionally upon the path of mortality intensity $\mu(t)$. Since R_2 and R_3 are all related to the mortality risk, we distinguish them by the name systematic risk factor R_2 and unsystematic risk factor R_3. As Dahl and Møller (2006) point out, the most significant difference between them is that the systematic mortality risk is a non-diversifiable risk, which does not disappear when the size of the portfolio increases, whereas the unsystematic mortality risk is diversifiable, i.e. it vanishes as the number of policyholders goes to infinity.

The corresponding compensated processes, i.e. the process less their $\mathbb{F}$-compensators, are denoted by $M_1^W(t)$, $M_2^W(t)$, and $M^N(t)$ respectively. The notion of the compensator comes from a classic theorem – Doob decomposition, which expresses a submartingale $Y(t)$ as the sum of a martingale $M(t)$ and an increasing predictable process $S(t)$. Namely, $Y(t) = M(t) + S(t)$. Then the process $S(t)$ is called the compensator of $Y(t)$. Schilling et al. (2015) obtain the following proposition.

Proposition 3.3.1 *Suppose the state process is* $X(t) = (X_1(t), X_2(t))'$*, the corre-*

sponding compensated process of $X_i(t)$ is given by

$$M_i^W(t) = \int_0^t \sigma_{i1}(s)dW_1(s) + \int_0^t \sigma_{i2}(s)dW_2(s),\ 0 \le t \le T^*, i = 1, 2.$$

The compensated process of $N(t)$ is

$$M^N(t) = N(t) - \int_0^t (m - N(s-))\mu(s)ds,\ 0 \le t \le T^*.$$

Proof. A proof can be found in Schilling et al. (2015), p.12. □

In fact, the idea behind Proposition 3.3.1 is to obtain martingales. $M_i^W(t)$ is a martingale because it is a sum of two Ito integrals. For the compensated process $M^N(t)$, Schilling et al. (2015) use conclusions from Bielecki and Rutkowski (2013) to prove that $M^N(t)$ is also a martingale.

Base on this proposition, Schilling et al. (2015) propose the so called MRT decomposition

$$L := L_0 - \mathrm{E}[L_0] = \sum_{i=1}^{2} \int_0^{T^*} \psi_i^W(t)dM_i^W(t) + \int_0^{T^*} \psi^N(t)dM^N(t) \tag{3.6}$$

$$= \sum_{i=1}^{2} R_i + R_3. \tag{3.7}$$

The idea and proof are constructed on the Martingale Representation Theorem, hence a decomposition of the form (3.6) is called MRT decomposition. Each integral in this decomposition is interpreted as the portion of the total randomness of L caused by the associated source of risk. Thus, the risk factors are given by $R_i := \int_0^{T^*} \psi_i^W(t)dM_i^W(t)$, $i = 1, 2$, and $R_3 := \int_0^{T^*} \psi_i^N(t)dM_i^N(t)$.

The following more general proposition justifies this decomposition.

Proposition 3.3.2 *Assuming that* $\det \sigma(t) \neq 0$ *in (3.1) and that random variable L_0 is square integrable, then there exists processes $\psi_1^W(t), \ldots, \psi_n^W(t), \psi^N(t)$ such that the MRT decomposition (3.6) holds. The representation is unique and $\int_0^{T^*} \psi^N(t)dM^N(t)$ is square integrable, i.e.*

$$\mathrm{E}\Big[\Big(\int_0^{T^*} \psi^N(t)dM^N(t)\Big)^2\Big] < \infty.$$

Proof. Proof can be found in Schilling et al. (2015), pp. 13. □

Proposition 3.3.2 ensures us the existence and uniqueness of the MRT decomposition, which is very important, but the calculation of different risk factors is of equivalent importance. Schilling et al. (2015) indeed also discuss this issue in their paper. Since processes $M_i^W(t)$ and $M^N(t)$ are given by Proposition 3.3.1, the remaining calculation amounts to the determination of integrands $\psi_1^W(t), \ldots, \psi_n^W(t), \psi^N(t)$ in (3.6). For each summand of L_0 defined in (3.4), they provide a corresponding calculation scheme and thus the MRT decomposition of L_0 itself can be obtained by summing up the individual decompositions.

Since our L_0 in (3.5) only involves discrete survival cash flows, we focus on the calculation of the decomposition for $C_{a,k}$.

Proposition 3.3.3 *Let $L_{0,k} = (m - N(t_k))C_{a,k}$, for $0 \leq t_k \leq T^*$ and assume that $C_{a,k}$ is of the form*

$$C_{a,k} = e^{-\int_0^{t_k} g(s,X(s))ds} h(X(t_k)),$$

where X is the state process specified in (3.1). Define f by

$$f_k(t,x) := \mathrm{E}\big[e^{-\int_t^{t_k}(\mu(s,X(s))+g(s,X(s)))ds} h(X(t_k))|X(t) = x\big]. \tag{3.8}$$

Then the unique integrands of the MRT decomposition of $L_{0,k} - \mathrm{E}[L_{0,k}]$ are given by

$$\begin{aligned}
\psi_i^W(t) &= \mathbb{1}_{[0,t_k]}(t)(m - N(t-))e^{-\int_0^t g(s,X(s))ds} \frac{\partial f}{\partial x_i}(t, X(t)),\ i = 1, 2, \\
\psi^N(t) &= -\mathbb{1}_{[0,t_k]}(t)e^{-\int_0^t g(s,X(s))ds} f(t, X(t)).
\end{aligned}$$

Proof. Proof can be found in Schilling et al. (2015), p. 23. □

Next we proceed to deriving the MRT decomposition of L_0 defined in (3.5). First note that $C_{a,k} = e^{-\int_0^{t_k} g(s,X(s))ds} h(X(t_k)) = e^{-\int_0^{t_k} r(s)ds}$ for $g \equiv r$ and $h \equiv 1$. Then we define function $f_k(t, r, \mu)$ as

$$\begin{aligned}
f_k(t,r,\mu) &= \mathrm{E}\big[e^{-\int_t^{t_k}(\mu(s,X(s))+r(s,X(s)))ds}|X(t) = x\big], \\
&= \mathrm{E}\big[e^{-\int_t^{t_k}(\mu(s)+r(s))ds}|r(t) = r, \mu(t) = \mu\big],\ 0 \leq t \leq t_k.
\end{aligned} \tag{3.9}$$

Notice that we omit $X(s)$ in the second line since $\mu(t)$ and $r(t)$ are components of the state process $X(t)$. This function can be simplified by using the independence between r and μ and the affine property of r.

Applying Proposition 3.3.3, we obtain the integrands of the MRT decomposition of $C_{a,k}$

$$\begin{aligned}
\psi^W_{1,k}(t) &= \mathbb{1}_{[0,t_k]}(t)(m - N(t-))e^{-\int_0^t r(s)ds}\frac{\partial f_k}{\partial r}(t,r,\mu),\\
\psi^W_{2,k}(t) &= \mathbb{1}_{[0,t_k]}(t)(m - N(t-))e^{-\int_0^t r(s)ds}\frac{\partial f_k}{\partial \mu}(t,r,\mu),\\
\psi^N_{k}(t) &= -\mathbb{1}_{[0,t_k]}(t)e^{-\int_0^t r(s)ds} f_k(t,r,\mu).
\end{aligned}$$

Summing up the decompositions of $C_{a,k}$, we obtain the MRT decomposition of L as

$$L = L_0 - \mathrm{E}[L_0] = R_1 + R_2 + R_3, \tag{3.10}$$

where the systematic risk factors R_1 and R_2 implied by r and μ are given by

$$R_1 = \sum_{k=0}^{n}\int_0^{t_k}(m - N(t-))e^{-\int_0^t r(s)ds}\frac{\partial f_k}{\partial r}(t,r,\mu)dM^W_1(t), \tag{3.11}$$

$$R_2 = \sum_{k=0}^{n}\int_0^{t_k}(m - N(t-))e^{-\int_0^t r(s)ds}\frac{\partial f_k}{\partial \mu}(t,r,\mu)dM^W_2(t), \tag{3.12}$$

respectively, and the unsystematic mortality risk factor R_3 is given by

$$R_3 = -\sum_{k=0}^{n}\int_0^{t_k} e^{-\int_0^t r(s)ds} f_k(t,r,\mu)dM^N(t). \tag{3.13}$$

Finally we obtain the explicit MRT decomposition of a whole life annuity, numerical calculations and detailed analysis will be carried out in the next section.

Chapter 4

Numerical Example

In this chapter, we demonstrate the applicability and usefulness of the MRT decomposition by presenting some numerical examples. To calculate risk factors R_1, R_2, and R_3, we first project the stochastic mortality model and the interest rate model. Then we give some details about the computation by deducing more explicit expressions for risk factors and conditional expectations. Finally, we analyze the base scenario of which the age is 65 years old and compare it with other scenarios.

4.1 Projection

4.1.1 Mortality Model

In Section 1.6.1, we derived a generalized RH model based on the stochastic process $\kappa(t)$. The mortality process follows the dynamic

$$\begin{aligned} \ln \mu(t,x) &= \beta_1(x+t) + \beta_2(x+t)\kappa(t) + \beta_3(x+t)\gamma, \\ d\kappa(t) &= \theta dt + \sigma dW_\mu(t), \end{aligned} \tag{4.1}$$

where x denotes the policyholder's age at time 0 and $W_\mu(t)$ is a Brownian motion. The initial mortality intensities $\mu(0,x)$ are calculated from the calibration procedure using the last estimated value for $\kappa_t^{(2)}$.

The first step is to fit the RH model (1.4) to Canadian mortality data – males aged between 40 and 99 over the period 1960 to 2011. Figure 4.1 shows the calibration

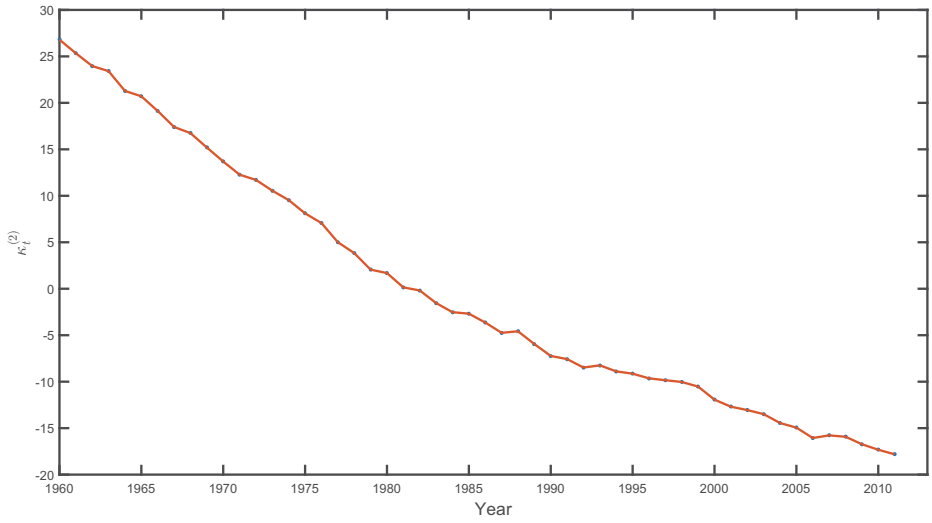

Figure 4.1: Estimated $\kappa_t^{(2)}$ over the fitting period 1960–2011 for ages 40–99.

results for $\kappa_t^{(2)}$. Based on the maximum likelihood estimators (1.3), the parameters θ and σ in (4.1) are given by

$$\theta = \frac{\kappa_{2011}^{(2)} - \kappa_{1960}^{(2)}}{2011 - 1960} = \frac{-17.8133 - 26.7997}{51} = -0.8748,$$
$$\sigma = \frac{1}{2011 - 1960} \sum_{t=1961}^{2011} (\kappa_t^{(2)} - \kappa_{t-1}^{(2)} - \theta)^2 = 0.5795,$$

where $\kappa(0) = \kappa_{2011}^{(2)} = -17.8133$.

Then we are ready to project $\kappa(t)$. Figure 4.2 shows the mean sample path and its 95% confidence interval generated by the time index $\kappa(t)$ where the time horizon is set to 35 years.

From the calibration procedure, we obtain the estimated $\hat{\beta}_x^{(1)}$, $\hat{\beta}_x^{(2)}$ and $\hat{\beta}_x^{(3)}$ where $x = 40, 41, \ldots, 99$. In continuous-regime, we need three continuous functions since those estimated values are just point estimates at each integer age $x + t$. One of the approaches to get smooth functions $\beta_1(x + t)$, $\beta_2(x + t)$ and $\beta_3(x + t)$ through polynomials. We present the estimated values and graphs of smoothed functions in Figures 4.3, 4.4, and 4.5. We assume that all policyholders are of age $x = 65$ at time 0, then the value of the cohort parameter γ is -13.7242 corresponding the cohort year 1946.

In order to give readers a picture of simulated mortality intensity paths, we use an

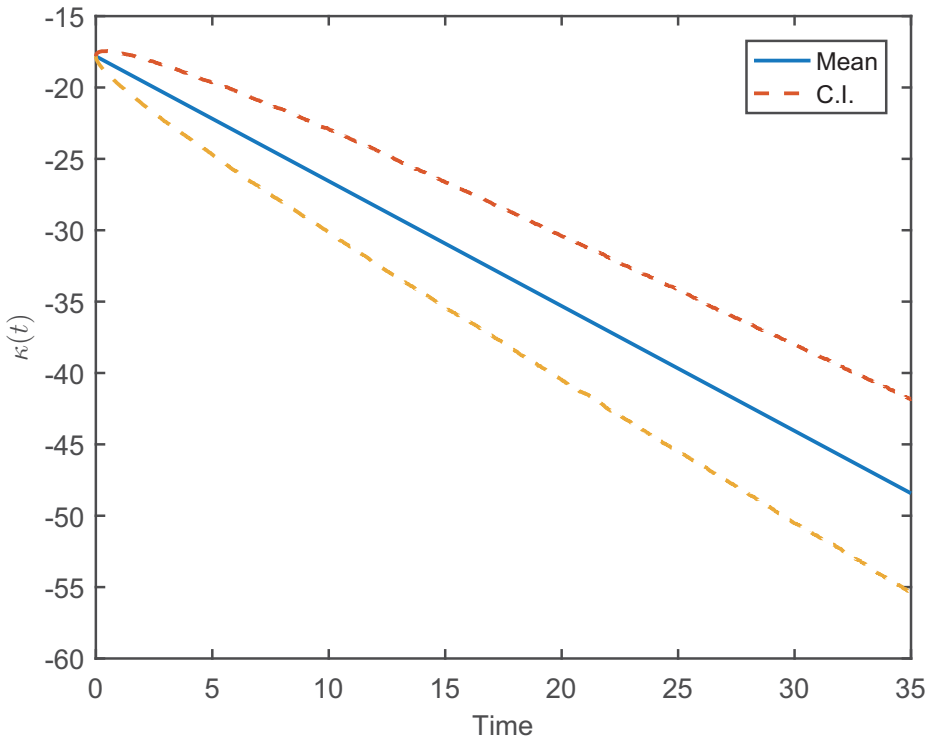

Figure 4.2: Mean sample path and its 95% confidence interval for $\kappa(t)$ over time $0 - 35$.

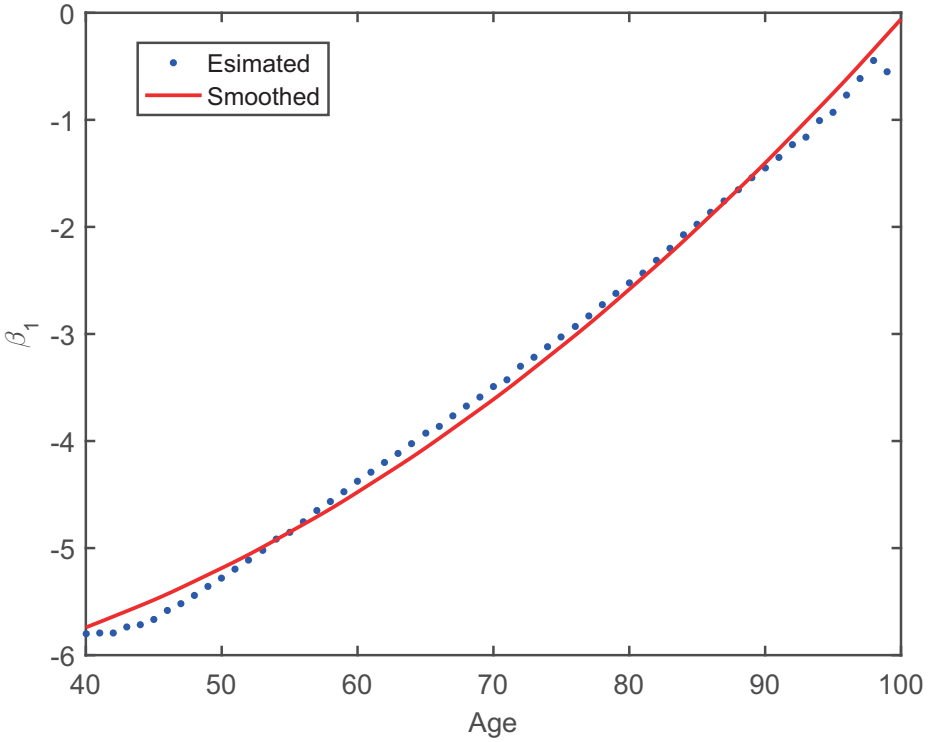

Figure 4.3: Estimated $\hat{\beta}_x^{(1)}$(dots) and smoothed function $\beta_1(x)$(line).

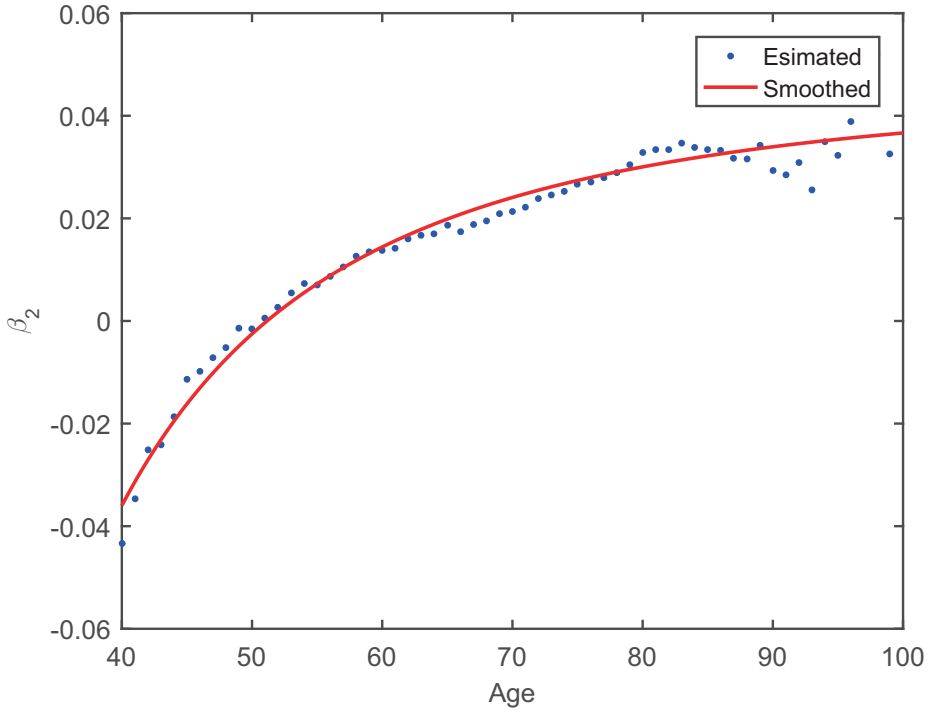

Figure 4.4: Estimated $\hat{\beta}_x^{(2)}$(dots) and smoothed function $\beta_2(x)$(line).

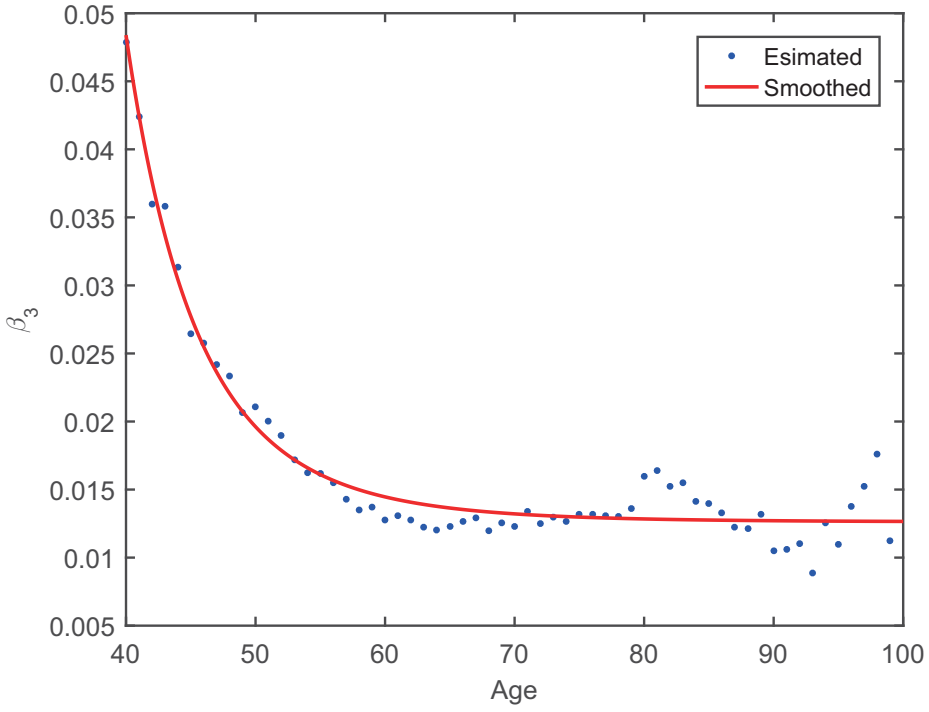

Figure 4.5: Estimated $\hat{\beta}_x^{(3)}$(dots) and smoothed function $\beta_3(x)$ (line).

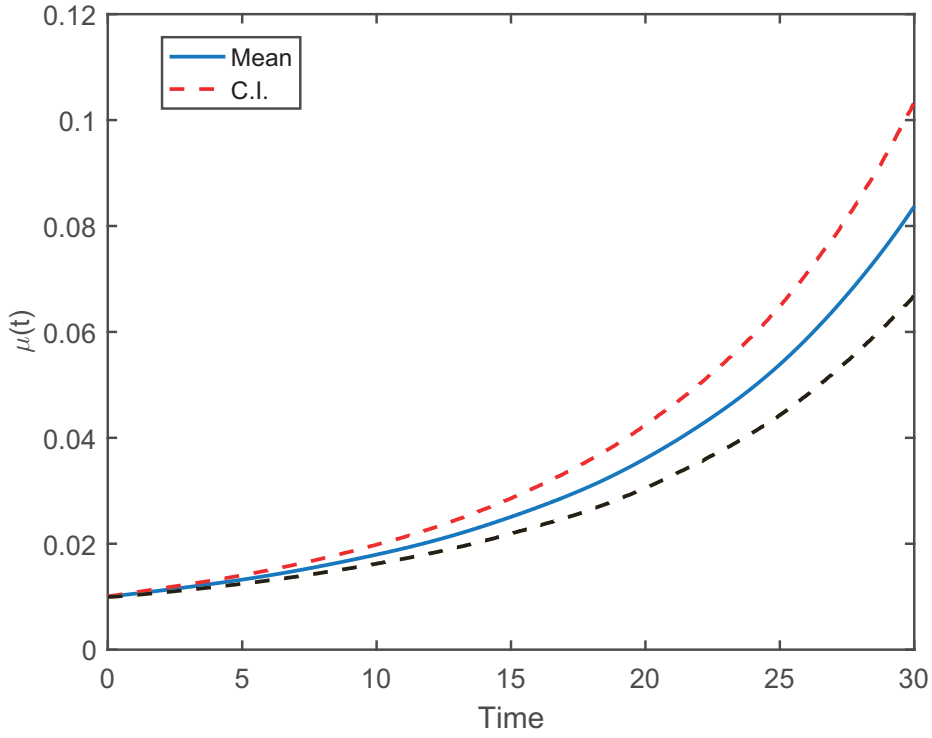

Figure 4.6: Mean sample path and its 95% confidence interval for $\mu(t)$ over time $0 - 30$.

Euler scheme with $n = 100$ time steps per year and perform $N = 10,000$ simulations. Figure 4.6 shows the averaged path of $10,000$ runs and the 95% confidence interval. In what follows, the initial age x is always fixed, so that we no longer indicate the dependency on the age cohort but just write $\mu(t)$ and $\beta(t)$. The limiting age ω is set to 115 years old so $\beta_1(t)$, $\beta_2(t)$, and $\beta_3(t)$ are extended according to respective smoothed functions when $t \geq 100$.

4.1.2 Interest Rate Model

We choose the CIR model $dr(t) = \kappa(\theta - r(t))dt + \sigma\sqrt{r(t)}dW_r(t)$ as our short rate model. The estimation results have been obtained in Chapter 2.

First note that model parameters κ, θ, and σ are all positive and that the Feller condition $2\kappa\theta \geq \sigma^2$ holds. Therefore the CIR process is generated according to the

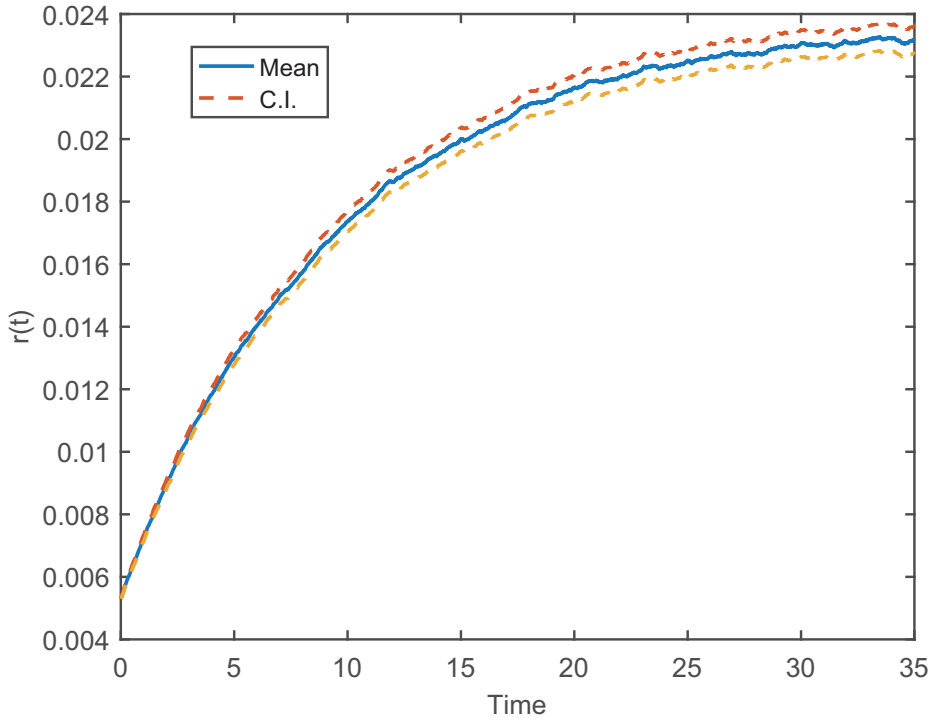

Figure 4.7: Mean sample path and its 95% confidence interval for $r(t)$

following Euler method:

$$r(t_i) = r(t_{i-1}) + \kappa\big(\theta - r(t_{i-1})\big)(t_i - t_{i-1}) + \sigma\sqrt{r(t_i)^+}\sqrt{t_i - t_{i-1}}Z_i, \tag{4.2}$$

where $0 = t_0 < t_1 < \cdots < t_n = T, i = 1, \ldots, n$. Z_i is a standard normal random variable. $r(t_i)^+ = \max\left(r(t_i), 0\right)$ represents the greater value of $r(t_i)$ and 0 since it is possible to generate negative $r(t_i)$'s in the simulation process. As in the mortality model, we use an Euler scheme with $n = 100$ time steps per year and perform $N = 10,000$ simulations. Figure 4.7 shows the mean sample path and its 95% confidence interval.

4.1.3 MRT Decomposition

We demonstrate the applicability and usefulness of the MRT decomposition by presenting some numerical examples involving the risk factors R_1, R_2, and R_3 defined by (3.10) – (3.12). Since we assume that $W_r(t)$ and $W_\mu(t)$ are independent Brownian motions, the volatility matrix of the state process $\sigma(t, r, \mu) = diag\{\sigma_r\sqrt{r(t)}, \sigma_\mu\beta_2(t)\mu(t)\}$

is a 2×2 diagonal matrix, which simplifies the compensated processes $M_1^W(t)$ and $M_2^W(t)$. Applying Proposition 3.3.1, we obtain the following equations:

$$dM_1^W(t) = \sigma_r \sqrt{r(t)} dW_r(t), \tag{4.3}$$

$$dM_2^W(t) = \sigma_\mu \beta_2(t) \mu(t) dW_\mu(t), \tag{4.4}$$

$$dM^N(t) = dN(t) - (m - N(t-))\mu(t)dt. \tag{4.5}$$

Therefore risk factors R_1, R_2, and R_3 are further simplified as

$$\begin{aligned} R_1 &= \sum_{k=0}^{n} \int_0^{t_k} (m - N(t-)) e^{-\int_0^t r(s)ds} \frac{\partial f_k}{\partial r}(t, r, \mu) dM_1^W(t) \\ &= \sum_{k=0}^{n} \int_0^{t_k} (m - N(t-)) e^{-\int_0^t r(s)ds} \frac{\partial f_k}{\partial r}(t, r, \mu) \sigma_r \sqrt{r(t)} dW_r(t). \end{aligned} \tag{4.6}$$

$$\begin{aligned} R_2 &= \sum_{k=0}^{n} \int_0^{t_k} (m - N(t-)) e^{-\int_0^t r(s)ds} \frac{\partial f_k}{\partial \mu}(t, r, \mu) dM_2^W(t) \\ &= \sum_{k=0}^{n} \int_0^{t_k} (m - N(t-)) e^{-\int_0^t r(s)ds} \frac{\partial f_k}{\partial \mu}(t, r, \mu) \sigma_\mu \beta_2(t) \mu(t) dW_\mu(t). \end{aligned} \tag{4.7}$$

$$\begin{aligned} R_3 &= -\sum_{k=0}^{n} \int_0^{t_k} e^{-\int_0^t r(s)ds} f_k(t, r, \mu) dM^N(t) \\ &= -\sum_{k=0}^{n} \int_0^{t_k} e^{-\int_0^t r(s)ds} f_k(t, r, \mu) \Big(dN(t) - (m - N(t-))\mu(t)dt \Big) \\ &= \sum_{k=0}^{n} \int_0^{t_k} (m - N(t-)) e^{-\int_0^t r(s)ds} \mu(t) dt - \sum_{k=0}^{n} \int_0^{t_k} e^{-\int_0^t r(s)ds} f_k(t, r, \mu) dN(t) \end{aligned} \tag{4.8}$$

Now that we have improved representations for R_1, R_2, and R_3, we could present more details about the calculations of risk factors before stepping into the analysis of numerical results. At first, we generate one sample path for $W_r(t)$. Then the short rate $r(t)$ can be projected applying Euler discretization method (4.2). Similarly, mortality intensity $\mu(t)$ can also be generated using (4.1). As for the counting process $N(t)$, we

first generate an exponential random variable E_i for each policyholder, $i = 1, \ldots, m$. Then the residual lifetime τ_x^i for each policyholder is determined by (3.2). Therefore $N(t)$ is calculated using (3.3).

Once we obtain sample paths for processes $r(t)$, $\mu(t)$, and $N(t)$, we can focus on the calculation of $f_k(t, r, \mu)$ and its partial derivatives. By the definition (3.9),

$$\begin{aligned} f_k(t, r, \mu) &= \mathrm{E}\big[e^{-\int_t^{t_k}(\mu(s)+r(s))ds} | r(t) = r, \mu(t) = \mu\big] \\ &= \mathrm{E}\big[e^{-\int_t^{t_k} r(s)ds} | r(t) = r\big]\mathrm{E}\big[e^{-\int_t^{t_k} \mu(s)ds} | \mu(t) = \mu\big] \\ &= A(t, t_k)e^{-B(t,t_k)r}\mathrm{E}\big[e^{-\int_t^{t_k} \mu(s)ds} | \mu(t) = \mu\big]. \end{aligned} \tag{4.9}$$

The second line follows from the independence of the short rate $r(t)$ and the mortality intensity $\mu(t)$. The third line is simplified by substituting the bond price (2.3). Since $\mu(t)$ does not possess the affine property, there is no closed-form formula for the conditional expectation with respect to $\mu(t)$. Therefore we resort to Monte Carlo simulations for the calculation of $\mathrm{E}\big[e^{-\int_t^{t_k} \mu(s)ds} | \mu(t) = \mu\big]$. We also need two partial derivatives of $f_k(t, r, \mu)$ in the integrands of R_1 and R_2. From (4.9), we obtain the partial derivative of $f_k(t, r, \mu)$ with respect to r as

$$\frac{\partial f_k}{\partial r}(t, r, \mu) = -A(t, t_k)B(t, t_k)e^{-B(t,t_k)r}\mathrm{E}\big[e^{-\int_t^{t_k} \mu(s)ds} | \mu(t) = \mu\big]. \tag{4.10}$$

As for the partial derivative with respect to μ, we define a function

$$g(t, \mu) = \mathrm{E}\big[e^{-\int_t^{t_k} \mu(s)ds} | \mu(t) = \mu\big]$$

to make notation concise. Thus the calculation of $\frac{\partial f_k}{\partial \mu}(t, r, \mu)$ reduces to find the derivative of $g(t, \mu)$, which is approximated using finite difference. More precisely, we use the equation

$$\frac{d}{d\mu}g(t, \mu) \approx \frac{g(t, \mu + h) - g(t, \mu)}{h}.$$

4.2 Analyses

Now we are able to calculate R_1, R_2, and R_3 since we have obtained building blocks in those integrands from above. For the numerical analysis, we consider $m = 100$

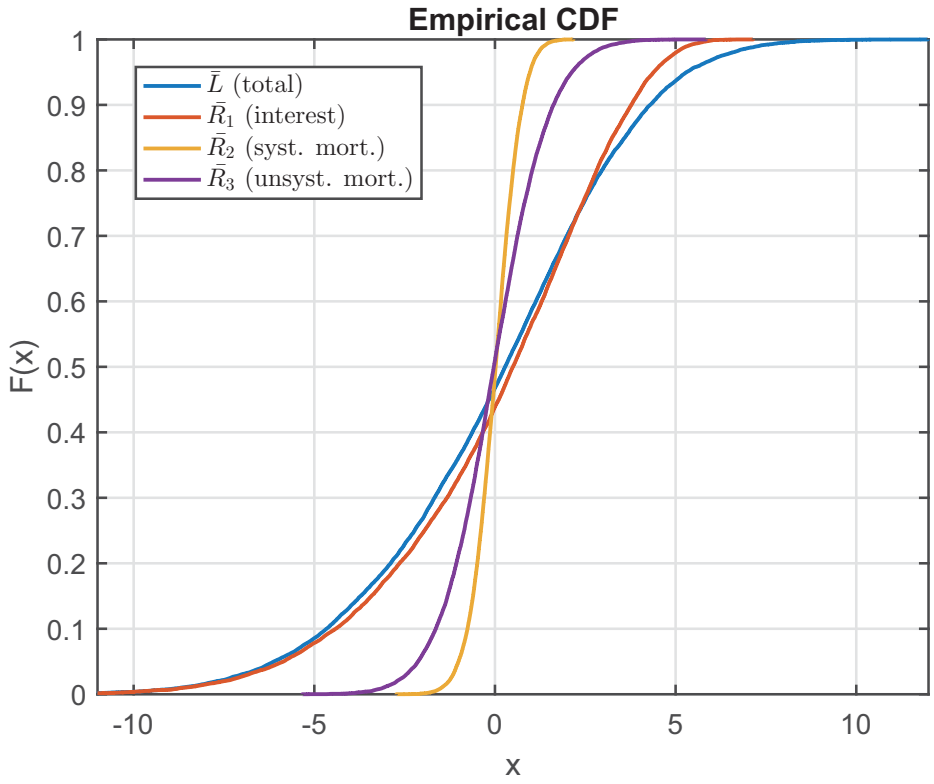

Figure 4.8: Empirical distribution functions at age 65.

policyholders and perform $N = 25{,}000$ simulations to estimate the distributions of L, R_1, R_2, and R_3. We use the Euler scheme with $n = 100$ time steps per year for projecting the risk drivers $r(t)$ and $\mu(t)$ as well as for approximating the stochastic integrals.

For our base scenario, policyholders are assumed to be 65 years old at time 0. We focus on the distributions scaled by the number of policyholders in the portfolio, i.e. we consider $\bar{L} := \frac{L}{m}$, $\bar{R}_i := \frac{R_i}{m}, i = 1, 2, 3$. The resulting empirical distribution functions of the total risk $\bar{L}$, the interest risk $\bar{R}_1$, the systematic mortality risk $\bar{R}_2$, and the unsystematic mortality risk $\bar{R}_3$ are shown in Figure 4.8. We find that the distribution function of the interest risk factor ($\bar{R}_1$) is right-skewed while the distribution functions of all other risk factors are approximately symmetric. Moreover, the plots indicate that the interest (R_1) is the most relevant risk driver since its distribution function is the most similar to the total risk $\bar{L}$.

The tails of risk factors are also of our concerns. We sort the respective outcomes

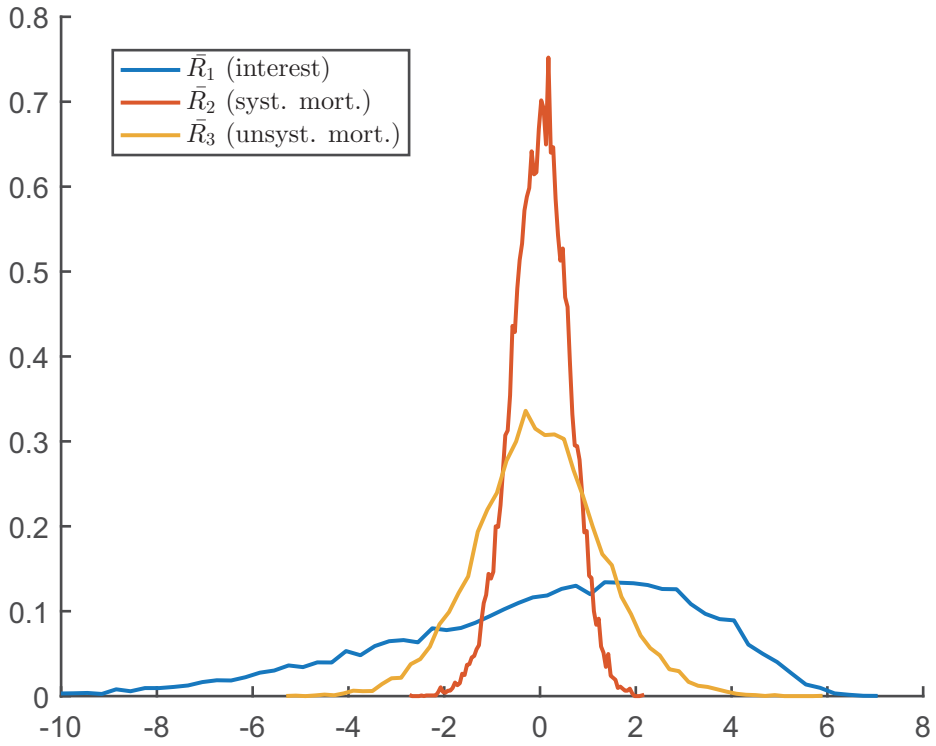

Figure 4.9: Probability density function estimates at age 65.

into spaced bins and plot the corresponding relative frequencies in Figure 4.9. We observe that the tails of the interest and the unsystematic mortality risk are the two heaviest ones among the three risk factors. The right-skewed characteristic of the interest risk factor R_1 and the symmetric characteristic of R_2 and R_3 is rather clear from Figure (4.9). If we confine our attention to the systematic mortality risk factor R_2 and unsystematic mortality risk factor R_3, we could find that the range of likely outcomes of R_3 is rather wide compared to the ranges of R_2. Therefore, R_3 is more variable then R_2.

In order to quantify contributions made by different risk factors and confirm our observations from the empirical distribution functions presented in Fig 4.8 and the relative frequencies Fig 4.9, we resort to the so-called Euler principle. More precisely, for a homogeneous risk measure ρ, Tasche (2007) proposes the following principle to uniquely determine the risk contribution of each risk factor with

$$\rho(R_i|L) = \frac{d\rho}{dh}(L + R_i h)|_{h=0}, \ i = 1, 2, 3. \tag{4.11}$$

	$\bar{L}$	$\bar{R}_1$	$\bar{R}_2$	$\bar{R}_3$
Std	3.4801	2.8960 (83.2%)	0.1092 (3.1%)	0.4743 (13.6%)
$\text{VaR}_{0.99}$	7.3065	4.3492 (60.3%)	0.4254 (5.9%)	2.4343 (33.8%)
$\text{TVaR}_{0.99}$	8.2240	4.9618 (60.3%)	0.6503 (7.9%)	2.6107 (31.7%)

Table 4.1: The total risk capital and the Euler risk contributions for a whole life annuity portfolio with $m = 100$ contracts at age 65.

We use three risk measures: standard deviation (Std), Value-at-Risk at the 99% level ($\text{VaR}_{0.99}$), and Tail-Value-at-Risk at the 99% level ($\text{TVaR}_{0.99}$) to report the total risk capital $\rho(\bar{L})$ as well as the risk contributions according to the Euler principle (4.11). Table 4.1 provides results for both absolute values and percentages of the sum of the three risk contributions.

The interest risk makes about 83% of the total risk capital in the standard deviation risk measure and about 60% in the VaR risk measure. It confirms our observation that R_1 is the significant factor. The unsystematic risk R_3 makes roughly 14% in the standard deviation risk measure then increases to 34% in the VaR risk measure and 32% in the TVaR measure. It suggests that R_3 becomes more important in the tail of the aggregate risk. The systematic risk factor R_2 also exhibits the same pattern which increases from 3% to 6%, and 8%, respectively. This increasing trend indicates that the systematic mortality risk is more sensitive in the tails.

4.2.1 Number of Policyholders

Table 4.2 provides results for the three risk measure where we set the number of policyholders to 1000. The most striking difference between Tables (4.1) and (4.2) is that the systematic risk factor R_2 is always larger than the unsystematic risk R_3. This is caused by the intrinsic difference between the systematic mortality risk and the unsystematic mortality risk. As we mentioned in Chapter 3, the unsystematic mortality risk is diversifiable so that it is not surprising to see that R_3 takes smaller percentages when the number of policyholders is increased to 1000. In this case, the

	$\bar{L}$	$\bar{R}_1$	$\bar{R}_2$	$\bar{R}_3$
Std	3.2146	3.0543 (95.0%)	0.1136 (3.5%)	0.0465 (1.4%)
$\text{VaR}_{0.99}$	6.0801	4.8044 (81.4%)	0.7890 (13.4%)	0.3073 (5.2%)
$\text{TVaR}_{0.99}$	6.6295	5.3515 (80.7%)	0.8769 (13.2%)	0.3995 (6.0%)

Table 4.2: The total risk capital and the Euler risk contributions for a whole life annuity portfolio with $m = 1000$ contracts at age 65.

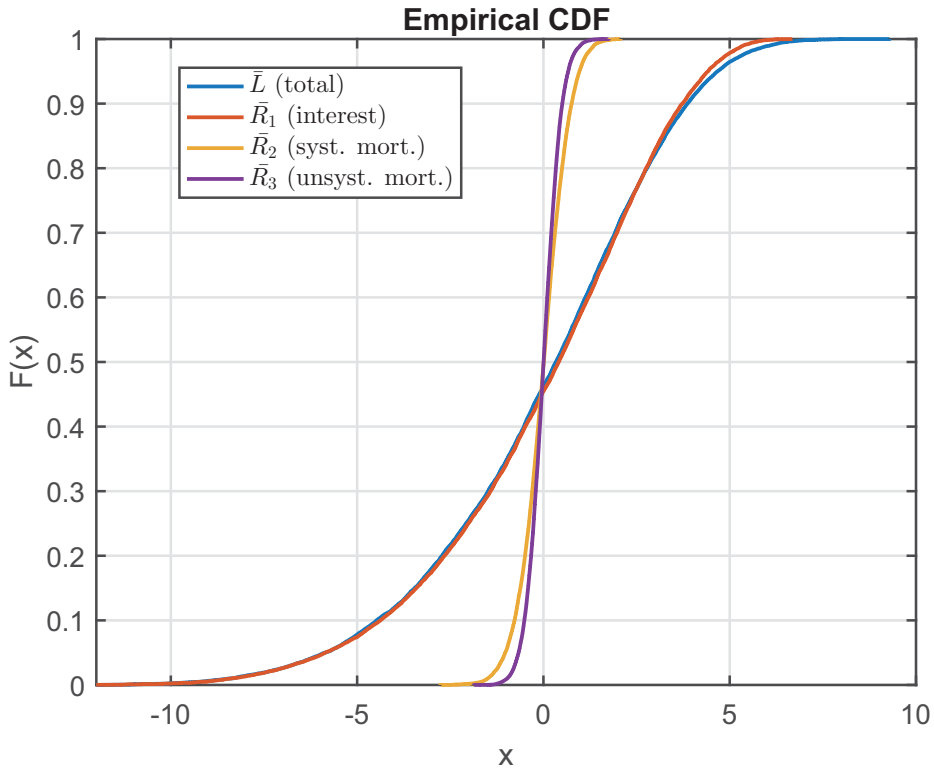

Figure 4.10: Empirical distribution functions at age 65 for $m = 1000$.

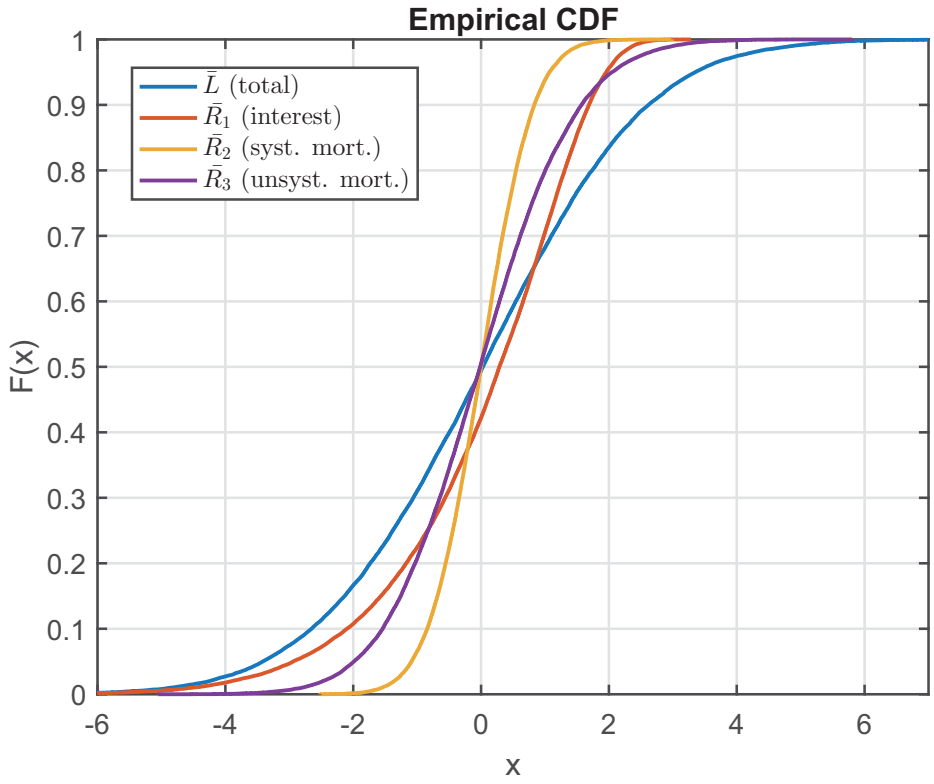

Figure 4.11: Empirical distribution functions at age 75.

interest risk is the most significant factor accounting for between about 81% and 95%, whereas unsystematic mortality risk exhibits the smallest risk contribution which is less than 6%. Figure 4.10 illustrates the distribution functions of the three risk factors and the total risk for a portfolio of 1000 contracts. Clearly, the distribution of interest risk R_1 very closely resembles the distribution of the total risk L. While the distribution functions of R_2 and R_3 are still symmetric, now the unsystematic mortality risk R_3 has a more concentrated distribution which is opposite from Fig 4.8 where R_2 is narrower than R_3.

4.2.2 Ages

In order to further explore interactions between three risk factors, we report the analysis for ages 75 and 85. Figures 4.11 and 4.12 show the corresponding empirical distribution functions. They look rather differently from the base scenario (Figure 4.8). More specifically, now the unsystematic mortality risk R_3 is the dominant risk

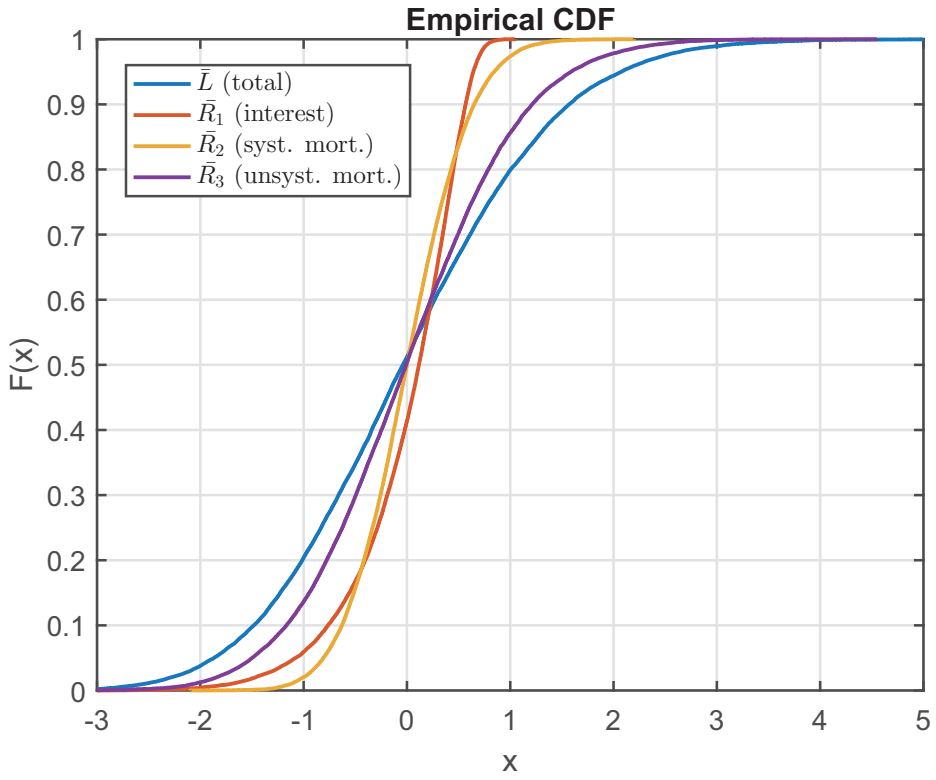

Figure 4.12: Empirical distribution functions at age 85.

factor since its distribution function is the closest to the distribution function of total risk L. Figure 4.12 further confirms our observation in that the distribution of R_3 almost coincide with the distribution of L. This seems intuitive since the unsystematic mortality risk factor will play a more important role in the portfolio as the cohort age increases. Since the limiting age ω is fixed to 115, increasing the cohort age is equivalent to reducing the active period of the annuity portfolio. However, these three empirical distribution function graphs still share some common features. If we look closely, we would find that R_2 and R_3 are still approximately symmetric and the interest risk factor R_1 is still skewed.

Tables 4.3 and 4.4 provide numerical allocated risk contributions for age 75 and 85, respectively. The results confirm our observations from a quantitative point of view. Firstly, the unsystematic risk R_3 is undoubtedly the most relevant risk factor except for the standard deviation risk measure. It makes up between about 35% and 52% of the total risk capital in Table 4.3. When the age increases to 85, R_3 makes

	$\bar{L}$	$\bar{R}_1$	$\bar{R}_2$	$\bar{R}_3$
Std	2.0640	1.1223 (54.4%)	0.2111 (10.2%)	0.7300 (35.4%)
$\text{VaR}_{0.99}$	4.8730	1.7041 (35.3%)	0.6214 (12.9%)	2.5001 (51.8%)
$\text{TVaR}_{0.99}$	5.5564	1.9037 (34.3%)	0.8277 (14.9%)	2.8203 (50.8%)

Table 4.3: The total risk capital and the Euler risk contributions for a whole life annuity portfolio at age 75.

	$\bar{L}$	$\bar{R}_1$	$\bar{R}_2$	$\bar{R}_3$
Std	1.1993	0.2520 (21.0%)	0.2103 (17.5%)	0.7367 (61.4%)
$\text{VaR}_{0.99}$	3.0287	0.2639 (9.0%)	0.5443 (18.6%)	2.1231 (72.4%)
$\text{TVaR}_{0.99}$	3.5363	0.5012 (14.2%)	0.7008 (19.8%)	2.3331 (66.0%)

Table 4.4: The total risk capital and the Euler risk contributions for a whole life annuity portfolio at age 85.

even larger percentage of the total risk capital accounting for more than 60% of the total risk capital. Secondly, from these two tables we can identify that the interest rate risk is the second-most significant factor in the case of $x = 75$ while systematic mortality risk R_2 is placed as the second in the case of $x = 85$. Figure 4.12 also supports this observation in that the distribution function of R_2 spreads more than R_1. Moreover, the total risk capital of L is strictly decreasing due to the increase of the cohort age.

4.2.3 Deferred Annuities

Now let us bring the deferred annuities to our attention. In order to compare with the whole life annuity issued at age 65, we consider 10-year and 20-year deferred annuity issued at age 55 and 45, respectively. Figures 4.13 and 4.14 may seem similar at first glance, but if we check the upper-right corner of both graphs we can find that interest rate risk factor R_1 and total risk L are closer to each other in Figure 4.14.

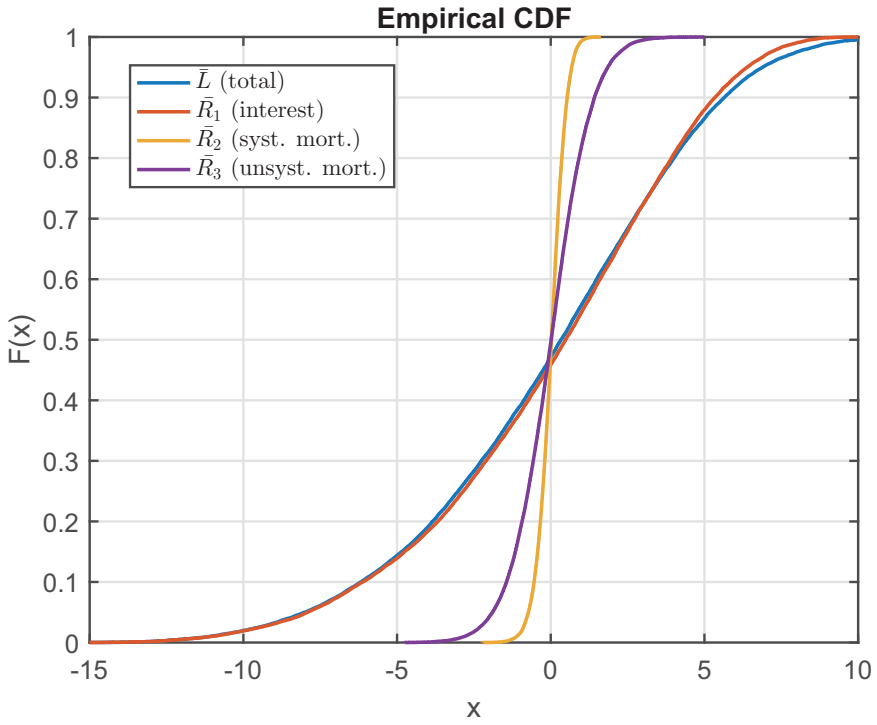

Figure 4.13: Empirical distribution functions for 10-year deferred annuity issued at age 55.

The systematic mortality risk R_2 and unsystematic mortality risk R_3 are also closer in Figure 4.14. The interest rate risk R_1 is the dominant risk in the two deferred annuity cases, however, it makes up larger percentage of the risk capital allocation in the 20-year deferred annuity.

Tables 4.5 and 4.6 provide risk capital allocation results for the two deferred annuity examples. As we expected, the interest rate risk R_1 is the most significant factor accounting for between 77% and 93% in 10-year deferred case. Moreover, it also makes a larger percent (between 86% and 96%) of total risk capital in the 20-year deferred annuity portfolio. The difference between systematic mortality risk R_2 and unsystematic mortality risk R_3 is obviously smaller in Table 4.6 as we inferred from the empirical distribution function graphs. It is not very surprising to see that the total risk capital L is increasing with respect to the deferral period in Tables 4.1, 4.5, and 4.6 because the active period of the portfolio is increasing. That is to say we would expect an annuity portfolio inherit more risk if it is exposed to the effects of

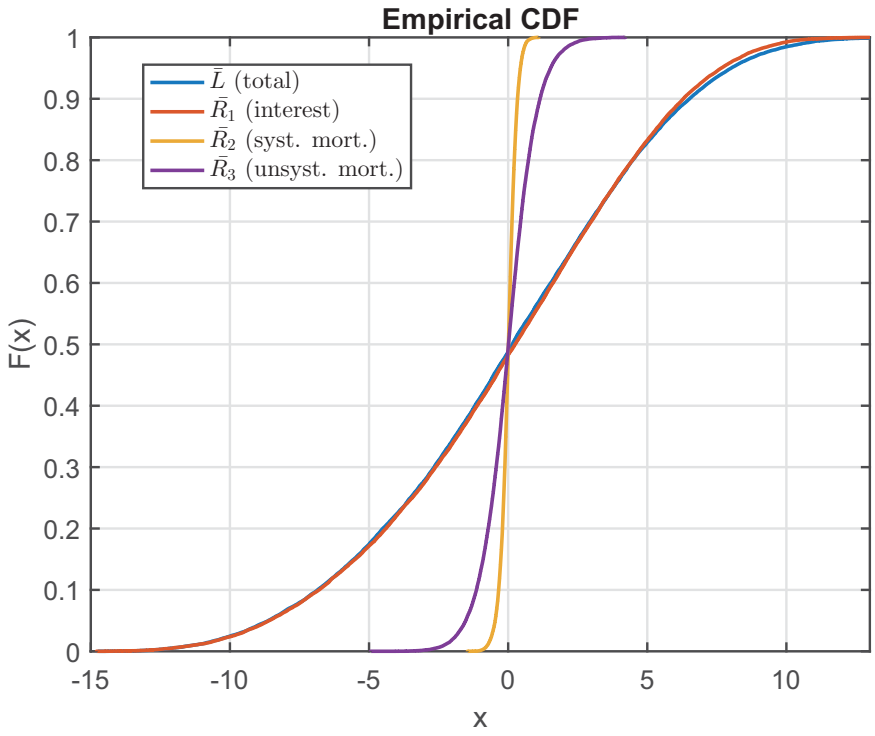

Figure 4.14: Empirical distribution functions for 20-year deferred annuity issued at age 45.

risk drivers longer. However, the last two columns of Tables 4.5 and 4.6 show that the systematic risk factor R_2 and the unsystematic risk factor R_3 are decreasing in all three risk measures as the length of the deferred phase of the annuity portfolio to increasing. The reasons for the smaller mortality risks are twofold; the first reason is that some of the policyholders in the portfolio will die during the deferred term. As the deferred term gets longer, the number of deaths during this period becomes larger. Therefore, the total number of policyholders exposed to the unsystematic mortality risk is smaller. Another possible reason is that mortality rates evolve according to the mortality model during the deferred term. The values of death rates q_{65}, $q_{66}, \ldots$ would become smaller as the deferred term become longer. As a result, the unsystematic mortality risk introduced by counting process $N(t)$ would become smaller.

Figure 4.15 illustrates the empirical distribution functions of the total risk L for three annuity portfolios in one graph. We see that the ranges of possible outcomes of L are increasing when the cohort age is decreasing from 65 to 45. At age 65, the

	$\bar{L}$	$\bar{R}_1$	$\bar{R}_2$	$\bar{R}_3$
Std	4.5271	4.2011 (92.8%)	0.0411 (0.9%)	0.2842 (6.3%)
$\text{VaR}_{0.99}$	9.1209	7.2142 (79.2%)	0.2651 (2.9%)	1.6301 (17.9%)
$\text{TVaR}_{0.99}$	10.0691	7.7558 (77.0%)	0.2954 (2.9%)	2.0202 (20.1%)

Table 4.5: The total risk capital and the Euler risk contributions for 10-year deferred annuity portfolio issued at age 55.

	$\bar{L}$	$\bar{R}_1$	$\bar{R}_2$	$\bar{R}_3$
Std	5.0394	4.8387 (96.0%)	0.0178 (0.4%)	0.1822 (3.6%)
$\text{VaR}_{0.99}$	10.5653	9.3471 (88.6%)	0.0236 (0.2%)	1.1766 (11.2%)
$\text{TVaR}_{0.99}$	11.6129	9.9942 (86.1%)	0.1555 (1.3%)	1.4577 (12.3%)

Table 4.6: The total risk capital and the Euler risk contributions for 20-year deferred annuity portfolio issued at age 45.

loss L_{65} exhibits the most concentrated distribution, while at age 45 it becomes the most scattered one. It confirms our observation since the distribution function of L_{45} is more risky than its counterparts. Similarly, Figures 4.16 shares the same pattern. The distribution function with the longest deferred term (20 years) is always the most scattered one. However, this order changes in the Figure 4.17 and 4.18. The plots of mortality distribution functions at age 45 is the most concentrated one. Again, it confirms our observation in Tables 4.5 and 4.6.

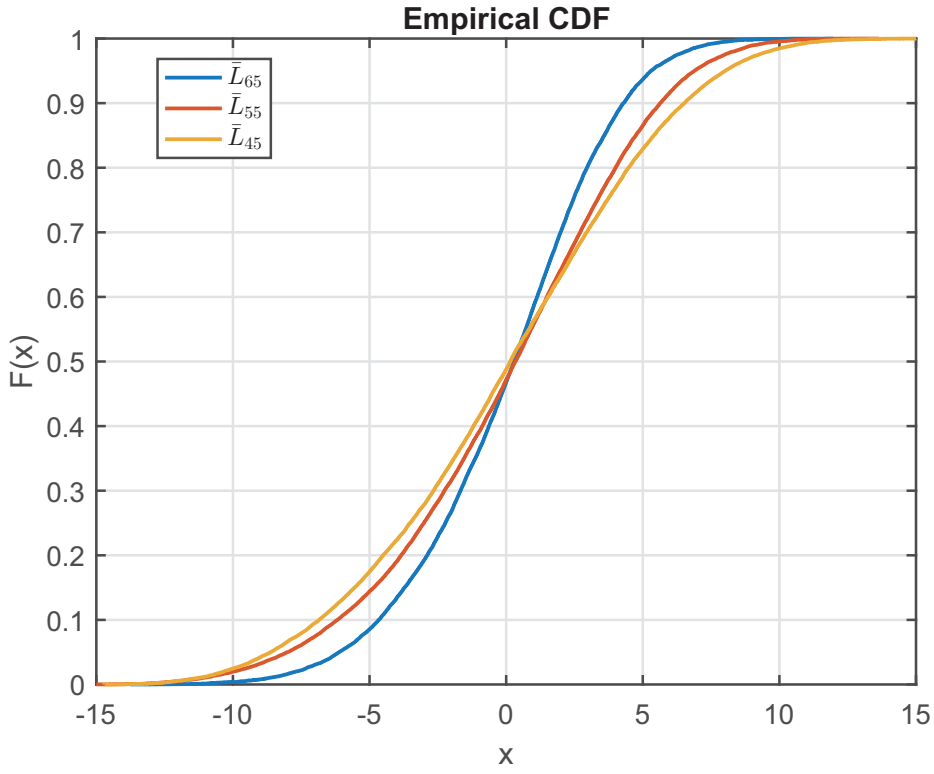

Figure 4.15: Empirical distribution functions for total risk L

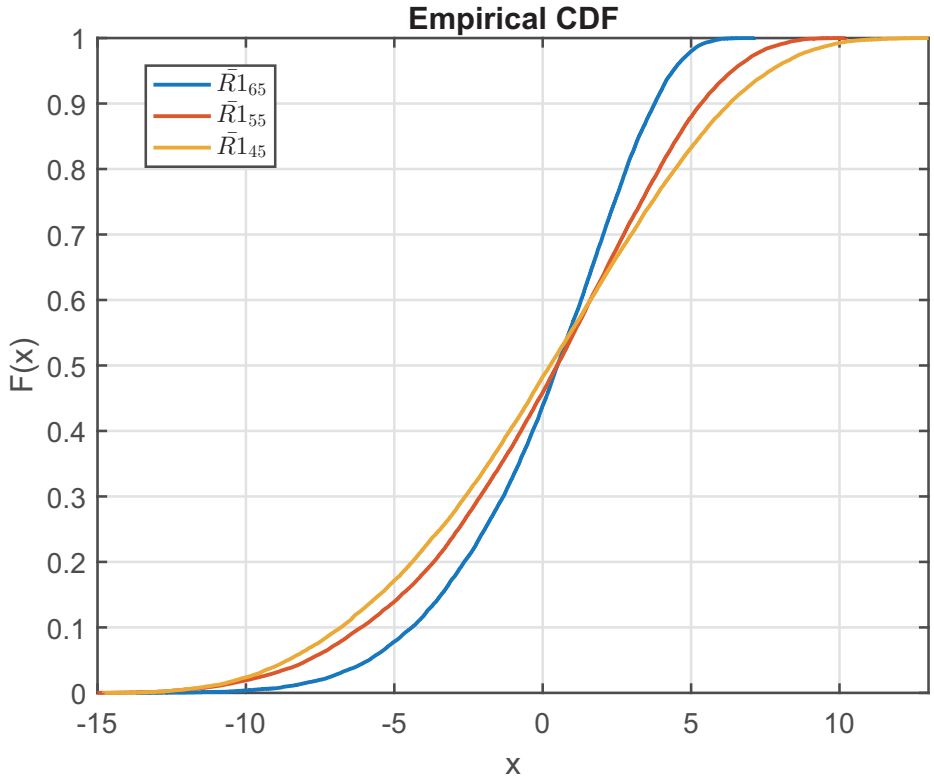

Figure 4.16: Empirical distribution functions for interest rate risk R_1

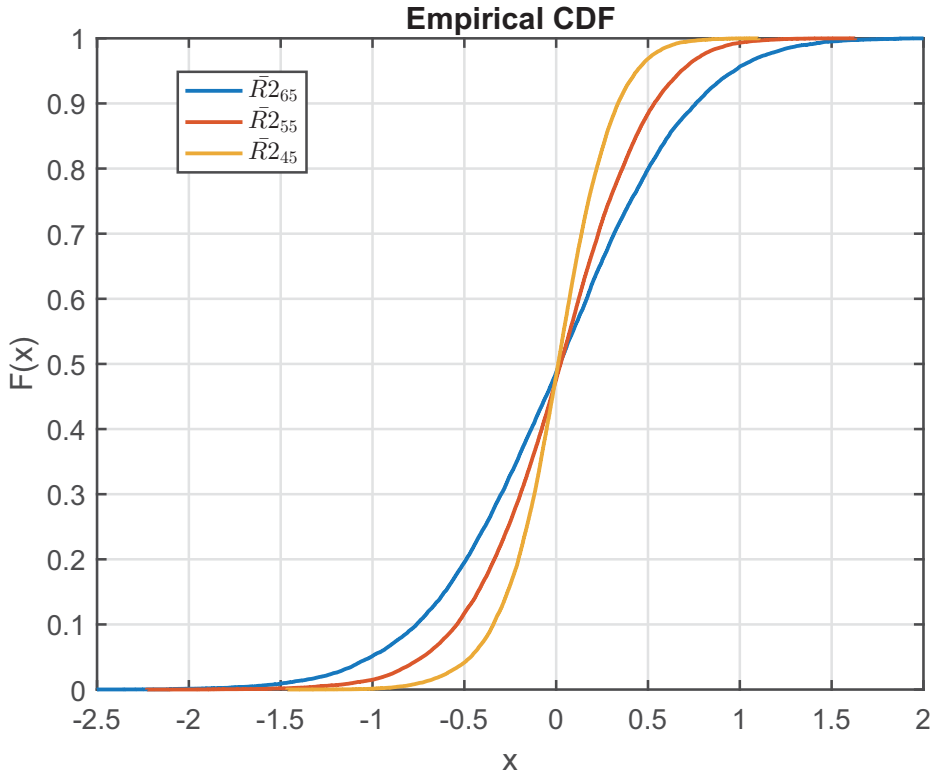

Figure 4.17: Empirical distribution functions for systematic mortality risk R_2

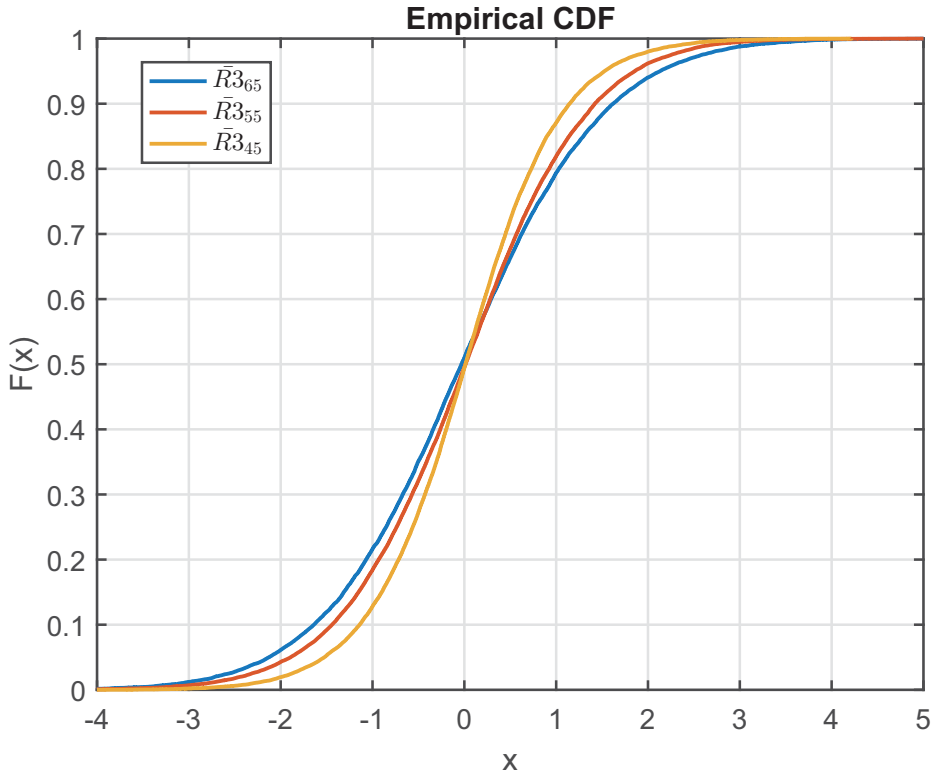

Figure 4.18: Empirical distribution functions for unsystematic mortality risk R_3

Conclusion

This thesis applies the MRT decomposition method to whole life annuities and deferred annuities. The risk factors obtained from the decomposition summarize risk in insurance liabilities associated with the different sources of uncertainty. Therefore, risk factors can be used to allocate risk capitals by using an appropriate allocation principle.

In Chapter 1 we introduce three candidates of stochastic mortality models that can serve as a mortality model used in the MRT decomposition. Renshaw–Haberman model is selected as the most suitable model. The generalization of RH model into continuous case is also discussed. Chapter 2 is devoted to stochastic interest models. We introduce and calibrate the Vasicek model and the CIR model, then choose the CIR model as our stochastic interest model used in the MRT decomposition.

In Chapter 3 we give a detailed discussion on the MRT decomposition (Schilling et al. (2015)) and obtain the explicit formula for our annuity portfolio. In Chapter 4 we provide an numerical example using the MRT decomposition to demonstrate its applicability and usefulness. It shows that for the whole life annuity, the interest rate risk R_1 is the most relevant risk factor for young cohorts. However, when the age of cohort is increasing, the unsystematic mortality risk R_3 becomes the most relevant risk. We also observe that the length of deferred term has a significant influence on the unsystematic mortality risk R_3 for deferred annuities.

Bibliography

T. R. Bielecki and M. Rutkowski. *Credit risk: modeling, valuation and hedging*. Springer Science & Business Media, 2013.

K. Black and H. D. Skipper. *Life and health insurance*. Prentice Hall, 2000.

H. Booth, J. Maindonald, and L. Smith. Applying lee-carter under conditions of variable mortality decline. *Population studies*, 56(3):325–336, 2002.

D. Brigo and F. Mercurio. *Interest rate models-theory and practice: with smile, inflation and credit*. Springer Science & Business Media, 2007.

N. Brouhns, M. Denuit, and J. K. Vermunt. A poisson log-bilinear regression approach to the construction of projected lifetables. *Insurance: Mathematics and economics*, 31(3):373–393, 2002.

H. Bühlmann. Life insurance with stochastic interest rates, in ottaviani g.(ed.)-financial risk in insurance, 1995.

A. J. Cairns, D. Blake, and K. Dowd. A two-factor model for stochastic mortality with parameter uncertainty: theory and calibration. *Journal of Risk and Insurance*, 73(4):687–718, 2006.

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

A. J. Cairns, D. Blake, K. Dowd, G. D. Coughlan, D. Epstein, and M. Khalaf-Allah. Mortality density forecasts: An analysis of six stochastic mortality models. *Insurance: Mathematics and Economics*, 48(3):355–367, 2011.

CMI. Stochastic projection methodologies: Lee–carter model features, example results and implications. *Working paper 25*, 2007.

J. C. Cox, J. E. Ingersoll Jr, and S. A. Ross. A theory of the term structure of interest rates. *Econometrica: Journal of the Econometric Society*, pages 385–407, 1985.

M. Dahl and T. Møller. Valuation and hedging of life insurance liabilities with systematic mortality risk. *Insurance: mathematics and economics*, 39(2):193–217, 2006.

S. H. Friedberg, A. J. Insel, and L. E. Spence. *Linear Algebra*. Pearson, 2003.

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

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

A. E. Renshaw and S. Haberman. A cohort-based extension to the lee–carter model for mortality reduction factors. *Insurance: Mathematics and Economics*, 38(3): 556–570, 2006.

K. Schilling, D. Bauer, M. C. Christiansen, and A. Kling. Decomposing life insurance liabilities into risk factors. 2015.

S. E. Shreve. *Stochastic calculus for finance II: Continuous-time models*, volume 11. Springer Science & Business Media, 2004.

D. Tasche. Capital allocation to business units and sub-portfolios: the euler principle. *arXiv preprint arXiv:0708.2542*, 2007.

O. Vasicek. An equilibrium characterization of the term structure. *Journal of financial economics*, 5(2):177–188, 1977.