Stochastic Mortality Modelling with Lévy Processes based on GLM's and Applications

Seyed Saeed Ahmadi

A Thesis
In the Department
of
Mathematics and Statistics

Presented in Partial Fulfillment of the Requirements
For the Degree of
Doctor of Philosophy (Mathematics) at
Concordia University
Montreal, Quebec, Canada

August 2013

CONCORDIA UNIVERSITY

SCHOOL OF GRADUATE STUDIES

This is to certify that the thesis prepared

By: **Seyed Saeed Ahmadi**

Entitled: **Stochastic Mortality Modelling with Lévy Processes based on GLM's and Applications**

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

DOCTOR OF PHILOSOPHY (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:

_____ Chair
Dr. S. Betton

_____ External Examiner
Dr. S.H. Cox

_____ External to Program
Dr. J. Etezadi

_____ Examiner
Dr. Y.P. Chaubey

_____ Examiner
Dr. L.G. Doray

_____ Thesis Co-Supervisor
Dr. P. Gaillardetz

_____ Thesis Co-Supervisor
Dr. J. Garrido

Approved by _____
Dr. J. Garrido, Graduate Program Director

August 5, 2013

_____
Prof. J. Locke, Interim Dean
Faculty of Arts and Science

Abstract

Stochastic Mortality Modelling with Lévy Processes based on GLM's and Applications

Seyed Saeed Ahmadi, Ph.D.
Concordia University, 2013

Mortality rates have shown a gradual and steady decline over the last decades. In this thesis, we propose a stochastic process for the force of mortality. Similarly to Renshaw et al. (1996), the force of mortality will be defined using an exponential function of Legendre polynomials. In order to model perturbations in the force of mortality, we use the approach of Ballotta and Haberman (2006) and add a stochastic process which follows a one-dimensional Ornstein-Uhlenbeck process.

We show how Generalized Linear Models can be used to estimate coefficients of the explanatory variables as well as the value of the coefficient of the Ornstein-Uhlenbeck process. For this purpose, the estimator of this coefficient is obtained by minimizing the residual deviance.

Next we change the structure of the perturbed term in the Ornstein-Uhlenbeck process by replacing Brownian motion with Lévy processes. We give some examples to clarify the fitting process and show the advantages of using stochastic forces of mortality. Predictions of the probabilities of death will be investigated to show how the model can be used in actuarial applications. Life annuities are then priced and compared using the proposed model based on Lévy processes and the model in Renshaw et al. (1996).

In this thesis, we also reconsider the two-factor stochastic mortality model introduced by Cairns, Blake and Dowd (2006). We first show that the underlying normality assumption of the error terms does not hold for the considered data set. We suggest to model the error terms using bivariate Generalized Hyperbolic distribution that includes four non-Gaussian, fat-tailed distributions. Our empirical analysis shows how the model can provide a better fit for the considered data.

In addition, we try to model age adjusted death rates embedded in the Swiss Re mortality bond using generalized least squares approach. We use the variable length Markov chains (VLMC) model proposed by M*ä*chler and B*ü*hlmann (2004) to model the incidence of catastrophic events. The proposed model is compared to the current recognized models in the literature. Finally, we perform a simulation study to estimate the market price of risk that can be used to fairly price the Swiss Re bond.

Acknowledgements

First and foremost, I would like to express my deepest appreciation to my supervisor Dr. Patrice Gaillardetz for his guidance and continuous support during my Ph.D. program at Concordia University. This thesis would not have been possible without his encouragement and patience.

Besides my supervisor, I will always be grateful to Prof. Jose Garrido for his invaluable suggestions and reviews both this thesis and my previous research papers.

In addition, I wish to record my profound respect and grateful thanks to the examining committee especially to the external examiner, Prof. Samuel H. Cox, for his insightful comments which have undoubtedly improved the first version of my doctorate dissertation.

Finally, my deepest gratitude is sent to my beloved mother, father and my dear siblings for their warm support.

I dedicate this thesis to my younger sister, Jeenossadat.

Contents

List of Figures

List of Tables

Introduction

Trends in Longevity

Life expectancy is the expected number of years of life remaining at a given age. As mentioned in Rorabaugh et al. (2004), during the early 1600s in England, life expectancy was only about thirty-five years, and two-thirds of all children died before the age of four. In the seventeenth century, bacterial stomach infections, intestinal worms, epidemic diseases, contaminated food and water, neglect and carelessness all contributed to explain why 40 percent of children failed to reach adulthood in New England. In 1900, the world life expectancy for a newborn was approximately 30 years and in 1985 it was about 62 years. During the industrial revolution, the life expectancy of children increased dramatically. The percentage of children born in London who died before the age of five decreased from 74.5% in 1730-1749 to 31.8% in 1810-1829; see Kumar (2011) and Buer (1926). During the 20th century, the average lifespan in the United States increased by more than 30 years. In other words, life expectancy has increased from 1950 till 2010. It is now clear that mortality rates have a decreasing trend over time. However the evolution of these rates differs for different ages and calendar years.

This behavior has an impact on two major areas: social policies and insurance companies. State welfare, social security and pensions are some examples of social policies. Social security systems have begun to experience some problems because of the evolution in life expectancy. Also, earlier defined benefit pension systems are experiencing sustainability problems due to the increased longevity. For instance, in pension plans, when life expectancy of retirees increases, the fund has to pay pensions for longer periods. This demographic process threatens the actuarial balance of the fund and changes are needed accordingly. Therefore, fluctuations in longevity have a considerable influence on the management of a large portfolios of elderly people.

If insurance companies that provide annuity products rely on current standard life tables and use them to obtain premium rates and reserves, then they may undercharge their policyholders. This is due to the fact that these life tables do not capture the mortality improvement over time and underestimate the liabilities

of the insurance company. Hence they will encounter some losses. In the United Kingdom, at the end of 1990's, an unanticipated decline in mortality rates as well as falling market interest rates has led to the bankruptcy of Equitable Life (the world's oldest life insurance company) in 2000 due to their insolvency problems. Among others, Ballotta and Haberman (2003), Pitacco (2003), Boyle and Hardy (2003) and Olivieri (2001) studied the effects of changing forces of mortality[1] on insurance products such as annuities and life insurances.

There is a vast literature on models for the force of mortality. See for example: De Moivre (1725), Gompertz (1825), Makeham (1860), Perks (1932), Weibull (1939), Heligman and Pollard (1980) and Anderson and Ashwood (1985). The most important concept missing in almost all these models is the time effect. In order to make the transition from one mortality model to another and get a good approximation of future life expectancy, we need to include a time effect.

Lee-Carter Model

Since 1980, many authors have tried to estimate dynamic (in time) mortality rates. Among previous works, Lee and Carter (1992) (LC) is the most prominent. They try to model central mortality rates as a linear combination of age effects and random periodic effects. Singular value decomposition and time series techniques are used in order to model the central death rates. In their model, they investigate the effect of time in the force of mortality. The LC model reduces the dimensionality of the problem using a single time index, which affects the force of mortality at each time for all ages simultaneously. As pointed out by Cairns et al. (2008, 2009, 2011), the LC age-period model has several disadvantages:

- It is a one-factor model leading to a correlation of mortality improvements at all ages (trivial correlation structure).
- The model gives a poor fit to historical data if a cohort effect is observed in the past.
- The uncertainty in future death rates is proportional to the average improvement rate. This can cause the uncertainty being too low for higher ages, since historical improvement rates have often been lower at higher ages.
- If the basic version of the model is used, then a lack of smoothness can occur in the estimated age effect.

Many authors have tried to fix the problems of the LC model by some additions or modifications, e.g. Lee and Miller (2001), Brouhns et al. (2002), Renshaw and

[1]Force of mortality is known under different names in reliability theory and survival analysis including *failure rate* or *hazard rate* as mentioned in Bowers et al. (1997, Chapter 3).

Haberman (2003a), Booth et al. (2002), Jong and Tickle (2006), Delwarde et al. (2007). Renshaw and Haberman (2006) extended the LC modeling framework through the introduction of a wider class of generalized, parametric, non-linear models. They add a cohort effect for the first time to the LC model. This leads to a significant fit improvement when a cohort effect is observed in the past data. However, as mentioned by Cairns et al. (2009, 2011) and Plat (2009), the model proposed by Renshaw and Haberman (2006) has some robustness problems in parameter estimates relative to the range of data employed. Currie (2006) considered a special case of the model in Renshaw and Haberman (2006).

Cairns et al. (2009) suppose eight different stochastic mortality models. The Bayes Information Criterion (BIC) is used to find the models that best fit the data. Haberman and Renshaw (2009) generalize LC's model to include an age modulated cohort index, in addition to the standard age modulated period index. They compare annuity value predictions as well as the life expectancy for different pensioner ages and for various periods. For a comprehensive review of mortality estimation see Booth and Tickle (2008). In addition to the problems mentioned above, the estimated parameters in age-periodic-cohort models are highly correlated, which leads to large variations in the forecasts.

Generalized Linear Modelling

Generalized Linear Models (GLM's) can be used also to forecast mortality rates. The most important difference between GLM and LC models is that the time effect in a GLM framework will appear as an explicit covariate random variable, whereas in LC models, time is a factor represented by a series. GLM's have been used in actuarial applications since the early 1980s, see McCullagh and Nelder (1989). They fit GLM models to different data sets, including motor insurance and marine insurance.

Prior to 1991, not much attention was paid to the link between GLM's and previous models used for graduation purposes. Renshaw (1991) stated that many of the models used by actuaries were special cases of GLMs. For example the Gompertz' model can be obtained by a Poisson GLM. Regarding these models in the context of GLM helps generalizations, such as extensions by including a time effect for predictions. Haberman and Renshaw (1996) reviewed some applications of GLM in actuarial problems. Renshaw, Haberman and Hatzoupoulos (1996) develop a model (RHH) which incorporates both an age variation and time trend in the force of mortality. They use a Poisson GLM to model the force of mortality as an exponential function. They show that for the considered data, this model provides a good fit. Sithole et al. (2000) predict the UK annuitant and pensioner

mortality with qualified success. They use a similar modelling structure as the RHH model. Renshaw and Haberman (2000) model the mortality improvement factor using GLM's. They also introduce some conditions under which the form of their model is the same as LC's model. A parallel methodology to the LC model based on GLM's is proposed by Renshaw and Haberman (2003b). The LC time series approach is adopted and modified in Renshaw and Haberman (2003c) to forecast the behavior of mortality reduction factors by an alternative regression type model. Renshaw and Haberman (2003a) generalize mortality forecasts on the basis of the first two sets of single value decomposition vectors rather than just on the first such set of vectors. Also, obtained forecasts have been compared with two similar approaches based on generalized linear and bilinear models with Poisson error structures.

Lévy Stochastic Mortality Modelling

Chen and Cox (2009) include a jump process into the original LC model and use it to forecast mortality rates. Biffis (2005) proposes affine jump diffusion processes to model asset prices and mortality dynamics. However, this model cannot guarantee a nonnegative force of mortality as pointed out by Chen and Cox (2009).

Wang et al. (2010) claim that mortality indices in the LC model have tails thicker than a normal distribution and appear to be skewed. Therefore they suggest two infinitely divisible distributions, Generalized Hyperbolic and Classical Tempered Stable distributions to model the mortality indices in the LC model. They also apply different criteria such as the Akaike information criterion, Bayes information criterion, Kolmogorov-Smirnov tests and mean absolute percentage errors, in their projections to show the advantages of two proposed models over the LC model.

Hainaut and Devolder (2008) consider the force of mortality as a combination of one deterministic component and a stochastic part. This dynamic part is the solution to a stochastic differential equation with respect to some Lévy processes. They investigate intensities that have a mean reverting stochastic property. Moreover, some special cases of Lévy processes have been considered to obtain closed expressions for the survival probabilities. In addition, they show that their proposed models were able to capture some of the observed mortality tendencies such as rectangularization. It means that the survival function has an increasing concentration of death around the mode (at old ages). This property entails that the shape of the survival function evolves toward a rectangle. However, they did not estimate parameters of the dynamic part appearing in the stochastic differential

equation. Obviously not enough applied mortality models have yet used Lévy processes and studied their adequacy.

Other Models

Currie et al. (2004) generalized P-spline (penalized b-splines) techniques to forecast two-dimensional mortality tables. They use a penalized generalized linear model (PGLM) with Poisson errors and show how to construct regression and penalty matrices appropriate for two-dimensional modelling. Cairns et al. (2006) propose a stochastic mortality model with two factors. Logistic regression is a special case of GLM and can be used to model mortality rates as well. Butler and Park (1987) use a logistic regression model for the analysis of proportionate mortality data. They suppose that the occupational exposure, if it has an effect on mortality, develops the rate of death by a multiplicative factor. Perks (1932) and Kannisto (1992) also suggest logistic models for the force of mortality which provide a better fit to mortality data of people aged over 85 than Makeham's model. However it is difficult to estimate the parameters of these two models by using maximum likelihood estimators. More recently Doray (2008) proposed a weighted least-squares estimator for Kannisto's model. Also the estimator has been shown to be consistent, asymptotically unbiased and normally distributed. Moreover, for Perks's model, Doray (2008) uses Taylor series expansion to reduce the estimation problem to a least-squares problem.

Much work has been done in estimating the parameters in the force of mortality, either with LC's model or within the GLM's framework, in contrast to the estimation of the parameters for the dynamic part of the model. Renshaw et al. (1996) develop a model which includes both age and time trends in the force of mortality. They show its advantages over the ad hoc method of first fitting a model with respect to age, and then trying to find the time trend in each model separately. Sithole et al. (2000) also defined similar models as Renshaw et al. (1996) concerning age and time. They show that if a trend adjustment term is linear in time on the log scale and its coefficient is linear in age effects, then the models proposed by Renshaw et al. (1996) could be used for prediction purposes. They used a quasi likelihood approach for their estimation. However, their models do not include a stochastic term that develops according to a stochastic differential equation (SDE).

In the following two chapters, we follow Ballotta and Haberman (2006) and add a perturbation term to the force of mortality. We believe that by including this perturbation term, the model can better capture the time trend. The perturbation term follows a one-dimensional Ornstein-Uhlenbeck (OU) process. The SDE itself

depends on 3 different parameters. As an example, we first simplify this SDE as those of Ballotta and Haberman (2006), by reducing the proposed SDE to a 1-parameter case (OU parameter).

Poisson GLM's have been considered to link this unknown parameter with other explanatory random variables. Here we estimate all the parameters appearing in the GLM as follows. We first fix the OU parameter and fit different GLM models by assuming a Poisson distribution. We record the corresponding residual deviances for each model. Residual deviances are minimized to obtain the best possible value for the OU parameter.

Our main goal in this chapter is to propose an estimation method for the OU parameter and all the coefficients in the GLM. In addition to the Poisson GLM, we will also use a perturbed logit regression model for the force of mortality. Projections will be carried out for both models and will be compared with real data.

As mentioned in Pitacco et al. (2009), the choice of calendar years and ages in the mortality fitting is a crucial step. This means that in future analysis, we may reduce dimensionality of the data used to get a better model fit.

The outline of Chapter 1 is as follows: in Section 1.2, we will specify the proposed models and the data used. The error term will be added to the model that follows an Ornstein-Uhlenbeck (OU) process based on Brownian motion (BM). Section 1.3 is for model fitting and testing model adequacy. Here we will explain how to choose the best model as well as how to project the death rates and compare them to the real data available in the life table. We also compare two proposed model in terms of their capabilities to do the projections. Section 1.4 gives the conclusions for Chapter 1.

In Chapter 2, we will replace the BM by a Lévy process. Specifically, in Section 2.1 we provide the definition of the Lévy process and explain why we use this instead of the BM. We summarize some additional definitions and results in Appendix A. In Section 2.2, we state our proposed model. Some numerical examples have been included in Section 2.3.3 to show how the Lévy process can lead to a significantly improved mortality fit in comparison to the results obtained in Chapter 1. The conclusions of Chapter 2 are given in Section 2.4.

In Chapter 3, we evaluate life annuities using the model developed in Chapter 2. The data set is introduced in Section 3.2. We fit our models in Section 3.3. Model comparisons are given in Section 3.4 and life annuities are evaluated in Section 3.5. The conclusions for this chapter are provided in Section 3.6.

In Chapter 4, we focus on two-factor stochastic mortality modeling developed by Cairns et al. (2006). The outline of this chapter is as follows. In Section 4.2, a brief review of the CBD model is given together with the data set that we used to fit

our model. Then, we test the assumptions given in the CBD model by applying Doornik-Hansen's multivariate normality test in Section 4.3. Next, generalized hyperbolic distributions are defined in Section 4.4. We propose and fit our model in Section 4.5. Section 4.6 is devoted to model comparisons. Conclusions are given in Section 4.7.

Chapter 5 is devoted to model age adjusted death rates embedded in the Swiss Re mortality bond. We use two approaches namely Generalized Least Squares and Variable Length Markov Chains to construct our model. The data is described in Section 5.2 and variable length Markov chains in Section 5.3. A brief review of the generalized least square approach is given in Section 5.4. The proposed model is explained in Section 5.5. We fit our model and present the results in Section 5.6. Model comparisons are carried out in Section 5.7. The design of the Swiss Re bond and estimation of the market price is given in Section 5.8. Section 5.9 gives conclusion of this study.

Chapter 1

Ornstein-Uhlenbeck Process

1.1 Introduction

The idea of having a perturbed term in the force of mortality is first introduced by Ballotta and Haberman (2006). They assume that this perturbation term follows a one-dimensional Ornstein-Uhlenbeck process. In this chapter, we follow the same idea and add a random variable to the force of mortality. Moreover, we provide a new approach to estimate the parameters of the model. The structure of the force of mortality given in this chapter is similar but not identical to that of Ballotta and Haberman (2006). However, the same argument as given in this chapter can be applied to their model.

In Section 1.2, we propose a Poisson GLM to model the force of mortality. Legendre polynomials, an alternative Logistic Regression Model and the required data are also introduced. Section 1.3 gives the details of the model fitting. Here we explain how to construct the linear form in the model that enables us to apply GLM's for the estimation and forecasting. Some examples are provided to show applications of the results. Section 1.4 gives conclusions of this chapter and discusses briefly some possible future research.

1.2 Model Specification

1.2.1 Poisson GLM

Renshaw et al. (1996) use a Poisson GLM to model the force of mortality with a log link[1] function, that is:

$$\ln \mu(x,t) = \beta_0 + \sum_{j=1}^{s} \beta_j L_j(x^{'}) + \sum_{i=1}^{r} \alpha_i t^{'^i} + \sum_{i=1}^{r}\sum_{j=1}^{s} \gamma_{ij} L_j(x^{'}) t^{'^i}, \quad x \in [x_1, x_2], \quad t \in [y_1, y_2], \tag{1.1}$$

where $\mu(x,t)$ is the force of mortality at age x in year t, $[x_1, x_2]$ and $[y_1, y_2]$ define the range of ages and calendar years, respectively. The unknown parameters α_i, β_j, γ_{ij} need to be estimated, while the $L_i(x^{'})$ are Legendre polynomials generated by

$$L_0(x) = 1, \quad L_1(x) = x, \quad (n+1)L_{n+1}(x) = (2n+1)xL_n(x) - nL_{n-1}(x), \quad n \geq 1.$$

Hence, $L_2(x) = \frac{3x^2-1}{2}$ and $L_3(x) = \frac{5x^3-3x}{2}$. Here $x^{'}$ and $t^{'}$ are the transformed ages and transformed calendar years that map x and t onto the interval [-1,1], as defined by:

$$x^{'} = \frac{2x - max(x) - min(x)}{max(x) - min(x)}, \quad t^{'} = \frac{2t - max(t) - min(t)}{max(t) - min(t)}.$$

Equation (1.1) explains the fluctuations in the force of mortality through three different variables. The first and second summations of (1.1) capture the age effect and the time effect, respectively. The last term of (1.1) has been added to model interactions between the age and time effects.

Renshaw et al. (1996) proposed a methodology to determine the optimum values for the Legendre polynomial orders r and s in (1.1). They monitor the improvement in model residual deviances as the values of r and s are increased. Here we skip this step to select a similar but not identical model as that suggested by Sithole et al. (2000), our model of choice, as we would like to compare our results to those of Ballotta and Haberman (2006) that use the estimated parameters of Sithole et al. (2000).

We add a conditional perturbation term to (1.1) to capture more efficiently the evolution of time in the model. Therefore, our proposed model structure is similar to those of Renshaw et al. (1996) and Sithole et al. (2000), except for this additional random variable, whose distribution depends on time.

[1]The log link function ensures that the force of mortality, $\mu(x,t)$, is always positive.

Let $D(x,t)$ and $\mu(x,t)$ be the number of deaths and the force of mortality occurring at aged x and in calendar year t, respectively. We considered $D(x,t)$ as a Poisson response random variable with $\mathbb{E}[D(x,t)] = r(x,t)\mu(x,t)$, where $\mathbb{E}[.]$ represents the expected value and $r(x,t)$ is a non-random constant equal to the number of people exposed to the risk of death. To show the similarity of our results with those of Ballotta and Haberman (2006), we first let $r(x,t) = 1$. However for projection purposes, we must consider $\ln(r(x,t))$ as an offset term[1] in our model. We assume that the conditional expected value of $D(x,t)$, given a random variable $Y_{t'}$, is

$$\mathbb{E}[D(x,t) \mid Y_{t'}] = r(x,t) \exp\Big(\beta_0 + \sum_{i=1}^{3} \beta_i L_i(x^{'}) + \beta_4 t^{'} + \sum_{i=1}^{2} \beta_{i+4} L_i(x^{'}) t^{'} + \beta_7 Y_{t'}\Big), \quad x \in [x_1, x_2], \quad t \in [y_1, y_2], \tag{1.2}$$

where Y_t follows an Ornstein-Uhlenbeck process defined as the solution of the following SDE:

$$\begin{cases} dY_t = a(b - Y_t)d_t + \gamma dX_t, & t > 0, \\ Y_0 = 0, \end{cases} \tag{1.3}$$

where a, b and γ are unknown constants and $(X_t : t \geq 0)$ is a standard one-dimensional Brownian motion. Equation (1.3) is more general than the SDE proposed by Ballotta and Haberman (2006), which can be obtained by setting $b = 0$ and $\gamma = 1$ in (1.3). This will enable the comparison of our results with those of Ballotta and Haberman (2006).

One of the main reasons for using Legendre polynomials is their computational tractability. Working with real calendar years and ages is time consuming, especially when fitting several models. Using these transformations, enables us to reduce computational time significantly. Also, as explained in De Jong and Heller (2008), if the age variable x is large, then overflow occurs when computing x^m for large values of m. Another consequence is that underflow occurs for the estimation of the corresponding $\hat{\beta}$. This problem can be avoided by scaling ages x and times t appropriately, using $x^{'}$ and $t^{'}$. Collinearity between higher order polynomials can also be avoided by using orthogonal polynomials. It should be mentioned that Legendre polynomials are orthogonal and have been used to overcome possible collinearity.

[1] An offset term is an explanatory random variable added to the GLM with a known coefficient of one.

1.2.2 Logistic Regression Model

Cairns et al. (2006) introduce a two factor stochastic mortality model:

$$\text{logit}(\tilde{q}(t,x)) = \ln\Big(\frac{\tilde{q}(t,x)}{1-\tilde{q}(t,x)}\Big) = A^{(1)}_{t+1} + A^{(2)}_{t+1}(x+t), \quad x \geq 60, \quad t \in [1961, 2002],$$

where $\tilde{q}(t,x)$ is the realized mortality rate at time t for the cohort aged x and $A^{(1)}_t$, $A^{(2)}_t$ are two stochastic processes. They show that the first factor affects the mortality rate dynamics at all ages in the same way, whereas the second factor affects mortality rates much more at higher ages than at lower ages ones.

In this section, we will propose another GLM approach to model the required death probabilities, rather than the number of deaths. Suppose that $D(x,t)$ has a binomial distribution with parameters $n = r(x,t)$ and $q(x,t)$. Here, the probability of success is equal to the corresponding one year probability of death at each age x and calendar year t, denoted by $q(x,t)$. Moreover, let

$$\begin{aligned} \text{logit}(q(x,t)) &= \ln\Big(\frac{q(x,t)}{1-q(x,t)}\Big), \quad x \in [x_1, x_2], \quad t \in [y_1, y_2] \\ &= \beta_0 + \sum_{i=1}^{3} \beta_i L_i(x') + \beta_4 t' + \sum_{i=1}^{2} \beta_{i+4} L_i(x') t' + \beta_7 \frac{\sigma^2(Y_{t'})}{2}, \end{aligned} \tag{1.4}$$

where the last term $\sigma^2(Y_{t'})$ is the variance of $Y_{t'}$ and $L_i(x')$ are Legendre polynomials defined in Subsection 1.2.1. The reason to include $\sigma^2(Y_{t'})$ in (1.4) will be explained later in Section 3.

1.2.3 Data

The data considered are 1x1 (i.e. by age[1] and year) female's periodic life tables, for the United Kingdom during calendar years 1922-2006 and for ages ranging from 0-110 years, obtained from the Human Mortality Database (HMD)[2]. In this data set, the most important information that we use to fit our process are: calendar year, age, $q(x,t)$ and the number of survivors and the number of deaths at exact age x at time t, denoted by $l(x,t)$ and $D(x,t)$ respectively. Details of the available HMD information as well as smoothing methods used to graduate probabilities of death can be found in the explanatory notes[3] or in Wilmoth et al. (2007).

[1] One-year age groups (or "by age") means 0, 1, 2,..., 109, 110+.

[2] The data are available online at www.mortality.org.

[3] Available at: http://www.mortality.org/Public/ExplanatoryNotes.php# CompleteDataSeries

1.3 Model Fitting

The first step is to solve the SDE in (1.3) for Y_t. By multiplying each side of (1.3) by e^{at}, we get,

$$e^{at}dY_t + ae^{at}Y_t d_t = abe^{at}d_t + \gamma e^{at}dX_t.$$

Hence $d(Y_t e^{at}) = bd(e^{at}) + \gamma e^{at}dX_t$ and therefore,

$$Y_t = b(1 - e^{-at}) + \gamma \int_0^t e^{-a(t-s)}dX_s. \tag{1.5}$$

Equation (1.2) is based on conditional expectation and contains an unobservable random variable Y_t. Consequently, we can not use a GLM approach directly here. Therefore, we need to find the distribution of Y_t in order to express $\mathbb{E}[D(x,t)]$ as a linear function of some observable explanatory random variables. We use the following lemma.

Lemma 1.1. *Let $(X_t : t \geq 0)$ be a standard one-dimensional Brownian motion and for every function $f \in L^2(\mathbb{R}^+)$[1], define the Wiener-integral as:*

$$Y_t = \int_0^t f(s)dX_s, \quad t \geq 0.$$

Then Y_t has a normal distribution with mean 0 and variance $\sigma^2(Y_t) = \int_0^t |f(s)|^2 d_s$.

Proof. See Applebaum (2009), Chapter 4 for the proof. □

To find the distribution of Y_t in (1.5), we use Lemma 1.1, where $f(s) = e^{-a(t-s)}$. Hence Y_t has a normal distribution with mean $b(1 - e^{-at})$ and variance $\sigma^2(Y_t) = \gamma^2\big(\frac{1-e^{-2at}}{2a}\big)$. Clearly, the variance of Y_t depends on the OU's parameters: a, γ and t. It means that adding the perturbation term will transfer some properties of the underlying SDE (1.3) into the GLM. Moreover, $\sigma^2(Y_t) = \frac{\gamma^2}{2a}$ for $a > 0$ as $t \to \infty$. This shows the long term behavior of Y_t. Using (1.2) with $\gamma^2 = \frac{\beta_8}{\beta_7^2}$, we get

$$\begin{aligned}\ln\Big(\frac{\mathbb{E}[D(x,t)]}{r(x,t)}\Big) = \beta_0 \;&+\; \sum_{i=1}^{3}\beta_i L_i(x') + \beta_4 t' + \sum_{i=1}^{2}\beta_{i+4}L_i(x')t' \\ &+\; \beta_7 b(1 - e^{-at'}) + \beta_8\big(\frac{1 - e^{-2at'}}{4a}\big).\end{aligned} \tag{1.6}$$

We set $\sigma^2(Y_{t'}) = \frac{1-e^{-2at'}}{2a}$ in (1.4) to be consistent with (1.6). Using (1.6) and $\mathbb{E}[D(x,t)] = r(x,t)\mu(x,t)$, we have

$$\begin{aligned}\mu(x,t) = exp\Big\{\beta_0 \;&+\; \sum_{i=1}^{3}\beta_i L_i(x') + \beta_4 t' + \sum_{i=1}^{2}\beta_{i+4}L_i(x')t' \\ &+\; \beta_7 b(1 - e^{-at'}) + \beta_8\big(\frac{1 - e^{-2at'}}{4a}\big)\Big\}.\end{aligned} \tag{1.7}$$

[1] L^2-space of equivalence classes of mappings to $\mathbb{R}^+$.

Our objective in this chapter is to show how we can use Equation (1.7) to estimate all the GLM's coefficients, β_i, appearing in (1.2). Moreover, we predict and compare the values of $q(x,t)$ as well as $\mu(x,t)$ in Equations (1.4) and (1.7), respectively, with the observed and smoothed $q(x,t)$. We end this section by introducing the residual deviance, $R(a)$. This is equivalent to a residual sum of squares, but in the context of GLM's. $R(a)$ is defined for the Poisson GLM as

$$R(a) = 2\sum_{x}\sum_{t}\Big(d(x,t)\ln\Big(\frac{d(x,t)}{\hat{d}(x,t)}\Big) - \big[d(x,t) - \hat{d}(x,t)\big]\Big), \tag{1.8}$$

where $\hat{d}(x,t)$ is the predicted number of death at age x and time t obtained by

$$\hat{d}(x,t) = r(x,t)\hat{\mu}(x,t). \tag{1.9}$$

In Equation (1.9),

$$\begin{aligned}\hat{\mu}(x,t) = exp\Big\{\hat{\beta}_0 \;&+\; \sum_{i=1}^{3}\hat{\beta}_i L_i(x^{'}) + \hat{\beta}_4 t^{'} + \sum_{i=1}^{2}\hat{\beta}_{i+4}L_i(x^{'})t^{'} \\ &+\; \hat{\beta}_7 b(1-e^{-at^{'}}) + \hat{\beta}_8\Big(\frac{1-e^{-2at^{'}}}{4a}\Big)\Big\},\end{aligned} \tag{1.10}$$

where $\hat{\beta}_i$ is the estimated value of β_i, for $i = 0, ..., 8$. The interested reader is referred to Renshaw (1991) for further study on the residual deviance.

Theorem 1.1. *The residual deviance of model (1.6) is invariable to homogeneous transformations $T^*(b) : (1-e^{-at^{'}}) \rightarrow b(1-e^{-at^{'}})$, where $b \neq 0$ and all other explanatory random variables remain the same as in (1.6).*

Proof: We show that under the new transformed model, $\hat{d}(x,t)$ and $R(a)$ consequently remain unchanged. For notational convenience, we define vectors $X_i = L_i(x^{'})$, for $i = 1, ..., 3$, $X_4 = t^{'}$, $X_{i+4} = L_i(x^{'})t^{'}$, for $i = 1, 2$, $X_7 = 1 - e^{-at^{'}}$, $X_8 = \frac{1-e^{-2at^{'}}}{4a}$ and $X_7^* = bX_7$. Now we rewrite (1.10) in the form of a linear combination:

$$\ln(\hat{\mu}(x,t)) = \hat{\beta}_0 + \sum_{j=1}^{8}\hat{\beta}_i X_j = \hat{\beta}_0 + \sum_{j=1}^{6}\hat{\beta}_i X_j + \hat{\beta}_7\frac{X_7^*}{b} + \hat{\beta}_8 X_8. \tag{1.11}$$

Equation (1.11) shows that under the transformation $T^*(b)$, the new coefficient for X_7^* i.e. $\hat{\beta}_7^*$, can be obtained by $\hat{\beta}_7^* = \frac{\hat{\beta}_7}{b}$. Therefore, $\hat{\mu}(x,t)$ in (1.10) remains the same as in model (1.6) and $\hat{d}(x,t)$ in (1.8) is not changed.

1.3.1 Estimation and Forecasting

In this section we explain how to estimate the unknown parameters. We consider ages from $[x_1, x_2]$ and calendar years from $[y_1, y_2]$ and refer to this subset of data sets as the historical data. From a practical point of view, we not only have to estimate death rates, but more importantly, we should forecast these rates. In general, we perform the following steps to estimate and forecast the mortality rates:

Step 1: Determine N, the number of years that we need to forecast.

Step 2: Extend all the explanatory random variables in (1.4) and (1.6) for N years.

Step 3: To compare our results with those of Ballotta and Haberman (2006), determine b and γ and set an appropriate interval for the parameter $a \in [a_1, a_2]\backslash\{0\}$.

Step 4: Fix the parameter $a \in [a_1, a_2]\backslash\{0\}$ and fit the models (1.6) and (1.4). Next, find the optimum value for the parameter a by minimizing the residual deviance $R(a)$.

Theorem 1.1 states that when this optimum value is obtained, then changing the values of b determined in Step 3, will not change the residual deviance. The response variables used are $d(x,t)$ and $\frac{d(x,t)}{l(x,t)}$ in models (1.6) and (1.4), respectively. Here, we consider $r(x,t) = l(x,t)$ in model (1.6) as an offset term. Moreover, for the logit model (1.4), since we are modeling the ratios, and not the number of successes, we use weights $= l(x,t)$ in our programming. By completing this step, we select two models for (1.6) and (1.4) with their corresponding estimated β_i's.

Step 5: Plug in estimated coefficients, the β_i's, obtained in Step 4, in Equations (1.10) and (1.4), to forecast. We emphasize the advantage of forecasting using (1.4) and (1.6). Because we can directly forecast required probabilities by using the logit model (1.4). Also predictions of the mortality rates can be obtained from model (1.6).

1.3.2 Examples

Table 1.1 shows the set of parameters for the first example. Data is restricted to the ages ranging from 65-100 and also calendar years 1958-1994. The reason for using this subset is that, our data in this case is close in terms of ages and calendar years to those of Sithole et al. (2000). We set $N = 0$, since we are not interested

Table 1.1: Parameters Set for Example 1

$x_1 = 65$	$x_2 = 100$	$y_1 = 1958$	$y_2 = 1994$	$N = 0$
$b = 0$	$\gamma = 1$	$a_1 = -1$	$a_2 = 2$	$r(x,t) = 1$

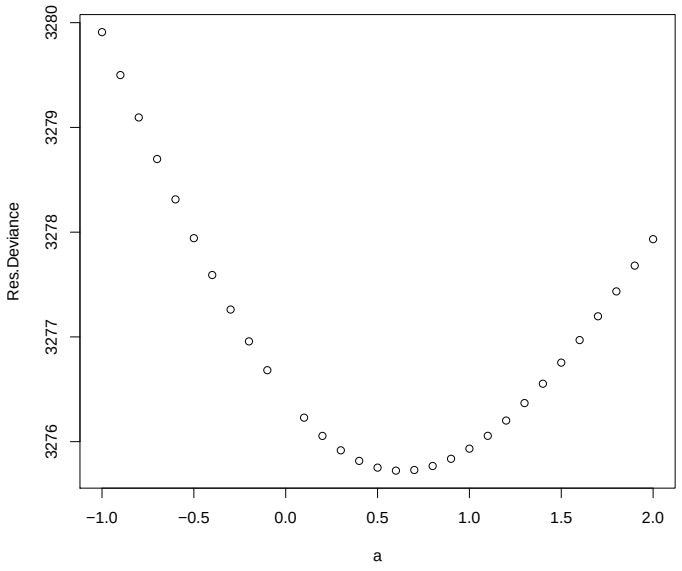

Figure 1.1: Residual deviances in (1.8) v.s. parameter a. Age ranges from 65 - 100, calendar year ranges from 1958 - 1994

in forecasting the mortality rates at this point and modelling is done only over the historical data. We also let $r(x,t) = 1$ to indicate that the number of deaths are modeled for the historical data. Moreover, we consider $b = 0$ and $\gamma = 1$ to reduce our model to the SDE proposed by Ballotta and Haberman (2006).

As we explained above, we first fix the parameter a in the interval, $[-1, 0) \cup (0, 2]$ and fit GLM models with the log as the link function in (1.6) and $r(x,t) = 1$. Then, the β_i's are estimated using the maximum likelihood method. The next step is to obtain the residual deviance. The plot of the residual deviance against parameter a is given in Figure 1.1. We seek the parameter a that minimizes the residual deviance for each fitted model. Then, we choose the model which has the smallest residuals deviance. However, other criteria such as the Akaike Information Criterion (AIC) can be used.

Figure 1.1 shows that the minimum value for the residual deviance in (1.8) occurs at $a = 0.6$. Although the model structure for the force of mortality in (1.6) is different from that of Ballotta and Haberman (2006), the point estimation for parameter a is unexpectedly close to their choice for this particular parameter.

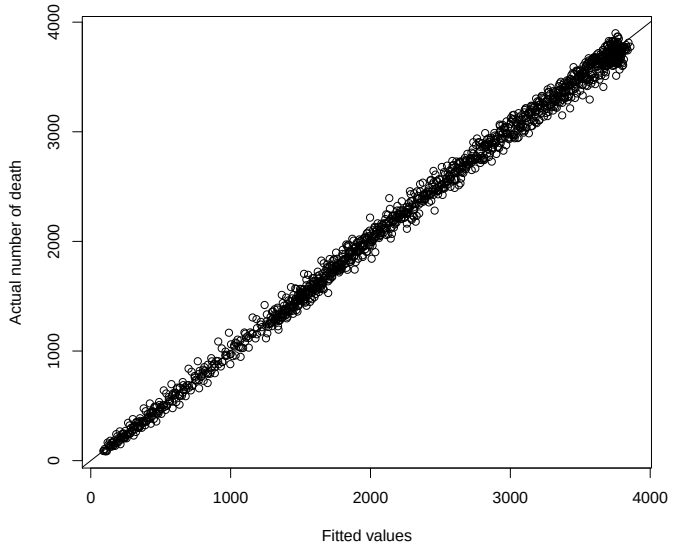

Figure 1.2: The observed death rates, $d(x,t)$ (y axis) v.s. fitted values, $\hat{d}(x,t)$ (x axis), ages 65-100, calendar years 1958-1994

Estimated parameters are given in Table 1.2. Next we check the model adequacy for this particular model.

Figure 1.2 plots the fitted values, $\hat{d}(x,t)$ that are calculated from the model using the estimated parameters against the actual number of deaths, $d(x,t)$, i.e. the historical and observed values for all the ages and calendar years used to fit the model. Also in Figure 1.3, a residual plot is given, for a fixed age $x = 70$. Here, the vertical axis represents the residuals i.e. the difference between the observed number of deaths and the fitted values. This graph does not show any trend.

Table 1.2: Estimated Parameters for $a = 0.60$

$Intercept$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_4$	$\hat{\beta}_5$	$\hat{\beta}_6$	$\hat{\beta}_8$
7.589	-0.565	-1.285	-0.381	0.109	0.473	0.248	0.027

We also obtain the ANOVA table for this model and check the contribution of each coefficient using a chi square test. Table 1.3 summarizes the results. The 3rd column in Table 1.3 (Deviance Residual) shows reduction in the residual deviances as the terms are added sequentially. All the estimated parameters are significant at $\alpha = 0.05$. The last row in this table shows that adding the OU process drops the residual deviance from 3301 to 3276. We should mention that the optimum value found for parameter a in this section completely depends on both the range of ages and calendar years used to fit the GLM.

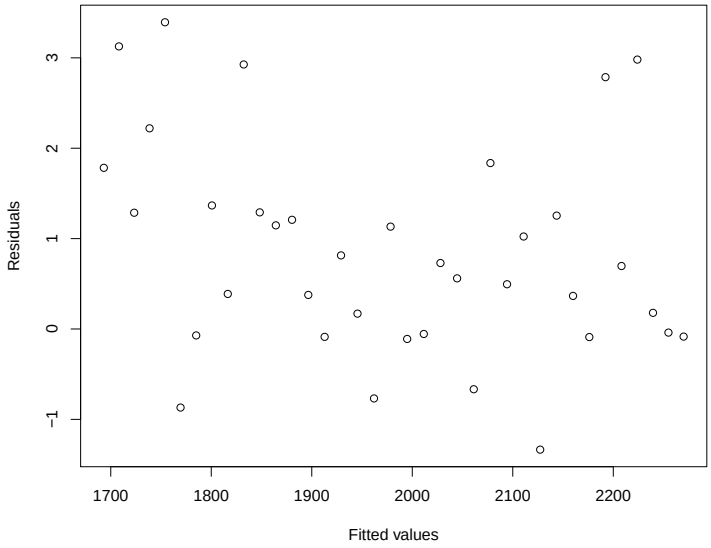

Figure 1.3: Residuals v.s. fitted values, x=70, calendar year ranges from 1958-1994

One of the main disadvantages of this model (i.e. $a = 0.6$) is the large values of the residual deviances[1] which are illustrated in Figure 1.1. Changing the parameters in Table 1.1 has a significant effect on residual deviance. We emphasize that the numbers r and s appearing in model (1.1) should be increased when modeling mortality rates for ages more than 50. This is necessary to capture the evolution of mortality rates for older ages. Hence we focus on younger ages and try to find 3 reasonable models for ages $x = 20, 30, 40$ years and decrease residual deviance in Figure 1.1 by using (1.6) and the logistic regression model (1.4).

Table 1.3: ANOVA Table, Significance Codes: '***' 0, '**' 0.001

	Df	$DevianceResidual$	Df	$ResidualDeviance$	$P(>\|Chi\|)$
$NULL$			1331	820992	
$\hat{\beta}_1$	1	70756	1330	750236	2.2×10^{-16} ***
$\hat{\beta}_2$	1	661239	1329	88997	2.2×10^{-16} ***
$\hat{\beta}_3$	1	39563	1328	49437	2.2×10^{-16} ***
$\hat{\beta}_4$	1	1576	1327	47858	2.2×10^{-16} ***
$\hat{\beta}_5$	1	36251	1326	11608	2.2×10^{-16} ***
$\hat{\beta}_6$	1	8307	1325	3301	2.2×10^{-16} ***
$\hat{\beta}_8$	1	25	1324	3276	0.001393 **

[1]The residual deviance was defined in (1.8).

Table 1.4: Parameters Set for x=20

$x_1 = 15$	$x_2 = 25$	$y_1 = 1948$	$y_2 = 2006$	$N = 20$
$b = 0, 1$	$\gamma = 1$	$a_1 = 0$	$a_2 = 40$	$r(x,t) = l(x,t)$

In Figure 1.4, we compare the model in (1.6) for two cases, $b = 0$ and $b = 1$ with the set of parameters provided in Table 1.4. Therefore, we reduce the data set to ages ranging from 15-25 years and calendar years 1948-2006. Parameter a ranges in the interval $(0, 40]$. The same argument as mentioned in the previous section is considered to select the best model. We fix the parameter a and fit the model in (1.6). The preferred model is the one that has minimum residual deviance. We apply an offset term $r(x,t) = l(x,t)$.

Panels (A) and (B) in Figure 1.4 plot the values of parameter a against the corresponding residual deviances. In panel (A) when $b = 1$, the residual deviance has attained its minimum value of 264 at $a = 24$. However, in panel (B) where $b = 0$, residual deviance is minimized at $a = 7$ with a minimum value of 313. This shows that we can decrease the residual deviance by about 16% when adding the parameter b into the SDE proposed by Ballotta and Haberman (2006).

Panels (C) and (D) of Figure 1.4 are assigned to show predicted force of mortalities (filled circles) for the optimum value of a and crude mortality rates (empty circles), respectively. Comparing predicted forces of mortality in panel (C) with panel (D) of Figure 1.4 indicates that, when $b = 1$, these forecasts are shifted downwards slightly to match the crude mortality rates better. This is due to the reduction of the residual deviance resulting from the inclusion of the term b in the OU process, as mentioned above. In addition to the said procedure, forecasts are carried out for 20 years. Both models in Figure 1.4 fit the historical data well. However, model (1.6) with $b = 1$ seems to be more adequate.

Figure 1.5 compares the logit model in (1.4) and the GLM in (1.6) with $b = 0$, where we fix age $x = 20$ years. The set of parameters is given in Table 1.4. The residual deviance is minimized at $a = 7$, with minimum value of 313, for both models (1.4) and (1.6) as shown in panels (A) and (B). We indicate predicted forces of mortality (filled circles) and crude mortality rates (empty circles) in panels (C) and (D) for the logit model in (1.4) and the GLM in (1.6), respectively. Forecasting for 20 years shows the same trend for both models. This comes from the fact that both the logit model in (1.4) and the GLM in (1.6) with $b = 0$, have identical explanatory random variables. Comparing minimum residual deviances in panels (A) of Figure 1.4 and Figure 1.5 reveals that the GLM (1.6) with $b = 1$ is the dominant model.

We fix age $x = 30$ in Figure 1.6 and compare the model (1.6) in two cases $b = 0$ and $b = 1$. We give the set of parameters for this example in Table 1.5. Since now we are interested in age 30, we changed the historical data to better capture the evolution of mortality. Ages vary from 25-35 and calendar years range from 1950-2006. The parameter γ is set to 1 and a is chosen in the interval $[-20, 40]\backslash\{0\}$.

Panel (A) of Figure 1.6 shows that for the case $b = 1$, the residual deviance is minimized at $a = 7$, with minimum value of 230, whereas in Panel (B) where $b = 0$, this minimization is found at $a = 5$ with corresponding minimum value of 233.

Panels (C) and (D) plot predicted forces of mortality (filled circles) for the optimum value of a found above as well as crude mortality rates (empty circles). We predict mortality rates for 30 years. One can compare the two upper panels (A), (B) and the two lower panels (C) and (D) in Figure 1.6, to conclude that both models have almost the same power in predicting mortality rates. We should choose model (1.6) with $b = 0$ if we are interested in a model with fewer parameters to estimate. However, we would like to emphasize that the model (1.6) with $b = 1$ would be a strong candidate, by reducing residual deviance significantly if we change the historical data.

The logit model in (1.4) behaves similarly to the model in (1.6) with $b = 1$, both in optimization for parameter a as well as in predictions and, therefore we did not include a graph for the logit model.

In Figure 1.7, we fix age $x = 40$ and examine the difference in model (1.6) for two cases, $b = 0$ and $b = 1$. We provide the set of parameters for this example in Table 1.6 and change the historical data for ages ranging from 37-43 and calendar years 1948-2006. Parameter a takes its value in the interval $(0, 70]$.

In panel (A) where $b = 1$, the residual deviance attains its minimum value of 214 at $a = 38$. However, the corresponding smallest value of the residual deviance when $b = 0$ is 221, and it occurs at $a = 11$. We present forecasting for the forces of mortality (filled circles) with $a = 38$ and $a = 11$, as well as crude mortality rates (empty circles) in panels (C) and (D), respectively. We carry out predictions from 2007-2036 for 30 years. Both models fit well, however based on the minimum value for the residual deviance, we select model (1.6) with $b = 1$.

Table 1.5: Parameters Set for x=30

$x_1 = 25$	$x_2 = 35$	$y_1 = 1950$	$y_2 = 2006$	$N = 30$
$b = 0, 1$	$\gamma = 1$	$a_1 = -20$	$a_2 = 40$	$r(x,t) = l(x,t)$

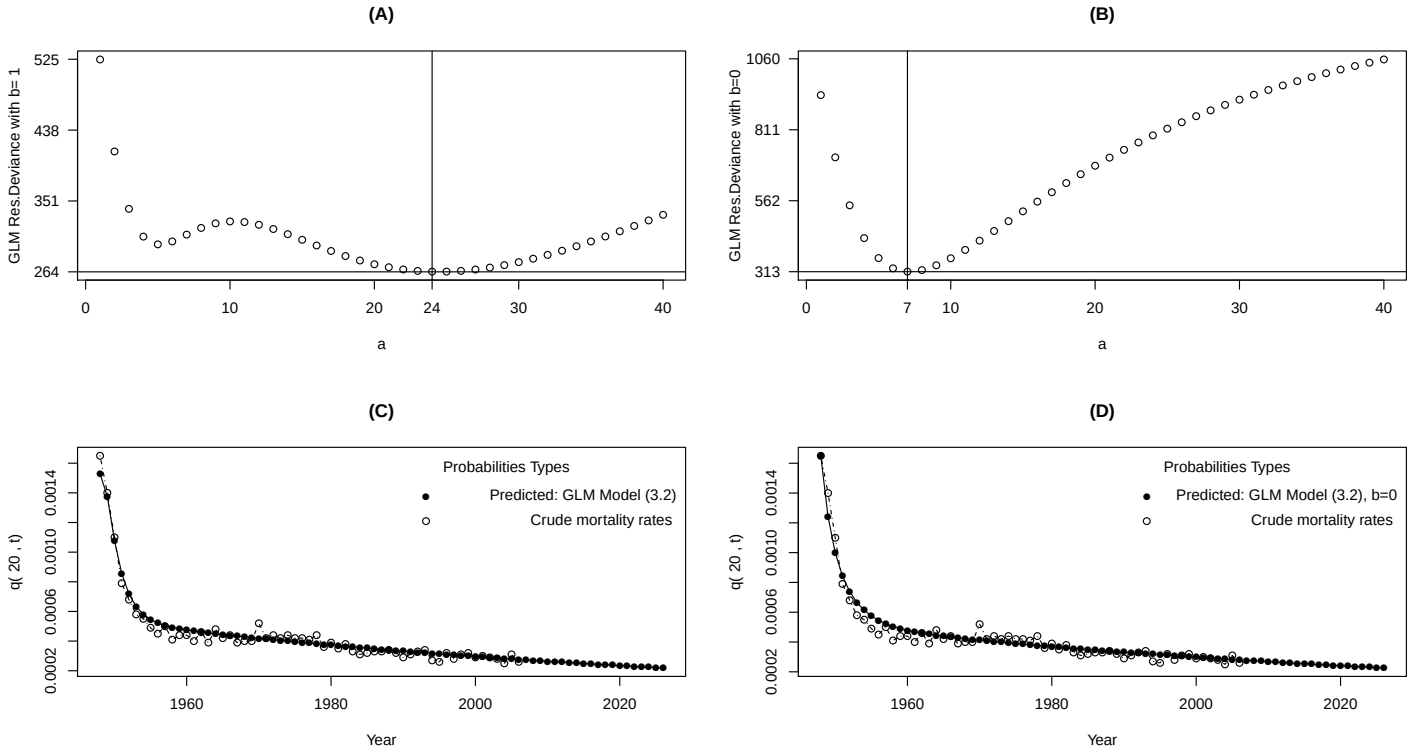

Figure 1.4: Comparison of model (3.2) with $b = 1$, $b = 0$, $x = 20$

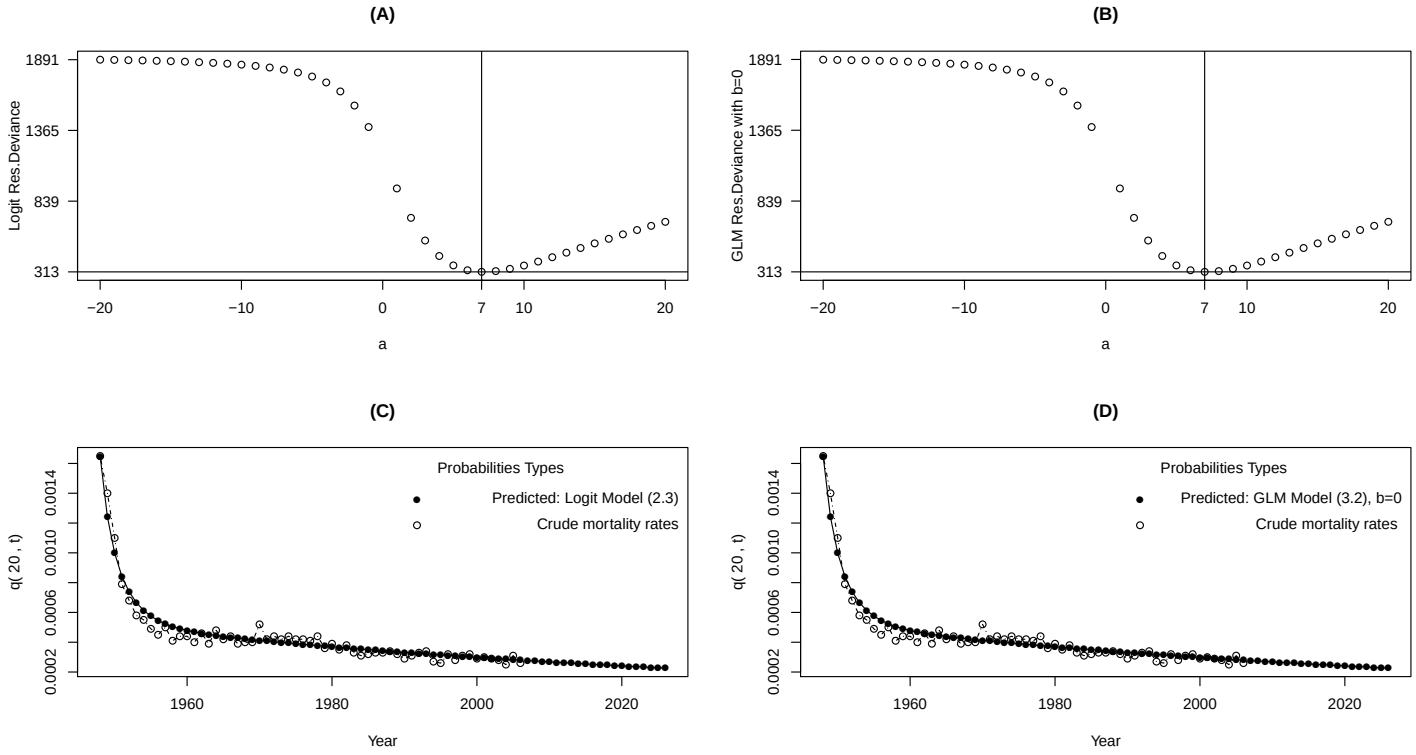

Figure 1.5: Comparison of model (3.2) with $b = 0$ and logit model (2.3)

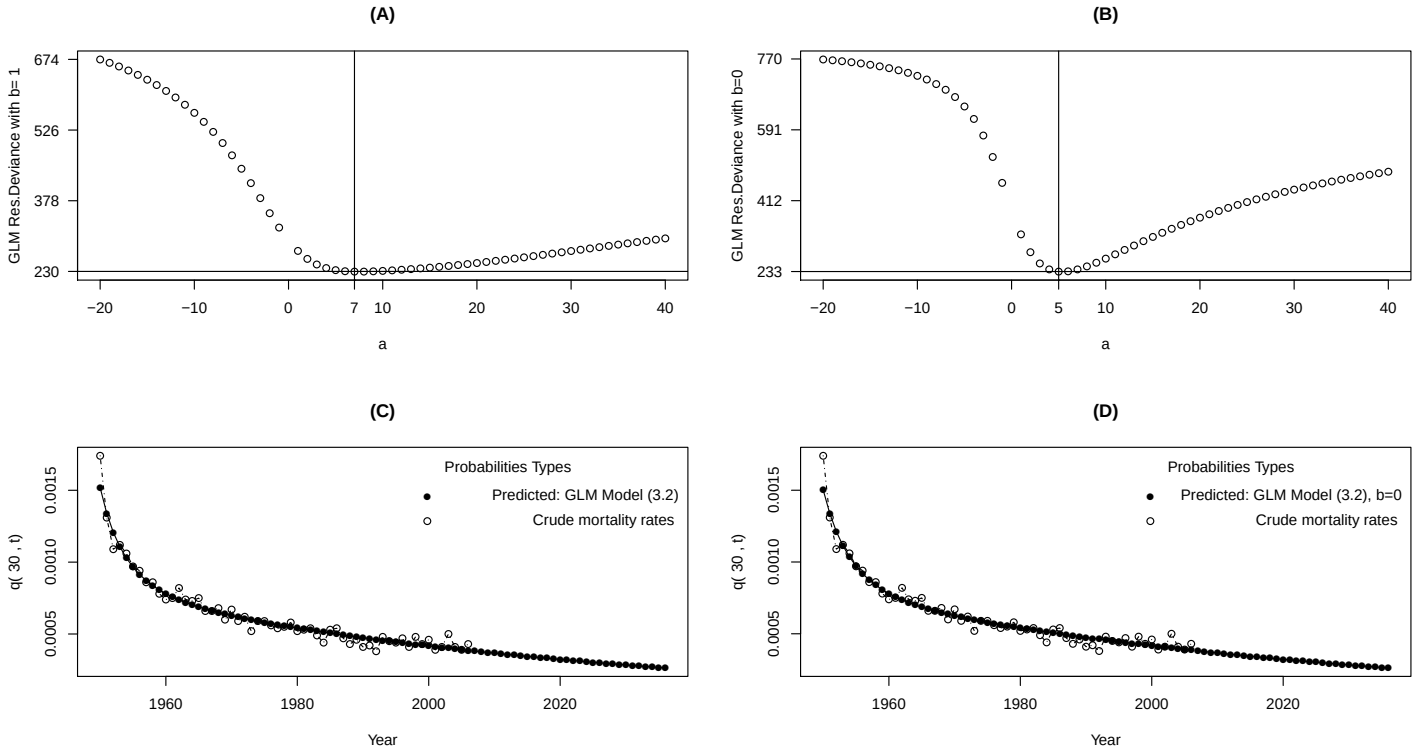

Figure 1.6: Comparison of model (3.2) with $b = 1$, $b = 0$, $x = 30$

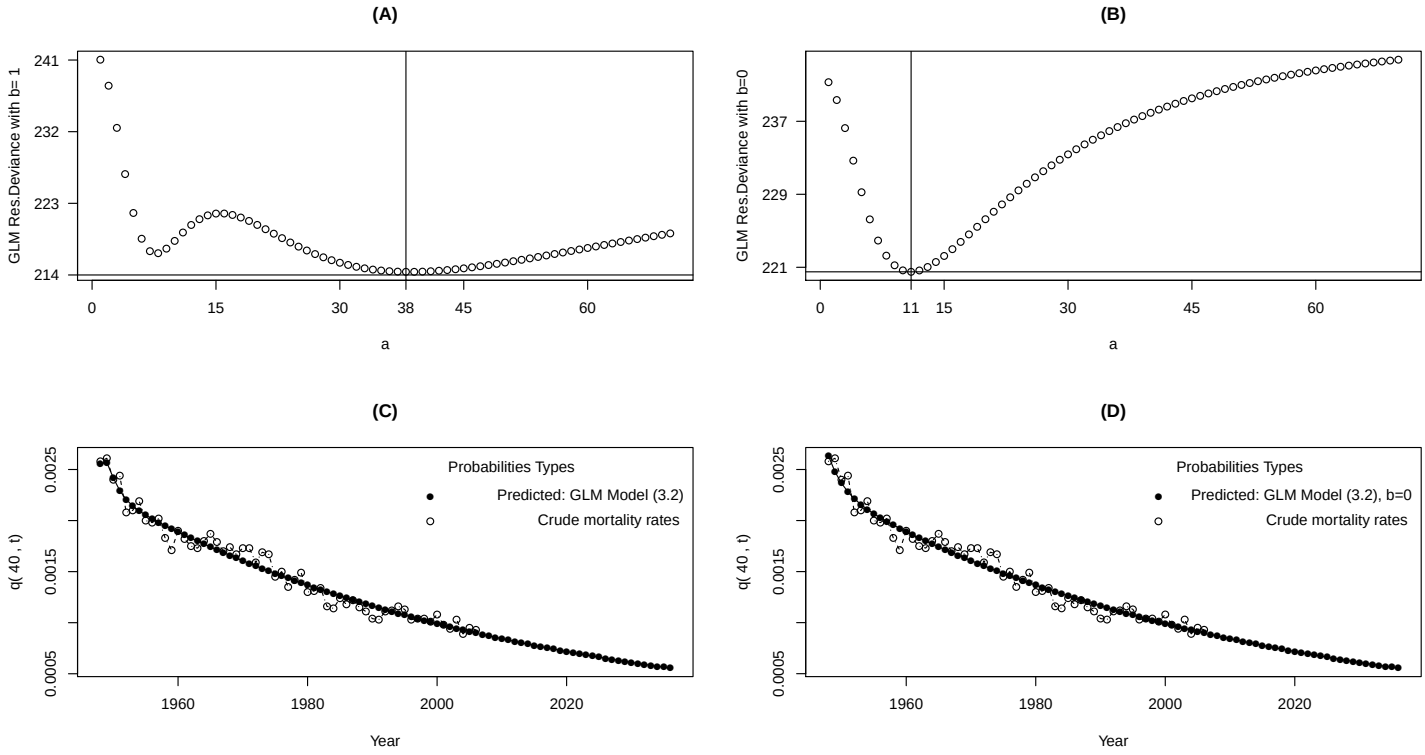

Figure 1.7: Comparison of model (3.2) with $b = 1$, $b = 0$, $x = 40$

Table 1.6: Parameters Set for x=40

$x_1 = 37$	$x_2 = 43$	$y_1 = 1948$	$y_2 = 2006$	$N = 30$
$b = 0, 1$	$\gamma = 1$	$a_1 = 0$	$a_2 = 70$	$r(x,t) = l(x,t)$

1.4 Conclusions

In this chapter, we added a random perturbation to the traditional GLM used to model mortality rates. The evolution of this random variable follows a one-dimensional Ornstein-Uhlenbeck process. Then, we used the distribution of the perturbed term to model the number of deaths. Moreover, we proposed a new approach for finding the parameter associated with the dynamic part of the force of mortality. We minimized the residual deviance for this purpose. It is worth mentioning that the optimum value found by this method depends, not only on the range of ages considered to do the modelling, but also the time period used to fit the model.

Two models were suggested to project mortality rates. The first model is based on modelling the number of deaths using GLM techniques and then trying to project the mortality rates. We consider two cases, $b = 0$ and $b = 1$ for this model. Also logistic regression was defined and used to directly predict the probability of death. A comparison of these models is carried out for 3 different ages, 20, 30 and 40 years. The performance of the logit model is similar to that of the GLM with $b = 0$. The adequacy of the GLM with $b = 1$ is much better than with $b = 0$ for ages 20 and 40. However, the two GLM models almost have the same power for age 30.

Further research should consider more complicated SDEs for the dynamic part of the model. In addition, other data sets can be compared by using this approach. Also, we can generalize the idea given in this chapter by adding cohort effects to the model to see their impact on the fit.

Chapter 2

Jump Diffusion Processes

2.1 Introduction

Jump diffusion processes are stochastic processes that include both jumps and diffusion. For instance these can be combinations of compound Poisson and Brownian motion. Recently, Lévy processes have been used to model mortality rates. Chen and Cox (2009) include a jump process into the original LC model. They apply this model to forecast mortality rates and examine mortality securitization. This model is useful to study mortality dependent cash flows of life insurance portfolios, annuity portfolios, and portfolios of mortality derivatives.

As mentioned by Wang et al. (2010), many of the models discussed in Section 1.1, do not include short-term catastrophe mortality shocks, such as the influenza pandemic in 1918 and the Indonesian tsunami in December 2004, which may lead to much higher (or lower) mortality rates. This can justify, to some extent, the reason why we include Lévy processes in modelling mortality rates. Therefore, we consider Lévy processes in this chapter to capture the effect of shocks in mortality. Parameters that characterize the Lévy processes can control jump intensities and play an important role when considering mortality rates at older ages. We give some examples in Section 2.3.3 to show how this process can provide a better fit to the data and captures the trend (compared to the model developed in the previous chapter).

We start with the following definition of a Lévy process given in Cont and Tankov (2004).

Definition 2.1. *A stochastic process $(Z_t)_{t\geq 0}$ on $(\Omega, \mathcal{F}, \mathbb{P})$ with values in $\mathbb{R}$ such that $Z_0 = 0$ is called a Lévy process if it posesses the following properties:*

1. *Independent increments: for every increasing sequence of times $t_0, ..., t_n$, the random variables $Z_{t_0}, Z_{t_1} - Z_{t_0}, ..., Z_{t_n} - Z_{t_{n-1}}$ are independent.*

2. *Stationary increments: the law of* $Z_{t+h} - Z_t$ *does not depend on* t.

3. *Stochastic continuity:* $\forall \epsilon > 0, \quad \lim_{h \to 0} \mathbb{P}(|Z_{t+h} - Z_t| \geq \epsilon) = 0.$

We provide some additional definitions and results from Kyprianou (2006) in Appendix A, that will be used in Section 2.2. This includes definitions of an infinitely divisible distribution, Lévy measure, Poisson random measure and subordinators. Moreover, we summarize in Appendix A some important results including the Lévy-Khintchine formula, the bounded variation form of Lévy processes and a multidimensional Itô formula.

2.2 Model Fitting

Mortality modelling in Chapter 1 is extended to include Lévy processes. We keep the same structure of conditional expectation as in (1.3). This consistency is necessary to determine the effect of the generalization in the case of Lévy processes. Therefore, we assume that

$$\begin{aligned} \mathbb{E}[D(x,t)|Y_{t'}] = r(x,t) \exp\Big(\beta_0 \quad &+ \quad \sum_{i=1}^{3} \beta_i L_i(x') + \beta_4 t' \\ &+ \quad \sum_{i=1}^{2} \beta_{i+4} L_i(x') t' + \beta_7 Y_{t'}\Big), \end{aligned} \tag{2.1}$$

where Y_t follows an Ornstein-Uhlenbeck process (OU) with respect to a Lévy process i.e.

$$\begin{cases} dY_t = a(b - Y_t)d_t + \gamma dZ_t \\ Y_0 = 0, \end{cases} \tag{2.2}$$

where a, b, γ are constant and $(Z_t : t \geq 0)$ is a subordinator without drift as defined in (A.4). Hence Z_t can be written as (A.3) with $d = 0$.

Proposition 2.1. *For* Y_t *defined in (2.2), we have,*

$$d(Y_t e^{at}) = e^{at} dY_t + a e^{at} Y_t d_t. \tag{2.3}$$

Proof. To show (2.3), we need Theorem A.3 in Appendix A from Oksendal and Sulem (2010). This is known as the multidimensional Itô's formula. Let $n = 2$ and $l = 1$ in Theorem A.3 and define $K(t) = f(t, \mathbf{Y}_t) = Y_t^{(1)} Y_t^{(2)}$, where $Y_t^{(i)}, \quad i = 1, 2$ are two Itô-Lévy processes of the form (A.4). Hence, $\frac{\partial f}{\partial t} = 0$, $\frac{\partial f}{\partial y^{(1)}} = y^{(2)}$, $\frac{\partial f}{\partial y^{(2)}} = y^{(1)}$, $\frac{\partial^2 f}{\partial y^{(1)} \partial y^{(2)}} = \frac{\partial^2 f}{\partial y^{(2)} \partial y^{(1)}} = 1$, $\frac{\partial^2 f}{\partial y^{(1)} \partial y^{(1)}} = \frac{\partial^2 f}{\partial y^{(2)} \partial y^{(2)}} = 0$. Now using Theorem A.3,

we obtain

$$\begin{aligned} dK(t) &= Y_t^{(2)}(\alpha_1 dt + \sigma_1 dB_t) + Y_t^{(1)}(\alpha_2 dt + \sigma_2 dB_t) + \frac{1}{2}2\sigma_1\sigma_2 dt \\ &+ \int_{|z|<R} \Big\{ f\Big(t, \begin{bmatrix} Y_{t^-}^{(1)} + \gamma_1(t,z) \\ Y_{t^-}^{(2)} + \gamma_2(t,z) \end{bmatrix}\Big) - f\Big(t, \begin{bmatrix} Y_{t^-}^{(1)} \\ Y_{t^-}^{(2)} \end{bmatrix}\Big) \\ &\quad -\gamma_1(t,z)Y_{t^-}^{(2)} - \gamma_2(t,z)Y_{t^-}^{(1)} \Big\}\nu(dz)dt \\ &+ \int_{\mathbb{R}} \Big\{ f\Big(t, \begin{bmatrix} Y_{t^-}^{(1)} + \gamma_1(t,z) \\ Y_{t^-}^{(2)} + \gamma_2(t,z) \end{bmatrix}\Big) - f\Big(t, \begin{bmatrix} Y_{t^-}^{(1)} \\ Y_{t^-}^{(2)} \end{bmatrix}\Big)\Big\}\bar{N}(dt,dz). \end{aligned}$$

Therefore,

$$\begin{aligned} dK(t) &= Y_t^{(2)}(\alpha_1 dt + \sigma_1 dB_t) + Y_t^{(1)}(\alpha_2 dt + \sigma_2 dB_t) + \sigma_1\sigma_2 dt \\ &+ \int_{|z|<R} \Big\{ \big(Y_{t^-}^{(1)} + \gamma_1(t,z)\big)\big(Y_{t^-}^{(2)} + \gamma_2(t,z)\big) \\ &\quad -Y_{t^-}^{(1)}Y_{t^-}^{(2)} - Y_{t^-}^{(2)}\gamma_1(t,z) - Y_{t^-}^{(1)}\gamma_2(t,z) \Big\}\nu(dz)dt \\ &+ \int_{\mathbb{R}} \Big\{ \big(Y_{t^-}^{(1)} + \gamma_1(t,z)\big)\big(Y_{t^-}^{(2)} + \gamma_2(t,z)\big) - Y_{t^-}^{(1)}Y_{t^-}^{(2)} \Big\}\bar{N}(dt,dz). \end{aligned}$$

Hence,

$$\begin{aligned} dK(t) &= Y_t^{(2)}(\alpha_1 dt + \sigma_1 dB_t) + Y_t^{(1)}(\alpha_2 dt + \sigma_2 dB_t) + \sigma_1\sigma_2 dt \\ &+ \int_{|z|<R} \gamma_1(t,z)\gamma_2(t,z)\nu(dz)dt + \int_{\mathbb{R}} \Big\{ \gamma_1(t,z)\gamma_2(t,z) \\ &\quad +Y_{t^-}^{(1)}\gamma_2(t,z) + Y_{t^-}^{(2)}\gamma_1(t,z) \Big\}\bar{N}(dt,dz). \end{aligned} \tag{2.4}$$

Now we let $Y_t^{(1)} = e^{at}$ and $Y_t^{(2)} = Y_t$, where Y_t has been defined in (2.2). As a result, $dY_t^{(1)} = ae^{at}dt$ and $dY_t^{(2)} = dY_t = a(b - Y_t)d_t + \gamma dZ_t$. Note that both $Y_t^{(1)}$ and $Y_t^{(2)}$ are Itô-Lévy processes with $\alpha_1(t,\omega) = ae^{at}$, $\sigma_1(t,\omega) = \gamma_1(t,z) = 0$, $\alpha_2(t,\omega) = a(b - Y_t)$, $\sigma_2(t,\omega) = 0$ and $\gamma_2(t,z) = \gamma z$. Next we plug in these values in (2.4) to obtain

$$d(e^{at}Y_t) = Y_t a e^{at}dt + e^{at}\Big(a(b - Y_t)dt\Big) + \int_{\mathbb{R}} z\gamma e^{at}\bar{N}(dt,dz). \tag{2.5}$$

We let $R = 0$ in

$$\bar{N}(dt,dz) = N(dt,dz) - \chi_{(|z|<R)}\nu(dz)dt$$

to have $\bar{N}(dt,dz) = N(dt,dz)$. Hence (2.5) can be written as:

$$\begin{aligned} d(e^{at}Y_t) &= Y_t a e^{at}dt + e^{at}\Big(a(b - Y_t)dt\Big) + \int_R z\gamma e^{at}N(dt,dz) \\ &= e^{at}\Big\{a(b - Y_t)dt + \gamma\int_R zN(dt,dz)\Big\} + aY_t e^{at}dt. \end{aligned} \tag{2.6}$$

Since we assumed that Z_t in (2.2) is a subordinator without drift, we can use representation (A.3) to have

$$Z_t = \int_{[0,t]} \int_{\mathbb{R}} zN(ds, dz), \quad t \geq 0,$$

or

$$dZ_t = \int_{\mathbb{R}} zN(dt, dz), \quad t \geq 0. \tag{2.7}$$

From (2.7) we can rewrite (2.2) as:

$$\begin{cases} dY_t = a(b - Y_t)d_t + \gamma \int_{\mathbb{R}} zN(dt, dz) \\ Y_0 = 0. \end{cases} \tag{2.8}$$

From (2.8) and (2.6), we obtain

$$d(e^{at}Y_t) = e^{at}dY_t + aY_t e^{at}dt. \tag{2.9}$$

□

Proposition 2.2. *The solution to (2.2) for $\gamma = \frac{1}{\beta_7}$ satisfies*

$$Y_t = b\Big(1 - e^{-at}\Big) + \frac{1}{\beta_7} \int_0^t e^{-a(t-u)} dZ_u.$$

Proof. We solve (2.2), for all $s > t$ and let $f(s, Y_s) = \eta_s = e^{as}(b - Y_s) = be^{as} - Y_s e^{as}$. Using Proposition 2.1 and (2.2) leads to,

$$\begin{aligned} d\eta_s &= abe^{as}ds - d(Y_s e^{as}) \\ &= abe^{as}ds - e^{as}dY_s - ae^{as}Y_s ds \\ &= ae^{as}(b - Y_s)ds - e^{as}dY_s \\ &= -\gamma e^{as}dZ_s. \end{aligned} \tag{2.10}$$

Hence $\eta_s = \eta_t - \gamma \int_t^s e^{au}dZ_u$. Since $\eta_s = e^{as}(b - Y_s)$, we get

$$\begin{aligned} Y_s &= b - e^{-as}\eta_s = b - e^{-as}\Big(\eta_t - \gamma \int_t^s e^{au}dZ_u\Big) \\ &= b - e^{-as}\Big(e^{at}(b - Y_t) - \gamma \int_t^s e^{au}dZ_u\Big) \\ &= b\Big(1 - e^{-a(s-t)}\Big) + e^{-a(s-t)}Y_t + \gamma \int_t^s e^{-a(s-u)}dZ_u. \end{aligned} \tag{2.11}$$

□

To find the moments of Y_t, we let $t = 0$, $\gamma = \frac{1}{\beta_7}$ and change the subscript s for t in (2.11) to have

$$Y_t = b\Big(1 - e^{-at}\Big) + \frac{1}{\beta_7} \int_0^t e^{-a(t-u)} dZ_u. \tag{2.12}$$

We need to find $\mathbb{E}(e^{\beta_7 Y_t})$. However this expectation is difficult to evaluate even for subordinators, Z_t. As a result we consider the class of tempered α-stable subordinators among positive Lévy processes and focus on two special cases, gamma and inverse Gaussian processes, in the next section.

2.3 Gamma and Inverse Gaussian Processes

The next definition is from Cont and Tankov (2004).

Definition 2.2. *A process Z_t is said to be an α-stable subordinator, with $\alpha \in (0,1]$ if for all $a > 0$, the following holds in distribution:*

$$\Big(\frac{Z_{at}}{a^{\frac{1}{\alpha}}}\Big) \stackrel{\mathrm{d}}{=} Z_t.$$

The Lévy measure of the α-stable subordinator is

$$\nu(z) = \frac{c}{z^{\alpha+1}} \cdot \chi_{z>0}, \quad \alpha \in (0,1].$$

The tempered stable subordinator is an exponentially tempered version of the α-stable subordinator. In this case we multiply the Lévy measure above with a decreasing exponential:

$$\nu(z) = \frac{ce^{-\lambda z}}{z^{\alpha+1}} \cdot \chi_{(z>0)}, \tag{2.13}$$

where c and λ are positive constants, $\alpha \in [0,1]$ and $\chi_{(z>0)} = 1$ if $z > 0$ and $\chi_{(z>0)} = 0$ otherwise. The case $\alpha = 0$ has been included for greater generality, although it cannot be obtained from an α-stable subordinator. As mentioned in Cont and Tankov (2004), the parameter c changes the intensity of jumps of all sizes simultaneously, λ fixes the decay rate of big jumps and α determines the relative importance of small jumps in the path of the process. If for the subordinator Z_t, we define the cumulant transform

$$k(\theta) = \log\Big(\mathbb{E}(e^{\theta Z_1})\Big), \tag{2.14}$$

then we get the following useful closed expressions:

$$k(\theta) = \int_0^{+\infty} \Big(e^{\theta z} - 1\Big)\nu(dz) = \begin{cases} c.\Gamma(-\alpha)\{(\lambda-\theta)^\alpha - \lambda^\alpha\} & \text{if } \alpha \neq 0 \\ -c.\log\Big(1 - \frac{\theta}{\lambda}\Big) & \text{if } \alpha = 0 \end{cases} \tag{2.15}$$

When $\alpha = 0$, we get the gamma process with probability density function:

$$p_{Z_t}(z) = \frac{\lambda^{ct}}{\Gamma(ct)} z^{ct-1} e^{-\lambda z}, \quad \forall z > 0.$$

For the case $\alpha = 0.5$, we get the inverse Gaussian process. The following theorem from Eberlein and Raible (1999) should be used in order to have $\mathbb{E}(e^{\beta_7 Y_t})$.

Theorem 2.1. *Let Z_t be a subordinator with cumulant transform $k(\theta)$ in (2.14) and $f : \mathbb{R}_+ \to \mathbb{C}$ be a complex valued left continuous function such that $|Re(f)| \leq M$ then*

$$\mathbb{E}\Big(\exp\Big(\int_0^t f(\theta) dZ_{\vartheta\theta} \Big) \Big) = exp\Big(\int_0^t \vartheta k(f(\theta)) d\theta \Big). \tag{2.16}$$

Proof. See Lemma 3.1 in Eberlein and Raible (1999). □

It follows from (2.12) that

$$\begin{aligned} \mathbb{E}(e^{\beta Y_t}) &= \mathbb{E}\Big(\exp\big(b\beta(1-e^{-at}) + \int_0^t e^{-a(t-u)} dZ_u \big) \Big) \\ &= e^{b\beta(1-e^{-at})} E\Big(\exp\big(\int_0^t e^{-a(t-u)} dZ_u \big) \Big). \end{aligned} \tag{2.17}$$

2.3.1 Gamma Process

Now for the case $\alpha = 0$, we can use (2.16) and (2.17) with $\beta = \beta_7$ to get

$$\begin{aligned} \mathbb{E}(e^{\beta_7 Y_t}) &= \exp\Big(b\beta_7(1-e^{-at}) + \int_0^t \{ -c \log(1 - \frac{e^{-a(t-\theta)}}{\lambda}) \} d\theta \Big) \\ &= \exp\Big(b\beta_7(1-e^{-at}) - c\frac{dilog(\frac{\lambda - e^{-at}}{\lambda}) - dilog(\frac{\lambda-1}{\lambda})}{a} \Big), \end{aligned} \tag{2.18}$$

where $dilog(x)$ is the dilogarithm function and is defined as $dilog(x) = \int_1^x \frac{\log(u)}{1-u} du$. Taking another expectation from (2.1) with (2.18) implies that

$$\begin{aligned} \mathbb{E}[D(x,t)] = r(x,t) \exp\Big(\beta_0 + \sum_{i=1}^{3} \beta_i L_i(x') + \beta_4 t' + \sum_{i=1}^{2} \beta_{i+4} L_i(x') t' \\ + b\beta_7(1 - e^{-at'}) - c\frac{dilog(\frac{\lambda - e^{-at'}}{\lambda}) - dilog(\frac{\lambda-1}{\lambda})}{a} \Big) \end{aligned} \tag{2.19}$$

2.3.2 Inverse Gaussian Process

Let $\alpha = 0.5$ and similarly for $t > 0$ we have:

$$\mathbb{E}(e^{\beta Y_t}) = \exp\Big(b\beta(1-e^{-at}) + c\Gamma(-\alpha) \int_0^t \big\{ \big(\lambda - e^{-a(t-\theta)}\big)^{0.5} - \lambda^{0.5} \big\} d\theta \Big), \tag{2.20}$$

with

$$\begin{aligned} \int_0^t \big\{ \big(\lambda - e^{-a(t-\theta)}\big)^{0.5} - \lambda^{0.5} \big\} d\theta &= \frac{-1}{a} \Big[\big\{ 2\sqrt{e^{-at}(\lambda e^{at} - 1)} \\ &\quad -2\sqrt{\lambda} \quad arctanh(\frac{\sqrt{e^{-at}(\lambda e^{at} - 1)}}{\sqrt{\lambda}}) - 2\sqrt{\lambda - 1} \big\} \\ &\quad +2\sqrt{\lambda} \quad arctanh(\frac{\sqrt{\lambda - 1}}{\sqrt{\lambda}} + at\sqrt{\lambda}) \Big], \end{aligned} \tag{2.21}$$

where *arctanh* is the hyperbolic arctangent function.

We have expressed the expected value of the number of deaths as an exponential function of some observable random variables in (2.19). Hence we are in a position to use GLM's to fit the mortality model.

Note that in (2.19) the parameter c can be treated as β_8 in modelling and estimating from the GLM, although this estimation may not preserve the positivity condition of c. Whenever this problem happens, we can change the sign of the corresponding explanatory random variable for parameter c in (2.19) to get a positive value for c.

Theorem 1.1 implies that the predicted number of deaths will not change and we can still forecast the mortality rates without having estimation problems. In addition, we should mention that the main goal of this research is forecasting the mortality rates rather than estimating the parameters appearing in (2.13).

2.3.3 Examples

In this section, we would like to compare models (1.6) and (2.19). In other words, we compare the GLM models based on Brownian motion (BM) and the Lévy tempered stable subordinator. The data has been already presented in Section 1.2.3. Specifically, we used the female mortality rates for the United Kingdom during calendar years 1922-2006 and for ages ranging from 0-110 years. The idea of selecting the best model is basically the same as explained there. The only difference in the case of the Lévy subordinator is its 3 different parameters, a, b, and λ. By Theorem 1.1, we can fix parameter $b = 1$, and try to find the parameters a and λ that minimize the residual deviance. In Table 2.1, we summarize the parameters values that have been used to obtain the best possible values for parameters $a \in [a_1, a_2]\backslash\{0\}$ and $\lambda \in [\lambda_1, \lambda_2]\backslash\{0\}$.

The optimization has been carried out over the set of parameters in Table 2.1. In Table 2.2, we provide some subset of the results obtained to compare the residual deviance in two cases where $b = 1$. In this table, the column Res.Dev.BM gives the residual deviance for model (1.6) and the column Res.Dev.Lévy, provides the residual deviance for model (2.19). The optimization has been carried out over a wider range of parameters for λ that is mentioned in Table 2.1. Table 2.2 shows that, for the OU process, model (1.6) with a BM term, we get the minimum value of 264 for the residual deviance at $a = 26$. Similarly, for the the model (2.19) with the Lévy subordinator, we obtain the minimum value of 268 for the residual deviance occurring at $\lambda = 1$ and $a = 10$.

Now we can fit the models (2.19) and (1.6) with $b = 1$, $\lambda = 1$ and compare the results. We used the parameters given in Table 2.3, to produce Figure 2.1. Panels

Table 2.1: Parameters of the OU Process Including BM and Lévy Subordinator, x=20

$x_1 = 15$	$x_2 = 25$	$y_1 = 1948$	$y_2 = 2006$	$b = 1$
$\lambda_1 = -15$	$\lambda_2 = 15$	$a_1 = -4$	$a_2 = 50$	$r(x,t) = l(x,t)$

(A) and (B) of Figure 2.1 plot the residual deviance in (1.8) vs. parameter a with $\lambda = 1$, for models (2.19) and (1.6), respectively. It shows that model (1.6) is slightly better than the Lévy model (2.19) based on the residual deviance. However comparing panels (C) and (D) in Figure 2.1, indicates that the Lévy model (2.19) can capture the trend of mortality rates for the historical data, especially over the calendar years 1960-1980. This property is in contrast to the almost monotone decreasing trend in forecastings of model (1.6). We have also calculated the error sum of squares

$$SS = \sum_{x=x_1}^{x_2} \sum_{t=t_1}^{t_2} \Big(q(x,t) - \hat{q}(x,t) \Big)^2,$$

where $\hat{q}(x,t)$ is the predicted probability of death obtained by one of the two models. For the Lévy model (2.19), $SS = 9.310722 \times 10^{-8}$ and for the BM model (1.6), $SS = 9.369733 \times 10^{-8}$. The model based on a Lévy subordinator has a smaller SS. Table 2.4 gives the ANOVA table for the Lévy model (2.19). All the parameters are significant at $\alpha = 0.05$. Note that for simplicity of notation, we set

$$\zeta(a, \lambda, t') = \frac{dilog(\frac{\lambda - e^{-at'}}{\lambda}) - dilog(\frac{\lambda - 1}{\lambda})}{a}.$$

If we increase the age, then the improvement in using Lévy processes is much more significant. For example consider modelling mortality rates for age $x = 50$. Exactly the same procedure as for age $x = 20$ was used to estimate parameters a and λ. With the parameter values given in Table 2.5 we obtain $\lambda = 4$, $a = 6$ for model (2.19) and $a = 4$ for model (1.6).

The next step is to forecast mortality rates based on these optimized parameters. The corresponding parameter values for predictions are given in Table 2.6. In Figure 2.2 , we compare the forecasts for the two models (2.19) and (1.6). Panel (A) shows that the residual deviance is minimimal at $a = 6$ with a minimum value of 537. While in panel (B) the corresponding value for parameter $a = 4$, with a minimum value of 656 for model (2.19). This means that by using a model based on a Lévy tempered subordinator, we were able to decrease the residual deviance by almost 18%. Panels (C) and (D) indicate that the Lévy model is much better in capturing fluctuations in mortality rates, especially during calendar years 1960-1980, when more variations in mortality rates have been observed. Finally we

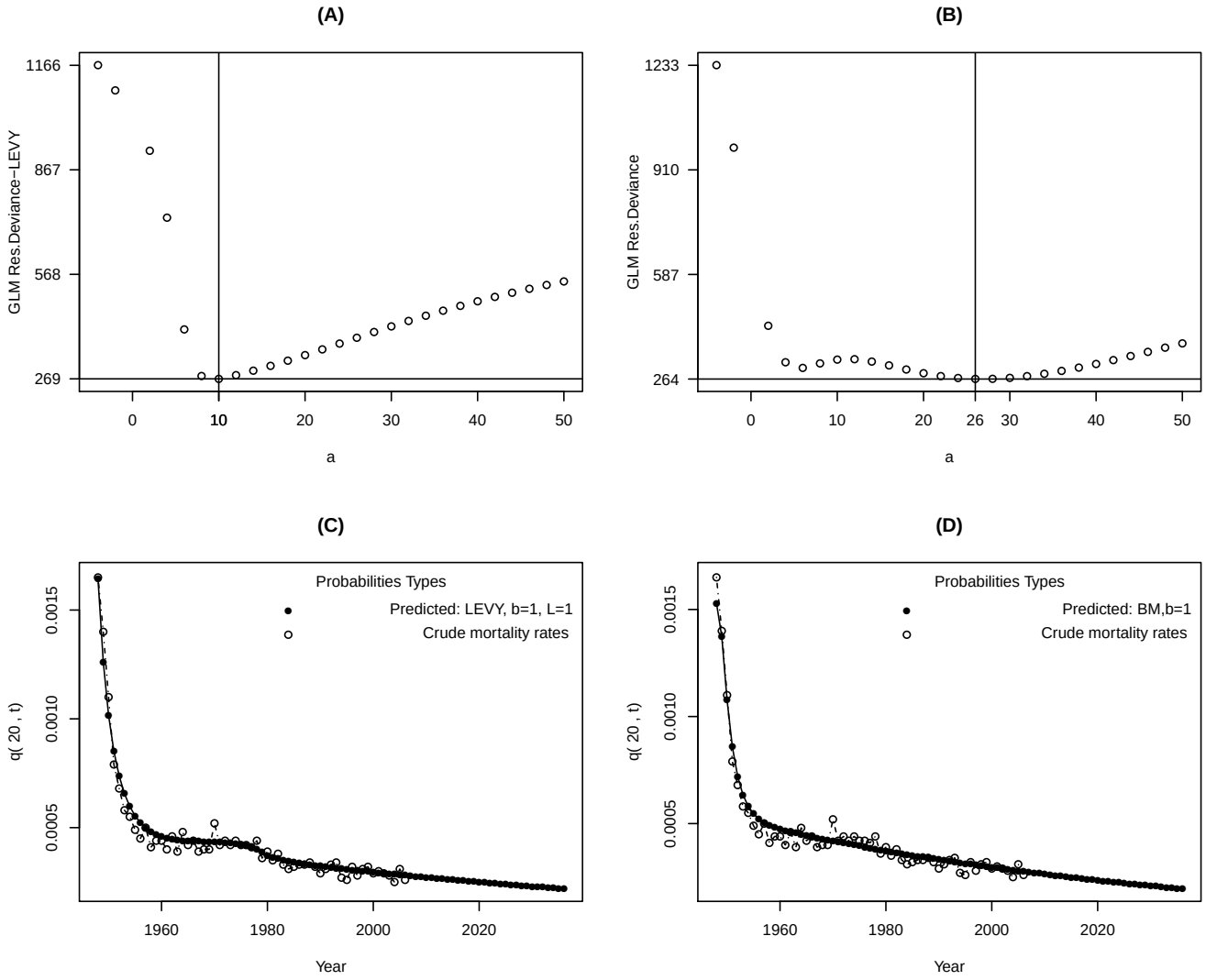

Figure 2.1: Comparison of models (1.6) and (2.19) for x=20, $\lambda = 1$, $b = 1$

mention that for model (2.19), $SS = 1.907015 \times 10^{-6}$. However the corresponding value for model (1.6), based on BM is $SS = 2.425177 \times 10^{-6}$. This implies that by using Lévy processes, we can reduce the error sum of squares of mortality rates by almost 21%.

We provide two more examples to predict mortality rates for the ages $x = 60$ and $x = 80$. Tables 2.7 and 2.8 give the set of parameter values used to find the optimum values for parameters a and λ. For age $x = 60$, we conclude that $a = 5$, for both models (2.19) and (1.6). Also $\lambda = 1$ is the obtained optimum value for the model in (2.19). Table 2.9 gives a brief summary of the results. Similarly at age $x = 80$, we found $a = -2$, $\lambda = 1$ in (2.19) and $a = -1$ in (1.6). Table 2.11 provides a subset of the results for this optimization.

Panels (A) and (B) of Figure 2.3, give the plot of the residual deviance vs.

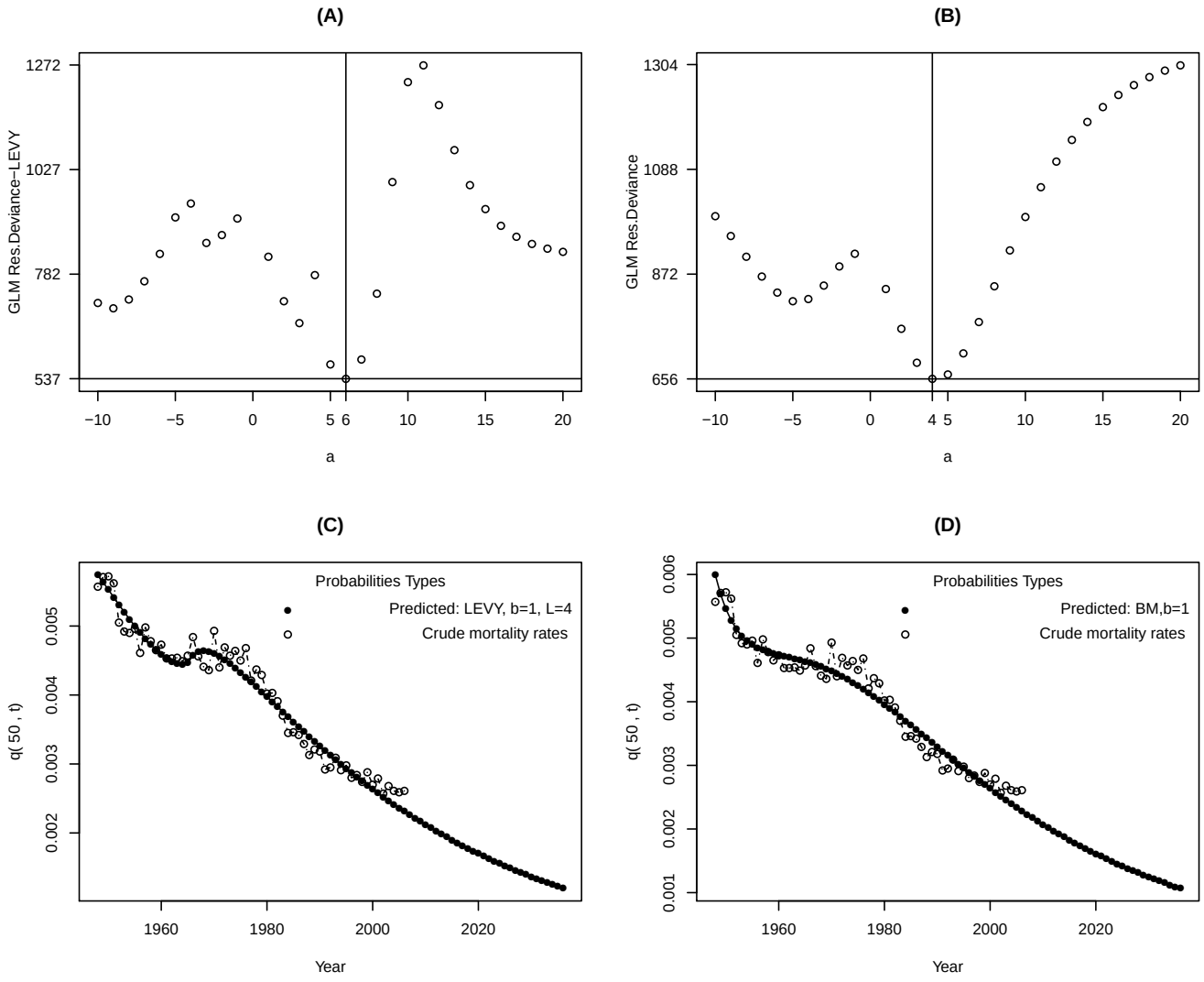

Figure 2.2: Comparison of models (1.6) and (2.19) for x=50, $\lambda = 4$, $b = 1$. Panels (A) and (B) show the minimization of residual deviance with respect to the parameter a in models (2.19) and (1.6), respectively. Panels (C) and (D) represent the observed mortality rates together with the fitted values in in models (2.19) and (1.6), respectively.

parameter a, with $\lambda = 1$ for models (2.19) and (1.6), respectively. A comparison of the fitted values (filled circles), i.e. the model values with fitted parameters, in panels (C) and (D) of Figure 2.3 shows that, prior to calendar year 1980, model (2.19) again can capture the trend slightly better than the model in (1.6). For the model (2.19) the sum of square errors of the fitted values is, $SS = 2.386065 \times 10^{-6}$, while the corresponding value for the model in (1.6) is $SS = 3.000105 \times 10^{-6}$. This indicates that Lévy processes can reduce our estimation error by almost 20%.

In Figure 2.4, panels (A) and (B) show the optimization over parameter a. The residual deviance is minimized at $a = -2$ for the model in (2.19), with a minimum value of 516. For the model in (1.6), we obtained $a = 5$ with minimum value of 552 for the residual deviance. This means that we reduce the residual deviance by almost 6% when applying Lévy processes. In the next chapter, we use AIC and BIC criteria to check for any over parametrization. Panel (C) in Figure 2.4 reveals that the fitting trend changes at calendar year 1980, while a gradual decreasing trend in the fitted values is observed in the Panel (D) of Figure 2.4. Moreover, we can see that as we increase the forecasting period after 2006, the predictions obtained by Lévy processes decline faster than those of model (1.6).

2.4 Conclusions

In this chapter, we have generalized model (1.6) to include more complex fluctuations based on Lévy tempered subordinators. First we solve the stochastic differential equation using Itô's formula. Then we get a closed form for the moments of the stochastic part, assuming a special case of Lévy processes, the so called tempered stable processes. For the gamma process presented in Section 2.3.1, we have optimized the parameters.

Finally, GLM's with Lévy tempered subordinators are compared with the model in Chapter 1 for 4 different ages $x = 20, 50, 60, 80$. We observe that the model in (2.19) based on Lévy processes is better than the OU model in (1.6) with a BM term to capture the trend, especially at older ages.

Further work on Lévy processes to be considered includes:

(a) Replacing the gamma process with an inverse Gaussian process by changing α to 0.5.

(b) Studying another mortality data set that includes mortality shocks such as the influenza pandemic in 1918 or the tsunami in the Indian Ocean of December 2004. This could explain whether Lévy processes can provide a better fit to the data with some mortality jumps.

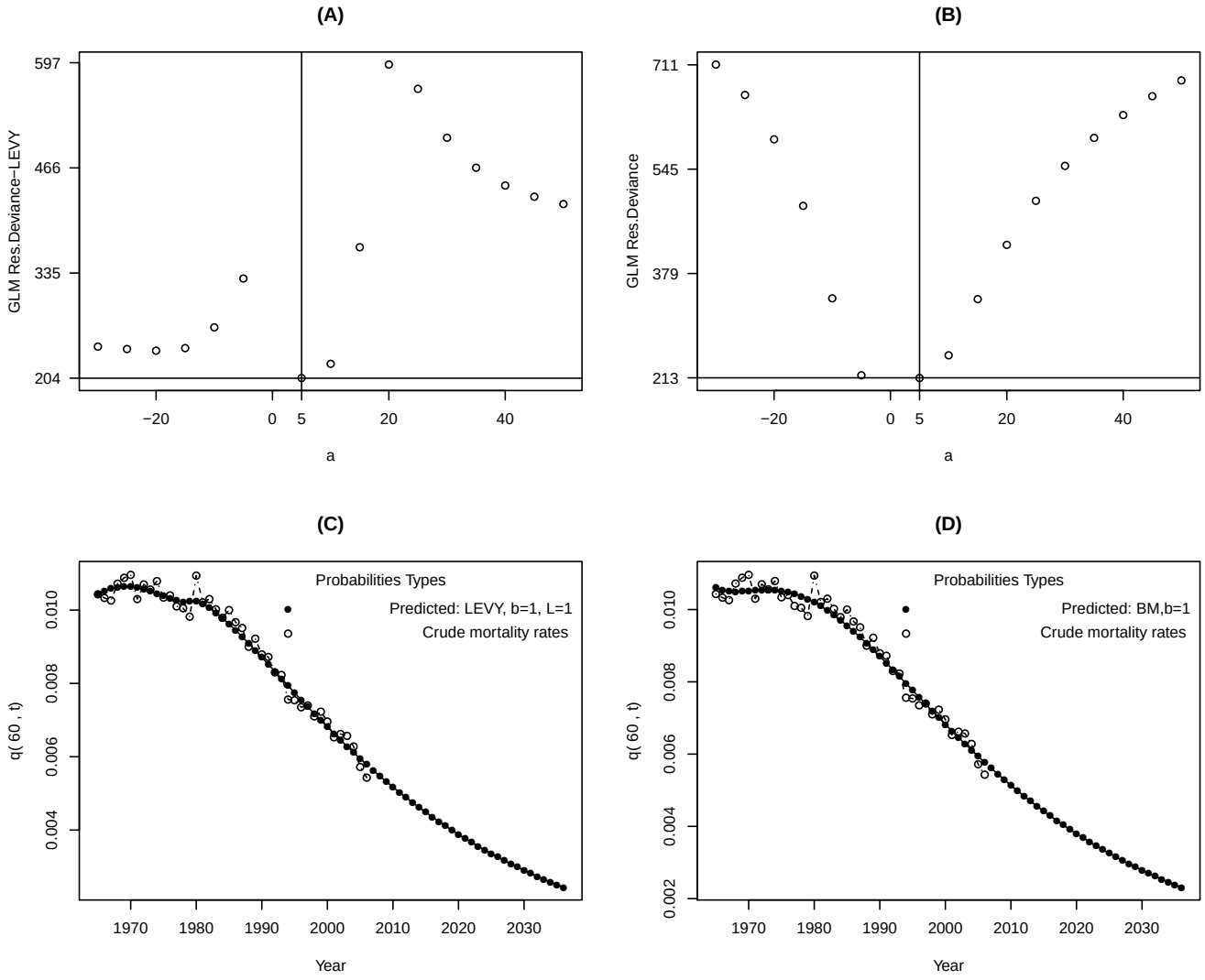

Figure 2.3: Comparison of models (1.6) and (2.19) for x=60, $\lambda = 1$, $b = 1$

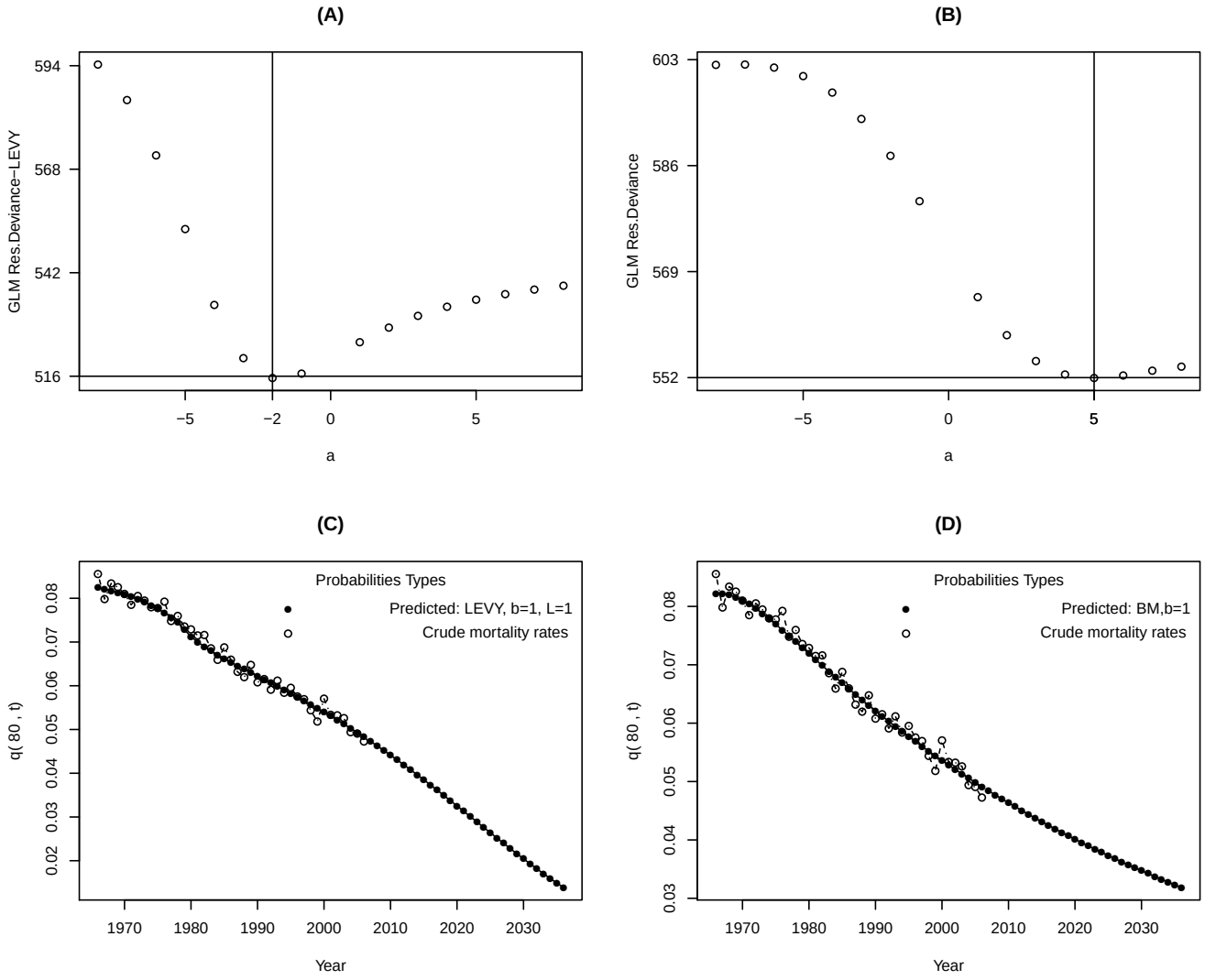

Figure 2.4: Comparison of models (1.6) and (2.19) for x=80, $\lambda = 1$, $b = 1$

(c) The Lee-Carter (LC) model can be compared with the results in this chapter. Note that the probabilities of death can be written in terms of central death rates, i.e. the quantity that the LC model forecasts.

(d) We can also consider some stochastic interest rates. Hence both the mortality and the interest return can be considered stochastic simultaneously. Then we can try to price some insurance products, such as annuities, using these dual stochastic models.

Table 2.2: Residual Deviance for Lévy and BM Case, x=20

a	**b**	**λ**	**Res.Dev.BM**	**Res.Dev.Lévy**
-4	1	1	1233.488	1166.486
-2	1	1	978.6552	1094.262
2	1	1	428.5019	921.5472
4	1	1	315.794	730.1289
6	1	1	298.4142	410.4302
8	1	1	312.3271	277.0891
10	1	1	324.013	268.8267
12	1	1	325.2171	279.5556
14	1	1	317.9034	292.3455
16	1	1	306.1661	306.1472
18	1	1	293.4437	321.0396
20	1	1	281.989	336.8455
22	1	1	273.0469	353.2447
24	1	1	267.1593	369.9094
26	1	1	264.424	386.5619
28	1	1	264.6813	402.988
30	1	1	267.6357	419.0321
32	1	1	272.933	434.5867
34	1	1	280.2069	449.5823
36	1	1	289.1054	463.9777
38	1	1	299.3047	477.7526
40	1	1	310.5159	490.9021
42	1	1	322.4865	503.432
44	1	1	334.9995	515.3556
46	1	1	347.8708	526.6914
48	1	1	360.9462	537.4612
50	1	1	374.098	547.6886

Table 2.3: Parameters of Models (1.6) and (2.19), $x = 20$

$x_1 = 15$	$x_2 = 25$	$y_1 = 1948$	$y_2 = 2006$	$b = 1$
$\lambda_1 = 1$	$a_1 = -4$	$a_2 = 50$	$r(x,t) = l(x,t)$	$N = 30$

Table 2.4: Analysis of Deviance for Lévy Model (2.19), Significance Level: '.' 0.1, '***' 0.001

Term	**Df**	**Deviance Resid.**	**Resid. Df**	**Resid. Dev.**	$\boldsymbol{P(>\lvert Chi\rvert)}$
NULL	1		648	6580.5	
$L_1(x^{'})$	1	547.5	647	6032.9	$< 2.2 \times 10^{-16}$ ***
$L_2(x^{'})$	1	50.4	646	5982.5	$< 2.2 \times 10^{-16}$ ***
$L_3(x^{'})$	1	27.3	645	5955.2	8.795×10^{-16} ***
$t^{'}$	1	3975.7	644	1979.5	$< 2.2 \times 10^{-16}$ ***
$(1 - e^{-at^{'}})$	1	962.8	643	1016.7	$< 2.2 \times 10^{-16}$ ***
$L_1(x^{'})t^{'}$	1	50.3	642	966.4	$< 2.2 \times 10^{-16}$ ***
$L_2(x^{'})t^{'}$	1	1.5	641	964.9	0.06225 .
$\zeta(a, \lambda, t^{'})$	1	696.1	640	268.8	$< 2.2 \times 10^{-16}$ ***

Table 2.5: Parameters of OU Process Including BM and Lévy Subordinator, x=50

$x_1 = 45$	$x_2 = 55$	$y_1 = 1948$	$y_2 = 2006$	$b = 1$
$\lambda_1 = -15$	$\lambda_2 = 15$	$a_1 = -10$	$a_2 = 20$	$r(x,t) = l(x,t)$

Table 2.6: Parameters of Models (1.6) and (2.19), $x = 50$

$x_1 = 45$	$x_2 = 55$	$y_1 = 1948$	$y_2 = 2006$	$b = 1$
$\lambda_1 = 5$	$a_1 = -10$	$a_2 = 20$	$r(x,t) = l(x,t)$	$N = 30$

Table 2.7: Parameters of OU Process Including BM and Lévy Subordinator, x=60

$x_1 = 57$	$x_2 = 62$	$y_1 = 1965$	$y_2 = 2006$	$b = 1$
$\lambda_1 = -20$	$\lambda_2 = 20$	$a_1 = -30$	$a_2 = 50$	$r(x,t) = l(x,t)$

Table 2.8: Parameters of OU Process Including BM and Lévy Subordinator, x=80

$x_1 = 78$	$x_2 = 82$	$y_1 = 1966$	$y_2 = 2006$	$b = 1$
$\lambda_1 = -15$	$\lambda_2 = 15$	$a_1 = -8$	$a_2 = 8$	$r(x,t) = l(x,t)$

Table 2.9: Residual Deviance for Lévy and BM Case, age $x = 60$

a	**b**	**λ**	**Res.Dev.BM**	**Res.Dev.Lévy**
-30	1	-20	711.4785	308.7345
15	1	-20	338.2158	311.6731
-30	1	-15	711.4785	299.5461
15	1	-15	338.2158	306.2938
-30	1	-10	711.4785	287.5206
15	1	-10	338.2158	297.7336
-30	1	-5	711.4785	269.5173
15	1	-5	338.2158	280.9678
-30	1	-1	711.4785	239.772
15	1	-1	338.2158	239.3899
5	1	1	212.6875	204.2213
10	1	1	248.86	221.8847
50	1	1	686.2203	420.9371
5	1	5	212.6875	213.4478
45	1	5	661.1342	510.7032
5	1	10	212.6875	212.6886
45	1	10	661.1342	550.8552
5	1	15	212.6875	212.6411
45	1	15	661.1342	576.8543
5	1	20	212.6875	212.6379
45	1	20	661.1342	597.1013
50	1	20	686.2203	554.7021

Table 2.10: Parameters of Models (1.6) and (2.19), $x = 60$

$x_1 = 57$	$x_2 = 62$	$y_1 = 1965$	$y_2 = 2006$	$b = 1$
$\lambda_1 = 1$	$a_1 = -30$	$a_2 = 50$	$r(x,t) = l(x,t)$	$N = 30$

Table 2.11: Residual Deviance for Lévy and BM Case, age $x = 80$

a	**b**	$\boldsymbol{\lambda}$	**Res.Dev.BM**	**Res.Dev.Lévy**
-8	1	-15	602.1565	603.676
-1	1	-15	580.2943	579.9368
7	1	-15	553.11	552.5092
-3	1	-10	593.485	592.1899
5	1	-10	551.9378	551.7029
-5	1	-5	600.3563	598.3264
3	1	-5	554.6568	555.5176
-7	1	-1	602.2089	596.1682
1	1	-1	564.911	567.1612
8	1	-1	553.7611	550.685
-2	1	1	587.5743	515.552
5	1	1	551.9378	535.225
-5	1	5	600.3563	589.5692
3	1	5	554.6568	553.8153
-7	1	10	602.2089	596.2432
1	1	10	564.911	564.4489
8	1	10	553.7611	555.335
-2	1	15	587.5743	588.3744
6	1	15	552.3451	552.9872

Chapter 3

Pricing Life Annuities with Lévy Processes

3.1 Introduction

In this chapter, we consider the pricing of whole life annuities due using the model developed in Chapter 2 based on Lévy subordinators. In particular, we focus on the gamma process and show how this model can provide a better fit. Similarly to Chapter 2, Generalized Linear Models are used to estimate coefficients of the explanatory variables and the gamma process. For this purpose, the coefficient of the gamma process is obtained by minimizing the residual deviance.

We use mortality data of males in Japan from 1996-2009, as well as Denmark and the U.S. over a long period of time in order to compare our results with the model proposed by Renshaw et al. (1996). As a reminder, we use RHH to abbreviate this model. Some preferences are indicated based on Akaike's information criterion, the Bayesian information criterion, the likelihood ratio test, Akaike weights, and the evidence ratio to support the proposed model. We then use a cubic smoothing spline method to fit the interest rate curve and illustrate some over(under) estimations in the prices of the annuities under the structure suggested by Renshaw et al. (1996).

For the Japan data over a short period of time, we find that the model proposed by Renshaw et al. (1996) can over estimate the liabilities. Also, our analysis shows that for U.S. mortality rates over a long period of time, the price of whole life annuities can be under-estimated for the ages above 60 based on the modeling structure in Renshaw et al. (1996). However, the model from Renshaw et al. (1996) slightly over-estimates prices for age 50. Finally, in Denmark, we observe an under-estimation of the prices for all the considered ages according to RHH model. We explain our data in the next subsection.

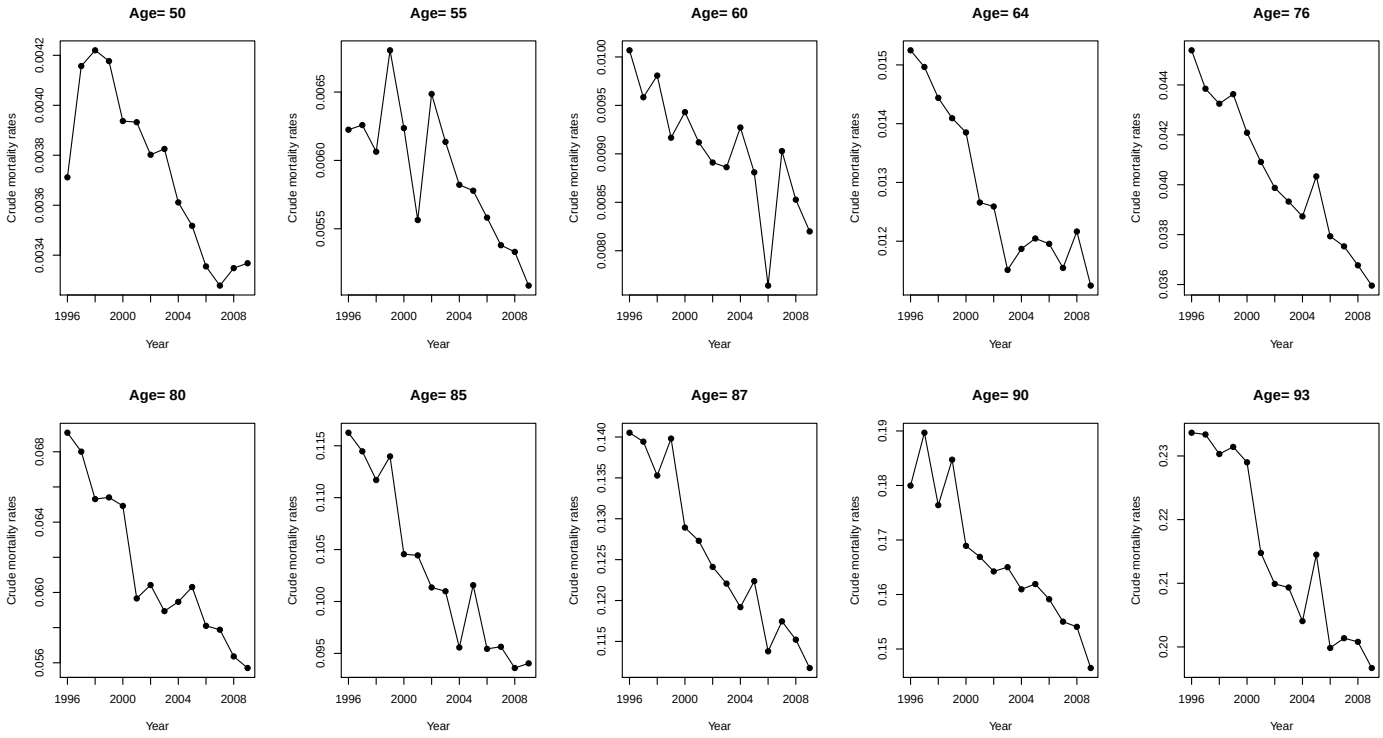

Figure 3.1: Crude mortality rates, $q(x,t)$, for Japan males during 1996-2009

3.2 Data

We use mortality rates for Japan, for males, over the period of 1996-2009 and for ages 50-100. The data sets are available at the Human Mortality Database[1]. Figure 3.1 shows the crude mortality rates for some selected ages from 1996 to 2009. In this plot, it is clear that in some particular calendar years, the observed mortality rates have had positive or negative fluctuations. For example, there is a dramatic increase in the mortality rates for age 50 during 1996-1997.

To visualize this positive jump clearly, we use a box plot for the differences in mortality rates of each consecutive year, as shown in Figure 3.2. The solid black segments indicate the median of the differences, while the lower and upper limits of the box represent the lower quartile (Q_1) and the upper quartile (Q_3), respectively. The upper (lower) whisker is drawn at the most extreme differences which is less (greater) than or equal to $Q_3 + 1.5 \times IQ$ ($Q_1 - 1.5 \times IQ$), where $IQ = Q_3 - Q_1$ is the inter-quartile. Any difference outside these whiskers can be considered as a possible outlier and is indicated with a dot. As an example, for age 50 in Figure 3.2, the dot shows that the difference in mortality rates between 1997 and 1996 is large enough to be assumed as a positive jump. We can also observe that four points have been indicated for age 90 in Figure 3.2, while there are only

[1] www.mortality.org.

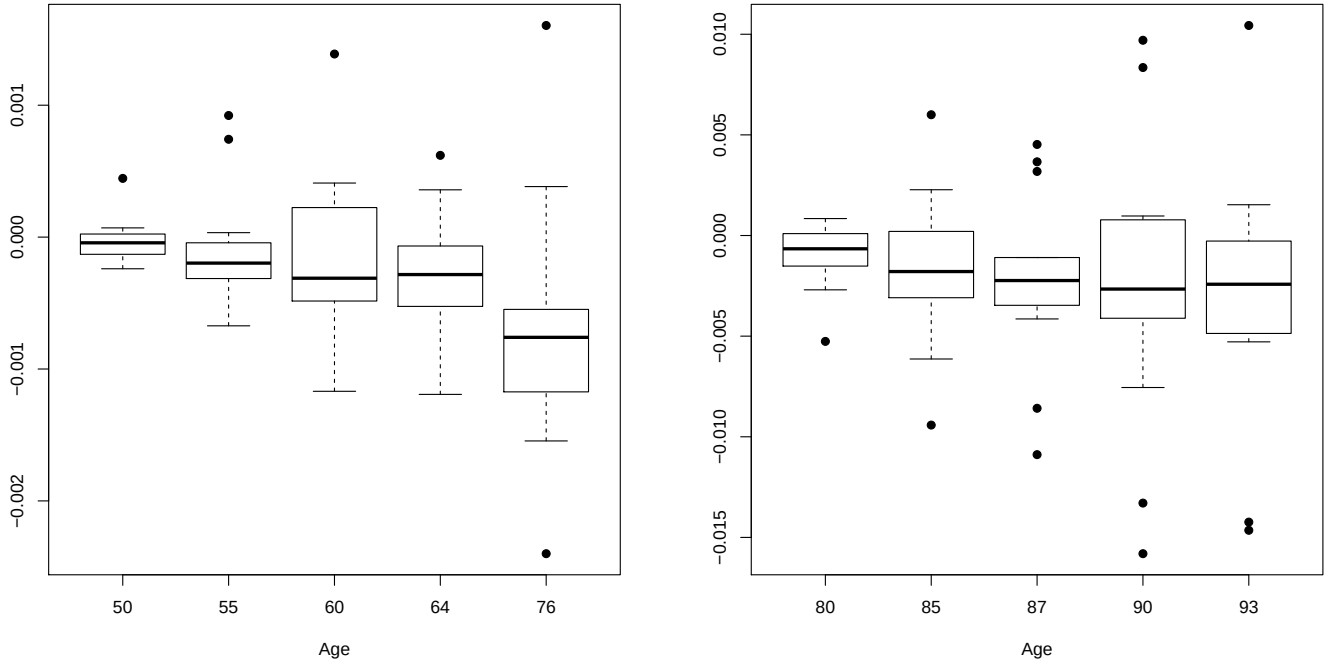

Figure 3.2: Box plot of the differences in mortality rates, $q(x, t+1) - q(x, t)$, for Japan males during 1996-2009

2 positive jumps from 1996-1997 and 1998-1999. The other two points show the corresponding declines from 1997-1998 and 1999-2000; as explained above, the consecutive differences have been taken into account to construct the boxplot. Therefore, this data set contains some outliers, so it is appropriate to fit to it the Lévy model developed in the previous chapter.

3.3 Fit of RHH Model

For the Japan data set, we first fit (1.1) and determine the optimum values of r, s, k and l by monitoring the improvements in the residual deviance. See Renshaw et al. (1996) and Sithole et al. (2000) for a complete explanation of this technique. The optimum values are $r = 2$, $s = 5$, $k = 2$, $l = 2$, in addition to the term with coefficient γ_{13}. The analysis of variance in Table 3.1 reveals that the contribution of the quadratic term in time, α_2, is not significant.

Therefore, we remove this term and fit the following model:

$$\mu(x,t) = \exp\left[\beta_0 + \sum_{j=1}^{5} \beta_j L_j(x') + \alpha_1 t' + \sum_{i=1}^{2}\sum_{j=1}^{2} \gamma_{ij} L_j(x') t'^{i} + \gamma_{13} L_3(x') t'\right]. \quad (3.1)$$

It is observed that all the remaining parameters are statistically significant. See

Table 3.1: Japan Male, Model (1.1), Significance Levels: '**' 0.01,'***' 0.001, ' ' 1

Coefficients	Estimate	Std. Error	z-value	p-value
β_0	-3.279529	0.001625	-2018.262	2×10^{-16} ***
α_1	-0.116506	0.001759	-66.246	2×10^{-16} ***
α_2	-0.001488	0.003162	-0.471	0.637896
β_1	2.353283	0.003244	725.500	2×10^{-16} ***
β_2	0.023110	0.004038	5.723	1.05×10^{-8} ***
β_3	-0.109831	0.003226	-34.046	2×10^{-16} ***
β_4	-0.064843	0.003348	-19.369	2×10^{-16} ***
β_5	0.016582	0.003542	4.681	2.85×10^{-6} ***
γ_{11}	0.016555	0.003694	4.482	7.40×10^{-6} ***
γ_{21}	0.031235	0.006053	5.160	2.46×10^{-7} ***
γ_{12}	0.032895	0.004244	7.752	9.06×10^{-15} ***
γ_{22}	-0.025159	0.007510	-3.350	0.000808 ***
γ_{13}	-0.019840	0.004829	-4.109	3.98×10^{-5} ***

Table 3.2 for the estimated parameter values, their standard deviations and the p-values in the revised model.

This is the model that fits the historical data best, regardless of any projections. As explained in Sithole et al. (2000), the difficulty is not only in finding the model that provides a good fit to the data, but also produces reasonable mortality projections. In order to price the whole life annuities accurately, we need to project mortality rates for 60 years within the range of ages 50-110 years. This means that we should project the mortality rates outside the range of ages modelled. Two important properties that the projections should possess are:

- As time increases, the predicted force of mortality should decrease for each fixed age.
- As age increases, the predicted force of mortality should increase for each fixed calendar year.

We first fitted (3.1) and found that the projected force of mortality for some of the considered ages is increasing in time. To fulfill the two properties above and to obtain reasonable projections, we revise (3.1) and find that adding quadratic coefficients for time in the multiplicative term, $L_j(x^{'})t^{'^i}$, can dramatically affect the projections. This property was mentioned in Sithole et al. (2000) and the polynomials with lower-order should be tested as well. Therefore, we remove the terms associated to γ_{21}, γ_{22}, sequentially and at each step we check the corresponding

Table 3.2: Japan Male, Revised Model (1.1), Significance Level: '***' 0.001

Coefficients	Estimate	Std. Error	z-value	p-value
β_0	-3.280096	0.001092	-3003.999	2×10^{-16} ***
α_1	-0.116490	0.001758	-66.272	2×10^{-16} ***
β_1	2.353777	0.003070	766.692	2×10^{-16} ***
β_2	0.022472	0.003805	5.906	3.51×10^{-9} ***
β_3	-0.109832	0.003226	-34.046	2×10^{-16} ***
β_4	-0.064838	0.003348	-19.368	2×10^{-16} ***
β_5	0.016581	0.003542	4.681	2.85×10^{-6} ***
γ_{11}	0.016438	0.003684	4.462	8.10×10^{-6} ***
γ_{21}	0.029943	0.005393	5.552	2.82×10^{-8} ***
γ_{12}	0.032876	0.004242	7.750	9.16×10^{-15} ***
γ_{22}	-0.023457	0.006581	-3.564	0.000365 ***
γ_{13}	-0.019888	0.004826	-4.121	3.77×10^{-5} ***

projections. The final fitted model is

$$\mu(x,t) = \exp\Big[\beta_0 + \sum_{j=1}^{5} \beta_j L_j(x^{'}) + \alpha_1 t^{'} + \sum_{j=1}^{3} \gamma_{1j} L_j(x^{'}) t^{'}\Big]. \tag{3.2}$$

Table 3.3 gives the details of a chi-square test for the reduction in residual deviance. All the remaining parameters in this model are statistically significant.

Table 3.3: Chi-Square Test, Revised Model (3.2), Significance Level: '***' 0.001

Terms	Deviance Residual	Df	Residual Deviance	p-value
Null		713	1618567	
α_1	853	712	1617714	2.2×10^{-16} ***
β_1	1614408	711	3306	2.2×10^{-16} ***
β_2	13	710	3293	0.0002846 ***
β_3	1780	709	1513	2.2×10^{-16} ***
β_4	357	708	1156	2.2×10^{-16} ***
β_5	21	707	1134	4.119×10^{-6} ***
γ_{11}	76	706	1059	2.2×10^{-16}***
γ_{12}	55	705	1003	9.685×10^{-14} ***
γ_{13}	17	704	986	3.393×10^{-5} ***

The crude mortality rates together with the predicted rates based on the model (3.2) are plotted against calendar years 1996-2009 for ages $60, 65, 70, 75, 80$ as shown in Figure 3.3. It shows that RHH model (3.2) fits the historical data well

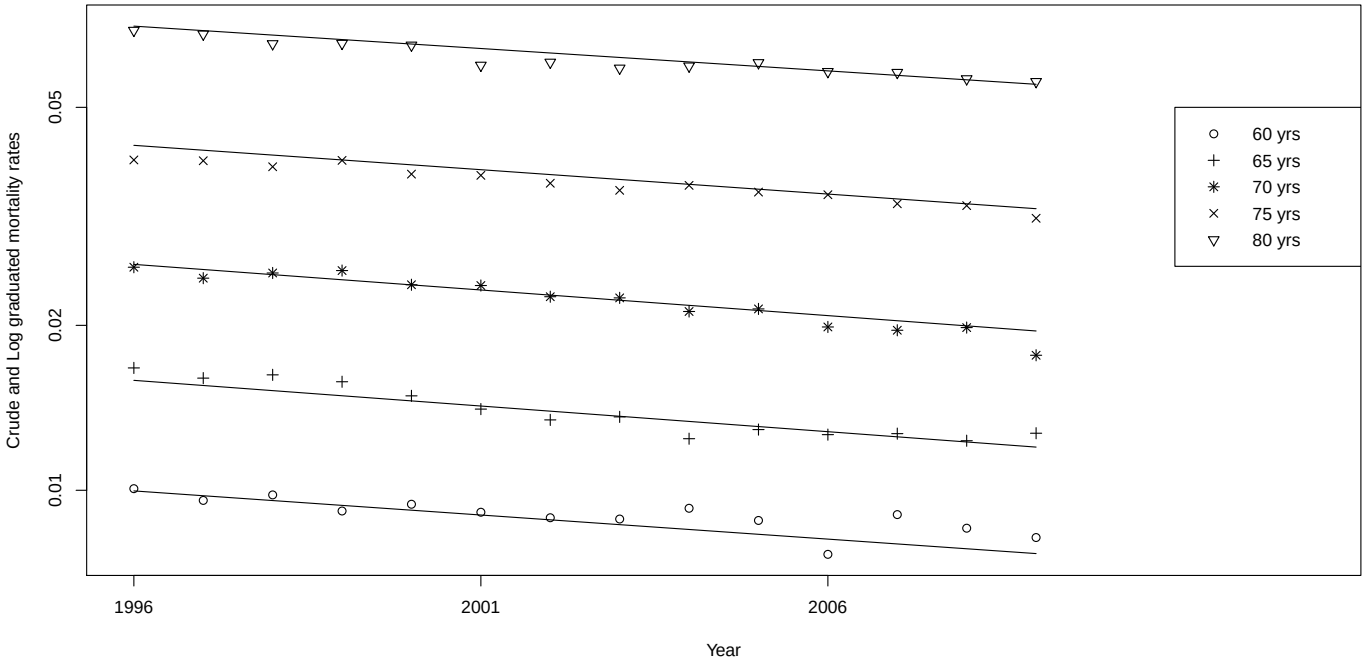

Figure 3.3: Crude mortality rates, $q(x,t)$, and predicted force of mortality, $\hat{\mu}(x,t)$, of model (3.2).

within the range of ages considered to fit the model. Figure 3.4 plots the projected force of mortality on the log scale for model (3.2) at 5-year age intervals from 1996-2079. The projections satisfy the properties mentioned above and progress at the same rate that is consistent with the historical annual mortality improvements. In the following section, we improve the obtained model and investigate the inclusion of a gamma process in the projections.

3.3.1 GLM with gamma Process

In this section, we study the effect of adding the last term in (2.19) into the RHH model (3.2). Therefore, we first exclude the term $b\delta_1(1-e^{-at})$ in (2.19) and model the force of mortality as:

$$\mu(x,t) = \exp\Big[\beta_0 \;+\; \sum_{j=1}^{5}\beta_j L_j(x^{'}) + \alpha_1 t^{'} + \sum_{i=1}^{2}\sum_{j=1}^{3}\gamma_{ij}L_j(x^{'})t^{'^i} \;-\; c\frac{dilog(\frac{\lambda-e^{-at}}{\lambda}) - dilog(\frac{\lambda-1}{\lambda})}{a}\Big]. \tag{3.3}$$

The residual deviance is minimized with respect to parameters a and λ by fitting (3.3) as shown in Figure 3.5. As a reminder, we rewrite the definition of residual

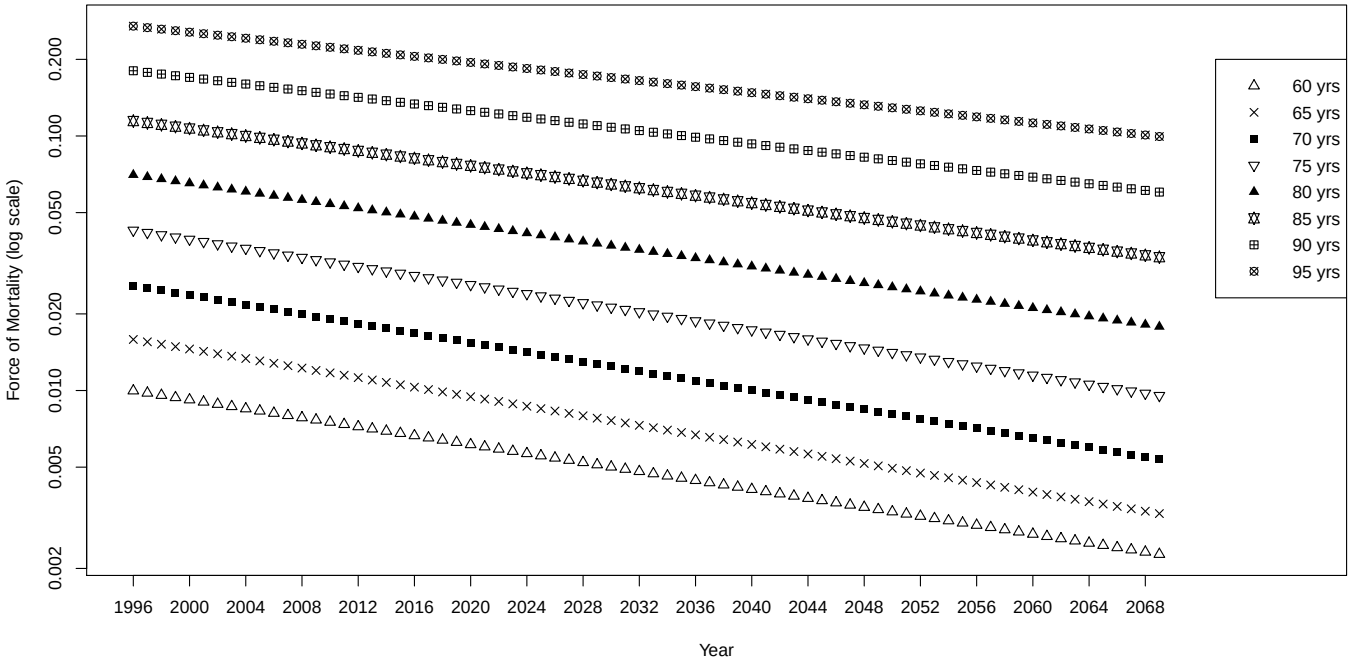

Figure 3.4: Projected force of mortality based on model (3.2)

deviance in (1.8), $R(a, \lambda)$ for the Poisson GLM as

$$R(a,\lambda) = 2\sum_x \sum_t \Big(d(x,t) \ln \Big(\frac{d(x,t)}{\hat{d}(x,t)} \Big) - [d(x,t) - \hat{d}(x,t)] \Big), \tag{3.4}$$

where $\hat{d}(x,t)$ is the predicted number of deaths at age x and time t obtained by

$$\hat{d}(x,t) = r(x,t)\hat{\mu}(x,t). \tag{3.5}$$

The minimization is carried out over the region $a \in [-80, 80]$ and $\lambda \in [-15, 15]$. We found that the minimum residual deviance is obtained at $a = 10$ and $\lambda = 15$ with the minimum value of 974.93. When we fit model (3.3), the parameter estimates for this particular pair of a and λ are given in Table 3.4.

We observe that the contribution of the stochastic term c is statistically significant. Figure 3.6 shows the predicted force of mortality and the crude rates for ages $60, 65, 70, 75, 80$. The plot of the projected forces of mortality, in the log scale at 5-year age intervals from 1996-2069, is given in Figure 3.7. We see that by adding the stochastic term, the forces of mortality progress smoothly for almost all the ages. Figure 3.7 also shows that the annual mortality improvement rates based on model (3.3) are lower than in RHH model (3.2).

We further analyze the reduction of the residual deviance by using a chi-square test for the fit of model (3.3). Table 3.5 summarizes the results. This test shows

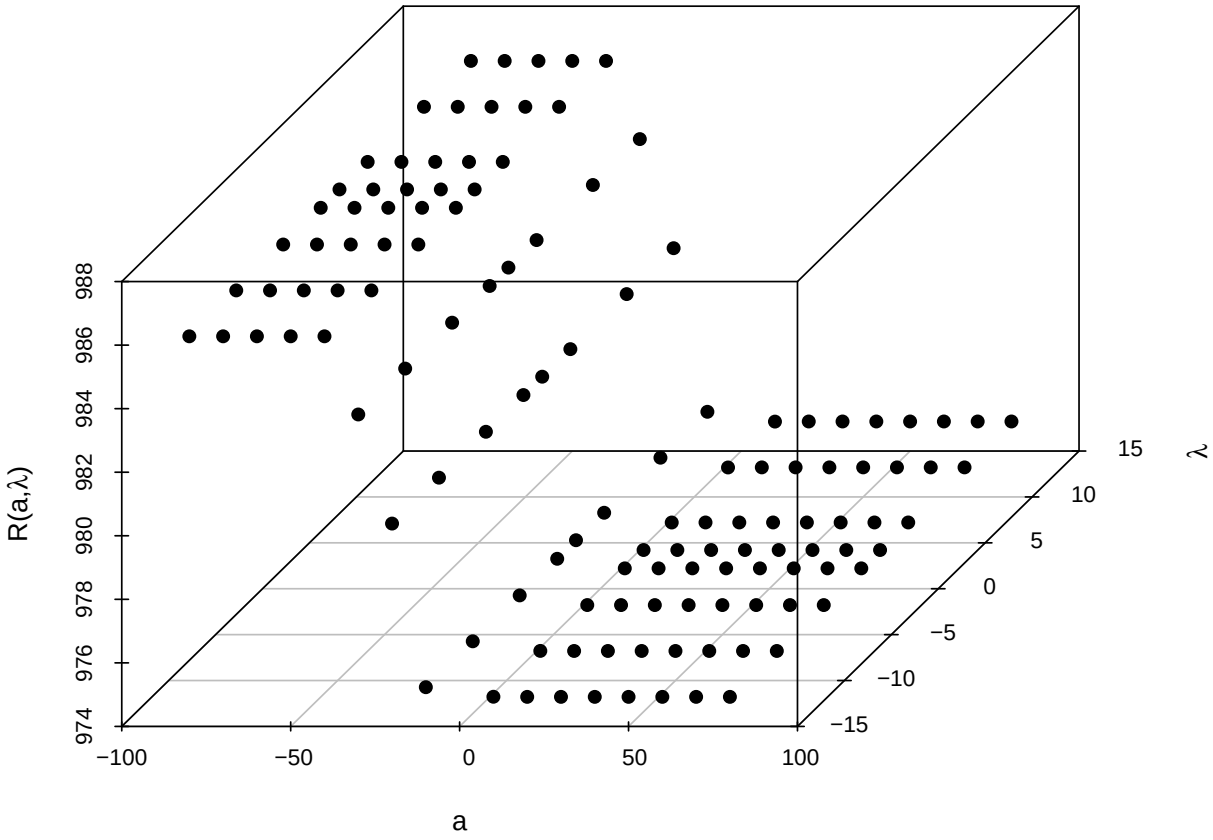

Figure 3.5: Minimization of $R(a, \lambda)$ using model (3.3)

that all the reductions in the residual deviances are statistically significant, including the term due to the gamma process.

Table 3.4: Japan Male, Model (3.3), Significance Levels: '***' 0.001, '*' 0.05

Coefficients	Estimate	Std. Error	z-value	p-value
β_0	-3.220402	0.054153	-59.468	2×10^{-16} ***
α_1	-0.334632	0.089326	-3.746	0.000180 ***
β_1	2.745631	0.039298	69.867	2×10^{-16} ***
β_2	-0.134689	0.042120	-3.198	0.001385 **
β_3	-0.426223	0.041781	-10.201	2×10^{-16} ***
β_4	-0.072493	0.013687	-5.297	1.18×10^{-7} ***
β_5	0.041257	0.008814	4.681	2.86×10^{-6} ***
γ_{11}	0.113158	0.045597	2.482	0.013075 *
γ_{12}	0.110641	0.046832	2.362	0.018153 *
γ_{13}	-0.194401	0.046609	-4.171	3.03×10^{-5} ***
c	0.053728	0.015942	3.370	0.000751 ***

Table 3.5: Chi-Square Test for Model (3.3), Significance Level: '***' 0.001

Terms	Deviance Residual	Df	Residual Deviance	p-value
Null		713	1618567	
α_1	853	712	1617714	2.2×10^{-16} ***
β_1	1614408	711	3306	2.2×10^{-16} ***
β_2	13	710	3293	0.0002846 ***
β_3	1780	709	1513	2.2×10^{-16} ***
β_4	357	708	1156	2.2×10^{-16} ***
β_5	21	707	1134	4.119×10^{-6} ***
γ_{11}	76	706	1059	2.2×10^{-16} ***
γ_{12}	55	705	1003	9.685×10^{-14} ***
γ_{13}	17	704	986	3.393×10^{-5} ***
c	11	703	975	0.0007541 ***

We test the normality assumption of the residuals for models (3.2) and (3.3) using the Shapiro-Wilk test of normality. Details of the Shapiro-Wilk's test can be found in Royston (1982a, 1982b, 1995). The p-values for RHH model (3.2) and the proposed model (3.3) are 0.2208 and 0.2391, respectively. Therefore, we cannot reject the normality assumption at a conventional significance level. We further check the residuals graphically for constant variance. We did not find any violation from this assumption as well.

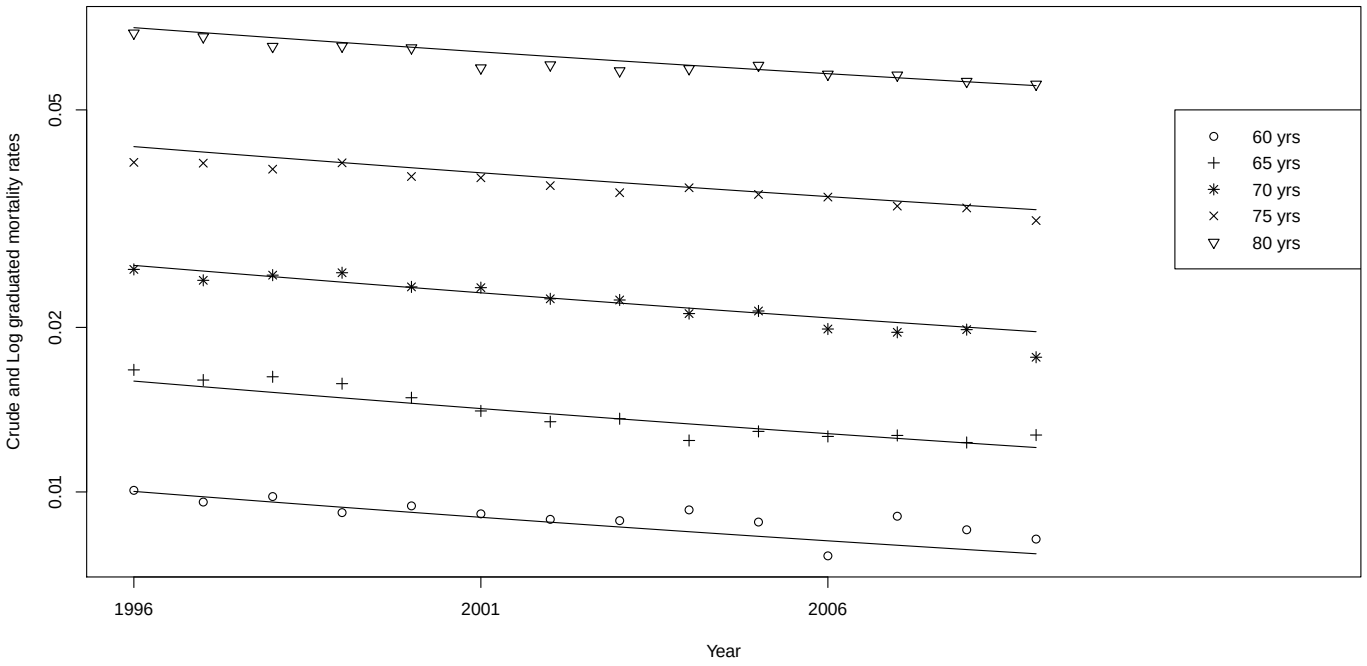

Figure 3.6: Crude mortality rates and predicted force of mortality for model (3.3), $a = 10$ and $\lambda = 15$.

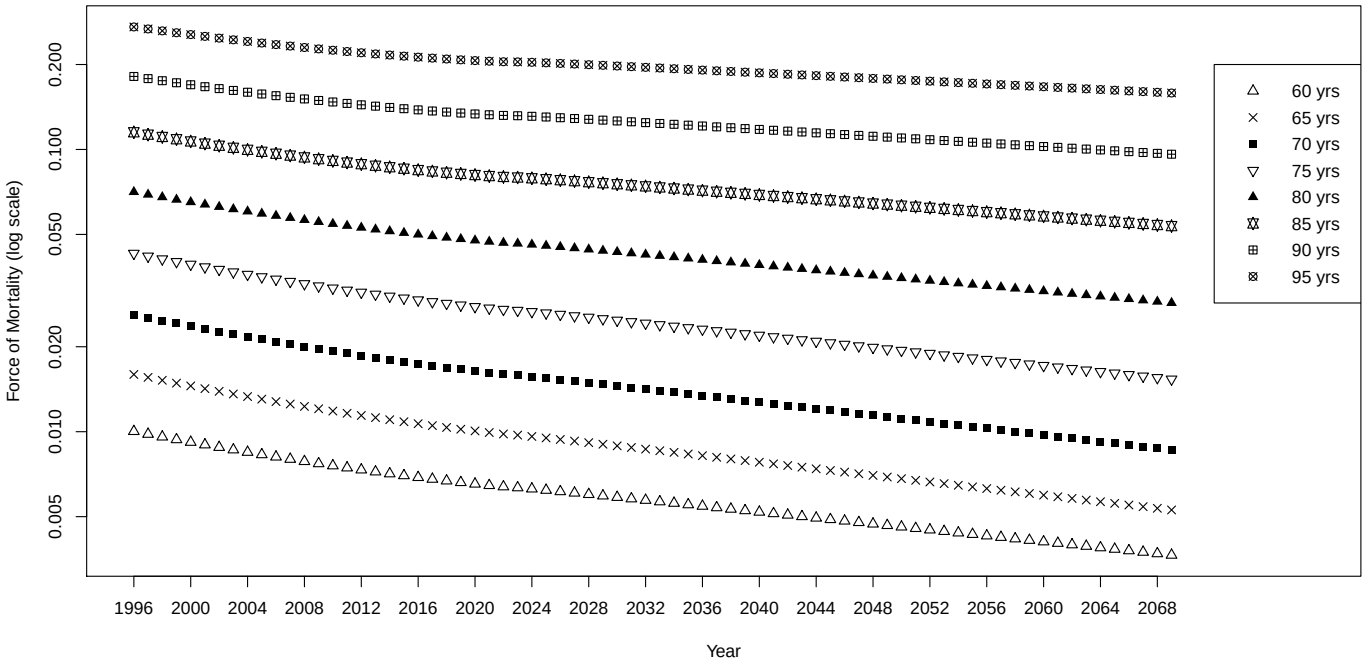

Figure 3.7: Projected force of mortality for model (3.3), $a = 10$ and $\lambda = 15$.

In addition to the Japanese data set, explained in Subsection (1.2.3), the proposed model in (2.19) and RHH model in (1.1) was fitted to the mortality data sets for the U.S.[1] and Denmark[2] as well. We use the range of ages 50-80 to fit models (3.2) and (3.3) over the period of 1910-2007 for the U.S. and 1910-2009 for Denmark. The projections was obtained for 70 years. The final fitted models based on the model suggested in Renshaw et al. (1996) for the U.S and Denmark are:

$$\mu(x,t) = \exp\Big[\beta_0 + \sum_{j=1}^{3}\beta_j L_j(x^{'}) + \sum_{i=1}^{2}\alpha_i t^{'^i} + \gamma_{11}L_1(x^{'})t^{'} + \sum_{i=1}^{2}\gamma_{i3}L_3(x^{'})t^{'^i}\Big], \tag{3.6}$$

$$\mu(x,t) = \exp\Big[\beta_0 + \sum_{j=1}^{2}\beta_j L_j(x^{'}) + \sum_{i=1}^{3}\alpha_i t^{'^i} + \gamma_{11}L_1(x^{'})t^{'} + \sum_{j=1}^{2}\gamma_{2j}L_j(x^{'})t^{'^2}\Big], \tag{3.7}$$

respectively. The corresponding proposed models using a Gamma process can be obtained by adding the last term in (2.19) into (3.6) and (3.7).

3.4 Model Comparisons

In this section, we compare the proposed model (3.3) and RHH model (3.2) by considering different criteria that includes: Akaike information criterion (AIC), Bayesian information criterion (BIC), likelihood ratio test, Akaike weights (ω_i), and evidence ratio. See Burnham and Anderson (2002) for the details of each criteria.

3.4.1 Akaike and Bayesian Information Criterion

The Akaike information criterion is the first criteria that we use to select the better model and also to check for any over parametrization. The AIC is defined as

$$AIC = -2LLF + 2NPS, \tag{3.8}$$

where LLF is the log-likelihood function and NPS is the effective number of parameters being estimated. The AIC not only rewards goodness of fit by considering

[1]Sources: Human Life-Table Database (http://www.lifetable.de) over the period of 1910-1932 and Human Mortality Database (www.mortality.org) over the period of 1933-2007.

[2]Source: Human Mortality Database.

Table 3.6: AIC and BIC for RHH Model (1.1) (R) and Proposed Model (2.19) (G)

Data	Model	AIC	BIC
Japan	R	7,520	7,566
	G	7,511	7,561
U.S.	R	43,827	43,881
	G	40,479	40,539
Denmark	R	62,829	62,884
	G	55,220	55,281

the log-likelihood function, but also includes a penalty that is an increasing function of the number of estimated parameters. The Bayesian information criterion is defined as

$$BIC = -2LLF + NPS \times \log(n), \tag{3.9}$$

where n is the number of observations. The BIC criterion is more conservative than AIC. Table 3.6 summarizes the AIC and BIC results base on the RHH model as well as the proposed model including the gamma process. There is not any evidence of over parameterization based on both AIC and BIC criteria for all three countries. The differences in the AIC values within each country are in favor of the proposed model based on the gamma process, according to the cutoff point given in Burnham and Anderson (2002). This reduction is more significant in the U.S and Denmark than Japan. Moreover, based on the BIC criteria, there is strong evidence (for Japan) and conclusive evidence (for the U.S and Denmark), for the superiority of the model (G) according to the cutoff value[1] mentioned in Raftery (1993).

3.4.2 Likelihood Ratio Test

The null hypothesis in the likelihood ratio test states that the correct model is RHH model (1.1). The alternative hypothesis includes a more complex model and states that the proposed model (2.19) is the true model. As an example for Japan, the null and alternative hypotheses suggest (3.2) and (3.3) as the correct models, respectively. Under certain regularity conditions, the test statistics, $\Lambda = 2(\hat{\ell}_{Lev} - \hat{\ell}_{Ren})$ is asymptotically chi-squared distributed with one degree of freedom, where $\hat{\ell}_{Lev}$ and $\hat{\ell}_{Ren}$ represent the maximum log likelihood for the fitted proposed model and RHH model. We reject the null hypothesis if $\Lambda > \chi^2_{(1,\alpha)}$, where $\chi^2_{(1,\alpha)}$

[1] Raftery (1993) suggests that a difference of 5 units in the BIC values indicates a "strong evidence" to support a model.

Table 3.7: Results of the Likelihood Ratio Test

Data	Λ	p-value
Japan	11.35	0.0007
The U.S.	3349.92	2.2×10^{-16}
Denmark	7611.07	2.2×10^{-16}

is the quantile of the chi-square distribution with one degree of freedom and α is the significance level. Table 3.7 contains the test statistics and the p-values. For all three countries, the null hypothesis is rejected in favor of the proposed model at $\alpha = 0.001$.

3.4.3 Akaike Weights and Evidence Ratio

The Akaike weights are easy to calculate and can be be used to assess the relative likelihood of the models (3.3) and (3.2). They are defined by:

$$\omega_i = \frac{\exp\left(-\frac{1}{2}\Delta_i\right)}{\sum_{r=1}^{2} \exp\left(-\frac{1}{2}\Delta_r\right)}, \tag{3.10}$$

where $\Delta_i = AIC_i - AIC_{min}$ is the difference of AIC for the ith model with the minimum AIC in all of the considered models. The ω_i can be easily interpreted. They show the weight of evidence in favor of model i. To explain further, first we can see from (3.10) that ω_i decreases as Δ_i increases. Also the sum of ω_i equals one and as a result we can compare them on a scale of 1. The ω_i can be interpreted as a probability that the ith model is the best among the considered models. For Japan, the Akaike weights for models (3.3) and (3.2) are $\omega_{Lev} = 0.991$ and $\omega_{Ren} = 0.009$, respectively. Based on these results, we can conclude that model (3.3) is better than RHH model (3.2) with a probability of almost 0.99, according to the Akaike weights. For the U.S and Denmark, this probability is nearly one due to the large differences in the AIC's.

In addition to Akaike weights, we can use the evidence ratio for Japan defined as $ER = \omega_{Lev}/\omega_{Ren}$. It provides some evidence to select the best model and can be interpreted as the relative value of evidence in favor of model (3.3). Here, the evidence ratio is $ER = 110$. Therefore in this case, with reasonably strong evidence, we can conclude that model (3.3) is likely to be the better model.

3.5 Pricing Life Annuities

A discrete life annuity due is a series of payments made at the beginning of equal intervals such as (months, quarters, years) while the beneficiary of initial age x

survives. These financial products play a major role in pension systems and are often purchased as part of a retirement plan to ensure income during the retirement years. Moreover, annuities have a role in disability and workers' compensation insurances as well. First, we state the formal definition of a whole life annuity due from Bowers et al. (1997, Chapter 5). A whole life annuity is a type of annuity that pays a unit amount at the beginning of each year that the annuitant age x survives. The present value of future payments is denoted by

$$Y = \sum_{j=0}^{K(x,t)} \nu(j), \tag{3.11}$$

where the random variable $K(x,t)$ is the curtate-future lifetime of age x at time t and $\nu(j) = (1+i)^{-j}$ is the discount factor, i.e. the present value of \$1 payable at time j. We use the traditional actuarial approach known as the *equivalence principle* to price the annuity. This is based on the fact that the insurance company providing annuities can diversify its mortality risk within a large portfolio of annuitants.

Let $\ddot{a}_{x,t}$ be the net single premium for a whole life annuity due evaluated at year t, based on the projected values in model (3.3). Moreover, we denote by $q_{x,t}$ the probability that an individual with age x at time t will die between t and $t+1$. The corresponding one year survival probability is defined by $p_{x,t} = 1 - q_{x,t}$. In order to obtain $q_{x,t}$, we assume that the force of mortality remains constant over each calendar year and over each integer age, i.e. $\mu(x,t) = \mu(x+u, t+s)$, $\forall\, x, t \in \mathbb{N}$ and $0 \leq u, s < 1$. As a result, we can transform the force of mortalities in (3.2) and (3.3) into $q_{x,t}$ by:

$$q_{x,t} = 1 - \exp[-\mu(x,t)]. \tag{3.12}$$

The fair price or the net single premium ($\ddot{a}_{x,t}$) without considering any expenses incurred by the insurance company, is in fact the actuarial present value of the annuity payments and is obtained by:

$$\ddot{a}^{M}_{x,t} = \mathbb{E}[Y] = \sum_{k=0}^{\infty} \sum_{j=0}^{k} \nu(j)\, {}_{k}p_{x,t}\; q_{x+k,t+k}, \tag{3.13}$$

where ${}_{k}p_{x,t}$ is the k year survival probability of a life aged x at year t defined by:

$${}_{k}p_{x,t} = \prod_{j=0}^{k-1} p_{x+j,t+j}, \quad k = 1, 2, \ldots \quad . \tag{3.14}$$

The survival probabilities in (3.14) can be recursively estimated by applying (3.12) together with the projected force of mortalities in RHH model (3.2) as well as the

proposed stochastic model (3.3). The pricing formula in (3.13) can also be obtained by using the current life tables for Japan and Denmark along with the U.S life table. Therefore, in this case (3.13) becomes:

$$\ddot{a}^{T}_{x,t} = \sum_{k=0}^{\infty} \sum_{j=0}^{k} \nu(j) \prod_{i=0}^{k-1} p_{x+i,t}\ q_{x+k,t}. \tag{3.15}$$

Both actuarial present values calculated in (3.13) and (3.15) use the same bond market assumption but different mortalities tables. In other words, (3.13) is based on a model table $q_{x+k,t+k}$ for $k = 0, ...$ and (3.15) is based on one of three current tables $q_{x+k,t}$ for $k = 0, ...$ where $t = 2007$ for the U.S. and $t = 2009$ for the Japan and Denmark. The M in (3.13) stands for model and the T in (3.15) is for current table. See Pitacco et al. (2009, Chapter 4, Subsection 4.4) for the details of using the projected tables. To evaluate the prices in (3.13) and (3.15), we assume that $\nu(j) = P(0, j)$, where $P(0, j)$ is the current price of a zero coupon bond that pays \$1 at maturity j. Therefore, the annual price of zero coupon bonds, $P(0, t)$, should be first estimated to find (3.13) and (3.15). However, the zero coupon bond rates are normally available at specific maturities. Therefore, to predict the yield curve for a zero coupon bond, we first obtain the Japanese government bond rates on March 4$^{\text{th}}$, 2012 from Bloomberg[1] as reported in Table 3.8. Then we use a cubic smoothing spline method and fit the yield curve (see Chambers and Hastie (1993) for the details). The yield rates for the U.S.[2] and Denmark[3] were obtained on June 12, 2012 and July 3th 2012, respectively.

Table 3.8: Japanese Government Bond Rates on March 4$^{\text{th}}$, 2012

Maturity (Years)	0.25	0.5	1	2	3	4	5	6	7	8	9	10	15	20	30
Yield (%)	0.10	0.10	0.11	0.11	0.14	0.21	0.30	0.41	0.56	0.67	0.82	0.99	1.43	1.76	1.95

We then evaluate the prices with three different approaches for each country. The first one is based on the current life tables according to (3.15) for $x = 55, 60, ..., 100$. The other present values are derived using (3.13) with $t = 2012$ where the projected rates are obtained based on the RHH model (3.2) and our proposed model (3.3). Since the first summation in (3.13) is not finite, we set a convergence tolerance of 1% to find the maximum age attained and evaluate the prices. In all three methods, we use the corresponding fitted yield curve for each country to obtain the price of a zero coupon bond, as mentioned above. Table

[1] http://www.bloomberg.com/markets/rates-bonds/government-bonds/japan/

[2] http://www.bloomberg.com/markets/rates-bonds/government-bonds/us/

[3] http://www.forexpros.com/rates-bonds/denmark-government-bonds?maturity_from=1&maturity_to=29

Table 3.9: $(\ddot{a}_{x,2012})$ prices for the Current Life Table and Percentage Changes for RHH Model (1.1) (R-H) and our Proposed Model (2.19) (G)

Data	Model	Age									
		55	**60**	**65**	**70**	**75**	**80**	**85**	**90**	**95**	**100**
Japan	Life table 2009 ($)	22.3	19.6	16.8	14.1	11.3	8.8	6.6	4.9	3.7	2.8
	R-H (%)	6	6	6	6	6	8	9	12	14	14
	G (%)	3	3	3	3	4	5	6	10	14	14
U.S.	Life table 2007 ($)	19.1	17.0	14.8	12.5	10.3	8.1	6.2	4.6	3.5	2.7
	R-H (%)	5	5	4	5	4	5	6	9	6	0
	G (%)	3	7	9	11	13	14	16	17	14	7
Denmark	Life table 2009 ($)	20.3	17.7	15.0	12.3	9.7	7.4	5.5	4.0	3.0	2.4
	R-H (%)	12	19	20	19	22	27	33	38	37	29
	G (%)	33	31	30	30	33	36	42	48	47	38

3.9 includes the prices based on (3.15) for the current life tables and the relative percentage changes in the prices, for RHH model (3.2) and our proposed model (3.3), defined by:

$$100 \times \frac{\ddot{a}^{M}_{x,2012} - \ddot{a}^{T}_{x,t}}{\ddot{a}^{T}_{x,t}}, \tag{3.16}$$

where $t = 2007$ for the U.S. and $t = 2009$ for Japan and Denmark. It is apparent that for all three countries and within each considered model, the prices gradually decrease as the age increases. This is due to the fact that younger annuitants tend to live longer than older people and therefore should pay more.

Age by age comparisons of the Japanese percentage changes shows that the RHH model over-estimates the prices by almost 3% for the ages 50 to 85 compared to the proposed model. This over-estimation will gradually decline as no differences can be observed in the prices of two models for the ages above 95.

Unlike Japan, we can see a different trend in the U.S. as the prices based on RHH model show an increase of 5 % at ages 55 and 60. However, the fair prices based on the proposed model rise by 3 % and 7% at ages 55 and 60, respectively. In other words, an over-estimation of 2% can be seen at age 55 according to the RHH model. From age 60 to 90, RHH model under-estimates the prices with an increasing trend starting from 2% at age 60 with a peak of 10% at age 85. The RHH model shows an improvement of 9% at age 90 and will reach the same price of $ 2.7 as provided by the U.S. Life table (2007) at age 100. The proposed model indicates an increase of 17% at age 90 with a reduction of 7% in the prices at age 100.

For Denmark, RHH model will under-estimate the prices for all the considered ages with a peak at age 55. The proposed model provides an increase of 33% in the annuity value at age 55 while RHH model suggests an increase of 12%. This

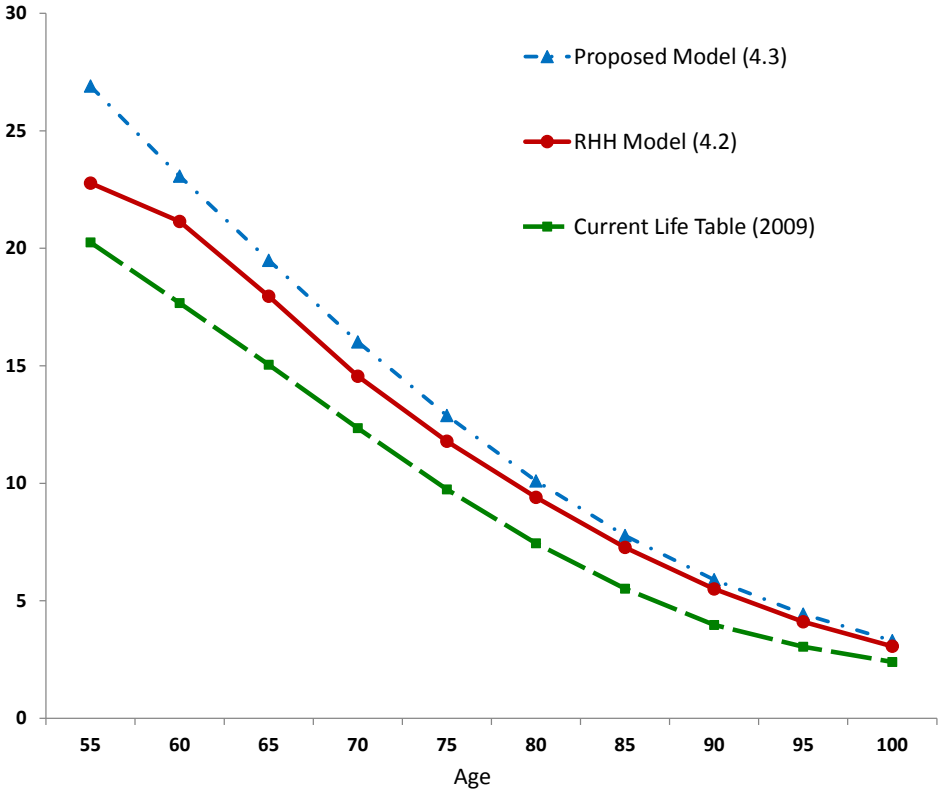

Figure 3.8: Price of whole life annuity due, Denmark males in 2012, proposed model (3.3), RHH model (3.2) & Life Table in 2009.

shows how modeling the mortality rate with the gamma process can significantly affect the annuity prices. We also see some large percentage changes in Denmark compared to Japan and the U.S. This means that the annual rates of mortality improvement in Denmark based on the fitted model are different from those in Japan and the U.S. Moreover, the highest percentage increase for both models with respect to the prices based on the current life table, occurs at age 90 with rises 38% and 48% for the RHH model and our proposed model, respectively.

We illustrate the resulting prices for Denmark based on three different models in Figure 3.8. The dash-dot line are the prices (triangles) based on the proposed model (3.3) evaluated at ages 55 to 100. Also, we indicate the prices (dots) according to RHH model (3.2) with the solid line. The long-dashed line represents the prices (squares) based on the current life table of Denmark in 2009. We observe that the under-estimation of the prices in RHH model (3.2) is higher at younger ages than older ages and this under-estimation gradually decreases towards age 100.

3.6 Conclusions

In this chapter, we show how to include a perturbation term into the modeling structure suggested by Renshaw et al. (1996). In particular we focus on a gamma process, as a special case of α-stable Lévy subordinators to express the force of mortality as a linear function in terms of regression covariates. The proposed stochastic model provides a better fit based on criteria like AIC, BIC, likelihood ratio test, Akaike weights and evidence ratio. Our analysis also indicates that in Japan with data over a short period of time, the projected rates based on the suggested model progress more smoothly than the model introduced by Renshaw et al. (1996).

We also study the pricing of whole life annuities based on three different mortality data under a deterministic interest rate. Our numerical analysis shows how RHH model can under/over-estimate the annuity prices, depending on the considered mortality data.

The conclusions given in this chapter are based on the assumption that the model that provides a better fit can lead to more reliable prices. Unfortunately, this assumption is not always true and there maybe some models that provide a good fit but are not useful for the purpose of pricing, due to their low power in projecting mortality rates. Therefore, comparing the projected rates and their consistency with historical data should be also considered when pricing the annuities.

Chapter 4

Two-Factor Stochastic Mortality Modeling

4.1 Introduction

The results in this chapter are given in Ahmadi and Gaillardetz (2014). Here, we reconsider the two-factor stochastic mortality model introduced by Cairns, Blake and Dowd (2006) (CBD). The error terms in the CBD model are assumed to form a two-dimensional random walk. We first use the Doornik and Hansen (2008) multivariate normality test to show that the underlying normality assumption does not hold for the considered data set. Ainou (2011) proposed independent univariate normal inverse Gaussian Lévy processes to model the error terms in the CBD model. We generalize this idea by introducing a possible dependency between the 2-dimensional random variables, using a bivariate Generalized Hyperbolic distribution. We propose four non-Gaussian, fat-tailed distributions: Student's t, normal inverse Gaussian, hyperbolic and generalized hyperbolic distributions. Our empirical analysis shows some preferences for using the new suggested model, based on Akaike's information criterion, the Bayesian information criterion and likelihood ratio test, as our in-sample model selection criteria, as well as the mean absolute percentage error for our out-of-sample projection errors. We also develop an alternative representation for the multivariate normal inverse Gaussian process in Appendix A.3.

As mentioned in Cairns et al. (2006), there are 3 different mortality risks for insurance companies that offer annuities and life insurance products. We summarize them here:

- *Mortality risk:* that means fluctuations of mortality rates over time.
- *Longevity risk:* which can be considered as any randomness in the long-term trend of mortality rates.

- *Short-term, catastrophic mortality risk:* that can be explained by any sudden phenomena like the influenza pandemic in 1918, the tsunami of December 2004 in Indonesia and of 2011 in Japan.

We model longevity risk according to Cairns et al. (2006) that includes two stochastic factors. They propose a bivariate normal distribution to model the dynamic of the stochastic factors. However, the error terms of the CBD model seem to have tails thicker than those of a normal distribution, as we show in our empirical analysis. Therefore, we consider a family of bivariate generalized hyperbolic (GH) distributions to model the error terms in the CBD model. This family of distributions has semi-heavy tails with a dependence structure. We use this desirable property to model longevity risk. Thus, we suggest using the bivariate generalized hyperbolic (GH) distribution in the CBD model.

Specifically, we consider four non-Gaussian distributions within the GH class, which include Student's t, normal inverse Gaussian, hyperbolic and generalized hyperbolic distributions. Next, in order to compare our model with the CBD model, we use the likelihood ratio test, Akaike's information criterion (AIC) and the Bayesian information criterion (BIC) as our in-sample model selection criteria. In addition, for the out-of-sample performance, we project mortality rates and apply the mean absolute percentage error to the proposed model in order to indicate some preferences.

4.2 Review of the CBD Model

In this section, we summarize the two-factor stochastic mortality model proposed by Cairns et al. (2006). The realization of the one-year survival probabilities for the cohort aged x and still alive at time t is denoted by $\tilde{p}(t,x)$. Furthermore, the realized mortality rate is defined by $\tilde{q}(t,x) = 1 - \tilde{p}(t,x)$. For their empirical analysis, they choose the following model for the mortality curve:

$$\tilde{q}(t,x) = \frac{e^{A^1_{t+1}+A^2_{t+1}(x)}}{1+e^{A^1_{t+1}+A^2_{t+1}(x)}}, \tag{4.1}$$

where they assume that $\boldsymbol{A}_t = \left(A^1_t, A^2_t\right)'$ is a two-dimensional random walk with drift. To make forecasts of the future distribution for $\boldsymbol{A}_t$, they propose to model the factors in $\boldsymbol{A}_t$ according to,

$$\boldsymbol{A}_{t+1} = \boldsymbol{A}_t + \boldsymbol{\mu} + C\boldsymbol{Z}_{t+1}, \tag{4.2}$$

where $\boldsymbol{\mu}$ is a constant 2×1 vector, C is a constant 2×2 upper triangular matrix and $\boldsymbol{Z}_t$ is a two-dimensional standard normal random variable.

Similarly to Cairns et al. (2006), we use an ordinary least square method to estimate A_t. Then, we change the distributional assumption in (4.2); next we apply the maximum likelihood method to estimate the dynamic properties of the two factors.

The first contribution of this chapter is that there is a historical data set for which the normal assumption (4.2) does not hold and should be tested before making any further inference. We also propose a GH distribution to use in the CBD model and show that it can provide a better fit. First we describe the data set that we use for our empirical illustration. We rely on mortality data for Males in Italy, 1969-2008 and ages 60 to 90. The data set was obtained from the Human Mortality Database (HMD[1]). We deliberately consider more recent data, since this period of time exhibit fewer levels of uncertainty compared to the mortality data for the first half or two-thirds of the 20th century. This allows us to have data with less volatility and develop our model accordingly.

4.3 Doornik-Hansen's Multivariate Normality Test

In order to estimate the mean and the variance-covariance matrix in (4.2), we first need to estimate $\boldsymbol{A}_t$. To do so, the ungraduated mortality rates for each t are transformed from $\tilde{q}(t,x)$ to

$$\log\Big(\frac{\tilde{q}(t,x)}{\tilde{p}(t,x)}\Big) = A^1_{t+1} + A^2_{t+1}(x), \tag{4.3}$$

where x ranges from 60-90 and t covers 1969-2008. Then, the linear regression is applied in (4.3) to estimate $\boldsymbol{A}_t$. It is clear from (4.2) that

$$\begin{aligned} \mathbb{E}[\boldsymbol{A}_{t+1} - \boldsymbol{A}_t] &= \boldsymbol{\mu}, \\ \mathbb{V}ar[\boldsymbol{A}_{t+1} - \boldsymbol{A}_t] &= CC'. \end{aligned} \tag{4.4}$$

Equation (4.4) shows that the mean and the variance of the first consecutive differences, $\boldsymbol{A}_{t+1} - \boldsymbol{A}_t$, can be used to estimate $\boldsymbol{\mu}$ and $V = CC'$, respectively. The estimations are given in Table 4.1. Generally, the negative value for μ_1 shows mortality improvement. At the same time, the positive value for μ_2 indicates that mortality rates at higher ages are improving at a slower rate. These results are consistent with those of Cairns et al. (2006) in the original CBD model. Before explaining the Doornik and Hansen normality test, we need to check our data (i.e. the increments $\boldsymbol{A}_{t+1} - \boldsymbol{A}_t$) for any serial autocorrelation, dependency as well as conditional heteroscedasticity (non-constant variance). These are requirements for the hypothesis tests that we use in this chapter.

[1] The data are available online at www.mortality.org.

Table 4.1: Estimated Mean and Variance Matrices for the CBD Model in (4.4)

Mortality data	$\hat{\boldsymbol{\mu}}$	$\hat{V}$
Italy	$\begin{bmatrix} -0.056037813 \\ 0.000529807 \end{bmatrix}$	$\begin{bmatrix} 0.0524230293 & -0.0006519287 \\ -0.0001527592 & 0.000002196178 \end{bmatrix}$

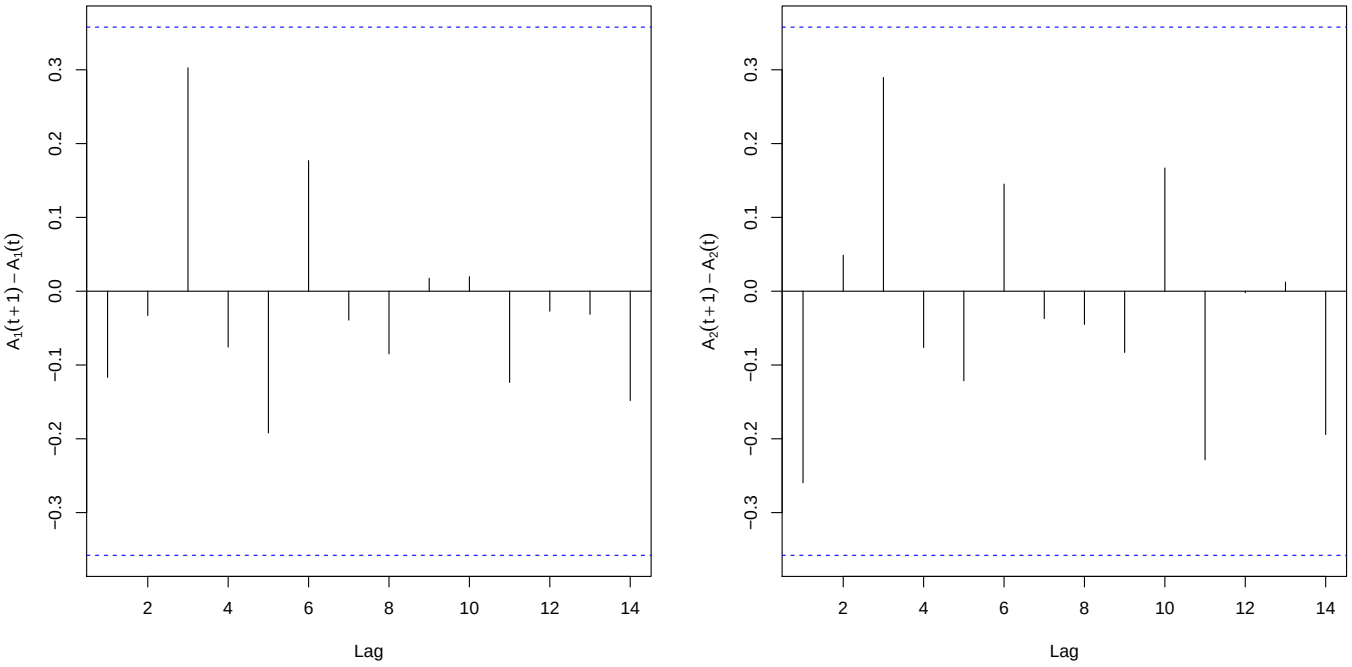

Figure 4.1: Sample ACF Plot for Italy-Data: 1969-1999

In order to test the serial autocorrelation, we obtain the sample autocorrelation function (ACF), r_k at lag k defined by:

$$r_k = \frac{\sum_{t=k+1}^{n}(Y_t - \bar{Y})(Y_{t-k} - \bar{Y})}{\sum_{t=k+1}^{n}(Y_t - \bar{Y})^2}, \; k = 1, 2, ..., \tag{4.5}$$

where $Y_t = A^i_{t+1} - A^i_t, \; i = 1, 2$. Figure 4.1 shows the sample ACF of $A^1_{t+1} - A^1_t$ and $A^2_{t+1} - A^2_t$ with the 95 % confidence limits (dotted lines). Looking at this figure, we can see that the sample ACF's are not statistically significant. Therefore, there is no indication of serially autocorrelated increments. The Ljung-Box test (Ljung and Box (1978)) can be applied to test the independence of the $A_{t+1} - A_t$ values. The test statistics is defined as:

$$Q(k) = n(n+2)\sum_{m=1}^{k}\frac{\hat{r}_k^2}{n-m}, \tag{4.6}$$

where n is the number of observation and $\hat{r}_k$ is the estimated sample ACF defined in (4.5). The null hypothesis is the linear independence in $A_{t+1} - A_t$. Under the null hypothesis, $Q(k)$ has an asymptotic chi-squared distribution with k degrees of freedom. The null hypothesis is rejected when the value of $Q(k)$ is greater than the selected critical value of chi-squared distribution with k degrees of freedom. Table 4.2 shows the value of the test statistic, $Q(k)$, the degrees of freedom, *df* and the p-value of the Ljung-Box test. Therefore, we cannot reject the null hypothesis of linear independence at a significance level of 0.05. We also check the increments

Table 4.2: Ljung-Box Test of Independency

Data	$Q(k)$	df	p-value
$A^1_{t+1} - A^1_t$	15.1886	24	0.9151
$A^2_{t+1} - A^2_t$	19.9168	24	0.7015

for the assumption of identical independent distribution (iid) by applying the test proposed by McLeod and Li (1983). The null hypothesis assumes that the data are iid. The test statistic is similar to Ljung-Box test (4.6), except that for the sample ACF, $\hat{r}_k$, which is replaced by the sample autocorrelations of the squared data. This is due to the fact that if the data are iid then square of the data are iid as well. The McLeod-Li test (McLeod and Li (1983)) can also be used to assess the conditional heteroscedasticity of the data. The test statistic is chi-square distributed with k degrees of freedom under the null hypothesis of iid. Figure 4.2 shows the p-values of the McLeod-Li test, evaluated up to lag 14. The dashed line represents the 5% confidence level. Based on this test, we cannot reject the null hypothesis of identical independent distribution. We can now test the bivariate normality assumption in CBD model by using the method, proposed by Doornik and Hansen (2008). The latter compares their suggested test with four other tests for multivariate normality and concludes that it has the best size and power properties over the other tests considered. This test is relatively simple, it controls most sizes well and can be applied for samples as low as 10 observations. The Doornik-Hansen's test for multivariate normality is based on the skewness and kurtosis of multivariate data that is transformed to ensure independence, as explained in Appendix A.2. The package and corresponding R functions that we used to perform the Doornik-Hansen's test are *asbio* and *DH.test*, respectively. The results are reported in Table 4.3.

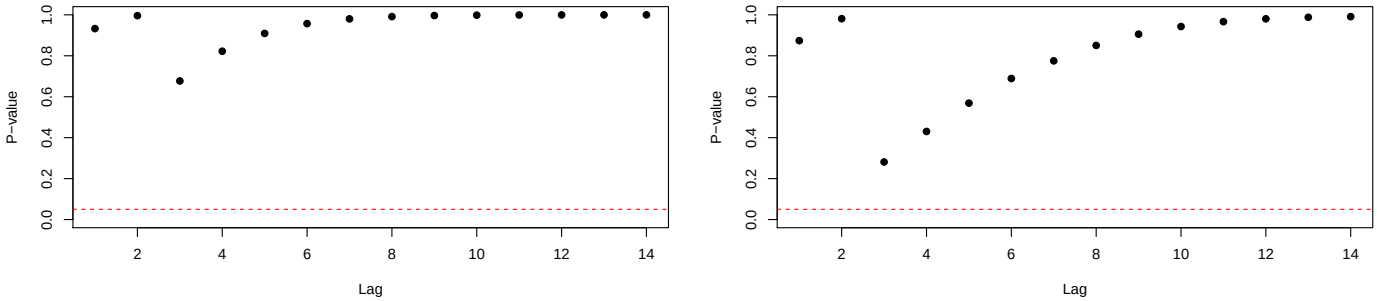

Figure 4.2: P-values of the McLeod-Li test for $A^1_{t+1} - A^1_t$ (left) and $A^2_{t+1} - A^2_t$ (right), Italy data

The multivariate section of Table 4.3 indicates that the test statistic is significant and based on the p-value of the test, the bivariate normality assumption is rejected at a significance level of 0.05. The table also contains the univariate tests for normality. The univariate normality assumption is rejected at the 5% significance level. Consequently, the multivariate normality assumption does not hold. Moreover, we applied a Shapiro-Wilk test for multivariate data sets using

Table 4.3: Doornik-Hansen Normality Test

Mortality data for Italy	Test statistics	df	p-value
Multivariate	25.20	4	4.6×10^{-4}
Univariate	19.74	2	5.2×10^{-4}
	5.46	2	0.065

the R package *mvnormtest*. Details of Shapiro-Wilk's test can be found in Royston (1982a, b, 1995). Table 4.4 summarizes the results. The normality assumption is rejected based on this multivariate test as well. Here, we address the importance of the measurement error, as explained in Cannon (2010). It originates from the fact that the CBD methodology first estimates the factors, then it analyses their dynamic properties. This may affect the statistical results given in this section.

To propose an appropriate model, we need a more flexible class of distributions, here we suggest the generalized hyperbolic distributions. We briefly review this class in the next section.

Table 4.4: Multivariate Shapiro-Wilk Normality Test

Mortality data	Test statistic	p-value
Italy	0.7629	1.48×10^{-4}

4.4 Generalized Hyperbolic Distributions

In this section, we review the definition of the family of generalized hyperbolic (GH) distributions of McNeil et al. (Chapter 3, 2005).
The random vector $\mathbf{X}$ is said to have a d-dimensional GH distribution with parameters $(\lambda, \chi, \psi, \boldsymbol{\mu}, \Sigma, \boldsymbol{\gamma})$, denoted as $\mathbf{X} \sim GH(\lambda, \chi, \psi, \boldsymbol{\mu}, \Sigma, \boldsymbol{\gamma})$ if

$$\mathbf{X} \stackrel{\mathrm{d}}{=} \boldsymbol{\mu} + W\boldsymbol{\gamma} + \sqrt{W}A\mathbf{Z}, \tag{4.7}$$

where $\stackrel{\mathrm{d}}{=}$ denotes equality in distribution and

(i) $\mathbf{Z} \sim N_k(\mathbf{0}, I_k)$,

(ii) $A \in \mathbb{R}^{d\times k}$,

(iii) $\boldsymbol{\mu}, \boldsymbol{\gamma} \in \mathbb{R}^d$,

(iv) $W \geq 0$ is a scalar-valued random variable which is independent of $\mathbf{Z}$ and has a generalized inverse Gaussian distribution, denoted GIG(λ, χ, ψ). (See Appendix A.3 for details.)

The joint density function of the GH distribution in the non-singular case (Σ has rank d) is

$$\begin{aligned} f_{\boldsymbol{X}}(\boldsymbol{x}) &= \int_0^\infty f_{\boldsymbol{X}|W}(\boldsymbol{x}|w) f_W(w) dw \\ &= \int_0^\infty \frac{e^{(\boldsymbol{x}-\boldsymbol{\mu})'\Sigma^{-1}\gamma}}{(2\pi)^{\frac{d}{2}}|\Sigma|^{\frac{1}{2}}} \exp\Big\{ -\frac{Q(\boldsymbol{x})}{2w} - \frac{\gamma'\Sigma^{-1}\gamma}{2/w} \Big\} f_W(w) dw \\ &= c \times \frac{K_{\lambda-\frac{d}{2}}\big(\sqrt{(\chi+Q(\boldsymbol{x}))(\psi+\gamma'\Sigma^{-1}\gamma)}\big) e^{(\boldsymbol{x}-\boldsymbol{\mu})'\Sigma^{-1}\gamma}}{\big(\sqrt{(\chi+Q(\boldsymbol{x}))(\psi+\gamma'\Sigma^{-1}\gamma)}\big)^{\frac{d}{2}-1}}, \quad -\infty < x < \infty, \end{aligned} \tag{4.8}$$

where the normalizing constant is

$$c = \frac{(\sqrt{\psi/\chi})^\lambda (\psi+\gamma'\Sigma^{-1}\gamma)^{\frac{d}{2}-\lambda}}{(2\pi)^{\frac{d}{2}}|\Sigma|^{\frac{1}{2}} K_\lambda(\sqrt{\chi\psi})},$$

and f_W is the density function of the GIG random variable W. Here $K_\lambda(.)$ is the modified Bessel function of the third kind and $Q(\boldsymbol{x})$ denotes the Mahalanobis distance $(\boldsymbol{x}-\boldsymbol{\mu})'\Sigma^{-1}(\boldsymbol{x}-\boldsymbol{\mu})$. These parameters admit the following interpretations:

- λ, χ, ψ specify the shape of the distribution and how much weight is assigned to the tails compared with the center. The larger those parameters are, the closer the distribution is to the normal distribution.
- $\boldsymbol{\mu}$ is the location parameter.
- $\Sigma = AA^{'}$ is the dispersion-matrix. It controls correlations between components of $\mathbf{X}$ and has to fulfill the usual conditions for covariance matrices, i.e., symmetry and positive definiteness as well as a full rank property.
- $\boldsymbol{\gamma}$ is the skewness parameter. If $\boldsymbol{\gamma} = \mathbf{0}$, then the distribution is symmetric.

The characteristic function of the GH distribution can be expressed as:

$$\phi_{\boldsymbol{X}}(\boldsymbol{u}) = \mathbb{E}\big(e^{i\boldsymbol{u}'\boldsymbol{X}}\big) = e^{i\boldsymbol{u}'\mu}\hat{H}\big(\frac{1}{2}\boldsymbol{u}'\Sigma\boldsymbol{u} - i\boldsymbol{u}'\boldsymbol{\gamma}\big), \tag{4.9}$$

where $\hat{H}(\theta) = \int_0^\infty e^{-\theta\nu} dF(\nu)$ is the Laplace-Stieltjes transform of the distribution function F of the GIG random variable W. See McNeil et al. (2005) for more details.

The GH distribution family includes some special cases under different names, listed as follows:

- If $\lambda = \frac{d+1}{2}$, we have a *multivariate hyperbolic* (hyp) distribution.
- If $\lambda = \frac{-1}{2}$, a *normal inverse Gaussian* (NIG) distribution is obtained.
- If $\chi = 0$, $\lambda > 0$, we have a *variance gamma* (VG) distribution.
- If $\psi = 0$, $\lambda < 0$, one gets a *generalized hyperbolic Student's t* distribution. The shape parameter for this particular case is $\nu = -2\lambda$, which determines the degrees of freedom.

In the next section, we use GH distributions to model the error terms in the CBD model.

4.5 Proposed Model

We first define the generalized hyperbolic Lévy process based on the GH distribution. Then we use this process to model the increments $\boldsymbol{A}_{t+1} - \boldsymbol{A}_t$. The generalized hyperbolic Lévy process is defined by

$$\boldsymbol{X}^{GH} = \{\boldsymbol{X}_t, t > 0\}, \tag{4.10}$$

where $\boldsymbol{X}_0 = 0$, with stationary and independent increments and $\boldsymbol{X}_t$ has characteristic function,

$$\mathbb{E}\big[\exp\big(i\boldsymbol{u}\boldsymbol{X}_t\big)\big] = \big(\phi_{\boldsymbol{X}}(\boldsymbol{u})\big)^t, \tag{4.11}$$

where $\phi_{\boldsymbol{X}}(\boldsymbol{u})$ is the characteristic function of the GH distribution, defined in (4.9).

Similarly to Cairns et al. (2006), we adopt the following mortality curve:

$$\log\Big(\frac{\tilde{q}(t,x)}{\tilde{p}(t,x)}\Big) = A^1_{t+1} + A^2_{t+1}(x), \tag{4.12}$$

and assume that for $t > 0$

$$\boldsymbol{A}_{t+1} - \boldsymbol{A}_t = \boldsymbol{X}_1, \tag{4.13}$$

where $\boldsymbol{X}_1$ is a bivariate GH Lévy process evaluated at a unit time scale. Here $\boldsymbol{A}_t = (A^1_t, A^2_t)'$ are two stochastic factors. In other words, we change the normality assumption in the CBD model to a bivariate GH Lévy process, while we keep the same structure as the CBD model for the evolution of survival probabilities. It is worth mentioning that the iid assumption of the increments for the GH Lévy process has already been tested for the selected data set in Section 4.3. Therefore, we can now fit the proposed model and compare it with the CBD model.

4.6 Model Comparisons

This section compares the CBD model with the proposed model. For the sake of comparison, we use the log-likelihood function (LLF), Akaike information criterion (AIC), Bayesian information criterion (BIC), as defined in (3.8) and (3.9), together with the likelihood ratio test (LRT).

The likelihood-ratio test can be used to check whether a special case of the GH distribution is the true underlying distribution. The LRT test statistic is defined as:

$$\Lambda = \frac{\sup\{L(\theta|\boldsymbol{Y}) : \theta \in \Theta_0\}}{\sup\{L(\theta|\boldsymbol{Y}) : \theta \in \Theta\}}, \tag{4.14}$$

where L denotes the likelihood function with respect to the parameter θ and data $\boldsymbol{Y}$, and Θ_0 is a subset of the parameter space Θ. The null hypothesis H_0 states that $\theta \in \Theta_0$ and the alternative hypothesis H_1 states that $\theta \in \Theta_0^c$, where Θ_0^c is the complement of Θ_0. Under the null hypothesis and certain regularity conditions, it can be shown that $-2\log\Lambda$ is asymptotically chi-square distributed with ν degrees of freedom. Here ν is the number of free parameters specified in Θ minus the number of free parameters specified in Θ_0. The null hypothesis is rejected if $-2\log\Lambda$ exceeds the confidence level-quantile of the chi-square distribution with ν degrees of freedom. In this study, H_0 is the bivariate normal distribution in the CBD model and H_1 is the special case of the GH distribution.

Table 4.5: In-sample Goodness of Fit Measures

GH Distribution	Symm.	LLF	AIC	BIC	LLF Rank	AIC Rank	BIC Rank
ghyp	T	233.39	-454.79	-446.38	2	1	1
NIG	T	231.84	-451.68	-443.27	4	2	3
ghyp	F	233.66	-451.32	-440.12	1	3	4
t	T	230.53	-451.06	-444.06	6	4	2
hyp	T	229.96	-447.93	-439.52	8	5	5
NIG	F	231.91	-447.81	-436.60	3	6	7
t	F	230.73	-447.46	-437.65	5	7	6
hyp	F	230.03	-444.07	-432.86	7	8	8
Normal	T	221.29	-432.57	-425.56	9	9	9

4.6.1 Empirical Analysis

For the purpose of measuring the in-sample model performance, we use the mortality data for Males in Italy from 1969-1999 and ages 60-90. Then, to assess the out-of-sample model performance, we forecast the development of the mortality rates for the 9 subsequent years. We first estimate $\boldsymbol{A}_t$ in (4.13) by using the least square technique, then we fit ten GH distributions with the maximum likelihood method. The considered distributions are: Student's t, NIG, hyp and generalized hyperbolic distributions (ghyp) with density function defined in (4.8), both in the symmetric and asymmetric cases. We use the R package *ghyp* in order to fit the above distributions for $\boldsymbol{X}_1$ defined by (4.13).

Table 4.5 provides the in-sample goodness of fit measures based on the LLF, the AIC and the BIC statistics, together with their corresponding ranks. A commonly used rule of thumb consists in considering that two models are significantly different if the difference in the AIC is larger than 10, as discussed in Burnham and Anderson (2002). Raftery (1995) suggests that a model significantly outperforms a competitor if the difference in their respective BIC values exceeds 5. Therefore, all three criteria show a preference for GH distributions when comparing to the normal distribution with the lowest rank. The symmetric ghyp distribution is the best distribution based on the BIC and the AIC. According to the LLF, the asymmetric ghyp distribution offers the best fit to our mortality data set.

We use the R function *lik.ratio.test*[1] to perform likelihood ratio tests. Table 4.6

[1]This function is available in the package *ghyp*.

Table 4.6: Likelihood-Ratio Test

Model	Symm.	L-statistic	df	p-value
ghyp	T	5.52×10^{-06}	1	8.62×10^{-07}
NIG	T	2.61×10^{-06}	1	4.30×10^{-06}
ghyp	F	4.20×10^{-06}	2	4.20×10^{-06}
hyp	T	1.7×10^{-04}	1	3.1×10^{-05}
NIG	F	2.40×10^{-05}	2	2.40×10^{-05}
t	F	7.91×10^{-05}	1	1.38×10^{-05}
hyp	F	1.59×10^{-04}	2	1.59×10^{-04}

provides the test statistics, degrees of freedom and the p-values. The likelihood ratio tests are statistically significant for the selected GH distributions. This table indicates that the considered mortality data set is more likely to come from the GH distribution than a bivariate normal distribution.

Overall, Tables 4.5 and 4.6 provide some evidence to support the use of GH distributions for modeling $\boldsymbol{X}_1$ in (4.13).

4.6.2 Mortality Projection

In this section, the out-of-sample performance of the proposed model is investigated. The reference cohort is the set of males aged 65 in 1999. We first explain how to use (4.12) in order to project the mortality rates for nine years corresponding to $t = 2000, 2001, ..., 2008$. We generate nine iid copies of $\boldsymbol{X}_1$ from the fitted symmetric ghyp distribution based on the mortality data over the period of 1969-1999. Then, we apply (4.13) together with the estimated value of $\boldsymbol{A}_{1999}$ in order to obtain $\boldsymbol{A}_t$ for $t = 2000, 2001, ..., 2008$. Next, we use (4.12) to project the mortality rates for the considered reference cohort, denoted by $\hat{q}(1999+i, 65+i)$, $i = 1, 2, ..., 9$. Finally, to evaluate the out-of-sample performance, we repeat the above procedure 20,000 times and record the projected mortality rates. Similarly to Wang et al. (2011), we find the mean absolute percentage error (MAPE) for each replication $j = 1, 2, ..., 20,000$, defined as follows:

$$MAPE_j = 100\% \times \frac{1}{9}\sum_{i=1}^{9}\Big|\frac{\tilde{q}(1999+i, 65+i) - \hat{q}(1999+i, 65+i)}{\tilde{q}(1999+i, 65+i)}\Big|, \qquad (4.15)$$

where $\tilde{q}(t, x)$ is the observed mortality rate at time t for the cohort aged x in simulation j. Table 4.7 illustrates the differences in mortality projection between

Table 4.7: Percentile of MAPE of Mortality Projection (Unit: %)

Model	Symm.	Mean	90%	95%
ghyp	T	8.76	14.48	16.55
Normal	T	9.74	16.13	18.32

the symmetric ghyp distribution and normal distribution, based on the mean, 90^{th} percentile, and the 95^{th} percentile of the MAPE. In this table, the model with a better predictive power will have lower mean and percentiles. We find that the symmetric ghyp distribution provides better mortality projection performance based on the mean, 90^{th} percentile, and the 95^{th} percentile of MAPE.

In addition to the Italian mortality data set used in this chapter, we also fitted our model to the Russian, Spanish and U.S. mortality data. Similarly to the results given in this section, we find that the ghyp distribution fits better to the data compared to the normal distribution, although these results have not been reported in this chapter.

4.7 Conclusions

In this chapter, we show that the bivariate normality assumption in the CBD model is sometimes appropriate for the considered mortality data. We test normality using the multivariate normality test of Doornik and Hansen (2008) and the multivariate Shapiro-Wilk normality test. We reject the bivariate normality assumption based on both tests.

Generalized hyperbolic distributions are proposed to model the increments of $\boldsymbol{A}_t$. We estimate the parameters by maximum likelihood. Four GH distributions, in the symmetric and asymmetric cases, are compared with the CBD model based on the AIC, BIC, LLF and the likelihood ratio test. We find that the symmetric ghyp distribution provides the best fit and also gives better mortality projections according to the MAPE.

To summarize the chapter, we suggest testing the bivariate normality assumption in the CBD model before using it for projections. In addition, the GH distributions can be considered as a viable alternative to the Normal distribution in the CBD model.

Chapter 5

Modeling Catastrophic Mortality Bonds

5.1 Introduction

In this chapter, we propose to model age adjusted death rates embedded in the Swiss Re mortality bond that was offered in 2003 using a generalized least squares approach. The model allows the error terms to be correlated and includes a binary random variable that represents the occurrence of mortality jumps (in particular influenza pandemics). We apply the *variable length Markov chains* (VLMC) model proposed by Mächler and Bühlmann (2004) to model the incidence of catastrophic events. The developed model is then compared to the currently recognized models in the literature. A simulation study is carried out to estimate the market price of risk in the Swiss Re bond using the Wang's transformation.

Catastrophe mortality bonds are designed to transfer a set of specific risks, such as extreme mortality, from a sponsor to investors. For instance, suppose we would like to price the Swiss Re bond introduced in December 2003 and consider the influenza pandemic in 1918 as the catastrophic event that significantly increased the mortality rates over a short period of time. The first and probably the most important question that comes to mind is if it could happen again? Modern virologists and numerous epidemiologists agree that similar events can actually happen again. See for example Webster and Walker (2003).

Moreover, since the conditions that caused the influenza pandemic in 1918 happened once before, similar conditions can lead to equally disastrous pandemics, as indicated in Taubenberger and Morens (2006). Also, the prevalence of the virus can be accelerated by air travels while medical centers can be overfilled. With all medical advances and the current knowledge in preventing diseases, the return of influenza pandemic virus in 1918 would probably kill more than 100 million people across the world, as addressed in Taubenberger and Morens (2006). It

is therefore crucial for both investors and (re)insurance companies to take into account mortality jumps when pricing mortality catastrophe bonds.

In this chapter, we focus on mortality shocks caused by severe influenza pandemics and study securitization in regard to extreme mortality risks. Different approaches have been proposed to model mortality linked securities. Milidonis et al. (2011) proposed a regime switching model with two states that follows a geometric Brownian motion. Deng et al. (2012) suggest a double-exponential jump diffusion model in order to model mortality jumps as well as cohort effects. Bauer and Kramer (2008) introduced a time-continuous model to price catastrophe mortality claims with stochastic force of mortality.

Lin and Cox (2008) use a geometric Brownian motion together with a log-normal jump size distribution to estimate the market price of Swiss Re bonds. They consider two consecutive periods of time where a mortality jump can occur. Each interval may contain a maximum of one jump. As a result, four possible cases are considered and the likelihood function can be obtained by conditioning on the occurrence of a mortality jump within each interval. They construct a dependency structure among the mortality indexes (in the log scale). They also approximate the correlation introduced by mortality jumps in the likelihood function with an additional independence assumption. The model we develop characterizes this dependency by embedding a time series approach within the context of generalized least squares. However, our model does not need the extra independence assumption of Lin and Cox (2008) since the log-likelihood function is maximized for a single observation that has a multivariate normal distribution. This approach enables us to jointly estimate the parameters while considering the correlation imposed by mortality jumps. Also as an application, our model shows that Swiss Re bonds (issued in 2003) exhibit lower risk premiums for the investors than those obtained in Lin and Cox (2008).

Furthermore, our model not only improves the models by Milidonis et al. (2011) and Lin and Cox (2008). We find lower risk premiums than Lin and Cox (2008). It also provides a clear fitting performance over historical data. To the best of our knowledge, this property may be difficult to fulfill (if not impossible) by applying the two above mentioned papers.

In this chapter, we focus on modeling the mortality index of the Swiss Re bonds. The model that we develop takes into account both the occurrence of jumps and their magnitude. Our approach is in two stages. In the first stage, we model the occurrence of jumps by using *variable length Markov chains* proposed by Mächler and Bühlmann (2004). These results are then applied, by generalized least squares, to model the magnitude of jumps (if any) in the mortality index. Overall, we believe that the model proposed in this chapter can be distinguished from the

current well known models in the literature by first showing a less restrictive algorithm to estimate the parameters and at the same time providing a better fit to the historical data. It also helps the reader to visually evaluate the performance of the fitted values.

5.2 Data

We use the weighted average of mortality data over five different countries including: U.K., France, Italy, Switzerland and the U.S., as introduced in the design of the Swiss Re mortality bonds in late December 2003. We construct the mortality index, q_t over the period of 1900-2008, according to Krutov (2010) as follows:

$$q_t = \sum_{j=1}^{5} C_j \sum_{i=1}^{12} A_i \big(G^m q_{ijt}^m + G^f q_{ijt}^f \big), \qquad t = 1900, ..., 2008, \tag{5.1}$$

where C_j is the weight assigned to country j, A_i is the weight assigned to age group i, G^m is the weight assigned to males, G^f is the weight assigned to females, q_{ijt}^m is the mortality rate for male lives in age group i, in country j, in year t, and q_{ijt}^f is the mortality rate for female lives in age group i, in country j, in year t. The weights assigned to each country are: 15% of U.K., 7.5% of France, 2.5% of Italy, 5% of Switzerland and 70% of the U.S. Here G^m and G^f are set to be 65% and 35%, respectively. Table 5.1 summarizes the weights assign to each age group.

Table 5.1: Weights Assign to Each Age Group

Age group (i)	A_i
20-24	1%
25-29	5%
30-34	12.5%
35-39	20%
40-44	20%
45-49	16%
50-54	12%
55-59	7%
60-64	3%
65-69	2%
70-74	1%
75-79	0.5%

The mortality rates for each gender in (5.1) are obtained by using $q_{ijt} = \frac{D_{ijt}}{Pop_{ijt}}$, where D_{ijt} and Pop_{ijt} are the total number of deaths and population size for

the age group i, country j, in year t, respectively. In order to find q_{ijt}, we use the Human Mortality Database[1] to obtain D_{ijt} and Pop_{ijt} for the U.K, France, Italy, Switzerland over the period 1900-2008 and during 1933-2008 for the U.S. We complete the data set for the U.S. for the period 1900-1932 by using the Human Life Table[2] database as our source.

Figure 5.1 illustrates the time series of q_t where $t = 1900, ..., 2008$. There have been four influenza pandemics during the 20th century including the Spanish flu (1918), Asian flu (1957), Hong Kong flu (1968) and Russian flu (1977). As Figure 5.1 shows, just the Spanish flu dramatically increased the mortality index.

It is possible to categorize influenza pandemics according to their case-fatality ratio[3] (CFR), when the data are available. The *pandemic severity index* introduced by the United States Department of Health and Human Services in 2007 classifies pandemics into 5 different categories according to their CFR, ranging from less that 0.1% in Category 1 to 2% or higher in Category 5. The Spanish flu of 1918 is by far the most severe influenza pandemic ever experienced and is placed in Category 5 while the Asian and Hong Kong flu are set into Category 2.

There have been 13 or more influenza pandemics since 1500. It is difficult to obtain the pandemic severity index for all the past influenza pandemics since the data are not readily available. Therefore, we rely on Taubenberger and Morens (2009) to identify the pandemics in 1557, 1580, 1729, 1889 and 1918 as severe pandemics with high fatality rates. The VLMC model as defined in Mächler and Bühlmann (2004) shall be used to predict the occurrence of severe influenza pandemics. For categorical time series data, the VLMC provides substantial additional flexibility when fitting and modelling Markov chains. We later use this approach in our simulation analysis, as detailed in Subsection 5.8.1.

5.3 Variable Length Markov Chain

This section gives the formal definition of VLMC according to Mächler and Bühlmann (2004). Let Flu_t be a stationary process for $t \in \mathbb{Z}$ taking values in a finite categorical space χ. In this chapter, Flu_t is a binary time series indicating the occurrence of highly fatal influenza pandemics at time t. The deterministic values are denoted by small letters and let $x_i^j = (x_j, x_{j-1}, x_{j-2}, ..., x_i)$, $i \leq j$, be the vector of deterministic values whose components are written in reverse order. Similarly, we can define a random vector as $Flu_i^j = (Flu_j, Flu_{j-1}, Flu_{j-2}, ..., Flu_i)$, $i \leq j$. A preliminary function c_{pre} is defined by: $c_{pre} : \chi^\infty \to \bigcup_{j=0}^{\infty} \chi^j \cup \chi^\infty \quad (\chi^0 = \emptyset)$

[1] The data are available online at www.mortality.org.

[2] The data are available online at www.lifetable.de.

[3] The percentage of deaths out of the total reported cases of the disease in the U.S.

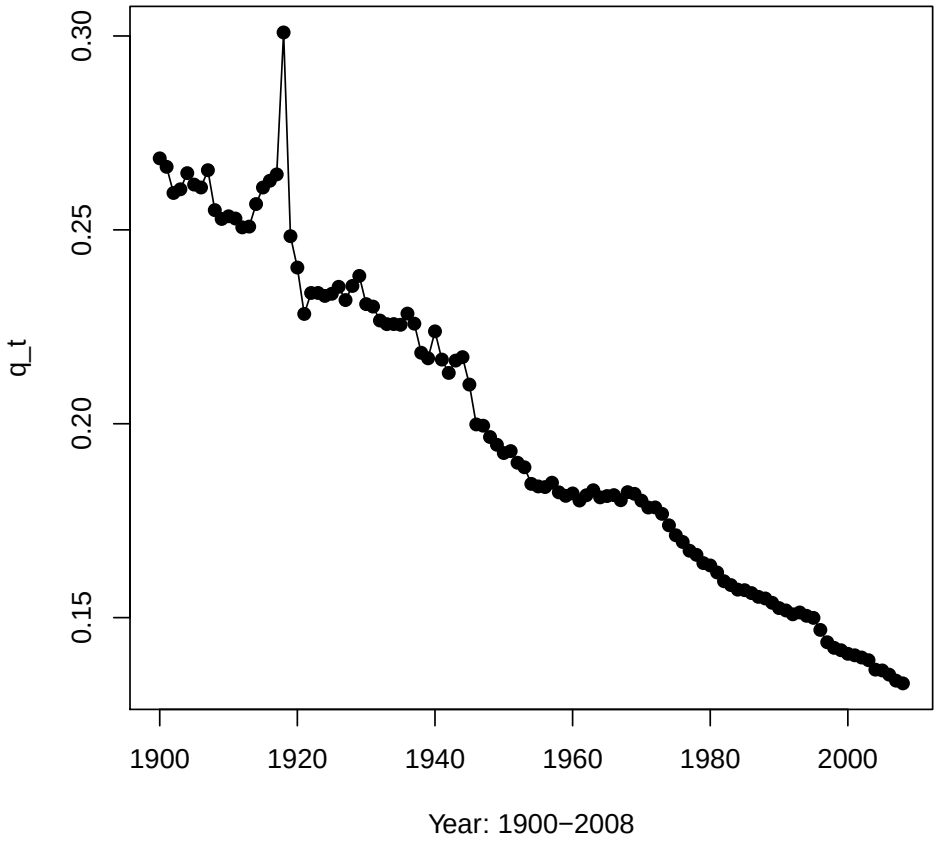

Figure 5.1: Mortality index of Swiss Re bond: 1900-2008

that maps an infinite sequence (the infinite past) to a possibly shorter string: $c_{pre} : x^0_{-\infty} \mapsto x^0_{-l+1}$, where l is defined by

$$\begin{aligned} l = l(x^0_{-\infty}) = \min\{ \quad k \quad & ; P[Flu_1 = x_1 | Flu^0_{-\infty} = x^0_{-\infty}] \\ = \quad & P[Flu_1 = x_1 | Flu^0_{-k+1} = x^0_{-k+1}], \forall x_1 \in \chi\}, \end{aligned} \tag{5.2}$$

and $l \equiv 0$ corresponds to independence. The function $c_{pre}(.)$ is called the preliminary context function. Let $0 \leq p \leq \infty$ be the smallest integer such that

$$card\big(c_{pre}(x^0_{-\infty})\big) = l(x^0_{-\infty}) \leq p, \quad \forall \quad x^0_{-\infty} \in \chi^\infty.$$

The number p is called the order of the preliminary context function $c_{pre}(.)$ and if $p < \infty$, Flu_t is called a stationary VLMC of order p. A VLMC is a full Markov chain of order p if the preliminary context function $c_{pre}(.)$ of order p is the full projection $x^0_{-\infty} \mapsto x^0_{-p+1}$ for all $x^0_{-\infty}$. The final form of a context function is denoted by $c(.)$ and combines the values of $c_{pre}(.)$ whose last symbols are the same.

The context tree τ is defined as $\tau = \tau_c = \{w; w = c(x^0_{-\infty}), x^0_{-\infty} \in \chi^\infty\}$, where $c(.)$ is the context function of a stationary VLMC. The structure of the variable length memory should be estimated to fit the VLMC. As this contains non-trivial steps, we refer interested readers to Section 6.1 of Mächler and Bühlmann (2004) for the details of fitting VLMC's. It is worth mentioning that in R (as our statistical software) the package *VLMC* can be used to estimate and predict variable length Markov chains.

5.3.1 Example

The following example clarifies the above definitions. In this example, H_t is a categorical time series that takes only values $\{A, B, C\}$ i.e. $\chi = \{A, B, C\}$. Consider the following Markov chain model of order 2 where

$$\begin{aligned} & P[H_t = h_t | H_{t-1} = h_{t-1}, H_{t-2} = h_{t-2}, ...] \\ & = \begin{cases} P[H_t = h_t | H_{t-1} = A] & if \quad h_{t-1} = A \\ P[H_t = h_t | H_{t-1} = B] & if \quad h_{t-1} = B \\ P[H_t = h_t | H_{t-1} = C, H_{t-2} \in \{A, B\}] & if \quad h_{t-1} = C, h_{t-2} \in \{A, B\} \\ P[H_t = h_t | H_{t-1} = C, H_{t-2} = C] & if \quad h_{t-1} = C, h_{t-2} = C. \end{cases} \end{aligned} \tag{5.3}$$

The preliminary context function with transition probabilities given in (5.3) is:

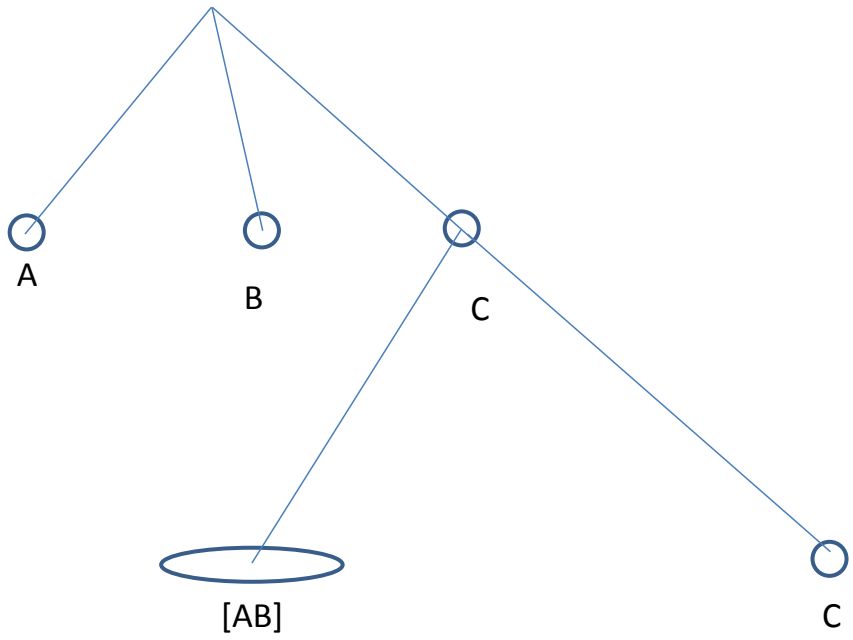

Figure 5.2: The tree representation of Example 5.3.1

$$c_{pre}(h^0_{-\infty}) = \begin{cases} A & if \quad h_0 = A, \quad h^{-1}_{-\infty} \quad \text{arbitrary} \\ B & if \quad h_0 = B, \quad h^{-1}_{-\infty} \quad \text{arbitrary} \\ CA & if \quad h_0 = C, \quad h_{-1} = A, \quad h^{-2}_{-\infty} \quad \text{arbitrary} \\ CB & if \quad h_0 = C, \quad h_{-1} = B, \quad h^{-2}_{-\infty} \quad \text{arbitrary} \\ CC & if \quad h_0 = C, \quad h_{-1} = C, \quad h^{-2}_{-\infty} \quad \text{arbitrary.} \end{cases} \tag{5.4}$$

The final representation of the context function can be obtained by aggregating the values of $c_{pre}(.)$ whose second last symbols are the same as given below:

$$c(h^0_{-\infty}) = \begin{cases} A & if \quad h_0 = A, \quad h^{-1}_{-\infty} \quad \text{arbitrary} \\ B & if \quad h_0 = B, \quad h^{-1}_{-\infty} \quad \text{arbitrary} \\ C[A,B] & if \quad h_0 = C, \quad h_{-1} \in \{A,B\} \quad h^{-2}_{-\infty} \quad \text{arbitrary} \\ CC & if \quad h_0 = C, \quad h_{-1} = C, \quad h^{-2}_{-\infty} \quad \text{arbitrary.} \end{cases} \tag{5.5}$$

Figure 5.2 show the tree representation of Example 5.3.1.

5.4 Review of Generalized Least Squares

The standard linear regression model assumes that

$$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \tag{5.6}$$

where $\boldsymbol{Y}$ is the $n \times 1$ response vector, $\boldsymbol{X}$ is the $n \times (k+1)$ matrix of predictor variables, $\boldsymbol{\beta}$ is the $(k+1) \times 1$ vector of regression coefficients and $\boldsymbol{\varepsilon}$ is an $n \times 1$ vector of regression errors. For this model, the error sum of squares $\boldsymbol{\varepsilon}'\boldsymbol{\varepsilon}$ can be minimized to obtain the ordinary least squares estimates of the regression coefficients i.e. $\hat{\boldsymbol{\beta}}_{ols} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{Y}$. Furthermore, if we assume that $\boldsymbol{\varepsilon} \sim N_n(\boldsymbol{0}, \sigma^2\boldsymbol{I}_n)$, i.e. the error terms are uncorrelated and follow an n-dimensional multivariate normal distribution with mean zero and constant variance, then the covariance matrix of regression coefficients can be written as $Var(\hat{\boldsymbol{\beta}}_{ols}) = \sigma^2(\boldsymbol{X}'\boldsymbol{X})^{-1}$. The assumption of uncorrelated errors may not always be true. In this case, we can use a generalized least squares approach as follows. Let $\boldsymbol{\varepsilon} \sim N_n(\boldsymbol{0}, \boldsymbol{\Sigma})$, where $\boldsymbol{\Sigma}$ is a symmetric and positive-definite matrix of covariance errors. Then, for the known $\boldsymbol{\Sigma}$, the maximum likelihood estimator of $\boldsymbol{\beta}$ can be written as

$$\hat{\boldsymbol{\beta}}_{gls} = (\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{Y}, \tag{5.7}$$

with the covariance matrix $Var(\hat{\boldsymbol{\beta}}_{gls}) = (\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{X})^{-1}$. From the practical point of view, we first need to estimate $\boldsymbol{\Sigma}$ from the data, in order to estimate the regression coefficients. This means that the correlation structure within the error terms should be correctly modeled prior to estimate the regression parameters. As an example, in the next subsection, we focus in a particular case where the error terms $\boldsymbol{\varepsilon}$ are not independent and will be considered as a time series with a serially autocorrelated structure.

5.4.1 Stationary Correlated Errors

In this subsection, we assume that the error terms follow a stationary process with mean zero and constant variance σ^2. Also, the covariance structure depends only on the distance of two error terms, i.e. $Cov(\varepsilon_t, \varepsilon_{t-s}) = Cov(\varepsilon_t, \varepsilon_{t+s}) = \sigma^2\rho_s$, where ρ_s is the error autocorrelation at lag s. Therefore, the error covariance matrix can be expressed as $\boldsymbol{\Sigma} = [a_{ij}]_{n\times n}$, where a_{ij} is the entry in the ith row and jth column that equals to $\sigma^2\rho_{|i-j|}$, for $i \neq j$, and σ^2 otherwise. This means that $\boldsymbol{\Sigma}$ (and subsequently $\hat{\boldsymbol{\beta}}_{gls}$ in (5.7)) can be obtained by estimating σ^2 and ρ_s. We model the error terms according to an Autoregressive Integrated Moving Average process, ARIMA(p, d, q), such that

$$\big(1 - \sum_{i=1}^{p} \phi_i L^i\big)\big(1 - L\big)^d \varepsilon_t = \big(1 + \sum_{i=1}^{q} \theta_i L^i\big)\nu_t, \tag{5.8}$$

where L is the lag operator, the ϕ_i are the parameters of the autoregressive part of the model, the θ_i are the parameters of the moving average. Here ν_t represents the white noise term that are normally distributed with mean zero and constant

variance σ_ν^2 and independent of ε_t. As an example, if the error terms follow an autoregressive process of order one, AR(1), i.e. $\varepsilon_t = \phi\varepsilon_{t-1} + \nu_t$, then we can easily show that: $\rho_1 = \phi$, $\rho_s = \phi^s$ and $\sigma^2 = \dfrac{\sigma_\nu^2}{1-\phi^2}$. The log-likelihood of the model can be written as:

$$\ln L(\boldsymbol{\beta}) = -\frac{n}{2}\ln 2\pi - \frac{1}{2}\ln(det\boldsymbol{\Sigma}) - \frac{1}{2}(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta}), \tag{5.9}$$

where n is the number of observations. In the case of AR(1), we can obtain the maximum likelihood estimates by maximizing the log-likelihood function in (5.9) to jointly estimate the regression coefficients, $\hat{\boldsymbol{\beta}}_{gls}$, ϕ and σ_ν^2.

5.5 Model Specification

For our in-sample analysis, we use the estimated mortality index, q_t in Section 5.2 during $1900-1998$[1]. For simplicity, we indicate this period of time by $t = 1, ..., 99$. We develop our primary model using the standard linear regression in (5.6) based on orthogonal polynomials of degree q as follows:

$$\begin{bmatrix} Y_1 \\ . \\ . \\ . \\ Y_{99} \end{bmatrix} = \begin{bmatrix} P_0(1) & P_1(1) & . & . & . & P_N(1) & 0 \\ P_0(2) & P_1(2) & . & . & . & P_N(2) & 0 \\ . & . & . & . & . & . & . \\ P_0(19) & P_1(19) & . & . & . & P_N(2) & 1 \\ . & . & . & . & . & . & . \\ P_0(99) & P_1(99) & . & . & . & P_N(99) & 0 \end{bmatrix} \times \begin{bmatrix} \beta_0 \\ \beta_1 \\ . \\ . \\ \beta_N \\ \beta_F \end{bmatrix} + \begin{bmatrix} \varepsilon_0 \\ \varepsilon_1 \\ . \\ . \\ \varepsilon_N \\ \varepsilon_F \end{bmatrix}, \tag{5.10}$$

where $Y_t = \text{logit}(q_t) = \log\left(\frac{q_t}{1-q_t}\right)$ for $t = 1, 2, ..., 99$ and $P_j(t)$ is an orthogonal polynomial[2] of degree j that can be calculated using a three-term recursion as given in Kennedy and Gentle (1980). See Appendix A.4 for the construction of $P_j(t)$. Equation (5.10) is equivalent to

$$Y_t = \text{logit}(q_t) = \sum_{j=0}^{N} \beta_j P_j(t) + \beta_F Flu_t + \varepsilon_t, \quad t = 1, 2, ..., 99, \tag{5.11}$$

where Flu_t is a binary sequence of $\{0, 1\}$ that indicates the occurrence of influenza pandemic at time t and ε_t is a sequence of uncorrelated and normally distributed random variables with zero-mean and constant variance, σ^2. As discussed in Section 5.2, the influenza pandemic in 1918 was much more severe than the pandemics in 1957 1968 and 1977. Therefore, we only mark the influenza pandemic in 1918 with $Flu_{19} = 1$ and $Flu_t = 0$, otherwise.

[1] We use this period of time to be consistent with Lin and Cox (2008).

[2] Using the orthogonal polynomials reduces significantly the multicollinearity in the polynomial regression analysis and avoids any possible numerical problems due to near-singularity.

Table 5.2: Ordinary Least Square Estimations for Model (5.11) with $N = 4$

Coefficient	Estimates	Std. Error	t value	p-value
β_0	-3.172514	0.005062	-626.728	2×10^{-16}
β_1	-3.860225	0.050413	-76.572	2×10^{-16}
β_2	-0.079582	0.050112	-1.588	0.11566
β_3	0.138954	0.050298	2.763	0.00691
β_4	-0.162132	0.050528	-3.209	0.00183
β_F	0.413613	0.051300	8.063	2.51×10^{-12}

There are two popular procedures to select the order of the polynomial in (5.11). The first one is to start with a very simple model (i.e. $N = 0$) and then try to add higher order of polynomials (forward selection). The second technique is based on choosing a sufficiently high order and then dropping terms sequentially (backward elimination). The criterion for adding or dropping terms is usually the t-test, although other criteria can be considered. We follow the second method and start by fitting an ordinary linear regression model (5.11) with $N = 4$. Table 5.2 summarizes the least squares estimation of the regression coefficients together with their standard errors, t-values and p-values. Based on the the t-test, all the predictable variables are statistically significant except the quadratic polynomial at $\alpha = 0.01$. Prior to drooping any terms for this model we first check adequacy of the model by inspecting the residuals of the fitted model for any serial autocorrelation, constant variance and the normality assumption. This is a crucial step; since if the residuals are serially autocorrelated, then the normal t-test fails to provide a reliable criteria.

Figure 5.3 shows the time series of the residuals in the fitted model. Here the residuals are clearly serially autocorrelated and therefore the assumption of independence is not satisfied. Also, there is a tendency for negative residuals to follow negative values and positive residuals to follow positive values. Moreover, the variance of the residuals seems to be constant over time. We use the normality test proposed by Jarque and Bera (1981) to check the normality assumption of the residuals. The test statistic is 0.308 with the p-value of 0.8573. Therefore, the normality assumption of the residuals is not rejected at a 5% level. To correct the autocorrelation problem, we model the residuals of the fitted regression equation as an ARIMA(p, d, q) process. Therefore, the orders p, d, and q should be properly selected.

We next obtain the sample autocorrelation function (ACF), and sample partial autocorrelation function (PACF) denoted by r_k and ϕ_{kk}, respectively, at lag k that

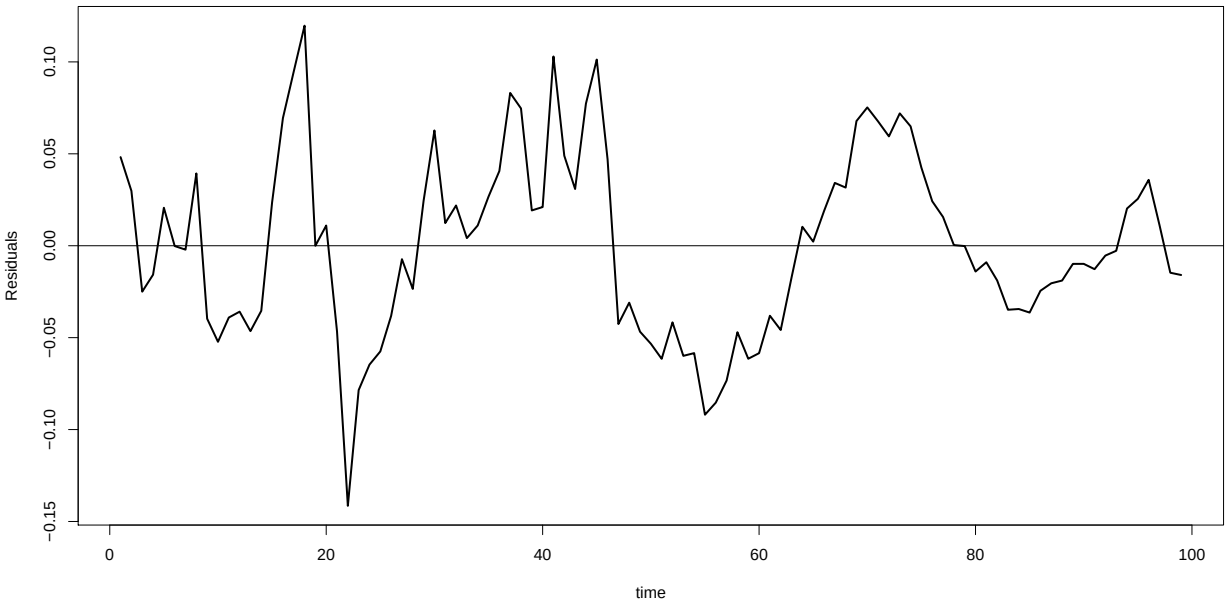

Figure 5.3: Residuals (ε_t) of model (5.11) with $N = 4$

is defined by:

$$r_k = \frac{\sum_{t=k+1}^{n}(\varepsilon_t - \bar{\varepsilon})(\varepsilon_{t-k} - \bar{\varepsilon})}{\sum_{t=k+1}^{n}(\varepsilon_t - \bar{\varepsilon})^2},\ k = 1, 2, ..., \tag{5.12}$$

$$\phi_{kk} = \frac{r_k - \sum_{j=1}^{k-1} \phi_{k-1,j}.r_{k-j}}{1 - \sum_{j=1}^{k-1} \phi_{k-1,j} r_j}, \tag{5.13}$$

where $\phi_{k,j} = \phi_{k-1,j} - \phi_{kk}\phi_{k-1,k-j}$ for $j = 1, 2, ..., k-1$. Figure 5.4 shows the sample ACF and PACF of the fitted regression model together with the 95 % confidence limits (dotted lines). There is a sinusoidal decay in the ACF plot and two spikes at lag 1 and 4 in the PACF. Therefore, this plot suggests an AR(1) with the coefficient close to one. Alternatively, we can use the Akaike information criterion (AIC) and Bayesian information criterion (BIC) to select the order p, d and q, as defined in (3.8) and (3.9). The model with the minimum AIC or BIC can be considered as the preferred model, given a set of candidate models. Overall, we fitted 27 ARIMA models with $p, d, q \in \{0, 1, 2\}$ and searched for the model with the minimum criteria. The best models according to the AIC and BIC criteria, within the fitted models, are ARIMA(2,0,1) and AR(1), respectively. Table 5.3 summarizes the results for these two candidate models. Here, the differences in the AIC values are less than one unit and negligible. However, the BIC criterion for AR(1) is smaller than the BIC for ARIMA(2,0,1) by almost 4.5. This suggests the use of AR(1) in modeling the residuals of the fitted regression model.

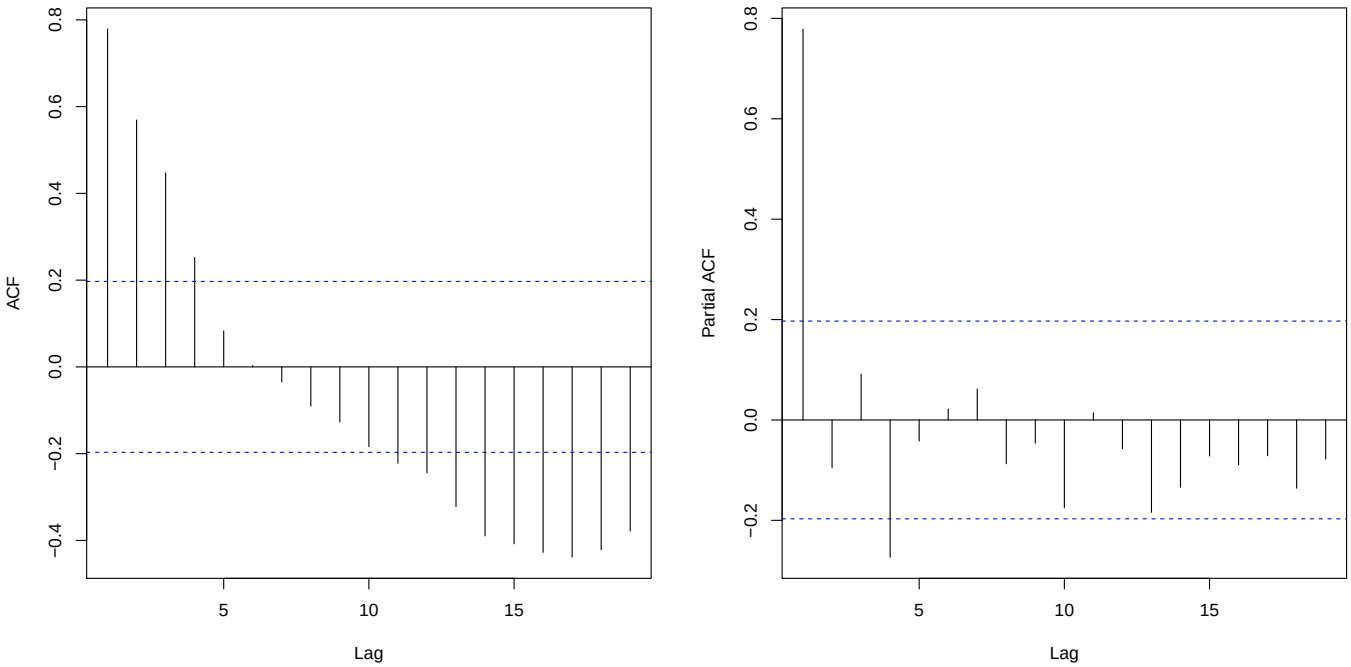

Figure 5.4: ACF and PACF plot of the residuals in model (5.11) with $N = 4$

Table 5.3: The Best Two Candidate Models Based on AIC and BIC Criteria

Model	AIC	BIC
AR(1)	-407.11	-401.92
ARIMA(2,0,1)	-407.94	-397.56

Table 5.4: Generalized Least Square Estimations for Model 5.14, $\hat{\phi} = 0.803$, $\hat{\sigma_\nu} = 0.0486$

Coefficient	Estimates	Std. Error	t value	p-value
β_0	-3.171424	0.01488672	-213.03714	0.0000
β_1	-3.890020	0.13943457	-27.89853	0.0000
β_2	-0.065357	0.13212128	-0.49468	0.6220
β_3	0.090311	0.11558114	0.78137	0.4366
β_4	-0.138227	0.10728056	-1.288469	0.2008
β_F	0.349926	0.02334694	14.98809	0.0000

5.6 Model Fit

We now use the generalized least squares approach explained in Subsection 5.4.1 to fit the model

$$Y_t = \text{logit}(q_t) = \sum_{j=0}^{4} \beta_j P_j(t) + \beta_F Flu_t + \varepsilon_t, \quad t = 1, 2, ..., 99, \tag{5.14}$$

where ε_t follows an AR(1) process. The maximum likelihood estimates of the regression parameters and their standard errors together with t-values and p-values are reported in Table 5.4. We find that the standard errors of the regression coefficients have been adjusted to take into account the correlation structure embedded in the model. Here, in contrast to the results given in Table 5.2, the t-test indicates that the contribution of the polynomials of order 2, 3 and 4 is not statistically significant. In order to select the order of polynomials, we update the model (5.14) by sequentially dropping the non-significant terms. Therefore, in addition to model (5.14), we fit three more models to the same data set using a generalized least squares[1] according to:

$$Y_t = \text{logit}(q_t) = \sum_{j=0}^{N} \beta_j P_j(t) + \beta_F Flu_t + \varepsilon_t, \quad t = 1, 2, ..., 99, \tag{5.15}$$

where ε_t follows an AR(1) process for $N = 1, 2, 3$.

Figure 5.5 shows the observed mortality index in the log scale (solid dots), q_t during $1900 - 2008$ and the fitted curves based on model (5.14) with $N = 4$ together with models (5.15) using polynomials of order $N = 1, 2, 3$, as described above. For each fitted model, we project the mortality index for the next 18 years (i.e. 1999-2016). We can now evaluate the out of sample performance of the

[1]Similarly to our fitting procedure for model (5.14) and prior to applying the generalized least squares approach, we first fit these three models using linear regression and found that the AR(1) is a suitable process to model the error terms.

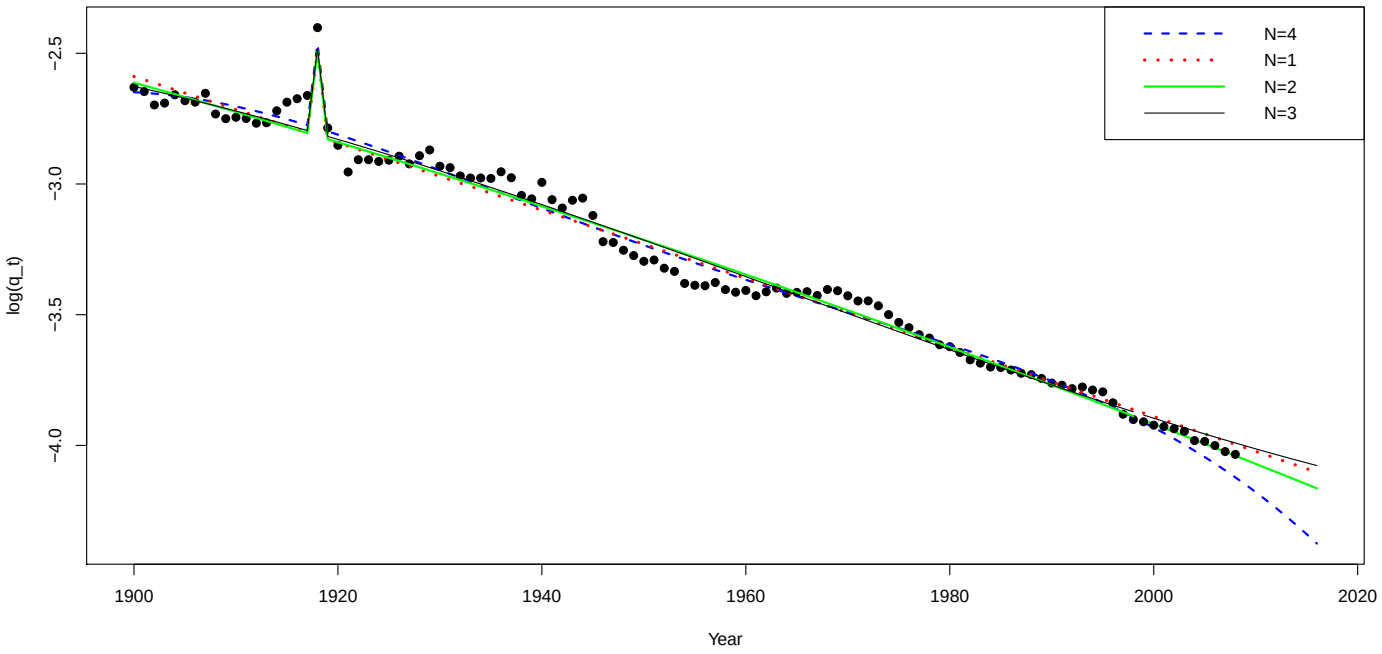

Figure 5.5: Observed mortality index and the fitted curves in model (5.15) using polynomials of order $N = 1, 2, 3$ and model (5.14) with $N = 4$

considered models. It is now clear that the model with $N = 4$ will under-estimates the mortality index over the projection period, while the models with $N = 1$ and $N = 3$ over-estimate the mortality index. The model with $N = 2$ seems to fit the data well and provide a reasonable projection. Therefore, we keep the quadratic term in the model. In this case, the fitted model is

$$\text{logit}(q_t) = -3.17P_0(t) - 3.86P_1(t) - 0.13P_2(t) + 0.35Flu_t + \varepsilon_t, \tag{5.16}$$

where $\varepsilon_t = 0.83\varepsilon_{t-1} + \nu_t$ and ν_t is normally distributed with mean zero and standard deviation of 0.053. The inclusion of the Flu_t in (5.16) is statistically significant at $\alpha = 0.001$. This shows that mortality jumps should be taken into account when modeling the mortality index. Although this conclusion is consistent with Lin and Cox (2008), here we used a different approach to show the importance of mortality jumps.

5.7 Model Comparisons

In this section, we compare our model with two other well known models proposed by Milidonis et al. (2011) and Lin and Cox (2008). Milidonis et al. (2011) proposed to model the log change rate of the mortality index with a regime switching model

Table 5.5: Comparison of AIC and BIC Criteria for RS-GBM, GMB and Model (5.16)

Model	AIC	BIC
RS-GBM	-390.50	-374.98
GMB	-383.06	-370.14
Model (5.16)	-404.95	-389.38

between two geometric Brownian motions (RS-GBM). They let $Z_t = \log(q_t/q_{t-1})$ and defined a Markov process with two states such that ${}^1Z_t = \mu_1 t + \sigma_1 W_t$ and ${}^2Z_t = \mu_2 t + \sigma_2 W_t$, where W_t is a standard Brownian motion process. Therefore, the mortality index changes according to two geometric Brownian motions. Lin and Cox (2008) use the geometric Brownian motion (GBM) with a discrete Markov process to model the mortality index. Specifically, the dynamic of mortality index at time t, when there is no one-time catastrophic event, is modeled as $d\tilde{q}_t = \alpha dt + \sigma \tilde{q}_t dW_t$, where α and σ represent the drift term and the instantaneous volatility of the mortality index, respectively. Also, the number of catastrophic events during year t is described by a discrete Markov chain, N_t with $N_0 = 0$ and

$$N_{t+1} = \begin{cases} N_t + 1 & \text{with probability } p, \\ N_t & \text{with probability } 1 - p, \end{cases} \tag{5.17}$$

where p is the probability of having a mortality jump during the period $[t, t+1]$. The percentage change in the mortality rate due to a random shock is denoted by $R_t - 1$. Lin and Cox (2008) considered a log-normal distribution with parameters r_1 and r_2 to model R_t such that $R_t = e^{r_1 + r_2 U_t}$, where U_t is a standard normal random variable. The mortality index q_t is then modeled as:

$$q_t = \begin{cases} \tilde{q}_t R_t & \text{with probability } p, \\ \tilde{q}_t & \text{with probability } 1 - p. \end{cases} \tag{5.18}$$

We fit our data over the period of $1900 - 1998$ to models RS-GBM and GBM. The AIC and BIC criteria for RS-GBM, GBM and our model in (5.16) are summarized in Table 5.5. Both criteria support the fitted model in (5.16).

5.7.1 The VLMC Fit

We fit the VLMC model as explained Section 5.3 to a binary time series of the highly fatal observed influenza pandemics from 1500-1998 that includes influenza pandemics in 1557, 1580, 1729, 1889, 1918. The estimated order of the Markov chain and the estimated number of leaves are $\hat{p} = 178$ and 33, respectively. The estimated context tree size, $\hat{\tau}$, is 390. The fitted VLMC could correctly predict four out of five considered severe influenza pandemics.

5.8 Swiss Re Bond

This section explains the design of the Swiss Re bond. In December 2003, Swiss Re issued a longevity bond with a principal of $400 million that had a maturity of three years. It provides the investors with quarterly coupons of three-month U.S. dollar LIBOR plus 135 basis points. The principal repayment depends on the performance of the mortality index q_t, for the years $2004, 2005, 2006$. Specifically, the loss at year t ($loss_t$), is given by:

$$loss_t = \begin{cases} 0 & if \quad q_t \leq 1.3q_0, \\ 1 - \dfrac{1.5q_0 - q_t}{0.2q_0} & if \quad 1.3q_0 < q_t \leq 1.5q_0, \\ 1 & if \quad q_t > 1.5q_0, \end{cases} \tag{5.19}$$

where q_t is the mortality index in year t defined in (5.1) and q_0 is the level of the mortality index in 2002. Equation (5.19) shows that the amount of principal paid to the investors is reduced as the realized mortality index exceeds the threshold of $1.3q_0$ and will be completely exhausted for a mortality index greater than 150% of the 2002 level. For this bond, there will be 12 coupons (i.e. quarterly during 3 years of the bond life time) with a coupon value of

$$C_j = \begin{cases} \Big(\dfrac{S + L_j}{4}\Big).F & if \quad j = \frac{1}{4}, \frac{2}{4}, ..., \frac{11}{4}, \\ \Big(\dfrac{S + L_j}{4} + \alpha\Big).F & if \quad j = 3, \end{cases} \tag{5.20}$$

where S and L_j represent the spread values and the LIBOR rates, respectively, F is the face value of the bond and α is defined as

$$\alpha = max\Big(0, 1 - \sum_{t=2004}^{2006} loss_t\Big). \tag{5.21}$$

Therefore, the discounted cash flow (DC) of the payments is equal to:

$$DC(r) = \sum_{i=1}^{12} \frac{C_{\frac{i}{4}}}{(1 + \frac{r}{4})^i}, \tag{5.22}$$

where r is the nominal annual interest rate.

5.8.1 Estimation of the Market Price

Equations (5.20) and (5.21) show a random payoff for the Swiss Re bond at the maturity of the contract. Also, we do not have access to any tradable mortality index to hedge the portfolio for this mortality derivative. As a result, we cannot use the complete markets approach to price the Swiss Re bond as mentioned in Lin and Cox (2008). Here, we can apply Wang's transform, see Wang (2002), to

handle these problems. Generally speaking, this method requires to first transform the distribution function of the random variable in (5.21). Then, a discounted expected value under this new transform distribution provides the price of the mortality bond. The Wang transform is defined as:

$$F_X^*(x) = \Phi[\Phi^{-1}(F_X(x)) + \lambda], \tag{5.23}$$

where X is a random payment with a distribution function F, Φ is the standard normal cumulative distribution function and λ is the market price that can be considered as the unhedgeable risk factor in the insurance industry. In practice, we do not know the true distribution function, F, in Equation (5.23) and it should be estimated. Therefore, Wang (2002) proposed the two-factor transformation as:

$$F_X^*(x) = Q[\Phi^{-1}(F_X(x)) + \lambda], \tag{5.24}$$

where Q is the t-distribution. In order to use (5.24) to price the Swiss Re bond, we should first estimate the distribution function of the random variable α in (5.21). Hence, we must project the mortality index, q_t in (5.16) over the period of $2000 - 2006$. As mentioned before, to fit the model (5.16), we used the data over the period of $1900 - 1998$. Now, we explain how to simulate the mortality index for $t = 100 - 107$:

- **Step 1**: Generate $P_j(t)$ in (5.16) for $j = 0, 1, 2$ and $t = 100 - 107$ by using the recursive formula given in the Appendix A.4.
- **Step 2**: Simulate ε_t by generating a normal random variable ν_t with mean zero and standard deviation 0.053 that was estimated in (5.16). The starting value of the process should be the final residual[1] of the fitted model in (5.16).
- **Step 3**: Set $Flu_t = 0$ for $t = 100 - 104$ as we have not observed any severe influenza pandemic during $1999 - 2003$.
- **Step 4**: Use the fitted VLMC model in Subsection 5.7.1 to simulate the occurrence of the random variable Flu_t in (5.16) over the period of $2004 - 2006$.
- **Step 5**: Plug in the obtained predictable variables $P_j(t)$, Flu_t and ε_t in (5.16) to predict q_t over $1999 - 2006$.

We run 50,000 simulations and evaluate the losses in (5.19). Lin and Cox (2008) approximate the payment at maturity by defining $q = max(q_{2004}, q_{2005}, q_{2006})$. This approach requires estimating the empirical distribution function of q that should

[1]Residuals is the difference of the observed q_t and the fitted value in (5.16)

be next transformed by using Wang's transform (5.24). In this chapter, as an alternate to the first approach, we first find the empirical probability distribution function of the total loss during $2004, ..., 2006$, i.e. $F_X(x)$ in (5.23), where $X = \sum_{t=2004}^{2006} loss_t$. Then, we transform $F_X(x)$ by using a starting value of the market price λ (that will be solved later), i.e.

$$F_X^*(x) = Q[\Phi^{-1}(F_X(x)) - \lambda]. \tag{5.25}$$

It is now possible to find the expected value of α in (5.21) under the transformed distribution function as follows:

$$\mathbb{E}_\lambda^*(\alpha) = 1 - \int_0^1 (1 - F_X^*(x))dx, \tag{5.26}$$

where $\mathbb{E}_\lambda^*$ is the expected value with respect to the transformed distribution function $F_X^*(x)$, given the market price λ. In addition, we evaluate the expected value of the discounted cash flow of the payments, $\mathbb{E}_\lambda^*\big[DC(r)\big]$ in (5.22) with the par spread of the Swiss Re bond $s = 1.35\%$, the face value (F) and constant LIBOR rates as $L_j = r = 1\%$ for $j = 1, 2, ..., 12$. Finally, we numerically solve the equation of $\mathbb{E}_\lambda^*\big[DC(r)\big] - F = 0$ for λ to estimate the market price as $\hat{\lambda} = -0.043$. Obtaining the loss using the approximation, i.e. $q = max(q_{2004}, q_{2005}, q_{2006})$, leads to the same results.

Wang (2004) estimates the average market price of risk for catastrophic risks as -0.45. Our estimated market price is lower than the one obtained in Lin and Cox (2008) in terms of absolute value. Therefore our analysis shows that, unlike the conclusion given in Lin and Cox (2008), the Swiss Re company did not pay high risk premiums to the investors despite its extremely successful introduction in 2003. It means that by assuming $r = 1\%$ the mortality bond spread of 1.35% was fair. The difference may be due to the fact that we used the weighted mortality index while Lin and Cox (2008) used the U.S. age adjusted death rates in order to estimate the market price.

In order to analyze the sensitivity of the market price with respect to the interest rate, we fit our model using different values of r, ranging from 1% to 4.5 %. Table 5.6 gives the estimated market prices for the selected interest rates. Based on this table, as r increases, the estimated market price rises. This can be explained as follows. For a fixed value of r, $\mathbb{E}_\lambda^*\big[DC(r)\big]$ is a decreasing function of λ for the assumed spread. This is true since $F_X^*(x)$ decreases as λ rises in (5.25). As a result $\mathbb{E}_\lambda^*(\alpha)$ decreases in (5.26). Consequently, $\mathbb{E}_\lambda^*\big[DC(r)\big]$ declines. In addition, we need to show that for $\lambda_1 = -0.043$ and $0.01 \leq r \leq 0.045$, $\mathbb{E}_{\lambda_1}^*\big[DC(r)\big]$ is an increasing function of r. To show this property, we first obtain $\mathbb{E}_{-0.043}^*(\alpha) = 0.9586706$.

Table 5.6: Estimated Market Price vs. Different Values of Interest Rates

$r(\%)$	1	1.5	2	2.5	3	3.5	4	4.5
$\hat{\lambda}$	-0.043	-0.038	-0.033	-0.028	-0.023	-0.018	-0.013	-0.008

Next, we take the derivative of

$$\mathbb{E}^*_{\lambda_1}\big[DC(r)\big] = \frac{F}{4}\sum_{i=1}^{12}\frac{0.0135+r}{(1+\frac{r}{4})^i} + F\times\frac{0.9586706}{(1+\frac{r}{4})^{12}}, \tag{5.27}$$

with respect to r as follows:

$$\frac{d}{dr}\mathbb{E}^*_{\lambda_1}\big[DC(r)\big] = \frac{F}{4}\Big(\sum_{i=1}^{12}\frac{1}{(1+\frac{r}{4})^i} - \frac{1}{4}\sum_{i=1}^{12}\frac{i(0.0135+r)}{(1+\frac{r}{4})^{i+1}}\Big) - \frac{3F(0.9586706)}{(1+\frac{r}{4})^{13}}. \tag{5.28}$$

We then try to find the roots of Equation (5.28) by using Mathematica 5.1 to obtain:

$$\begin{aligned}\frac{d}{dr}\mathbb{E}^*_{\lambda_1}\big[DC(r)\big] \;=\; & F.(-0.0135)\times(-0.629346+r)\times(14.2515-0.915218r+r^2)\times\\ & (39.942+4.09275r+r^2)\times(71.0296+10.9624r+r^2)\times\\ & (98.7726+17.3423r+r^2)\times(115.062+21.1472r+r^2)\times\\ & (4+r)^{-13}.\end{aligned} \tag{5.29}$$

The only positive critical root on the real line is $r = 0.629346$. Therefore, the expected value in (5.27) is increasing for all $r < 0.629346$, as shown in Figure 5.6. In this plot, the y-axis represents the expected value of discounted cash flows for $\lambda_1 = -0.043$ in (5.27) and the x-axis represents the interest rates, r ranging from 0.01 to 0.9. So we can conclude that $\mathbb{E}^*_{\lambda_1}\big[DC(r)\big]$ is an increasing function of r for $0.01 \le r \le 0.045$.

We are now in the position to justify Table 5.6. As we have already shown, for $r_1 = 0.01$, the estimated market price is $\lambda_1 = -0.043$, obtained numerically by solving $\mathbb{E}^*_{\lambda_1}\big[DC(r_1)\big] - F = 0$. If we increase r_1 to $r_2 \le 0.045$, then by the above arguments, we must have $\mathbb{E}^*_{\lambda_1}\big[DC(r_2)\big] \ge \mathbb{E}^*_{\lambda_1}\big[DC(r_1)\big]$ and that leads to $\mathbb{E}^*_{\lambda_1}\big[DC(r_2)\big] \ge F$. Therefore, to estimate the new market price λ_2 corresponding to r_2, we should decrease the expected value of discounted cash flows by increasing the market price, such that it is equal to the face value of the contract, i.e. F. This means that λ_2 should be greater than λ_1.

To summarize, the results given in this part show that by increasing the interest rates, the estimated market price will not change substantially.

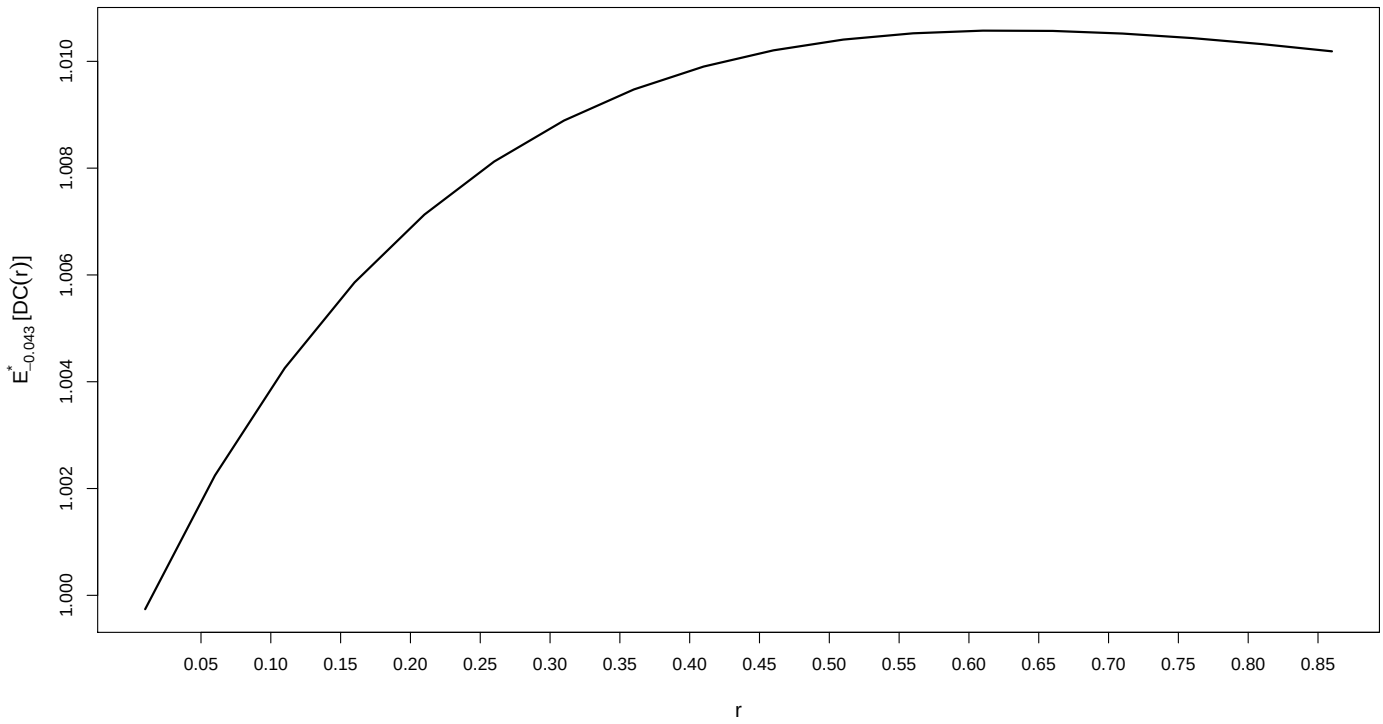

Figure 5.6: Expected value of discounted cash flows for $\lambda_1 = -0.043$ in (5.27) (y-axis) v.s. the interest rates $0.01 \leq r \leq 0.9$. (x-axis)

5.9 Conclusions

In this chapter, we consider modeling the mortality index by combining variable length Markov chains and the generalized least squares technique. First, the mortality index has been constructed based on the design of the Swiss Re bonds. Then, we develop a polynomial regression model using the generalized least squares. An AR(1) process has been identified as the correlation structure of the error terms. Next, we fit our model and compare it to the models proposed by Milidonis et al. (2011) and Lin and Cox (2008). Both the AIC and BIC criteria support the suggested model in this chapter. In addition, we estimate the market price of risk for the Swiss Re bond by performing a simulation study. Our estimation of the market price was lower (in absolute value) than the estimated market price in Lin and Cox (2008). In summary, we suggest modeling the error terms as a time series model.

Conclusions

The contributions of the current thesis are divided into four main parts. In the first part, we add a random variable that follows a one-dimensional OU process to the force of mortality proposed by Ballotta and Haberman (2006). Similarly to Renshaw et al. (1996), a Poisson GLM is used to model the force of mortality that includes both age effects and time trends. The residual deviance is minimized to estimate coefficient of the explanatory variables and the process. Some examples explain how to estimate and forecast the mortality rates in details.

In the 2nd part, we extend the idea of including a perturbed term by conditionally adding an α-stable Lévy subordinator in the force of mortality to capture the effect of shocks in mortality. In particular, we focus on gamma process and provide some examples to support the proposed approach. The price of whole life annuities due are compared based on RHH model, our suggested model and the current life table. The results illustrate some over/under estimations in the prices of annuities by the RHH model.

In the 3rd part, we first review the CBD model that relies on the Gaussian distribution to model the error terms. The normality assumption is tested and rejected for the considered data set. Bivariate GH distributions are proposed to model the error terms in the CBD model. We show how the suggested model can provide a better fit to the considered mortality data based on AIC, BIC, LLF and the likelihood ratio test. In addition, we evaluate the out of sample performance of the proposed model by obtaining MAPE. These facts support the developed model according to GH distributions. Furthermore, we develop an alternative representation for the multivariate normal inverse Gaussian process.

In the last part, we study a catastrophic mortality bond issued by Swiss Re. The mortality index is constructed according to their design. The incidence of catastrophic events (severe influenza pandemics) is modeled based on VLMC approach. We use the generalized least squares technique to model the evolution of mortality index. In this model, we add the time effects based on orthogonal polynomials. We compare our proposed model with RS-GBM model introduced by Milidonis et al. (2011) as well as GMB model suggested by Lin and Cox (2008).

For the future work, it may be possible to consider credit risk modeling to estimate the market price of risk for the companies with lower credit rating than Swiss Re. Here, the liabilities should be modeled by taking into account the rating of the company.

References

[1] Ahmadi, S. S., Gaillardetz, P., 2014. *Two Factor Stochastic Mortality Modeling with Generalized Hyperbolic Distribution*. Journal of Data Science 12, 1, 1-18.

[2] Ainou, V., 2011. *A two factor Lévy model for stochastic mortality*. Working paper, available on-line at: http://iriaf.univ-poitiers.fr/colloque2011/article/v1s1a3.pdf.

[3] Anderson, M., Ashwood, F., 1985. *Projection of mortality rates for the elderly*. Population Trends 42, 22-29.

[4] Applebaum, D., 2009. *Lévy Process and Stochastic Calculus*. Cambridge University Press.

[5] Ballotta, L., Haberman, S., 2003. *Valuation of guaranteed annuity conversion options*. Insurance: Mathematics and Economics 33, 87-108.

[6] Ballotta, L., Haberman, S., 2006. *The fair valuation problem of guaranteed annuity options: The stochastic mortality environment case*. Insurance: Mathematics and Economics 38, 195-214.

[7] Barndorff-Nielsen, O. E., 1997. *Normal inverse Gaussian distribution and stochastic volatility modelling*. Scandinavian Journal of Statistics 24, 1-13.

[8] Bauer,. D., Kramer, F. W., 2008. *Risk and valuation of mortality contingent catastrophe bonds*. Discussion paper PI-0805, Pensions Institute, Cass Business School, City University, United Kingdom.

[9] Biffis, E., 2005. *Affine processes for dynamic mortality and actuarial valuation*. Insurance Mathematics and Economics 37, 443-468.

[10] Booth, H., Maindonald, J., Smith, L., 2002. *Applying Lee-Carter under conditions of variable mortality decline*. Population Studies 56, 325-336.

[11] Booth, H., Tickle, L., 2008. *Mortality modeling and forecasting: A review of methods*. Annals of Actuarial Science 3 (1&2), 3-43.

[12] Bowers, N. L., Gerber, H. U., Hickman, J. C., Jones, D. A., Nesbitt, C. J., 1997. *Actuarial Mathematics.* Society of Actuaries.

[13] Boyle, P. P., Hardy, M. R., 2003. *Guaranteed annuity options.* ASTIN Bulletin 33, 125-152.

[14] Brouhns, N., Denuit, M., Vermunt, J. K., 2002. *A Poisson log-bilinear regression approach to the construction of projected life tables.* Insurance: Mathematics and Economics 31, 373-393.

[15] Buer, M. C., 1926. *Health, Wealth and Population in the Early Days of the Industrial Revolution.* George Routledge & Sons.

[16] Burnham, K. P., Anderson, D. R., 2002. *Model Selection and Multi-model Inference: A Practical Information-theoretic Approach.* Springer-Verlag.

[17] Butler, W. J., Park, R. M., 1987. *Use of logistic regression model for the analysis of proportionate mortality data.* American Journal of Epidemiology 125, 3, 534-535.

[18] Cairns, A.J.G., Blake, D., Dowd, K., 2006. *A two-factor model for stochastic mortality with parameter uncertainty: Theory and calibration.* Journal of Risk and Insurance 73, 687-718.

[19] Cairns, A.J.G., Blake, D., Dowd, K., 2008. *Modelling and management of mortality risk: A review.* Scandinavian Actuarial Journal, 2-3, 79-113.

[20] Cairns, A.J.G., Blake, D., Dowd, K., Coughlan, G. D., Epstein, D., Khallaf-Allah, M., 2011. *Mortality density forecasts: An analysis of six stochastic mortality models.* Insurance: Mathematics and Economics 48, 355-367.

[21] Cairns, A.J.G., Blake, D., Dowd, K., Coughlan, G.D., Epstein, D., Ong, A., Balevich, I., 2009. *A quantitative comparison of stochastic mortality models using data from England & Wales and the United States.* North American Actuarial Journal 13, 1, 1-35.

[22] Cannon, E., 2010. *Estimation and pricing with the Cairns-Blake-Dowd model of mortality.* Discussion paper PI-1003, Pensions Institute.

[23] Chambers, J. M., Hastie, T. J., 1993, *Statistical Models in S.* Chapman and Hall/CRC.

[24] Chen, H., Cox, S. H., 2009. *Modelling mortality with jumps: Application to mortality securitization.* Journal of Risk and Insurance 76, 3, 727-751.

[25] Cont, R., Tankov, P., 2004. *Financial Modelling with Jump Processes*. Chapman and Hall/CRC.

[26] Currie, I. D., Durban, M., Eilers, P. H. C.,2004. *Smoothing and forecasting mortality rates*. Statistical Modelling 4, 279-298.

[27] Currie, I. D., 2006. *Smoothing and forecasting mortality rates with P-splines*. Talk given at the Institute of Actuaries, June 2006. Available at http://www.ma.hw. ac.uk/∼iain/research/talks.html.

[28] Delwarde, A., Denuit, M., Eilers, P., 2007. *Smoothing the Lee-Carter and Poisson log- bilinear models for mortality forecasting: A penalized log-likelihood approach*. Statistical Modelling 7, 29-48.

[29] De Jong., P., Heller, G. Z., 2008. *Generalized Linear Models for Insurance Data*. International Series on Actuarial Science, Cambridge University Press.

[30] De Moivre, A., 1725. *In Annuities on Lives*. 2nd edn., W.P. London.

[31] Deng, Y., Brockett, P. L., MacMinn, R. D., 2012. *Longevity/mortality risk modeling and securities pricing*. The Journal of Risk and Insurance 79, 3, 697-721.

[32] Doray, L. G., 2008. *Inference for logistic-type models for the force of mortality*. International Symposium on Living to 100 and Beyond, SOA monograph M-LI08-01, 18p.

[33] Doornik, J. A., Hansen, H., 2008. *An omnibus test for univariate and multivariate normality.* Oxford Bulletin of Economics and Statistics 70, 927-939.

[34] Eberlein, E., Raible, S., 1999. *Term structure models driven by general Lévy processes*. Mathematical Finance 9, 1, 31-53.

[35] Gompertz, B., 1825. *On the nature of the function expressive of the law of human mortality, and on a new mode of determining the value of life contingencies*. Philosophical Transactions of Royal Society, (Series A) 115, 513-583.

[36] Haberman, S., Renshaw, A. E., 1996. *Generalized linear models and actuarial science*. The Statistician 45, 4, 407-436.

[37] Haberman, S., Renshaw, A. E., 2009. *On age-period-cohort parametric mortality rate projections*. Insurance: Mathematics and Economics 45, 255-270.

[38] Hainaut, D., Devolder, P., 2008. *Mortality modelling with Lévy processes*. Insurance: Mathematics and Economics 49, 409-418.

[39] Heligman, L., Pollard, J. H., 1980. *The age pattern of mortality.* Journal of the Institute of Actuaries 107, 49-80.

[40] Jarque, C. M., Bera, A. K., 1980. *Efficient tests for normality, homoscedasticity and serial independence of regression residuals.* Economics Letters 6, 3, 255-259.

[41] Jong, P. D., Tickle, L., 2006. *Extending the Lee-Carter model of mortality projection.* Mathematical Population Studies 13, 1-18.

[42] Kannisto, V. 1992. Presentation at a workshop on old-age mortality, Odense University, Odense, Denmark.

[43] Kennedy, W. J. Jr., Gentle, J. E., 1980. *Statistical Computing.* Marcel Dekker.

[44] Krutov, A., 2010. *Investing in Insurance Risk: Insurance-Linked Securities-A Practitioner's Perspective.* Risk Books.

[45] Kumar, K., 2011. *Modernization-population change.* Encyclopedia Britannica. Available on-line at: http:www.britannica.com/EBchecked/topic/387301/modernization.

[46] Kyprianou, A., E., 2006. *Introductory Lectures on Fluctuations of Lévy Processes with Applications.* Springer.

[47] Lee, R. D., Carter, L. R., 1992. *Modelling and forecasting U.S. mortality.* Journal of the American Statistical Association 87, 419, 659-671.

[48] Lee, R., Miller, T., 2001. *Evaluating the performance of the Lee-Carter model for forecasting mortality.* Demography 38, 537-549.

[49] Lin, Y., Cox, S. H., 2008. *Securitization of catastrophe mortality risks.* Insurance: Mathematics and Economics 42, 628-637.

[50] Ljung, G. M., Box, G. E. P., 1978. *On a measure of lack of fit in time series models.* Biometrika 65, 297-303.

[51] Mächler, M., Bühlmann, P., 2004. *Variable length Markov chains: methodology, computing, and software.* Journal of Computational and Graphical Statistics 13, 2, 435-455.

[52] Makeham, W. M., 1860. *On the law of mortality and the construction of annuity tables.* Journal of Institute of Actuaries 8, 301-310.

[53] McCullagh, P., Nelder, J. A., 1989. *Generalized Linear Models*. 2nd edn., Chapman and Hall.

[54] McLeod, A. I., W. K. Li, 1983. *Diagnostic checking ARMA time series models using squared residual autocorrelations*. Journal of Time Series Analysis 4, 269-273.

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

[56] Milidonis, A., Lin, Y., Cox,. S. H., 2011. *Mortality regimes and pricing*. North American Actuarial Journal 15, 2, 266-289.

[57] Oksendal, B., Sulem, A., 2010. *Applied Stochastic Control of Jump Diffusions*. 2nd ed., Springer.

[58] Olivieri, A., 2001. *Uncertainty in mortality projections: an actuarial perspective*. Insurance: Mathematics and Economics 29, 231-245.

[59] Perks, W., 1932. *On some experiments in the graduation of mortality statistics*. Journal of the Institute of Actuaries 63, 1240.

[60] Pitacco, E., Denuit, M., Haberman, S., Olivieri, A., 2009. *Modelling Longevity Dynamics for Pensions and Annuity Business*. Oxford University Press.

[61] Pitacco, E., 2003. *Survival models in actuarial mathematics: From Halley to longevity risk*. Invited talk at 7th International Congress Insurance: Mathematics and Economics, ISFA, Lyon.

[62] Plat, R., 2009. *On stochastic mortality modeling*. Insurance: Mathematics and Economics 45, 393-404.

[63] Raftery, A. E., 1993. Bayesian Model Selection in Structural Equation Models. In *Testing Structural Equation Models*, ed. K. A. Bollen & J. S. Long, 163-180.

[64] Raftery, A. E., 1995. *Bayesian model selection in social research*. Sociological Methodology 25, 111-163.

[65] Renshaw, A. E., 1991. *Actuarial graduation practice and generalised linear and non-linear models*. Journal of the Institute of Actuaries 118, 295-312.

[66] Renshaw, A. E., Haberman, S., Hatzoupoulos, P., 1996. *The modelling of recent mortality trends in United Kingdom male assured lives*. British Actuarial Journal 2 , II, 449-477.

[67] Renshaw, A. E., Haberman, S., 2000. *Modelling for mortality reduction factors (Actuarial Research Paper No. 127)*. Department of Actuarial Science and Statistics, City University, London.

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

[69] Renshaw, A. E., Haberman, S., 2003b. *Lee-Carter mortality forecasting: a parallel generalised linear modelling approach for England and Wales mortality projections*. Applied Statistics 52, 119-137.

[70] Renshaw, A. E., Haberman, S., 2003c. *On the forecasting of mortality reduction factors*. Insurance: Mathematics and Economics 32, 379-401.

[71] Renshaw, A. E., Haberman, S., 2006. *A cohort-based extension to the Lee-Carter model for mortality reduction factors*. Insurance: Mathematics and Economics 38, 556-570.

[72] Rorabaugh, W. J., Critchlow, D. T., Baker, P., 2004. *America's Promise: a Concise History of the United States*. Rowman and Littlefield.

[73] Royston, P., 1982a. *Algorithm AS 181: The W test for normality*. Applied Statistics 31, 176-180.

[74] Royston, P., 1982b. *An extension of Shapiro and Wilk's W test for normality to large samples.* Applied Statistics 31, 115-124.

[75] Royston, P., 1995. *A remark on Algorithm AS 181: The W test for normality.* Applied Statistics 44, 547-551.

[76] Schoutens, W., 2003. *Lévy Process in Finance Pricing Financial Derivatives.* Wiley Series in Probability and Statistics.

[77] Sithole, T. Z., Haberman, S., Verrall, R. J., 2000. *An investigation into parametric models for mortality projections,with applications to immediate annuitants' and life office pensioners' data*. Insurance: Mathematics and Economics 27, 285-312.

[78] Taubenberger J., Morens D., 2006. *1918 influenza: the mother of all pandemics*. Emerging Infectious Diseases 12, 1, 15-22.

[79] Taubenberger, J., Morens, D., 2009. *Pandemic influenza - including a risk assessment of H5N1*. Revue Scientifique et Technique 28, 1, 187-202.

[80] Wang, C.W., Huang, H.C., Liu, I.C., 2011. *A quantitative comparison of the Lee-Carter model under different types of non-Gaussian innovations*. The Geneva Papers on Risk and Insurance 36, 675-696.

[81] Wang, S.S., 2002. *A universal framework for pricing financial and insurance risks*. ASTIN Bulletin 32, 2, 213234.

[82] Wang, S. S., 2004. *Cat bond pricing using probability transforms* (Insurance and the State of the Art in Cat Bond Pricing). The Geneva Papers on Risk and Insurance - Issues and Practice 19-29 (special issue).

[83] Webster, R. G., Walker, E. J., 2003. *Influenza*. American Scientist 91, 2, 122-129.

[84] Weibull, W., 1939. *A statistical theory of the strength of materials*. Ingeniörs Vetenskaps Akademien Handligar 151, 1-45.

[85] Wilmoth, J. R., Andreev, K., Jdanov, D., Glei, D. A., 2007. *Methods Protocol for the Human Mortality Database*. Available at: http://www.mortality.org/Public/Docs /MethodsProtocol.pdf.

Appendix A

A.1 Lévy Processes: Preliminaries

Definition A.1. *We say that a real-valued random variable Θ has an infinitely divisible distribution if for each $n = 1, 2, ...$ there exist a sequence of i.i.d. random variables $\Theta_{1,n}, ..., \Theta_{n,n}$ such that*

$$\Theta \stackrel{\mathrm{d}}{=} \Theta_{1,n} + ... + \Theta_{n,n},$$

where $\stackrel{\mathrm{d}}{=}$ is equality in distribution.

Definition A.1 can also be written based on the characteristic exponent; suppose that Θ has characteristic exponent $\Psi(u) := -log\mathbb{E}(e^{iu\Theta})$ for all $u \in \mathbb{R}$. Then Θ has an infinitely divisible distribution if for all $n \geq 1$ there exists a characteristic exponent of a probability distribution, say Ψ_n, such that $\Psi(u) = n\Psi_n(u)$ for all $u \in \mathbb{R}$. The expression for the characteristic exponent Ψ of an infinitely divisible distribution will be stated in the following theorem, known as the Lévy-Khintchine formula.

Theorem A.1. *A probability law μ of a real-valued random variable is infinitely divisible with characteristic exponent Ψ,*

$$\int_{\mathbb{R}} e^{i\theta x}\mu(dx) = e^{-\Psi(\theta)},$$

for all $\theta \in \mathbb{R}$, if and only if there exists a triplet (a, σ, Π), where $a \in \mathbb{R}$, $\sigma \geq 0$ and Π is a measure concentrated on $\mathbb{R}\backslash\{0\}$ satisfying $\int_{\mathbb{R}}(1 \wedge x^2)\Pi(dx) < \infty$, such that

$$\Psi(\theta) = -log\mathbb{E}(e^{i\theta X}) = ia\theta + \frac{1}{2}\sigma^2\theta^2 + \int_{\mathbb{R}}(1 - e^{i\theta x} + i\theta x\chi_{(|x|<1)})\Pi(dx), \quad \forall\theta \in \mathbb{R},$$

where

$$\chi_{(|x|<1)} = \begin{cases} 1, & |x| < 1 \\ 0, & o.w. \end{cases}$$

Definition A.2. *The measure Π is called the Lévy (characteristic) measure.*

We should mention that for any $t > 0$, Z_t in Definition 2.1 is a random variable belonging to the class of infinitely divisible distributions. Moreover the Lévy process Z_t has the property that for all $t \geq 0$, $E(e^{i\theta Z_t}) = e^{-t\Psi(\theta)}$, where $\Psi(\theta) := \Psi_1(\theta)$ is the characteristic exponent of Z_1, which has an infinitely divisible distribution.

Definition A.3. *Let $(S, \mathcal{S}, \eta)$ be an arbitrary σ-finite measure space and $N : \mathcal{S} \to \{1, 2, ..., \} \bigcup \{\infty\}$ in such a way that $\{N(A) : A \in \mathcal{S}\}$ is a family of random variables defined on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Then N is called a Poisson random measure on $(S, \mathcal{S}, \eta)$ (or sometimes a Poisson random measure on $\mathcal{S}$ with intensity η) if*

(i) *for mutually disjoint $A_1, ..., A_n$ in $\mathcal{S}$, the variables $N(A_1), ..., N(A_n)$ are independent,*

(ii) *for each $A \in \mathcal{S}$, $N(A)$ is Poisson distributed with parameter $\eta(A)$ (here we allow $0 \leq \eta(A) \leq \infty$),*

(iii) *$\mathbb{P}$-almost surely N is a measure.*

Theorem A.2. *A Lévy process with Lévy-Khintchine exponent corresponding to the triplet (a, σ, Π) has paths of bounded variation if and only if*

$$\sigma = 0 \quad \textit{and} \quad \int_{\mathbb{R}} (1 \wedge |x|)\Pi(dx) < \infty. \tag{A.1}$$

The finiteness property of the integral in (A.1) allows us to rewrite the Lévy-Khintchine exponent of any such bounded variation process as

$$\Psi(\theta) = -id\theta + \int_{\mathbb{R}} (1 - e^{i\theta x})\Pi(dx), \tag{A.2}$$

where the constant $d \in \mathbb{R}$ relates to the constant a and Π via

$$d = -\Big(a + \int_{|x|<1} x\Pi(dx)\Big).$$

Hence, in this case, the Lévy process can be written in the form of

$$Z_t = dt + \int_{[0,t]} \int_{\mathbb{R}} zN(ds, dz), \quad t \geq 0. \tag{A.3}$$

Definition A.4. *A Lévy process is a subordinator if and only if $\Pi(-\infty, 0) = 0$, $\int_{(0,\infty)} (1 \wedge x)\Pi(dx) < \infty$, $\sigma = 0$, $d = -\Big(a + \int_{(0,1)} x\Pi(dx)\Big) \geq 0$.*

Condition $\Pi(-\infty, 0) = 0$ means that the Lévy process has no negative jumps. If we further assume that $\int_{(0,\infty)} (1 \wedge x)\Pi(dx) < \infty$, $\sigma = 0$ and $d \geq 0$ in representation (A.2) of the characteristic exponent, we will obtain that the Lévy process has non-decreasing paths. Hence, Lévy processes whose paths are almost surely non-decreasing (or simply non-decreasing for short) are called subordinators. The gamma process and the inverse Gaussian process are two examples of subordinators.

Theorem A.3. *Let $\mathbf{Y}_t \in \mathbb{R}^n$ be an Itô-Lévy process of the form*

$$d\mathbf{Y}_t = \alpha(t,\omega)dt + \sigma(t,\omega)dB_t + \int_{\mathbb{R}^l} \gamma(t,z,\omega)\bar{N}(dt,dz), \tag{A.4}$$

where $\alpha : [0,T] \times \Omega \to \mathbb{R}^n$, $\sigma : [0,T] \times \Omega \to \mathbb{R}^{n\times m}$, and $\gamma : [0,T] \times \mathbb{R}^l \times \omega \to \mathbb{R}^{n\times l}$ are adapted processes such that the integral exists. Let $f \in C^{1,2}([0,T] \times \mathbb{R}^n; \mathbb{R})$, i.e., the space of all continous functions with first and 2nd derivatives. B_t in (A.4) is an m-dimensional Brownian motion and

$$\begin{aligned}
\bar{N}(dt,dz)^T &= \Big(\bar{N}_1(dt,dz_1),...,\bar{N}_l(dt,dz_l)\Big) \\
&= \Big(N_1(dt,dz_1) - \chi_{(|z_1|<R_1)}\nu_1(dz_1)dt, ..., N_l(dt,dz_l) \\
&\quad -\chi_{(|z_l|<R_l)}\nu_l(dz_l)dt\Big),
\end{aligned}$$

where $\{N_j\}$ are independent Poisson random measures with Lévy measures ν_j coming from l independent (one-dimensional) Lévy processes $\eta_1, ..., \eta_l$ and $\chi_{(|z|<R)}$ is defined as:

$$\chi_{(|z|<R)} = \begin{cases} 1, & |z| < R \\ 0, & otherwise. \end{cases} \tag{A.5}$$

Put $K(t) = f(t, \mathbf{Y}_t)$. Then

$$\begin{aligned}
dK(t) &= \frac{\partial f}{\partial t}dt + \sum_{i=1}^n \frac{\partial f}{\partial y^{(i)}}(\alpha_i dt + \sigma_i dB_t) + \frac{1}{2}\sum_{i,j=1}^n (\sigma\sigma^T)_{ij}\frac{\partial^2 f}{\partial y^{(i)}\partial y^{(j)}}dt \\
&\quad + \sum_{k=1}^l \int_{|z_k<R_k|} \Big\{ f(t, Y_{t^-} + \gamma^{(k)}(t,z_k)) - f(t,Y_{t^-}) \\
&\qquad - \sum_{i=1}^n \gamma_i^{(k)}(t,z_k)\frac{\partial f}{\partial y^{(i)}}(Y_{t^-})\Big\}\nu_k(dz_k)dt \\
&\quad + \sum_{k=1}^l \int_{\mathbb{R}} \Big\{ f(t, Y_{t^-} + \gamma^{(k)}(t,z_k)) - f(t,Y_{t^-})\Big\}\bar{N}_k(dt,dz_k),
\end{aligned} \tag{A.6}$$

where $\gamma^{(k)} \in \mathbb{R}^n$ is column number k of the $n \times l$ matrix $\gamma = [\gamma_{ik}]$ and $\gamma_i^{(k)} = \gamma_{ik}$ is the coordinate number i of $\gamma^{(k)}$.

A.2 Doornik-Hansen's Test

In this section, we briefly explain the multivariate normality test proposed by Doornik and Hansen (2008). Denote a $p \times n$ matrix of n observations on a p-dimensional vector by $X' = (\boldsymbol{x_1}, ..., \boldsymbol{x_n})$ with sample mean $\bar{\boldsymbol{x}} = n^{-1}\sum_{i=1}^n \boldsymbol{x_i}$ and sample covariance matrix $S = n^{-1}\sum_{i=1}^n (\boldsymbol{x_i} - \bar{\boldsymbol{x}})(\boldsymbol{x_i} - \bar{\boldsymbol{x}})'$. Let $V = diag(\hat{\sigma_1}, ..., \hat{\sigma_p})$ be the diagonal matrix which has the variances on the diagonal and obtain the

correlation matrix $C = V^{\frac{-1}{2}} S V^{\frac{-1}{2}}$. The eigenvalues of C can then be used to define a diagonal matrix $\Lambda = diag(\lambda_1, ..., \lambda_p)$. Next, each observation is transformed according to $\boldsymbol{y_i} = H\Lambda^{\frac{-1}{2}} H' V^{\frac{-1}{2}} (\boldsymbol{x_i} - \bar{\boldsymbol{x}})$ to obtain a $p \times n$ transformed matrix $Y' = (\boldsymbol{y_1}, ..., \boldsymbol{y_n})$. Here, the columns of H are the corresponding eigenvectors of C, such that $H^{'} H = I_p$ and $\Lambda = H^{'} C H$. Denote the univariate skewness and kurtosis for each of the p-transformed vectors of n observations by $\boldsymbol{B_1}^{'} = (\sqrt{b_{11}}, ..., \sqrt{b_{1p}})$ and $\boldsymbol{B_2}^{'} = (b_{11}, ..., b_{1p})$, respectively. The multivariate Doornik-Hansen's test statistic is:

$$E_p = \boldsymbol{Z_1}^{'} \boldsymbol{Z_1} + \boldsymbol{Z_2}^{'} \boldsymbol{Z_2},$$

that has approximately a chi-square distribution with $2p$ degrees of freedom, where $\boldsymbol{Z_1}^{'} = (z_{11}, ..., z_{1p})$ and $\boldsymbol{Z_2}^{'} = (z_{21}, ..., z_{2p})$ are determined by (A.7) and (A.8) as follow:

$$\begin{aligned}
\beta &= \frac{3(n^2 + 27n - 70)(n+1)(n+3)}{(n-2)(n+5)(n+7)(n+9)}, \\
w^2 &= -1 + \Big\{ 2(\beta - 1) \Big\}^{\frac{1}{2}}, \\
\delta &= \frac{1}{\{\log(\sqrt{w^2})\}^{\frac{1}{2}}}, \\
y &= \sqrt{b_1} \Big\{ \frac{w^2 - 1}{2} \frac{(n+1)(n+3)}{6(n-2)} \Big\}^{\frac{1}{2}}, \\
z_1 &= \delta \log \{ y + (y^2 + 1)^{\frac{1}{2}} \},
\end{aligned} \tag{A.7}$$

$$\begin{aligned}
\delta &= (n-3)(n+1)(n^2 + 15n - 4), \\
a &= \frac{(n-2)(n+5)(n+7)(n^2 + 27n - 70)}{6\delta}, \\
c &= \frac{(n-7)(n+5)(n+7)(n^2 + 2n - 5)}{6\delta}, \\
k &= \frac{(n+7)(n+5)(n^3 + 37n^2 + 11n - 313)}{12\delta}, \\
\alpha &= a + b_1 c, \\
\chi &= (b2 - 1 - b1) 2k, \\
z_2 &= \Big\{ \Big(\frac{\chi}{2\alpha} \Big)^{\frac{1}{3}} - 1 + \frac{1}{9\alpha} \Big\} (9\alpha)^{\frac{1}{2}}.
\end{aligned} \tag{A.8}$$

A.3 The Generalized Inverse Gaussian Distribution

The $GIG(\lambda, \chi, \psi)$, is defined by

$$f_{GIG}(\omega) = \Big(\frac{\psi}{\chi} \Big)^{\frac{\lambda}{2}} \frac{\omega^{\lambda - 1}}{2K_\lambda(\sqrt{\chi\psi})} \exp \Big\{ -\frac{1}{2} \Big(\frac{\chi}{\omega} + \psi\omega \Big) \Big\}, \quad \omega > 0, \tag{A.9}$$

with parameters satisfying
$(\chi > 0, \psi \geq 0, \quad \lambda < 0)$, or $(\chi > 0, \psi > 0, \quad \lambda = 0)$, or else $(\chi \geq 0, \psi > 0, \quad \lambda > 0)$.
Special cases of the GIG distribution are the gamma distribution, when $\chi = 0$ and $\lambda > 0$, as well as the inverse gamma distribution, with $\psi = 0$ and $\lambda < 0$. We start by giving the definition the inverse Gaussian (IG) random variable, from Schoutens (2003).

Definition A.5. *A random variable Y is said to have an IG distribution with parameters a and b, written as $IG(a, b)$, if it has a characteristic function of the form:*

$$\phi_Y(u; a, b) = \exp\big\{ - a(\sqrt{-2iu + b^2} - b)\big\}.$$

The IG distribution is infinitely divisible and hence we can define the IG process $I = \{I_t, t \geq 0\}$, with parameters $a, b \geq 0$, as the process which starts at zero and has independent and stationary increments such that

$$E[\exp(iuY_t)] = \phi_Y(u; at, b) = \exp\big\{ - at(\sqrt{-2iu + b^2} - b)\big\}.$$

We also mention that for $\lambda = -\frac{1}{2}$ the $GIG(\lambda,\, a^2,\, b^2)$ reduces to the $IG(a, b)$.

In the following theorem, we develop an alternative representation for the multivariate NIG process that is more convenient to perform Monte Carlo simulations especially when the problem is path dependent.

Theorem A.4. *For $t > 0$ define*

$$\boldsymbol{X}_t = \delta^2 \Delta\, \boldsymbol{\beta} I_t + \delta I_d \boldsymbol{W}_{I_t} + t\boldsymbol{\mu}, \tag{A.10}$$

where I_t is an IG process with parameters $a = 1$ and $b = \delta\sqrt{\alpha^2 - \boldsymbol{\beta}^{'}\Delta\boldsymbol{\beta}}$ such that $\alpha,\ \delta > 0$ and $\boldsymbol{\beta} \in \{\boldsymbol{x} \in \mathbb{R}^d : \alpha^2 - \boldsymbol{\beta}^{'}\Delta\boldsymbol{\beta} > 0\}$ with $\Delta \in \{A \in \mathbb{R}^{d\times d} : |A| = 1\}$. Here I_d is a d-dimensional identity matrix, $\boldsymbol{W}_t$ is a d-dimensional Brownian motion without drift, and diffusion parameter Δ and $\boldsymbol{\mu} \in \mathbb{R}^d$. Then $\boldsymbol{X}_t \sim NIG(\alpha, \boldsymbol{\beta}, t\boldsymbol{\mu}, t\delta, \Delta)$.

Proof: For simplicity we first assume that $\boldsymbol{\mu} = \mathbf{0}$, then the conditional distribution of W_{I_t} is a d-dimensional Gaussian, i.e. given I_t, $W_{I_t} \sim N_d(\mathbf{0}, \delta^2\Delta I_t)$. Hence given I_t, $\boldsymbol{X}_t \sim N_d(\delta^2\Delta\boldsymbol{\beta}I_t, \delta^2\Delta I_t)$. Therefore, for any $\boldsymbol{u} \in \mathbb{R}^d$, we have that

$$\mathbb{E}[\exp\{\delta^2\boldsymbol{u}^{'}\boldsymbol{X}_t|I_t\}] = \exp\big\{\delta^2(\boldsymbol{u}^{'}\Delta\boldsymbol{\beta} + \frac{1}{2}\boldsymbol{u}^{'}\Delta\boldsymbol{u})I_t\big\}.$$

Using the characteristic function, or equivalently the moment generating function of an IG process, we obtain

$$\begin{aligned}
\mathbb{E}\big[\exp\{\delta^2 \boldsymbol{u}' \boldsymbol{X}_t\}\big] &= \mathbb{E}\Big[\exp\big\{\delta^2(\boldsymbol{u}'\Delta\boldsymbol{\beta} + \frac{1}{2}\boldsymbol{u}'\Delta\boldsymbol{u})I_t\big\}\Big] \\
&= \exp\Big(-t\delta\Big\{\sqrt{-2\boldsymbol{u}'\Delta\boldsymbol{\beta} - \boldsymbol{u}'\Delta\boldsymbol{u} + \alpha^2 - \boldsymbol{\beta}'\Delta\boldsymbol{\beta}} \\
&\quad -\sqrt{\alpha^2 - \boldsymbol{\beta}'\Delta\boldsymbol{\beta}}\Big\}\Big) \\
&= \exp\Big(-t\delta\Big\{\sqrt{\alpha^2 - (\boldsymbol{\beta}+\boldsymbol{u})'\Delta(\boldsymbol{\beta}+\boldsymbol{u})} \\
&\quad -\sqrt{\alpha^2 - \boldsymbol{\beta}'\Delta\boldsymbol{\beta}}\Big\}\Big). \qquad \text{(A.11)}
\end{aligned}$$

Here (A.11) is shown to be the moment generating function of the multivariate NIG distribution in Barndorff-Nielsen (1997). The location parameter $\boldsymbol{\mu}$ can be added to the process and by the properties of the conditional expectation, Equation (A.10) holds.

A.4 Orthogonal Polynomials

The orthogonal polynomials in (5.10) can be obtained by the following recursion formula:

$$\begin{aligned}
P_{-1}(t) &= 0, \quad P_0(t) = 1, \qquad \text{(A.12)} \\
P_{j+1}(t) &= (t - \rho_{j+1})P_j(t) - \gamma_j P_{j-1}(t), \quad (j = 0, 1, ..., q-1),
\end{aligned}$$

where for $j > 0$,

$$\rho_{j+1} = \frac{\sum_{t=1}^{99} t[p_j(t)]^2}{\sum_{t=1}^{99}[p_{j-1}(t)]^2}, \qquad \text{(A.13)}$$

with $\rho_1 = \bar{t}$ and

$$\gamma_j = \frac{\sum_{t=1}^{99}[p_j(t)]^2}{\sum_{t=1}^{99}[p_{j-1}(t)]^2}, \quad (j = 1, 2, ..., 98), \qquad \text{(A.14)}$$

where $\gamma_0 = 0$.