A Pseudo Non-Parametric Bühlmann Credibility Approach to Modeling Mortality Rates

by

Xiang Luan

B.B.A., University of Wisconsin Madison, 2013

Project Submitted in Partial Fulfillment

of the Requirements for the Degree of

Master of Science

in the

Department of Statistics and Actuarial Science

Faculty of Science

SIMON FRASER UNIVERSITY

Summer 2015

Approval

Name: Xiang Luan

Degree: Master of Science (Actuarial Science)

Title: *A Pseudo Non-Parametric B*ü*hlmann Credibility Approach to Modeling Mortality Rates*

Examining Committee: **Dr. Tim Swartz** (chair)
Professor

Dr. Cary Chi-Liang Tsai
Associate Professor
Senior Supervisor

Dr. Gary Parker
Associate Professor
Supervisor

Dr. Yi Lu
Associate Professor
Internal Examiner

Date Defended: 29 July 2015

Abstract

Credibility theory is applied in property and casualty insurance to perform prospective experience rating, i.e., to determine the future premiums to charge based on both past experience and the underlying group rate. Insurance companies assign a credibility factor Z to a specific policyholder's own past data, and put $1 - Z$ onto the prior mean which is the group rate determined by actuaries to reflect the expected value for all risk classes. This partial credibility takes advantage of both policyholder's own experience and the entire group's characteristics, and thus increases the accuracy of estimated value so that the insurance companies can stay competitive in the market. Faced with its popular applications in property and casualty insurance, this project aims to apply the credibility theory to projected mortality rates from three existing mortality models. The approach presented in this project violates one of the conditions, and thus produces the pseudo non-parametric Bühlmann estimates of the forecasted mortality rates. Numerical results show that the accuracy of forecasted mortality rates are significantly improved after applying the non-parametric Bühlmann method to the Lee-Carter model, the CBD model, and the linear regression-random walk (LR-RW) model. A measure of mean absolute percentage error (MAPE) is adopted to compare the performances in terms of accuracy of mortality prediction.

Keywords: Credibility Theory; Non-Parametric Bühlmann Estimate; Lee-Carter Model; CBD Model; Linear Regression Model; Mortality Rates; MAPE

Dedication

This project is dedicated to my parents for their endless love, unconditional support, and unceasing encouragement.

Acknowledgements

First and foremost I would like to express my sincere gratitude to my supervisor, Dr. Cary Tsai, whose expertise in not only academic literature but also industry experience has helped and taught me immensely in the past two years. His continuous support and comprehensive guidance help me in all the time of research and writing of this project. I would never have been able to finish my project without his tremendous help, and I really appreciate his unlimited patience and kindness to me.

At the same time, I would also like to thank my committee members, Dr. Gary Parker and Dr. Yi Lu, not only for their insightful comments and reviews on this project, but also for the invaluable help they offer to me all the time. They have been supportive in every way including passing on knowledge as well as providing assistance in terms of job hunting and career development.

I wish to thank Dr. Tim Swartz for taking precious time out from his busy schedule to serve as the chair, and my sincere thanks also go to the whole department, all the faculties and staff who provided me all kinds of help and support. I thank my fellow office mates for the stimulating discussions along the way, for all the hard-working time that we worked together, and for all the fun we have had in the last two years. I have been blessed with such a friendly and cheerful group of classmates, office mates, and friends, and we have made a wonderful journey during our two-year studies at Simon Fraser University.

A special acknowledgement goes to my boyfriend Will, who accompanied me throughout the whole writing process and kept providing endless encouragement and valuable technical support.

Table of Contents

Approval ii

Abstract iii

Dedication iv

Acknowledgements v

Table of Contents vi

List of Tables viii

List of Figures ix

1 Introduction 1

1.1 Background & Introduction to Mortalities . . . 2
1.2 Motivation . . . 3

2 Literature Review 5

2.1 Mortalities . . . 5
2.1.1 Lee-Carter Model . . . 5
2.1.2 CBD Model . . . 7
2.1.3 Linear Relational Model . . . 8
2.2 Credibility Theory . . . 9

3 Applying Credibility Theory to Existing Models 12

3.1 Lee-Carter Model . . . 13
3.1.1 Notations & Assumptions . . . 13
3.1.2 Estimation of Model Parameters . . . 14
3.1.3 Model Prediction . . . 16
3.2 CBD Model . . . 18
3.2.1 Notations & Assumptions . . . 18
3.2.2 Estimation of Model Parameters . . . 18
3.2.3 Model Prediction . . . 20
3.3 LR-RW Model . . . 22

3.3.1 Notations & Assumptions 22
3.3.2 Estimation of Model Parameters 23
3.3.3 Model Prediction 23
3.4 Bühlmann Credibility Theory 27
3.4.1 Credibility Estimation 27
3.4.2 Parametric Bühlmann Model 28
3.4.3 Non-parametric Bühlmann Model 30
3.5 Applying Pseudo Non-Parametric Bühlmann Estimation 32
3.5.1 Construction of Non-Parametric Bühlmann Estimation 33
3.5.2 Notation and Interpretations 34

4 Numerical Results **36**
4.1 Model Specification 37
4.1.1 Stage I: Grouped vs Ungrouped Age Span in Models 37
4.1.2 Stage II: Grouped vs Ungrouped Age Span for Bühlmann Estimation . . 39
4.2 Numerical Results 51

5 Applications **55**

6 Conclusion **61**

Bibliography **63**

List of Tables

Table 3.1 Estimation for μ, v and a 31
Table 3.2 Framework for the non-parametric Bühlmann estimate 34

Table 4.1 Summary of 3 forecasting year spans 36
Table 4.2 Framework for Stage I and Stage II 37
Table 4.3 $MAPE^{VW}$ values for 10-year forecasting 51
Table 4.4 $MAPE^{VW}$ reduction ratios for 10-year forecasting 51
Table 4.5 $MAPE^{VW}$ values for 30-year forecasting 52
Table 4.6 $MAPE^{VW}$ reduction ratios for 30-year forecasting 52
Table 4.7 $MAPE^{VW}$ values for 50-year forecasting 53
Table 4.8 $MAPE^{VW}$ reduction ratios for 50-year forecasting 53

List of Figures

Figure 3.1 Framework of the Lee-Carter and CBD models 13
Figure 3.2 Estimates of parameters for the Lee-Carter model 15
Figure 3.3 $ln(\hat{m}_{x,t_0+\tau})$ and $\hat{V}ar\left[ln(\tilde{m}_{x,t_0+\tau})\right]$ for $x = 40$ based on 5 fitting year spans 17
Figure 3.4 Estimates of parameters for the CBD model 19
Figure 3.5 $logit(\hat{q}_{x,t_0+\tau})$ and $\hat{V}ar\left[logit(\tilde{q}_{x,t_0+\tau})\right]$ for $x = 40$ based on 5 fitting year spans 21
Figure 3.6 Framework of the LR-RW model 22
Figure 3.7 Estimates of parameters for the LR-RW model 24
Figure 3.8 $ln(\hat{m}_{x,t_0+\tau})$ and $\hat{V}ar\left[ln(\tilde{m}_{x,t_0+\tau})\right]$ for $x = 40$ based on 5 fitting year spans 26
Figure 3.9 Framework of collecting data for the non-parametric Bühlmann model . 33

Figure 4.1 Framework of the grouped age spans 38
Figure 4.2 $MAPE_j$'s vs $1950+j$ for UK with $T = 10$ and year span = $[1950+j, 2000]$ 40
Figure 4.3 $MAPE_j$'s vs $1950+j$ for US with $T = 10$ and year span = $[1950+j, 2000]$ 42
Figure 4.4 $MAPE_j$'s vs $1950+j$ for $Japan$ with $T = 10$ and year span = $[1950+j, 2000]$ 43
Figure 4.5 $MAPE_j$'s vs $1950+j$ for UK with $T = 30$ and year span = $[1950+j, 1980]$ 44
Figure 4.6 $MAPE_j$'s vs $1950+j$ for US with $T = 30$ and year span = $[1950+j, 1980]$ 45
Figure 4.7 $MAPE_j$'s vs $1950+j$ for $Japan$ with $T = 30$ and year span = $[1950+j, 1980]$ 47
Figure 4.8 $MAPE_j$'s vs $1950+j$ for UK with $T = 50$ and year span = $[1950+j, 1960]$ 48
Figure 4.9 $MAPE_j$'s vs $1950+j$ for US with $T = 50$ and year span = $[1950+j, 1960]$ 49
Figure 4.10 $MAPE_j$'s vs $1950+j$ for $Japan$ with $T = 50$ and year span = $[1950+j, 1960]$ 50

Figure 5.1 $MAPE_x$ for 30-year term life insurance vs age x for US males 57
Figure 5.2 $MAPE_x$ for 30-year endowment vs age x for US males 59
Figure 5.3 $MAPE_x$ for 30-year temporary life annuity immediate vs age x for US males 60

Chapter 1

Introduction

Mortality rate is a measure of the number of deaths in a particular population per unit of time. It is the core concern that all financial institutes need to incorporate when designing not only traditional life insurance products, but also newly emerged mortality-linked financial derivatives. There are many kinds of life insurance products prevalent in the market, with diverse contracts for benefit amounts, durations, and all other conditions specified. Despite of these distinct requirements, they all provide protections, in exchange for premiums, for the insureds from the uncertainty of death occurrence in the future. In this sense, insurance companies heavily rely on their forecasted mortality rates to design life insurance products and determine their premiums. As a consequence, the ability to conduct accurate forecasting of the future mortality rates turns out to be a critical task for life insurance companies: once they are able to make mortality predictions with high-level accuracy, they can accordingly determine competitive premiums, conduct proper reserve analysis, and thus operate a stronger business in the market. However, at the same time, it is quite a challenging task to forecast the future mortality rates due to the random movements and irregular patterns in mortality rates, which seem not to be easily captured in some mortality models. This project employs three mortality models: Lee-Carter model, CBD model, and linear relational model; afterwards we apply the non-parametric Bühlmann credibility approach based on these models to performing estimation with regards to mortality rates. Outputs show that the implementation of the non-parametric Bühlmann credibility approach based on original mortality models can provide higher degree of accuracy and more stable prediction results. Chapter 3 will present a more elaborated demonstration of the mortality models as well as the Bühlmann credibility theory.

All the mortality data applied in this project come from the Human Mortality Database (HMD) at http://www.mortality.org/. HMD is a public database that provides detailed mortality and population information for 37 countries or areas around the world. The database contains calculations of death rates and life tables, which consists of death counts from vital statistics, plus census counts, birth counts, and population estimates from various sources. Again, this project focuses on the topic of mortality forecasting and proposes an approach which embeds the non-parametric Bühlmann credibility theory into a mortality model for forecasting mortality rates.

1.1 Background & Introduction to Mortalities

The procedure of our approach goes as follows: first we apply a mortality model to make mortality forecasting from datasets for an age span and different fitting year spans; thereafter, we implement the non-parametric Bühlmann credibility method using the forecasting results we got to achieve the final estimation. As mentioned above, this projects mainly focuses on three mortality models and collects forecasting values from each of the three. However, the implementation of the non-parametric Bühlmann credibility approach is not restricted to these three models; instead, it can be conducted with any sample data generated from any other alternative models. This is also a major advantage since such approach possesses high-level flexibilities, which make it more applicable in reality. For illustration purpose, in this project we only present illustrations using Lee-Carter model, CBD model, and linear relational model.

The Lee-Carter model proposed by Lee and Carter (1992) has been recognized as the most popular mortality model with wide applications in literature. There are numerous extensions and modifications to the Lee-Carter model with the purpose of improving its forecasting performance. The CBD model proposed by Cairns et al. (2006) was designed for modeling mortality at higher ages based on the historical observations that the logit function of one-year death probability is approximately linear in age for ages beyond 40. The CBD model also contains the dynamic features of a stochastic model as the Lee-Carter model does, and is regarded as a robust model with generally good fit. The linear relational model is an innovative model proposed by Tsai and Yang (2015), which applies linear regression approach to modeling mortality rates from observing a linear relation between two mortality sequences. One of the favorable attributes of the linear relational model is its simple and straightforward construction. There are two versions under the linear relational model in Tsai and Yang (2015): LR-LR model and LR-RW model; in this project we only focus on the LR-RW model.

Credibility theory is a significant cornerstone in property and casualty insurance. One of its popular applications is to conduct statistical estimations towards aggregate loss amount of insurance claims. In a general case, insurers observe loss experience under a risk class and obtain a prior mean for this risk class. At the same time, they might also have gathered a specific policyholder's own experience data. With the chance that the prior mean from the risk class might seem very different from the observed data from the policyholder, insurers have to decide how credible the prior mean is and how much credit they should put onto the policyholder's historical data in order to determine an appropriate premium for this policyholder. In this sense, they apply a credibility factor which represents the weight given to the policyholder's historical data, and assign the rest of the weight to the prior mean from the risk class. Chapter 3 will present a more detailed demonstration on how to construct credibility framework. As mentioned earlier, the credibility theory is one of the most important and popular strategies applied in prospective rate making, and has wide-ranging applications in property and casualty insurance. This project extends the credibility theory to the exercise of mortality forecast area and applies the method with special implementation and interpretation; more explanations are provided in Chapter 3.

1.2 Motivation

Providing mortality forecast with high-degree accuracy as well as stable forecasting performance has critical meanings to both life insurers and reinsurers. Precise forecasting can help reduce mortality/longevity risk and ensure a smoother operation of business. Underestimating future mortality rates would lead to insufficient premium levels while overestimation could severely affect insurers' competitiveness and gradually pull the insurers off the market. Retaining appropriate premium levels takes even more important role at this point since the economic condition is still in downturn. In this sense, the selection of mortality models turns out to be more determinant for life insurance companies to accomplish accurate forecasting.

As the essential foundation for all different kinds of mortality-linked insurance products or financial derivatives, mortality rate suggests critical implications in financial market, and thus this project aims to develop an appropriate approach that can provide highly accurate and stable forecasting results for life insurance companies. There are numerous studies that have been conducted to improve the forecasting performance based on specific mortality models, and we will elaborate more on this in Chapter 2. However, a large number of research studies exclusively focused on one specific mortality model and they achieved better forecasting results by adding more complex factors and thus raising the difficulty levels for practical implementation. This project, on the other hand, concentrates on the forecasted mortality rates that have been already generated from existing mortality models. It provides a simple approach to applying the non-parametric Bühlmann credibility method, and eventually produces satisfactory estimation results. It is a creative approach to forecasting mortality rates with no previous research references, and can bring better forecasting results to life insurance companies.

The structure of this project proceeds as follows: Chapter 2 focuses on literature review on the development of Lee-Carter model, CBD model, and linear relational model; it also elaborates the derivation of credibility theory as well as its current applications and modifications in industry. Chapter 3 presents the three mortality models mentioned above in mathematical formulas, and after a concise introduction to Bühlmann credibility theory, we apply the non-parametric Bühlmann method based on the sample data collected from each of the three models. Numerical results are summarized in tables in Chapter 4; figure comparisons and analysis are also provided to support the conclusion that applying the credibility theory leads to better forecasting results with lower errors. Chapter 5 illustrates several practical applications with our forecasted mortality rates; we apply the estimated outcomes to price three types of life insurance and annuity products, and compare the results with those from the original models, which eventually confirms the conclusion declared in Chapter 4. Chapter 6 is the conclusion and it summarizes the whole project with suggestions and further considerations.

Chapter 2

Literature Review

2.1 Mortalities

2.1.1 Lee-Carter Model

The Lee-Carter mortality model was proposed by Lee and Carter (1992). It is a stochastic model with dynamic features, and it concentrates on two factors: age effect and time effect when modeling and forecasting mortality rates. Although it has been recognized as one of the most popular and effective mortality models that can generally produce quite a good fit and forecasting results, the Lee-Carter model has also been called into questions along the way and consequently modified in order to achieve better forecasting results. For instance, Lee and Miller (2001) examined Lee-Carter model's actual and hypothetical forecast errors. Their results suggested that the Lee-Carter model tends to under-project gains and furthermore, the actual ever-changing age patterns also affect the model fitting. Brouhns et al. (2002) implemented a new computational method for fitting, extrapolating, and improving the Lee-Carter model. They expanded the model from the original Poisson regression setting, applied a time-varying index for mortality forecasting, and updated the model by deriving Brass-type relational model which is quite helpful in terms of estimating the cost of adverse selection in the whole life annuity market.

With the original age-effect and time-trend parameters in question, the Lee-Carter model has been extended and modified to different extents in the past decades. Delwarde et al. (2007) examined the associated parameters in the Lee-Carter model, and discovered the irregular patterns of the age-effect parameters exhibited in most cases, which might lead to deviated forecasting values. To deal with this concern, they developed an approach to smoothing the estimated parameter by applying Poisson log-bilinear models for mortality projection. They also determined the optimal value of the smoothing with the help of cross validation. Mitchell et al. (2013) put forward a modification which utilizes the data with the growth rates of mortality rates, rather than the regular mortality rates as in the original Lee-Carter model. Their numerical results prove that such modification can help the model achieve better prediction performance in terms of capturing mortality dynamics, as well as many of the subsequent variants.

As a predominant method to describe and define the evolution of mortality rates, the Lee-Carter model can perform well in most cases. However, the traditional Lee-Carter model assumes normal distributions in the error terms, while there exists certain situations where the normal distribution assumption can no longer hold, such as the occurrence of unusual worldwide disaster or epidemic. Faced with the truth that such events changed the empirical mortality distribution to a significant extend, Giacometti et al. (2009) included a non-Gaussian distributions for the error terms in the Lee-Carter model so that it can exhibit such deviant patterns, which can not be expressed as normal any more.

Besides the non-Gaussian distribution which helps the model better capture the occurrences of jumps, there are also other extensions conducted to the Lee-Carter model. For instance, Renshaw and Haberman (2006) incorporated an additional term into the model for the sake of cohort effect, which refers to the observation that for some generations, people born in the same year or same period of time display very similar pattern in terms of mortality dynamics as well as life expectancy improvements.

As mortality-linked securities become more and more prevalent in financial derivatives markets, the requirements for a meticulous model which can allow accurate martingale measure for mortality/longevity-linked securities have been rising in the past decade. Based on this tendency, there are also abundant studies to link the Lee-Carter model with financial derivative pricing strategies in order to corporate diverse demands in financial industry. Chuang and Brockett (2014) made further extensions based on the progress from Mitchell et al. (2013); they brought forward Esscher transform and Levy process in order to link the modification in Mitchell et al. (2013) with martingale pricing. Chuang and Brockett (2014) also conducted a q-forward pricing and found that their approach can produce similar in-sample estimations without any complicated adjustments. Such computational expediency is very helpful and beneficial for life insurance companies since they can directly apply this strategy in pricing, and furthermore, their model can also achieve lower costs of longevity hedging.

The Lee-Carter model, as well as most other mortality models proposed in recent literature, is constructed based on the assumption that the associated time-trend parameters follow a standard autoregressive integrated moving average (ARIMA) model, in particular, a random walk with drift. One flaw regarding to this assumption is that the forecasted values will be highly sensitive to the calibration period. Faced with this concern, Berkum et al. (2015) analyzed the impact of multiple structural changes on a large collection of mortality models. Although they did not make a general conclusion that work for all the cases, they did find several cases where improvements in estimates and more robust projections are verified with structural changes.

Besides the ARIMA model involved in the Lee-Carter model, there are also questions raised towards other assumptions. For instance, the Lee-Carter model assumes an invariant age effect with linear time component. However, Booth et al. (2002) found that such assumption does not fit Australian data. Faced with significant departures from linearity in the time effect, they adjusted the time-trend parameter to reproduce the age distribution of deaths, rather than total deaths. They also determined the optimal fitting period in order to address non-linearity in the time effect. In terms of Australian case, such modification produces higher forecast life expectancy than the original Lee-Carter method with a high reduction in error.

2.1.2 CBD Model

The CBD model was proposed by Cairns, Blake and Dowd (2006). The framework of the CBD model follows similar patterns with most of other stochastic mortality models; there are two time-trend parameters involved in the CBD model, which can be estimated by fitting the model into historical data. It is generally believed that the CBD model was mainly developed to capture mortality dynamics for higher ages, and thus it cannot provide a fit as good as the Lee-Carter model does. However, in the current market, mortality/longevity risk has become a newly social concern which has potentially big effects on life insurance companies. Longevity risk particularly refers to the unexpected longer life expectancy, which consequently leads to higher pay-out amounts and thus heavier financial burdens for life insurers. Faced with such underestimation possibility, there are plenty of hedging instruments developed in order to protect the insurers from the potential liabilities. For example, Cairns (2011) proposed a method using the Taylor expansion to conduct approximated longevity-contingent values, and by applying the functions to the CBD model, Cairns (2011) derived a closed-form formula which is a significant progress, since the previous research can only rely on simulations.

As the longevity risk catches more attention, the development of longevity-linked derivatives becomes more prevalent in the market, and thus more researchers begin to realize the CBD model's favorable properties in terms of this perspective. For example, Chan et al. (2014) found that, among various stochastic mortality models with time-trend parameters, the CBD model turns out to be the most appropriate candidate to indicate longevity risk levels at different time points. Based on this judgment, they expanded the CBD model by jointly modeling a more general class under a multivariate time-series setting so that their modification can take consideration of cross-correlations. Chan et al. (2014) applied this expanded model to conduct joint prediction region, which can function as a graphical longevity risk metric. This metric works as a helpful tool for life insurers to be able to compare the longevity risk exposures coming from different portfolios, and thus they can accomplish more intuitive and beneficial pricing strategies.

Similar to the Lee-Carter model, the associated parameters involved in the CBD model have also been modified along the way. There are questions raised towards the original assumption since historical data show that the two time-trend factors in the CBD model do not follow a random walk. Instead, Sweeting (2011) proposed that they should be modeled as a random fluctuation around a trend which varies periodically, and applied statistical techniques to quantitatively determine the periodicity, which can be used to project the frequency of future changes.

Both the Lee-Carter and the CBD models are robust stochastic models with good fits, and those extensions and modifications mentioned above can improve their forecasting performance to different extents. However, the concern is that we cannot find the universally best model with the most suitable extension that works out for all datasets. The best choice of mortality models and possibly corresponding modifications all depend on the data. For example, Cairns et al. (2009) quantitatively compared the results from eight stochastic models in terms of fitting and forecasting mortality rates, based on the mortality data of England and Wales and in the United States. As the numerical results indicate, for higher ages, an extension of the CBD model that incorporates a cohort effect fits the England and Wales males data best, while for U.S. males, the extension by Renshaw and Haberman (2006) to the Lee-Carter model that also allows for a cohort effect provides the best fit. This increases the challenges for life insurers to select an appropriate model with possible modifications when they need to perform pricing. Since different populations reflect distinct features, which cannot be captured using a unified mortality model, life insurers have to test different candidates to find the most suitable one for each specific case. To make the situation even worse, not all of these extensions mentioned above are implementable when they come to mortality forecasting.

2.1.3 Linear Relational Model

The linear relational model, an innovative model proposed by Tsai and Yang (2015), exploits the linear relation between two mortality sequences of equal length. Based on this foundation, Tsai and Yang (2015) applied a simple linear regression to relate the target mortality rate sequence to the base mortality sequence. For each of the estimated slope and intercept parameters, they conducted simple linear regression (LR-LR) and random walk with drift (LR-RW) models. The sequences of the fitted slope and intercept parameters can be then applied for forecasting future deterministic and stochastic mortality rates. They also compared the forecasting performances generated from the LR-LR and LR-RW models as well as the Lee-Carter and CBD models, and applied a measure called hedge effectiveness rate in order to quantify risk reduction amount per dollar spent for hedging mortality and longevity risks with a mortality-linked security under the underlying mortality models.

Hedging mortality and longevity risks has caught more and more attentions due to the flourishing market for mortality-linked securities, a brand new class of financial securities. As mentioned earlier, mortality/longevity risk has raised more attention in financial market; however, the fact is that there remains considerable unresolved issues for ensuring successful development in a financial market. Accordingly, there are numerous research papers addressing these concerns and offering resolutions for life insurers and annuity providers to manage their exposure to such risks. For instance, Wills and Sherris (2010) studied the securitization of longevity risk and derived an approach to structuring and pricing a longevity bond using techniques developed for the financial markets, particularly for mortgages and credit risk. Besides the securitization issue, Blake et al. (2006) analyzed the main characteristics of longevity bonds and demonstrated various forms into which these bonds can evolve. They addressed the sensitivity concerns and corresponding valuation issues in order to expand longevity bonds into a professional risk management tool for hedging longevity risk. To be more specific, Blake et al. (2006) analyzed two famous mortality-linked securities: the Swiss Re mortality bond issued in December 2003 and the EIB/BNP longevity bond issued in November 2004. They broke down the contracts and examined each specific condition, and also provided exhaustive comments regarding to such products' potential uses and implementation issues.

The Lee-Carter, CBD, and linear relational models are the three mortality models that we will focus on in this project. However, there are many other favorable mortality models and their extensions and modifications from different perspectives to fit and forecast mortality rates. In this sense, which model to choose turns out to be an indispensable question for both insurers and researchers. Therefore, Dowd et al. (2010) designed a framework to evaluate the goodness of fit of six stochastic mortality models using English & Welsh male mortality data. While they recognized notably deviated values generated from different models, they did not find a model that can perform well in all tests, and thus they concluded that no model clearly dominates the others.

To further expand the conclusion by Dowd et al. (2010), Cairns et al. (2011) continued to evaluate and compare the suitability of six stochastic mortality models which vary in different perspectives. They assessed these models individually in terms of forecasting performance at different ages, and established a set of qualitative criteria such as forecasting plausibility, biological reasonableness, the robustness of the forecasts relative to the sample period and so on. After an exhaustive comparison and analysis, they found out that even though a model might work as a good fit to historical data, it does not guarantee that the model will produce accurate forecasting values. Moreover, those models satisfying their presumed criteria can still generate significant deviations when forecasting mortality rates at different ages.

2.2 Credibility Theory

Credibility theory takes a vitally important role in property and casualty (P&C) insurance as a popular and effective approach to providing predictions in claim so that companies can reply on such estimates to determine the prospective premiums. Bühlmann (1967) proposed credibility premiums with a statistical framework which formalized the basis and principles for credibility theory, and also proposed the procedures to compute Bühlmann credibility factors Z in the credibility model with equal exposure units. Bühlmann and Straub (1970) achieved another significant practical improvement by allowing unequal exposure units within a portfolio, and this affords more flexibilities and puts the model forward to wider applications in practice.

Based on the result that Bühlmann (1967) developed, in the past few decades there have been extensive efforts paid by researchers in order to extend the model to more broad cases and deal with more intricate conditions. For example, Gerber and Jones (1975) and Frees et al. (1999) were dedicated to credibility models with time dependence. Frees and Wang (2005) modified the credibility theory by applying copulas which can be used to model the dependencies over time. They also extended the linear regression to the generalized linear models with respect to loss distributions. Particularly, they are the first to apply t-copula, the copula associated with the multivariate t-distribution, under such generalized linear model setting. Such link between credibility and loss distributions demonstrates favorable advantages since t-copula gives rise to easily computable predictive distributions that insurance companies can use to generate credibility predictors.

Yeo and Valdez (2006) extended the model by Frees and Wang (2005), and continued to extend the credibility theory within the frameworks of copula models. The accomplishment they made is to exempt the theory from claim independence assumption; namely, the classic model commonly assumes that claims are independent of each other, but Yeo and Valdez (2006) managed to apply the credibility theory without this restriction, and apply the concept of copula models to derive the predictive claim expressions so that they can conduct claim prediction.

Besides the modification and extensions mentioned above regarding to the model itself, there are also abundant studies to promote computational convenience. Generally speaking, the credibility factor can be calculated using two popular approaches, Bayesian and B$\ddot{u}$hlmann. It has been proved that under certain conditions, the credibility factor calculated from the Bühlmann and Straub model is equal to that from the empirical Bayes method (Ericson, 1970). However, with explicit prior and posterior loss distributions required, it is usually inconvenient to implement the Bayes method into practical computations. The Bühlmann credibility factor, on the other hand, can be calculated from some general statistical packages, and there are numerous reports presenting different methods to calculate the Bühlmann credibility factor. For example, Tolley et al. (1999) proposed a method of calculating the Bühlmann credibility factor using one-way Analysis of Variance (ANOVA). ANOVA is a popular statistical model to analyze the differences among group means and variations. It can be easily implemented with various programming tools including Excel, and thus provides a more open and accessible platform for companies to run the model. Based on the ANOVA setting, Tolley et al. (1999) applied the method of moments, and thus simply constructed and computed the credibility factor. Such advancement has important meanings regarding to business applications since the method of credibility theory is extensively applied in P&C insurance for actuaries to determine the premiums for all different types of business lines.

The two paragraphs above have presented abundant academic research with regards to mortality models and the credibility theory, with elaborated modifications and extended applications in insurance industry. However, the literature combining these two is quite limited. The credibility model is mainly applied in P&C insurance while mortality models and all their corresponding modifications primarily concentrate on life insurance or financial derivatives market. With different areas, no one tried to apply the credibility theory to improve the estimation of mortality rates. In this sense, this project develops a link to combine these two, derives the formula to conduct estimation, presents numerical results to compare forecasting performance, and makes decent contributions to the literature of mortality modeling. The forecasting and estimating methods illustrated in this project are applicable to products issued by both traditional life insurance companies and financial derivative markets, and can be applied under various product pricing scenarios as well.

Chapter 3

Applying Credibility Theory to Existing Models

Lee-Carter model, CBD model, and linear relational models are the three mortality models embedded with the credibility theory in this project. These three models possess favorable performances in terms of capturing the trend of mortality rates. The Lee-Carter model is a simple robust model in which the mortality rate dynamics is driven by both age effect and time effect; it provides a good fit over wide age ranges. The CBD model uses two time-trend factors, level and slope, to capture the movements of mortality rates, particularly taking advantage of the linear pattern reflected from the logit function of one-year death probability at ages above 40, so it was designed to model mortality rates for higher ages. Although the CBD model does not fit as well as the Lee-Carter model in general cases, it is still a robust model with simple age effects and convenience to incorporate parameter uncertainty. The linear relational model, on the other hand, is an innovative approach to modeling mortality rates. It exploits a linear relationship between two mortality sequences, and can produce decent forecasting performance in terms of hedging effectiveness; it even outperforms the Lee-Carter model for the mortality datasets studied in Tsai and Yang (2015). All these three models are stochastic models which provide more flexibility to deal with the random deviation of the deterministic forecast in mortality rates. Although some deterministic models can capture the overall declining trend of mortality rates, the real declines will be quite volatile; the true mortality rates will always stay unknown until they really happen in the future, and thus the uncertainty needs to be taken into account when we conduct forecasting.

Figure 3.1 displays the framework for the Lee-Carter and CBD models. In the figure, m represents the length of the age span and n is the length of the fitting year span. In this project, t_0 stays fixed, i.e., the ending year of the fitting year span remains the same, while n changes so that the length of the fitting year span varies. Changing fitting year spans will return diverse forecasted results which are the observations of the sample data employed into the credibility method in later sections.

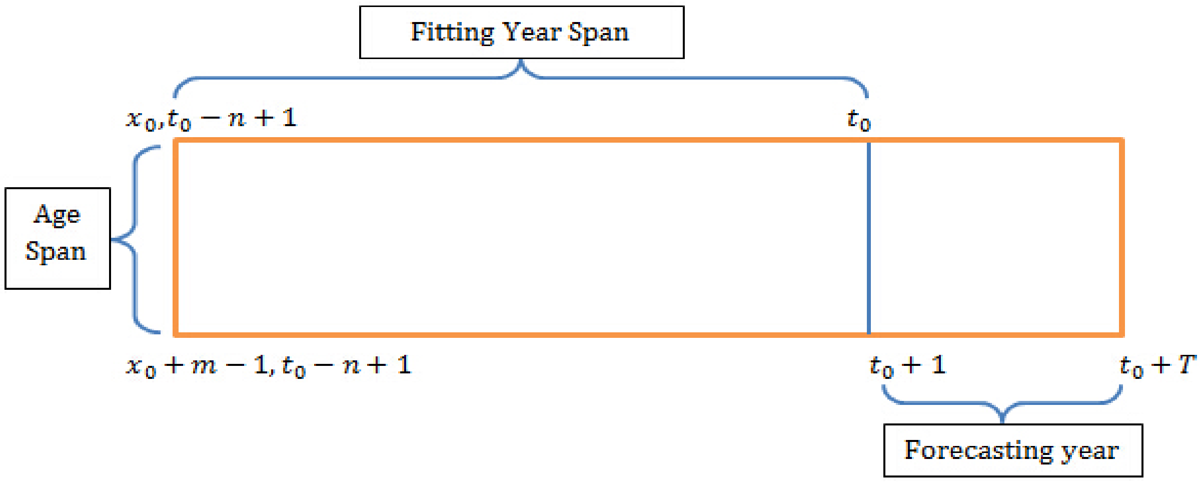


Figure 3.1: Framework of the Lee-Carter and CBD models

The following sections illustrate the three models in more details. Based on the fact that different fitting year spans lead to distinct mortality forecasts, we treat these forecasts as a random sample, and thereafter apply the non-parametric Bühlmann method to the sample. Justification, model interpretations, and numerical results are also provided.

3.1 Lee-Carter Model

The Lee-Carter model is the most cited and used model in actuarial literature to model mortality rates. It models the logarithm of central death rate, $ln(m_{x,t})$, with a selected age span and a fitting year span. The deterministic part represents the mean of the fitted or forecasted mortality rate, and the stochastic part describes all the errors that are not captured in the model. This section briefly goes over the essentials of the model, and only focuses on the deterministic part, from which we can apply the credibility theory, get the non-parametric Bühlmann estimates, and achieve a better prediction.

3.1.1 Notations & Assumptions

The Lee-Carter model captures the dynamics of mortality rates taking both age effect and time (year) effect into account, and performs forecasting based on this layout. The classical Lee-Carter model is presented as

$$ln(m_{x,t}) = a_x + b_x \times k_t + \varepsilon_{x,t}, \quad x = x_0, x_0 + 1, ..., x_0 + m - 1, \quad t = t_0 - n + 1, ..., t_0,$$

where a_x is the average age-specific mortality factor, k_t is the general mortality level in year t, b_x is the age-specific reaction to k_t at age x, and $\varepsilon_{x,t}$ represents the error term, which is assumed to be independent and identically distributed (i.i.d) normal for $t = t_0 - n + 1, \cdots, t_0$ with mean 0 and variance $\sigma^2_{\varepsilon_x}$. Furthermore, it is assumed that $\varepsilon_{x,t}$ is independent with $\varepsilon_{z,t}$ for $x \neq z$.

The equation above is subject to two constraints: $\sum_{x=x_0}^{x_0+m-1} b_x = 1$ and $\sum_{t=t_0-n+1}^{t_0} k_t = 0$. To fit the model and estimate its parameters, we can use singular value decomposition (SVD) method; alternatively, we can estimate parameters as follows:

$$\begin{aligned}
&\sum_{t=t_0-n+1}^{t_0} ln(m_{x,t}) = n \times a_x + b_x \times \sum_{t=t_0-n+1}^{t_0} k_t = n \times a_x \\
&\Rightarrow \hat{a}_x = \frac{1}{n} \sum_{t=t_0-n+1}^{t_0} ln(m_{x,t}), \quad x = x_0, x_0+1, ..., x_0+m-1,
\end{aligned} \tag{3.1}$$

and

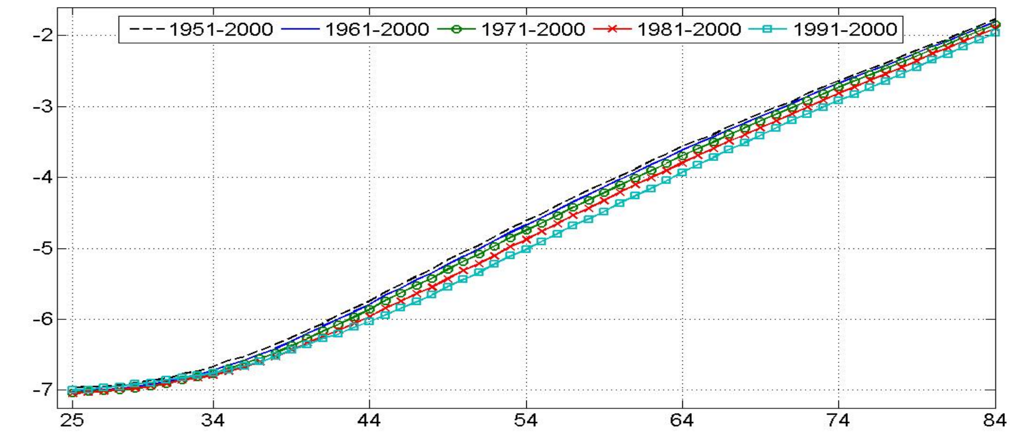
$$\begin{aligned}
&\sum_{x=x_0}^{x_0+m-1} [ln(m_{x,t}) - \hat{a}_x] = k_t \times \sum_{x=x_0}^{x_0+m-1} b_x = k_t \times 1 \\
&\Rightarrow \hat{k}_t = \sum_{x=x_0}^{x_0+m-1} [ln(m_{x,t}) - \hat{a}_x], \quad t = t_0-n+1, \cdots, t_0.
\end{aligned} \tag{3.2}$$

To get $\hat{b}_x$, we regress $[ln(m_{x,t}) - \hat{a}_x]$ on $\hat{k}_t$ without the constant term involved for each x.

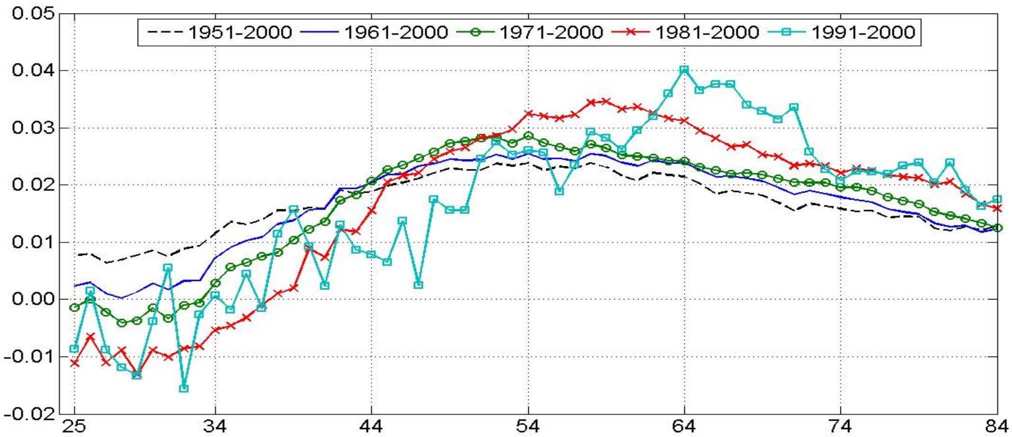
3.1.2 Estimation of Model Parameters

Figure 3.2 displays Lee-Carter model's parameters $\hat{a}_x$, $\hat{b}_x$, and $\hat{k}_t$. Throughout this chapter, we use UK male mortality data to conduct parameter illustrations. Five fitting year spans are selected to illustrate mortality forecasting for $2001 - 2010$, which are $1951 - 2000, 1961 - 2000, 1971 - 2000, 1981 - 2000$, and $1991 - 2000$. The age span is set to be $[25, 84]$ with $x_0 = 25$ and $m = 60$. Based on this layout, there are 5 paths corresponding to 5 fitting year spans for $\hat{a}_x$'s, $\hat{b}_x$'s and $\hat{k}_t$'s given in Figures 3.2 (a), (b) and (c), respectively.

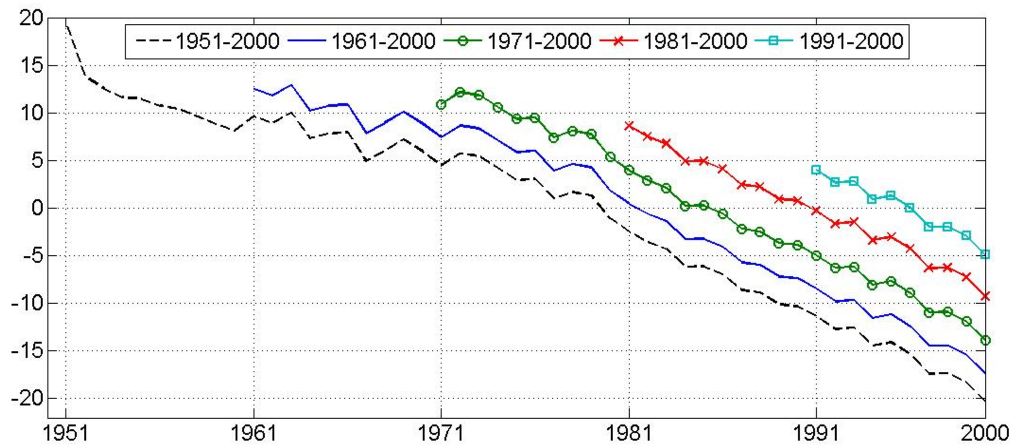
Figure 3.2 also provides us a good picture for the parameter movements. It can be seen that $\hat{a}_x$'s follow quite a consistent pattern with a neat and clear uptrend; the 5 paths turn out very close to each other, and $\hat{a}_x$'s increase in x. $\hat{k}_t$'s also show a neatly linear pattern but with an overall down trend instead; the 5 paths look parallel with each other. $\hat{b}_x$'s, on the other hand, reflect more fluctuations and do not have an evident linear relationship with age x; however, it can still be seen that the parameter values increase with age for the first half, and then start decreasing at some higher age.



(a) $\hat{a}_x$



(b) $\hat{b}_x$



(c) $\hat{k}_t$

Figure 3.2: Estimates of parameters for the Lee-Carter model

3.1.3 Model Prediction

From Figure 3.2 (c) it can be seen that $\hat{k}_t$'s follow an approximately linear pattern, so we apply $ARIMA(0,1,0)$ to model $\hat{k}_t$, that is, a random walk model with drift θ: $\hat{k}_t = \hat{k}_{t-1} + \theta + \epsilon_t$, where ϵ_t is the time-trend error and follows independent and identically distributed (i.i.d.) normal with mean 0 and variance σ_ϵ^2. In this sense, $\hat{\theta}$ can be estimated by

$$\hat{\theta} = \frac{1}{n-1} \sum_{t=t_0-n+2}^{t_0} \left(\hat{k}_t - \hat{k}_{t-1}\right) = \frac{1}{n-1} \left(\hat{k}_{t_0} - \hat{k}_{t_0-n+1}\right). \tag{3.3}$$

Hence, $\hat{k}_{t_0+\tau}$ for year $t_0+\tau$ can be projected as

$$\hat{k}_{t_0+\tau} = \hat{k}_{t_0} + \tau \times \hat{\theta}, \quad \tau = 1, 2, ..., T, \tag{3.4}$$

where τ represents the number of years beyond t_0 such that $[t_0+1, t_0+T]$ is the forecasting year span as shown in Figure 3.1. Based on the equations above, we can forecast the deterministic mortality rate for age x in year $t_0+\tau$ as

$$ln(\hat{m}_{x,t_0+\tau}) = \hat{a}_x + \hat{b}_x \times (\hat{k}_{t_0} + \tau \times \hat{\theta}) = ln(\hat{m}_{x,t_0}) + \hat{b}_x \times \tau \times \hat{\theta}, \tag{3.5}$$

where $ln(\hat{m}_{x,t_0}) = \hat{a}_x + \hat{b}_x \times \hat{k}_{t_0}$ stands for the fitted value for year t_0. For the stochastic part, we need to add two error terms, the model error $\varepsilon_{x,t}$ from $ln(m_{x,t})$ and the time-trend error ϵ_t from $\hat{k}_t$, to get

$$\begin{aligned} ln(\tilde{m}_{x,t_0+\tau}) &= \hat{a}_x + \hat{b}_x \times (\hat{k}_{t_0} + \tau \times \hat{\theta} + \sum_{t=1}^{\tau} \epsilon_{t_0+t}) + \varepsilon_{x,t_0+\tau} \\ &= ln(\hat{m}_{x,t_0+\tau}) + \hat{b}_x \times \sum_{t=1}^{\tau} \epsilon_{t_0+t} + \varepsilon_{x,t_0+\tau}. \end{aligned} \tag{3.6}$$

It follows that the variance of $ln(\tilde{m}_{x,t_0+\tau})$ is

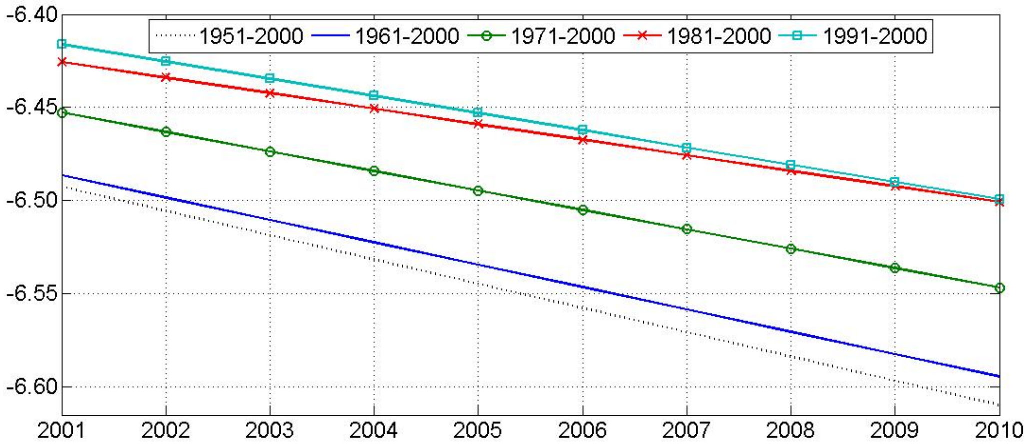
$$\sigma_{x,\tau}^2 = Var\left[ln(\tilde{m}_{x,t_0+\tau})\right] = \tau \times \hat{b}_x^2 \times \sigma_\epsilon^2 + \sigma_{\varepsilon_x}^2, \tag{3.7}$$

where σ_ϵ^2, the variance of the time-trend error, can be estimated by

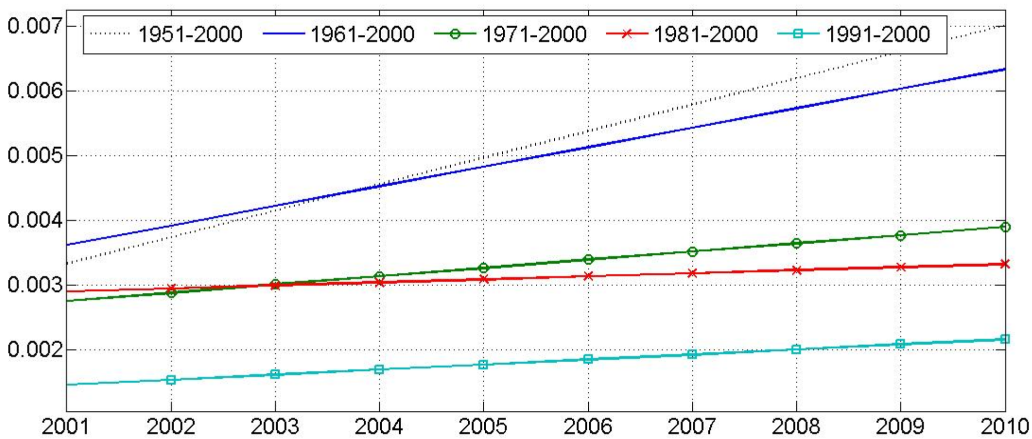
$$\hat{\sigma}_\epsilon^2 = \frac{1}{n-2} \sum_{t=t_0-n+2}^{t_0} \epsilon_t^2 = \frac{1}{n-2} \sum_{t=t_0-n+2}^{t_0} \left(\hat{k}_t - \hat{k}_{t-1} - \hat{\theta}\right)^2, \tag{3.8}$$

and $\sigma_{\varepsilon_x}^2$, the variance of the model error, can be estimated as

$$\hat{\sigma}_{\varepsilon_x}^2 = \frac{1}{n-2} \sum_{t=t_0-n+1}^{t_0} \varepsilon_{x,t}^2 = \frac{1}{n-2} \sum_{t=t_0-n+1}^{t_0} \left[ln(\hat{m}_{x,t}) - \left(\hat{a}_x + \hat{b}_x \times \hat{k}_t\right)\right]^2. \tag{3.9}$$



(a) $ln(\hat{m}_{x,t_0+\tau})$



(b) $\hat{\sigma}^2_{x,\tau} = \hat{V}ar\left[ln(\tilde{m}_{x,t_0+\tau})\right]$

Figure 3.3: $ln(\hat{m}_{x,t_0+\tau})$ and $\hat{V}ar\left[ln(\tilde{m}_{x,t_0+\tau})\right]$ for $x = 40$ based on 5 fitting year spans

Figure 3.3 displays the forecasted $ln(\hat{m}_{x,t_0+\tau})$ and estimated $\hat{V}ar\left[ln(\tilde{m}_{x,t_0+\tau})\right]$ for $x = 40$. The same 5 fitting year spans, $1951-2000, 1961-2000, 1971-2000, 1981-2000$, and $1991-2000$, are selected to forecast mortality rates for years $2001-2010$. It is obvious from Figure 3.3 (a) that different fitting year spans give different $\hat{a}_x$, $\hat{b}_x$ and $\hat{k}_{t_0}$ to form distinct fitting value $ln(\hat{m}_{x,t_0}) = \hat{a}_x + \hat{b}_x \times \hat{k}_{t_0}$ for year t_0, and then yield varying forecast $ln(\hat{m}_{x,t_0+\tau}) = ln(\hat{m}_{x,t_0}) + \hat{b}_x \times \tau \times \hat{\theta}$. Same situation can also be found in Figure 3.3 (b) for $\hat{\sigma}^2_{x,\tau} = \hat{V}ar\left[ln(\tilde{m}_{x,t_0+\tau})\right] = \tau \times \hat{b}^2_x \times \hat{\sigma}^2_\epsilon + \hat{\sigma}^2_{\varepsilon_x}$, and such distinct values of variance will affect simulation results.

3.2 CBD Model

3.2.1 Notations & Assumptions

Built on the observation that the logarithm of $q_{x,t}/p_{x,t}$ is approximately linear at ages above 40, the CBD model was designed to model mortality rates for higher ages. Although the CBD model is less popular than the Lee-Carter model, yet it possesses its own superiority. In fact, Chan et al. (2014) concluded that the CBD model is most suitably used as indexes to indicate levels of longevity risk at different time points among all different kinds of time-varying models. The model applies two time-varying parameters to capture the movements in mortality rates. With the same framework demonstrated in Figure 3.1, the CBD model is presented as

$$logit(q_{x,t}) = k_t^1 + k_t^2(x - \bar{x}) + \varepsilon_{x,t}, \quad x = x_0, \cdots, x_0 + m - 1, \quad t = t_0 - n + 1, \cdots, t_0, \tag{3.10}$$

where $logit(q_{x,t}) = ln\left(\frac{q_{x,t}}{1-q_{x,t}}\right)$, $\bar{x} = \dfrac{1}{m}\sum_{x=x_0}^{x_0+m-1} x$, and $\varepsilon_{x,t}$ is the model error which is assumed i.i.d. normal for $t = t_0 - n + 1, \cdots, t_0$ with mean 0 and variance $\sigma^2_{\varepsilon_x}$. In addition, it is assumed that $\varepsilon_{x,t}$ is independent of $\varepsilon_{z,t}$ (i.e., $\varepsilon_{x,t} \perp \varepsilon_{z,t}$) for $x \neq z$.

3.2.2 Estimation of Model Parameters

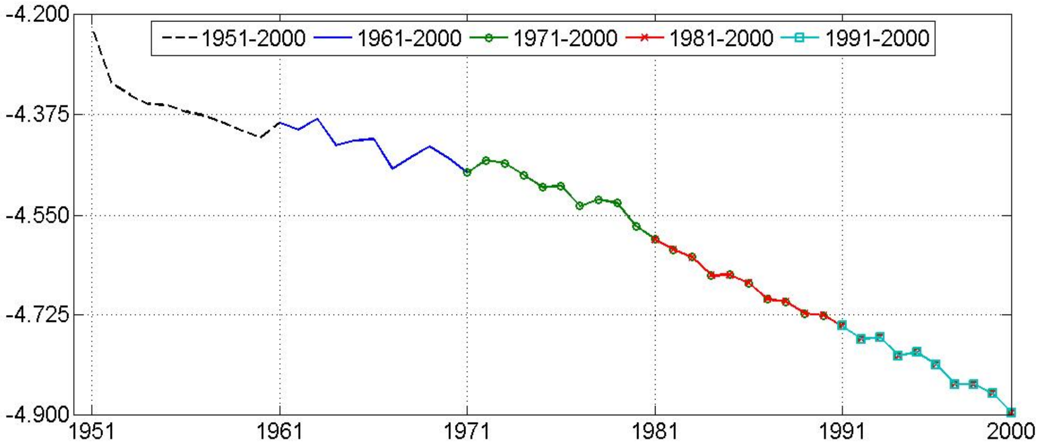
Both estimates of k_t^1 and k_t^2 can be obtained by regressing $logit(q_{x,t})$ on $(x - \bar{x})$ for each t to get

$$\begin{aligned}\sum_{x=x_0}^{x_0+m-1} logit(q_{x,t}) &= \sum_{x=x_0}^{x_0+m-1} k_t^1 + \sum_{x=x_0}^{x_0+m-1} k_t^2(x - \bar{x}) \\ \Rightarrow \hat{k}_t^1 &= \frac{1}{m}\sum_{x=x_0}^{x_0+m-1} logit(q_{x,t}),\end{aligned} \tag{3.11}$$

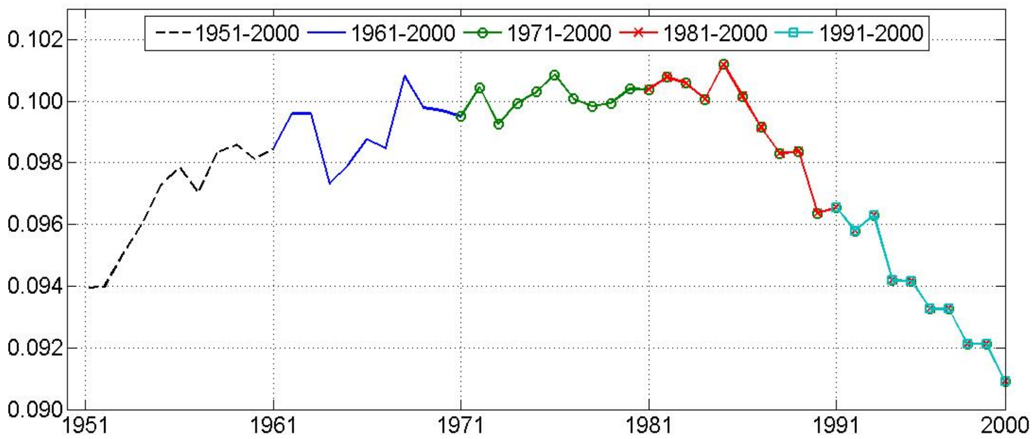
and

$$\begin{aligned}\sum_{x=x_0}^{x_0+m-1} [logit(q_{x,t}) \times (x - \bar{x})] &= \sum_{x=x_0}^{x_0+m-1} k_t^1(x - \bar{x}) + \sum_{x=x_0}^{x_0+m-1} k_t^2(x - \bar{x})^2 \\ \Rightarrow \hat{k}_t^2 &= \frac{\sum_{x=x_0}^{x_0+m-1} [logit(q_{x,t}) \times (x - \bar{x})]}{\sum_{x=x_0}^{x_0+m-1} (x - \bar{x})^2}.\end{aligned} \tag{3.12}$$

Figures 3.4 (a) and (b) exhibit $\hat{k}_t^1$'s and $\hat{k}_t^2$'s, respectively. The age span remains at $[25, 84]$, and the 5 fitting year spans are still $1951 - 2000, 1961 - 2000, 1971 - 2000, 1981 - 2000$, and $1991 - 2000$. From (3.11) and (3.12) above, it can be concluded that for $i = 1, 2$, $\hat{k}_t^i$ will return the same value as long as the fitting year span covers the same t. For example, the fitting year spans $1951 - 2000$ and $1961 - 2000$ share the same historical data for the period from 1961 to 2000, and since $\hat{k}_t^i$ only depends on $logit(q_{x,t})$ and $(x - \bar{x})$, as a consequence, $\hat{k}_t^i$ for $1961 - 2000$ overlaps with $\hat{k}_t^i$ for $1951 - 2000$ for $i = 1, 2$, respectively. Same overlapped patterns also apply



(a) $\hat{k}_t^1$



(b) $\hat{k}_t^2$

Figure 3.4: Estimates of parameters for the CBD model

to other fitting year spans, which can be clearly seen in Figure 3.4.

The CBD model assumes that the time trends $\hat{k}_t^1$ and $\hat{k}_t^2$ are independent and both follow a random walk with drift θ_i, $i = 1, 2$. Specifically,

$$\hat{k}_t^i = \hat{k}_{t-1}^i + \theta_i + \epsilon_t^i, \quad i = 1, 2, \tag{3.13}$$

where the time-trend error $\epsilon_t^i \overset{i.i.d}{\sim} N\left(0, \sigma_{\epsilon^i}^2\right)$ for $i = 1, 2$. In addition, we assume that the two time-trend errors are independent of each other, i.e., $\epsilon_t^1 \perp \epsilon_t^2$. The drift θ_i can be estimated as

$$\hat{\theta}_i = \frac{1}{n-1} \sum_{t=t_0-n+2}^{t_0} \left(\hat{k}_t^i - \hat{k}_{t-1}^i\right) = \frac{1}{n-1} \left(\hat{k}_{t_0}^i - \hat{k}_{t_0-n+1}^i\right), \quad i = 1, 2. \tag{3.14}$$

There is noticeable difference between Figures 3.4 (a) and (b). While random walk seems to be a good fit for $\hat{k}^1$, it does not seem appropriate for $\hat{k}^2$. We should realize that although the CBD model suggests the random walk with drift to model the time-trend parameters $\hat{k}^1$ and $\hat{k}^2$, yet the random walk assumption might not work out well for all cases. It may not be suitable for some age spans, year spans, or population datasets to be fitted into the random walk model. We are conscious of a critical inference that there does not exist a universally ideal model which fits all datasets well; instead, a model's forecasting performance heavily depends on the data, and different datasets may return quite diverse forecasting results.

3.2.3 Model Prediction

By extrapolating $\hat{k}_t^1$ and $\hat{k}_t^2$ to the future, we can make mortality prediction. The *logit* function of the forecasted $q_{x,t_0+\tau}$ for age x in year $t_0+\tau$ is

$$\begin{aligned} logit\left(\hat{q}_{x,t_0+\tau}\right) &= \left(\hat{k}_{t_0}^1 + \tau \times \hat{\theta}_1\right) + \left(\hat{k}_{t_0}^2 + \tau \times \hat{\theta}_2\right) \times (x-\bar{x}) \\ &= logit\left(\hat{q}_{x,t_0}\right) + \tau \times \left[\hat{\theta}_1 + \hat{\theta}_2 \times (x-\bar{x})\right], \end{aligned} \tag{3.15}$$

where $logit\left(\hat{q}_{x,t_0}\right) = \hat{k}_{t_0}^1 + \hat{k}_{t_0}^2 \times (x-\bar{x})$ represents the fitted value for year t_0. The stochastic part is

$$\begin{aligned} logit\left(\tilde{q}_{x,t_0+\tau}\right) &= \left(\hat{k}_{t_0}^1 + \tau\hat{\theta}_1 + \sum_{t=1}^{\tau} \epsilon_{t_0+t}^1\right) + \left(\hat{k}_{t_0}^2 + \tau\hat{\theta}_2 + \sum_{t=1}^{\tau} \epsilon_{t_0+t}^2\right) \times (x-\bar{x}) + \varepsilon_{x,t_0+\tau} \\ &= logit\left(\hat{q}_{x,t_0+\tau}\right) + \sum_{t=1}^{\tau}\left[\epsilon_{t_0+t}^1 + \epsilon_{t_0+t}^2(x-\bar{x})\right] + \varepsilon_{x,t_0+\tau}, \end{aligned} \tag{3.16}$$

and the variance of $logit\left(\tilde{q}_{x,t_0+\tau}\right)$ is

$$\sigma_{x,\tau}^2 = Var\left[logit\left(\tilde{q}_{x,t_0+\tau}\right)\right] = \tau \times \sigma_{\epsilon^1}^2 + \tau \times \sigma_{\epsilon^2}^2 \times (x-\bar{x})^2 + \sigma_{\varepsilon_x}^2, \tag{3.17}$$

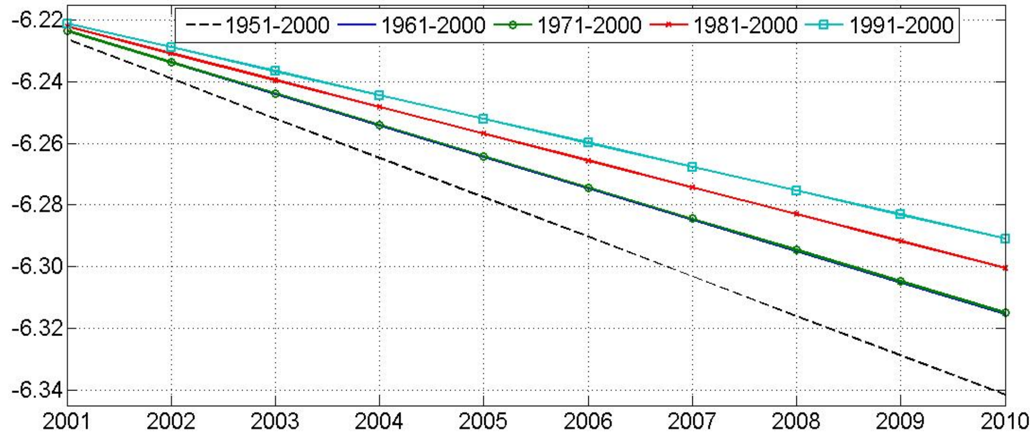
where $\sigma_{\epsilon^i}^2$, the variance of the time-trend error, can be estimated by

$$\hat{\sigma}_{\epsilon^i}^2 = \frac{1}{n-2} \sum_{t=t_0-n+1}^{t_0-1} \left(\hat{k}_{t+1}^i - \hat{k}_t^i - \hat{\theta}_i\right)^2, \quad i=1,2, \tag{3.18}$$

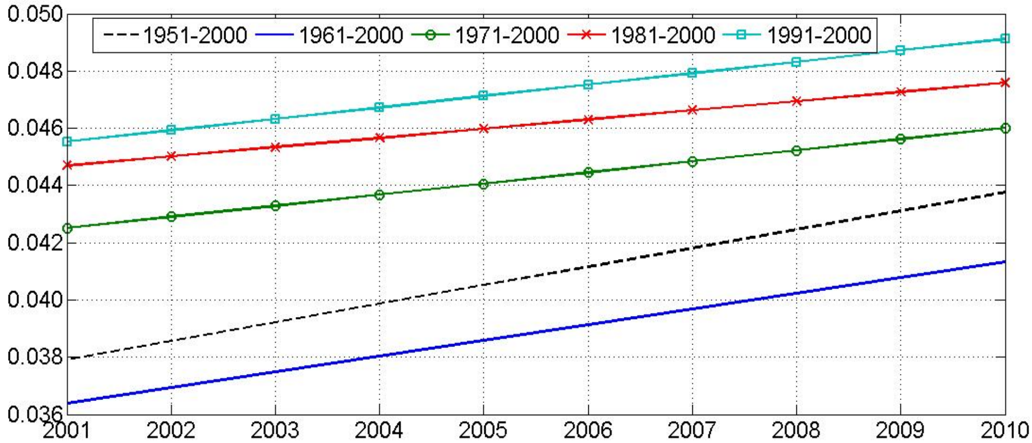
and $\sigma_{\varepsilon_x}^2$, the variance of the model error, can be estimated by

$$\hat{\sigma}_{\varepsilon_x}^2 = \frac{1}{n-2} \sum_{t=t_0-n+1}^{t_0} \left[logit\left(q_{x,t}\right) - \hat{k}_t^1 - \hat{k}_t^2 \times (x-\bar{x})\right]^2. \tag{3.19}$$

Figure 3.5 displays 5 groups of forecasting results with the CBD model being applied to forecast mortality rates for years $2001-2010$ from 5 fitting year spans: $1951-2000, 1961-2000, 1971-2000, 1981-2000$, and $1991-2000$ for $x=40$. It is obvious that different fit-



(a) $logit(\hat{q}_{x,t_0+\tau})$



(b) $\hat{\sigma}^2_{x,\tau} = \hat{V}ar\left[logit(\tilde{q}_{x,t_0+\tau})\right]$

Figure 3.5: $logit(\hat{q}_{x,t_0+\tau})$ and $\hat{V}ar\left[logit(\tilde{q}_{x,t_0+\tau})\right]$ for $x = 40$ based on 5 fitting year spans

ting year spans generate divergent slopes and discrepant forecasted mortality rates accordingly. Such discrepancy may be caused by the models for $\hat{k}^1_t$ and $\hat{k}^2_t$. Figure 3.4 (a) displays a relatively linear pattern, so a linear regression model is able to capture $\hat{k}^1_t$'s dynamics; however, $\hat{k}^2_t$'s pattern reflected in Figure 3.4 (b) is quite volatile, in which case it is not a good idea to apply the random walk with drift θ. As mentioned above, θ_i is estimated by $\hat{\theta}_i = \frac{1}{n-1}\sum_{t=t_0-n+1}^{t_0}\left(\hat{k}^i_t - \hat{k}^i_{t-1}\right) = \dfrac{1}{n-1}\left(\hat{k}^i_{t_0} - \hat{k}^i_{t_0-n+1}\right)$ for $i = 1, 2$; so different selections of n lead to very distinct slopes and corresponding forecasted results, which is reflected in Figure 3.5 (a).

3.3 LR-RW Model

The linear relational model is an innovative model developed from a different layout compared with that of the Lee-Carter and CBD models. The relational model relates one mortality sequence to the other by $u_{x,t_0+t} = f(u_{x,t_0};\Theta) + E_{\epsilon,t}$ for some function f and its associated parameter set Θ. $u_{x,t}$ can be $q_{x,t}$, $p_{x,t}$, $m_{x,t}$, $ln(m_{x,t})$, or $logit(q_{x,t})$. With such wide-range choices and high-level flexibility, this model can provide favorable forecasting performance. For this project, we will focus on the linear regression and random walk (LR-RW) model, a specific form under the linear relational model category. Specifically, $f(u_{x,t_0})$ is a linear function of $u_{x,t}$. We fit u_{x,t_0+t} by simple linear regression, and then construct a random walk with drift to model the movements of the parameters of f.

Figure 3.6 is the framework demonstration for the LR-RW model where m represents the length of age span and n is the length of the fitting year span. We keep t_0, the ending year of the fitting year span, to be fixed, and change the length n in order to make forecasting based on different fitting year spans.

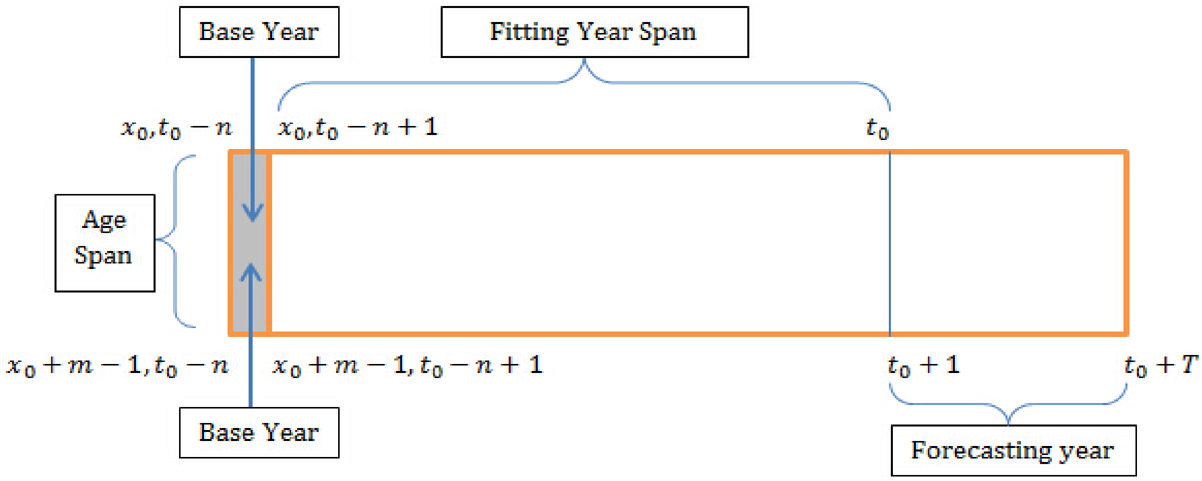


Figure 3.6: Framework of the LR-RW model

3.3.1 Notations & Assumptions

As demonstrated in the figure above, the LR-RW model employs a base year $t_0 - n$ in order to construct the model, and assumes a linear relationship between two mortality sequences for the base year and a year in the fitting year span. The base year always stays one year before the selected fitting year span. For example, if the fitting year span is chosen to be $1951 - 2000$, then the base year would be 1950. Linear regression is applied to fit u_{x,t_0-n+t} as

$$u_{x,t_0-n+t} = \beta_{0,t} + \beta_{1,t} \times u_{x,t_0-n} + \varepsilon_{x,t}, \quad t = 1, 2, ..., n, \tag{3.20}$$

where $\varepsilon_{x,t}$ is the model error which is assumed independent and identically distributed normal for $t = 1, 2, ..., n$, i.e., $\varepsilon_{x,t} \overset{i.i.d}{\sim} N(0, \sigma^2_{\varepsilon_x})$. Let u_{x,t_0-n+t} be the logarithm of central death rate for age x in year $t_0 - n + t$, $t = 1, 2, ..., n$, and the LR-RW model is given by

$$ln(m_{x,t_0-n+t}) = \beta_{0,t} + \beta_{1,t} \times ln(m_{x,t_0-n}) + \varepsilon_{x,t}, \quad t = 1, 2, ..., n, \tag{3.21}$$

where $\beta_{1,t}$ is the slope parameter that signalizes the time-varying relationship between the fitting mortality sequence for year $t_0 - n + t$ and that for the base year $t_0 - n$, $\beta_{0,t}$ is the intercept parameter, and $\varepsilon_{x,t}$ represents the model error term which is assumed to be independent and identically distributed (i.i.d) normal for $t = 1, \cdots, n$ with mean 0 and variance $\sigma^2_{\varepsilon_x}$. Furthermore, it is assumed that $\varepsilon_{x,t}$ is independent of $\varepsilon_{z,t}$ for $x \neq z$.

3.3.2 Estimation of Model Parameters

Figure 3.7 displays the parameter sequence $\hat{\beta}_{i,t}$ estimated from linear regression with 5 paths corresponding to 5 fitting year spans for UK males. From the figure we can see that the 5 $\hat{\beta}_{i,t}$'s from different fitting year spans follow quite similar pattens for $i = 1, 2$. However, it does not necessarily imply that the shortest fitting year span will return the biggest value of $\hat{\beta}_{i,t}$ as that for the Lee-Carter model, in which the shorter the fitting year span, the bigger the $\hat{k}_t$ as shown in Figure 3.2 (c).

3.3.3 Model Prediction

Let $B_i = \{\hat{\beta}_{i,1}, \quad \cdots, \quad \hat{\beta}_{i,n}\}$, $i = 0, 1$. To model B_i, Tsai and Yang (2015) assumed a random walk with drift θ_i as follows:

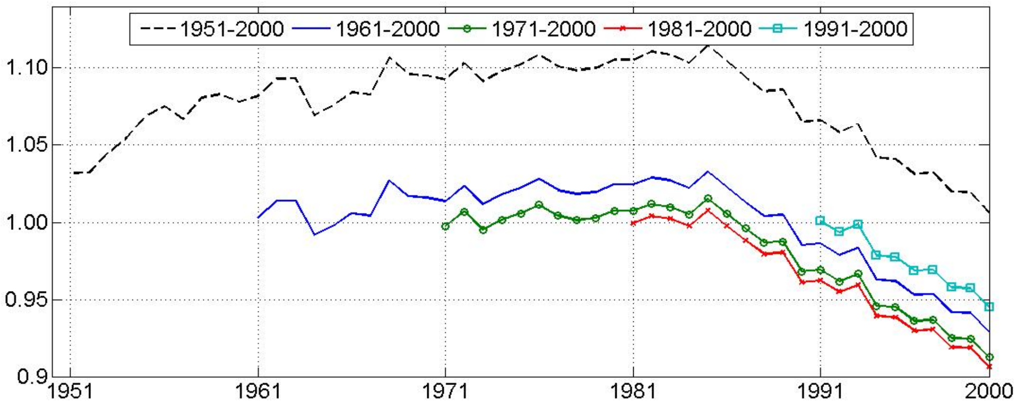
$$\hat{\beta}_{i,t} = \hat{\beta}_{i,t-1} + \theta_i + \epsilon^i_t, \tag{3.22}$$

where θ_i can be estimated by $\hat{\theta}_i = \dfrac{1}{n-1}\left(\hat{\beta}_{i,t_0} - \hat{\beta}_{i,t_0-n+1}\right)$, and the time-trend error ϵ^i_t are assumed to follow i.i.d $N\left(0, \sigma^2_{\epsilon^i}\right)$ for $i = 0, 1$. In addition, we assume that the two time-trend errors are independent of each other, i.e., $\epsilon^1_t \perp \epsilon^0_t$. The variance of the time-trend error, $\sigma^2_{\epsilon^i}$, can be estimated by

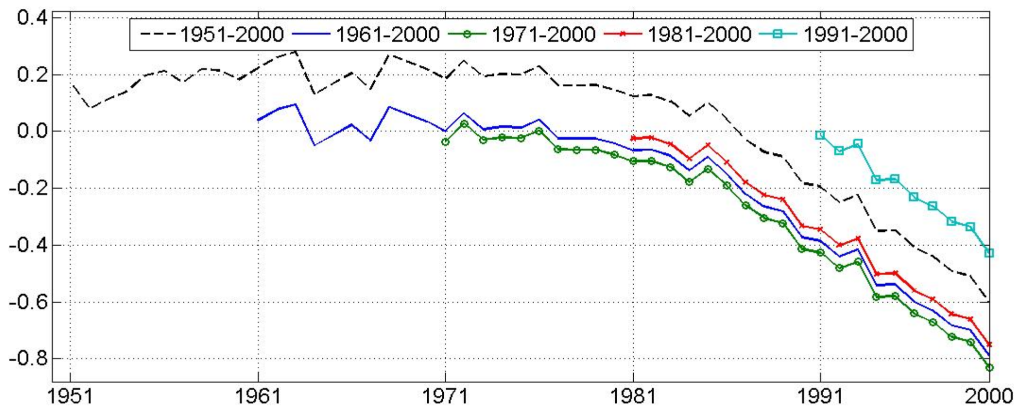
$$\hat{\sigma}^2_{\epsilon^i} = \frac{1}{n-2}\sum_{t=2}^{n}\left[\hat{\beta}_{i,t_0-t+2} - \hat{\beta}_{i,t_0-t+1} - \hat{\theta}_i\right]^2, \tag{3.23}$$

so that the time-trend parameter $\hat{\beta}_{i,t}$ for year $t_0 + \tau$ can be projected as

$$\tilde{\hat{\beta}}_{i,t_0+\tau} = \hat{\beta}_{i,t_0} + \tau \times \hat{\theta}_i + \sum_{t=1}^{\tau} \epsilon^i_{t_0+t}, \quad \tau = 1, 2, ..., T. \tag{3.24}$$



(a) $\hat{\beta}_{1,t}$



(b) $\hat{\beta}_{0,t}$

Figure 3.7: Estimates of parameters for the LR-RW model

Thus, the deterministic prediction for age x in year $t_0 + \tau$ is

$$ln\left(\hat{m}_{x,t_0+\tau}\right) = \left(\hat{\beta}_{1,t_0} + \tau \times \hat{\theta}_1\right) \times ln\left(\hat{m}_{x,t_0-n}\right) + \left(\hat{\beta}_{0,t_0} + \tau \times \hat{\theta}_1\right), \tag{3.25}$$

and the stochastic one is

$$\begin{aligned} ln\left(\tilde{m}_{x,t_0+\tau}\right) &= \left(\hat{\beta}_{1,t_0} + \tau\hat{\theta}_1 + \sum_{t=1}^{\tau} \epsilon^1_{t_0+t}\right) \times ln\left(m_{x,t_0-n}\right) + \left(\hat{\beta}_{0,t_0} + \tau\hat{\theta}_0 + \sum_{t=1}^{\tau} \epsilon^0_{t_0+t}\right) + \varepsilon_{x,t_0+\tau} \\ &= ln\left(\hat{m}_{x,t_0+\tau}\right) + \sum_{t=1}^{\tau}\left[ln\left(m_{x,t_0-n}\right) \times \epsilon^1_{t_0+t} + \epsilon^0_{t_0+t}\right] + \varepsilon_{x,t_0+\tau}. \end{aligned} \tag{3.26}$$

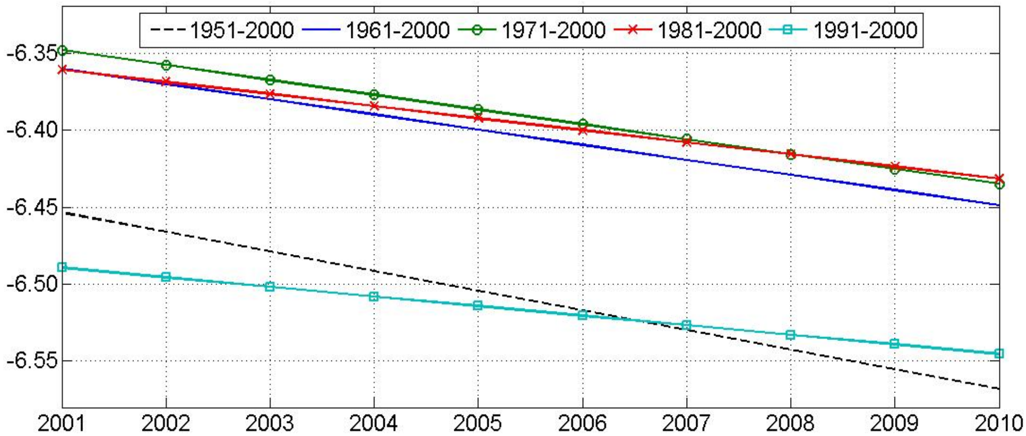
The variance of the stochastic forecast $ln\left(\tilde{m}_{x,t_0+\tau}\right)$ is

$$\sigma^2_{x,\tau} = Var[ln\left(\tilde{m}_{x,t_0+\tau}\right)] = \tau \times \left[ln\left(m_{x,t_0-n}\right)^2 \times \sigma^2_{\epsilon^1} + \sigma^2_{\epsilon^2}\right] + \sigma^2_{\varepsilon_x}, \tag{3.27}$$

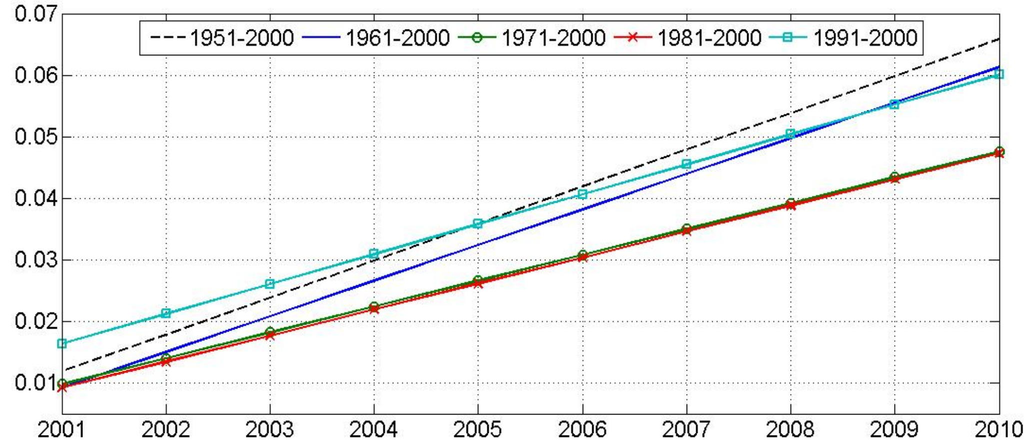
where $\sigma^2_{\varepsilon^i}$ is estimated in (3.23), and $\sigma^2_{\varepsilon_x}$, the variance of the model error, is estimated by

$$\hat{\sigma}^2_{\varepsilon_x} = \frac{1}{n(m-2)} \sum_{j=1}^{n} \sum_{i=0}^{m-1} \left[ln\left(m_{x+i,t_0-n+j}\right) - \hat{\beta}_{0,t_0-n+j} - \hat{\beta}_{1,t_0-n+j} \times ln\left(m_{x+i,t_0-n}\right)\right]^2. \tag{3.28}$$

Figure 3.8 exhibits the results for 5 fitting year spans. We apply the LR-RW model to forecast mortality rates for $2001 - 2010$, generated from 5 fitting year spans: $1951 - 2000, 1961 - 2000, 1971 - 2000, 1981 - 2000$, and $1991 - 2000$, for a 40-year-old UK male.



(a) $ln(\hat{m}_{x,t_0+\tau})$



(b) $\hat{\sigma}^2_{x,\tau} = \hat{V}ar\left[ln(\tilde{m}_{x,t_0+\tau})\right]$

Figure 3.8: $ln(\hat{m}_{x,t_0+\tau})$ and $\hat{V}ar\left[ln(\tilde{m}_{x,t_0+\tau})\right]$ for $x = 40$ based on 5 fitting year spans

3.4 Bühlmann Credibility Theory

With wide applications in property and casualty (P&C) insurance, credibility theory works as an approach to performing prospective rating, i.e., to calculating different premium levels that will be charged to policyholders. Take P&C prospective rating as an example, it combines the policyholder's own experience with the expected value of the risk class which the policyholder belong to. Under the credibility theory, a partial weight (the credibility factor Z) is assigned to the policyholder's own data (the sample mean of these observations, denoted as $\bar{Y}$), and the rest of the weight $(1-Z)$ is assigned to the expectation of the risk class (denoted as μ). Accordingly, the prospective premium to be charged is $C = Z \times \bar{Y} + (1 - Z) \times \mu$.

The reason to assign such a credibility factor is that, although the risk class is assumed homogeneous with respect to the underwriting characteristics, yet there still remains a little heterogeneity when it comes to each specific policyholder's behavior and risk characteristics. Insurance company needs to take care of both sides of information in order to determine an appropriate premium level. In this sense, credibility method not only captures the risk characteristics of the whole risk class, but also takes each policyholder's specific experience into consideration. Thus in this way, it can return better results and reduce the average error. Bühlmann (1967) developed a credibility model with a statistical framework, called "the Bühlmann credibility model" (Klugman, Panjer and Willmot, 2012). The subsections below will demonstrate in details how to apply B*ü*hlmann's credibility model to quantitatively formulate the experience data as well as the estimator which is produced from other information.

3.4.1 Credibility Estimation

Suppose that we have collected observed values $Y_{x,j}$ for age $x = x_0, x_0+1, \cdots, x_0+m-1$ (m represents the length of the age span) from the j^{th} fitting year span, $j = 1, 2, \cdots, J$. It can also be interpreted as a random sample of size J for age x available for our credibility estimation. We have an expected rate μ corresponding to this risk class (age) and plan to determine an estimation also based on all these different observed values. To proceed with the Bühlmann credibility framework, we assume that the mortality rate in the risk class is characterized by a risk parameter Θ_x. Furthermore, we assume that the mortality rates in the sample corresponding to different fitting year spans are conditionally independent and identical; specifically, the mortality rates $Y_{x,1}|\Theta_x, \cdots, Y_{x,J}|\Theta_x$ are independent and identical as $Y_{x,*}|\Theta_x$. This layout provides us a model-based framework to determine the credibility estimate.

While there are many different approaches to calculating the credibility factor, Bühlmann (1967) derived a formula for the credibility factor using a linear estimator that minimizes the expected squared error. It is suggested that since we have observed $Y_{x,1}, \cdots, Y_{x,J}$, we can conduct an approximation of $\mu(\Theta_x) = E[Y_{x,j}|\Theta_x] = E[Y_{x,*}|\Theta_x]$ by a linear function of the observed data. That is, we will choose α_0, α_1, $\cdots$, α_J to minimize the mean of squared error Q

as follow (see Klugman et al., 2012):

$$Q = E\left\{\left[\mu(\Theta_x) - \alpha_0 - \sum_{j=1}^{J} \alpha_j Y_{x,j}\right]^2\right\},$$

where the expectation is over the joint distribution of $Y_{x,1}, \cdots, Y_{x,J}$, and Θ_x. In this sense, the squared error is the average over all possible values of Θ_x and all sample values available. To minimize Q, we take the derivatives and set it to be 0; then we get

$$E[Y_{x,*}] = E[\mu(\Theta_x)] = \hat{\alpha}_0 + \sum_{j=1}^{J} \hat{\alpha}_j \cdot E[Y_{x,j}], \tag{3.29}$$

where $\hat{\alpha}_0, \hat{\alpha}_1, \cdots, \hat{\alpha}_J$ are the values of $\alpha_0, \alpha_1, \cdots, \alpha_J$ which minimize Q. For $i = 1, 2, \cdots, J$, we have

$$E[\mu(\Theta_x) \cdot Y_{x,i}] = E[Y_{x,*} \cdot Y_{x,i}], \tag{3.30}$$

and

$$Cov(Y_{x,i}, Y_{x,*}) = \sum_{j=1}^{J} \hat{\alpha}_j \cdot Cov(Y_{x,i}, Y_{x,j}). \tag{3.31}$$

Equation (3.29) and the J equations (3.31) together are called the normal equations. These equations can be solved for $\hat{\alpha}_0$, $\hat{\alpha}_1$, $\cdots$, $\hat{\alpha}_J$ to reach the credibility estimate $\hat{\alpha}_0 + \sum_{j=1}^{J} \hat{\alpha}_j \cdot Y_{x,j}$ which is the best linear estimator of the hypothetical mean $\mu(\Theta_x)$.

Special case: If $E[Y_{x,j}] = \mu$, $Var(Y_{x,j}) = \sigma^2$ and $Cov(Y_{x,i}, Y_{x,j}) = \rho \cdot \sigma^2$ for $i \neq j$, where the correlation coefficient ρ satisfies $-1 < \rho < 1$. Then

$$\hat{\alpha}_0 = \frac{(1-\rho)\cdot\mu}{1-\rho+n\rho}, \quad \hat{\alpha}_j = \frac{\rho}{1-\rho+n\rho}, \quad j = 1, 2, \cdots, J.$$

The credibility estimate is then

$$\hat{\alpha}_0 + \sum_{j=1}^{J} \hat{\alpha}_j \cdot Y_{x,j} = \frac{1-\rho}{1-\rho+n\rho}\cdot\mu + \frac{\rho}{1-\rho+n\rho}\cdot\sum_{j=1}^{J} Y_{x,j} = (1-Z)\cdot\mu + Z\cdot\bar{Y_x},$$

where $Z = n\rho/(1-\rho+n\rho)$ and $\bar{Y_x} = \frac{1}{J}\sum_{j=1}^{J} Y_{x,j}$. Thus, if $0 < \rho < 1$, then $0 < Z < 1$ and the credibility estimate is a weighted average of the sample mean $\bar{Y_x}$ and the expectation $E[Y_{x,*}] = \mu$.

3.4.2 Parametric Bühlmann Model

This section will present the inference and logics to construct the parametric Bühlmann model (Bühlmann, 1967) and compute the Bühlmann credibility factor so that we can extend the model under a more complex context based on similar procedures. Assume that for a specific age x, $Y_{x,1}|\Theta_x$, $Y_{x,2}|\Theta_x$, $\ldots$, $Y_{x,J}|\Theta_x$ are independent and identically distributed, and that

$\Theta_x, \cdots, \Theta_{x+m-1}$ are independent and identically distributed. In order to apply the parametric Bühlmann model, both distributions of $Y_{x,j}|\Theta_x$ and Θ_x must be given.

Denote
(1) the hypothetical mean $\mu(\Theta_x) = E[Y_{x,j}|\Theta_x]$;
(2) the process variance $v(\Theta_x) = Var[Y_{x,j}|\Theta_x]$;
(3) the expected value of the hypothetical means (the collective premium) $\mu = E[\mu(\Theta_x)]$;
(4) the expected value of the process variances $v = E[v(\Theta_x)] = E[Var(Y_{x,j}|\Theta_x)]$; and
(5) the variance of the hypothetical means $a = Var[\mu(\Theta_x)] = Var[E(Y_{x,j}|\Theta_x)]$.

The mean and variance of $Y_{x,j}$ can be obtained as follows:

$$E[Y_{x,j}] = E[E(Y_{x,j}|\Theta_x)] = E[\mu(\Theta_x)] = \mu,$$

and

$$\begin{aligned} Var[Y_{x,j}] &= E[Var(Y_{x,j}|\Theta_x)] + Var[E(Y_{x,j}|\Theta_x)] \\ &= E[v(\Theta_x)] + Var[\mu(\Theta_x)] = v + a, \quad j = 1, 2, \ldots, J. \end{aligned}$$

Also, for $i \neq j$, $Cov[Y_{x,i}, Y_{x,j}] = Var[\mu(\Theta_x)] = a$. From the special case, we have $\sigma^2 = v + a$ and $\rho = a/(v + a)$. The credibility premium becomes

$$\hat{\alpha}_0 + \sum_{j=1}^{n} \hat{\alpha}_j \cdot Y_{x,j} = Z \cdot \bar{Y}_x + (1 - Z) \cdot \mu, \tag{3.32}$$

a linear function of $\bar{Y}_x$ with the slope Z and the intercept $(1 - Z) \cdot \mu$, where

$$Z = \frac{J\rho}{1 - \rho + J\rho} = \frac{J}{J + v/a} = \frac{J}{J + k} \tag{3.33}$$

is called the Bühlmann credibility factor, and

$$k = \frac{v}{a} = \frac{E[v(\Theta_x)]}{Var[\mu(\Theta_x)]} = \frac{E[Var(Y_{x,j}|\Theta_x)]}{Var[E(Y_{x,j}|\Theta_x)]}.$$

From (3.32), we can see that it is a weighted average of sample mean $\bar{Y}_x$ and the expected rate μ. This formula turns out quite desirable for insurance companies since both sample values and expected value are taken into account. From (3.33) for the credibility factor Z, it can be seen that Z approaches to 1 as J increases to infinity, which implies that companies will pay their attention to the sample mean. From the other side, if all the $\mu(\Theta_{x_0}), \cdots, \mu(\Theta_{x_0+m-1})$ display a homogeneity to a great extent, then the hypothetical means $\mu(\Theta_x) = E[Y_{x,j}|\Theta_x]$ will not vary much from μ. In this case, such small variation will produce a small a and a large k, and thus a low Z. It implies that insurance companies will put little focus on the sample data, but instead concentrate on the risk characteristics of ages $x_0, \cdots, x_0 + m - 1$.

3.4.3 Non-parametric Bühlmann Model

Let $Y_{x,j}$ denote the random variable representing the forecasted mortality rate for age x in some year from the j^{th} fitting year span for $x = x_0, x_0 + 1, \ldots, x_0 + m - 1$ and $j = 1, \ldots, J$, where $J \geq 2$ and $m \geq 2$. Also, let $Y_x = (Y_{x,1}, \ldots, Y_{x,J})$ be the mortality random sample that insurance companies already collected by forecasting mortality rates from J fitting year spans for age x in some year. We would like to compute the non-parametric Bühlmann estimate $E[Y_{x,\circ}|Y_x]$ for all x, $x = x_0, x_0 + 1, \ldots, x_0 + m - 1$, where $Y_{x,\circ}$ is the true mortality rate for age x. For the non-parametric Bühlmann model, the data needs to satisfy all the assumptions below:

(1) $Y_{x_0}, \ldots, Y_{x_0+m-1}$ are independent;
(2) For $x = x_0, x_0 + 1, \ldots, x_0 + m - 1$, the distribution of each element $Y_{x,j}$ $(j = 1, \ldots, J)$ of Y_x depends on an (unknown) risk parameter Θ_x;
(3) $\Theta_{x_0} \ldots \Theta_{x_0+m-1}$ are independent and identically distributed random variables;
(4) Given x, $Y_{x,1}|\Theta_x, \ldots, Y_{x,J}|\Theta_x$ are independent; and
(5) Each combination of age and fitting year span has an equal number of underlying exposure units.

For $x = x_0, x_0 + 1, \ldots, x_0 + m - 1$ and $j = 1, \ldots, J$, define the hypothetical mean by $\mu(\Theta_x) = E[Y_{x,j}|\Theta_x]$ and the process variance as $v(\Theta_x) = Var[Y_{x,j}|\Theta_x]$. Let
$\mu = E[\mu(\Theta_x)] = E\{E[Y_{x,j}|\Theta_x]\} = E[Y_{x,j}]$, the expected value of the hypothetical means,
$a = Var[\mu(\Theta_x)] = Var[E(Y_{x,j}|\Theta_x)]$, the variance of the hypothetical means, and
$v = E[v(\Theta_x)] = E[Var(Y_{x,j}|\Theta_x)]$, the expected value of the process variances.

In this case, the non-parametric Bühlmann estimate of the mortality rate for age x in some year τ is

$$E[Y_{x,\circ}|Y_x] = \hat{Z} \times \bar{Y}_x + (1 - \hat{Z}) \times \hat{\mu}, \quad x = x_0, x_0 + 1, \ldots, x_0 + m - 1,$$

where $\bar{Y}_x = \dfrac{1}{J} \sum_{j=1}^{J} Y_{x,j}$, $\hat{Z} = \dfrac{J}{J + \hat{k}}$ and $\hat{k} = \dfrac{\hat{v}}{\hat{a}}$. Note that it is possible that $\hat{a}$ could be negative due to the substraction. When that happens, it is customary to set $\hat{a} = 0$, implying $\hat{Z} = 0$, and the Bühlmann estimate becomes $\hat{\mu} = \bar{Y}$. Below are the formulas to estimate μ, v and a.

In this project, we apply the principles and structures of the non-parametric Bühlmann model; however, assumption (4) is violated since our forecasted mortality rates $Y_{x,1}|\Theta_x, \ldots, Y_{x,J}|\Theta_x$ partially share the same fitting year span containing same fitted data, so our approach is denominated as "pseudo non-parametric Bühlmann estimation" in this sense. More details will be demonstrated in next section.

Table 3.1: Estimation for μ, v and a

$(Y_{x_0,1} \quad Y_{x_0,2} \quad \cdots \quad Y_{x_0,J})$	$\bar{Y}_{x_0} = \frac{1}{J}\sum_{j=1}^{J} Y_{x_0,j}$, $\hat{v}_{x_0} = \frac{1}{J-1}\sum_{j=1}^{J}(Y_{x_0,j} - \bar{Y}_{x_0})^2$;
$(Y_{x_0+1,1} \quad Y_{x_0+1,2} \quad \cdots \quad Y_{x_0+1,J})$	$\bar{Y}_{x_0+1} = \frac{1}{J}\sum_{j=1}^{J} Y_{x_0+1,j}$, $\hat{v}_{x_0+1} = \frac{1}{J-1}\sum_{j=1}^{J}(Y_{x_0+1,j} - \bar{Y}_{x_0+1})^2$;
$\vdots \quad \vdots \quad \vdots \quad \vdots$	$\vdots$
$(Y_{x_0+m-1,1} \quad Y_{x_0+m-1,2} \quad \cdots \quad Y_{x_0+m-1,J})$	$\bar{Y}_{x_0+m-1} = \frac{1}{J}\sum_{j=1}^{J} Y_{x_0+m-1,j}$, $\hat{v}_{x_0+m-1} = \frac{1}{J-1}\sum_{j=1}^{J}(Y_{x_0+m-1,j} - \bar{Y}_{x_0+m-1})^2$.
$\hat{a} = \frac{1}{m-1}\sum_{x=x_0}^{x_0+m-1}(\bar{Y}_x - \bar{Y})^2 - \frac{\hat{v}}{J}$	$\hat{\mu} = \bar{Y} = \frac{1}{m}\sum_{x=x_0}^{x_0+m-1}\bar{Y}_x$, $\hat{v} = \frac{1}{m}\sum_{x=x_0}^{x_0+m-1}\hat{v}_x$ $= \frac{1}{m(J-1)}\sum_{x=x_0}^{x_0+m-1}\sum_{j=1}^{J}(Y_{x,j} - \bar{Y}_x)^2$.

3.5 Applying Pseudo Non-Parametric Bühlmann Estimation

From (a) of Figures (3.3), (3.5) and (3.8), it can be seen that different lengths of the fitting year span will generate distinct slopes when forecasting future mortality rates for year $t_0 + \tau$ if t_0, the ending year of the fitting year span, is fixed. Roughly speaking, the longer the fitting year span we choose, the steeper the forecasted slope it will reflect from the figures. Faced with these varying forecasted values, which fitting year span should we choose in order to achieve the best forecasting result? This implies significant financial concerns for insurance companies since accurate mortality estimation is one of the most essential parts for them to conduct product pricing. An appropriate premium level helps insurance operate stronger business and stay more competitive in the market. Thus, companies need to develop an approach to determining the fitting year span which can return an accurate mortality estimate.

For a specific case, companies can apply all possible fitting year spans, i.e., $1951 - 2000$, $1952 - 2000$, $1953 - 2000$, ..., until $1999 - 2000$. They can apply all these fitting year spans into a mortality model separately to get the forecasted mortality rates for years $2001 - 2010$, and compare the results with the true mortality rates by the mean absolute percentage error (MAPE) defined in (4.2). Through the MAPE comparisons, companies can find out which fitting year span does the best job; the fitting year span that returns the smallest MAPE implies the highest prediction accuracy. However, this "best fitting year span" will only apply to the age span from a certain population selected by the company; it is not guaranteed to remain the best fitting year once the age span or the population changes. As a matter of fact, there does not exist a universally best fitting year span that will always returns the most accurate prediction result. Each case has its own best fitting year span.

To deal with such varying forecasting results, this project collects forecasted mortality rates for the same age span but different fitting year spans with an equal ending year, applies the non-parametric Bühlmann credibility to the forecasted values, and calculates corresponding estimates. In this way, it offers a "compromise" among all possible forecasted values, which, from the outcome, will generate an overall better forecasting performance than those generated from different fitting year spans. Below are sections demonstrating the layout and formulas to construct such model with supportive numerical results illustrated in the next chapter.

3.5.1 Construction of Non-Parametric Bühlmann Estimation

Figure 3.9 demonstrates how the sample is constructed in this case. From each of the different fitting year spans, we take the corresponding forecasted mortality rates for the same future year to form a sample for the year; then we conduct the non-parametric Bühlmann estimation based on the sample. For example, assume we have 48 fitting year spans: $1951-2000$, $1952-2000$, ..., until $1998-2000$, each of which can be used to forecast mortality rates for years $2001-2010$. In this sense, we can collect 48 columns of forecasted mortality rates to form a matrix for each of years $2001-2010$. In the matrix, each column represents the forecasted mortality rates for the year across the same age span. Then the non-parametric Bühlmann estimation will be applied based on the matrix.

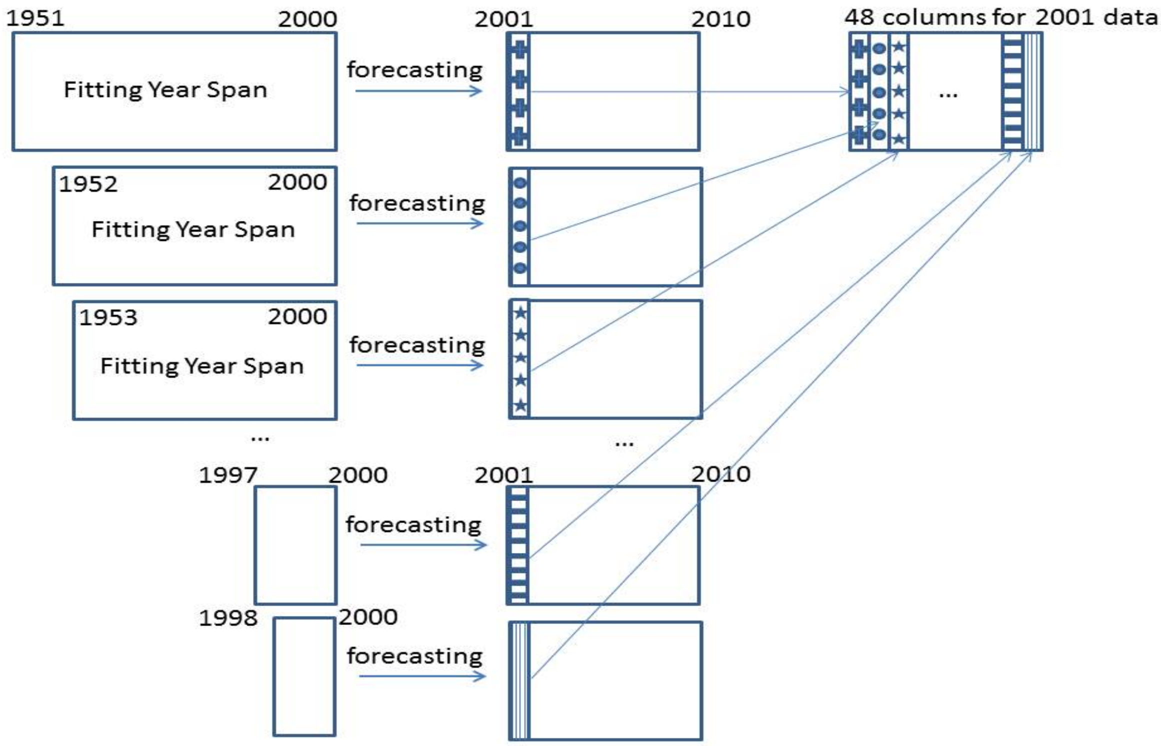


Figure 3.9: Framework of collecting data for the non-parametric Bühlmann model

Denote $Y_{x,j,\tau}$ the forecasted mortality rate for age x in year $t_0+\tau$ generated from the j^{th} fitting year span, $\tau = 1, 2, ..., T$ and $j = 1, 2, ..., J$. For the Lee-Carter and LR-RW models, $Y_{x,j,\tau} = ln(\hat{m}_{x,j,t_0+\tau})$ is given in (3.5) and (3.25), respectively. For the CBD model, $Y_{x,j,\tau} = logit(\hat{q}_{x,j,t_0+\tau})$ is given in (3.15). Table 3.2 is the layout to conduct the non-parametric Bühlmann estimation for year $t_0+\tau$.

In credibility theory setting, μ_τ is the expected value of the hypothetical means, a_τ is the variance of the hypothetical means, and v_τ is the expected value of the process variances. The $\hat{\mu}_\tau$, $\hat{a}_\tau$, and $\hat{v}_\tau$ in Table 3.2 are unbiased estimators, and thus, the non-parametric Bühlmann

Table 3.2: Framework for the non-parametric Bühlmann estimate

$(Y_{x_0,1,\tau} \quad Y_{x_0,2,\tau} \quad \ldots \quad Y_{x_0,J,\tau})$	$\bar{Y}_{x_0,\cdot,\tau} = \frac{1}{J}\sum_{j=1}^{J} Y_{x_0,j,\tau}$, $\hat{v}_{x_0,\tau} = \frac{1}{J-1}\sum_{j=1}^{J}\left(Y_{x_0,j,\tau} - \bar{Y}_{x_0,\cdot,\tau}\right)^2$;
$(Y_{x_0+1,1,\tau} \quad Y_{x_0+1,2,\tau} \quad \ldots \quad Y_{x_0+1,J,\tau})$	$\bar{Y}_{x_0+1,\cdot,\tau} = \frac{1}{J}\sum_{j=1}^{J} Y_{x_0+1,j,\tau}$, $\hat{v}_{x_0+1,\tau} = \frac{1}{J-1}\sum_{j=1}^{J}\left(Y_{x_0+1,j,\tau} - \bar{Y}_{x_0+1,\cdot,\tau}\right)^2$;
$\vdots$	$\vdots$
$(Y_{x_0+m-1,1,\tau} \quad Y_{x_0+m-1,2,\tau} \quad \ldots \quad Y_{x_0+m-1,J,\tau})$	$\bar{Y}_{x_0+m-1,\cdot,\tau} = \frac{1}{J}\sum_{j=1}^{J} Y_{x_0+m-1,j,\tau}$, $\hat{v}_{x_0+m-1,\tau} = \frac{1}{J-1}\sum_{j=1}^{J}\left(Y_{x_0+m-1,j,\tau} - \bar{Y}_{x_0,\cdot,\tau}\right)^2$;
$\hat{a}_\tau = \frac{1}{m-1}\sum_{x=x_0}^{x_0+m-1}\left(\bar{Y}_{x,\cdot,\tau} - \bar{Y}_{\cdot,\cdot,\tau}\right)^2 - \hat{v}_\tau/J$	$\hat{\mu}_\tau = \bar{Y}_{\cdot,\cdot,\tau} = \frac{1}{m}\sum_{x=x_0}^{x_0+m-1} Y_{x,\cdot,\tau}$, $\hat{v}_\tau = \frac{1}{m}\sum_{x=x_0}^{x_0+m-1}\hat{v}_{x,\tau}$ $= \frac{1}{m(J-1)}\sum_{x=x_0}^{x_0+m-1}\sum_{j=1}^{J}\left(Y_{x,j,\tau} - \bar{Y}_{x,\cdot,\tau}\right)^2$.

estimate for a age x in year $t_0 + \tau$ is

$$C_{x,\tau} = \hat{Z}_{x,\tau} \times \bar{Y}_{x,\cdot,\tau} + \left(1 - \hat{Z}_{x,\tau}\right) \times \bar{Y}_{\cdot,\cdot,\tau},$$

where $\hat{Z}_{x,\tau} = \frac{J}{J+\hat{k}_\tau}$ and $\hat{k}_\tau = \hat{v}_\tau/\hat{a}_\tau$.

3.5.2 Notation and Interpretations

According to the credibility theory, $P_c = Z \times \bar{Y} + (1 - Z) \times \mu$, where $\bar{Y}$ stands for the mean of the observed data $Y_1, \cdots, Y_J$ from a policyholder, Y_j is the observed value for the j^{th} year, and μ is the group rate that indicates the risk characteristic for the group. The non-parametric Bühlmann model is popularly applied in property and casualty insurance in order to perform prospective rating, i.e., to determine the premium to charge for the $(J+1)^{th}$ year. In this sense, the credibility premium charged to the policyholder for the $(J+1)^{th}$ year is denoted by $E[Y_{J+1}|Y_1, \ldots, Y_J]$. However, the Bühlmann estimation applied in this project has different implication, and it is of vital importance to realize the different interpretation. In our case, the forecasted values are made for the same year by some specific mortality model, but generated from different fitting year spans. In this sense, we treat the forecasted mortality rates as our sample drawn for the policyholders aged x in the same age span, and want to get a "compromise average" from the sample. For the case demonstrated in Figure 3.9, we collect 48 column vectors of forecasted mortalities for the year 2001, generated from 48 fitting year spans: $1951 - 2000$, $1952 - 2000$, $\cdots$, and $1998 - 2000$. Applying the non-parametric Bühlmann estimation to these 48 column vectors of values can help us determine the final mortality estimate for year 2001.

This interpretation violates one of the assumptions required for the non-parametric Bühlmann model, as our forecasted mortality rates $Y_{x,1}|\Theta_x, \ldots, Y_{x,J}|\Theta_x$ are obtained from the mortality data in partially overlapped fitting year spans, so they cannot be regarded as conditionally independent. However, we still apply the non-parametric Bühlmann estimation, and follow the same procedures to compute the credibility factor; so our approach is denominated as "pseudo non-parametric Bühlmann estimation." Such conduction is practically applicable for insurance companies; for example, when an insurance institute plans to estimate the mortality rates for an age span in 2010, and it already collects J columns of different estimates for the age span provided by J insurance companies. In this situation we can apply the pseudo non-parametric Bühlmann estimation from these J columns of values, and thus determine the final estimate of the forecasted mortality rate for the age span in 2010.

Given a sequence of projected mortality rates for some unknown age x in a specific year from different fitting year spans, we apply a pseudo non-parametric Bühlmann approach to getting the Bühlmann estimate for that age. Model specifications and numerical results are presented in the next chapter.

Chapter 4

Numerical Results

This chapter presents the numerical results of the models constructed in Chapter 3 with both visualized plots and table summaries, followed by corresponding comparison and analysis. We apply the data from 3 countries, UK, USA and Japan, for male (M) and female (F) separately, and conduct mortality forecasting under the Lee-Carter (hereafter the LC model), CBD, and LR-RW models, respectively. The age span remains to be $25 - 84$, but the fitting year span varies. Table 4.1 displays 3 forecasting year spans conducted in this project, where T represents the length of the forecasting year span, t_0 is the ending year of the fitting year span, and J stands for the total number of the fitting year spans.

Table 4.1: Summary of 3 forecasting year spans

T	10	30	50
t_0	2000	1980	1960
J	48	28	8
Fitting year spans	1951-2000 1952-2000 ⋮ 1998-2000	1951-1980 1952-1980 ⋮ 1978-1980	1951-1960 1952-1960 ⋮ 1958-1960
Forecasting year spans	2001-2010	1981-2010	1961-2010

Note that both the LC and CBD models require at least two-year data to estimate the associated parameters and make predictions. In this case, the theoretical shortest fitting year span is 2 years. However, a two-year fitting practically returns quite large forecasting errors that turn out to be a distraction to our comparison and analysis. Based on this fact, we set the shortest fitting year span to be 3 years long, and that is why $J = 48$, 28, 8 in the table above, which will generate 48, 28 and 8 column vectors of forecasted mortality rates, respectively.

4.1 Model Specification

Based on the forecasted mortality rates given in (3.5), (3.15) and (3.25) from the three mortality models, we also apply a grouping strategy to the age span and compare the corresponding performances of the strategies with or without grouping. To be more specific, there are 2 stages in this project: in Stage I the original age span $[25, 84]$ is either split into 12 age groups (the strategy is denoted by G) or ungrouped (the strategy is denoted by U) when fitting and forecasting mortality rates with a mortality model; in Stage II, the same grouping strategy is adopted again when applying the non-parametric Bühlmann method to a matrix of forecasted mortality rates (see Figure 3.9). Then 4 combinations are generated as shown in Table 4.2 below:

Table 4.2: Framework for Stage I and Stage II

Stage I Forecasting using mortality models: LC, CBD and LR-RW	Stage II Forecasting using the non-parametric Bühlmann credibility theory
Ungrouped age span (U)	Ungrouped age span (UU) Grouped age span (UG)
Grouped age span (G)	Ungrouped age span (GU) Grouped age span (GG)

The purpose of conducting such a grouping strategy is to better identify homogeneous risks within each group. As mentioned before, the Bühlmann credibility factor Z is a weight assigned to individual observations and $1 - Z$ stays with the group rate which represents the risk characteristics from the group. It is theoretically assumed that the wider range a group covers, the higher degree of heterogeneity the data in the group will reflect. In this sense, we want to study if there will be any noteworthy improvements after dividing the data into 12 smaller groups. In the ideal case, such a smaller group is supposed to display more homogeneity and better represent the risk characteristics, and thus reduce the variability in observations. Such a grouping strategy will generate a more meticulous group rate μ for each of the 12 groups, and thus we can compare the corresponding forecasting performance with that based on the original (ungrouped) age span. The following two subsections illustrate this grouping strategy in more details; numerical results are also provided at the end of this chapter.

4.1.1 Stage I: Grouped vs Ungrouped Age Span in Models

As mentioned above, Stage I focuses on the 3 mortality models with 12 grouped age spans or the original age span. For the age span $[25, 84]$ selected in this project, we split it into 12 groups with 5 ages being compiled into one group (the 12 groups are $[25, 29]$, $[30, 34]$, $[35, 39]$, $\cdots$, $[80, 84]$), and apply the LC, CBD and LR-RW models to both the original age span and the 12 grouped age spans. Consequently, there will be two matrices of forecasted mortality rates generated from each of the three mortality models in Stage I. Figure 4.1 exhibits the framework of the grouped age spans.

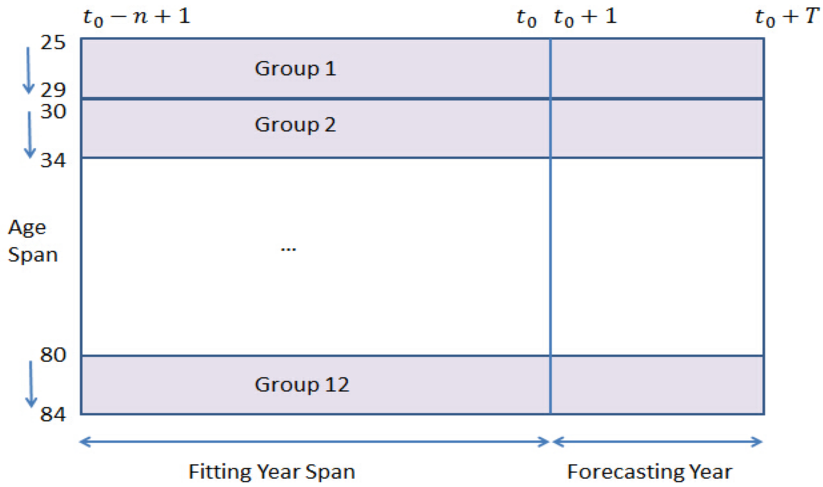


Figure 4.1: Framework of the grouped age spans

The methods of estimating parameters of the three mortality models remain the same under such a grouping arrangement except for different starting and ending values of the summation function. For example, $\hat{a}_x^G$ and $\hat{a}_x^U$ for the LC model are exactly the same since $\hat{a}_x^G = \frac{1}{n}\sum_{t=t_0-n+1}^{t_0} ln(m_{x,t})$ for $x = x_0, x_0+1, ..., x_0+m-1$; the formula has nothing to do with age x so $\hat{a}_x$'s value is invariant with a grouping strategy. However, the estimate of k_t depends on the age span; $\hat{k}_t$'s under the grouping strategy are calculated as follows:

$$
\begin{aligned}
\hat{k}_t^{G_1} &= \sum_{x=25}^{29}\left[ln(m_{x,t}) - \hat{a}_x^G\right], \quad t = t_0 - n + 1, \cdots, t_0; \\
\hat{k}_t^{G_2} &= \sum_{x=30}^{34}\left[ln(m_{x,t}) - \hat{a}_x^G\right], \quad t = t_0 - n + 1, \cdots, t_0; \\
&\vdots \\
\hat{k}_t^{G_{12}} &= \sum_{x=80}^{84}\left[ln(m_{x,t}) - \hat{a}_x^G\right], \quad t = t_0 - n + 1, ..., t_0.
\end{aligned} \tag{4.1}
$$

For each age $x = x_0, \cdots, x_0 + m - 1$, $\hat{b}_x^G$ is obtained by regressing $\left[ln(m_{x,t}) - \hat{a}_x^G\right]$ on $\hat{k}_t^{G_i}$ where x is in the i^{th} group for some $i \in \{1, 2, \cdots, 12\}$. The same procedure of estimating parameters under the grouping strategy also applies to the CBD and LR-RW models. The dimension of the matrix (see Figure 3.9) of forecasted mortality rates from J fitting year spans remains the same no matter the age span is grouped or ungrouped. More specifically, for $\tau = 1, \cdots, T$, a mortality model will always generate a matrix of forecasted mortality rates of dimension $60 \times J$ from J fitting year spans, so there will be $60 \times J \times T$ predicted mortality rates in total for all forecasting years and all fitting year spans regardless of grouped or ungrouped age span.

4.1.2 Stage II: Grouped vs Ungrouped Age Span for Bühlmann Estimation

For Stage II we also apply the same grouping strategy to the age spans with the non-parametric Bühlmann method: either splitting the 60-year long age span into 12 groups with 5 consecutive ages being assigned into one group, or keeping the original age span unchanged. Based on such setting, there will be 4 combinations for Stage II, UU, UG, GU and GG, as illustrated in Table 4.2. With grouped and ungrouped age spans, Stage I returns two corresponding matrices of dimension $60 \times J \times T$. For each $\tau = 1, \cdots, T$, by fitting either a matrix of $60 \times J$ once (U) or 12 matrices of $5 \times J$ (G) separately into the non-parametric Bühlmann model, we get two corresponding column vectors of Bühlmann estimates for the age span $[25, 84]$.

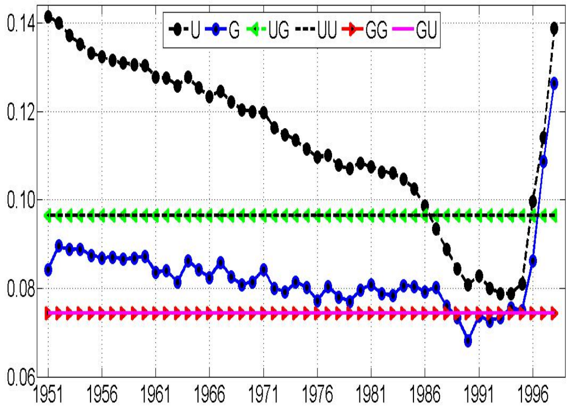
As mentioned earlier, the non-parametric Bühlmann estimate "smoothes" the unpredictable variations resulting from the selection of different fitting year spans, and thus can provide a better forecasting outcome. In order to have a sense of how well the non-parametric Bühlmann estimate performs, we apply the mean absolute percentage error (MAPE) to the forecasted mortality rates to compare the forecasting performances. MAPE is a statistic index defined below to measure the accuracy of the fitted/forecasted values with the true values. Before applying MAPE, we need to convert the forecasted $ln(\hat{m}_{x,t_0+\tau,j})$'s (for the LC and LR-RW models) and $logit(\hat{q}_{x,t_0+\tau,j})$'s (for the CBD model) to the corresponding $\hat{q}_{x,t_0+\tau,j}$'s by $\hat{q}_{x,t_0+\tau,j} = 1 - exp[-e^{\ln(\hat{m}_{x,t_0+\tau,j})}]$ and $\hat{q}_{x,t_0+\tau,j} = e^{logit(\hat{q}_{x,t_0+\tau,j})}/[1 + e^{logit(\hat{q}_{x,t_0+\tau,j})}]$, respectively, and then compare them with the true $q_{x,t_0+\tau}$'s. For the V strategy ($V = U, G$) in Stage I,

$$\begin{aligned}
MAPE^V_{x,j} &= \frac{1}{T}\sum_{\tau=1}^{T}\left|\frac{\hat{q}^V_{x,t_0+\tau,j} - q_{x,t_0+\tau}}{q_{x,t_0+\tau}}\right|, \quad x = x_0, \cdots, x_0+m-1, \quad j = 1, \cdots, J; \\
MAPE^V_{\tau,j} &= \frac{1}{m}\sum_{x=x_0}^{x_0+m-1}\left|\frac{\hat{q}^V_{x,t_0+\tau,j} - q_{x,t_0+\tau}}{q_{x,t_0+\tau}}\right|, \quad \tau = 1, \cdots, T, \quad j = 1, \cdots, J; \\
MAPE^V_{j} &= \frac{1}{T}\sum_{\tau=1}^{T} MAPE^V_{\tau,j} = \frac{1}{m}\sum_{x=x_0}^{x_0+m-1} MAPE^V_{x,j}.
\end{aligned} \tag{4.2}$$

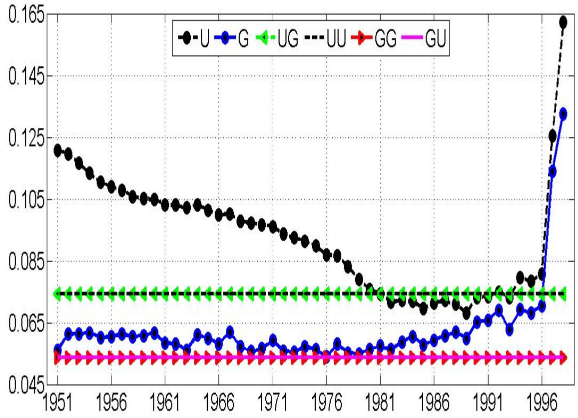
Note that the $MAPE^V_j$ in (4.2) above is based on the j^{th} fitting year span for the V strategy in Stage I. For the VW strategy in Stage II where $(V, W) = (U, U)$, (U, G), (G, U) and (G, G), the corresponding MAPE formula is

$$MAPE^{VW} = \frac{1}{mT}\sum_{\tau=1}^{T}\sum_{x=x_0}^{x_0+m-1}\left|\frac{\hat{q}^{VW}_{x,t_0+\tau} - q_{x,t_0+\tau}}{q_{x,t_0+\tau}}\right| \tag{4.3}$$

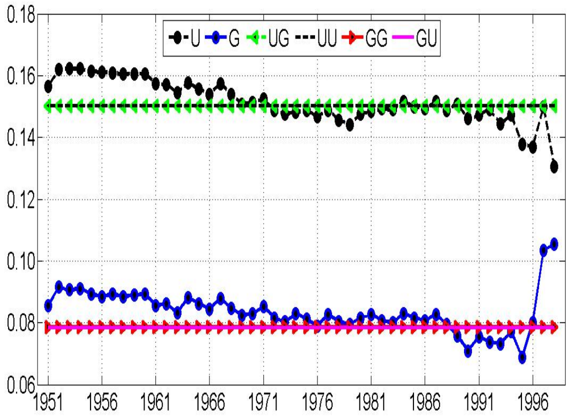
where $\hat{q}^{VW}_{x,t_0+\tau}$ is the non-parametric Bühlmann estimate of $q_{x,t_0+\tau}$ for age x in year $t_0 + \tau$ using the VW strategy. As illustrated in Table 4.1, J is the total number of all fitting year spans involved, and T is selected to be 10, 30 and 50 with corresponding $J = 48$, 28 and 8, respectively. Figures 4.2-4.10 are the $MAPE^V_j$ and $MAPE^{VW}$ plots for 3 countries (UK, USA and Japan) and 2 genders regarding to the 3 pairs of (T, J)'s, $(10, 48)$, $(30, 28)$ and $(50, 8)$.



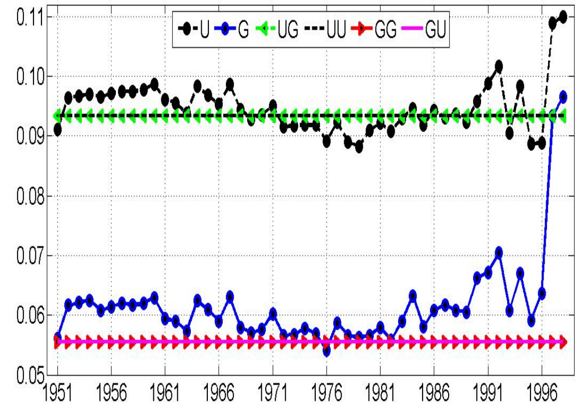
(a) UK_M, LC



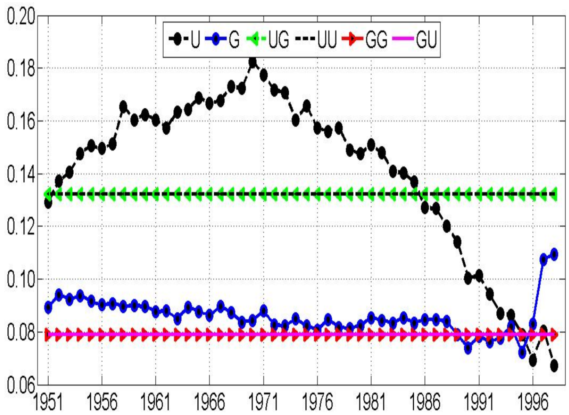
(b) UK_F, LC



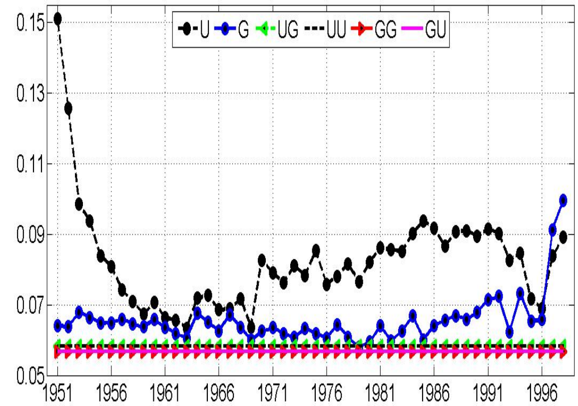
(c) UK_M, CBD



(d) UK_F, CBD



(e) $UK_M, LR-RW$



(f) $UK_F, LR-RW$

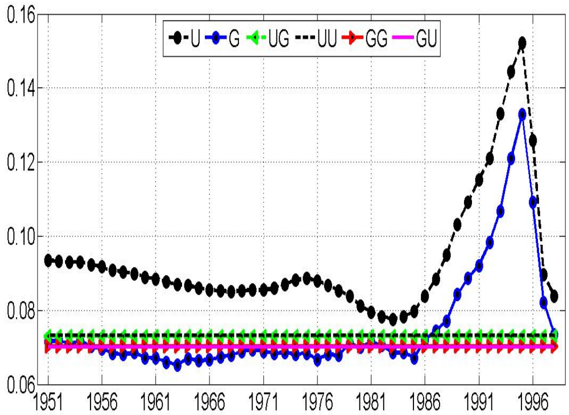
Figure 4.2: $MAPE_j$'s vs $1950+j$ for UK with $T=10$ and year span $=[1950+j, 2000]$

Figure 4.2 demonstrates the $MAPE$ plots with the forecasting year span 2001−2010 ($T = 10$) and $J = 48$ for UK males and females under each of the three mortality models. We can see that compared with $MAPE_j^U$'s, $MAPE_j^G$'s help to smooth the fluctuations by producing a more stable plot with smaller $MAPE$ values. However, while $MAPE_j^G$ generally returns better outcomes than $MAPE_j^U$ for 48 $MAPE_j$ values corresponding to 48 fitting year spans, both of them depend on the selection of fitting year spans. Such fluctuations particularly occur for very short fitting year spans. For example, for the fitting year spans 1996 − 2000, 1997 − 2000 and 1998 − 2000, the $MAPE_j^V$ values under the LC model jump sharply as shown in the steep right tails of Figures 4.2 (a) and (b), while the previous $MAPE_j^V$ values display a smoothly decreasing trend.

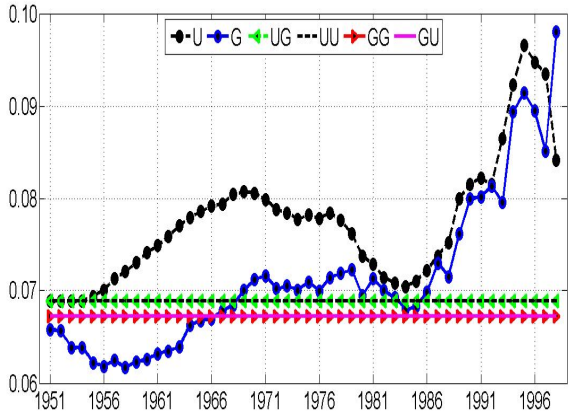
Besides the fluctuations shown from $MAPE_j^U$'s and $MAPE_j^G$'s, it can be also seen that the three mortality models reveal different reactions towards Stage I and Stage II. In Figure 4.2, GU and GG strategies produce much lower $MAPE$ values than UU and UG strategies except for the UK females under the LR-RW model, which implies that the grouping strategy in Stage I takes more important role in reducing forecasting errors and improving prediction performance. Besides, GU and GG strategies produce quite similar results with each other that they look as overlapped from the plots, and the same pattern also happens to UU and UG strategies. Such feature suggests that it is not necessary to repeat grouping at Stage II if it is already applied in earlier stage, since they will produce similar results in the end. Grouping strategy in Stage I significantly reduce the $MAPE$ values, and this can be reflected from all the subfigures of Figure 4.2. The reason for this significant improvement might be that, after conducting the grouping strategy in Stage I, the associated parameters from each of the mortality models can better represent the dynamics of the mortality rates, and thus will generate a more accurate forecast result.

There is a belief that, within a certain range, the shorter the fitting year span we choose, the better forecasting performance we will achieve, since recent years' data can reflect mortality changes better and thus will be more able to capture and forecast the pattern of future mortality rates. On the other hand, the more past fitting years are included, the more faraway information they contain, and thus the prediction performance with too much faraway information may not work out very well. However, from these $MAPE_j^V$ plots, the length of the fitting year spans does not have any relationship with the forecasting performance. As a matter of fact, the forecasting performance in terms of $MAPE_j^V$ values totally depends on the dataset.

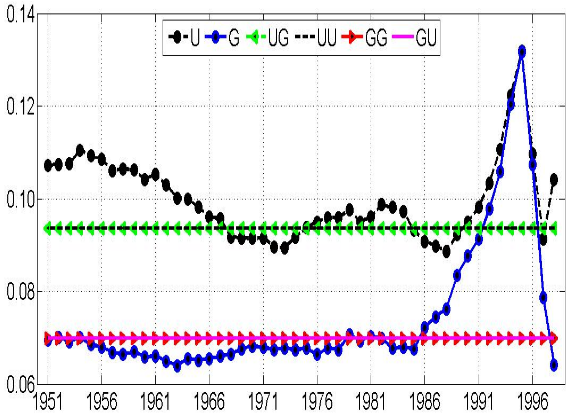
In order to show the irrelevance between the fitting year span and its corresponding forecasting performance, we compare the plots for these three countries and two genders under the LC model for $T = 10$ and $J = 48$. Figure 4.2 (a) demonstrates that under LC model, the smallest $MAPE_j^G$ for UK males occurs at the fitting year span 1990 − 2000, and after that $MAPE_j^G$ steeply increases when the fitting year span shortens until it shrinks to 1998 − 2000. This might lead to a specious conclusion with respect to the relationship between the forecasting performance and the fitting year span. However, once we change the population dataset, the pattern



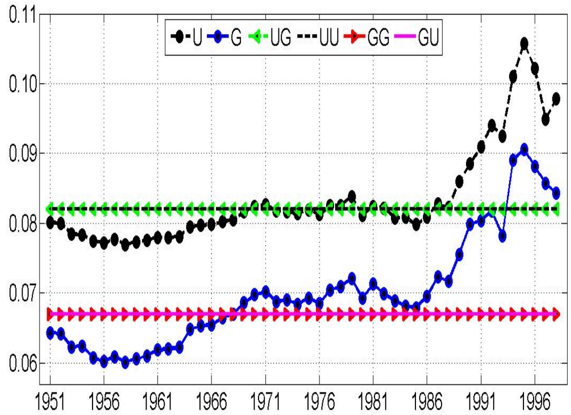
(a) US_M, LC



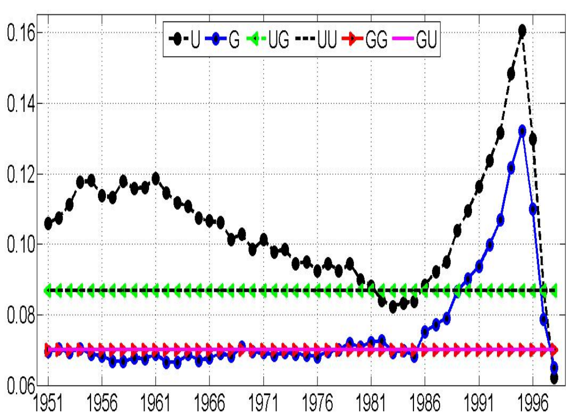
(b) US_F, LC



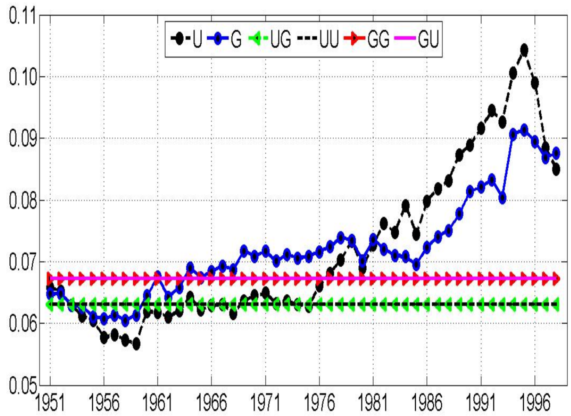
(c) US_M, CBD



(d) US_F, CBD

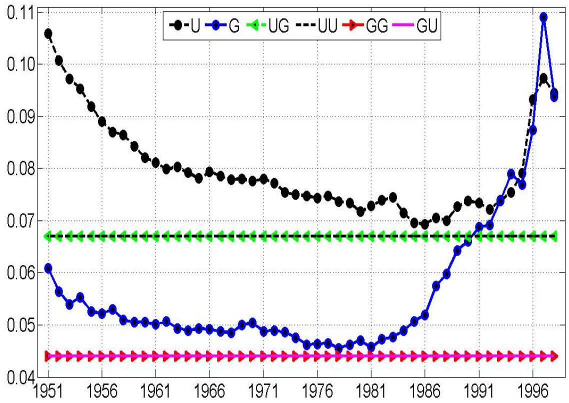


(e) $US_M, LR-RW$

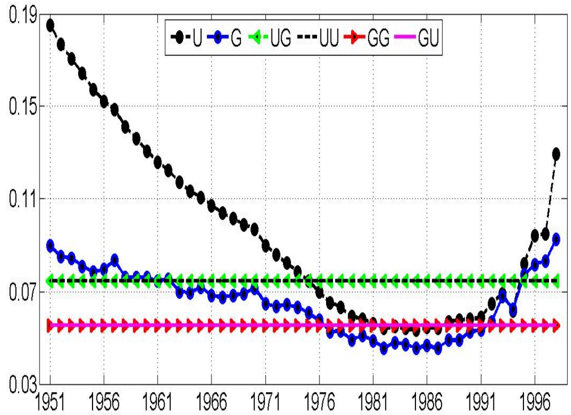


(f) $US_F, LR-RW$

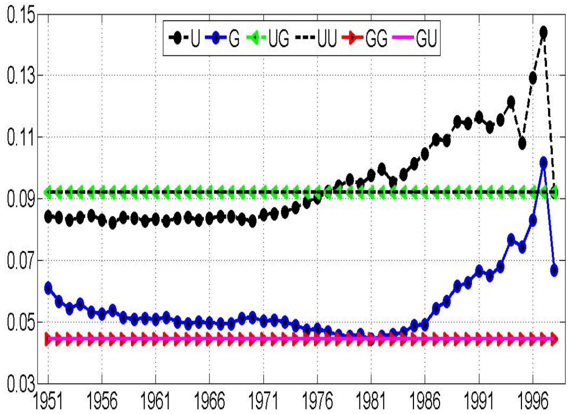
Figure 4.3: $MAPE_j$'s vs $1950+j$ for US with $T=10$ and year span $= [1950+j, 2000]$



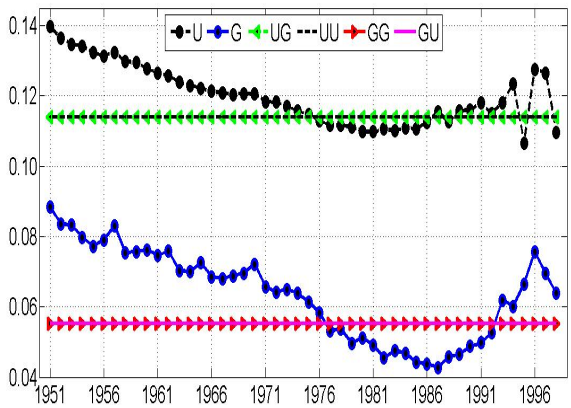
(a) $Japan_M, LC$



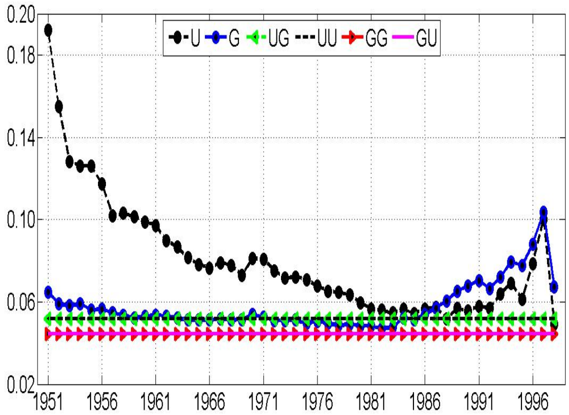
(b) $Japan_F, LC$



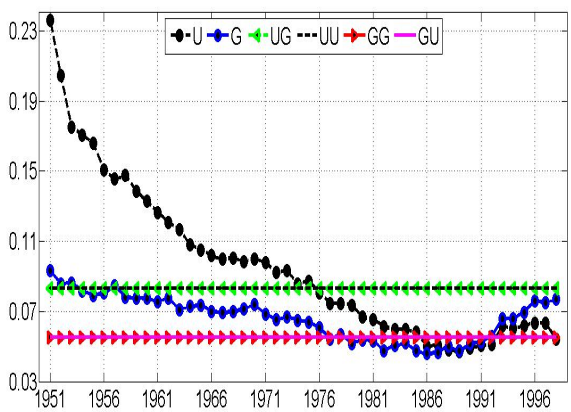
(c) $Japan_M, CBD$



(d) $Japan_F, CBD$

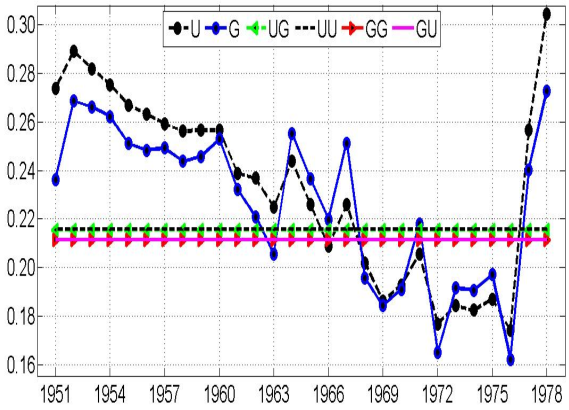


(e) $Japan_M, LR-RW$

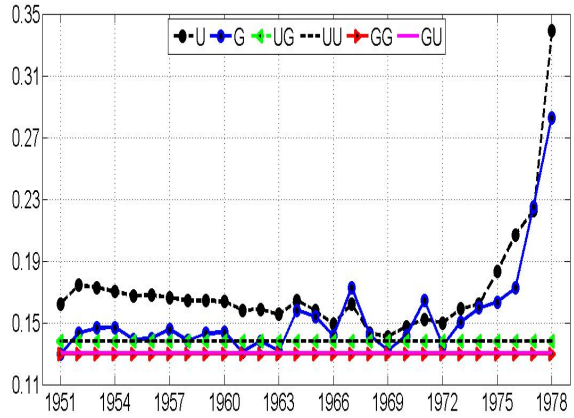


(f) $Japan_F, LR-RW$

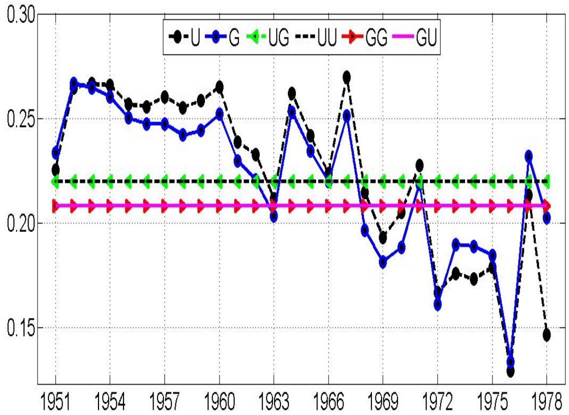
Figure 4.4: $MAPE_j$'s vs $1950 + j$ for $Japan$ with $T = 10$ and year span $= [1950 + j, 2000]$



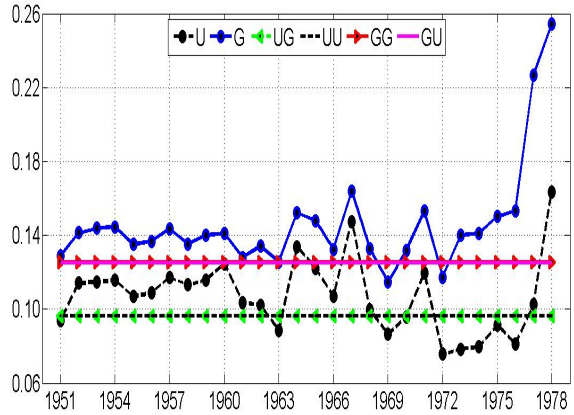
(a) UK_M, LC



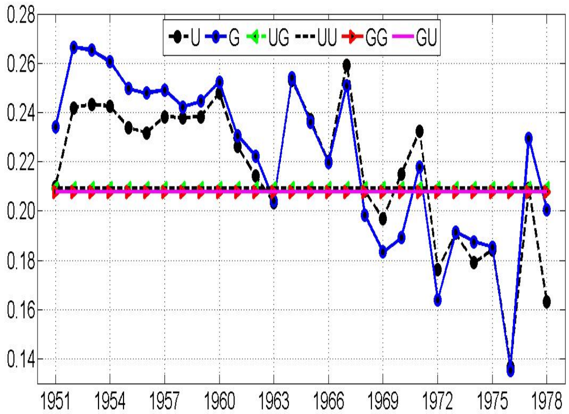
(b) UK_F, LC



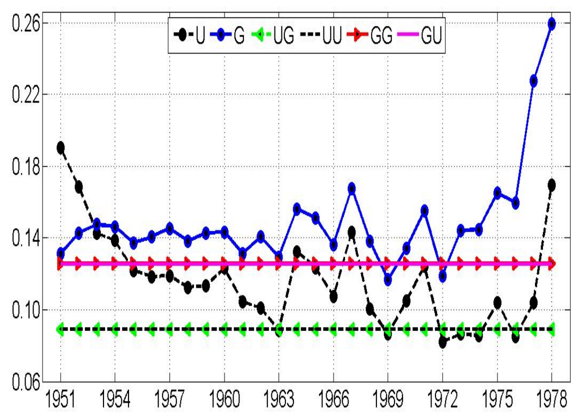
(c) UK_M, CBD



(d) UK_F, CBD

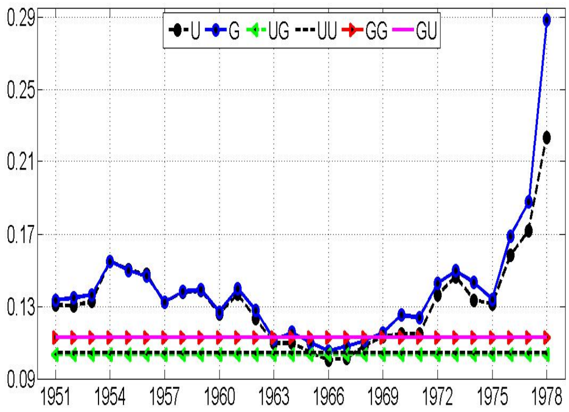


(e) $UK_M, LR-RW$

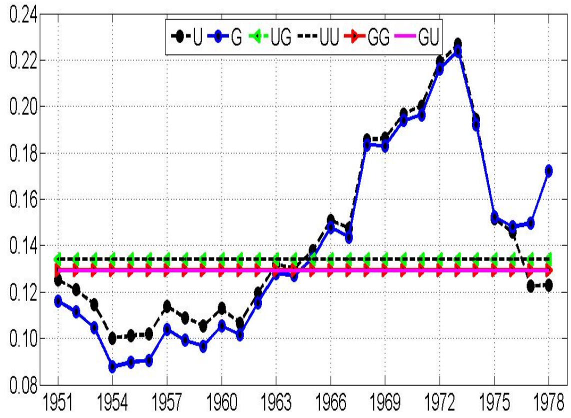


(f) $UK_F, LR-RW$

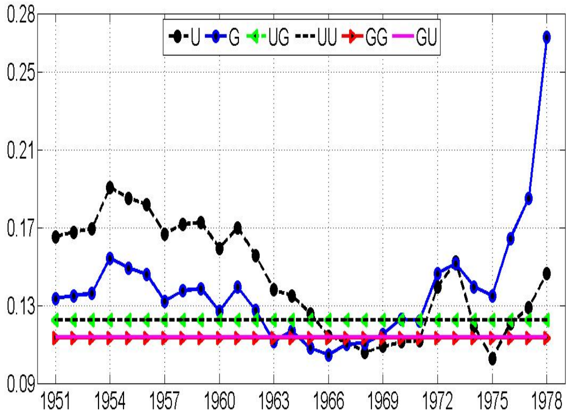
Figure 4.5: $MAPE_j$'s vs $1950+j$ for UK with $T = 30$ and year span $= [1950+j, 1980]$



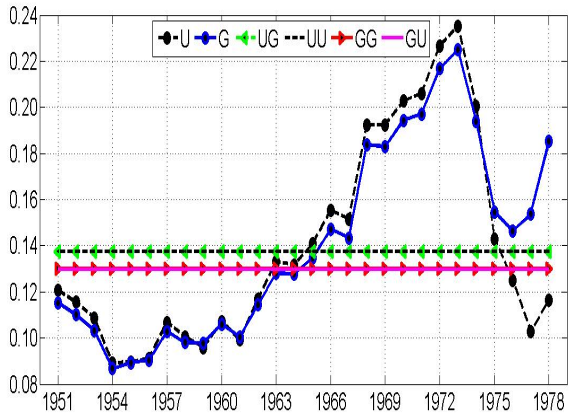
(a) US_M, LC



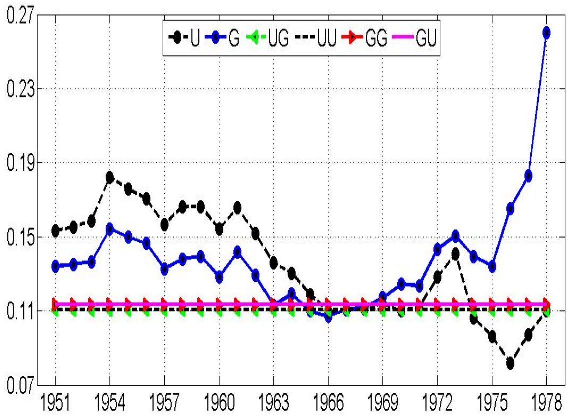
(b) US_F, LC



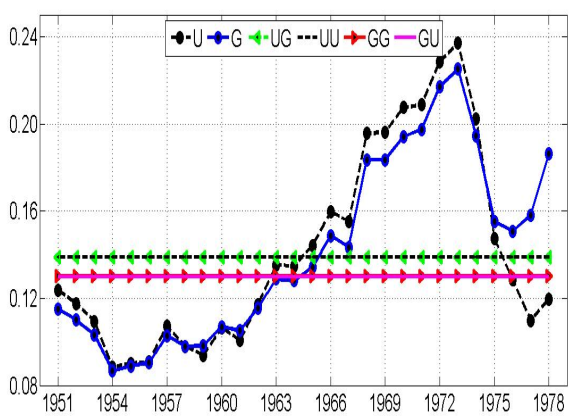
(c) US_M, CBD



(d) US_F, CBD



(e) $US_M, LR-RW$



(f) $US_F, LR-RW$

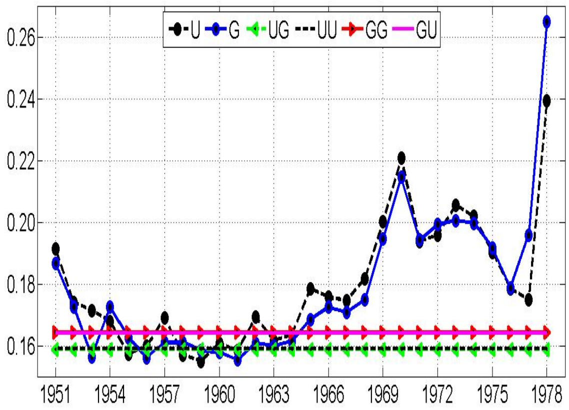
Figure 4.6: $MAPE_j$'s vs $1950 + j$ for US with $T = 30$ and year span $= [1950 + j, 1980]$

varies accordingly. For example, $MAPE_j^G$ in Figure 4.3 (a) for US males stays approximately constant until $1985-2000$, then increases to its peak at $1995-2000$, and falls after that. What is more , $MAPE_j^G$ in Figure 4.4 (a) for Japan males displays a convex pattern where the minimum $MAPE_j^G$ occurs at the fitting year span $1981-2000$.

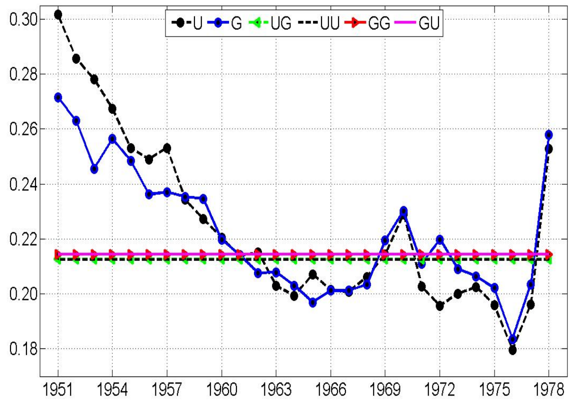
The preceding paragraph shows that there does not exist an evident relationship between the forecasting performance and the length of the fitting year span; instead, the forecasting performance completely depends on the dataset. This conclusion also applies to the CBD and LR-RW models, as well as for $T = 30$ (forecasting $1981-2010$) and $T = 50$ (forecasting $1961-2010$).

The Figures 4.5-4.7 demonstrate the $MAPE_j^V$ plots for UK, USA, and Japan under each of the three mortality models with $T = 30$ and $J = 28$. These three figures display some different patterns compared with Figures 4.2-4.4 for $T = 10$ and $J = 48$. Generally speaking, the gap between $MAPE_j^U$ and $MAPE_j^G$ significantly narrows down regardless of which specific mortality model we are applying. This shrinked gap indicates that the grouping strategy in Stage I does not help a lot to produce better forecasting outcomes in this case compared with what it does for $T = 10$. Even more, there are several cases where $MAPE_j^G$ is larger than $MAPE_j^U$, implying that the grouping strategy in Stage I even produces worse forecasting results. An extreme example can be seen from Figure 4.5 (d) for UK females, where the $MAPE_j^U$ is clearly smaller than $MAPE_j^G$.

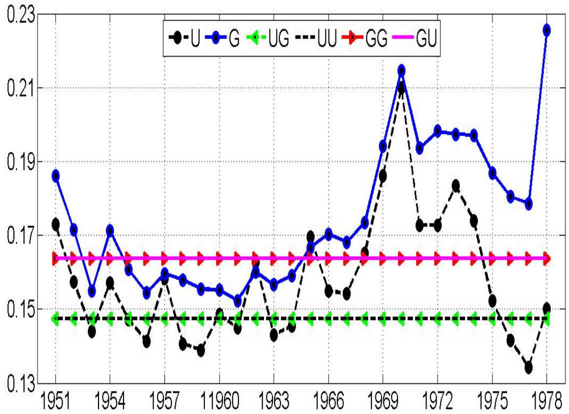
By comparing the forecasting performance across different mortality models, countries and genders, we can see that the GU and GG strategies generally produces the best forecasting performance for $T = 10$. This indicates that the grouping strategy in Stage I turns out to be the most effective way to reduce forecasting errors. However, for $T = 30$, the GU and GG strategies do not necessarily produce the best results. For example, Figures 4.5 (d) and (f) show larger $MAPE$ values generated from GU and GG strategies, compared with those from UU and UG strategies. In these two cases, applying UU and UG strategies will produce a better outcome: less estimation effort with quite a low $MAPE$ value. For $T = 30$ and $J = 28$, the $MAPE_j^V$ $(V = U, G)$ plots under the CBD model reflects quite fluctuating pattens which turn out quite opposite as what they do for $T = 10$ and $J = 48$. For example, by comparing Figures 4.2 (c) and 4.5 (c) for UK males, we can see that $MAPE_j^V$ under the CBD model in Figure 4.5 (c) produces larger errors with more irregularities. Although the performance depends on countries and genders to a large extent and sometimes the pseudo non-parametric Bühlmann estimation cannot guarantee a lowest $MAPE$ for all j's, yet it yields better performances on average. More importantly, it smooths such variations and produces a stable result, which is definitely a favorable feature for life insurance companies.



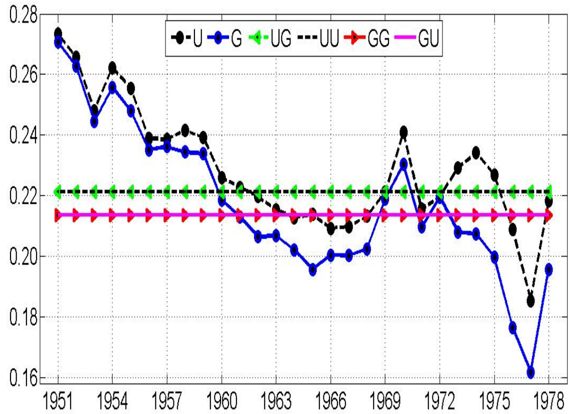
(a) $Japan_M, LC$



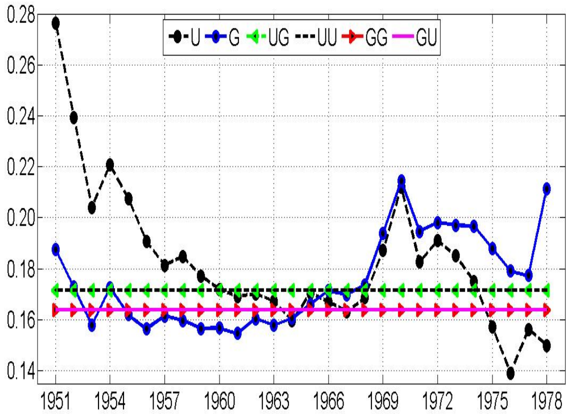
(b) $Japan_F, LC$



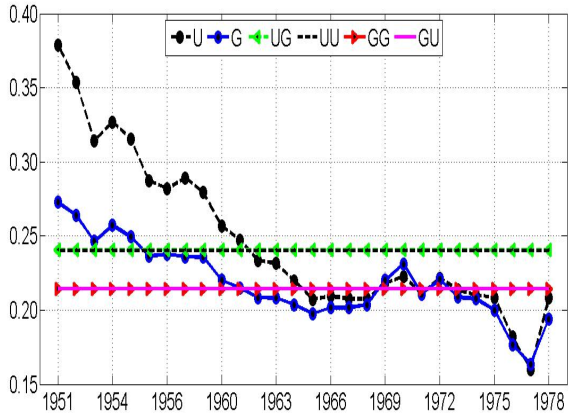
(c) $Japan_M, CBD$



(d) $Japan_F, CBD$

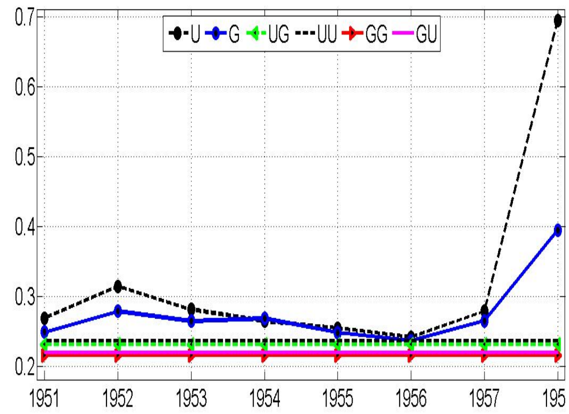


(e) $Japan_M, LR-RW$

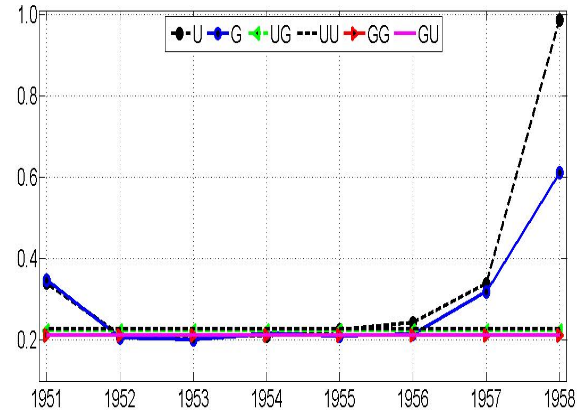


(f) $Japan_F, LR-RW$

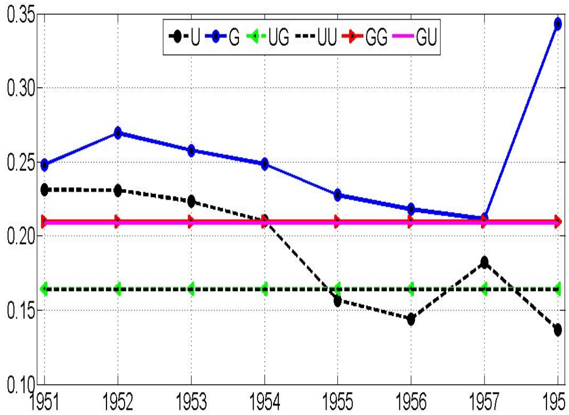
Figure 4.7: $MAPE_j$'s vs $1950 + j$ for *Japan* with $T = 30$ and year span $= [1950 + j, 1980]$



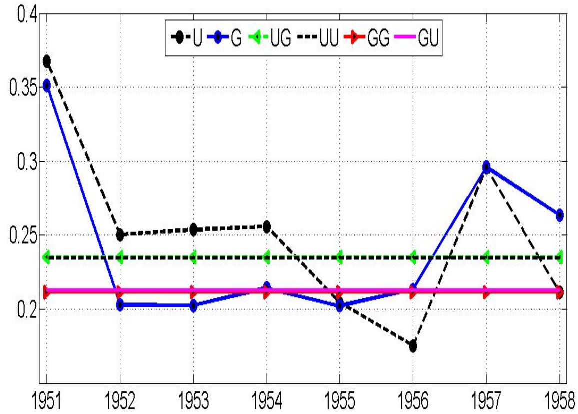
(a) UK_M, LC



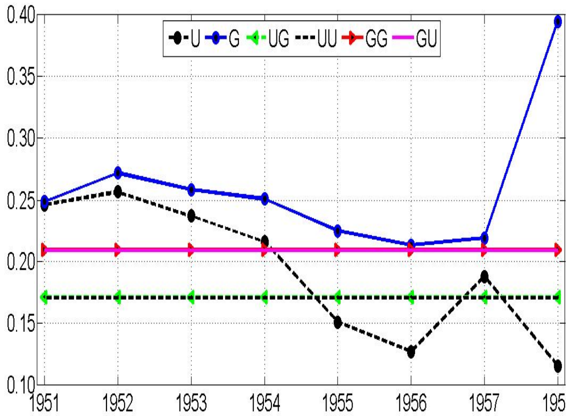
(b) UK_F, LC



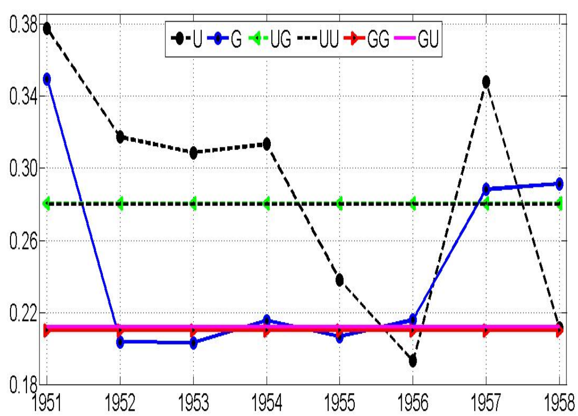
(c) UK_M, CBD



(d) UK_F, CBD

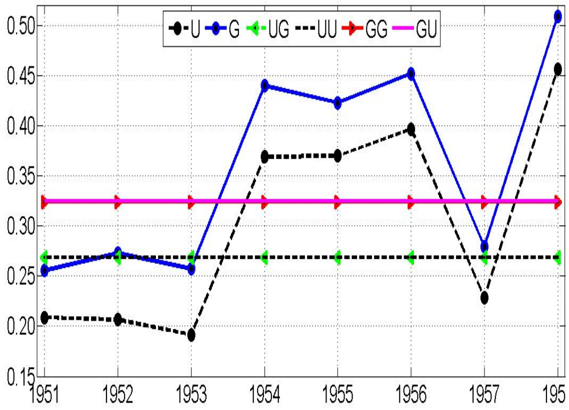


(e) $UK_M, LR-RW$

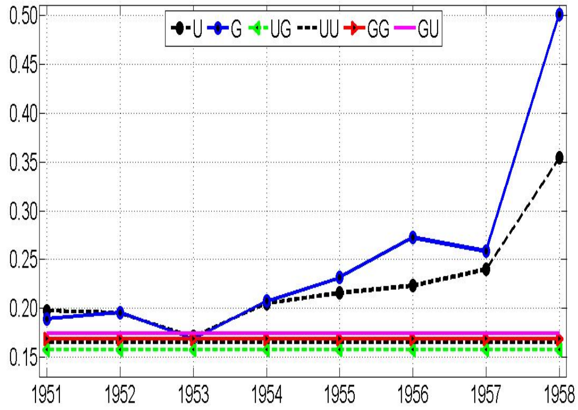


(f) $UK_F, LR-RW$

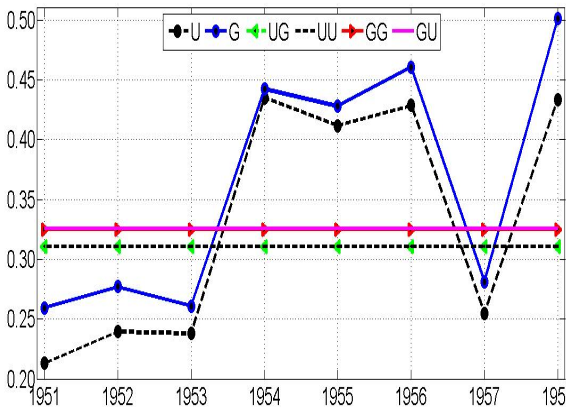
Figure 4.8: $MAPE_j$'s vs $1950+j$ for UK with $T=50$ and year span $=[1950+j, 1960]$



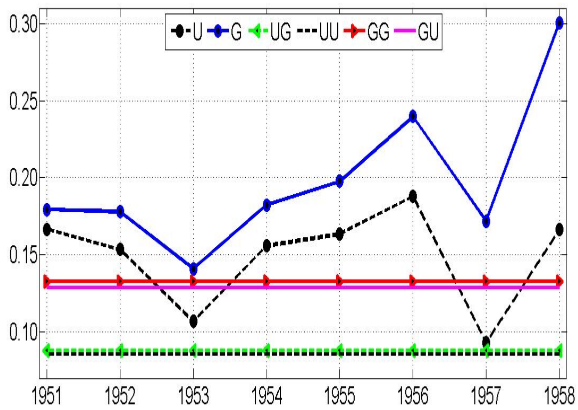
(a) US_M, LC



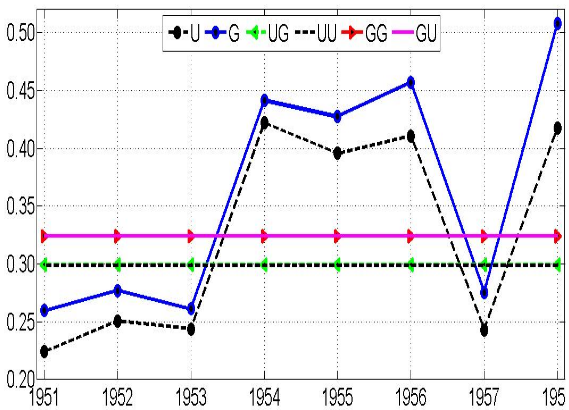
(b) US_F, LC



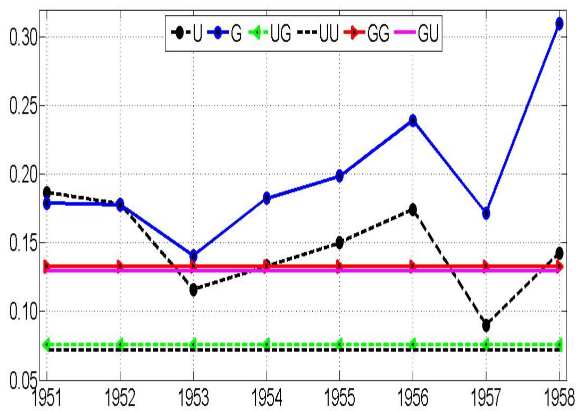
(c) US_M, CBD



(d) US_F, CBD

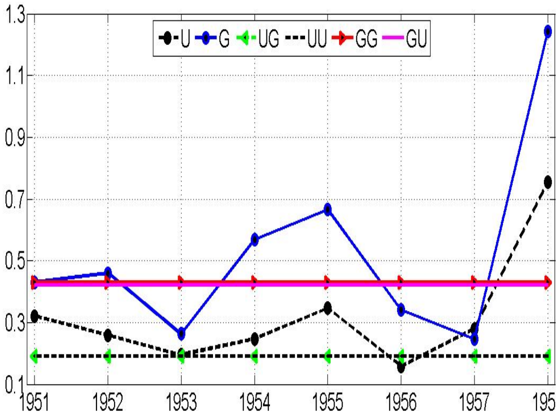


(e) $US_M, LR-RW$

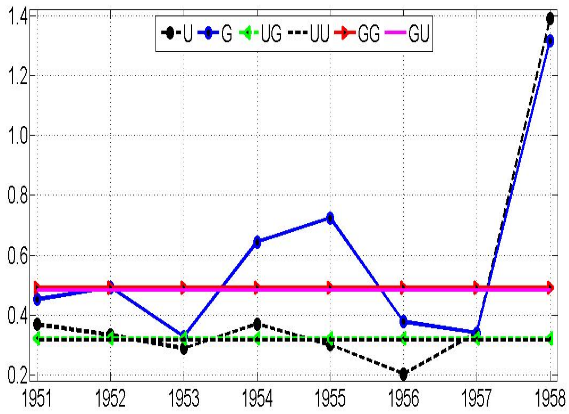


(f) $US_F, LR-RW$

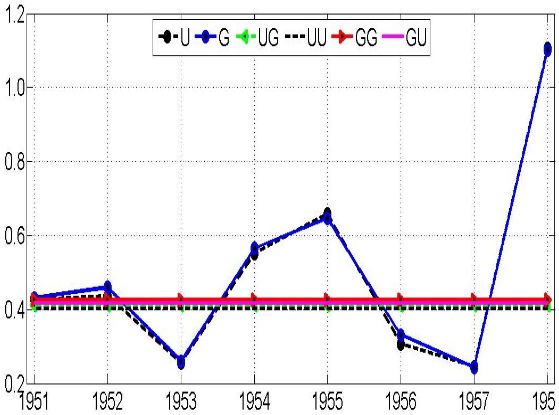
Figure 4.9: $MAPE_j$'s vs $1950+j$ for US with $T = 50$ and year span $= [1950+j, 1960]$



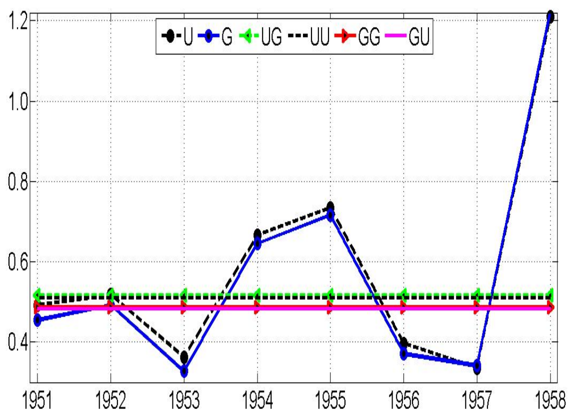
(a) $Japan_M, LC$



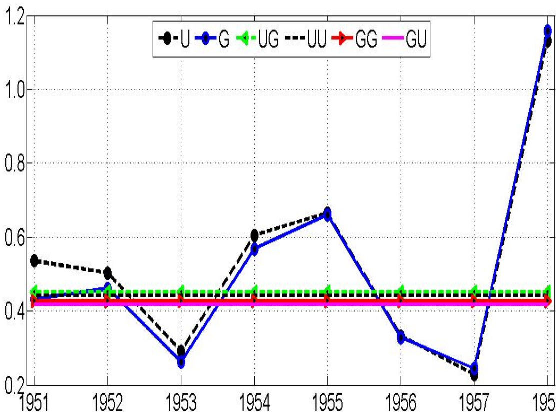
(b) $Japan_F, LC$



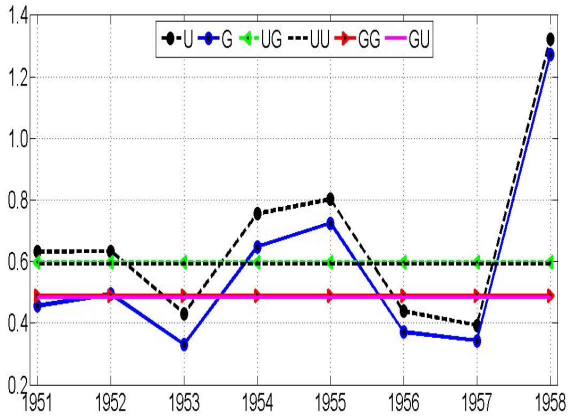
(c) $Japan_M, CBD$



(d) $Japan_F, CBD$



(e) $Japan_M, LR-RW$



(f) $Japan_F, LR-RW$

Figure 4.10: $MAPE_j$'s vs $1950+j$ for $Japan$ with $T=50$ and year span $=[1950+j, 1960]$

4.2 Numerical Results

The $MAPE^{VW}$ values and corresponding $MAPE^{VW}$ reduction ratios for $T = 10, 30$ and 50 with the smallest $MAPE^{VW}$ and the biggest $MAPE^{VW}$ reduction ratio being highlighted for each population under each mortality model are summarized in Tables 4.3-4.8, where $MAPE^{VW}$ is given in (4.3) and $(V, W) = (U, U)$, (U, G), (G, U) and (G, G). $MAPE^{VW}$ reduction ratio indicates the extent of relative improvements in $MAPE$ using the VW strategy in Stage II with respect to the average of $MAPE_j^V$'s over J fitting year spans using the V strategy in Stage I. Specifically, the $MAPE^{VW}$ reduction ratio for the strategy VW is $1 - \frac{MAPE^{VW}}{\frac{1}{J}\sum_{j=1}^{J} MAPE_j^V}$ where $MAPE_j^V$ is given in (4.2).

Table 4.3: $MAPE^{VW}$ values for 10-year forecasting

10 years	MAPE	UK_M	UK_F	USA_M	USA_F	Japan_M	Japan_F
LC	UU	0.096405	0.074328	0.073048	0.068887	0.067027	0.074448
	UG	0.096550	0.074400	0.072919	0.068912	0.067048	0.074491
	GU	**0.074408**	0.053749	**0.070137**	**0.067185**	0.044048	**0.055165**
	GG	0.074416	**0.053746**	0.070167	0.067208	**0.044014**	0.055166
CBD	UU	0.150090	0.093302	0.093585	0.082051	0.092035	0.113843
	UG	0.150100	0.093301	0.093573	0.082057	0.092023	0.113850
	GU	**0.078441**	**0.055469**	0.069938	**0.066997**	**0.044270**	**0.055121**
	GG	0.078562	0.055485	**0.069928**	0.067014	0.044271	0.055169
LR-RW	UU	0.132135	0.058297	0.086863	0.063076	0.051654	0.082970
	UG	0.132255	0.058301	0.086865	0.063098	0.051841	0.083018
	GU	**0.078881**	**0.056687**	0.070037	**0.067256**	**0.044327**	**0.054938**
	GG	0.079080	0.056822	**0.070029**	0.067273	0.044451	0.055006

Table 4.4: $MAPE^{VW}$ reduction ratios for 10-year forecasting

10 years	method	UK_M	UK_F	USA_M	USA_F	Japan_M	Japan_F	Average
LC	UU	**14.87%**	**19.92%**	21.67%	**11.04%**	16.12%	**22.22%**	**17.64%**
	UG	14.74%	19.84%	**21.80%**	11.01%	16.10%	22.17%	17.61%
	GU	10.06%	13.99%	6.83%	5.71%	21.71%	15.85%	12.36%
	GG	10.05%	14.00%	6.79%	5.68%	**21.77%**	15.85%	12.36%
CBD	UU	0.92%	1.58%	6.25%	1.68%	2.94%	5.19%	3.09%
	UG	0.92%	1.58%	**6.27%**	1.67%	2.96%	5.18%	3.09%
	GU	**6.15%**	**10.04%**	5.55%	**4.31%**	**19.39%**	**13.49%**	**9.82%**
	GG	6.00%	10.01%	5.56%	4.29%	19.39%	13.41%	9.78%
LR-RW	UU	6.07%	**29.32%**	**17.20%**	**12.30%**	**35.64%**	13.20%	**18.96%**
	UG	5.99%	29.32%	17.20%	12.27%	35.41%	13.15%	18.89%
	GU	**8.19%**	13.48%	7.43%	6.48%	23.01%	**17.20%**	12.63%
	GG	7.95%	13.27%	7.44%	6.46%	22.80%	17.10%	12.50%

Comparing Tables 4.3 and 4.4, we observe that while the lowest $MAPE$ values always occur at either GU or GG, these two strategies do not necessarily produce the highest reduction ratios. Furthermore, the difference between the lowest and the highest $MAPE$ is not significantly huge, so the insurers need to consider computational effort to conduct such grouping strategy in practice. If the grouping strategy turns out to be time-/effort-consuming, they can only apply the non-parametric Bühlmann estimation without grouping the age span and still get a respectable improvement, reflected in the $MAPE$ reduction ratio. Also, the results vary largely with the selection of countries and genders, as we see that mortality models exposed to the same grouping strategy and the same credibility method can still produce very distinct responses in terms of their $MAPE$ values and $MAPE$ reduction ratios.

Table 4.5: $MAPE^{VW}$ values for 30-year forecasting

30 years	MAPE	UK_M	UK_F	USA_M	USA_F	Japan_M	Japan_F
LC	UU	0.215638	0.138213	0.104037	0.134060	0.158833	**0.212429**
	UG	0.215469	0.137659	**0.103427**	0.134062	**0.158774**	0.212627
	GU	0.211492	0.130453	0.112650	**0.129232**	0.163904	0.214291
	GG	**0.211306**	**0.129948**	0.112626	0.129312	0.164146	0.214460
CBD	UU	0.219551	**0.096419**	0.122542	0.137485	**0.147286**	0.221189
	UG	0.219572	0.096436	0.122446	0.137486	0.147302	0.221277
	GU	0.208105	0.124942	0.113829	**0.129738**	0.163635	**0.213599**
	GG	**0.207968**	0.125319	**0.113416**	0.130026	0.163773	0.213632
LR-RW	UU	0.208993	0.088878	0.110342	0.138836	0.171562	0.240170
	UG	0.208992	**0.088843**	**0.110300**	0.138819	0.171639	0.240380
	GU	0.207886	0.125182	0.113580	**0.129790**	**0.163816**	**0.214336**
	GG	**0.207828**	0.125818	0.113176	0.130186	0.163940	0.214382

Table 4.6: $MAPE^{VW}$ reduction ratios for 30-year forecasting

30 years	method	UK_M	UK_F	USA_M	USA_F	Japan_M	Japan_F	Average
LC	UU	7.64%	19.13%	21.63%	5.70%	11.52%	**5.37%**	11.83%
	UG	7.71%	**19.45%**	**22.09%**	5.69%	**11.56%**	5.29%	**11.96%**
	GU	6.82%	15.24%	19.12%	**7.57%**	8.25%	3.63%	10.11%
	GG	**6.90%**	15.57%	19.14%	7.51%	8.11%	3.56%	10.13%
CBD	UU	2.03%	10.10%	14.80%	1.21%	**6.77%**	**3.29%**	6.37%
	UG	2.02%	10.09%	14.86%	1.21%	6.76%	3.25%	6.37%
	GU	5.95%	**14.45%**	17.58%	**7.54%**	6.49%	1.85%	**8.98%**
	GG	**6.01%**	14.20%	**17.88%**	7.33%	6.41%	1.84%	8.94%
LR-RW	UU	3.59%	24.15%	18.00%	1.65%	6.25%	**2.58%**	9.37%
	UG	3.59%	**24.18%**	18.03%	1.66%	6.21%	2.49%	9.36%
	GU	6.29%	16.36%	17.97%	**7.99%**	**6.54%**	2.07%	**9.54%**
	GG	**6.31%**	15.93%	**18.26%**	7.71%	6.47%	2.05%	9.46%

The scenario for $T = 30$ show more variations and irregularities in terms of the distribution of smallest $MAPE$'s. In this case, UG and UU strategies take more important roles compared with what they do for $T = 10$, and this is also consistent with what we observe from Figures 4.5 (d) and (f), where the lowest $MAPE$ paths are generated from UU and UG strategies. For application in reality, it is a wise decision to not to conduct grouping since insurers could save time and cost and still achieve favorable forecasting improvements by applying this pseudo non-parametric Bühlmann estimation approach; for example, the reduction ratios for the UU strategy can reach 10%-24% for UK females and USA males under all three mortality models and for Japan males under the LC model.

Table 4.7: $MAPE^{VW}$ values for 50-year forecasting

50 years	MAPE	UK_M	UK_F	USA_M	USA_F	Japan_M	Japan_F
LC	UU	0.237005	0.227854	0.268602	0.164896	**0.189809**	**0.314934**
	UG	0.231657	0.223325	**0.268572**	**0.157344**	0.190882	0.320814
	GU	0.219842	0.212082	0.324142	0.174500	0.419674	0.482245
	GG	**0.216052**	**0.211558**	0.323244	0.168564	0.429905	0.492173
CBD	UU	**0.163630**	0.234207	0.310390	**0.085199**	**0.401134**	0.509598
	UG	0.164286	0.234862	**0.310236**	0.087596	0.410283	0.516667
	GU	0.208432	0.212875	0.325244	0.128633	0.415556	**0.482082**
	GG	0.209817	**0.211565**	0.324576	0.132380	0.426439	0.487139
LR-RW	UU	0.170722	0.280204	**0.298340**	**0.071553**	0.442342	0.590521
	UG	0.171254	0.280626	0.299485	0.075723	0.450890	0.596104
	GU	**0.209277**	0.212026	0.323853	0.129594	**0.417582**	**0.483266**
	GG	0.209829	**0.210002**	0.324165	0.133010	0.427570	0.487836

Table 4.8: $MAPE^{VW}$ reduction ratios for 50-year forecasting

50 years	method	UK_M	UK_F	USA_M	USA_F	Japan_M	Japan_F	Average
LC	UU	27.12%	34.04%	11.42%	26.74%	**40.65%**	**29.90%**	28.31%
	UG	**28.76%**	**35.35%**	**11.43%**	30.10%	40.32%	28.59%	**29.09%**
	GU	20.35%	27.06%	10.19%	31.04%	20.31%	17.51%	21.08%
	GG	21.72%	27.24%	10.44%	**33.39%**	18.36%	15.81%	21.16%
CBD	UU	13.63%	6.95%	6.47%	**42.85%**	**19.45%**	13.52%	17.14%
	UG	13.28%	6.69%	6.51%	41.24%	17.61%	12.32%	16.28%
	GU	**17.62%**	12.49%	10.58%	35.26%	17.63%	**15.60%**	**18.20%**
	GG	17.08%	**13.03%**	**10.77%**	33.38%	15.48%	14.72%	17.41%
LR-RW	UU	11.24%	2.87%	8.52%	**51.12%**	17.57%	12.62%	17.32%
	UG	10.96%	2.72%	8.17%	48.27%	15.98%	11.79%	16.32%
	GU	**19.63%**	14.12%	**10.88%**	35.19%	**18.82%**	**16.65%**	**19.21%**
	GG	19.42%	**14.94%**	10.79%	33.48%	16.88%	15.86%	18.56%

From Table 4.8 we can see that the MAPE reduction ratios for $T = 50$ are generally higher than those for $T = 10$ and 30, which might lead to a conjecture that the longer the forecasting year, the better forecasting performance we can achieve by applying the pseudo non-parametric Bühlmann estimation approach. However, the performance varies with the population dataset. Those variations reflected in the tables above also confirm our belief that there does not exist a universally best model or strategy which will always generate the most satisfactory forecasting performances for all population datasets. Actually, the forecasting performance depends on the underlying mortality data to a great extent.

Besides MAPE, the mean absolute error $MAE = \frac{1}{m \times T} \sum_{x=x_0}^{x_0+m-1} \sum_{t=t_0+1}^{t_0+T} |\hat{q}_{x,t} - q_{x,t}|$ and the root mean squared error $RMSE = \left[\frac{1}{m \times T} \sum_{x=x_0}^{x_0+m-1} \sum_{t=t_0+1}^{t_0+T} \left(\hat{q}_{x,t} - q_{x,t}\right)^2\right]^{1/2}$ are another two commonly used indexes to measure mortality forecasting accuracy. However, these two indexes are scale-dependent and cannot be used to make comparisons among different age levels. In credibility theory, the mean squared prediction error $MSPE = E\left[\left(\hat{q}_{x,t} - q_{x,t}\right)^2\right]$ is one of the most commonly applied index and it refers to the same meaning with $RMSE$ in our mortality modeling. In this project we are interested in comparing the $\hat{q}_{x,t}$ instead of the original $ln(\hat{m}_{x,t})$ or $logit(\hat{q}_{x,t})$; furthermore, in the process of transferring the $ln(\hat{m}_{x,t})$ or $logit(\hat{q}_{x,t})$ back to the corresponding $\hat{q}_{x,t}$, to minimize $ln(\hat{m}_{x,t})$ or $logit(\hat{q}_{x,t})$ does not necessarily imply minimizing the corresponding $\hat{q}_{x,t}$. In this sense, we decide to apply MAPE since it is expressed in generic percentage term. It measures the percentage error which is age-independent, and this relative error measurement allows us to compare the forecasting performance across different age levels.

Chapter 5

Applications

Capturing the dynamics of mortality rates has critical meanings for both life insurers and reinsurers due to its important financial implications. Mortality is one of the core factors to take care of for life insurance companies to price a product, for pension consulting firms to conduct valuations and adjust their pension annuity amounts, for government to manage social security programs, and for the financial investment institutions to issue or analyze mortality-linked securities. However, it is not an easy task to provide a good forecast, and it turns out even more challenging nowadays due to the fact the population of developed countries has improved rapidly over the last decades. Mortality rates are affected by many factors, and it is challenging for the life insurance companies to accurately capture and estimate future mortality rates.

From the outcomes in Chapter 4, we already confirm that applying the non-parametric Bühlmann credibility theory to a mortality model can help companies improve their forecasting performances to a significant extent. This chapter takes one step further and applies those mortality estimates to price life insurance and annuity products. Although there are numerous life insurance and annuity products in the market, we will mainly focus on term life insurance, endowment and annuity.

In this chapter we price the three T-year products for age x_0 in year t using the cohort mortality sequence $q_{x_0,t}^{T,c} = \{q_{x_0,t}, q_{x_0+1,t+1}, \cdots, q_{x_0+T-1,t+T-1}\}$ where "c" represents cohort effect. That is, we take the diagonal entries of the matrix of forecasted mortality rates for T years generated from the LC, CBD and LR-RW models, respectively, in Stage I (the results are denoted by U and G), and also do the same thing for the non-parametric Bühlmann estimates in Stage II (the results are denoted by UU, UG, GU and GG).

Assuming that the pricing interest rate $i = 2\%$, we calculate the net single premium (NSP) of each of the 3 life insurance and annuity products. More specifically, the 3 NSPs and their explicit formulas are

(1) NSP_x of T-year fully discrete term life insurance, $\sum_{k=0}^{T-1} {}_kp_{x,t} \times q_{x+k,t+k} \times v^{k+1}$;
(2) NSP_x of T-year fully discrete endowment, $\sum_{k=0}^{T-1} {}_kp_{x,t} \times q_{x+k,t+k} \times v^{k+1} + {}_T p_{x,t} \times v^T$;
(3) NSP_x of T-year temporary life annuity immediate, $\sum_{k=1}^{T} {}_kp_{x,t} \times v^k$;
where $v = \dfrac{1}{1+i}$ and ${}_kp_{x,t} = p_{x,t} \times \cdots \times p_{x+k-1,t+k-1}$.

The computations are conducted for each of the LC, CBD and LR-RW models with
(1) $T = 10$, $t = 2001$, $x = 25, ..., 75$;
(2) $T = 30$, $t = 1981$, $x = 25, ..., 55$; and
(3) $T = 50$, $t = 1961$, $x = 25, ..., 35$.

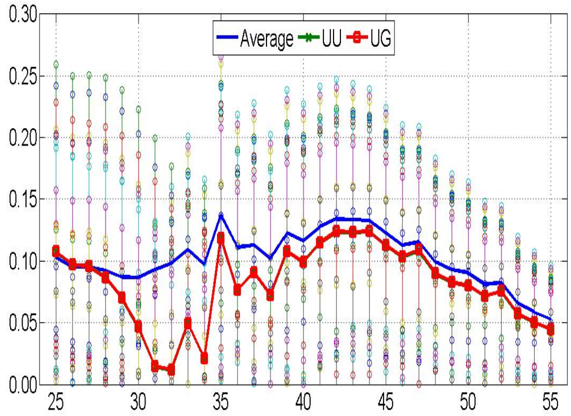
Chapter 4 has shown that, with the pseudo non-parametric Bühlmann estimation being applied, the forecasted mortality rates turn out to be more accurate and stable; we use the MAPE index to quantitatively support our conclusion. In this chapter we adopt the same index to compare and analyze the performances resulting from applying the non-parametric Bühlmann approach.

In this case, $MAPE_x^j = \left|\frac{N\hat{S}P_x^j - \text{true } NSP_x}{\text{true } NSP_x}\right|$, where $N\hat{S}P_x^j$ is the NSP for age x based on the forecasted mortality rates from the j^{th} fitting year span. For instance, if $T = 30$ and $J = 28$, then there are 28 fitting year spans. From the formula it can be seen that every time we switch to a new fitting year span, we get a corresponding $N\hat{S}P_x^j$ and thus an $MAPE_x^j$ value. As a consequence, for each age x, there are 28 $MAPE_x^j$ values.

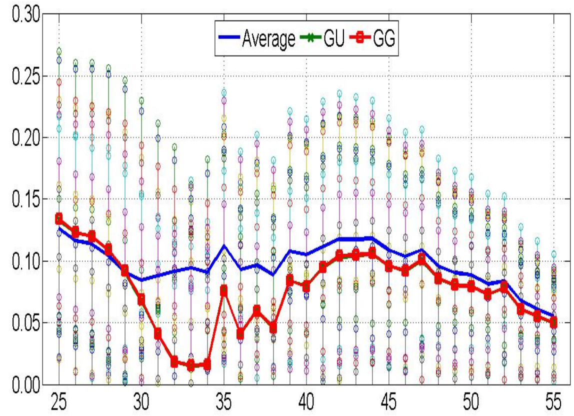
Figures 5.1-5.3 illustrate $MAPE_x^j$ against x under the LC, CBD and LR-RW models using mortality data of US males. For illustration purpose, we only present the scenario for $T = 30$ and $J = 28$ in this project. As mentioned above, there are 28 $MAPE_x^j$ values for each x, and we also calculate $MAPE_x^{Avg}$, the average of these 28 values, which is denoted as "Average" in the plots. The subscripts "U" and "G" following each of the LC, CBD and LR-RW models help to distinguish the two results generated from the ungrouped and grouped age spans in Stage I. We then apply the mortality rates produced from UU, UG, GU and GG strategies, respectively, in Stage II to compute corresponding NSP_x^{VW}'s and $MAPE_x^{VW}$'s where $MAPE_x^{VW} = \left|\frac{N\hat{S}P_x^{VW} - \text{true } NSP}{\text{true } NSP}\right|$ and $(V, W) = (U, U)$, (U, G), (G, U) and (G, G).

From all the subfigures, the difference between $MAPE_x^{UU}$ and $MAPE_x^{UG}$, as well as the difference between $MAPE_x^{GU}$ and $MAPE_x^{GG}$, is almost invisible. The two curves reflect such similar pattens that they almost overlap with each other in the subfigures.

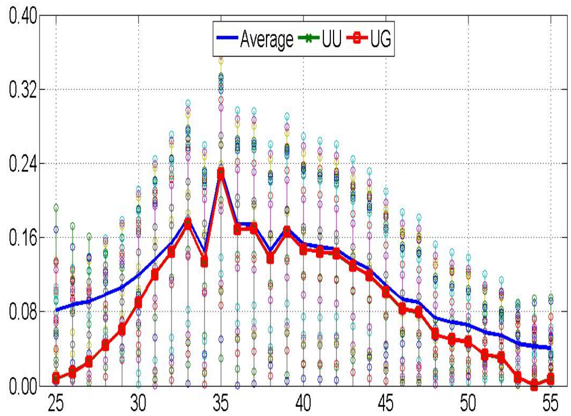
It is noticeable that, regardless of U case on the left side or G case on the right, the $MAPE_x^W$ path using the pseudo non-parametric Bühlmann credibility approach is almost lower than $MAPE_x^{Avg}$, the average of 28 $MAPE_x^j$'s, for each age x. Lower $MAPE_x$ value implies higher accuracy, that is, the estimate is close to the true value. However, we notice that the gap between



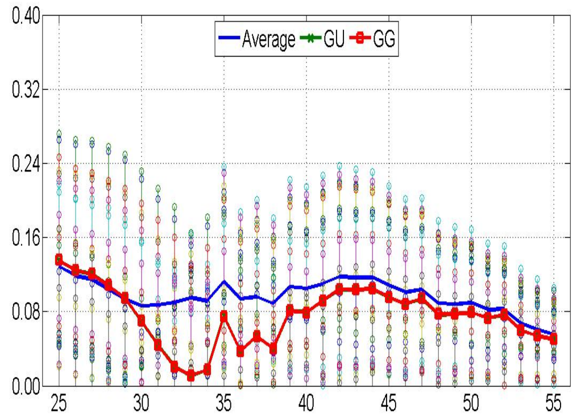
(a) LC_U



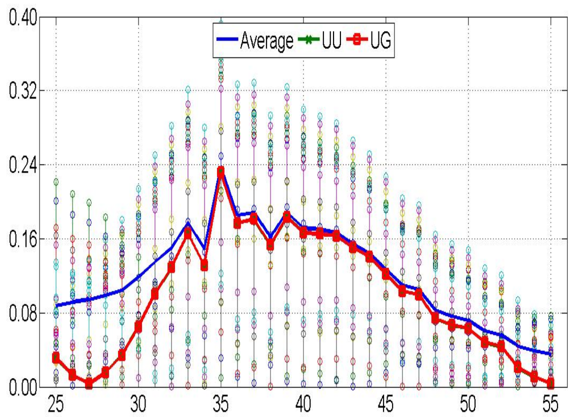
(b) LC_G



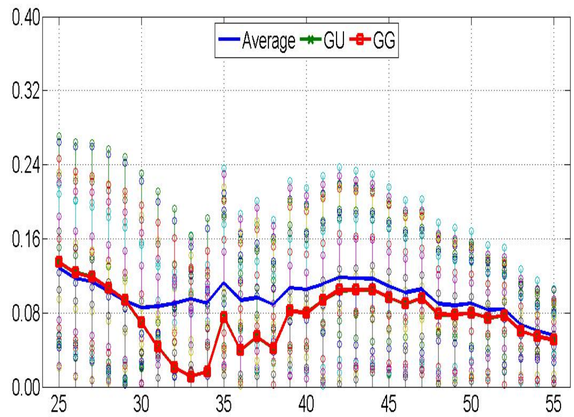
(c) CBD_U



(d) CBD_G



(e) $LR - RW_U$



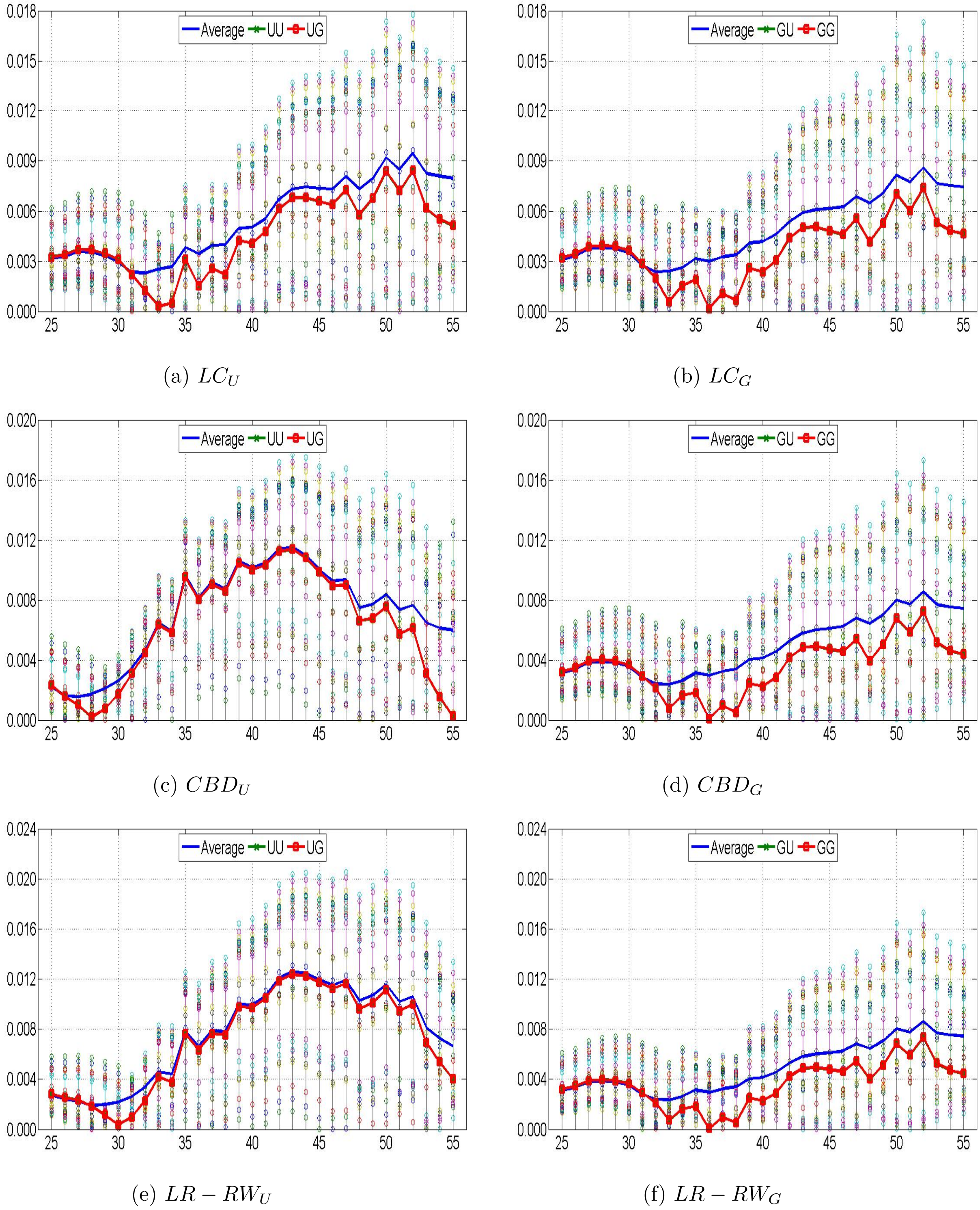
(f) $LR - RW_G$

Figure 5.1: $MAPE_x$ for 30-year term life insurance vs age x for US males

$MAPE_x^W$ and the average $MAPE_x^{Avg}$ is not as significant as what has been shown in mortality prediction. For example, when comparing Figure 4.6 with 5.1 for US males with $T = 30$, we observe that the pseudo non-parametric Bühlmann approach obviously reduces the $MAPE$ values for all 3 mortality models as shown in Figures 4.6 (a), (c) and (e), while the difference between the $MAPE_x^{GG}$ curve and the $MAPE_x^{Avg}$ curve is not big for most ages as displayed in Figures 5.1 (a), (b) and (c). This disparity can be explained by the different computation methods of MAPE. In Figure 4.6, the MAPE index is $MAPE_j^V (V = U, G)$ (see (4.2)) measuring accuracy of forcasted mortality rates based on the j^{th} fitting year span, which is averaged MAPE over all ages $x = x_0, \cdots, x_0 + m - 1$ and all forecasting years $t = t_0 + 1, \cdots, t_0 + T$, while in Figure 5.1, the MAPE refers to $MAPE_x$ which measures accuracy of forecasted NSP of a life insurance product issued to x; the $p_{x+i-1,t+i-1}$'s in ${}_kp_{x,t}$ under-estimating and over-estimating the true values can offset each other. Besides the different computation method, another possible explanation is the cohort effect. As mentioned earlier, we apply cohort mortality rate sequence in this pricing scenario. As a consequence, when we take the mortality rates from the diagonal entries of a forecasted mortality rate matrix, the $MAPE_j^V$ index does not serve as a helpful reference.

Comparing the G, GU and GG strategies, we can also find that the grouping strategy adopted in Stage I makes the 3 plots generated from three different mortality models look alike. For example, Figures 5.1 (b), (d) and (f) demonstrate quite similar patterns though the numerical values involved in each subfigure are not exactly the same. This similarity is caused by the grouping strategy which, on the other hand, also helps to eliminate or reduce the model risk for pricing since the three different mortality models can produce very close NSPs of the three products if the grouping strategy is adopted in Stage I.

From Figure 5.1 for 30-year term life insurance, it can be observed that $MAPE_x^j$'s are smaller and denser at higher ages for both U and G strategies. All the 6 subfigures show that the $MAPE_x$ paths have a downtrend beyond $x = 42$. However, opposite patterns are observed for the LC_G, CBD_G, $LR - RW_G$ and LC_U for endowment and annuity products as shown in Figures 5.2 and 5.3. Except this inconsistency, all other patterns remain the same: the grouping strategy in Stage I makes the 3 plots from three different mortality models look similar, and also contributes to generating a bigger gap between $MAPE_x^{VW}$ and $MAPE_x^{Avg}$, which implies a favorable pricing performance for this case. A more accurate forecast in mortality rates supports our conclusion before, and such a satisfactory prediction in premiums can also return potentially huge financial benefits to life insurance companies.

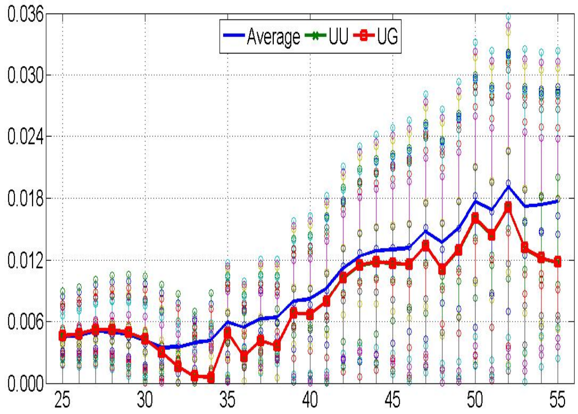


(a) LC_U (b) LC_G

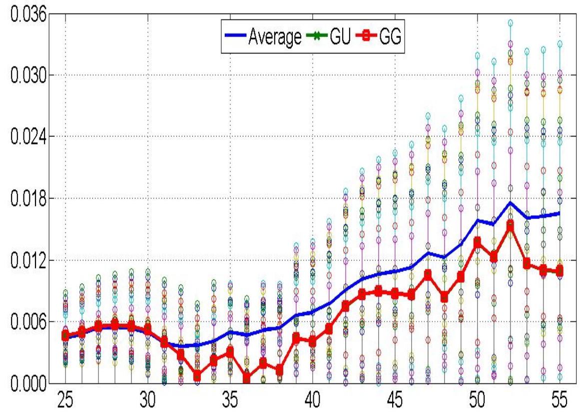
(c) CBD_U (d) CBD_G

(e) $LR-RW_U$ (f) $LR-RW_G$

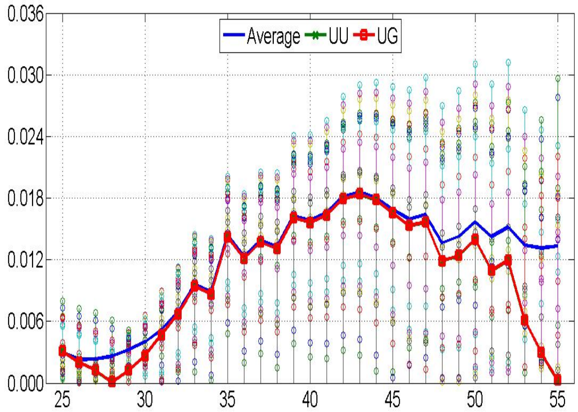
Figure 5.2: $MAPE_x$ for 30-year endowment vs age x for US males



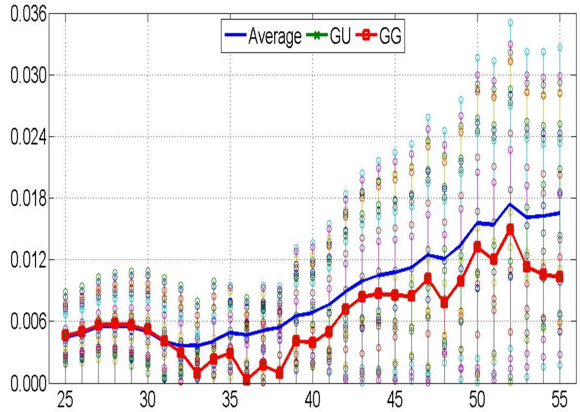
(a) LC_U



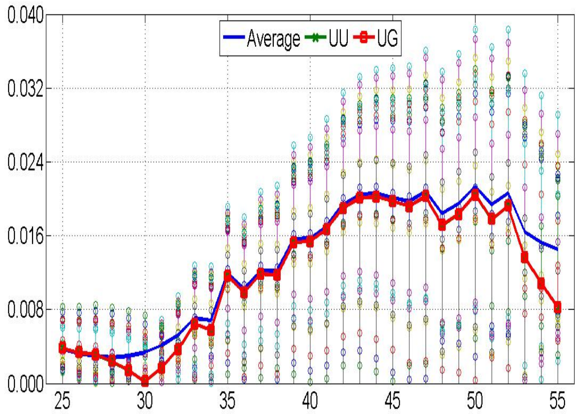
(b) LC_G



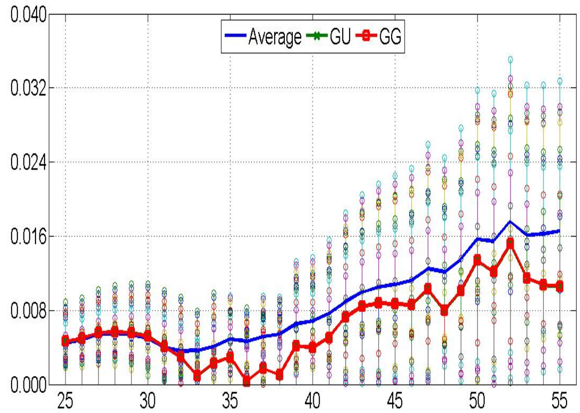
(c) CBD_U



(d) CBD_G



(e) $LR - RW_U$



(f) $LR - RW_G$

Figure 5.3: $MAPE_x$ for 30-year temporary life annuity immediate vs age x for US males

Chapter 6

Conclusion

This project proposes an effective approach to estimating future mortality rates. With quite satisfactory results of predicting mortality rates and pricing life insurance and annuity products, our approach makes contributions to the literature in mortality modeling as well as mortality-linked insurance pricing. In this project we retain the major structure of 3 well-known mortality models, Lee-Carter, CBD and LR-RW models, with only slight modifications in Stage I by dividing the original age span into twelve smaller age groups and conduct mortality modeling. Based on Stage I results, we then apply the pseudo non-parametric Bühlmann estimation to these forecasted mortality rates produced from different fitting year spans. By fitting this pseudo non-parametric Bühlmann model we can calculate a weighted average of the mean of the sample values for age x and the mean of the sample values for the age span. We summarize the numerical results and compare plots using different population datasets; all these outputs have shown that, with such a grouping strategy, applying the pseudo non-parametric Bühlmann credibility approach can significantly reduce the forecasting errors and thus improve the forecasting performance compared with the original mortality models.

As introduced at the beginning, there exists a large number of mortality models to fit and forecast future mortality rates, with various extensions and modifications in order to make them more adaptable to various pricing scenarios. However, as there are more and more complicated insurance products coming out in the market, it becomes more inconvenient for life insurers to apply such sophisticated models. It is previously realized by researchers that there is not a universally best model that can fit all mortality data well. This project takes a further step to point out that there does not exist any identifiable relationships between the choice of fitting year spans and the corresponding forecasting performance. Based on our testings for 3 countries (UK, USA, and Japan) and 2 genders with all possible fitting year spans, we do not find any general rule to conclude that which fitting year span would produce the best forecasting results. As a matter of fact, different selection of fitting year spans will generate very divergent results in some cases; the forecasting performances totally depend on dataset.

In this sense, the approach proposed in this project turns out to be quite a pleasant solution, and this is another favorable advantage our approach possesses: besides the accurate estimation results with low prediction errors, this pseudo non-parametric Bühlmann credibility approach is also quite convenient to implement and can produce stable forecasting performances. Insurers do not need to worry about any intricate modifications or various choices of different fitting year spans. Instead, they can pick up whichever fitting year spans, long or short, or recent or far-off, for a mortality model, that are the most convenient for them to exercise; and after finishing mortality forecasting using the original mortality models, they can easily take the outcomes into this pseudo non-parametric Bühlmann credibility framework and eventually get compromised results. Numerical results have shown that such an approach can work as a stable choice with decent outcomes. We also confirm this conclusion by bringing our approach into a pricing scenario. By computing the net single premiums of 3 types of traditional life insurance and annuity products and comparing the outcomes with the original ones, we support the conclusion that such pseudo non-parametric Bühlmann credibility approach can provide estimates of higher quality; this will certainly help the life insurance companies to exercise more competitive business and earn potential higher profits.

We can extend our work to wider fitting year intervals as well. The numerical results presented in the project demonstrate different performances that continuous fitting year span sequences can produce, i.e. the first fitting year span $1951-2000$ and the second one $1952-2000$ has only one year gap in between. However, it is not necessary for insurance companies to use such fitting year spans; they can definitely choose a 5-year or even 10-year gap between two fitting year spans. Although it is the general case to apply fitting year spans with a gap of 1, 5 or 10 years, it is actually up to the insurance companies; they can select any fitting year spans with which they feel comfortable. Overall, the application of this pseudo non-parametric Bühlmann approach to existing mortality models can help predict mortality rates more accurately, and can also protect the life insurers from some extraordinary deviations caused by selecting a certain fitting year span. With such accurate and stable forecasting performances, this approach serves as an effective way for the life insurance companies to conduct appropriate pricing for all kinds of insurance products.

Bibliography

[1] Berkum, F., Antonio, K., and Vellekoop, M. The impact of multiple structural changes on mortality predictions. *Scandinavian Actuarial Journal* (2015), to appear.

[2] Blake, D., Cairns, A., Dowd, K., and MacMinn, R. Longevity bonds: Financial engineering, valuation, and hedging. *The Journal of Risk and Insurance 73* (2006), 647–672.

[3] Blake, D., Cairns, A. J. G., and Dowd, K. Living with mortality: Longevity bonds and other mortality-linked securities. *British Actuarial Journal 12* (2006), 153–228.

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

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

[6] Buhlmann, H. Experience rating and credibility. *ASTIN Bulletin 4* (1967), 99–207.

[7] Buhlmann, H., and Straub, E. Glaubwurdigkeit fur schadensatze. *Bulletin of the Association of Swiss Actuaries 70* (1970), 111–133.

[8] Cairns, A. Modelling and management of longevity risk: Approximations to survivor functions and dynamic hedging. *Insurance: Mathematics and Economics 49* (2011), 438–453.

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

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

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

[12] Chan, W., Li, J., and Li, J. The CBD mortality indexes: Modeling and applications. *North American Actuarial Journal 18* (2014), 38–58.

[13] Chuang, S., and Brocketta, P. Modeling and pricing longevity derivatives using stochastic mortality rates and the esscher transform. *North American Actuarial Journal 18* (2014), 22–37.

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

[15] Dowd, K., Cairns, A., Blake, D., Coughlan, G. D., Epstein, D., and Khalaf-Allah, M. Evaluating the goodness of fit of stochastic mortality models. *Insurance: Mathematics and Economics 47* (2010), 255–265.

[16] Ericson, W. On the posterior mean and variance of a population mean. *Journal of the American Statistical Association 65* (1970), 649–652.

[17] Frees, E., and Wang, P. Credibility using copulas. *North American Actuarial Journal 9* (2005), 31–48.

[18] Frees, E., Young, V., and Luo, Y. A longitudinal data analysis interpretation of credibility models. *Insurance: Mathematics and Economics 24* (1999), 229–247.

[19] Gerber, H., and Jones, D. Credibility formulas of the updating type. *Transactions of the Society of Actuaries 27* (1975), 31–52.

[20] Giacometti, R., Ortobelli, S., and Bertocchi, M. I. Impact of different distributional assumptions in forecasting Italian mortality rates. *Investment Management and Financial Innovations 6* (2009), 186–193.

[21] Klugman, S., Panjer, H., and Willmot, G. *Loss models : From data to decisions.* John Wiley Sons, Inc., 2012.

[22] Lee, R., and Carter, L. Modeling and forecasting the time series of U.S. mortality. *Journal of the American Statistical Association 87* (1992), 659–671.

[23] Lee, R., and Miller, T. Evaluating the performance of the Lee-Carter method for forecasting mortality. *Demography 38* (2001), 537–549.

[24] Mitchell, D., Brockett, P., Mendoza-Arriage, R., and Muthuraman, K. Modeling and forecasting mortality rate. *Insurance: Mathematics and Economics 52* (2013), 275–285.

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

[26] Sweeting, P. J. A trend-change extension of the Cairns-Blake-Dowd model. *Annals of Actuarial Science 5* (2011), 143–162.

[27] Tolley, H., Nielsen, M., and Bachler, R. Credibility calculations using analysis of variance computer routines. *Journal of Actuarial Practice 7* (1999), 223–237.

[28] Tsai, C., and Yang, S. A linear regression approach to modeling mortality rates of different forms. *North American Actuarial Journal 19* (2015), 1–23.

[29] Wills, S., and Sherris, M. Securitization, structuring and pricing of longevity risk. *Insurance: Mathematics and Economics 46* (2010), 173–185.

[30] Yeo, K., and Valdez, E. Claim dependence with common effects in credibility models. *Insurance: Mathematics and Economics 38* (2006), 609–629.