Joint Mortality Modelling and Forecasting: A New Joint Model based on the Wang Transform

Jianhui Xu

A thesis submitted in fulfilment of the requirements for the degree of Doctor of Philosophy

Doctor of Philosophy in Actuarial Studies

Department of Applied Finance and Actuarial Studies

Faculty of Business and Economics

Macquarie University

Australia

October 28, 2014

Declaration Page

I certify that the work in this thesis entitled "Joint Mortality Modelling and Forecasting: a New Joint model based on the Wang Transform" has not previously been submitted for a degree nor has it been submitted as part of the requirements for a degree to any other university or institution other than Macquarie University.

I also certify that the thesis is an original piece of research and it has been written by me. Any help and assistance that I have received in my research work and the preparation of the thesis itself have been appropriately acknowledged.

In addition, I certify that all information sources and literature used are indicated in the thesis.

Signature: Jianhui Xu

Jianhui Xu

Acknowledgements

I appreciate the kind help of Piet de Jong and Leonie Tickle to me in pursuing my PhD degree. Their carefulness and seriousness encouraged me to fight the bitterness, hardships and frustration. On the way to my PhD degree, Piet and Leonie guided me with penetrating feedback and incisive recommendations. Also, I have learnt from them how to be a respectable academic and the attitude to be an honourable person.

I also wish to thank many people who gave support to my study. Agnieszka Baginska, Lin Bai and Eddy S Dharmadji provided administrative support. Laura Ficorilli, Valerie Williams and Florence Ma have improved my writing. Ana Debón and Maria Russolillo provided the R code of their models. Colleagues in the Department of Applied Finance and Actuarial Studies, such as Xian Zhou, Ken Sui, David Pitt and Timothy Kyng, have provided me with the teaching positions which allowed me to obtain income and working experience; also, Angela Chow and Maree Moses helped me with administrative work. I also thank my friends and PhD colleagues – Stanley Choi, Monica Ren, Frances Chang, Gong Sun, Ade Palupi, Simon Wang and Colin Zhang – who have supported me in my life and academic work. Finally, I thank my family, who are always my optimistic supporters who do not expect repayment.

Looking back on my study during the past few years, the road to my PhD degree has been joyful and unforgettable. This experience is the everlasting wealth in my life.

Table of Contents

Al	ostrac	et in the second s	xi
1	Intr	oduction	1
	1.1	Motivation and findings	1
	1.2	Thesis structure	9
2	Mor	tality Patterns and Trends	11
	2.1	Outline	11
	2.2	Mortality rates have declined over time	12
	2.3	Mortality rates have converged across populations	13
		2.3.1 Mortality rates converge across countries	14
		2.3.2 Mortality rates converge between females and males	15
	2.4	Variation in mortality decline	16
3	Exis	sting Mortality Models – A Review	17
	3.1	Outline	17
	3.2	Individual models	24
		3.2.1 Non-specific model – the Wang Transform model	25
		3.2.2 Age-specific models	30
		3.2.3 Age-period-specific model – the Lee-Carter model	31

		3.2.4	Improvements to the LC model	35
		3.2.5	Age-period-specific model – the Lee-Carter model using func- tional data analysis	40
		3.2.6	Age-period-specific model – the two time factors model	44
		3.2.7	Age-period-cohort-specific models	47
	3.3	Joint n	nodels	49
		3.3.1	Joint Lee-Carter model	50
		3.3.2	Joint Lee-Carter model with three-way decomposition	54
		3.3.3	Joint Lee-Carter model with parallels	57
		3.3.4	Joint age-period-cohort model	59
		3.3.5	Other joint mortality models	62
	3.4	Conclu	sion of existing individual and joint models	63
4	Join	t Wang	Transform Model – the Proposed New Joint Model	65
4	Join 4.1	t Wang Object	Transform Model – the Proposed New Joint Model	65 65
4	Join 4.1 4.2	t Wang Object The JV	Transform Model – the Proposed New Joint Model ives of the new model VT model	65 65 67
4	Join 4.1 4.2	t Wang Object The JV 4.2.1	Transform Model – the Proposed New Joint Model ives of the new model	65 65 67 67
4	Join 4.1 4.2	t Wang Object The JV 4.2.1 4.2.2	Transform Model – the Proposed New Joint Model ives of the new model	65 65 67 67 69
4	Join 4.1 4.2	t Wang Object The JV 4.2.1 4.2.2 4.2.3	Transform Model – the Proposed New Joint Model ives of the new model VT model Age-period-specific WT model Extending the WTAP model to a joint model Alternative joint Wang Transform models	 65 67 67 69 70
4	Join 4.1 4.2	t Wang Object The JW 4.2.1 4.2.2 4.2.3 Fitting	Transform Model – the Proposed New Joint Model ives of the new model VT model Age-period-specific WT model Extending the WTAP model to a joint model Alternative joint Wang Transform models and forecasting mortality using the JWT model	 65 65 67 67 69 70 71
4	Join 4.1 4.2 4.3	t Wang Object The JV 4.2.1 4.2.2 4.2.3 Fitting 4.3.1	Transform Model – the Proposed New Joint Model ives of the new model VT model VT model Age-period-specific WT model Extending the WTAP model to a joint model Alternative joint Wang Transform models and forecasting mortality using the JWT model Estimation of parameters in the JWT model	 65 65 67 67 69 70 71 71
4	Join 4.1 4.2	t Wang Object The JW 4.2.1 4.2.2 4.2.3 Fitting 4.3.1 4.3.2	Transform Model – the Proposed New Joint Model ives of the new model	 65 65 67 67 69 70 71 71 74
4	Join 4.1 4.2 4.3	t Wang Object The JV 4.2.1 4.2.2 4.2.3 Fitting 4.3.1 4.3.2 4.3.3	Transform Model – the Proposed New Joint Model ives of the new model	 65 65 67 67 69 70 71 71 74 77
4	Join 4.1 4.2 4.3	t Wang Object The JV 4.2.1 4.2.2 4.2.3 Fitting 4.3.1 4.3.2 4.3.3 Advan	Transform Model – the Proposed New Joint Model ives of the new model	 65 65 67 67 69 70 71 71 74 77 79

5	Metl	ods and Measures to Evaluate Models 8	85
	5.1	Outline	35
	5.2	Selection of models for evaluation	39
	5.3	Data	90
	5.4	Methods used to forecast time trends	91
	5.5	The jump-off rates	92
	5.6	Transforming mortality scales	94
	5.7	Bayesian information criterion	95
	5.8	Forecast accuracy: mean error and mean absolute error	98
		5.8.1 Mean error	98
		5.8.2 Mean absolute error	99
		5.8.3 Comparative mean absolute error	99
	5.9	Test of significant differences between models	00
	5.10	Convergence properties of mortality across populations	01
	5.11	Rate of mortality decline)1
6	Eval	uating the Fits and Forecasts of Individual and Joint Mortality Models 10)2
	6.1	Introduction to the evaluation and the main findings)2
	6.2	"Across country" analysis, separately for females and males 10)6
		6.2.1 Goodness of fit according to BIC)7
		6.2.2 Overall forecast accuracy	10
		6.2.3 Forecast accuracy over time	21
		6.2.4 Convergence properties of mortality across countries	24
	6.3	"Across sex" analysis, separately for each country	30
		6.3.1 Goodness of fit according to BIC	31

		6.3.2	Overall forecast accuracy – ME	133
		6.3.3	Overall forecast accuracy – MAE	136
		6.3.4	Forecast accuracy over time	140
		6.3.5	Convergence properties of mortality between females and males within a country	144
	6.4	Conclu	sion of the fits and forecasts	149
7	Disc	ussion a	and Further Research Directions	155
Re	eferen	ces		159
Aŗ	pend	ix A E	stimates of the JWT model	175
Aŗ	opend	ix B ".	Across country'' analysis, separately for females and males	181
Aŗ	pend	ix C ".	Across sex'' analysis, separately for each country	187

List of Figures

1.1	Death rates of females and males
1.2	Annual difference of death rates for females and males in the USA from 1950 to 2009 4
3.1	Family tree of mortality models
3.2	Relationships between models
4.1	Estimates of a_x and k_t for females and males in the "across country" and "across sex" analyses $\ldots \ldots \ldots$
4.2	Average of annual differences in the log death rate and the z–scores across ages for females and males in Australia
4.3	The z-scores of Australian females and Dutch males between 1948 and 1994
6.1	Average logit (q_{xti}) over t (1948-1994) of the USA and Japan 110
6.2	Estimates and forecasts of time trends $(k_{ti}^{(j)})$ of Australian females with the HU model using ARIMA process
6.3	Annual rate of forecast log death rate decline in 1995-2009 of Australian females and males using the WT model
6.4	Estimates and forecasts of k_t and k_{ti} using the JLC model for females \therefore 118
6.5	ME of the forecast log death rate over time for females and males 122
6.6	MAE of the forecast log death rate over time for females and males 122
6.7	Standard deviation of log death rate across countries for females 125

6.8	Standard deviation of log death rate across countries for males	126
6.9	Estimates and forecasts of k_{ti} with the JLC model for females \ldots \ldots	129
6.10	ME and MAE over time of the forecast log death rate	141
6.11	Standard deviation of log death rate across females and males in Italy	145
6.12	Observed and forecast log death rate for ages 0 and 49 of Italian females and males using the HU model	148
6.13	Estimates and forecasts of k_{ti} with the JLC model for Italy \ldots	148

List of Tables

3.1	The models reviewed in this research	21
3.2	Limitations of existing joint models and limitations of the relevant studies	64
5.1	The models applied in this research	89
5.2	Orders and forecast methods used by the HU model to forecast time trends	92
5.3	The forecast formulae of the models	94
5.4	Numbers of parameters of the models	96
5.5	Number of parameters (\mathbb{K}) of the selected models $\ldots \ldots \ldots \ldots \ldots$	97
5.6	Number of parameters (\mathbb{K}) used by the HU model for individual populations	98
6.1	BIC and log-likelihood of the models in the fitting period	107
6.2	ME, MAE and CMAE of the forecast log death rate	111
6.3	Rate of mortality decline (ROMD) of observed, fitted and predicted log death rate	114
6.4	Average BIC and log-likelihood across the 13 countries in the fitting period	131
6.5	ME of the forecast log death rate of each of the 13 countries	134
6.6	Ranking of ME of the 13 countries	134
6.7	MAE of the forecast log death rate of the 13 countries	136
6.8	Ranking of MAE of the 13 countries	137
6.9	CMAE of the forecast log death rate of each of the 13 countries	137

6.10	Overall MAE of the forecast log death rate for evaluations of "across country" (Section 6.2.2) and "across sex" (Section 6.3.3)
A.1	Estimates of the Parameters of the JWT model using GLM (across countries)175
B .1	ME of the forecast log death rate for females of the 13 countries 182
B.2	ME of the forecast log death rate for males of the 13 countries
B.3	Ranking of ME for females of the 13 countries
B.4	Ranking of ME for males of the 13 countries
B.5	MAE of the forecast log death rate for females of the 13 countries 184
B.6	MAE of the forecast log death rate for males of the 13 countries 184
B.7	Ranking of MAE for females of the 13 countries
B.8	Ranking of MAE for males of the 13 countries
B.9	CMAE based on the LC model for females of the 13 countries 186
B.10	CMAE based on the LC model for males of the 13 countries
C.1	Log-likelihood of each of the 13 countries
C.2	BIC of each of the 13 countries
C.3	Ranking of BIC

Abstract

Mortality models are mathematical approaches used to facilitate understanding and analysis of mortality patterns and trends, and to provide a basis for mortality forecasting. In an environment in which mortality is continuing to decline, there is considerable interest in developing mortality models that are flexible enough to capture variations in mortality by age, time and various other factors, and robust enough to produce reliable forecasts. Over recent years, there has been growing interest in the development of joint mortality models. Joint models aggregate similar populations to jointly fit and forecast mortality. Such models are able to incorporate the relationships among multiple populations and to ensure that forecast relationships remain reasonable over the long-term. However, existing models – both individual and joint – do suffer from shortcomings.

This research develops a new model – the joint Wang Transform (JWT) model – which aims to address shortcomings and improve upon existing models. As a joint model, the JWT model is able to capitalise on information from similar populations and to ensure that sensible relationships are maintained in the forecasts for such populations. The JWT model allows for a flexible rate of mortality decline over time, which is more realistic than the fixed rate of mortality decline assumed in other widely-used methods. The JWT model has a simple form, reducing the risk of over-parameterization and of unreliable forecasts. The JWT model has flexibility yet builds in constraints, such as ensuring nondivergence of forecasts and hence appears to be appropriate for modelling and forecasting across multiple populations.

This research applies the JWT model and seven existing individual and joint models to fit and forecast the mortality of 13 countries. The joint models are applied by pooling both "across country" separately for each sex, and "across sex" separately for each individual country. Model performance is evaluated by considering goodness of fit, forecasting accuracy, and ability to ensure a sensible relationship between forecasts of similar countries in the long-term (primarily avoiding forecast divergence).

In the analyses of both data sets, the JWT model produces the best forecast accuracy of all eight models according to the evaluation measures. While the evaluation has been conducted only for a selection of developed countries and has compared only a selection of models, the strong performance of the JWT model suggests its potential for further use and evaluation.

Chapter 1

Introduction

1.1 Motivation and findings

This research develops a new model to jointly fit and forecast mortality rates, and evaluates the new joint model by comparing it to existing individual and joint models. The rationale of the research is as follows. Firstly, joint models have advantages for mortality projection over individual models. Secondly, existing joint models have shortcomings in application which the new joint model aims to address. Thirdly, the existing literature lacks a comprehensive evaluation of joint models' performance.

A mortality model is a mathematical or statistical approach to simplify and quantify observed variation in mortality. Mortality data typically consist of a matrix of rates of mortality classified according to ages and years – often called a mortality matrix. Demographers may have difficulties in understanding the patterns or variations in mortality using such a matrix. Mortality models use a limited number of parameters to describe the patterns and variations in mortality in a simple and quantitative way. For example, the Wang Transform model (De Jong and Marshall, 2007) uses one parameter to describe the

variations in mortality for a single mortality matrix. Demographers can then more easily understand and interpret the patterns of mortality without looking at the mortality data directly.

In addition, models facilitate the prediction of future mortality, and mathematical or statistical models enable the estimation of prediction intervals. As described in Section 2.2, mortality rates have decreased throughout the world (Cairns et al., 2009; Leon, 2011; Richel, 2003). This decline varies from one country to another, due to differences in economic conditions, culture and lifestyle (Mackenbach, 2012; McMichael et al., 2004; Meslé et al., 2002). A large number of mortality models have been proposed during the past decades (Booth, 2006; Booth and Tickle, 2008; Cairns et al., 2008) with the aim to better understand and forecast mortality.

Existing mortality models can be classified as individual or joint models. Individual models are applied to individual populations, and therefore implicitly assume that the mortality in one country is unrelated to that in another country. In contrast, joint models jointly analyse mortality by incorporating similar populations. With joint models, populations are assumed to be related and may share trends in mortality.

Individual models are reviewed in Section 3.2. The Lee-Carter model (Lee and Carter, 1992) and the Wang Transform model (De Jong and Marshall, 2007) are two examples of individual models.

Individual models present shortcomings, such as projecting increasing differences in mortality across populations in the long term (Li and Lee, 2005; Li, 2013), referred to as divergence. These shortcomings mainly arise due to the fact that individual models ignore relationships between populations, for example, between countries, or between females and males within a country. Living in the same area and sharing the same culture, females and males in a certain country would be expected to behave similarly and to experience

related mortality trends (Liu et al., 2012; Nolte et al., 2000a,b).

Figure 1.1 illustrates the he nature of the relationship between populations using female and male populations as an example. Displayed are trends in mortality in four countries at selected ages. Disregarding the different scales, females (the dashed curves) and males (the solid curves) experience similar trends and other variation, particularly for the recent decades (since the 1950s).



Figure 1.1: Death rates of females and males

Figure 1.2 illustrates that the relationship also exists in short-term patterns. The figure shows the annual difference in death rates for ages 0, 1, 20 and 80 in the USA from 1950 to 2009. In the same periods, the difference in death rates simultaneously increases or decreases between females and males. Therefore, females and males experience high correlation in mortality development. The high correlation suggests that a coherent fit and forecast of mortality between females and males and males within a country is reliable and necessary. Similar situations occur in other age groups and other countries.





Figure 1.2: Annual difference of death rates for females and males in the USA from 1950 to 2009

Similarities between countries would also be expected to exist. In addition, globalization might be expected to create increasing similarity in mortality trends due to increasing similarity in economic conditions, culture and lifestyle (Li and Lee, 2005). By ignoring the relationships in mortality across countries or between females and males, individual models fail to use correlated information from similar populations, and thus do not fully capture the mortality information available. Individual models may also project unrealistic relationships between the forecasts in related populations, such as divergence. Consequently, individual models and in particular separate projections based on individual models may not be an optimal way to model and forecast mortality of populations with similar characteristics (Li and Lee, 2005).

As described in Section 3.2, joint models are therefore proposed to incorporate relationships by aggregating or disaggregating populations. On the one hand, joint models constrain the differences in mortality between populations, by aggregating similar populations and jointly forecasting their mortality rates (Li and Lee, 2005; Li, 2013). On the other hand, joint models may produce better estimation of mortality rates than individual models, by disaggregating a population and explicitly modelling differences and similarities of its sub-populations (Debón et al., 2011). Compared with modelling individual populations separately, joint models have the potential to more reliably estimate and predict mortality levels and trends (Li and Lee, 2005).

Existing joint models suffer from several shortcomings. The joint models proposed by Li and Lee (2005) and Li (2013) may cause uncertainty in projection due to the use of multiple time factors. The models proposed by Biatat and Currie (2010) and Russolillo et al. (2011) may cause unreliability in projection due to irregular time trends. Some joint models, such as those proposed by Cairns et al. (2011a) and Jarner and Kryger (2009), are suitable for two-population analysis and only used for older ages. The joint model proposed by Debón et al. (2011) is only suitable for geographically close populations, where these populations are assumed to share patterns of mortality and the trends of mortality decline, but possess different scales. In summary, the two major drawbacks of existing joint models are unreliable projection and applicability only to specific populations.

Additionally, many existing individual and joint models ignore an important feature of variation in mortality as described in Section 2.4 – the fact that the rate of mortality decline varies over time. For instance, females in many countries have experienced a smaller rate of mortality decline after the 1970s than before the 1970s (Liu et al., 2012). However, many existing models structure the expected rate of mortality decline as constant over time. For example, the Lee-Carter model (Lee and Carter, 1992) assumes that mortality (the log death rate) varies linearly over time at each age, by using a random walk with a drift process to model and forecast the time factor. If applying a model with the assumption of a constant rate of mortality decline to forecast female mortality, the forecast mortality will therefore generally be lower than the observed mortality. Therefore, the assumption used in many existing individual and joint models is inflexible, and may result in increasing bias in a long-term projection.

To address the shortcomings of existing models, this research proposes a new model as described in Chapter 4. The proposed model is a joint model and is able to jointly fit and forecast mortality of a group of similar populations. Furthermore, the new joint model permits mortality to vary flexibly over time. The new joint model aims to forecast mortality reliably by aggregating similar populations and by applying one time trend. In addition, the new joint model can be generally applied to all age groups or populations. The scale or metric used to model mortality – the z–scores of the survival probability – is different to existing models. Modelling mortality in terms of this metric imposes constraints on the progression of actual mortalities. The empirical results of this thesis suggest that the metric is suitable and practical.

The proposed new joint model is an extension of the individual Wang Transform

model (De Jong and Marshall, 2007). The assumptions of the Wang Transform model are that the z–score of the survival probability to each age is time-varying with a constant drift and that all age groups share the same constant drift. The Wang Transform model applies a constant drift as the unique parameter to model and forecast z–scores. The Wang Transform model is non-specific by disregarding the effects of age, time and cohort.

The Wang Transform model has two main advantages over existing individual models, and thus is appropriate to be extended to a joint model. On one hand, the model is concise, requiring only one parameter for mortality modelling and forecasting. Inheriting the property of simplification, the extension of the Wang Transform model can also be a simple model with convenient application. On the other hand, the Wang Transform model describes changes in mortality on a z–score scale. In the original Wang Transform model, the change in z–scores is a constant drift; however, the extension can incorporate additional parameters to permit this constant drift to vary over time. Therefore, the extension of the Wang Transform model may use a flexible rate of mortality decline to fit and forecast mortality.

The new joint model developed and evaluated in this thesis generalizes the Wang Transform model to jointly fit and forecast mortality, and is referred as the joint Wang Transform (JWT) model. The JWT model is developed in two steps. Firstly, the Wang Transform model is generalized to include age-period-specific factors, allowing for more flexible evolution of z–score behaviour. Hence, the constant drift of the z–score is now replaced by a drift that varies by age and time. Secondly, the age-period specific Wang Transform model is transformed into a joint model from an individual model, by assuming that similar populations share a common drift over time.

The JWT model has advantages over existing joint models in forecasting mortality. First of all, the use of the z–score measure and of a time-varying drift permits a flexible increase of z–scores over time, consistent with observed mortality trends (Lee, 2000; Lee and Miller, 2001; Richel, 2003). Additionally, the JWT model inherits its concise application from the Wang Transform model, using few parameters.

A further shortcoming addressed by the present study is the lack, in the existing literature, of an overall evaluation of joint models' forecast accuracy. For example, Li and Lee (2005), Li (2013), Cairns et al. (2011a) and Jarner and Kryger (2009) did not compare the forecast accuracy of their own joint models to other joint models. Debón et al. (2011) did not produce projections. Biatat and Currie (2010) and Russolillo et al. (2011) provided neither comparison nor projection.

This research evaluates the JWT model and compares it with other existing widelyused individual and joint models, addressing some of the limitations of the existing studies. The JWT and the selected existing models are applied to fit and forecast mortalities of 13 developed countries. The joint models are applied to jointly fit and forecast mortalities "across country" and "across sex" separately. "Across country" refers to jointly fitting and forecasting across the 13 countries, where females and males are separated. "Across sex" indicates joint fitting and forecasting of females and males within each of the 13 countries. The details of mortality data are introduced in Section 5.3.

The evaluation uses two main criteria of the models' performance: goodness of fit and forecasting performance. Goodness of fit reflects the match between observed and fitted mortality data in the fitting period. Forecasting performance assesses forecast accuracy and divergence control. The measures applied in explaining the models' performance will be introduced in Sections 5.7 to 5.11.

The analysis presented in Chapter 6 finds that the JWT model produces the highest forecast accuracy in both "across country" and "across sex" forecasts, for the evaluation measures used. The JWT model – like the existing joint models – yields a constant difference in mortality across countries ("across country") and between females and males

("across sex") in long-term projection, ensuring non-divergence. This research also finds that individual models generally yield lower forecast accuracy than joint models, and produce increasing divergence of mortality in both evaluations in long-term projection.

1.2 Thesis structure

The thesis consists of three parts. Part 1 (Chapters 2 and 3) provides a background of mortality modelling and forecasting, and demonstrates the limitations of existing mortality models and forecasting methods, and the gaps of existing studies in evaluating those mortality models. To address the limitations and fill these gaps, Part 2 (Chapters 4 and 5) proposes a new model – the JWT model – and outlines measures to comprehensively evaluate this new model. Part 3 (Chapters 6 and 7) illustrates the evaluations of this new model using the measures introduced in Part 2 and presents the conclusions.

The background of mortality modelling and forecasting involves both an analysis of variation in mortality (Chapter 2) and an outline of existing mortality forecasting methods (Chapter 3). Mortality rates vary over age, time and other factors, and mathematical models and methods have been developed to reflect this variation and to enable the forecasting of mortality. Methods were usually developed for individual populations, and ignore the additional information that could be incorporated from similar populations and the fact that mortality tends to converge across countries (Wilmoth, 1998; Wilson, 2001) (Section 3.2). More recently, joint models have been developed that jointly fit and forecast mortality of similar populations (Section 3.3). Joint models have advantages over individual models, but may also possess potential shortcomings – for example, uncertain or unreliable forecasts and limited usefulness for specific groups of populations (Section 3.4). Furthermore, the rate of mortality decline by age has varied over time (Section 2.4), whereas most existing individual and joint models assume that the rate of mortality

decline is constant over time, disregarding this important feature of temporal variation.

To address the above shortcomings of existing models, this research develops a new model (Chapter 4). The new model, which is extended from the Wang Transform model (De Jong and Marshall, 2007), permits the rate of mortality to flexibly vary over time (Section 4.2.1) and jointly fits and forecasts mortality by aggregating a group of similar populations (Section 4.2.2). This joint Wang Transform (JWT) model aims to produce reliable projection applying a unique time trend. The JWT model is suitable for application to a range of scenarios, and is not restricted to certain age groups or populations.

A series of measures are introduced to evaluate the JWT model (Chapter 5). The evaluation compares the JWT model with a selection of existing individual and joint models (Section 5.2), using mortality data sets of 13 countries (Section 5.3). The methods used to forecast mortality are outlined (Section 5.4) along with the measures to evaluate fit (Section 5.7) and forecasting performance (Section 5.8 and 5.10).

The new model and the selected existing models are applied to fit and forecast mortality for evaluations (Chapters 6). The evaluations consist of an "across country" analysis (Section 6.2) and an "across sex" analysis (Section 6.3). The "across country" analysis jointly fits and forecasts mortality across the 13 countries, where females and males are analysed separately. The "across sex" analysis jointly fits and forecasts mortality of females and males for each of the 13 countries. Each analysis illustrates the models' goodness of fit (Sections 6.2.1 and 6.3.1), forecast accuracy (Sections 6.2.2 and 6.2.3, Sections 6.3.2, 6.3.3 and 6.3.4) and divergence control (Sections 6.2.4 and 6.3.5). Conclusions are then drawn from these two analyses (Section 6.4) and from the research as a whole (Chapter 7).

Chapter 2

Mortality Patterns and Trends

2.1 Outline

This chapter introduces and discusses mortality patterns and trends in order to demonstrate features that must be dealt with by mortality modelling and forecasting. Mortality varies over ages, time, populations and other factors, where time and populations are two of the main dimensions of variation in mortality. Variation of mortality over time describes mortality trends. Mortality rates have declined throughout the world (Section 2.2). The decline in mortality has been variable and the future progression is uncertain. Variation of mortality across populations explains the cross sectional relationships between rates of mortality, for example, across countries or between females and males (Section 2.3). Variation across cohorts deals with differences across groups of people born at different points of time.

These different sources of variation in mortality are typically modelled mathematically and these mathematical structures are called mortality models (Chapter 3). Individual models – incorporating the variation in mortality over time – are discussed in Section 3.2. Joint models – incorporating the variation in mortality over time and across populations – are introduced in Section 3.3.

Many models, both individual and joint, do not incorporate an important feature of mortality as described in Section 2.4 – the fact that the rate of mortality decline varies over time. Most existing mortality models view this rate as constant, which typically results in increasing bias when forecasting mortality.

2.2 Mortality rates have declined over time

Mortality rates have declined, or life expectancy at birth has increased, throughout the world since the 1800's. From 1840 to 2009, the average life expectancy at birth among developed countries rose from under 40 to 80 years (Vaupel et al., 2011). From 1900 to 1988, in the USA, the combined sex life expectancy rose from 47 to 75 years (Lee and Carter, 1992). From 1901-10 to 1999, in Australia, female life expectancy increased from 57 to 80 years (Booth et al., 2002a). From the beginning of the 1900's to 2000, in the Netherlands, male life expectancy at birth increased from 47 to 76 years, and that of female's increased from 50 to 80 years (Hári et al., 2008).

To demonstrate the importance of joint models, the relationships in mortality across populations are emphasised in this section. Population refers to females, males or combined females and males in an area, such as a country. The trend in mortality decline has become increasingly similar between countries and between genders since the 1950s (Wilmoth, 1998; Wilson, 2001). This increasing similarity in mortality decline calls for mortality models that incorporate similar populations. Separate projections are often unrealistic and result in increasing differences in mortality between populations (Li and Lee, 2005). The universal decline in mortality rate is mainly due to the development of medical and biological technology, and improved public health care and economic conditions (Cutler and Meara, 2001; McMichael et al., 2004; Moser et al., 2005; Klenk et al., 2007; Gero, 2010; Finch, 2010; Vaupel et al., 2011). Economic conditions are the foundation for all development and underpin the other two factors (Crimmins and Finch, 2006; Cutler and Kadiyala, 1999; Cutler and Meara, 2001; Floyd et al., 2012; Preston et al., 1976; Vaupel et al., 2011). The advancement in medical and biological technology has improved human beings' ability to fight various diseases. The public health care system creates a platform to help the whole population in heath care and in achieving a public hygienic environment. Improvements in public health care particularly benefit the old ages and low income populations.

2.3 Mortality rates have converged across populations

Relationships in mortalities between populations can be expressed by "convergence" and "divergence". Convergence (divergence) is defined as the similar (different) mortality decline trends between populations over time (Mathers et al., 2001; Moser et al., 2005). The convergence or divergence can be across countries or between females and males. Throughout the world, there has been convergence across most countries and between females and males within most countries since the 1950s (Wilmoth, 1998; Wilson, 2001). In an environment in which convergence exists, joint mortality modelling and forecasting offers significant advantages (Jarner and Kryger, 2009; Li and Lee, 2005).

2.3.1 Mortality rates converge across countries

Due, in part, to the global spread of mass schooling, information technology and literacy, and convergence in economic conditions and public health measures, countries have experienced increasing similarity in mortality decline (McMichael et al., 2004; Wilson, 2011). Overall, trends in mortality have become increasingly similar throughout the world – across 184 territories and countries of the United Nations – in the past 50 years (Wilson, 2001).

Since economic growth is an important factor in mortality decline, countries with different economic conditions have experienced different patterns of mortality decline (Mathers et al., 2001; Meslé et al., 2002). Developed countries have experienced increasingly similar economic conditions, and thus increasingly similar mortality rates, resulting in convergence since the 1950s (Coale, 1996; White, 2002; Wilmoth, 1998).

Developing countries have experienced a higher rate of growth in their economies than developed countries after the 1950s, and the difference in mortality between developed and developing countries has also been narrowing (Elo and Preston, 1992; McMichael et al., 2004). The convergence in mortality across countries has resulted in a narrowing life expectancy between developing and developed countries from 1950 to 2010 (from 28 to 17 years for females and from 31 to 14 years for males) (Wilson, 2011).

In poor areas, like sub-Saharan Africa (McMichael et al., 2004), Central Europe (Mackenbach, 2012) and Eastern Europe (Moser et al., 2005; Shkolnikov et al., 1998; Wilson, 2011), the mortality rate has stagnated or increased. These odd patterns of mortality decline trends have been mainly due to poverty and lack of medical measures. For instance, AIDS contributed to over 40% of deaths in poor African countries (Dorrington, 2006), for example, Northern Malawi (Jahn et al., 2008), Mwanza of Tanzania (Urassa et al., 2001) and KwaZulu Natal of South Africa (Hosegood et al., 2004). Malaria contributed to about 50% of the total mortality in poor African countries (Murray et al., 2012), but it is treatable in developed and developing countries.

In summary, convergence in mortality dominates throughout the world. Most countries – developed and developing – have similar patterns in mortality decline trends. The exception is that in some countries with poor socio-economic and environmental conditions, mortality rates are stagnating or increasing.

2.3.2 Mortality rates converge between females and males

Mortality rates between females and males have become more and more similar in the past few decades (Hosseinpoor et al., 2012; Liu et al., 2012). The sex difference in life expectancy generally increased before the 1960s, stagnated between the 1970s and 1980s, and decreased after the 1980s in most developed countries and some developing countries (Glei and Horiuchi, 2007; Liu et al., 2012; McCartney et al., 2011; Saito et al., 2012; Trovato and Heyen, 2006).

However, the sex difference in mortality is different in Japan, Turkey and the former Soviet Union countries. Japanese females have had the largest rate of change in mortality throughout the world from 1950 to 1990, and thus the sex difference in life expectancy has increased in Japan since 1950 (Liu et al., 2012, 2013). Turkish females and males have held a relatively constant rate of change in mortality, and thus the sex difference in life expectancy has stagnated in Turkey since 1960 (Liu et al., 2012, 2013). Heavy consumption of alcohol has caused the increasing mortality rate in males in the former Soviet Union countries since 1990. For these reasons, the sex difference in life expectancy has increased in these countries (Cockerham, 1997; McMichael et al., 2004; Men et al., 2003; Shkolnikov et al., 1998).

2.4 Variation in mortality decline

As seen in the previous two sections, the rate of mortality decline is not constant. For example, the overall death rate in the USA decreased by 1% p.a. between 1900 and 1940, 2% p.a. between 1940 and 1955, 0% between 1955 and1965 and 1%-1.5% since 1965 (Cutler and Meara, 2001). Also, young and old age groups experience different rates of mortality decline over time (Cutler and Meara, 2001; Lee, 2000). In addition, as described in Section 2.3, the convergence of mortality across populations indicates that the rate of mortality decline varies over time. For example, Japan had a smaller rate of mortality decline after the 1990s than before the 1990s, but other developed countries experienced a larger rate of mortality decline after the 1990s than before the 1990s than before the 1990s than before the 1990s.

Over time, the rate of mortality decline varies. Most existing mortality forecasting methods assume that the rate of mortality decline is constant. This assumption may be an acceptable approximation for certain populations, but may result in increasing underestimation or overestimation for general projection as discussed in Section 3.4.

Chapter 3

Existing Mortality Models – A Review

3.1 Outline

Mortality models are mathematical and statistical approaches that describe the variation in mortality. Mortality data comprise of a matrix of mortality rates classified according to age and time. These numerics are difficult to understand without using a simple mathematical expression. Hence, mortality models aim to intuitively reflect mortality data using limited demographical explanation parameters. For example, a mortality model may use relatively simple age and time factors to quantify the variation in mortality across age and over time. Thus, mortality models simplify the expression of mortality data and increase demographers' understanding of mortality patterns. Furthermore, mortality models may provide intuitive ways to forecast mortality. A mortality model may assume mortality to vary linearly over time and use a time series – for example, random walk with drift – to model the time factors. Such usefulness drives the development of mortality models.

This chapter reviews the existing commonly used mortality models. To develop a new model, this chapter illustrates the advantages and disadvantages of existing models,

by presenting the commonly used models. The new model, then, absorbs these advantages and avoids these disadvantages. Therefore, the review of models is foundational to develop a new model as detailed Chapter 4.

Models are then selected from the reviewed models, as detailed in Section 5.2, to compare with the new model for a comprehensive evaluation as introduced in Chapter 6. The selected models may be either directly related to the new model or the most commonly used. The selected models constitute a subset of the reviewed models, and the reviewed models comprise a subset of all existing models. The selected models represent the general performance of existing mortality models in fitting and forecasting. Comparison with the selected models reflects the practicability and performance of the new model.

Figure 3.1 displays a family tree of existing mortality models. The models to be reviewed are displayed in green rectangles and the new model – the joint Wang Transform (JWT) model – is displayed in the red rectangle. The mathematical form and relevant sections are shown in Table 3.1.



Figure 3.1: Family tree of mortality models

To model and forecast mortality in a reliable way, numerous methods have emerged (Booth and Tickle, 2008), since Gompertz (1825) proposed his mortality model. Mortality forecasting methods may be classified as expectation, explanation and extrapolation (Booth, 2006; Booth and Tickle, 2008). The expectation methods rely on experts' opinions to determine the future path of mortality trends. This approach may cause bias in short-term projections (Booth, 2006; Lee and Carter, 1992; Waldron, 2005) and may be conservative and thus underestimate the rate of mortality decline (Alho and Spencer, 1990; Lee and Carter, 1992; Lee and Miller, 2001; Waldron, 2005). The explanation methods apply structural or epidemiological models to quantify the relationships between the causes of death and the total mortality by using regression methods (Tabeau, 2002). This approach may produce poor accuracy in the long-term projections (Booth and Tickle, 2008; Keyfitz, 1982) and may result in unstable forecasts due to the unstable relationships between mortality and the independent factors (Booth, 2006). The extrapolative methods assume that past mortality trends will continue in future, and thus past trends are encapsulated in the model and the model aims to limit subjective judgement. Extrapolation models are widely used (Booth, 2006).

Most models to date have been applied to individual populations, referred to as individual models. Mortality experience in countries has become increasingly similar as described in Section 2.3, and thus the separate fitting and forecasting of mortality has become relatively unreliable (Li and Lee, 2005). Models are now proposed to incorporate mortalities of multiple similar populations, referred to as joint models.

This chapter focuses on extrapolative individual models, the most widely used type of model. Extrapolative models can be divided into four categories regarding the number of factors: non-specific, age-specific, age-period-specific and age-period-cohort-specific models. This chapter reviews the non-specific and age-period-specific models in detail. The Wang Transform model (De Jong and Marshall, 2007) is a non-specific model and is the model from which the new model is derived as detailed in Chapter 4. Age-periodspecific models are the most widely used and are foundational for many existing joint models. Age-specific and age-period-cohort-specific models are unreliable in forecasts (Booth, 2006; Booth and Tickle, 2008) as described in Section 3.2.2. However, the disadvantages of age-specific and age-period-cohort-specific models can be considered when developing the new model.

Therefore, this chapter reviews the individual models in detail for the non-specific and age-period-specific models, and briefly for the age-period-cohort-specific models (Section 3.2). The joint models derived from these reviewed individual models are also detailed (Section 3.3). The review of each model comprises the introduction, estimation and evaluation. The "introduction" describes the definitions of the model. The "estimation" explains the methods used to estimate the parameters. The "evaluation" assesses the advantages and disadvantages of the model.

Table 3.1 lists the models to be reviewed. Column 1 displays the number of factors. Columns 2, 3 and 4 give the names and the mathematical forms, where the error item, ϵ , is omitted. Columns 5 and 6 illustrate the relevant sections and references.

Variables a and b, k and γ denote age, time and cohort factors. The a denotes the basic age-specific factor, and b denotes the age-specific coefficient response to k or γ . The ϵ denotes error terms. The ρ denotes the population effects. Subscripts x, t and i denote the index of ages, year and population, respectively, where x = 1, ..., m, t = 1, ..., n and $i = 1, ..., \ell$. Often the first age is 0, and then x = 1 denotes age 0.

Variables in joint models may have more than one subscript – often age x and/or time t – with the population effect subscript i. For example, in a joint model, a_{xi} denotes the basic age-specific factor for age x of population i and a_x denotes the common basic age specific factor for age x across all populations. The ρ_i denotes the effect of population i.

			Table 3.1: The models reviewed in this rest	earch	
Factors	Name	Scale	Mathematical form	Section	Reference
Zero	WT		X	3.2.1	De Jong and Marshall (2007)
	WTAP	Δxt	$a_x + k_t$	4	developed by this research
	ГС			3.2.3	Lee and Carter (1992)
	ΓM			3.2.4	Lee and Miller (2001)
Two	BMS		$ax \perp axht$	3.2.4	Booth et al. (2002b)
D M T	BDV	$\log(m_{xt})$		3.2.4	Brouhns et al. (2002)
	DT		$Xa_x + Xb_xk_t$	3.2.4	De Jong and Tickle (2006)
	ASE		$a_x + \sum_{j=1}^{r} b_x^{(j)} k_t^{(j)}$	3.2.4	Renshaw and Haberman (2003c)
	ΗU		$a(x) + \sum_{j=1}^{J} b^{(j)}(x) k_t^{(j)}$	3.2.5	Hyndman et al. (2007)
	CBD		$k_t^{(1)} + k_t^{(2)}(x - ar{x})$		Cairns et al. (2006)
	CBD1	$logit(a_{rt})$	$k_t^{(1)} + k_t^{(2)}(x - \bar{x}) + \gamma_c$	3.2.6	
	CBD2	(am1) 0	$k_{t}^{(1)} + k_{t}^{(2)}(x - \bar{x}) + k_{t}^{(3)}\left\{(x - \bar{x})^{2} - \hat{\sigma}_{x}\right\} + \gamma_{c}$		Cairns et al. (2009)
Three	CBD3		$k_t^{(1)} + k_t^{(2)}(x - \bar{x}) + \gamma_c(x^* - x)$		
	RH		$a_x + b_x^{(1)}k_t + b_x^{(2)}\gamma_c$	<i>τ τ</i> ε	Renshaw and Haberman (2006)
	Currie	$\log(m_{xt})$	$a_x + k_t + \gamma_c$	7.7.6	Currie (2006)
	Plat		$a_x + k_t^{(1)} + (\bar{x} - x)k_t^{(2)} + (\bar{x} - x)^+ k_t^{(3)} + \gamma_c$	3.2.7	Plat (2009a)
	SJLC		$a_{xi} + b_x k_t$		Li and Lee (2005)
	JLC		$\alpha + \pm b + b + b + b$	3.3.1	Li and Lee (2005)
	PJLC		$u_{xi} \top u_{xi} h_{ti} \top u_{xh}$		Li (2013)
	TLC	$\log(m_{xti})$	$a_{xi} + B_x K_t \rho_i$	3.3.2	Russolillo et al. (2011)
Joint	CBDCK		$a_{xi} + m^{-1}k_{ti} + m^{-1}\gamma_{ci}$	3.3.4	Cairns et al. (2011a)
	CLC		$a_x + b_x k_t$	3 1	Lee and Carter (1992)
	ILC		$a_{xi} + b_{xi}k_t$	7.1	Li and Hardy (2011)
	PLC	$logit(q_{xti})$	$a_x + b_x k_t + \rho_i$	3.3.3	Debón et al. (2011)
	JWT	λ_{xti}	$a_x + k_t$	4	developed by this research

Table 3-1. The models reviewed in this res

The Lee-Carter (LC) model can be viewed as a "base" or benchmark model. The common time trend LC model (CLC) and the integration time trend LC model (ILC) are included to complete the framework, but are not reviewed further. The CLC model is between the simple joint LC (SJLC) and parallel LC (PLC) models, assuming that all populations share the same patterns and trends of mortality. The ILC model is between the LC and SJLC models, assuming that populations share a common time trend.

Table 3.1 illustrates an overview of some of the existing popular mortality models, where the LC model and its extensions constitute a LC family. The models of the LC family generally directly model mortality – the central death rate m or one year death rate q. However, the WT and JWT models are different from the LC models and its extensions, as they apply different mortality scales λ , where λ denotes the change of z–scores of the survival probability across years. The advantages of such an application are discussed in Section 3.2.1 and incorporated in this research when proposing the new model as described in Chapter 4.

Figure 3.2 displays the relationships between the models listed in Table 3.1. The yellow, blue and green colours represent zero-, two- and three-factor models. The rectangles with round and square corners denote individual and joint models. The single-head arrows between models express the derivation directions, and items around the arrows show the change being made. The words "General" and "Restrictive" represent the directions of derivation. The \mathcal{F} denotes coefficient.

For population *i*, the LC model can be written as:

$$\log(m_{xti}) = a_{xi} + b_{xi}k_{ti} + \epsilon_{xti}$$

where m_{xti} denotes the central death rates for the age x, year t of population i and


Figure 3.2: Relationships between models

 $log(m_{xti})$ denotes the logarithm of the central death rate, named the log death rate in the following parts of this thesis. The other symbols are detailed in Section 3.2.3.

Note in this version all effects are population specific and hence there is no "communality" across populations. Models may be derived from another model by specializing or generalizing, for example, the age base level (a_{xi}) , and the age-specific coefficient response to the time trend (b_{xi}) and time trend (k_{ti}) . The ILC model is derived from the LC model, by equating the k_{xi} across i: $k_{xi} = k_x$. The SJLC model is derived from the ILC model by equating the b_{xi} across all i: $b_{xi} = b_t$. Thus, the SJLC model is more restrictive than the ILC and LC models. The JLC model is derived from the SJLC model by adding the individual factors, $b_{xi}k_{ti}$, so the JLC model is more general than the SJLC model.

Figure 3.2 illustrates two clusters – the LC family and the WT family. Many reviewed models belong to the LC family. However, the WT model does not fit into the generic LC family – it operates on a completely different scale – the z–score derived from the survival probability. By using a different mortality scale, the WT family model provides another way to model and describe the properties of mortality.

3.2 Individual models

Individual models deal with a single population in isolation of any other population. Individual models may have one or more factors and can be classified as non-specific, age-specific, age-period-specific and age-period-cohort-specific in each case detailing the number of factors and their type – zero, one, two or three – used to explain mortality variations. The non-specific, age-period-specific and age-period-cohort-specific models are practical and commonly used. Hence, the individual models reviewed in this section are confined to these three classifications of models. Non-specific models structure mortality without specific reference to age, time or other factors. Non-specific models are simple to apply but lack flexibility. An example of non-specific models is introduced in Section 3.2.1.

Age-specific models assume that mortality is a function of age. This function consists of different parameters. To forecast, estimates of a parameter can be modelled and predicted using a time series process. Due to impracticability, age-specific models are not commonly used (Booth and Tickle, 2008), and thus are not considered further here.

Age-period-specific models assume that mortality varies across the age and time dimensions. Age factors depend on age groups and are generally constant over time. Time factors follow time series processes typically shared by all age groups. Age-periodspecific models are the most widely used in mortality modelling and forecasting (Booth and Tickle, 2008). The commonly used age-period-specific models are introduced in Sections 3.2.3 to 3.2.6.

Age-period-cohort-specific models incorporate cohort effects, where a common cohort effect is shared by to those born in the same year. Cohort effects trends are often irregular and difficult to forecast (Booth and Tickle, 2008). Several age-period-cohortspecific models are introduced in Sections 3.2.6 to 3.2.7.

3.2.1 Non-specific model – the Wang Transform model

The WT model

De Jong and Marshall (2007) applied the Wang Transform proposed by Wang (2000) to fit and forecast mortality. This is called the WT model. To adjust an insurance premium for "risk," each exceedance probability for the risk can be transformed using the normal cumulative probability to a z–score. The actual exceedance probability function is then

adjusted for risk by uniformly shifting each of the z–scores (Wang, 2000) and using these risk-adjusted z–scores, and the implied exceedance probabilities in the actual premium calculation. This process of adjusting excellence probabilities via shifting the z–scores is often called the Wang Transform method.

De Jong and Marshall (2007) applied the Wang Transform in the context of mortality survival probabilities. A survival probability and z-score are calculated for each age and time period. Consecutive z-scores corresponding to consecutive periods are then assumed to be related by a fixed shift. In an empirical study (De Jong and Marshall, 2007), the z-score shift appears to be virtually independent of age and time leading to the conclusion that improvements in mortality can be reliably modelled using one "z-score shift" parameter. The parameter – called the z-score drift – captures the dynamics across the entire age spectrum. Thus, in the simplest case, all ages are assumed to share the same constant drift on the z-score sover time is called the Wang Transform (WT) model. The WT model has the unique parameter – the constant drift of z-scores – and is one of the simplest existing mortality models.

The WT model is written as follows:

$$z_{xt} - z_{x,t-1} = \lambda_{xt} \approx \lambda + \epsilon_{xt} = \rho + \epsilon_{xt} , \qquad (3.1)$$

where z_{xt} denotes the z-score for age x in year t, λ_{xt} denotes the change or drift of zscore between year t - 1 and t and λ is the unique parameter of the WT model. For a given population, λ is assumed to be constant. Typically one would expect λ to vary across populations, although this was not considered by De Jong and Marshall (2007). To be consistent with the other population effects models surveyed and reviewed in this thesis, λ can be written as ρ .

Estimation of parameter in the WT model

The WT model has two steps for modelling and forecasting mortality. Step 1 calculates the survival probability (s_{xt}) :

$$s_{xt} = p_{0t}p_{1t}\cdots p_{x-1,t} = \prod_{j=0}^{x-1} p_{jt} , \qquad s_{1t} = p_{0t} , \qquad s_{0t} = 1 , \qquad (3.2)$$

 s_{xt} denotes the probability of surviving from birth to age x and p_{xt} is the one year survival rate and denotes the probability of surviving between age x and x + 1. There is $p_{xt} = 1 - q_{xt}$, where q_{xt} is the one year death rate and denotes the probability of dying between age x and x + 1. The q_{xt} can be replaced by, for example, m_{xt} , where m_{xt} is the mid year death rate and denotes the probability of age x dying at the middle of year t. Both q_{xt} and m_{xt} can be obtain directly from a life table. All rates are those applying in year t.

Step 2 calculates z-scores:

$$s_{xt} = \Phi(z_{xt}) , \qquad z_{xt} = \Phi^{-1}(s_{xt}) , \qquad (3.3)$$

where Φ denotes the cumulative function of normal distribution and Φ^{-1} is its inverse mapping function. Using CDF simplifies the expression of survival probabilities.

Step 3 estimates the change in the z–score via the following computations:

$$\lambda_{xt} = z_{xt} - z_{x,t-1}, \qquad \hat{\lambda}_x = \frac{1}{n-1} \sum_t \lambda_{xt}, \qquad \hat{\lambda} = \frac{1}{m} \sum_x \hat{\lambda}_x, \qquad (3.4)$$

where $\hat{\lambda}_x$ and $\hat{\lambda}$ denote the estimates of λ_x and λ and $\hat{\lambda}$ is the estimated overall parameter.

After estimating $\hat{\lambda}$, the fitted z-scores and the survival probabilities are

$$\hat{z}_{xt} = z_{x1} + (t-1)\hat{\lambda}$$
, $\hat{s}_{xt} = \Phi(\hat{z}_{xt}) = \Phi\left\{z_{x1} + (t-1)\hat{\lambda}\right\}$. (3.5)

where \hat{s}_{xt} is the estimate of the survival probability, s_{xt} , and z_{x1} is the observation of the z-scores at the first year of the fitting period.

The forecasts of z-scores and the survival probability are

$$\hat{z}_{x,n+h} = z_{xn} + h\hat{\lambda} , \qquad \hat{s}_{x,n+h} = \Phi\left(\hat{z}_{x,n+h}\right) = \Phi\left(z_{xn} + h\hat{\lambda}\right) , \qquad (3.6)$$

where h denotes the h year ahead forecast. The fitted and forecast q_{xt} are calculated from the fitted or forecast surival probabilities as follows:

$$\hat{q}_{xt} = 1 - \hat{p}_{xt}$$
, $\hat{p}_{xt} = \frac{\hat{s}_{x+1,t}}{\hat{s}_{xt}}$, $\hat{p}_{0t} = \hat{s}_{1t}$, $\hat{s}_{0t} = s_{0t} = 1$, (3.7)

where \hat{p}_{xt} and \hat{s}_{xt} denote the estimates of p_{xt} and s_{xt} .

Advantages of the WT model in forecasting mortality

The WT model describes the variation in mortality in terms of z-scores, differing from most existing models which use the log death rate. The z-scores are transformed from the survival probabilities, and the survival probabilities are the product of the one year survival rates across age. Over time, survival probabilities vary more smoothly than the mortality rates by individual age and year. Modelling smooth scales is often more reliable than modelling jagged scales, for example, the log death rates.

Furthermore, for each age, z-scores vary linearly over time. The linearity is easy to model by using a linear mathematical expression. The WT model captures the trends of z-scores by using a constant drift λ or ρ , where all age groups share the same constant drift of z-scores over time. The WT model uses few parameters to describe mortality in terms of z-scores, and thus is one of the simplest existing mortality models. Therefore, the WT model is simpler than many other existing mortality models, for example, the Lee-Carter model (Lee and Carter, 1992) as described in Section 3.2.3.

Thirdly, although λ is assumed to be constant, the WT model potentially allows for a flexible drift of z–scores (De Jong and Marshall, 2007). The rate of mortality varies over time as described in Section 2.4. This property enables the WT model to be extended for wider application as discussed in Chapter 4.

Limitations of the WT model in forecasting mortality

A shortcoming of the WT model is that in some applications the assumption of a constant λ may be unreasonable. The original application found that the $\hat{\lambda}$ overestimates the change of z_{xt} for the young age groups but underestimates that for the old age groups (De Jong and Marshall, 2007, Figure 3). Furthermore, the z_{xt} fluctuates over time more seriously for the old age groups than for the young age groups. Therefore, the simple WT model may not be suitable for the old age groups and may need to be extended to cover these apparent extra sources of variation.

Moreover, transforming z_{xt} to the log death rate, $log(m_{xt})$, may result in curious projection patterns over the long term. When transforming z_{xt} to $log(m_{xt})$, the constant increase in z_{xt} over t results in the accelerated decline in $log(m_{xt})$ over t. Therefore, when using the WT model, mortality (the log death rate) varies nonlinearly over time. Consequently, the properties of the WT model conflict with the common view that mortality – in the log death rate scales – behaves linearly over time (Cairns et al., 2006; Carter and Lee, 1992; Lee and Carter, 1992).

In addition, if the mortality data set does not start from age 0, the WT model is impractical. The $\hat{\lambda}_x$ sequence increases across ages (De Jong and Marshall, 2007, Figure 6), if the start age is 50, 65 or 80. For these cases, $\hat{\lambda}$ is not the average of $\hat{\lambda}_x$ across x, and thus the WT model is not applicable. Lastly, since λ depends on populations, the WT model allows individual populations to possess their own drift of z-scores. This individual drift may result in increasing differences in z-scores between populations in projection.

3.2.2 Age-specific models

Age-specific models assume that annual mortality rate can be expressed by a function of age. This function consists of different parameters. For one parameter, when applying the function to annual mortality rates alternatively, there are a series values of this parameter which can can be modelled as a time series process. Therefore, mortality rates are projected vis the projection of every parameter.

Heligman and Pollard (1980) applied a function with eight parameters, called the HP model, to capture the age patterns of mortality. The HP model combines three functions to describe mortality – a rapidly declining exponential function for mortality in childhood, a lognormal function for mortality of the 'accident hump' in young age groups and a Gompertz exponential function for mortality of adults and in the old age group. Rogers and Planck (1983) extended the HP model, called the RP model, by incorporating a constant parameter for all ages and using a double exponential function for the 'accident hump' in the young age group. Similar age specific models have been proposed (Carriere, 1992; Hannerz, 1999, 2001a,b; Siler, 1983).

An advantage of age-specific models is that mortality rates are smooth across ages (Booth and Tickle, 2008). Furthermore, age-specific models are able to interpret the relationships between across ages (Bell, 1997).

However, age-specific models have two main disadvantages. Firstly, estimated parameters correlated to other parameters and these correlations affect the parameters' interpretability (Forfar and Smith, 1985). Secondly, the series of parameters may be irregular and thus difficult to forecast (Booth and Tickle, 2008) and cause unreliable mortality projection (Bell, 1997). Therefore, the age-specific models are not detailed in this research.

3.2.3 Age-period-specific model – the Lee-Carter model

The LC model

Lee and Carter (1992) introduced an age-period-specific model to forecast mortality, referred to as the Lee-Carter (LC) model. The LC model considers age and period factors and can be viewed as an improvement of the age-period mortality models given by Bell and Monsell (1991) and Bell (1992). Differing from the traditional method used by the Social Security Administration (SSA) and the US Census Bureau (USCB), the LC model relies on mortality data and excludes experts' ideas. For the LC model, the mortality trend in the future is the extension of that in the past obtained via mortality data.

The LC model applies two age factors and a time factor to express the log death rate. Age factors are constant over time. The time trend can be modeled as a time series of a random walk with drift. All age groups share the same trend in mortality decline.

The LC model is:

$$\log(m_{xt}) = a_x + b_x k_t + \epsilon_{xt} , \qquad (3.8)$$

where $\log(m_{xt})$ is the observed log death rate at age x and year t. The definition of a_x , b_x , k_t and ϵ_{xt} are described in Section 3.1. The a_x , b_x and k_t are unknown parameters. The a_x sequence is the "base" level of mortality. The k_t sequence describes the overall change of $\log(m_{xt})$ over time, and k_t sequence is assumed as a random walk with drift. The b_x sequence is the age-specific coefficient response to k_t , reflecting the rate of mortality decline for the specific age group x.

Estimation of parameters in the LC model

The a_x is estimated to be the average of $\log(m_{xt})$ across t,

$$\hat{a}_x = \frac{1}{n} \sum_{t=1}^n \log(m_{xt}) ,$$
(3.9)

where \hat{a}_x is the estimate of a_x . For any non-zero c, $b_x k_t = (b_x c) \times (k_t/c)$, and thus b_x and k_t are not unique. Constraints are introduced to ensure the uniqueness of b_x and k_t :

$$\sum_{x} b_x = 1 , \qquad \sum_{t} k_t = 0 . \qquad (3.10)$$

The b_x and k_t are estimated via singular value decomposition (SVD). Let Y denote a matrix with dimension $m \times n$ and entries y_{xt} , where

$$y_{xt} = \log(m_{xt}) - \hat{a}_x , \qquad Y = \begin{pmatrix} y_{11} & \cdots & y_{1n} \\ \vdots & \ddots & \vdots \\ y_{m1} & \cdots & y_{mn} \end{pmatrix} ,$$

Let b_x and k_t be expressed in terms of U, D and V,

$$Y' = U_{[n \times n]} D_{[n \times n]} V'_{[n \times m]} , \qquad \tilde{b}_x = V_{x1} , \qquad \tilde{k}_t = D_{11} U_{t1} , \qquad (3.11)$$

where Y' is the transposition matrix of Y. Matrices U and V are the left and right singular vectors of Y', and D is a diagonal matrix containing the singular values of Y'. Dimensions of U, D and V are shown in square brackets.

The \tilde{b}_x and \tilde{k}_t are initial estimates of b_x and k_t . The constraints in (3.10) are applied

to adjust \tilde{b}_x and \tilde{k}_t to avoid the identifiability problem,

$$\hat{b}_x = \frac{\tilde{b}_x}{\sum_x \tilde{b}_x}, \qquad \qquad \hat{k}_t = \tilde{k}_t \sum_x \tilde{b}_x, \qquad (3.12)$$

where \hat{b}_x and \hat{k}_t are final estimates of b_x and k_t .

To forecast mortality, \hat{k}_t is assumed to be a random walk with drift:

$$\hat{k}_t = \hat{k}_{t-1} + \delta + \varepsilon_t , \qquad \delta = \frac{\hat{k}_n - \hat{k}_1}{n-1} , \qquad (3.13)$$

where δ is the average drift of the \hat{k}_t sequence.

The \hat{k}_1 and \hat{k}_n denote the first and last value in the sequence \hat{k}_t . The forecast of \hat{k}_t is

$$\hat{k}_{n+h} = \hat{k}_{n+h-1} + \delta + \varepsilon_h , \qquad (3.14)$$

where h denotes the h^{th} year in the forecasting period. The forecast of mortality is

$$\log(\hat{m}_{x,n+h}) = \log(\hat{m}_{xn}) + \hat{b}_x \left(\hat{k}_{n+h} - \hat{k}_n\right) , \qquad (3.15)$$

where $\log(\hat{m}_{xn})$ is the jump-off rate. The jump-off rate is defined as the mortality rate for the final year in the fitting period (Booth et al., 2002b; Lee and Miller, 2001).

$$\log(\hat{m}_{xn}) = \hat{a}_x + \hat{b}_x \hat{k}_n \,.$$

Advantages of the LC model in forecasting mortality

The LC model has lower uncertainty than the traditional method used by the SSA and the USCB. Moreover, minimizing subjective judgement is another advantage of the LC model (Lee and Carter, 1992; Lee, 2000). The traditional method relies on experts' views and is conservative (Booth and Tickle, 2008), while the LC model relies on data and is more objective in projection than the traditional methods.

In the application to the USA (Lee and Carter, 1992) for ages 0-109 (grouped as 0, 1-4, 5-9, ... 105-109) and for years 1900-1989, and forecast mortality for years 1990-2065, the LC model yields a narrower confidence interval (CI) than the traditional method. By 2065, the LC model yields a CI (95%) of 6.8 years for the forecast of life expectancy, whereas the traditional method produces a CI of 9.3 years. The k_t sequence is assumed to be linear, and the projection of k_t produces a narrower CI of life expectancy than the traditional method. By year 2065, the LC model forecasts life expectancy of 86.05 years, higher than the 80.45 years predicted by the traditional method.

Limitations of the LC model in forecasting mortality

The forecast of k_t causes increasing difference in mortality between populations. Let k_{ti} and k_{tj} denote the time factors of populations i and j and $k_{ti} > k_{tj}$ for t < s, where 1 < s < n. Due to the constraints of the time factors, $\sum_t k_{ti} = \sum_t k_{tj} = 0$, and thus $k_{ti} < k_{tj}$ for t > s. Therefore, the k_{ti} and k_{tj} sequences have different slopes and cause increasing differences between i and j in the long-term projection, referred to as divergence. Moreover, the random walk with drift may not be suitable for the k_t sequence (Jarner et al., 2008).

The assumption of constant b_x over time is unrealistic when forecasting mortality. In practice \hat{b}_x varies in different periods, indicating that an age group has a different rate of mortality decline over time (Lee, 2000).

The \hat{b}_x sequence is normally jagged and may include irregular values. The jagged \hat{b}_x will result in jagged forecast mortality across age, and these differences across ages will

increase over time (Czado et al., 2005; De Jong and Tickle, 2006; Delwarde et al., 2007).

Additionally, the LC model does not incorporate cohort factors. For populations where cohort effects exist, the LC model may yield poor accuracy in estimation and projection (Renshaw and Haberman, 2006).

3.2.4 Improvements to the LC model

Due to the advantages over the traditional methods, the LC model has been widely applied, including analyse and forecast the mortality of Australia (Booth et al., 2002a), Austria (Carter and Prskawetz, 2000), Canada and England and Wales (Chan et al., 2008), Denmark (Jarner et al., 2008), China and Korea (Li et al., 2004), the G7 countries (Tuljapurkar et al., 2000), England and Wales (Renshaw and Haberman, 2003a), and so on.

However, the LC model has shortcomings in applications. To address these shortcomings, methods have been proposed to improve the LC model as described in this subsection. The improvements of the LC model can be classified as three types: applying methods rather than the SVD to estimate parameters (Booth et al., 2002b; Brouhns et al., 2002; Lee and Miller, 2001), introducing additional parameters into the LC model (Booth et al., 2002b; Girosi and King, 2007; Koissi et al., 2006; Renshaw and Haberman, 2003c) and smoothing the LC model (Czado et al., 2005; De Jong and Tickle, 2006; Delwarde et al., 2007).

The improved methods criticize the LC model from different angles and provide different measures to address the relevant shortcomings. By briefly introducing the improved methods, this research considers the potential advantages and disadvantages, when developing and evaluating a new model.

Adjusting time trend incorporating life expectancy – the Lee-Miller model

Lee and Miller (2001) pointed out three shortcomings of the LC model as follows. Firstly, the forecast mortality and the observed mortality in the fitting period are not consistent. Secondly, the decrease of time trend and the increase of life expectancy are not consistent. Thirdly, there is large bias of the forecast life expectancy.

Three measures are consequently introduced to improve the LC model, referred to as the Lee-Miller (LM) model. Firstly, mortality data after 1950 as recommended by Tuljapurkar et al. (2000) are used for fitting and forecasting, where the b_x sequence is relatively constant and is reliable for projection (Lee, 2000). Secondly, k_t is adjusted by minimizing errors between the observed and fitted life expectancy. Thirdly, the observed log death rates, $\log(m_{xn})$, are used as the jump-off rates.

Finding a linear time trend and adjusting the time trend incorporating number of deaths – the Booth-Maindonald-Smith model

Booth et al. (2002b) pointed out two shortcomings of the LC model as follows. Firstly, the time trend may not be linear, and thus the forecast is unreliable. Secondly, the errors in the log death rate are assumed to be constant, whereas the errors relate to the population size.

Two measures are then applied to improve the LC model, referred to as the Booth-Maindonald-Smith (BMS) model. Firstly, k_t is adjusted by minimizing the errors between the observed and estimated age pattern of deaths to avoid irregular noise of small size age groups. Secondly, a subset of the total time range in the data set is selected to use as the fitting period, based on optimising the linear fit to k_t .

The \hat{a}_x and \hat{b}_x are obtained using processes introduced in Section 3.2.3. Poisson re-

gression (Brillinger, 1986) is applied to adjust \hat{k}_t by minimizing \hat{D}_t :

$$\hat{D}_t = \sum_x \left\{ d_{xt} \log \left(\frac{d_{xt}}{\hat{d}_{xt}} \right) - \left(d_{xt} - \hat{d}_{xt} \right) \right\} , \qquad (3.16)$$

where \hat{D}_t is the residuals in the number of deaths across x and d_{xt} and \hat{d}_{xt} denote the observed and estimated number of deaths, where e_{xt} is the observed mid-year population

$$\hat{d}_{xt} = e_{xt} \exp\left(\hat{a}_x + \hat{b}_x \hat{k}_t\right) .$$
(3.17)

A ratio R(s) is introduced to assess the linearity of the k_t sequence, where $1 \le s < n$;

$$R(s) = \frac{\sum_{t=s}^{n} \left(\tilde{D}_t / r_1 \right)}{\sum_{t=s}^{n} \left(\hat{D}_t / r_2 \right)} ,$$

where \hat{D}_t is defined in (3.16), and \tilde{D}_t is also calculated via (3.16), but replacing \hat{k}_t by \tilde{k}_t , where the \tilde{k}_t sequence is the fitted straight line of the \hat{k}_t sequence. The r_1 and r_2 denote the degrees of freedom of \hat{D}_t and \tilde{D}_t respectively, where $r_1 = (m-1)(n-s-1)$ and $r_2 = m(n-s-1)$. Trying s from 1 to n-1, the s resulting in the smallest R(s) is selected.

Estimating parameters using Poisson regression - the Brouhns-Denuit-Vermunt model

Brouhns et al. (2002) argued that the LC model did not consider the population's size when estimating the parameters. A Poisson log-bilinear regression approach (Brillinger, 1986) was then introduced to estimate all parameters of the LC model, referred to as the Brouhns-Denuit-Vermunt (BDV) model, as where the number of deaths is assumed to follow a Poisson distribution, the estimates of parameters could be robust. The maximum log-likelihood estimation (MLE) is:

$$L(a, b, k) = \sum_{x, t} \left\{ d_{xt}(a_x + b_x k_t) - e_{xt} \exp(a_x + b_x k_t) \right\} + C , \qquad (3.18)$$

where L denotes the log-likelihood about a, b and k, and C is a constant. The Newton-Raphson iterative process is used as follows:

$$\begin{split} \hat{a}_{x}^{(0)} &= 0, \qquad \hat{b}_{x}^{(0)} = 0, \qquad \hat{k}_{t}^{(0)} = 0, \\ \hat{a}_{x}^{(v+1)} &= \hat{a}_{x}^{(v)} - \frac{\sum_{t} \left(d_{xt} - \hat{d}_{xt}^{(v)} \right)}{-\sum_{t} \hat{d}_{xt}^{(v)}}, \qquad \hat{b}_{x}^{(v+1)} = \hat{b}_{x}^{(v)}, \qquad \hat{k}_{t}^{(v+1)} = \hat{k}_{t}^{(v)} \\ \hat{k}_{t}^{(v+2)} &= \hat{k}_{t}^{(v+1)} - \frac{\sum_{x} \left(d_{xt} - \hat{d}_{xt}^{(v+1)} \right) \hat{b}_{x}^{(v+1)}}{-\sum_{x} \hat{d}_{xt}^{(v+1)} \left(\hat{b}_{x}^{(v+1)} \right)^{2}}, \qquad \hat{a}_{x}^{(v+2)} = \hat{a}_{x}^{(v+1)}, \qquad \hat{b}_{x}^{(v+2)} = \hat{b}_{x}^{(v+1)} \\ \hat{b}_{x}^{(v+3)} &= \hat{b}_{x}^{(v+2)} - \frac{\sum_{t} \left(d_{xt} - \hat{d}_{xt}^{(v+2)} \right) \hat{k}_{t}^{(v+2)}}{-\sum_{t} \hat{d}_{xt}^{(v+2)} \left(\hat{k}_{t}^{(v+2)} \right)^{2}}, \qquad \hat{a}_{x}^{(v+3)} = \hat{a}_{x}^{(v+2)}, \qquad \hat{k}_{t}^{(v+3)} = \hat{k}_{t}^{(v+2)} \end{split}$$

where \hat{d}_{xt} is defined in (3.17) and v denotes the v^{th} step in the iterative process. The iteration stops at the step where the difference between $L^{(v-1)}$ and $L^{(v)}$ is very small, for example, 10^{-10} , and

$$L^{(v)} = L^{(v)}\left(\hat{\theta}^{(v)}\right) , \qquad \hat{\theta}^{(v+1)} = \hat{\theta}^{(v)} - \frac{\partial L^{(v)}/\partial\theta}{\partial^2 L^{(v)}/\partial\theta^2} , \qquad (3.19)$$

where θ denotes the parameters a, b or k.

Age-specific enhancement with additional parameters

The LC model involves a pair of b_x and k_t . More pairs of b_x and k_t are introduced to generalise the LC model (Booth et al., 2002b; Girosi and King, 2007; Koissi et al., 2006; Renshaw and Haberman, 2003c). These generalizations of the LC model use additional

parameters to enhance the age-specific effects, referred to as the age-specific enhancement (ASE) model. The ASE model is

$$\log(m_{xt}) = a_x + \sum_{j=1}^{J} b_x^{(j)} k_t^{(j)} + \epsilon_{xt} , \qquad (3.20)$$

where the *j* denotes the *j*th order of b_x and k_t . If J = 1 or $b_x^{(j)}k_t^{(j)} = 0$ for j > 1, the ASE model becomes the LC model. Generally, the fact that J = 2 enables the ASE model to account for 90% of the variance for the fitted data (Renshaw and Haberman, 2003c). The estimation of ASE can be the SVD method applied by Koissi et al. (2006) or the Poisson regression applied by Renshaw and Haberman (2003c).

Due the additional parameters, the ASE model yields better fit than the LC model (Booth et al., 2002b; Girosi and King, 2007; Koissi et al., 2006; Renshaw and Haberman, 2003c). However, the $k_t^{(j)}$ sequence is normally irregular (Renshaw and Haberman, 2003c), and thus may be difficult to forecast.

The smooth Lee-Carter model

The LC model has jagged sequences of age factors and may yield jagged forecasts (Czado et al., 2005; De Jong and Tickle, 2006; Delwarde et al., 2007). To address this shortcoming, approaches have been formulated to create a "smooth" LC model. A smooth LC model has at least one smooth factor. For example, due to the jagged shape, the b_x sequence is smoothed by applying a Bayesian Poisson log-bilinear approach (Czado et al., 2005) or by applying a penalised log-likelihood approach (Delwarde et al., 2007).

Among the smooth LC models, the model proposed by De Jong and Tickle (2006), referred to as the De Jong-Tickle (DT) model, is widely evaluated (Booth et al., 2006; Shang et al., 2010). The DT model involves smooth a_x and b_x by using a "design" matrix. Let y_{xt} denote the log death rate $log(m_{xt})$, then the LC model is written as:

$$y_t = a + bk_t + \epsilon_t , \qquad (3.21)$$

where y_t , a, b and ϵ_t are column vectors with entries y_{xt} , a_x , b_x and ϵ_{xt} , respectively. The k_t is defined the same as in the LC model.

Let X be a "design" matrix with more rows than columns, then the DT model is as follows:

$$y_t = Xa + (Xb) \times k_t + \epsilon_t . \tag{3.22}$$

Since X has fewer columns compared to rows, the DT model has fewer age parameters than the LC model. The effect of X is to make Xa and Xb relatively smooth. The smoothness enables the DT model to avoid the non smooth estimation of age effects and to reduce the risk in projection. Therefore, the DT model and other smooth LC models yield more reasonable fits and more reliable forecasts as compared to the LC model (Czado et al., 2005; De Jong and Tickle, 2006; Delwarde et al., 2007).

3.2.5 Age-period-specific model – the Lee-Carter model using functional data analysis

The HU model

Hyndman et al. (2007) extended the LC model by incorporating two features: age-specific enhancement (like the ASE model) and smoothness (like the DT model), referred to as the Hyndman-Ullah (HU) model. The HU model uses a functional data analysis (Ramsay and Silverman, 2005) to estimate parameters and aims to provide robust statistics to fit and forecast mortality.

The HU model assumes that the log death rate is a smooth and continuous function, and then the HU model is

$$\log(m_{xt}) = a(x) + \sum_{j=1}^{J} b^{(j)}(x)k_t^{(j)} + \epsilon_t(x) + \sigma_t(x)\varepsilon_{xt} , \qquad (3.23)$$

where a(x) and $b^{(j)}(x)$ are smooth and continuous functions of a_x and $b_x^{(j)}$ and $J \le n-1$. The error items $\sigma_t(x)\varepsilon_{xt}$ and $\sigma_t(x)$ are defined as follows:

$$\log(m_{xt}) = f_t(x) + \sigma_t(x)\varepsilon_{xt} , \qquad f_t(x) = \log(\hat{m}_{xt}) + \epsilon_t(x) ,$$

where $f_t(x)$ is the smooth function of $\log m_{xt}$ of x, estimated with penalised regression splines (Wood, 2000). The $\log(\hat{m}_{xt})$ is the fitted log death rate using (3.23). The ε_{xt} is an independent and identically distributed (i.i.d.) noise and $\sigma_t(x)$ is the noise related to x.

Estimation of parameters in the HU model

The a(x) is the average of the smooth function, $f_t(x)$, across t:

$$\hat{a}(x) = \frac{1}{n} \sum_{t=1}^{n} f_t(x) \; .$$

Let $f_t^*(x)$ denote the difference between $f_t(x)$ and $\hat{a}(x)$,

$$f_t^*(x) = f_t(x) - \hat{a}(x) = \sum_{r=1}^R \alpha_{ts}(x)\xi_r(x) ,$$

where $\alpha_{tr}(x)$ denotes the entries of a matrix A with dimensions $n \times R$ and $\xi_r(x)$ relates to the $R \times R$ matrix V with entries $v_{\mathcal{I}\mathcal{J}} = \int \xi_{\mathcal{I}}(x)\xi_{\mathcal{J}}(x)$, where $1 \leq \mathcal{I}, \mathcal{J} \leq R$. Let V = U'U, applying the Cholesky decomposition to V, there is

$$\hat{b}^{(j)}(x) = \left(U^{-1}g^{(j)}\right)'\xi(x) ,$$

where $\xi(x) = \{\xi_1(x), ..., \xi_r(x)\}'$ and $g^{(j)}$ is the j^{th} normalized eigenvector of $(U^{-1})'VSV'U^{-1}$, and where $S = (n-1)^{-1}A'W^2A$, W is a diagonal matrix of dimensions $n \times n$ with entries w_t , the weights of populations' size in year t. Constraints of $b^{(j)}(x)$ are

$$\int \left\{ b^{(j)}(x) \right\}^2 d_x = 1 \qquad J = 1 ,$$
$$\int \left\{ b^{(j)}(x) b^{(j-1)}(x) \right\} d_x = 0 \qquad J \ge 2 .$$

Let F denote the $n \times m$ matrix with entries $f_t^*(x)$ and B denote the $m \times (n-1)$ matrix with entries $\hat{b}^{(j)}(x)$. There is K = FB, where K is a matrix with entries $\hat{k}_t^{(j)}$.

In the HU model, the order J is determined according to the forecast accuracy. Given that s is a break year in the fitting period and 1 < s < n, the HU model is applied to fit mortality data for the sub-fitting period 1, ..., s and then to forecast mortality in the sub-forecasting period s + 1, ..., n. The forecast accuracy is denoted as follows:

$$\epsilon_{s+h}(x) = \log(m_{x,s+h}) - \log(\hat{m}_{x,s+h}) ,$$

where h denote the h^{th} year in the sub-forecasting period, and $\log(m_{x,s+h})$ and $\log(\hat{m}_{x,s+h})$ denote the observed and forecast log death rate in year s + h. There are

$$SSE = \sum_{h=1}^{n-s} \sum_{x} \epsilon_{s+h}^2(x) , \qquad SSSE = \sum_{s=S}^{n-1} \sum_{h=1}^{n-s} \sum_{x} \epsilon_{s+h}^2(x) ,$$

where SSE is the sum of square errors of the forecast log death rate and SSSE is the sum of SSE when shifting s from a given year, denoted by S, until n - 1.

Trying a different J and obtaining a series of SSSE, the J resulting in the smallest SSSE is the most appropriate for the specific mortality data set. The process of finding the appropriate order is as follows:

- 1) Set initial order as J = 1;
- 2) Set initial *s*;
- 3) Calculate SSE;
- 4) Increase s by 1 and go to step 3) until s = n 1;
- 5) Sum these SSE as as SSSE and record this SSSE for the current order;
- 6) Increase order by 1 and go to step 2) until the given maximum order;
- 7) Obtain a series of SSSE for orders from J = 1 to the given maximum order, for example, J = n 1.

The J corresponding to the lowest SSSE is chosen.

Hyndman et al. (2007) produced a function called "fdm" to use the HU model to fit mortality and a function called "forecast.fdm" to forecast mortality. These two functions are included by the Demography package in R. This "forecast.fdm" function provides more than one method to forecast the time trends, including the random walk with drift and the ARIMA process.

Advantages of the HU model in estimating and forecasting mortality

The HU model uses the functional data analysis approach, where smoothness is applied. The smoothness reduces the inherent randomness of observations and avoids the unexpected patterns in mortality improvements caused by, for example, war, epidemic or catastrophes. Thus, the HU model may have be more robust estimation than the LC model. Moreover, the HU model may use more parameters than the LC model. The order J is determined by a series of trials, the process of determining J enables the HU model to be used for a flexible application. Due to the trials, the HU model captures the properties of a given mortality data set better than the LC model.

Hyndman et al. (2007) applied the HU model to fit and forecast the French mortality, and Booth et al. (2006) and Shang et al. (2010) applied the HU model to multiple countries. In these analyses, the HU model – similar to the DT model – performs better than the LC, LM and BMS models, particularly in forecast accuracy.

Limitations of the HU model in fitting and forecasting mortality

If J > 1, the HU model has more than one time trend. The additional time trends may increase the uncertainty and unreliability of projection as in the description of the ASE model. The selection of J increases the computation but may not result in significantly better performance compared with similar models, such as the LM, BMS and DT models (Booth et al., 2006; Shang et al., 2010).

3.2.6 Age-period-specific model – the two time factors model

The CBD models

The LC model is suitable for all age groups, but generally fits poorly or in any case worse for the old age groups (Cairns et al., 2009). To model the mortality rate of old age groups, Cairns et al. (2006) developed a stochastic model with two time factors, referred to as the Cairns-Blake-Dowd (CBD) model. In the CBD model, mortality rates of ages 60-89 vary linearly across ages and vary stochastically over time. The CBD model is extended to CBD1, CBD2 and CBD3 models (Cairns et al., 2009). These three extensions include cohort effects and are age-period-cohort-specific models. The CBD model and its extensions are called the CBD models.

The CBD model includes two time trends:

$$logit(q_{xt}) = k_t^{(1)} + k_t^{(2)}(x - \bar{x}) + \epsilon_{xt} , \qquad (3.24)$$

where \bar{x} denotes the average age of the age vector, and

$$\operatorname{logit}(q_{xt}) = \frac{q_{xt}}{1 - q_{xt}} , \qquad (3.25)$$

where $k_t^{(1)}$ reflects general mortality decline for all ages over time and $k_t^{(2)}$ reflects the rate of mortality decline for specific age groups.

The CBD1, CBD2 and CBD3 models (Cairns et al., 2009) are displayed as (3.26), (3.27) and (3.28):

$$logit(q_{xt}) = k_t^{(1)} + k_t^{(2)}(x - \bar{x}) + \gamma_c + \epsilon_{xt} , \qquad (3.26)$$

$$logit(q_{xt}) = k_t^{(1)} + k_t^{(2)}(x - \bar{x}) + k_t^{(3)} \left\{ x - \bar{x} \right\}^2 - \hat{\sigma}_x \left\{ + \gamma_c + \epsilon_{xt} \right\}, \quad (3.27)$$

$$logit(q_{xt}) = k_t^{(1)} + k_t^{(2)}(x - \bar{x}) + \gamma_c(x^* - x) + \epsilon_{xt} , \qquad (3.28)$$

where $\hat{\sigma}_x$ in (3.29) is the estimated variance of ages,

$$\hat{\sigma}_x = \frac{1}{m} \sum_{x=1}^m (x - \bar{x})^2 , \qquad (3.29)$$

and x^* in (3.28) is a constant age and needs to be estimated. The subscript c of the cohort factor is c = t - x. The range of c is from $t_1 - x_m$ to $t_n - x_1$, where t_1 and t_n , x_1 and x_m denote the exact year and ages in the vectors of years and ages. The length of the γ_c sequence is $(t_n - x_1) - (t_1 - x_m) + 1 = m + n - 1$. The CBD model is age-period-specific, and the CBD1, CBD2 and CBD3 models are age-period-cohort-specific.

Estimation of parameters in the CBD models

The observed mortality scale is

$$q_{xt} = 1 - p_{xt} = 1 - \exp\left(-\mu_{xt}\right) = 1 - \exp\left(-m_{xt}\right) , \qquad (3.30)$$

where μ_{xt} denotes the force of mortality and $\mu_{xt} = m_{xt}$ for the old age groups. Constraints are set as follows: $\sum_c \gamma_c = 0$ and $\sum_c c\gamma_c = 0$ for the CBD1 model, $\sum_c \gamma_c = 0$, $\sum_c c\gamma_c = 0$ 0 and $\sum_c c^2 \gamma_c = 0$ for the CBD2 model and $\sum_c \gamma_c = 0$ for the CBD3 model.

The Poisson regression with the Newton-Raphson iterative process (Brouhns et al., 2002) is applied to estimate the parameters in the CBD models.

Advantages of the CBD models in estimating mortality

Differing from the log death rate $(\log(m_{xt}))$ used by many other existing models, the CBD models use the logit death rate $(\log i(q_{xt}))$. By using the logit death rate, the CBD models assume linear relationships in the logit of mortality across age, particularly appropriate for the older age groups where linearity in log mortality is typically inappropriate. Linearity is convenient for modelling, and thus the CBD models have better fit for older age groups than many other existing models, such as the LC, Currie and RH models (Cairns et al., 2009, 2011b; Gogola, 2014). Due to the advantages in older age groups, the CBD models are widely used to deal with longevity problems (Cairns et al., 2006, 2014) and financial pricing (Cairns et al., 2014; Chan et al., 2014).

Limitations of the CBD models in estimating mortality

The CBD models have time and cohort trends. The multiple trends increase the uncertainty in long-term projection. The cohort trends of the CBD1, CBD2 and CBD3 models are irregular and may be unreliable in prediction. In addition, the CBD models are only for the older ages and perform poorly for general age groups (Plat, 2009a).

3.2.7 Age-period-cohort-specific models

Cohort effects aim to model the common experience in mortality for people born at the same point of time. Such common experience may be important and any model that ignores cohort effects, such as the LC model may yield poor fit and forecast. Since the LC model is one of the most widely used mortality models (Booth, 2006; Girosi and King, 2007; Tuljapurkar et al., 2000), cohort factors have been introduced into the LC framework yielding the age-period-cohort-specific (APC) model. The models proposed by Renshaw and Haberman (2006), Currie (2006) and Plat (2009a) are representatives of the APC models. APC models generally yield better fit than the LC model.

However, the APC models involve at least a time factor and a cohort factor. Cohort, time and age factors are not orthogonal and may generally lead to uncertainty in estimation and forecast. In addition, cohort trends are normally irregular and difficult to forecast (Booth, 2006), and hence, while there is a basis for cohort effects, the forecast mortality may be unreliable.

The Renshaw-Haberman model

Renshaw and Haberman (2006) extended the LC model by adding a cohort factor, referred

to as the Renshaw-Haberman (RH) model:

$$\log(m_{xt}) = a_x + b_x^{(1)} k_t + b_x^{(2)} \gamma_c + \epsilon_{xt} , \qquad (3.31)$$

where γ_c denotes the cohort effect, and c represents the year of birth: c = t - x and is detailed in Section 3.2.6.

Using cohort factors, the RH model avoids producing clusters of diagonal residuals when fitting mortality data (Cairns et al., 2009). In the analyses given by Cairns et al. (2009), Cairns et al. (2011b) and Gogola (2014), the RH model also performs well for older age groups according to the Bayesian information criterion (BIC) measure. However, the $\hat{\gamma}_c$ sequence of the RH model is not robust for forecasting (Cairns et al., 2011b).

The Currie model

Currie (2006) simplified the RH model, by setting the coefficients of k_t and γ_c as 1, referred to as the Currie model:

$$\log(m_{xt}) = a_x + k_t + \gamma_c + \epsilon_{xt} . \tag{3.32}$$

The Currie model inherits the shortcomings of the RH model. Moreover, the Currie model is less flexible than the RH model, due to the constant coefficients.

The Plat model

Plat (2009a) developed an age-period-cohort-specific model, by introducing a time factor to describe the mortality patterns for the young age groups and absorbing the "good" features from the LC, RH (or Currie) and CBD models. This combined age-period-cohortspecific model is referred to as the Plat model:

$$\log(m_{xt}) = a_x + k_t^{(1)} + (\bar{x} - x)k_t^{(2)} + (\bar{x} - x)^+ k_t^{(3)} + \gamma_c + \epsilon_{xt} , \qquad (3.33)$$

where the a_x , $k_t^{(1)}$ and γ_c come from the Currie model, $(\bar{x} - x)k_t^{(2)}$ comes from the CBD model, and $(\bar{x} - x)^+$ was introduced by Plat (2009a). The $(\bar{x} - x)^+$ ensures that $k_t^{(3)}$ describes the overall mortality varying trends for the younger age groups in the data set, where $(\bar{x} - x)^+ = \max(\bar{x} - x, 0)$.

The Plat model uses three time factors to reflect different information in mortality. The $k_t^{(1)}$ reflects the general trend in mortality decline. The $k_t^{(2)}$ reflects the relationships between ages and the mortality decline. The $k_t^{(3)}$ captures particular trends in young age groups. Young age groups have a higher risk of mortality from external factors, such as drugs, AIDS, alcohol abuse and violence. Since the dynamics of mortality rates at younger ages (below 50) are significantly different than those at older ages, if $k_t^{(3)}$ is removed, the Plat model is for old age groups only.

The Plat model performs well for whole age ranges in the analysis given by Plat (2009a) according to BIC. However, by using a large number of parameters, the Plat model may yield large uncertainty of projection. Furthermore, $\hat{k}_t^{(2)}$, $\hat{k}_t^{(3)}$ and $\hat{\gamma}_c$ are irregular and may be difficult to predict. Then, the forecast mortality may not be reliable.

3.3 Joint models

With globalization, countries are becoming increasingly similar in terms of lifestyle, culture and socio-economic conditions (Li and Lee, 2005), and thus experience increasingly similar mortality improvements (Mathers et al., 2001; Moser et al., 2005). Given the context of a worldwide convergence of mortality, combined mortality projection of different countries appears more cogent than treating each country separately and in isolation.

Joint mortality models aim to pool mortality information and trends across populations so as to provide a more robust and sound analysis compared to that provided by separate analyses based on segmented data. Joint models can describe both similarities and differences between populations. Incorporating or pooling across similar populations with joint models aims to achieve four goals – reducing divergence in forecasting, improving reliability in forecasting for small populations, analysing relationships between populations and increasing goodness of fit. In addition, incorporating more information, joint models – although not comprehensively assessed but illustrated by this research as detailed in Chapter 6 – may yield better forecast accuracy than individual models.

3.3.1 Joint Lee-Carter model

As described previously, the LC model and other similar individual models normally produce increasing differences in mortalities between populations in long-term projection. The projection of individual models is inconsistent with observations that mortalities in similar populations do not indefinitely diverge, whereas mortalities converge worldwide convergence in mortality as described in Section 2.3.

To address this shortcoming, Li and Lee (2005) extended the LC model to a joint form by incorporating common factors, referred to as the joint LC (JLC) model. The JLC model captures the common properties across similar populations. The common items consist of the common time trend and the common age-specific response to the common time trend. The JLC model stabilises the differences in mortality across populations in long-term projection.

The JLC model

Incorporating *i* as a population index, the LC model can be written as follows:

$$\log(m_{xti}) = a_{xi} + b_{xi}k_{ti} + \epsilon_{xti} , \qquad (3.34)$$

where $i = 1, ... \ell$ and a_{xi}, b_{xi} and k_{ti} are defined the same as a_x, b_x and k_t in the LC model with the *i* index stressing the fact that, initially, all the coefficients are distinct for different *i*.

A simplified joint LC (SJLC) model is

$$\log(m_{xti}) = a_{xi} + b_x k_t + \epsilon_{xti} ,$$

where b_x and k_t denote the common time trend and common age-specific coefficient response to the common time trend. The SJLC model assumes that populations have individual "base" levels (a_{xi}) but share a common trend b_x and k_t in mortality decline.

A general joint LC model, referred to as the JLC model, is

$$\log(m_{xti}) = a_{xi} + b_{xi}k_{ti} + b_xk_t + \epsilon_{xti} , \qquad (3.35)$$

where b_{xi} and k_{ti} denote the individual factors as defined in (3.34). The JLC model allows populations to possess both individual and common trends to model mortality declines.

Estimation of parameters in the JLC model

The common trend is dominant in the mortality decline across populations (Li and Lee, 2005), and thus common factors are estimated prior to individual factors. Therefore, step

1 estimates b_x and k_t . Step 2 estimates a_{xi} , b_{xi} and k_{ti} .

Estimation of common factors Applying the LC model to the aggregated population:

$$\log(m_{xt}) = \hat{a}_x + \hat{b}_x \hat{k}_t + \epsilon_{xti} , \qquad (3.36)$$

where

$$m_{xt} = \frac{\sum_{i=1}^{p} d_{xti}}{\sum_{i=1}^{p} e_{xti}} \qquad \qquad \hat{a}_x = \frac{\sum_{t=1}^{n} \log(m_{xt})}{n} , \qquad (3.37)$$

where d_{xti} and e_{xti} are defined in (3.16) and (3.17), and \hat{b}_x and \hat{k}_t are estimated using SVD as described in (3.11).

Estimation of individual factors The \hat{a}_{xi} is estimated the same as (3.9),

$$\hat{a}_{xi} = \frac{\sum_{t=1}^{n} \log(m_{xti})}{n} , \qquad (3.38)$$

Let Y_i denote a matrix with entries y_{xti} :

$$y_{xti} = \log(m_{xti}) - \hat{a}_{xi} - \hat{b}_x \hat{k}_t = b_{xi} k_{ti} + \epsilon_{xti} .$$
(3.39)

Then, (3.11) and (3.10) are applied to Y_i to obtain \hat{b}_{xi} and \hat{k}_{ti} .

Advantages of the JLC model in forecasting mortality

The JLC and LC models are evaluated by applying them to two groups of populations. Each group consists of multiple countries. The JLC model is applied to jointly fit and forecast mortality across countries in each group. Two measures are applied to evaluate the models' forecasting performance: the standard deviation of forecast life expectancy across countries and the average 95% confidence interval (CI) of the forecast life expectancy across countries.

Group 1 (the low mortality group) contains Austria, Canada, Denmark, England, Finland, France, Germany (West), Italy, Japan, the Netherlands, Norway, Spain, Sweden, Switzerland and the USA. In this analysis, the JLC model yields a smaller standard deviation and a smaller average CI of the forecast life expectancy across countries than the LC model.

Group 2 (the high mortality group) includes Bulgaria, Czech Republic, Germany (East), Hungary, Lithuania and Russia. In this analysis, the JLC model produces a smaller standard deviation but larger average CI of the forecast life expectancy across countries.

For low mortality countries or developed countries, mortalities converge across countries. The JLC model captures the convergence and yields higher certainty than the LC model. For high mortality countries or developing countries, mortalities diverge across countries as described in Section 2.3. The JLC model is not appropriate.

Since the AR(1) process requires k_{ti} to converge at zero, in the long term, member populations experience the same mortality decline $b_x k_t$. In the long term, the JLC model produces constant difference in the log death rate across populations, referred to as nondivergence. The SJLC model excludes individual time trends, and thus yields a nondivergent forecast log death rate in both the short term and the long term.

Limitations of the JLC model in forecasting

The JLC model uses a common time factor (k_t) and an individual time factor (k_{ti}) to describe the mortality trends of each population. The k_{ti} sequence is modelled as a AR(1) process. In the short term, the multiple trends may increase the forecast uncertainty.

Li and Lee (2005) introduced a ratio of similarity to determine whether populations are similar enough to jointly fit and forecast, but the criterion is determined subjectively. The JLC inherits some shortcomings of the LC model – the assumption of constant b_x and b_{xi} , lack of smoothness in forecast rates, and the lack of cohort effects.

Improving the estimation of the JLC model using Poisson regression

Li (2013) applied Poisson regression (Brouhns et al., 2002) to estimate the common factors of the JLC model, referred to as the Poisson JLC (PJLC) model. The individual factors are estimated by using the SVD method. By considering the aggregating population's size, the PJLC model enhances the effects of the whole group.

The PJLC model performs more reliably than the LC model when forecasting mortality in the long term. The PJLC model yields relatively constant differences – the maleto-female death ratio – between Australian males and females. However, the LC model yields an increasing male-to-female death ratio over time. Furthermore, the PJLC model produces a smaller difference of forecast life expectancy between females and males.

However, the comparison is only between the PJLC and LC models, disregarding the original JLC model proposed by Li and Lee (2005). Hence, it is unknown whether and how much the Poisson regression improves the JLC model. Furthermore, the forecast accuracy was not examined. In addition, the PJLC model is an enhancement of the JLC model, and thus contains the shortcomings of the JLC model.

3.3.2 Joint Lee-Carter model with three-way decomposition

The LC model applies to two-dimensional data sets – ages and years. To incorporate an extra dimension – for example, the population effect – Russolillo et al. (2011) extended

the LC model by introducing a three-way approach, referred to as the TLC model. The TLC model incorporates factors of ages, years and populations and thus describes the relationships in mortality between populations.

The TLC model

The TLC model is

$$\log(m_{xti}) = a_{xi} + b_x k_t \rho_i + \epsilon_{xti} , \qquad (3.40)$$

where ρ_i denotes the population specific factor of *i*. The a_{xi} , b_x and k_t are defined in (3.3.1). The TLC model allows populations to have the age-specific "base" level, a_{xi} , but share the common factors, b_x and k_t , and b_x and k_t are affected by ρ_i .

Estimation of parameters in the TLC model

The a_{xi} is estimated using (3.38). The item $b_x k_t \rho_i$ is three-dimensional and Russolillo et al. (2011) introduced a three-way SVD approach – the Tucker approach (Tucker, 1966). Let Y denote a cube with dimensions $m \times n \times \ell$ and entries y_{xti} , where

$$y_{xti} = \log(m_{xti}) - \hat{a}_{xi} = b_x k_t \rho_i + \epsilon_{xti} , \qquad (3.41)$$

In the Tucker approach, y_{xti} can be written as follows:

$$y_{xti} = \sum_{f=1}^{F} \sum_{g=1}^{G} \sum_{j=1}^{J} b_{xf} k_{tg} p_{ij} \psi_{fgj} , \qquad (3.42)$$

where the new factor ψ_{fgj} is the entry of the "designed" cube Ψ . The Tucker approach rearranges the Y to a matrix \mathbb{Y} with dimensions $m \times (n\ell)$,

$$\mathbb{Y}_{[m \times (n\ell)]} = \mathbb{B}_{[m \times F]} \Psi_{[F \times G \times J]} (\mathbb{P}'_{[J \times \ell]} \otimes \mathbb{K}'_{[G \times n]}) , \qquad (3.43)$$

where \mathbb{B} , \mathbb{K} and \mathbb{P} are a matrix with entries b_{xf} , k_{tg} and p_{ij} . The b_x , k_t and ρ_i in (3.40) denote the x^{th} , t^{th} and i^{th} row of the matrix \mathbb{B} , \mathbb{K} and \mathbb{P} . The core array, Ψ , captures the relationships between the F components of mode \mathbb{B} , the G components of mode \mathbb{K} and the J components of mode \mathbb{P} . The \mathbb{P}' and \mathbb{K}' denote the transposition matrix of \mathbb{P} and \mathbb{K} . The dimensions of Ψ are determined via trials of experiments (Kiers and Kinderen, 2003). The notation \otimes denotes the Kroneker product.

Advantages of the TLC model in analysing relationships between mortalities

Russolillo et al. (2011) applied the TLC model to 10 European countries for the ages 0, 1-4, 5-9, ... 95-99 and the years 1950-2000 with combined genders. The countries were Austria, Belgium, Denmark, Finland, France, Italy, the Netherlands, Spain, Sweden and the United Kingdom (UK). The TLC model illustrates a high similarity of mortality trends in mortality across the countries, with a correlation coefficient of 0.9876. By applying the Tucker approach, the TLC model is able to analyse the relationship in mortality among countries. Mortality of ages 20-40 and 45-94 is similar for all countries, Belgium, Finland, France, Italy and the UK experience overall similarity in mortality, while the Netherlands, Denmark and Spain differ from the other member countries.

Limitations of the TLC model in forecasting mortality

The TLC model may have more than one time factor. The additional time trends are normally irregular and difficult to forecast as can be seen in Russolillo et al. (2011, Figure 3). Secondly, the TLC model may cause increasing differences in mortality between populations. Since $b_x k_t \rho_i \approx b_x k_{ti}$, the k_{ti} sequences result in different mortality decline of individual populations, and then lead to increasing differences in mortality between populations in long-term projection as described in Section 3.2.3.

3.3.3 Joint Lee-Carter model with parallels

The PLC model

Debón et al. (2011) extended the LC model to a joint form. In this joint LC model, geographically close member populations share the same basic pattern of mortality which can be expressed by the LC model. For given populations, the differences between member populations is constant for all age groups. Therefore, curves of age-specific mortality are parallel across populations. This joint LC model is referred to as the PLC model. Contrary to the TLC model, the population effect factor in the PLC model is "added" not "multiplied". The "parallel" property is less flexible for general application.

The PLC model is

$$logit(q_{xti}) = a_x + b_x k_t + \rho_i + \epsilon_{xti} , \qquad (3.44)$$

where $logit(q_{xti})$ is defined in (3.25) and other symbols are defined the same as in (3.40). For a give age group, the curves of age-specific mortality are parallel between given populations. Populations share the same patterns of mortality, but with different shifts.

Estimation of parameters in the PLC model

Debón et al. (2011) applied a generalised nonlinear model (GNM) to estimate parameters in the PLC model. Under the GNM, the constraints described in (3.10) are not applied to b_x and k_t . The initial values of b_x and k_t are set as 1 and 0, respectively. The application of the GNM method to the LC model was detailed by Turner and Firth (2012). During the estimation, the weights of each age group and populations are incorporated. The whole group is viewed as an independent group.

Advantages of the PLC model in estimating mortality

The PLC model yields the best fit in comparison with the JLC and TLC models, and with a model proposed by Brass et al. (1971), when applying to fit mortality data sets – 17 of the 19 Spanish regions for the ages 0-97 and the years 1980-2006. The PLC model generates the smoothest b_x and k_t and the lowest overall errors in the number of deaths for both females and males. Errors generated by the PLC model were distributed normally and centralized around 0 across ages, time and populations.

Furthermore, by using a unique index – the population effect, ρ_i , the PLC models simplifies the differences in mortality across populations. The PLC model (3.44) can written as

$$logit(q_{xti}) = (a_x + \rho_i) + b_x k_t + \epsilon_{xti} \approx a_{xi} + b_x k_t + \epsilon_{xti} , \qquad (3.45)$$

the PLC model is similar to the SJLC model (3.3.1). If applying (3.45) to forecast mortality, populations possess individual initial difference a_{xi} , but share the same trend in projection. Therefore, (3.45) yields constant difference in logit death rate over time, referred to as non-divergence.
Limitations of the PLC model in estimating and forecasting mortality

The assumption of "parallel" of age-specific logit death rate across countries may not be appropriate for a general application. For geographically close populations, for example, where the regions in a country share similar economic, cultural and environmental conditions and experience improvements in mortality, this assumption could be suitable. However, for worldwide populations, for a given age group, the age-specific mortality is not parallel across countries, such as the USA and Belgium (Lee, 2000), Japan and the Netherlands (Biatat and Currie, 2010). Therefore, the assumption of "parallel" may be reliable for forecasts with non-divergent demands, but may be inappropriate for fits.

3.3.4 Joint age-period-cohort model

The CBDCK model

Cairns et al. (2011a) extended the Currie model to jointly forecast mortality between two populations, referred to as the Cairns-Blake-Dowd-Coughlan-Khalaf-Allah (CBDCK) model. The CBDCK model is separately applied to a large population and to a small population. The parameters estimated for the small population are adjusted using parameters estimated for the large population via a Bayesian framework. The CBDCK model improves the forecast reliability of the small population, by referencing a similar large population.

Cairns et al. (2011a) changed the Currie model (3.32) as:

$$\log(m_{xti}) = a_{xi} + \frac{1}{N_x} k_{ti} + \frac{1}{N_x} \gamma_{ci} + \epsilon_{xti} , \qquad (3.46)$$

where N_y is constant and denotes the length (in years) of the fitting period.

The CBDCK model is designed to jointly forecast mortality between two similar pop-

ulations. These two populations are denoted as Population 1 (the large population) and population 2 (the small population). For instance, Population 2 (Sydney) can be one of the subsets of Population 1 (Australia). With a small size, Population 2 may have unreliable estimation of parameters and may result in unreliable projection of mortality. Therefore, Population 2 relies on Population 1 to forecast mortality.

The CBDCK model assumes that Population 2 possesses its own mortality decline trend in the short-term, but shares the common trend with Population 1 in the long-term. According to this assumption, the CBDCK model applies a Bayesian method to adjust the projection of Population 2, using the parameters estimated for Population 1.

Estimation of parameters in the CBDCK model

Estimates of a_{xi} , k_{ti} and γ_{ci} are introduced in Section 3.2.7. The Bayesian framework is used to adjust the projection of k_{t2} and γ_{c2} , using k_{t1} and γ_{c1} as follows:

$$R_k(t) = k_{t1}, \ S_k(t) = k_{t1} - k_{t2}, \ R_{\gamma}(c) = \gamma_{c1}, \ S_{\gamma}(c) = \gamma_{c1} - \gamma_{c2},$$

where $R_k(t)$ is modelled as a random walk and $R_k(t_1) = 0$. The $S_k(t)$ sequence follows an AR(1) time series with mean reversion 0. The innovations for $R_k(t)$ and $S_k(t)$ are i.i.d bivariate normal from one to the next year with non-zero correlation. The series $R_{\gamma}(c)$ is an AR(2) process around a deterministic linear trend around 0. The $S_{\gamma}(c)$ sequence is an AR(2) process with mean reversion 0. The innovation of $R_{\gamma}(t)$ and $S_{\gamma}(t)$ are i.i.d from one year to the next year with small correlation. The Markov chain Monte Carlo (MCMC) method is applied to produce a Bayesian posterior distribution of parameters in the CBDCK model. The MCMC method deals with the missing data using the posterior distribution estimation and incorporating uncertainty.

Advantages of the CBDCK model in forecasting mortality

Compared with the Currie model, the CBDCK model increases the describability for a small population. With the application to England and Wales (the large population) and continuous mortality investigation (CMI) (the small population) males for the ages 60-89 and the years 1961-2005, with forecasts for 2006 - 2050, the CBDCK model produces less fluctuation in mortality forecasts for the CMI males than the Currie model. Since the Bayesian method reduced the uncertainty of projection for the CMI data, the CBDCK model yielded more smooth time trends and cohort trends than the Currie model. The CBDCK model provides consistent central projections for the CMI data, with a smaller width of the fans of projections for the small population, compared with the Currie model.

Limitations of the CBDCK model in forecasting mortality

The CBDCK model inherits the shortcomings of the Currie model. The cohort trend is irregular and is difficult to forecast. The $\frac{1}{N_x}$ is constant, and thus causes inflexible application and may result in an unreliable long-term projection. The Currie model may be unreliable for projection, and thus the CBDCK model may also be unreliable for projection. Furthermore, the CBDCK model was designed for a two-population – a large population and a small population – projection. If these two populations have similar sizes or there are more than two populations, the CBDCK model may not be suitable. The CBDCK model is for old age groups and may not be appropriate for general age groups. Hence, the CBDCK model is limited for a general application.

3.3.5 Other joint mortality models

Villegas and Haberman (2014) proposed a two-step model to model the differences of mortality across subpopulations, referred to as the VH model. The VH model is:

$$\log(m_{xt}) = a_x + b_x k_t + \gamma_c + \epsilon_{xt} , \qquad (3.47)$$

$$\log(m_{xti}) = \log(m_{xt}) + a_{xi} + b_{xi}k_{ti} + \epsilon_{xti} , \qquad (3.48)$$

where (3.47) is a simplification of the APC model, (3.31), and describes the mortality of the whole population. Equation (3.48) is an extension of the LC model, (3.8), and models the mortality of the subpopulation *i* in the whole population. The VH model assumes that the subpopulations share the same cohort trend, γ_c , of the whole population but experience individual time trend, k_{ti} and the age-specific factors, a_{xi} and b_{xi} . The VH model inherits the disadvantages of cohort factors as described in Section 3.2.7, where (3.47) may cause unreliable projection. Furthermore, the VH model may not be appropriate for modelling mortality across countries (Villegas and Haberman, 2014).

Hatzopoulos and Haberman (2013) introduce a ratio to model the relationship of mortality between females and males within a country and jointly model the female and male mortality:

$$\log(\mu_{xt}) = \alpha_x + \sum_{j=1}^J \beta_x^{(j)} \kappa_t^{(j)} + \xi_{xt} , \qquad (3.49)$$

$$\log(m_{xt1}) = a_{x1} + \sum_{j=1}^{J} b_{x1}^{(j)} k_{t1}^{(j)} + \sum_{j=1}^{J} b_{x}^{(j)} k_{t}^{(j)} + \epsilon_{xt1} , \qquad (3.50)$$

$$\log(m_{xt2}) = a_{x2} + \sum_{j=1}^{J} b_{x2}^{(j)} k_{t2}^{(j)} + \sum_{j=1}^{J} b_{x}^{(j)} k_{t}^{(j)} + \epsilon_{xt2} + \sum_{j=1}^{J} \beta_{x}^{(j)} \kappa_{t}^{(j)} , \quad (3.51)$$

where the three models constitute the HH model. Equation (3.49) models the ratio of central death rate between two populations – for example, females and males, and

$$\log(\mu_{xt}) = \log\left(\frac{\mu_{xt1}}{\mu_{xt2}}\right)$$
.

Equations (3.50) and (3.51) are extended from the HU model, (3.23), and the JLC model, (3.35) and model the mortality of populations 1 and 2, respectively. Equation (3.51) is based on (3.50) and incorporates the trends of ratio, $\sum_{j=1}^{J} \beta_x^{(j)} \kappa_t^{(j)}$, to ensure a non-divergent projection between populations 1 and 2. The HH model inherits the disadvantages of the HU and JLC models as described in Section 3.2.5 and 3.3.1, referred to as the problems of over-parameterization and the irregular time trends.

3.4 Conclusion of existing individual and joint models

This chapter highlights the WT, LC, HU, JLC, TLC, PLC and CBDCK models in detail, briefly introducing some improved methods of the LC model and briefly reviewing the age-period-cohort-specific (APC) models. Understanding and criticizing the properties of an existing mortality model provide an appropriate way to propose a new model.

The WT model describes mortality in terms of z-scores, using a different mortality scale from many existing models. The advantages of using z-scores are described in Section 3.2.1. Thus, using z-scores can be a way to establish a new joint model.

The LC model is widely used and evaluated mortality forecasting model. This chapter also briefly introduces the LC model's improved methods – the LM, BMS, BDV and DT models. The HU model is an extension of the LC model, by absorbing features – additional parameters and an application of smooth function for age factors. As a result, the HU model performs better than the LC model and its improved methods (Booth et al., 2006; Shang et al., 2010).

The JLC, TLC and PLC models are derived from the LC model, inheriting advantages and disadvantages of the LC model. Then, the new joint model can be extended from an existing individual model, considering the advantages and avoiding or minimizing the disadvantages. The CBDCK model is based on a APC model – the Currie model, and thus the CBDCK model inherits the strength and weakness of the APC models.

However, joint models have shortcomings as shown in Table 3.2. The JLC uses more than one time trend, and may result in uncertainty in projection as criticized in Section 3.3.1. The PLC model is appropriate for subnational populations as described in Section 3.3.3. As described in Section 3.3.4, the CBDCK model uses a cohort trend and may yield unreliable forecasts, and being designed for old age groups and for two-population analysis. The TLC model allows population effects to influence the common factors, and thus may produce divergence in projection as described in Section 3.3.2.

Model	Irregular Trends	Multiple Trends	Limited Application		
JLC		\checkmark		3.3.1	
PJLC		\checkmark		3.3.1	
TLC	\checkmark			3.3.2	
PLC			\checkmark	3.3.3	
CBDCK	\checkmark	\checkmark	\checkmark	3.3.4	

Table 3.2: Limitations of existing joint models and limitations of the relevant studies

Note: The "trends" include time and cohort trends.

Based on the shortcomings of existing individual and joint models, the new model developed in Chapter 4 is a joint model and therefore ensures non-divergence of forecasts. Furthermore, the new model does not involve cohort factors to ensure a reliable projection and uses one time trend to minimize the forecast uncertainty. In addition, the new model is suitable for all age groups and is appropriate for general populations.

Chapter 4

Joint Wang Transform Model – the Proposed New Joint Model

4.1 Objectives of the new model

This section develops a new joint model to describe the variations in mortality. This new model is an extension of the WT model introduced in Section 3.2.1 and has the following three features which address the deficiencies of existing mortality models.

Firstly, the new model is a joint model, enabling the joint fitting and forecasting of the mortality for groups of similar populations. As described in Section 2.3, mortalities throughout the world are increasingly similar or converging. Therefore, separate fitting and forecasting of mortality is inappropriate. Existing joint models have various short-comings as described in Section 3.4, such as unreliable projection due to multiple trends and limited application for specific age groups and populations. The new model uses one time trend to minimize the uncertainty of projection yet captures the improvements across all ages – the new model aims to apply to all age groups and populations.

Secondly, the new model permits flexible mortality trends over time. As described in Section 2.4, mortality rates have declined with a fair amount of uncertainty. However, existing individual and joint mortality models generally assume mortality to vary linearly over time (Lee and Carter, 1992; Renshaw and Haberman, 2003a; Tuljapurkar and Boe, 1998). The assumptions of mortality models conflict with actual mortality trends over time. To address this shortcoming, the new model permits mortality in terms of z–scores to vary flexibly over time, which is detailed in Section 4.2.

Thirdly, the new model is simple to apply. Existing individual and joint models generally use a large number of parameters to describe mortality, which may result in over– parameterization, and yield non-robust forecasts (Cairns et al., 2008). There are advantages in keeping models simple (Cairns et al., 2009; Plat, 2009b).

The new model is extended from the WT model. The WT model possesses two main advantages over other existing individual mortality models. Firstly, the WT model describes changes in mortality over time in terms of z-scores as described in Section 3.2.1. Existing mortality models almost always model the log death rate. By using another mortality scale – the changes in z-scores – the new model may describe the features of mortality from another angle, differing from the existing models. Secondly, although the shift of the z-scores is assumed to be constant over time, the WT model can be extended to allow a flexible shift of the z-scores. Additionally, by using one parameter, the WT model is one of the simplest models. The extension of the WT model inherits this simplicity.

The new model is a joint model extended from the WT model, referred to as the joint Wang Transform (JWT) model. The JWT model allows a flexible mortality trend and has a simple form. With these features, the JWT model is expected to perform better than existing individual and joint models. The evaluation in Chapter 6 suggests that the JWT model meets the above three targets and is a reliable joint model.

4.2 The JWT model

The JWT model is developed in two steps. Step 1 extends the WT model from a nonspecific to an age-period-specific model, referred to as the Wang Transform Age Period (WTAP) model. Step 2 extends the WTAP model from an individual model to a joint model.

4.2.1 Age-period-specific WT model

The WT model assumes that the z-scores (z_{xti}) experience a constant drift over time. The definition of z-scores is introduced in Section 3.1 and detailed in Section 3.2.1. Introducing the population index, *i*, to the WT model, Equation (3.1) becomes

$$z_{xti} = z_{x,t-1,i} + \lambda_i + \epsilon_{xti} , \qquad (4.1)$$

where the subscript *i* indicates that quantities, including λ_i , are dependent on population *i*. Thus the expected change in the z-score of population *i* may be different from the change or improvement in other populations. Hence this model supposes separate improvements for different populations – there is no communality across populations.

With λ_i depending only on *i*, an unbiased estimate of λ_i is derived as

$$\hat{\lambda}_i = \frac{1}{m(n-1)} \sum_{x,t} \hat{\lambda}_{xti} , \qquad \hat{\lambda}_{xti} = z_{xti} - z_{x,t-1,i} .$$

In this situation, while λ_i is population dependent, it tends to be restrictive as described in Section 3.2.1. For example, if the age range of the mortality data set is different from 0 to 100, then it is often found that the shift in z-scores tends to be non-constant across age.

A very general extension of the WT model permits both age and time effects as

$$z_{xti} = z_{x,t-1,i} + \lambda_{xti} + \epsilon_{xti} .$$

$$(4.2)$$

where (4.1) uses few parameters to describe the change of z–scores, while (4.2) uses many parameters. Equation (4.2) is impractical, since it is over-parameterized leading to overfitting and non-robust forecasting.

A possible extension of the population specific WT model is between the extremes of (4.1) and (4.2). The λ_{xti} is expressed by a function of age and time, referred to as the Wang Transform Age Period (WTAP) model:

$$\lambda_{xti} = a_{xi} + k_{ti} , \qquad (4.3)$$

where a_{xi} and k_{ti} are age and period factors for population *i*. Thus, (4.3) is an age-periodspecific model linking the progression of ages and time periods within a given population *i*. In fitting, the a_{xi} permit age effects – i.e. different progressions in z-scores for different ages – which are especially useful if the base age is not zero. Furthermore, k_{ti} is expected to be a stationary process indicating that the z-score processes are, in terms of time series terminology, at most integrated of order 1.

The WTAP model uses age-specific factors, a_{xi} . There are no age-time effects b_{xi} as with the LC model – saving on parameters. The LC model describes the mortality (the log death rate) directly and (3.8) is written as

$$E\{\log(m_{xti})\} = a_{xi} + b_{xi}k_{ti} , \qquad (4.4)$$

indicating that the expected age specific mortality in population i at time t has both "intercept" a_{xi} and "slope" effects b_{xi} . The WTAP model describes changes in z-scores. The age-specific factor, a_{xi} , reflects the rate of z-score increase over time for age x, and thus plays a role of "slope" as the b_{xi} does in (4.4) in describing variation in mortality over time. Hence, b_{xi} is not necessary in (4.3).

The time trend k_{ti} in the WTAP model monitors and models the mortality trends. When the k_{ti} sequence is relatively constant over time, z–scores increase constantly over time. When the k_{ti} sequence increases (decreases) over time, z–scores increase with acceleration (deceleration).

4.2.2 Extending the WTAP model to a joint model

The λ_{xti} as developed above fluctuates across years and ages. Pooling across similar populations can improve forecasting and yield insight into similarities and differences between populations. The joint form of the WTAP model is called the JWT model:

$$\lambda_{xti} = a_x + k_t \,, \tag{4.5}$$

where a_x and k_t are common age and time factors for all members in the group – that is the a_x and k_t do not depend on *i*. In this formulation z–score shifts for each age and time are common across all considered populations. Since λ_{xti} is a shift or change in z–scores, k_t is expected to follow a stationary process. The k_t sequence is expected to be stationary with an expected unconditional mean zero implying, in the long run, the z–scores shift by a constant amount a_x .

An expected zero value for k_t ensures the a_x and k_t are identifiable. In particular on the right hand side of (4.5), the a_x and k_t can be replaced by $a_x + c$ and $k_t - c$, respectively, indicating lack of identification unless a_x and/or k_t are normalised. The normalisation where k_t has unconditional expectation is convenient and aids interpretation.

The JWT model uses only common factors to describe the change in z-score over time. As a result, the JWT model assumes that populations share a common variation in mortality over time. This assumption is the same as the SJLC model (3.3.1) and the PLC model (3.44). In forecasting the z-score, however, the JWT model allows populations to use population and age-specific z-scores as jump-off rates. Therefore, individual populations possess individual age-specific mortality "base" levels, but share a common mortality trend. The forecast methods are discussed in Section 4.3.

4.2.3 Alternative joint Wang Transform models

Other models are possible for λ_{xti} and were tried. These alternative models have similar forms to the joint models introduced in Section 3.3 and allow for more intricate age, period and population effects and their possible interaction:

$$a_x + b_x k_t , \qquad (4.6a)$$

$$a_x + k_t + \rho_i , \qquad (4.6b)$$

$$\lambda_{xti} = \begin{cases} k_t + \rho_i , \qquad (4.6c) \end{cases}$$

$$a_x + b_x k_t \rho_i , \qquad (4.6d)$$

$$a_x + b_x k_t + \rho_i , \qquad (4.6e)$$

$$a_x + k_t \rho_i . (4.6f)$$

However, when fitting and forecasting mortality as introduced in Chapter 6, the generalisations embodied in Equations (4.6a) to (4.6f) do not fit well, or have insignificant coefficients. This is partly the result of the smoothing inherent in moving from mortality m_{xti} or q_{xti} to the z-scores. Therefore, the JWT model, for fitting and forecasting mortality, appears to have advantages over (4.6a) to (4.6f), particularly in forecasting.

The age-specific coefficient, b_x , does not significantly influence fit and forecast. In-

stead, the use of b_x may cause over-parameterization, indicating that b_x is largely redundant in modelling the changes in z-scores and therefore that (4.6a) is inappropriate.

The population effect, ρ_i , models the individual effect in a population as distinct from another population. Using ρ_i , the JWT model may yield a better fit, but was found to lead to inferior forecasts. Furthermore, the use of ρ_i generally produces divergence in the z-scores in long-term projection. Take (4.6b) as an example, for populations *i* and *j*,

$$\lambda_{xti} - \lambda_{xtj} = \rho_i - \rho_j$$
.

The differences in z-scores between i and j depend on the accumulative differences in λ between i and j over time. The accumulative difference in λ between i and j is increasing over time, and thus results in divergence of z-scores between i and j. The models (4.6c) to (4.6f) suffer from the similar problem accounted by (4.6b). With convergence demands, the population effect, ρ_i , results in divergence of z-scores across populations, and is therefore inappropriate to model the changes in z-scores.

4.3 Fitting and forecasting mortality using the JWT model

This section introduces the use of the generalised linear model (GLM) to estimate the parameters (a_x and k_t) in the JWT model. The methods to obtain the fitted and forecast mortality using these estimated parameters are then introduced.

4.3.1 Estimation of parameters in the JWT model

The model

$$z_{xti} - z_{x,t-1,i} = a_x + k_t + \epsilon_{xti}$$

can be viewed as a two-way Analysis of Variance (ANOVA) setup with different countries i serving as "within cell" replications. Since replications relate to different populations, they can be considered independent observations. However changes in z-scores for different ages within a population are related since they are derived from the same survival curve. Thus the ϵ_{xti} , for given i are related for different x. Any time series structure however is expected to be picked up in the k_t .

To estimate the unique values of a_x and k_t , constraints are required. As described in Section 4.2.2, the k_t varies in stationary manner over time. Then, the expectation of k_t is zero, $E(k_t) = 0$, ensuring k_t to be around zero over time. Therefore, constraints, $\sum_t k_t = 0$ and $k_1 = 0$ are applied to the GLM estimation. This research applies the glm package in R. The ages and the years are set as factors.

Since the JWT model is linear in the a_x and k_t , the generalised linear model (GLM) can be applied to estimate parameters. Similar applications of the GLM to estimate parameters in mortality models have been undertaken (Currie et al., 2004; Debón et al., 2008; Renshaw and Haberman, 2003a). The GLM model is generically provided by commonly used statistical programs, in for example, R, S-Plus and Matlab. Therefore, parameters of the JWT model are readily estimated by existing software packages.

GLM fitting routines generally assume the errors are independent and identical distribution (i.i.d.) and, at best, normal. In the current context, this assumption as stated above is inappropriate, since the adjacent "observations" $z_{xti} - z_{x,t-1,i}$ come from correlated transformed variables:

$$\hat{\lambda}_{xti} = z_{xti} - z_{x,t-1,i} = \Phi^{-1}(s_{xti}) - \Phi^{-1}(s_{x,t-1,i}) ,$$

where, s_{xti} is the product of p_{xti} across ages 0 - x as introduced in (3.2). As a result, s_{xti} correlates to $s_{\tilde{x}ti}$ where $\tilde{x} \neq x$. Hence, λ_{xti} is cross sectionally and serially correlated.

Accordingly, λ_{xti} does not represent independent observations and any estimation is appropriately interpreted as a smoothing or fitting exercise rather than formal estimation in an i.i.d. framework.

Nevertheless, the GLM estimation generally delivers unbiased estimates. Weighting with the population's size reflects the desire to give more weight to those ages and times with the most observations and ignore those ages and times where there are few observations. In this way GLM fitting is used to arrive at a reasonably coherent fit. In the application of the GLM estimation, the survivors out of the radix, l_{xti} (Wilmoth et al., 2007), are used to weight the population's size corresponding to λ_{xti} . The change of z-scores, λ , varies around a constant, the mean of λ_{xt} across x and t as used by De Jong and Marshall (2007). Hence, λ can be viewed as a variable of a normal distribution, and the family in the GLM function is set as Gaussian (Forfar et al., 1988; McCullagh and Nelder, 1989).

Note further that in Equation (4.5), there are no age factors b_x serving to multiply and modulate k_t , for the sake of saving parameters. The b_x does not appear to add significantly to the explanatory power of the model. Equation (4.5) is thus viewed as a basis to fit a surface to in the (x, t) plane, independent of the population i, and where the error terms serve to model deviations that a population has from the surface. The structure to any set of deviations, for example, related to a particular population, may be used to explore specific features of the given population, indicating where the model is deficient, and suggesting structural peculiarities of the population.

After obtaining \hat{a}_x and \hat{k}_t , the estimated $\hat{\lambda}_{xt}$ is

$$\hat{\lambda}_{xt} = \hat{a}_x + \hat{k}_t$$

The estimate of JWT model across countries is displayed in Table A.1 as an example.

4.3.2 Example of fitting mortality using the JWT model

Figure 4.1 displays examples of the estimated \hat{a}_x and \hat{k}_t sequences, when fitting the JWT model to jointly fit mortality across populations. The populations are described in Section 5.3. The mortality data sets of the populations are for the years 1948-1994. Since this research aims to evaluate the performance in a general application for common age groups, the age groups are selected for 0-89. Over ages 95, the z–scores may not be linear (De Jong and Marshall, 2007), and thus the modelling of λ for ages above 95 could be undertaken in a further study. The application and evaluation of the JWT model are described in Chapter 6. The JWT model is, thus evaluated and illustrated prior to a detailed analysis in subsequent chapters.



Figure 4.1: Estimates of a_x and k_t for females and males in the "across country" and "across sex" analyses

The top two panels of Figure 4.1 display \hat{a}_x and \hat{k}_t from an "across country" analysis, pooling across 13 countries for females and for males, separately. The 13 countries are introduced in Section 5.3 and the "across country" analysis is detailed in Section 6.2. The curves denote the \hat{a}_x or \hat{k}_t sequence of a group, in this case, females or males.

The bottom two panels of Figure 4.1 display the \hat{a}_x and \hat{k}_t from an "across sex" analysis for 13 different countries, pooling across males and females within each of the 13 countries introduced in Section 5.3. The "across sex" analysis is detailed in Section 6.3. The curves denote the \hat{a}_x or the \hat{k}_t sequence of a group, in this case, a country.

The top left panel of Figure 4.1 displays the \hat{a}_x sequence versus age in the "across country" analysis. For both males and females, the \hat{a}_x follows a V-shape across ages. The \hat{a}_x is initially high, drops to a minimum at age 65, and then increases with increasing age. The minimum of a_x is for the age group of 65 years old and is common for different sexes and countries. The age group of 65 years old experiences the smallest change rate of z-scores over time. For females, the old age groups, particularly ages over 80, experience a large drift of z-scores. However, for males, the old age groups experience a similar drift of z-scores as the young age groups. The plots of \hat{a}_x illustrate different patterns of variation in mortality in terms of the log death rates.

The top right panel of Figure 4.1 displays the \hat{k}_t sequence versus year in the "across country" analysis. For both males and females, the \hat{k}_t sequences generally decline but with cyclical trends, reflecting the fluctuation of the drift of the z-scores over time. The downward tendency of \hat{k}_t sequences indicate that the drift of the z-scores become smaller over time, agreeing the findings that developed countries have experienced a decrease of rate of mortality (the log death rate) decline (McMichael et al., 2004; Wilson, 2011).

By using the time-varying \hat{k}_t , the JWT model offers the advantage of permitting a flexible drift of z–scores in the short term. The assumption of stationarity ensures movement to a constant drift of z–scores in the long term.

The bottom left panel of Figure 4.1 displays \hat{a}_x in the "across sex" analysis. For all 13 countries, the \hat{a}_x sequences experience similar shapes in the left panel of Figure 4.1, although of a different magnitude, resulting in a series of somewhat parallel curves. The curves follow a shallow V-shape similar to that in the top left panel of Figure 4.1.

The bottom right panel of Figure 4.1 displays \hat{k}_t in the "across sex" analysis. For all 13 countries, the \hat{k}_t sequences have an initial downward tendency, and then fluctuate around zero. The curves of \hat{k}_t are less similar than the curves displayed in the top right panel of Figure 4.1. For individual years, \hat{k}_t may be positive or negative, but generally tend to converge at zero. Therefore, the \hat{k}_t sequence in the "across sex" analysis, supports the assumption that \hat{k}_t varies stationarily over time.

In the left panels of Figure 4.1, a_x obtains the lowest values for age groups in the 60s, denoting that the z-scores experience the smallest change for such age groups. The finding of these two panels is consistent with the conclusion drawn by Debón et al. (2012) where people aged 60 years old contribute the least to life expectancy. Therefore, the JWT model is basically practical in describing mortality in terms of z-scores.

Figure 4.2 displays the average drifts of the log death rate and z-scores for females and males in Australia, using data for ages 0-89 and years 1948-1994. The correlation coefficient between females and males is 0.81 for the left panel and 0.88 for the right panel. By using both log death rate and z-scores, females and males are highly correlated. On one hand, z-scores capture the features of mortality as well as the log death rate. On the other hand, females and males experience homogeneous variation in mortality, suggesting that the "across sex" analysis is reliable. The coherent analysis and joint forecast for females and males have been proposed in the literature and have been found to enhance forecasts (Li and Lee, 2005; Li, 2013; Hyndman et al., 2013)



Figure 4.2: Average of annual differences in the log death rate and the z–scores across ages for females and males in Australia

4.3.3 Forecast of mortality using the JWT model

Forecasting mortality using the JWT model involves three steps.

Step 1 forecasts the \hat{k}_t sequence using an appropriate time series model fitted to the \hat{k}_t . Ideally, the fit of the time series model is undertaken simultaneously with the estimation of the a_x and k_t . However, in this thesis, we have not done this: instead we have estimated the k_t using GLMs and then used time series methods to model the \hat{k}_t . Separating forecast from estimate is generally applied by many of existing mortality models. For example, in the LC model, \hat{k}_t is estimated via the SVD, Poisson process or other methods, and then a random walk with drift is applied to model and forecast the \hat{k}_t sequence. Consider, for example, the AR(1) process. As described in Sections 4.2.1 and 4.3.2, using a non-stationary process to forecast \hat{k}_t may yield infinite z-scores in the long-term projection. However, when \hat{k}_t trends to zero, the z-scores shift with a constant drift, \hat{a}_x , for age x. This constant drift of z-scores is similar to the assumption of a random walk with drift for the time trend in the LC model as introduced in Section 3.2.3. Therefore, the stationary \hat{k}_t is reliable. Furthermore, using a stationary process to forecast \hat{k}_t allows the drift of z-scores to vary in the short-term, but to be constant in the long-term. The JWT model is more flexible than models that use a fixed rate of mortality decline over time as, for instance, with the LC model.

Step 2 forecasts the $\hat{\lambda}_{xti}$. Suppose *n* denotes the latest observed time point. Since the JWT model assumes that all populations share the same mortality trends in the forecasting horizon,

$$\hat{\lambda}_{x,n+h,i} = \hat{\lambda}_{x,n+h} = \hat{a}_x + \hat{k}_{n+h} , \qquad (4.7)$$

where \hat{k}_{n+h} and $\hat{\lambda}_{x,n+h,i}$ denote h step ahead forecasts of \hat{k}_t and $\hat{\lambda}_{xti}$ at time t = n.

Step 3 forecasts the z-scores. The JWT model permits individual populations to possess a "base" level of mortality. The forecast of z-scores uses the latest z-scores z_{xni} of individual populations at each age x as the jump-off rates:

$$\hat{z}_{x,n+h,i} = z_{xni} + \left(\hat{\lambda}_{x,n+1} + \dots + \hat{\lambda}_{x,n+h}\right) = z_{xni} + \left(h\hat{a}_x + \hat{k}_{n+1} + \dots + \hat{k}_{n+h}\right)$$
(4.8)

where z_{xni} is the observation of the z-scores in year n and the future $\hat{\lambda}$ s are forecast as in (4.7). Note that with a zero mean process $\hat{k}_{n+h} \to 0$ as $h \to \infty$. In the long term, $\hat{k}_{n+h} \to 0$ and the z-scores increase by \hat{a}_x per period.

The JWT model (4.5) models the change in the z–scores. This emphasises the critical nature of jump-off rates. For example consider the case where a different *i* refers to males and females in a given country. Then, (4.5) models the common change from given z–

score levels for males and for females. Equation (4.8) uses z_{xni} as the jump-off rates and reflects the principal and actual differences in mortality across populations at the beginning of the forecasting period. In spite of these differences, mortalities are expected to converge across populations during the long term, for example, the forecasting period.

Step 4 transforms the forecast z-scores to the forecast survival probability:

$$\hat{s}_{x,n+h,i} = \Phi(\hat{z}_{x,n+h,i})$$

where $\hat{s}_{x,n+h,i}$ is the *h* step ahead forecast of the survival probability for age *x* of population *i*. In turn, (3.7) can be used to calculate one year death rates such as $\hat{q}_{x,n+h,i}$, $\hat{m}_{x,n+h,i}$ or $\log(\hat{m}_{x,n+h,i})$ for direct comparison to the output used by other models – for example, the LC model (3.8).

4.4 Advantages of the JWT model

The JWT model possesses a number of novel features and possible advantages. Unlike the existing models that model mortality directly, the JWT model describes the change of mortality in terms of the z–scores. The use of z–scores has advantages as described in Section 3.2.1. Analysis shows that z–score trend to more linearly over time. Linear time series are easy to deal with and in particular to forecast. Therefore, the scales of z–score describe mortality data in an efficient way, where the variations in mortality are smooth and linear, and thus are easy to capture. Figure 4.3 displays the z–scores of Australian females and Dutch males between 1948 and 1994 for ages 0-89. The z-scores vary almost linearly for each age group over time.



Figure 4.3: The z-scores of Australian females and Dutch males between 1948 and 1994

By applying k_t to model the change in z-scores, the JWT model permits a flexible trends of z-scores. An increase in k_t indicates an accelerated increase in the z-scores. Similarly a decrease in k_t indicates a decelerating increase in the z-scores. If k_t is zero for all t, the z-scores increase by the same amount a_x depending on age x. Existing mortality models assume log-mortality varies with constant drift: the expected change is constant. For example, the LC model, as described in Section 3.2.3, gives a random walk with drift for the time trends. Therefore, the JWT model operates on a different "scale" and can be viewed both as more flexible or less flexible depending on the scale viewpoint.

As a joint model, the JWT model yields non-divergent forecast of mortality across populations. From (4.8), the current differences – for instance, the jump-off differences in z–scores – between populations do not change. Given populations *i* and *j*the year n+h, the difference in z–scores is estimated to be

$$\hat{z}_{x,n+h,i} - \hat{z}_{x,n+h,j} = \left\{ z_{xni} + g(h) \right\} - \left\{ z_{xni} + g(h) \right\} = z_{xni} - z_{xnj} , \qquad (4.9)$$

where g(h) is a function of h and denotes the cumulative growth of z-scores from year n to n + h. From (4.8),

$$g(h) = \hat{\lambda}_{x,n+1} + \dots + \hat{\lambda}_{x,n+h} = h\hat{a}_x + \hat{k}_{n+1} + \dots + \hat{k}_{n+h} .$$
(4.10)

Hence, for any a_x , k_t or h, the difference in the forecast z-scores between populations iand j is constant for given x. Since z-scores increase over time, g(h) > 0 for h > 0.

The constant differences in z-scores across populations imply the convergence of the survival probability functions of the different populations. Let f(z) and F(z) denote the probability density function and the cumulative distribution function of the standard normal distribution of z, with the properties of f(z) as:

$$f(z_2) < f(z_1) , \qquad -\infty < z_2 < z_1 < 0 ,$$

$$f(z_2) > f(z_1) , \qquad 0 < z_2 < z_1 < +\infty .$$
(4.11)

Given populations i and j and $0 < z_{xnj} < z_{xni} < +\infty$,

$$s_{xni} - s_{xnj} = \Phi(z_{xni}) - \Phi(z_{xnj}) = F(z_{xni}) - F(z_{xnj}) = \int_{z_{xnj}}^{z_{xni}} f(z)dz , \qquad (4.12)$$

Similarly,

$$\hat{s}_{x,n+h,i} - \hat{s}_{x,n+h,j} = \int_{\hat{z}_{x,n+h,j}}^{\hat{z}_{x,n+h,i}} f(z) dz .$$
(4.13)

When the difference between z_{xni} and z_{xnj} is small, (4.12) and (4.13) become

$$s_{xni} - s_{xnj} \approx f\left(\frac{z_{xni} + z_{xnj}}{2}\right) \left(z_{xni} - z_{xnj}\right) ,$$
$$\hat{s}_{x,n+h,i} - \hat{s}_{x,n+h,j} \approx f\left\{\frac{z_{xni} + z_{xnj}}{2} + g(h)\right\} \left(z_{xni} - z_{xnj}\right) .$$

Since g(h) > 0 and $0 < z_{xnj} < z_{xni} < +\infty$, from (4.11),

$$f\left(\frac{z_{xni}+z_{xnj}}{2}\right) > f\left\{\frac{z_{xni}+z_{xnj}}{2}+g(h)\right\}$$

and thus

$$s_{xni} - s_{xnj} > \hat{s}_{x,n+h,i} - \hat{s}_{x,n+h,j}$$

Therefore, the difference in survival probability between populations i and j decreases over time, when the difference in z-scores between i and j is constant over time.

Given $z_{xni} = 1.2$, $z_{xnj} = 1$ and g(h) = 0.5, from the standard normal distribution table, the difference in survival probability between *i* and *j* in year *n* is

$$s_{xni} = \Phi(z_{xni}) = \Phi(1) = 0.8849 ,$$

$$s_{xnj} = \Phi(z_{xnj}) = \Phi(1.2) = 0.8413 ,$$

$$s_{xni} - s_{xnj} = 0.0436 ,$$

and the difference in survival probability between i and j in year n + h is

$$\hat{s}_{x,n+h,i} = \Phi\left(\hat{z}_{x,n+h,i}\right) = \Phi\left(1.5\right) = 0.9554 ,$$
$$\hat{s}_{x,n+h,j} = \Phi\left(\hat{z}_{x,n+h,j}\right) = \Phi\left(1.7\right) = 0.9332 ,$$
$$\hat{s}_{x,n+h,i} - \hat{s}_{x,n+h,j} = 0.0222 ,$$

Then, $s_{xni} - s_{xnj} > \hat{s}_{x,n+h,i} - \hat{s}_{x,n+h,j}$. The constant differences in z-scores across populations result in convergence in survival probability across populations. Therefore, the JWT model captures the convergence properties of mortality across populations.

Finally, the JWT model is simple for estimation and application. The JWT model is an extension of the WT model which, in its simplest form, uses one parameter to fit and forecast mortality. Hence, the JWT model uses much fewer parameters than existing individual or joint models. Furthermore, the JWT model is linear. With a linear format, the parameters of the JWT model are easy to fit, and thus many of the existing fitting tools can be applied. Additionally, the unique time trend has less forecast divergence than the models with multiple time trends.

4.5 Disadvantages of the JWT model

Transformed from the survival probability, the z–scores are not a direct mortality indicator and hence may not be easily understood. However, the z–scores also fully reflect the mortality information available as shown in Figure 4.2. In addition, the z–scores can be transformed into any well-understood mortality scales, for example, the log death rate. To transform from z–scores to log death rate is simply introduced in Step 4 in Section 4.3.3 with a reverse process to (3.7).

From (3.2), the survival probability (s_{xt}) is a product of the one year survival rate (p_{xt}) across age x. Hence, the survival probabilities correlate with each other across ages. However, the dependent survival probability can be converted back to the independent one year death rates. Therefore, modelling the correlated z-scores is acceptable in this scenario.

In addition, the JWT model has the risk of permitting z-scores to cross over by age. For observed z-scores, the curves of z_{xti} and z_{yti} do not cross over, where *i* is a given population and *x* and *y* are two given age groups. The WT model assumes that all age groups share the same constant drift of z-scores over time, and thus avoids the risk of cross over. Nevertheless, the JWT model permits different age groups to vary by a_x , and then a crossover between the curves of z_{xti} and z_{yti} is possible. However, this risk does not significantly influence the application of the JWT model. On one hand, the a_x is estimated using non-crossing z-scores, and the estimates of z-scores are expected to be similarly consistent. On the other hand, in the application of the JWT model as introduced in Chapter 6, the estimates of a_x are similar across x as displayed in the left panels of Figure 4.1. Incorporating the different values between z_{xni} for different x, the small differences of a_x across x do not appear to result in crossover of z-scores in the forecasting horizon n + h.

Chapter 5

Methods and Measures to Evaluate Models

5.1 Outline

This chapter introduces measures to evaluate the models, including the JWT model developed in Chapter 4 and some other existing models introduced in Chapter 3. Existing literature may not comprehensively evaluate joint models. The measures introduced in this chapter indicate the performance of the JWT model versus existing models, and individual versus joint models. These comparisons illustrate the usefulness, efficacy and advantages of the JWT model. In this comparative analysis we also cover individual models and hence provide answers as to the utility, practicability and efficacy of joint as opposed to individual models.

Comparative evaluation include both fit and forecast performance. To make the relevant comparisons requires measures to evaluate models, especially for forecast accuracy. To benchmark the strengths and weaknesses of the JWT model, a range of existing individual and joint models are selected for comparison. In particular we select seven existing models, including three individual models and four joint models, to compare with the JWT model (Section 5.2). To ensure a fair comparison, the eight models are applied to the same scenarios – the same fitting period, jump-off rates and forecasting period.

The eight models are applied to fit and forecast mortality of a group of populations (Section 5.3). Due to the reliable record of deaths, the mortality data of developed countries are more reliable than those of developing or poor countries. As a result, mortality data of developed countries are usually applied by demographers to evaluate the performance of mortality models, such as the G7 countries (Tuljapurkar et al., 2000), Australia (Tickle and Booth, 2014), Denmark (Jarner et al., 2008), Italy (D'Amato et al., 2011), England & Wales (Renshaw and Haberman, 2003b, 2006), the USA (Cairns et al., 2009; Lee and Carter, 1992), Sweden (Lundström and Qvist, 2004) and Japan (Wilmoth, 1993). Conversely, less developed countries experience various mortality trends over time and might not be robust for the evaluation of joint models Li and Lee (2005). Therefore, developed countries with large populations are reliable for the evaluation of models. This research selects the 13 developed countries used by Li and Lee (2005) and Russolillo et al. (2011)

Joint models are applied to jointly fit and forecast mortality "across country" and "across sex". The "across country" analysis jointly fits and forecasts mortality of the 13 countries for females and then for males. The "across sex" analysis jointly fits and forecasts mortality of the females and males for each of the 13 countries. Thus, in the "across country" analysis, mortality experience across countries is pooled. In the "across sex" analysis experience is pooled across the two sexes.

To evaluate forecast performance of the models, the time horizon is divided into two parts: the fitting period (1948-1994) and the forecasting period (1995-2009 or 1995-2044). Model parameters are estimated using data from 1948 to 1994, 47 years in total. The goodness of fit is measured in the fitting period. Models are then applied to forecast mortality from 1995 to 2009, a 15-year projection, where the forecast accuracy is measured. For an out-of sample prediction, the length of fitting period is three times of that of the forecasting period. To evaluate the models' ability in controlling divergence of mortality across populations, models are applied to forecast mortality from 1995 to 2044, a 50-year projection, where the convergence properties of mortality across populations are measured. Although the forecasting period (50 years) is longer than the fitting period (45 years), this prediction does not evaluate the forecast accuracy by the convergence properties. Furthermore, all these models are under the same scenario, and thus the prediction is acceptable and fare for the selected models.

Methods are applied consistently to ensure that the evaluation assesses the models rather than the specific details of its application (Section 5.4). For example, for all selected models, actual mortality is used as the jump-off rate in order to reduce the bias in forecasting (Booth et al., 2002b; Lee and Miller, 2001). However, since the actual mortality rate is jagged across ages and may cause high bias for specific age groups (Hyndman et al., 2013), the jump-off log death rate is smoothed across age by using b-splines prior to projection. Since all models are based on the same conditions, smooth jump-off rates are expected to improve the performance of all models similarly.

Goodness of fit is measured using the Bayesian information criterion (BIC) (Section 5.7). BIC is widely used to describe the match between a model and data. BIC measures the size of the fitting error – the difference between observed and estimated number of deaths and penalises for the number of parameters in a model.

Forecasting performance is based on forecast accuracy (Section 5.8) and convergence properties of mortality across populations (Section 5.10). The measures of forecast accuracy reflect the match between observed and forecast mortality. Individual models may forecast increasing differences in mortality between populations, referred to as divergence. The convergence properties examine the models' ability in controlling divergence when forecasting mortality. Thus BIC decreases with lower fitting error and a decrease in number of parameters.

The forecast accuracy is based on the analysis of errors, including mean error (ME) and mean absolute error (MAE). The ME can be positive, negative or zero, indicating underestimation, overestimation and zero error between observation and fit. The MAE is non-negative and low MAE indicates high forecast accuracy. Total ME and MAE are averaged across ages and over time, and reflect the overall forecast accuracy without specific reference to age or time. The ME and MAE over time illustrate the trends of errors over the forecasting period.

The comparative MAE (CMAE) is the percentage difference in the MAE between a model and a given benchmark model. The LC model is widely used and understood, and is thus used as the benchmark. For example, a CMAE of 50% indicates the MAE of the current model is 50% higher than that of the LC model. A positive CMAE indicates that the current model has worse forecast accuracy than the LC model, while negative a CMAE means that the current model produces better forecast accuracy than the LC model.

The convergence properties of mortality across populations describes the relative differences in mortality of populations as described in Section 5.10. The convergence property is described by the standard deviation of the forecast of mortality across different populations. A model causes divergence across populations when forecasting mortality, if such a standard deviation increases over time. Mortality is increasingly similar among most countries as described in Section 2.3, referred to as convergence. However, individual models may project an increasing difference, referred to as divergence. The conflict between divergence and convergence requires measures to be used to evaluate the performance of models. The convergence properties of mortality across populations assess whether a model produces an increasing difference in mortality across populations. Generally, joint models are expected to produce non-divergent projections.

5.2 Selection of models for evaluation

The models discussed in the previous chapters are listed in Table 5.1 with a symbol " \checkmark " indicating those models used in the empirical evaluation. This research selects three individual models (WT, LC and HU) and four joint models (SJLC, JLC, TLC and PLC) and these seven models are compared to the JWT model.

Factors	Name	Selected	Section	Reference		
Zero	WT	\checkmark	3.2.1	De Jong and Marshall (2007)		
Two	WTAP		4	Proposed by this research		
	LC	\checkmark	3.2.3	Lee and Carter (1992)		
	HU	\checkmark	3.2.5	Hyndman et al. (2007)		
	LM		3.2.4	Lee and Miller (2001)		
	BMS		3.2.4	Booth et al. (2002b)		
	BDV		3.2.4	Brouhns et al. (2002)		
	DT		3.2.4	De Jong and Tickle (2006)		
	ASE		3.2.4	Renshaw and Haberman (2003c)		
	CBD			Cairns et al. (2006)		
	CBD1		326			
Three	CBD2		5.2.0	Cairns et al. (2009)		
	CBD3					
	RH		377	Renshaw and Haberman (2006)		
	Currie		5.2.1	Currie (2006)		
	Plat		3.2.7	Plat (2009a)		
Joint	SJLC					
	JLC	\checkmark	3.3.1	Li and Lee (2005)		
	PJLC					
	TLC	\checkmark	3.3.2	Russolillo et al. (2011)		
	PLC	\checkmark	3.3.3	Debón et al. (2011)		
	CBDCK		3.3.4	Cairns et al. (2011a)		
	ILC		2.1	Proposed by this research		
	CLC		3.1			
	JWT	\checkmark	4			

Table 5.1: The models applied in this research

The LC model is widely used, and thus the LC model is generally chosen as a benchmark to compare with a new model. Furthermore, the LC model is a "base" model for many existing mortality models. These models may inherit the advantages and disadvantages of the LC model. The use of a Poisson model for deaths for estimation is widely applied to the LC model and its extensions (Delwarde et al., 2007; Dowd et al., 2010; Li et al., 2006; Wang and Lu, 2005). In the analyses detailed in Chapter 6, Poisson regression is applied to the LC model and to the common factors in the SJLC and JLC models.

The new joint model proposed in this research – the JWT model – is derived from the WT model, so the WT model is also selected. The HU model is a classical extension of the LC model, combining the good features of the LC model and its improved methods, and thus is selected. The SJLC, JLC, TLC and PLC models are selected as the joint models for comparison.

5.3 Data

This research selects 13 low mortality and high income countries from the Human Mortality Database (HMD) according to existing studies (Crimmins et al., 2011; Li and Lee, 2005; Russolillo et al., 2011). The 13 developed countries (and their acronyms) are: Australia (AUS), Canada (CAN), England and Wales (ENW), Finland (FIN), France (FRA), Italy (ITA), Spain (ESP), Sweden (SWE), Switzerland (CHE), the United States (USA), Japan (JPN), Denmark (DNK) and the Netherlands (NLD). Developed countries have more reliable mortality data records, compared to developing countries.

The evaluation splits the available records into fitting and forecasting periods. Goodness of fit depends on the fitting period: years 1948-1994. Forecast accuracy relates to the short forecasting period: years 1995-2009. The convergence properties of mortality across populations are assessed using the long forecasting period: years 1995-2044. The ages are between 0 and 89. The data extracted is exposed to risk and central mortality rates by year and single year of age.

5.4 Methods used to forecast time trends

The data analysis in this research is completed using R for all estimates and forecasts. The LC model is modelled using the open source R code provided by life metrics (www.jpmorgan.com/lifemetrics) and widely used by academics, such as Cairns et al. (2009), Cairns et al. (2011b) and Plat (2009a). The HU model is modelled by its original R demography package, provided by Hyndman et al. (2007). The TLC model is based on the R package, ThreeWay. The PLC model comes from the R codes provided by Debón et al. (2011). The JLC model is partially completed in this research: the common factors are the same with the LC model and the individual factors are amended from the demography package. The WT and JWT models are written in this research using R language, but the fitted process of the JWT model is based on the glm package. All predictions are based on the arima package.

The WT model uses constant drift as described in Section 3.2.1. The time trend of the JWT model is assumed to be stationary as described in Section 4.3, and thus an AR(1) process is applied. Due to fluctuation, b-splines are applied to smooth the k_t sequence prior to the application of the AR(1) process, in order to reduce the forecast uncertainty.

The random walk with drift (RW) was introduced by Lee and Carter (1992), and is widely accepted by the LC model and its extensions (Booth et al., 2002b; Cairns et al., 2006; Lee, 2000; Tuljapurkar et al., 2000). Although the RW may not be acceptable in certain populations, for a general and simple application, this research applies RW to most LC family models. The RW is applied to the SJLC, PLC and TLC models. For the

JLC model, RW is applied to the common trend and a AR(1) process is applied to the individual time trends (Li and Lee, 2005; Li, 2013).

The HU model uses different processes and orders for specific data sets. The selection of order J is discussed in Section 3.2.5. Let s denote the break year and $1980 \le s \le 1992$. To ensure that the HU model is applied as intended, this research uses the "fdm" and "forecast.fdm" functions included in the Demography R package provided by Hyndman et al. (2007) to fit and forecast mortality. Table 5.2 displays the orders and forecast methods for time trends determined for the HU model.

Table 5.2. Orders and forecast methods used by the fife moder to forecast time trends								
Females	Country	AUS	CAN	ENW	FIN	FRA	ITA	ESP
	Order	4	2	1	1	1	1	1
	Method	ARIMA	RW	RW	RW	RW	RW	ARIMA
	Country	SWE	CHE	USA	JPN	DNK	NLD	
	Order	1	1	1	2	1	1	
	Method	ARIMA	ARIMA	ARIMA	ARIMA	ARIMA	ARIMA	
Males	Country	AUS	CAN	ENW	FIN	FRA	ITA	ESP
	Order	2	3	3	3	3	3	1
	Method	ARIMA	RW	RW	RW	RW	RW	ARIMA
	Country	SWE	CHE	USA	JPN	DNK	NLD	
	Order	1	1	1	1	1	1	
	Method	ARIMA	ARIMA	ARIMA	ARIMA	ARIMA	ARIMA	

Table 5.2: Orders and forecast methods used by the HU model to forecast time trends

5.5 The jump-off rates

This research applies smooth actual jump-off rates, by applying b-spline methods to smooth the log death rates across age. The jump-off rates of the PLC and WT, JWT models are transformed from the smooth actual log death rate. Using smooth mortality rates provides more certainty in projection (Delwarde et al., 2007; Hyndman et al., 2007). Also, using the same jump-off rates across all models ensures that comparisons of forecasting performance are on a consistent basis. Jump-off can be from either fitted rates or actual rates (Booth et al., 2002b; Lee and Miller, 2001). Fitted jump-off rates refer to the estimated mortality rates in year n using a given mortality model, where n is the last year in the fitting period. Actual jump-off rates refer to the observed mortality rates in year n. Smooth actual jump-off uses methods, for example, b-spline, to smooth the actual jump-off rates across age.

Fitted jump-off rates are generated by models, and thus do not provide a consistent basis for the comparison of forecasting performance across models. Also, the differences of jump-off rates across populations may be caused by models, and then the analysis of convergence properties across populations as described in Section 5.10 is not reliable. Also, fitted jump-off rates may cause cumulative bias over the forecasting horizon (Lee and Miller, 2001).

Actual rates avoid jump-off error/bias and reflect the actual initial differences in mortality across different populations. Actual jump-off rates enable a clear assessment of forecast accuracy and of non-divergence to be made. Joint models are expected to yield non-increasing differences of the forecast mortality in the long term. Therefore, the use of actual jump-off rates clearly illustrates models' ability in controlling divergence. However, actual jump-off rates may be jagged due to unexpected outliers for individual age groups in certain year. For example, the death rate was zero for age 8 in year 1994 of Swedish females. Smooth jump-off rates avoids the errors caused by such outlier values.

Table 5.3 displays the formulae used for projection. The WT and JWT models apply three steps to obtain $\hat{\lambda}_{x,n+h,i}$, $\hat{z}_{x,n+h,i}$, and the $\hat{s}_{x,n+h,i}$. The \tilde{z}_{xni} is transformed from $\widetilde{\log(m_{xni})}$, where $\widetilde{\log(m_{xni})}$ denotes the smooth $\log(m_{xni})$ across x. The PLC model uses $\operatorname{logit}(\tilde{q}_{xni})$ as jump-off rates, due to the better forecast accuracy of (3.45) than (3.44),¹ where $\operatorname{logit}(\tilde{q}_{xni})$ is transformed from $\log(\tilde{m}_{xni})$.

¹The PLC model has MAE of 0.223 (females) and 0.236 (males) if using logit(\tilde{q}_{xn}) in Section 6.2.2.

ModelsForecast of mortalityWT $\hat{\lambda}_{x,n+h,i} = \hat{\lambda}_i$ $\hat{x}_{x,n+h,i} = z_{xni} + \sum_{t=n+1}^{n+h} \hat{\lambda}_{xti} = \tilde{z}_{xni} + h\hat{\lambda}$ $\hat{s}_{x,n+h,i} = \Phi\left(\hat{z}_{x,n+h,i}\right) = \Phi\left(\tilde{z}_{xni} + h\hat{\lambda}\right)$ LC $\log\left(\hat{m}_{x,n+h,i}\right) = \log\left(\underline{m}_{xni}\right) + \hat{b}_{xi}\hat{k}_{n+h,i}$ HU $\log\left(\hat{m}_{x,n+h,i}\right) = \log\left(\underline{m}_{xni}\right) + \sum_{j} \hat{b}_{xi}^{(j)} \hat{k}_{n+h,i}^{(j)}$ SJLC $\log\left(\hat{m}_{x,n+h,i}\right) = \log\left(\underline{m}_{xni}\right) + \hat{b}_{x}\hat{k}_{n+h}$ JLC $\log\left(\hat{m}_{x,n+h,i}\right) = \log\left(\underline{m}_{xni}\right) + \hat{b}_{x}\hat{k}_{n+h} + \hat{b}_{xi}\hat{k}_{n+h,i}$ TLC $\log\left(\hat{m}_{x,n+h,i}\right) = \log\left(\underline{m}_{xni}\right) + \hat{b}_{x}\hat{k}_{n+h} \rho_{i}$	Table 5.5. The forecast formulae of the models				
$\begin{split} & \text{WT} \qquad \begin{array}{l} \hat{\lambda}_{x,n+h,i} = \hat{\lambda}_i \\ \hat{z}_{x,n+h,i} = z_{xni} + \sum_{t=n+1}^{n+h} \hat{\lambda}_{xti} = \tilde{z}_{xni} + h\hat{\lambda} \\ \hat{z}_{x,n+h,i} = \Phi\left(\hat{z}_{x,n+h,i}\right) = \Phi\left(\tilde{z}_{xni} + h\hat{\lambda}\right) \\ \hline \text{LC} \qquad \log\left(\hat{m}_{x,n+h,i}\right) = \log\left(\underline{m}_{xni}\right) + \hat{b}_{xi}\hat{k}_{n+h,i} \\ \hline \text{HU} \qquad \log\left(\hat{m}_{x,n+h,i}\right) = \log\left(\underline{m}_{xni}\right) + \sum_j \hat{b}_{xi}^{(j)}\hat{k}_{n+h,i}^{(j)} \\ \hline \text{SJLC} \qquad \log\left(\hat{m}_{x,n+h,i}\right) = \log\left(\underline{m}_{xni}\right) + \hat{b}_x\hat{k}_{n+h} \\ \hline \text{JLC} \qquad \log\left(\hat{m}_{x,n+h,i}\right) = \log\left(\underline{m}_{xni}\right) + \hat{b}_x\hat{k}_{n+h} + \hat{b}_{xi}\hat{k}_{n+h,i} \\ \hline \text{TLC} \qquad \log\left(\hat{m}_{x,n+h,i}\right) = \log\left(\underline{m}_{xni}\right) + \hat{b}_x\hat{k}_{n+h} \rho_i \\ \hline \end{array}$	Models	Forecast of mortality			
$\begin{array}{c c} \text{LC} & \log \left(\hat{m}_{x,n+h,i} \right) = \log \left(m_{xni} \right) + \hat{b}_{xi} \hat{k}_{n+h,i} \\ \text{HU} & \log \left(\hat{m}_{x,n+h,i} \right) = \log \left(m_{xni} \right) + \sum_{j} \hat{b}_{xi}^{(j)} \hat{k}_{n+h,i}^{(j)} \\ \hline \text{SJLC} & \log \left(\hat{m}_{x,n+h,i} \right) = \log \left(m_{xni} \right) + \hat{b}_{x} \hat{k}_{n+h} \\ \hline \text{JLC} & \log \left(\hat{m}_{x,n+h,i} \right) = \log \left(m_{xni} \right) + \hat{b}_{x} \hat{k}_{n+h} + \hat{b}_{xi} \hat{k}_{n+h,i} \\ \hline \text{TLC} & \log \left(\hat{m}_{x,n+h,i} \right) = \log \left(m_{xni} \right) + \hat{b}_{x} \hat{k}_{n+h} \rho_{i} \\ \hline \end{array}$	WT	$ \hat{\lambda}_{x,n+h,i} = \hat{\lambda}_i \hat{z}_{x,n+h,i} = z_{xni} + \sum_{t=n+1}^{n+h} \hat{\lambda}_{xti} = \tilde{z}_{xni} + h\hat{\lambda} \hat{s}_{x,n+h,i} = \Phi\left(\hat{z}_{x,n+h,i}\right) = \Phi\left(\tilde{z}_{xni} + h\hat{\lambda}\right) $			
$\begin{array}{c c} \text{HU} & \log\left(\hat{m}_{x,n+h,i}\right) = \log\left(m_{xni}\right) + \sum_{j} \hat{b}_{xi}^{(j)} \hat{k}_{n+h,i}^{(j)} \\ \hline \text{SJLC} & \log\left(\hat{m}_{x,n+h,i}\right) = \log\left(m_{xni}\right) + \hat{b}_{x} \hat{k}_{n+h} \\ \hline \text{JLC} & \log\left(\hat{m}_{x,n+h,i}\right) = \log\left(m_{xni}\right) + \hat{b}_{x} \hat{k}_{n+h} + \hat{b}_{xi} \hat{k}_{n+h,i} \\ \hline \text{TLC} & \log\left(\hat{m}_{x,n+h,i}\right) = \log\left(m_{xni}\right) + \hat{b}_{x} \hat{k}_{n+h} \rho_{i} \\ \hline \end{array}$	LC	$\log\left(\hat{m}_{x,n+h,i}\right) = \log\left(\underbrace{m_{xni}}_{i}\right) + \hat{b}_{xi}\hat{k}_{n+h,i}$			
$\begin{array}{c c} \textbf{SJLC} & \log\left(\hat{m}_{x,n+h,i}\right) = \log\left(m_{xni}\right) + \hat{b}_x \hat{k}_{n+h} \\ \hline \textbf{JLC} & \log\left(\hat{m}_{x,n+h,i}\right) = \log\left(m_{xni}\right) + \hat{b}_x \hat{k}_{n+h} + \hat{b}_{xi} \hat{k}_{n+h,i} \\ \hline \textbf{TLC} & \log\left(\hat{m}_{x,n+h,i}\right) = \log\left(m_{xni}\right) + \hat{b}_x \hat{k}_{n+h} \rho_i \\ \hline \end{array}$	HU	$\log\left(\hat{m}_{x,n+h,i}\right) = \log\left(\tilde{m}_{xni}\right) + \sum_{j} \hat{b}_{xi}^{(j)} \hat{k}_{n+h,i}^{(j)}$			
$\begin{array}{c c} \textbf{JLC} & \log\left(\hat{m}_{x,n+h,i}\right) = \log\left(m_{xni}\right) + \hat{b}_x \hat{k}_{n+h} + \hat{b}_{xi} \hat{k}_{n+h,i} \\ \hline \textbf{TLC} & \log\left(\hat{m}_{x,n+h,i}\right) = \log\left(m_{xni}\right) + \hat{b}_x \hat{k}_{n+h} \rho_i \\ \hline \end{array}$	SJLC	$\log\left(\hat{m}_{x,n+h,i}\right) = \log\left(\widetilde{m}_{xni}\right) + \hat{b}_x \hat{k}_{n+h}$			
TLC $\log(\hat{m}_{x,n+h,i}) = \log(\underline{m}_{xni}) + \hat{b}_x \hat{k}_{n+h} \rho_i$	JLC	$\log\left(\hat{m}_{x,n+h,i}\right) = \widetilde{\log\left(m_{xni}\right)} + \hat{b}_x\hat{k}_{n+h} + \hat{b}_{xi}\hat{k}_{n+h,i}$			
	TLC	$\log\left(\hat{m}_{x,n+h,i}\right) = \log\left(\widetilde{m}_{xni}\right) + \hat{b}_x \hat{k}_{n+h} \rho_i$			
PLC $\left \text{ logit } \left(\hat{q}_{x,n+h,i} \right) = \text{ logit } \left(q_{xni} \right) + \hat{b}_x \hat{k}_{n+h} \right.$	PLC	$\log \left(\hat{q}_{x,n+h,i}\right) = \widetilde{\log \left(q_{xni}\right)} + \hat{b}_x \hat{k}_{n+h}$			
JWT $ \begin{aligned} \hat{\lambda}_{x,n+h} &= \hat{a}_x + k_{n+h} \\ \hat{z}_{x,n+h,i} &= z_{xni} + \sum_{t=n+1}^{n+h} \hat{\lambda}_{xti} \\ \hat{s}_{x,n+h,i} &= \Phi\left(\hat{z}_{x,n+h,i}\right) = \Phi\left(\tilde{z}_{xni} + \sum_{t=n+1}^{n+h} \hat{\lambda}_{xti}\right) \end{aligned} $	JWT	$ \hat{\lambda}_{x,n+h} = \hat{a}_x + k_{n+h} \hat{z}_{x,n+h,i} = z_{xni} + \sum_{t=n+1}^{n+h} \hat{\lambda}_{xti} \hat{s}_{x,n+h,i} = \Phi\left(\hat{z}_{x,n+h,i}\right) = \Phi\left(\tilde{z}_{xni} + \sum_{t=n+1}^{n+h} \hat{\lambda}_{xti}\right) $			

Table 5.3: The forecast formulae of the models

5.6 Transforming mortality scales

The measure introduced in Section 5.7 is calculated using the number of deaths (d). Observed d is obtained from data sets. Fitted (or forecast) d is calculated using (3.17) as $\hat{d}_{xt} = e_{xt}\hat{m}_{xt}$, where e_{xt} is the observed exposure for age x and year t and \hat{m}_{xt} denotes the fitted or forecast central death rate using a given mortality model.

Measures introduced in Section 5.8 are calculated using the log death rate $(\log(m))$. However, the WT and JWT models use the drifts of z-scores (λ), and the PLC model uses the logit death rate $(\log it(q))$. Equations (3.2) to (3.7) illustrate the transform between λ and q. The transform between m and q (Wilmoth et al., 2007) is

$$q_x = \frac{m_x}{1 + (1 - \zeta_x)m_x} , \qquad (5.1)$$

where the subscripts t and i are omitted. The ζ_x is the average number of years lived in the age interval [x, x + 1) for people dying at that age, and $\zeta_x = 0.5$ for $1 \le x \le 109$ and
$\zeta_x = 1/m_x$ for $x \ge 110$. If the forecasting horizon is beyond recorded data, ζ_x in year n is used for the transform between the forecast q_x and m_x .

For x = 0, ζ_0 can be estimated as (Coale et al., 1983; Preston et al., 2001):

$$\begin{cases} \zeta_0^F = 0.350 , \quad \zeta_0^M = 0.330 , \\ \zeta_0^F = 0.053 + 2.800 m_0^F , \quad \zeta_0^M = 0.045 + 2.684 m_0^M , \\ m_0 < 0.107 \end{cases}$$

where the superscripts F and M denote females and males. For sex combination T,

$$\zeta_0^T = \frac{\zeta_{0ti}^F d_0^F + \zeta_0^M d_0^M}{d_0^F + d_0^M} \; .$$

5.7 Bayesian information criterion

The Bayesian information criterion (BIC) is a standardized measure used to evaluate models' performance. The BIC reflects the difference between observation and estimation, considering the penalty of number of parameters. This research applies the BIC as:

$$BIC = \mathbb{L} + \mathbb{K}\log(\mathbb{N}),$$
 (5.2)

where $\mathbb{L} = -2L$ and L denotes the log-likelihood introduced by Brouhns et al. (2002):

$$L = \sum_{xti} \left\{ d_{xti} \log \left(e_{xti} \hat{m}_{xti} \right) - e_{xti} \hat{m}_{xti} - \log(d_{xti}!) \right\} , \qquad (5.3)$$

where $d_{xti}!$ denotes the product of d_{xti} . Since $\log(\hat{m}_{xti})$ and $-e_{xti}\hat{m}_{xti}$ are negative, L is negative and \mathbb{L} is positive. The \mathbb{N} denotes the number of observations, $\mathbb{N} = N_x \times N_t \times N_i$, where N_x , N_t and N_i denote the number of ages, years and populations. The \mathbb{K} denotes the number of parameters in a model for a group of populations, calculated via Table 5.4.

	Decuon	3.2.1	4	3.2.3	3.2.4	3.2.4	3.2.4	3.2.4	3.2.4	3.2.5		376	0.4.0		207	7.2.1	3.2.7	3.3.1	3.3.1	3.3.2	3.3.3	3.3.4	3.1	3.1	4
	Comment							*1		*2		*3	*3	*3	*3	*3	*3			*4		*3			*5
Ш.	VI VI	N_i	$(N_x + N_t)N_i$		$(0 M \perp M) M$	$\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$		$(2N'_x + N_t)N_i$	$(3N_x+2N_t)N_i$	$\sum_i (N_x + \mathcal{O}_i N_t + \mathcal{O}_i N_x)$	$2N_tN_i$	$(2N_t+N_c)N_i$	$(3N_t+N_c)N_i$	$(2N_t+N_c)N_i$	$(3N_x + N_t + N_c)N_i$	$(N_x + N_t + N_c)N_i$	$(N_x + 3N_t + N_c)N_i$	$N_x N_i + N_x + N_t$	$(2N_x + N_t)N_i + N_x + N_t$	$N_x N_i + N_x N_{\Psi_F} + N_t N_{\Psi_G} + N_i N_{\Psi_J}$	$2N_x + N_t + N_i$	$(N_x + N_t + N_c)N_i$	$2N_x + N_t$	$2N_xN_i + N_t$	$N_x + N_t'$
Pop	$ ho_i$																			N_i	N_i	N_i		N_i	N_i
	С																	t 2	t 2		t 2	2	$_t$ 2	t 1	t 2
mmon	$\frac{k_t}{k}$																	x N	x	. x	x		x N	N_{i}	N
C	FactorModelIndividualCommonPopK a_{xi} b_{xi} k_{ti} γ_{ci} \mathcal{C} a_x b_x k_t \mathcal{C} a_{xrr} a_{xrr} a_x b_x k_t \mathcal{C} a_x b_x b_x																	N	N	$\bigvee_x N$	$\bigvee_x N$		$V_x N$		V_x
			5					5	4	\mathcal{O}_i	5	2	3	1	6	2	3		5				I	1	I
	$\forall ci$									2		V_c	V_c	V_c	V_c	V_c	V_c					V_c			
dual	ti		V_t		7.	2,		V_t	V_t	N_t	V_t	V_t]	V_t]	V_t 1	V_t 1	V_t i	V_t]		V_t			V_t 1			
Indivi	k				ĸ	F			2	r	2	2^{1}	ŝ	2		Z	37					V			
	b_{xi}				N.	$\tau \mathbf{v} t$		N'_x	$2N_x$	${\cal O}_i N_i$					$2N_x$				N_x					N_x	
	a_{xi}		N_x		N	x v v		N'_x	N_x	N_x					N_x	N_x	N_x	N_x	N_x	N_x		N_x		N_x	
Madal	Model	WT	WTAP	ГC	ΓM	BMS	BDV	DT	ASE	ΗU	CBD	CBD1	CBD2	CBD3	RH	Currie	Plat	SJLC	JLC	TLC	PLC	CBDCK	CLC	ILC	JWT
	ractor	Zero				Two								Three						Joint					

nodels
f the 1
eters o
oarame
of Į
Numbers
4.
Table 5

* 1 N'' is the number of knots of the age sequence.
* 2 O_i are the orders of the HU model used for a population i.
* 3 N_c = N_x + N_t - 1 denotes the length of the cohort sequence.
* 4N_{Ψ_F} = F, N_{Ψ_G} = G and N_{Ψ_J} = J are the three dimensions of the core cube, Ψ.
* 5 N'_t = N_t - 1 is for the WT and JWT models.

The number of parameters of the TLC model relies on the dimensions of the core cube, Ψ . In Tucker method, the dimensions of Ψ are determined by trials of experiments (Kiers and Kinderen, 2003). This research applies F = G = J = 1 to Ψ for two reasons. Firstly, the $1 \times 1 \times 1$ of Ψ produces the lowest MAE in the forecasting period. With trials from 1 to 5 applied to F, G and J, the increase of F, G or J results in a decrease of forecast accuracy. Secondly, when G > 1, the TLC model has multiple time trends and thus has difficulties in forecasting.

Table 5.5 displays the number of parameters (\mathbb{K}), according to the data introduced in Section 5.3. The \mathbb{K} denotes the number of parameter used by a model for a group. In this case, a group means females (or males) of the 13 countries in Section 6.2 or the combined sex (females and males) within a country in Section 6.3. Across the eight models, the WT model uses the fewest parameters. As a non-specific model, the number of parameters in the WT model depends on the number of populations. The JWT model uses the second fewest parameters. By assuming populations to share the same trend in mortality decline, the JWT model has the same number of parameters (136) in both analyses.

Evolution		Individual	-			Joint		
Evaluation	WT	LC	HU	SJLC	JLC	TLC	PLC	JWT
Section 6.2	13	2,951	*	1,308	3,088	1,320	240	136
Section 6.3	2	454	*	317	591	319	229	136
Ranking	8	3	1	5	2	4	6	7

Table 5.5: Number of parameters (\mathbb{K}) of the selected models

* See Table 5.6.

Table 5.6 illustrates the details of \mathbb{K} used by the HU model for individual populations. The \mathbb{K} is 3,636 for females and 4,458 for males in the "across country" analysis (Section 6.2). The \mathbb{K} in the "across sex" analysis (Section 6.3) is from 454 to 1,002 and displayed as the bold figures in the "Overall".

	Country	AUS	CAN	ENW	FIN	FRA	ITA	ESP
s	Order	4	2	1	1	1	1	1
lale	\mathbb{K}	638	364	227	227	227	227	227
fem	Country	SWE	CHE	USA	JPN	DNK	NLD	Total
щ	Order	1	1	1	2	1	1	
	\mathbb{K}	227	227	227	364	227	227	3,636
	Country	AUS	CAN	ENW	FIN	FRA	ITA	ESP
	Order	2	3	3	3	3	3	1
lles	\mathbb{K}	364	501	501	501	501	501	227
M	Country	SWE	CHE	USA	JPN	DNK	NLD	Total
	Order	1	1	1	1	1	1	
	K	227	227	227	227	227	227	4,458
-	Country	AUS	CAN	ENW	FIN	FRA	ITA	ESP
eral	K	1002	865	728	728	728	728	454
Ő	Country	SWE	CHE	USA	JPN	DNK	NLD	Total
	K	454	454	454	591	454	454	8,094

Table 5.6: Number of parameters (\mathbb{K}) used by the HU model for individual populations

5.8 Forecast accuracy: mean error and mean absolute error

Forecast accuracy reflects the difference between the observed and the forecast mortality and is measured by mean error (ME) and mean absolute error (MAE). The ME reflects underestimation or overestimation. The MAE illustrates the size of the deviation.

5.8.1 Mean error

The error is the difference between observed and forecast mortalities:

$$\varepsilon_{xti} = \log(m_{xti}) - \log(\hat{m}_{xti}) , \qquad (5.4)$$

where ε_{xti} is the difference between the observed $(\log(m_{xti}))$ and the forecast $(log(\hat{m}_{xti}))$ log death rates. Negative ε_{xti} indicates overestimation and positive ε_{xti} indicates underestimation. The following averages of ε_{xti} provide different dimensions of the ME:

$$\varepsilon_{\cdot ti} = \frac{1}{m} \sum_{x} \varepsilon_{xti} , \qquad \varepsilon_{\cdot \cdot i} = \frac{1}{mn} \sum_{x,t} \varepsilon_{xti} , \qquad (5.5)$$

where $\varepsilon_{\cdot ti}$ denotes the ME over time for population *i* and $\varepsilon_{\cdot i}$ denotes the overall ME for population *i*. The averaged MEs of a model across populations are as follows:

$$\varepsilon_{\cdot t \cdot} = \frac{1}{\ell} \sum_{i} \varepsilon_{\cdot t i}, \qquad \varepsilon_{\cdot \cdot \cdot} = \frac{1}{\ell} \sum_{i} \varepsilon_{\cdot \cdot i}.$$
 (5.6)

5.8.2 Mean absolute error

Corresponding to (5.5), the MAEs are as follows:

$$\tilde{\varepsilon}_{\cdot ti} = \frac{1}{m} \sum_{x} |\varepsilon_{xti}|, \qquad \tilde{\varepsilon}_{\cdot \cdot i} = \frac{1}{mn} \sum_{x,t} |\varepsilon_{xti}|,$$
(5.7)

where $|\varepsilon_{xti}|$ is the absolute value of ε_{xti} , $\tilde{\varepsilon}_{ti}$ denotes the MAE over time for population *i* and $\tilde{\varepsilon}_{..i}$ denotes the overall MAE for population *i*. The averaged MAEs of a model across populations are as follows:

$$\tilde{\varepsilon}_{\cdot t} = \frac{1}{\ell} \sum_{i} \tilde{\varepsilon}_{\cdot ti} , \qquad \tilde{\varepsilon}_{\cdots} = \frac{1}{\ell} \sum_{i} \tilde{\varepsilon}_{\cdots i} . \qquad (5.8)$$

5.8.3 Comparative mean absolute error

The comparative MAE (CMAE) reflects the percentage increase (decrease) in the MAE compared with a "base" model – in this research, the LC model:

$$\varrho_i^{(j)} = \frac{\tilde{\varepsilon}_{\cdot\cdot i}^{(j)} - \tilde{\varepsilon}_{\cdot\cdot i}^{(LC)}}{\tilde{\varepsilon}_{\cdot\cdot i}^{(LC)}} \times 100 , \qquad (5.9)$$

where $\varrho_i^{(j)}$ denotes the CMAE of model *j*. Positive (negative) $\varrho_i^{(j)}$ indicates that model *j* has higher (lower) MAE expressed as a percentage than the LC model.

5.9 Test of significant differences between models

The test of significant differences examines the differences of forecast accuracy between two models, by using *t*-test to compare $\tilde{\varepsilon}_{xt}^{(j)}$ for different *j*:

$$\tilde{\varepsilon}_{xt\cdot}^{(j)} = \frac{1}{\ell} \sum_{i} \tilde{\varepsilon}_{xti}^{(j)} , \qquad (5.10)$$

where j denotes the models selected in Section 5.2 and $\tilde{\varepsilon}_{xti}^{(j)}$ is the absolute error yielded by model j. For example, the differences of forecast accuracy between models 7 and 8 can be measured by applying *t*-test to two matrices with entries $\tilde{\varepsilon}_{xt}^{(7)}$ and $\tilde{\varepsilon}_{xt}^{(8)}$.

To test the differences of forecast accuracy between two groups of models, *t*-test is used to compare $\tilde{\varepsilon}_{xt}^{(g)}$ for different *g*:

$$\bar{\tilde{\varepsilon}}_{xt\cdot}^{(g)} = \frac{1}{N_g} \sum_j \tilde{\varepsilon}_{xt\cdot}^{(j)} , \qquad (5.11)$$

where g denotes the group and N_g denotes the number of models included in group g. Let g = 1, 2 denotes individual models (j = 1, 2, 3) and joint models (j = 4, 5, 6, 7, 8):

$$\bar{\tilde{\varepsilon}}_{xt\cdot}^{(1)} = \frac{1}{3} \sum_{j=1}^{3} \tilde{\varepsilon}_{xt\cdot}^{(j)} , \qquad \bar{\tilde{\varepsilon}}_{xt\cdot}^{(2)} = \frac{1}{5} \sum_{j=4}^{8} \tilde{\varepsilon}_{xt\cdot}^{(j)} ,$$

where $\bar{\tilde{\varepsilon}}_{xt.}^{(1)}$ and $\bar{\tilde{\varepsilon}}_{xt.}^{(2)}$ are the averages of $\tilde{\varepsilon}_{xt.}^{(j)}$ across individual models and across and across joint models. To verify the differences of forecast accuracy between individual and joint models, the *t*-test is used to examine the two matrices with entries $\bar{\tilde{\varepsilon}}_{xt.}^{(1)}$ and $\bar{\tilde{\varepsilon}}_{xt.}^{(2)}$.

5.10 Convergence properties of mortality across populations

Convergence properties of mortality across populations are measured by the differences in forecast mortalities across populations. This research uses the standard deviation of forecast $\log(m)$ across populations to reflect the differences in the forecast mortality:

$$\vartheta_{xt} = \sqrt{\frac{1}{\ell - 1} \sum_{i} \left(\log(\hat{m}_{xti}) - \overline{\log(\hat{m}_{xt\cdot})} \right)^2}, \quad \overline{\log(\hat{m}_{xt\cdot})} = \frac{1}{\ell} \sum_{i} \log(\hat{m}_{xti}), \quad (5.12)$$

where ϑ_{xt} denotes the difference of forecast log death rate across populations for age xin year t, $\log(\hat{m}_{xti})$ denotes the forecast log death rate and $\overline{\log(\hat{m}_{xt})}$ is the average of forecast log death rate across i for the age x and the year t.

5.11 Rate of mortality decline

Rate of mortality decline (ROMD) reflects an averaged annual change in log death rate across ages and over years:

$$ROMD_i = -\frac{1}{m(n-1)} \sum_{x,t} \left\{ \log(m_{xti}) - \log(m_{x,t-1,i}) \right\} .$$

where $\log(m_{xti}) > \log(m_{x,t-1,i})$ in the long term, and thus $\sum_{x,t} \{\log(m_{xti}) - \log(m_{x,t-1,i})\}$ is negative. The sign "-" ensures a positive ROMD to simplify the comparison. Replacing $\log(m_{xti})$ by $\log(\hat{m}_{xti})$, ROMD represents the average annual rate of fitted (or predicted) log death rate across ages and over year. A large ROMD indicates that the log death rate is declining quickly.

Chapter 6

Evaluating the Fits and Forecasts of Individual and Joint Mortality Models

6.1 Introduction to the evaluation and the main findings

This chapter evaluates and compares a variety of individual and joint models. The evaluation aims to address shortcomings in the existing literature where, to date, there has been limited evaluation and comparison of joint models. Generally, it is shown that the performance of the JWT model is superior to that of the other considered models.

As discussed in Section 3.4, two of the major shortcomings of the existing literature are the lack of evaluation of forecast accuracy of joint models and the lack of comparison across these same joint models. Addressing these two shortcomings is necessary for two reasons. Firstly, new models aim to have advantages over existing models. Therefore, the new models must be compared with existing models to demonstrate advantages or disadvantages. Secondly, mortality models are used for projection. Therefore, forecast accuracy is a critical component of a model's performance. Evaluation and comparison of fitting and forecasting are conducted using annual data for the period 1948-2009. The period is divided into two parts. The first part, from 1948 to 1994, is used as the fitting period. The second part, from 1995 to 2009, is used as the forecasting period. The fitting period is used to estimate the parameters and assess the goodness of fit. Goodness of fit reflects the difference between observations and estimates of the number of deaths in the fitting period, with relatively small differences indicating a good fit and an appropriate model. The forecasting period is used to assess forecasting performance. Forecast accuracy is measured by the error between observed and forecast mortality (the log death rate), with a small error (in absolute value) indicating a good forecast and a possibly appropriate and/or robust model.

Convergence properties of mortality forecasts across populations are also evaluated. Mortalities of different developed countries around the world have common trends and have tended to become more similar as described in Section 2.3, referred to as convergence. In the long run, differences in mortality across populations are expected to be reasonably constant or declining. Constant differences or declines in forecast therefore suggest appropriate models. Convergence properties of mortality across countries is a long-term assessment, and the forecasting period of 15 years (1995-2009) may not be sufficiently long. Hence, to demonstrate whether a model yields convergent or divergent long-run forecasts, the forecasting period is extended from 15 years to 50 years (1995-2044) to evaluate convergence properties.

This chapter presents the evaluation of eight models:

- The three "individual" models: WT, LC, and HU
- The five "joint" models: SJLC, JLC, TLC, PLC and JWT

Section 5.2 outlines the reasons for selection of these models. The evaluation includes an assessment of how these eight models perform in the fitting period, as well as an assessment of accuracy and convergence properties in the forecasting period.

The evaluation demonstrates the advantages of the JWT and other joint models based on the comparison of forecasting performance. Firstly, the JWT model produces the best forecast accuracy of the eight models. Secondly, joint models have better accuracy than individual models in forecasting. Thirdly, the SJLC, JLC, PLC and JWT models yield constant differences in mortality across populations in the long run, and therefore, produce forecasts with favourable convergence properties.

The eight models are applied to female and male mortality of the 13 developed countries selected in Section 5.3: Australia (AUS), Sweden (SWE), Canada (CAN), Switzerland (CHE), England and Wales (ENW), the United States of America (USA), Finland (FIN), Japan (JPN), France (FRA), Denmark (DNK), Italy (ITA), the Netherlands (NLD) and Spain (ESP). To evaluate the models on a consistent basis, parameters of the selected models are estimated using the mortality data for years 1948-1994 and ages 0-89, and the forecasting periods are 1995-2009 for measuring forecast accuracy and 1995-2044 for measuring convergence properties.

Joint models are applied in two different ways – pooling across countries ("across country", Section 6.2) for each sex, or pooling across sex ("across sex", Section 6.3) for each of the 13 countries. The "across country" approach therefore consists of two analyses – females and males – each combining 13 countries, permitting common trends across countries for females, and then for males. The "across sex" approach consists of 13 analyses each combining the two sexes, allowing common trends for females and males and considering each country independently.

Goodness of fit is measured by the Bayesian information criterion (BIC) as introduced in Section 5.7. The BIC is a standard measure used to evaluate models' performance and is built up from the log-likelihood and the number of parameters. The log-likelihood reflects the match between observations and estimations in the fitting period. The BIC penalises the log-likelihood for the number of parameters when assessing a model's performance.

Forecast accuracy refers to "out of sample" forecasting accuracy. The "out of sample" period in this case is 1995-2009. In this chapter, forecast accuracy is measured by mean error (ME) and mean absolute error (MAE) as described in Section 5.8, with averaging over ages, time and populations. The ME can be positive, negative or zero, indicating that the forecast mortality is below (underestimation), above (overestimation) or equal to observed mortality. This research also uses comparative MAE (CMAE), as described in Section 5.8, to illustrate the percentage increase or decrease in the MAE of a model as compared to the LC model benchmark.

The overall ME, MAE and CMAE reflect overall forecast accuracy. These measures average over populations, ages and time points. Each measure can be decomposed by averaging over, for example, just populations and ages in which case a time series indicating the trend of the statistic over time is produced. For example, the ME over time reflects whether the underestimation or overestimation is stable or is getting worse. The MAE over time reflects whether a model's forecast accuracy is limited to certain time points or extends over the whole range of time. We could similarly average over other dimensions to gain an understanding of behaviour over different populations or ages.

To analyse the differences in projected mortality across populations, this research uses the standard deviation of mortality across populations as described in Section 5.10. If a model forecasts an increasing standard deviation over time, the model admits divergence in projection. Individual models project individual mortality trends over time and therefore generally produce increasing differences in mortality across similar populations in the long run (Li and Lee, 2005). Joint models, on the other hand, allow populations to share a more or less stringent common mortality trend and aim to reduce or control any difference. Joint models are expected to have smaller divergence than individual models.

6.2 "Across country" analysis, separately for females and males

This section illustrates the evaluation of the eight models to fit and forecast mortality of the 13 countries. The joint models are separately applied to females and males, with the combined country experience used for the common experience.

This section demonstrates the advantages of the JWT model and other joint models in jointly forecasting mortality. Using BIC, in spite of the penalty accorded to the number of parameters, models with more parameters generally fit better than those with fewer parameters. Individual models are found to fit better than joint models, and joint models incorporating more individual factors fit better than those with fewer individual factors. These findings apply to both females and males (Section 6.2.1).

However, joint models generally deliver superior forecasts than individual models (Sections 6.2.2 and 6.2.3) and the JWT model – with relatively few parameters – has the best forecast accuracy in projection across the eight models. The JWT model allows for a flexible rate of mortality decline over time where the flexibility is accorded not through the number of parameters but rather through the functional form of the parameterizations (i.e. the z–score). Joint models generally produce non-divergent forecasts whereas individual models produce divergent forecasts.

Although individual models use more parameters and experience better goodness of fit, the joint models – particularly, the JWT model – behave better in the forecast accuracy and divergence properties.

6.2.1 Goodness of fit according to BIC

Table 6.1 illustrates the overall fit of the models. The BIC and log-likelihood (\mathbb{L}) are for the group of 13 countries. A small \mathbb{L} or BIC indicates good performance and results in a high ranking. For example, the LC model has the smallest BIC (0.81) for females and is ranked first, indicating that the LC model performs the best of the eight models according to BIC. Note that the WT and JWT model have almost the same \mathbb{L} and BIC, due to the small number of parameters used.

	Models		Individua	1			Joint									
	widdels	WT	LC	HU	SJLC	JLC	TLC	PLC	JWT							
K		13	2,951	*	1,308	3,088	1,320	240	136							
	N				$90 \times 47 \times 13$											
les	$L(10^{6})$	1.57	0.78	0.86	2.19	0.93	1.47	3.61	2.31							
ma	$BIC(10^{6})$	1.57	0.81	0.90	2.23	0.96	1.48	3.61	2.31							
Fe	Ranking	5	1	2	6	3	4	8	7							
es	$L(10^{6})$	2.13	0.93	0.98	2.27	0.98	1.71	3.91	2.45							
1al	$BIC(10^{6})$	2.13	0.96	1.03	2.30	1.02	1.73	3.91	2.45							
4	Ranking	5	1	3	6	2	4	8	7							
all	$\mathbb{L}(10^{6})$	1.85	0.85	0.92	2.23	0.96	1.59	3.76	2.38							
ver	$BIC(10^{6})$	1.85	0.89	0.96	2.26	0.99	1.61	3.76	2.38							
0	Ranking	5	1	2	6	3	4	8	7							

Table 6.1: BIC and log-likelihood of the models in the fitting period

Note:

- The number of parameters, K, is calculated via formulae displayed in Table 5.4.
- The \mathbb{K} of the HU model is 3,636 for females and 4,458 for males.
- The number of observed data, points \mathbb{N} , is described in Section 5.3: ages 0-89 (90) and years 1948-1994 (47) and the selected countries (13). Therefore, $\mathbb{N} = 90 \times 47 \times 13$.
- The \mathbb{L} is calculated according to (5.3): $\mathbb{L} = -2 \sum_{xti} \left\{ d_{xti} \log \left(e_{xti} \hat{m}_{xti} \right) - e_{xti} \hat{m}_{xti} - \log(d_{xti}!) \right\}.$
- BIC is calculated according to (5.2): $BIC = \mathbb{L} + \mathbb{K} \log(\mathbb{N})$.

Note the following highlights in Table 6.1; further explanations are shown in subsequent sections.

- The BIC rankings of the models are almost the same between females and males, indicating that the relative performance of the model is consistent across both populations. Despite the number of parameters penalising goodness of fit using BIC, models with more parameters generally perform better than those with fewer parameters.
- 2) The general finding that models with more parameters have a lower BIC does not always apply. Exceptions are the HU (versus LC), JLC (versus LC) and PLC (versus WT) models. The HU, JLC and PLC models use more parameters but yield worse fit than the comparison models displayed in the brackets. Additional parameters usually, but do not always, result in better fit.

Models perform consistently in different populations – models with more parameters fit better than those with fewer parameters

The ranking of the eight models is almost the same for females and males. Females and males are biologically different and the same ranking of a model between females and males indicates that the models perform relatively consistently in different populations. The number of parameters significantly influences a model's fit. A decrease in the number of parameters is generally associated with an increase in BIC across the LC, HU, JLC, TLC, SJLC, PLC, JWT and WT models.

Within the individual models, the WT, LC and HU models span the range from a very small to a very large number of parameters. There is an inverse relationship between the number of parameters and BIC, although the HU model has more parameters and a larger BIC than the LC model.

Within the joint models, the number of parameters is not as clearly related to BIC. A joint model may involve both individual and joint factors. Individual (or joint) factors refer to the parameters used by a model to describe mortality of an individual population (or all member populations). When the joint factors are inappropriate or the model is not suitable, a model with more parameters may not perform better than one with fewer parameters. The JLC and PLC models are two examples of this case.

Additional parameters usually, but do not always, result in better fits

Models use parameters to describe the features of mortality rates. Generally, adding parameters enables a model to better describe the mortality rates. A model with more parameters generally fits better than a model with fewer parameters, based on the BIC.

However, specific models may not follow this rule. For example, the HU model has more parameters but a worse BIC than the LC model. Parameters of the LC model are estimated using the log death rate, while those of the HU model are estimated via the smoothed log death rate of ages. The HU model experiences additional bias between the estimated and smooth rates, and thus yields a worse BIC than the LC model.

Reviewing two pairs of models – the LC and JLC models and the WT and PLC models, additional common factors may not result in better fit. Recall the fomulae for – LC (3.8) and JLC (3.35), WT (3.1) and PLC (3.44):

$$\begin{cases} \log(m_{xti}) &= a_{xi} + b_{xi}k_{ti} + \epsilon_{xti} ,\\ \log(m_{xti}) &= a_{xi} + b_{xi}k_{ti} + b_{x}k_{t} + \epsilon_{xti} ,\\ \end{bmatrix} \begin{cases} \lambda_{xti} &= \rho_i + \epsilon_{xti} ,\\ \log it(q_{xti}) &= a_x + b_xk_t + \rho_i + \epsilon_{xti} \end{cases}$$

The additional common factors $(b_x k_t \text{ and } a_x + b_x k_t)$ do not enhance the fits of the JLC and PLC models. This may be due to the estimation of the common factors not being robust from the aggregated mortality data sets.

Furthermore, inappropriate application may influence a model's performance. The

PLC model assumes that age-specific logit death rates, $logit(q_{xti})$, are parallel between populations *i*. The assumption may be reasonable for geographically close populations, for example, regions in Spain, which is the purpose for which this model was developed (Debón et al., 2011). However, this assumption may not be generalised across countries. Using the PLC model, Figure 6.1 displays the observed and fitted average $logit(q_{xti})$ over *t* (1948-1994) for the USA and Japan. Solid curves (observed) are not parallel between the USA and Japan, but dashed curves (fitted) are parallel between these two countries. Therefore, the PLC model imposes an unreasonable parallel relationship for the fitted rates, and thus produces the worst fits of the eight models.



Figure 6.1: Average logit (q_{xti}) over t (1948-1994) of the USA and Japan

6.2.2 Overall forecast accuracy

Forecast accuracy reflects the difference between the observed and the forecast of log death rate produced by a model, and is measured by the ME, MAE and CMAE as described in Section 5.8. The ME indicates whether forecast mortality is higher or lower than the observed mortality in the forecasting period (1995-2009). A value of ME close

to zero indicates an overall small forecast bias and a high ranking. A small ME can arise from an averaging of large positive and large negative values. The MAE reflects the absolute departure of forecast mortality from observed mortality in 1995-2009. A small overall MAE indicates a good forecast accuracy and a high ranking. The CMAE describes the percentage of difference between a model and the LC model. A model with negative CMAE is better than the LC model, while a model with a positive CMAE is worse than the LC model.

Table 6.2 illustrates the overall ME, MAE and CMAE of the 13 countries, separately for females and males.

	Models		Individua	.1	Joint									
	Models	WT	LC	HU	SJLC	JLC	TLC	PLC	JWT					
	K	13	2,951	3,636	1,308	3,088	1,320	240	136					
	ME	0.081	0.018	-0.007	0.045	0.099	0.023	0.021	-0.004					
les	Ranking	2	6	8	4	1	7	5	3					
ma	MAE	0.175	0.158	0.145	0.151	0.183	0.148	0.143	0.142					
Fe	Ranking	7	6	3	5	8	4	2	1					
	CMAE(%)	10.86	0.00	-8.35	-4.25	15.53	-6.67	-9.57	-10.21					
	ME	-0.025	-0.077	-0.105	-0.066	-0.020	-0.079	-0.077	-0.037					
es	Ranking	2	6	8	4	1	7	5	3					
1al(MAE	0.153	0.176	0.172	0.156	0.155	0.163	0.159	0.147					
2	Ranking	2	8	7	4	3	6	5	1					
	CMAE(%)	-12.71	0.00	-1.76	-11.02	-11.84	-7.03	-9.63	-16.04					
	ME	0.028	-0.030	-0.056	-0.011	0.039	-0.028	-0.028	-0.020					
all	Ranking	5	6	8	1	7	4	3	2					
ver	MAE	0.164	0.167	0.159	0.154	0.169	0.155	0.151	0.145					
0	Ranking	6	7	5	3	8	4	2	1					
	CMAE(%)	-1.54	0.00	-4.88	-7.81	1.13	-6.86	-9.60	-13.27					

Table 6.2: ME, MAE and CMAE of the forecast log death rate

Note:

- The number of parameters, \mathbb{K} , is the same as shown in Table 6.1.
- The K of the HU model is 3,636 for females and 4,458 for males.
- The full information of ME, MAE, CMAE and the rankings of MAE of each of the 13 countries is shown in Tables B.1 to B.10 in Appendix B.
- The "Overall" means the average of females and males.

Note the following highlights in Table 6.2; explanations are given below the highlights.

- 1) The JWT model has the smallest MAE for both females (0.142) and males (0.147), and provides the most reliable forecasts for both females and males.
- 2) The HU model has the most parameters (3,636 for females and 4,458 for males), but relatively poor forecast accuracy (fifth overall according to MAE). The HU model may involve more than one time factor and these additional time factors are difficult to forecast. As a result, the HU model may be unreliable to forecast mortality for specific populations.
- 3) The WT model has different rankings between females (7) and males (2). Since the WT model uses one parameter to describe mortality for all ages over time, the unique parameter results in very constrained and inflexible forecasting.
- 4) The JLC model performs the worst of the eight models for females (though third for males). The SJLC model excludes k_{ti} and performs better than the JLC model, and thus k_{ti} appears to reduce the forecast accuracy. The JLC model may be unreliable for short-term projection.
- 5) Overall, simple models perform better than complex models. The JWT, TLC, PLC and SJLC models use fewer parameters but obtain better forecast accuracy than the LC, HU and JLC models. Using fewer parameters may avoid over-parameterization and reduce the risk of inaccurate forecasts.
- 6) Six of the eight models yield positive ME for females and all eight models produce negative ME for males. Females experience a smaller rate of mortality decline (ROMD) in the forecasting period than in the fitting period, while males experience a larger ROMD in the forecasting period. Most models, therefore, underestimate mortality of females and overestimate that of males.

7) The average of overall MAE across joint models is 0.153 for females and 0.156 for males, smaller than those – 0.159 for females and 0.169 for males – of individual models. In projection, the joint models perform better than the individual models, using the *t*-test to evaluate the significance of the difference between individual and joint models (Section 5.9).

The advantages of the JWT model when forecasting mortality

The JWT model performs best when forecasting. This may be due to one or more of three reasons. Firstly, the JWT model is a joint model, and thus incorporates multiple similar populations. By aggregating similar populations, the JWT model exploits and capitalises on the relationships of mortality between populations. Aggregating populations also enables the JWT model to make use of more information and to give a more reliable forecast.

Secondly, the JWT model is flexible in projection. The sequence k_t , as described in Section 4.4, adjusts the drift of z-scores in the short term. Therefore, the JWT model permits z-scores to flexibly vary over time. In contrast, the other seven models have a lack of flexibility. For example, the WT model assumes that the z-scores increase with a constant drift, and the LC model assumes that the log death rates decline linearly. The SJLC, JLC, TLC and PLC models are joint models, but do not allow a flexible ROMD. Therefore, the JWT model is more flexible and performs more accurately than these joint models.

Table 6.3 displays the rate of mortality decline (ROMD) in the log death rate as introduced in Section 5.11. For the "observed" column, ROMD is calculated using observed log death rate. For the models columns, ROMD is calculated using the fitted and forecast log death rate by the relevant models.

	/T	FCA	0.022	0.022	0.022	0.022	0.022	0.022	0.022	0.022	0.022	0.022	0.022	0.022	0.022	0.022	۲ ۲	0.023	0.023	0.023	0.023	0.023	0.023	0.023	0.023	0.023	0.023	0.023	0.023	0.023	0.023	
Observed WT LC HU SJLC JLC TLC PLC JWT FIT FCA FIT <td>0.025</td> <td></td> <td>0.018</td> <td></td>	0.025	0.025	0.025	0.025	0.025	0.025	0.025	0.025	0.025	0.025	0.025	0.025	0.025		0.018	0.018	0.018	0.018	0.018	0.018	0.018	0.018	0.018	0.018	0.018	0.018	0.018	0.018				
	Ų	FCA	0.024	0.024	0.024	0.024	0.024	0.024	0.024	0.024	0.024	0.024	0.024	0.024	0.024	0.024	^	0.017	0.017	0.017	0.017	0.017	0.017	0.017	0.017	0.018	0.017	0.018	0.017	0.017	0.017	
Observed WT LC HU SJLC JLC TLC PLC JWT FIT FCA FIT <td>0.024</td> <td></td> <td>0.017</td> <td></td>	0.024	0.024	0.024	0.024	0.024	0.024	0.024	0.024	0.024	0.024	0.024	0.024	0.024		0.017	0.017	0.017	0.017	0.017	0.017	0.017	0.017	0.017	0.017	0.017	0.017	0.017	0.017				
Observed WT LC HU SJLC JLC TLC PLC JWT FIT FCA FIT <td>0.021</td> <td>0.022</td> <td>0.019</td> <td>0.030</td> <td>0.023</td> <td>0.030</td> <td>0.034</td> <td>0.021</td> <td>0.023</td> <td>0.017</td> <td>0.046</td> <td>0.015</td> <td>0.019</td> <td>0.025</td> <td>\uparrow</td> <td>0.016</td> <td>0.017</td> <td>0.014</td> <td>0.022</td> <td>0.014</td> <td>0.020</td> <td>0.022</td> <td>0.016</td> <td>0.016</td> <td>0.012</td> <td>0.032</td> <td>0.009</td> <td>0.014</td> <td>0.017</td> <td></td>	0.021	0.022	0.019	0.030	0.023	0.030	0.034	0.021	0.023	0.017	0.046	0.015	0.019	0.025	\uparrow	0.016	0.017	0.014	0.022	0.014	0.020	0.022	0.016	0.016	0.012	0.032	0.009	0.014	0.017			
	0.022	0.019	0.030	0.023	0.030	0.034	0.021	0.023	0.017	0.046	0.015	0.019	0.025		0.016	0.017	0.014	0.022	0.014	0.020	0.022	0.016	0.016	0.012	0.032	0.009	0.014	0.017	_			
	0.032	0.033	0.030	0.031	0.031	0.031	0.033	0.033	0.039	0.020	0.040	0.038	0.033	ĸ	0.021	0.023	0.023	0.020	0.023	0.020	0.026	0.025	0.026	0.028	0.011	0.031	0.025	0.023				
	0.023	0.022	0.029	0.025	0.028	0.031	0.023	0.024	0.017	0.044	0.017	0.019	0.025		0.017	0.016	0.015	0.023	0.016	0.019	0.022	0.014	0.016	0.011	0.035	0.009	0.011	0.017				
Observed WT LC HU SJLC JLC TLC PLC JWT FIT FCA FIT <td>0.027</td> <td>\uparrow</td> <td>0.019</td> <td></td>	0.027	0.027	0.027	0.027	0.027	0.027	0.027	0.027	0.027	0.027	0.027	0.027	0.027	0.027	\uparrow	0.019	0.019	0.019	0.019	0.019	0.019	0.019	0.019	0.019	0.019	0.019	0.019	0.019	0.019			
	0.027	0.027	0.027	0.027	0.027	0.027	0.027	0.027	0.027	0.027	0.027	0.027	0.027	0.027		0.019	0.019	0.019	0.019	0.019	0.019	0.019	0.019	0.019	0.019	0.019	0.019	0.019	0.019			
	0.021	0.022	0.021	0.030	0.023	0.028	0.018	0.021	0.023	0.016	0.025	0.015	0.017	0.022	۲	0.017	0.015	0.015	0.024	0.015	0.019	0.012	0.013	0.015	0.011	0.014	0.006	0.012	0.014	1		
	0.021	0.022	0.021	0.030	0.023	0.028	0.032	0.024	0.023	0.016	0.043	0.015	0.017	0.024		0.016	0.015	0.015	0.024	0.015	0.019	0.022	0.014	0.015	0.011	0.033	0.007	0.011	0.017			
	Ŋ	FCA	0.020	0.020	0.019	0.031	0.023	0.031	0.035	0.020	0.022	0.015	0.050	0.010	0.015	0.024	\uparrow	0.016	0.016	0.015	0.026	0.015	0.021	0.021	0.017	0.016	0.011	0.035	0.004	0.014	0.017	
Observed WT LC HU SJLC JLC TLC PLC JW FIT FCA FIT FCA <td>0.020</td> <td>0.020</td> <td>0.019</td> <td>0.031</td> <td>0.023</td> <td>0.031</td> <td>0.035</td> <td>0.020</td> <td>0.022</td> <td>0.015</td> <td>0.050</td> <td>0.010</td> <td>0.015</td> <td>0.024</td> <td></td> <td>0.016</td> <td>0.016</td> <td>0.015</td> <td>0.026</td> <td>0.015</td> <td>0.021</td> <td>0.021</td> <td>0.017</td> <td>0.016</td> <td>0.011</td> <td>0.035</td> <td>0.004</td> <td>0.014</td> <td>0.017</td> <td></td>	0.020	0.020	0.019	0.031	0.023	0.031	0.035	0.020	0.022	0.015	0.050	0.010	0.015	0.024		0.016	0.016	0.015	0.026	0.015	0.021	0.021	0.017	0.016	0.011	0.035	0.004	0.014	0.017			
	T	FCA	0.026	0.030	0.025	0.039	0.034	0.042	0.045	0.027	0.032	0.020	0.061	0.021	0.023	0.033	К	0.020	0.023	0.020	0.032	0.024	0.029	0.033	0.020	0.023	0.015	0.047	0.014	0.016	0.024	
Observed WT LC HU SJLC JLC TLC PLC FIT FCA FIT FCA <td>0.025</td> <td>0.021</td> <td>0.032</td> <td>0.028</td> <td>0.033</td> <td>0.035</td> <td>0.023</td> <td>0.026</td> <td>0.018</td> <td>0.045</td> <td>0.018</td> <td>0.020</td> <td>0.026</td> <td></td> <td>0.017</td> <td>0.019</td> <td>0.017</td> <td>0.026</td> <td>0.020</td> <td>0.023</td> <td>0.026</td> <td>0.017</td> <td>0.019</td> <td>0.013</td> <td>0.036</td> <td>0.013</td> <td>0.014</td> <td>0.020</td> <td></td>	0.025	0.021	0.032	0.028	0.033	0.035	0.023	0.026	0.018	0.045	0.018	0.020	0.026		0.017	0.019	0.017	0.026	0.020	0.023	0.026	0.017	0.019	0.013	0.036	0.013	0.014	0.020				
	trved	FCA	0.025	0.019	0.019	0.019	0.025	0.031	0.030	0.018	0.032	0.014	0.023	0.028	0.023	0.024	ĸ	0.028	0.028	0.025	0.029	0.030	0.037	0.036	0.019	0.039	0.023	0.024	0.036	0.031	0.030	
	Obse	FIT	0.021	0.023	0.021	0.030	0.024	0.029	0.032	0.023	0.025	0.017	0.044	0.016	0.018	0.025		0.016	0.016	0.015	0.024	0.015	0.019	0.022	0.016	0.016	0.011	0.034	0.009	0.012	0.017	_
			AUS	CAN	ENW	FIN	FRA	ITA	ESP	SWE	CHE	USA	Ndſ	DNK	NLD	Mean	Change	AUS	CAN	ENW	FIN	FRA	ITA	ESP	SWE	CHE	USA	Ndſ	DNK	NLD	Mean	-
				Females									I							sə	lal	N						-				

Table 6.3: Rate of mortality decline (ROMD) of observed, fitted and predicted log death rate

Note:

- FIT and FCA denote the fitting (1948-1994) and forecasting (1995-2009) periods.
- Under the models, the rate of mortality decline (ROMD) is calculated using the fitted and predicted mortality.
 - Symbols \searrow , \rightarrow and \nearrow represent the change of ROMD between the fitting and forecasting periods.

From Table 6.3, in the observed column, females have a smaller ROMD in the forecasting period than in the fitting period (0.024 vs. 0.025) and males are the reverse (0.030 vs. 0.017). Since the JWT model permits the ROMD to vary over time as described in Section 4.4, the changed directions of ROMD produced by the JWT model (the arrow notations in Table 6.3) are consistent with those of the observed log death rate while the other seven models assume an expected constant rate of mortality decline. This property enables the JWT model to perform better than the other seven models.

Thirdly, the JWT model has a simpler form than the other models (except for the WT model), incorporating only an age factor and a time factor. Among the eight models, the LC, HU and JLC models have the most parameters and produce the best fit as described in Section 6.2.1. However, a larger number of parameters may result in over-parameterization, and thus increase the unreliability when forecasting mortality. The WT model has too few parameters and thus has a lack of flexibility. The JWT model uses an appropriate number of parameters and keeps a balance between over- and underparameterization. The JWT model, with a simple form, obtains high forecast accuracy.

Relatively weak performance of the HU model

The HU model is ranked fifth according to overall MAE as shown in Table 6.2. Amongst the eight models, the HU model has the most parameters (3,636 for females and 4,458 for males). However, the large number of parameters does not result in superior forecasting performance. This may be due to the following two main factors.

Firstly, the forecast of time trends may not be reliable. Figure 6.2 displays the estimate and forecast of $k_{ti}^{(j)}$ for Australian females with the HU model using an ARIMA time series. For $k_{ti}^{(1)}$, the forecast may be reliable, but the patterns of $k_{ti}^{(2)}$, $k_{ti}^{(3)}$ and $k_{ti}^{(4)}$ are irregular and the forecasts may therefore not be reliable.



Figure 6.2: Estimates and forecasts of time trends $(k_{ti}^{(j)})$ of Australian females with the HU model using ARIMA process

The HU model forecasts well for females (ranked third according to overall MAE), but poortly for males (ranked seventh). Using the ARIMA process to forecast $k_{ti}^{(1)}$ results in a smaller decline than using random walk with drift. From Table 5.2, the HU model uses a ARIMA process to forecast $k_{ti}^{(j)}$ for nine out of the 13 countries. Then, the HU model yields a decrease of ROMD from the fitting period to the forecasting period. The decrease of ROMD happens to match the reality of females, but differs from that of males as can be seen in Table 6.3. Therefore, the HU model forecasts well for females, but poorly for males.

Secondly, the order J – the maximum of j – of $b_{xi}^{(j)}$ and $k_{ti}^{(j)}$ is determined via trials

as described in Section 3.2.5 and as shown in Table 5.2 in Section 5.4. However, since the variation in mortality varies over time, good forecast accuracy in the sub-forecasting period (1980-1994) does not indicate consistent performance in the actual forecasting period (1995-2009). Therefore, the selection of orders may not be reliable.

Potentially reliable forecast of the WT model

The WT model is ranked higher for males (second) than for females (seventh). The different performance between females and males is due to the fact that the WT model always results in an increasing ROMD as can be seen in Table 6.3. The increasing ROMD occurs because the WT model assumes that z-scores increase over time constantly, causing a constant decrease in m_{xt} . Since the logarithm function is a concave function, the constant decrease in m_{xt} results in an accelerating decrease in $\log(m_{xt})$.

Figure 6.3 displays the forecast annual rate of mortality decline (AROMD) in 1995-2009 for Australian females and males using the WT model. Each curve denotes the AROMD for an age group. The top curve in both panels denotes the AROMD of new born babies. The bottom curves in both panels denote the AROMD of older age groups. In spite of the different scales, all age groups experience the same increase in AROMD. When using the WT model to forecast mortality, the improvement of log death rate is not linear but accelerative.

In the long term, mortality of both females and males may decline with a smaller rate in future than at present (1995-2009) (Denison, 2011; Kogan et al., 2011; Richel, 2003), indicating that the increase of male ROMD may not last. Therefore, the WT model matches the mortality improvements of males currently, but may be inappropriate in long-term projection.



Figure 6.3: Annual rate of forecast log death rate decline in 1995-2009 of Australian females and males using the WT model

Unreliable short-term forecast of the JLC model

The JLC model is uncertain in forecasting due to the individual time trends. Figure 6.4 displays the estimates and forecasts of the common time trend, k_t , and the individual time trends, k_{ti} , of the 13 countries for females as an example.



Figure 6.4: Estimates and forecasts of k_t and k_{ti} using the JLC model for females

The solid black curve denotes sequence k_t and other colour curves represent k_{ti} . The vertical grey dashed line separates the horizon into fitting (1948-1994) and forecasting (1995-2009) periods. The k_{ti} sequences are assumed to follow the AR(1) processes as introduced in Section 3.3.1. Therefore, the forecast of k_{ti} converges to zero in the forecasting period.

Another time series ARIMA(1,1,0) is tried to forecast k_{ti} , and gives an overall MAE of 0.152 for females and 0.156 for males. Compared to using the AR(1) process, the ARIMA(1,1,0) process results in better accuracy for females and marginally worse accuracy for males than the AR(1) process. However, the ARIMA(1,1,0) process may cause uncertain projection and result in increasing differences in mortality across populations, referred to as divergence. The AR(1) process may not be appropriate for short-term projection, and the ARIMA(1,1,0) process is not suitable for long-term projection.

Furthermore, in Table 6.3, the JLC model has a larger ROMD in the forecasting period (0.033 for females and 0.023 for males) than in the fitting period (0.023 for females and 0.017 for males). In Figure 6.4, the estimates of k_{ti} sequences are upward for all countries except Japan. Using the AR(1) process, the forecast k_{ti} sequences decline in the forecast-ing period. The decline of k_t and k_{ti} results in an increase of ROMD in the short-term projection. In the long term, the forecast k_{ti} sequences converge at zero, that is at 2030. Then, after 2030, the ROMD will be constant and similar to the ROMD in 1948-1994.

In Table 6.2, the SJLC model has a lower overall MAE than the JLC model (0.151 for females and 0.156 for males) by excluding individual factors. The SJLC model has a simpler format but better forecast accuracy than the JLC model. The individual time trends, k_{ti} , reduce the forecast accuracy at least in the short term.

Advantages of simple models when forecasting mortality

Models of simple form perform better than models of complex form. In Section 6.2.1, models with more parameters have a better fit than those with fewer parameters. On the contrary, Table 6.2 shows that the JWT, TLC, PLC and SJLC models use fewer parameters but obtain better forecast accuracy than the LC, HU and JLC models.

Using a large number of parameters may result in a good fit, but may result in overparameterization and irregular time trends. These two risks lead to unreliable projection. An example is the HU model in this research. Using few parameters may avoid overparameterization and reduce the risk of unreliable projection.

Underestimation of female rates and overestimation of male rates

The average ME across the eight models is 0.034 for females and -0.061 for males. The different signs indicate that the models underestimate the log death rate for females and overestimate that for males. Females and males experience increasingly similar ROMD, referred to as convergence across sex as mentioned by Liu et al. (2012).

The underestimation of female rates and overestimation of male rates are due to the different ROMD between the fitting and forecasting periods. In Table 6.3, the ROMD decreases for females and increases for males from the fitting period to the forecasting periods. Most models use constant ROMD when forecasting mortality, and thus underestimate mortality for females and overestimate that for males.

The LC, SJLC, TLC and PLC models use constant ROMD and yield underestimation. The HU model also uses constant ROMD, but the use of the ARIMA process yields a smaller ROMD in the fitting period than in the forecasting period. Therefore, the HU model overestimates mortality for both females and males.

Significant differences between joint and individual models

From Table 6.2, the average overall MAE across individual models (WT, LC and HU) is 0.163 (0.159 for females and 0.167 for males) and the related value for joint models (SJLC, JLC, TLC, PLC and JWT) is 0.155 (0.153 for females and 0.156 for males). The difference between individual and joint models, although small, is highly significant using a *t*-test (see Section 5.9) for both females and males.¹ Despite the fact that the JLC model is the worst performing of the eight models, the advantages of joint models are apparent and significant (Cairns et al., 2011a; Jarner and Kryger, 2009; Li and Lee, 2005; Li, 2013).

6.2.3 Forecast accuracy over time

This section shows the trends in ME and MAE over the forecasting period and provides more information than the overall ME and MAE to evaluate the models. Section 6.2.2 described the performance in the forecasting period of 15 years, which does not reflect performance in the longer term. An analysis of forecast accuracy over time complements this shortcoming by describing the trends of ME and MAE over time, allowing us to anticipate performance over a longer forecasting period.

The ME over time is introduced in Section 5.8.1 and denoted by $\varepsilon_{\cdot ti}$, reflecting the trends of difference between observed and forecast log death rate of population *i* over *t*. The MAE over time is introduced in Section 5.8.1 and denoted by $\tilde{\varepsilon}_{\cdot ti}$, reflecting the absolute difference between observed and forecast log death rate of population *i* over *t*. The slopes of the fitted straight lines of the curves are defined as the average difference between $\tilde{\varepsilon}_{\cdot ti}$ and $\tilde{\varepsilon}_{\cdot,t-1,i}$, reflecting the trend in MAE in the forecasting period. The trend in MAE is helpful to assess a model's performance in a long-term projection. For example,

¹A *t*-test comparing average MAE across countries and for individual versus joint models (for ages 0-89, years 1995-2009) gives p-value = 0.0015 for females and p-value = 0.0046 for males.

the forecasting period is 1995-2009, while the slope of $\tilde{\varepsilon}_{\cdot ti}$ over t indicates the patterns of MAE after 2009.

Figure 6.5 displays the ME of the forecast log death rate over time of the overall 13 countries, which is the average of $\varepsilon_{\cdot ti}$ across *i*, denoted by $\varepsilon_{\cdot t.}$. Figure 6.6 displays the MAE of the forecast log death rate over time of the overall 13 countries, which is the average of $\tilde{\varepsilon}_{\cdot ti}$ across *i*, denoted by $\tilde{\varepsilon}_{\cdot t.}$. For example, the WT model yields an increasing $\tilde{\varepsilon}_{\cdot t.}$ of the 13 countries over time with a slope 1.12 (×10⁻²) for females.



Figure 6.5: ME of the forecast log death rate over time for females and males



Figure 6.6: MAE of the forecast log death rate over time for females and males 122

Note the following highlights in Figures 6.5 and 6.6 and the explanations given below the highlights.

- 1) The PLC model has nearly constant $\varepsilon_{.t.}$ for females (the left panel of Figure 6.6) and the JWT model has nearly constant $\varepsilon_{.t.}$ for males (the right panel of Figure 6.6), indicating possible use of these models for hedging.
- 2) The JWT model has the smallest slopes for both females (0.58) and males (0.95) among the eight models in Figure 6.6 and would therefore be expected to retain its advantages in a longer-term projection.

Hedging mortality risk using the JWT and PLC models

Mortality derivatives (i.e. q-forward and longevity bonds) rely on the forecast mortality. A model with good forecast accuracy is the best basis for hedging mortality derivatives. In this study, the JWT and PLC models yield constant ME over time for males and females, respectively. The stable bias indicates a predictable hedging risk.

The LC and TLC models have constant ME over time for females (Figure 6.5), but are inappropriate to hedge mortality risks due to the poor forecast accuracy and the divergent projection across populations as described in Section 6.2.4. The divergent projection contravenes the finding of increasing similarities in mortality decline throughout the world.

Reliable long-term projection of the JWT model

The JWT model produces the highest forecast accuracy of the eight models over the 15year projection period. The advantage of the JWT model arises due to its property of flexible ROMD as described in Section 6.2.2. Figure 6.6 illustrates the JWT model's advantage for short-term and medium-term projection – for example, the 15-year projection in this research. This advantage of the high forecast accuracy of the JWT model may also be maintained in the longer-term projection. Information in Figure 6.6 supports the finding in Table 6.2.2 and further illustrates that the JWT model is better in projection than the other seven models.

6.2.4 Convergence properties of mortality across countries

This section illustrates the differences in the forecast mortality across countries, measured by the standard deviation of the forecast log death rate across populations – in this case countries, denoted by ϑ_{xt} as introduced in Section 5.10. Mortalities of countries are converging throughout the world, ϑ_{xt} is expected to be reducing or remaining constant over time.

As described in Section 3.3, individual models permit populations to have individual mortality trends and allow for divergence. Joint models, in contrast, constrain populations to share common mortality trends and do not permit divergence. If a model yields constant or decreasing ϑ_{xt} over t, the model is verified to control the divergence in forecast mortality and is suitable for the convergence properties of mortality across countries.

Figures 6.7 and 6.8 display the standard deviation of the log death rate across the 13 countries for years 1948-2044 and randomly selected ages 0, 10, 20, 30, 49, 59, 69, 79 and 89 for females and males, respectively. Every panel displays the ϑ_{xt} of a model. In the fitting period (1948-1994), the ϑ_{xt} is calculated using the observed log death rate, and thus the eight models have the same patterns of ϑ_{xt} . In the forecasting period (1995-2044), ϑ_{xt} is calculated using the forecast log death rate produced by a specific model.



Note:

- From 1948 to 1994, standard deviation is calculated using observed log death rate.
- From 1995 to 2044, standard deviation is calculated using forecast log death rate.
- Plotted curves are the standard deviation at ages 0, 10, 20, 30, 49, 59, 69, 79 and 89.

Figure 6.7: Standard deviation of log death rate across countries for females



Note:

- From 1948 to 1994, standard deviation is calculated using observed log death rate.
- From 1995 to 2044, standard deviation is calculated using forecast log death rate.
- Plotted curves are the standard deviation at ages 0, 10, 20, 30, 49, 59, 69, 79 and 89.

Figure 6.8: Standard deviation of log death rate across countries for males

Note the following highlights in Figures 6.7 and 6.8 and the explanations given below the highlights.

- In the fitting period (1948-1994), the standard deviations for the younger ages decrease for both females and males.
- 2) In the forecasting period, the WT, LC, HU and TLC models yield increasing standard deviation in the forecast rates of mortality over time. Therefore, individual models (WT, LC, HU) and the TLC model produce apparently unreliable results as their forecasts are likely to diverge, which is inconsistent with preconceptions about the future trajectory of mortality across countries.
- 3) In the forecasting period, the SJLC, PLC and JWT models produce constant standard deviation over time. The SJLC, PLC and JWT models may be reliable for mortality projection, due to the non-divergence across populations.
- 4) In the forecasting period, the JLC model produces, in the short term, and for differing age groups, differing directions in the standard deviation over time. The JLC model is reasonable in long-term mortality forecasting, since this model permits divergence in mortality across populations in short-term projection and keeps a constant difference across populations in long-term projection.

Mortality converges in young age groups

The improvements of mortality for younger ages mainly benefit from medical progress, while the older ages mainly benefit from public health care (Cutler and Meara, 2001; Deaton and Paxson, 2001; Wilmoth, 1998). Both females and males of younger ages benefit from medical progress similarly across countries, and thus experienced similar improvements in mortality. (Mathers et al., 2001; Nolte et al., 2000a,b). Therefore, the standard deviation in log death rate across countries decreases over time for both females

and males of younger ages. In particular, newborn babies in simultaneously benefited from medical factors (Finch, 2010). The infant death rate declined and the differences in the infant death rate across countries decreased (the black curves in Figures 6.7 and 6.8).

Unreliable long-term projection of the TLC model and the individual models

The WT, LC and HU models produce increasing standard deviation during the forecasting period, referred to as divergence. These three models are individual models and allow populations to hold an individual ROMD. The individual ROMD is different across populations and results in increasing differences in mortality across populations. Divergence in forecasts is one of the shortcomings of individual models as discussed in Section 3.3.

The TLC model has a problem in controlling divergence as described in Section 3.3.2, where $k_t \rho_i \approx k_{ti}$. Individual populations effectively have individual time trends, and thus the TLC model causes increasing differences in forecast mortality over time. The JWT and SJLC models do not incorporate population effects and the ρ_i in the PLC model does not influence the common time trend. Therefore, as joint models, the JWT, SJLC and PLC models yield non-divergent forecasts which the TLC model is unable to do.

Reliable long-term projection of the SJLC, PLC and JWT models

Joint models reduce the divergence in forecasting, by applying common trends among the member populations. Figures 6.7 and 6.8 display consistent results with the hypothesis introduced in Section 3.3. The SJLC, PLC and JWT models use one unique time trend, k_t , to forecast mortality for all populations. As a result, member populations have the same path in mortality variation over time. The same rate of mortality decline results in the same difference in mortalities of the 13 countries over time. Consequently, the three joint models have horizontal standard deviation across populations over time.

Reliable long-term projection of the JLC model

The JLC model is a joint model. Compared to the SJLC model, the JLC model involves more parameters – the individual factors, b_{xi} and k_{ti} . The SJLC model produces constant standard deviation of forecasts over time, while the JLC model produces various directions of standard deviation for different ages, due to the individual time trends, k_{ti} .

Figure 6.9 displays the estimates and forecasts of k_{ti} of the JLC model from 1995 to 2044 for females using an AR(1) process. During 1995-2009, k_{ti} sequences are converging and varying over time. As a result, the standard deviations are fluctuating during that period. During 2020-2044, k_{ti} sequences are almost converged and the standard deviation trends to stable over time.



Figure 6.9: Estimates and forecasts of k_{ti} with the JLC model for females

The standard deviations may fluctuate in the short term and be constant in the long term. In the short term, the JLC model permits populations to hold individual mortality trends, whereas in the long term, populations share the common trend. Therefore, the JLC model permits mortality divergence in the short term, but constrains divergence in the long term.

6.3 "Across sex" analysis, separately for each country

This section illustrates the evaluation of the eight models to fit and forecast female and male mortality within each of the 13 countries as shown in Section 5.3. The eight models can be classified as individual (WT, LC and HU) and joint (SJLC, JLC, TLC, PLC and JWT) models. The joint models are applied to jointly fit and forecast mortality, with the combined female and male experience used for the common experience. This section demonstrates the advantages of the JWT model and other joint models in jointly forecast-ing mortality.

Using BIC, models with more parameters have better fit than those with fewer parameters (Section 6.3.1). The individual models perform better than the joint models and the joint models that incorporate more individual factors perform better than those with fewer individual factors. These findings apply to most of the 13 countries.

However, using ME and MAE, joint models produce better forecasting accuracy than individual models (Sections 6.3.2, 6.3.3 and 6.3.4). The JWT model – with few parameters – yields the best forecast accuracy across the eight models. The JWT model performs the best because it allows for a flexible rate of mortality decline (ROMD) over time.

In the analysis of convergence properties of mortality across females and males, the models perform differently (Section 6.2.4). The SJLC, PLC and JWT models yield constant differences of mortality between females and males over time (non-divergence). The JLC model produces various directions of difference in the short term, but yields constant difference in the long term, whereas the individual models and the TLC model yield divergence.

The relationships between females and males are discussed in Section 2.3. Due to sharing the same culture and lifestyle, females and males in the same area generally expe-
rience similar mortality trends. Joint forecasting helps reduce the difference in mortality between females and males in the long-term projections. The finding from this analysis are very similar to those for the "across country" analysis in Section 6.2.

6.3.1 Goodness of fit according to BIC

Table 6.4 illustrates the overall performance of the models in the fitting period. The eight selected models are grouped as individual (WT, LC and HU) and joint (SJLC, JLC, TLC, PLC and JWT) models. The \mathbb{L} and BIC are the average across 13 countries, with a small \mathbb{L} or BIC indicating good performance and a high ranking. For example, the LC model has the smallest BIC (0.135) and is ranked first, indicating the best performance among the eight models.

Table 6.4: Average BIC and log-likelihood across the 13 countries in the fitting period

Models \mathbb{K} \mathbb{N} $\mathbb{L}(10^6)$ $BIC(10^6)$		Individua	1	Joint					
	WT	LC	HU	SJLC	JLC	TLC	PLC	JWT	
K	2	454	*	317	591	319	229	136	
\mathbb{N}		$90 \times 47 \times 2$							
$L(10^{6})$	0.285	0.131	0.142	0.195	0.138	0.200	0.280	0.279	
$BIC(10^{6})$	0.285	0.135	0.147	0.200	0.143	0.203	0.282	0.280	
Ranking	8	1	3	4	2	5	7	6	

Note:

- Both \mathbb{K} and \mathbb{N} are for females and males within a country.
- The HU model has K from 454 to 1,002 for each country as can be seen in Section 5.6.
- The number of parameters, K, is calculated via formulae displayed in Table 5.4.
- The number of observed data, points \mathbb{N} , is described in Section 5.3: ages 0-89 (90) and years 1948-1994 (47) and females and males within a country (2).
- The log-likelihood, \mathbb{L} , is calculated via (5.3): $\mathbb{L} = -2\sum_{xti} \left\{ d_{xti} \log \left(e_{xti} \hat{m}_{xti} \right) - e_{xti} \hat{m}_{xti} - \log(d_{xti}!) \right\}.$
- BIC is calculated via (5.2): $BIC = \mathbb{L} + \mathbb{K} \log(\mathbb{N})$.
- The full information of L, BIC and the BIC ranking are displayed in Tables C.1, C.2 and C.3 in Appendix C.

Note the following highlights in Table 6.4 and the explanations given below the highlights.

- Models have similar rankings in Table 6.4 to those in Table 6.1, suggesting that models perform consistently in different populations. In general, there is a pattern that models with more parameters have a lower BIC.
- 2) The PLC and JWT models perform better than the WT model, whereas this was not the case for the "across country" analysis in Section 6.2. The WT model is an individual model and its performance is, therefore, the same in the two analyses. The better performance of PLC and JWT models indicates that the assumptions of these two models are more appropriate for "across sex" analysis than for "across country" analysis.

Further evidence – models perform consistently in different populations

The models perform consistently in this analysis as in the "across country" analysis (Section 6.2.1); the rankings in Tables 6.1 and 6.4 are similar. Furthermore, the eight models perform similarly across countries (Table C.3, Appendix C).

Models with more parameters generally perform better than those with fewer parameters. In both analyses, the LC, HU and JLC models use the most parameters and perform the best according to BIC, while the WT, PLC and JWT models use the least parameters and perform the worst. Assumptions of the PLC and JWT models are more appropriate for "across sex" than for "across country"

In this analysis, the JWT and PLC models perform better than the WT model, whereas the JWT and PLC models perform worse than the WT model in the "across country" analysis. The WT model is an individual model, and thus is applied in the same way for the "across country" and "across sex" analyses. The similarities of mortality thus appear to be more easily captured by the JWT and PLC models in the case of sex rather than country.

6.3.2 Overall forecast accuracy – ME

Table 6.5 displays the overall ME of the forecast log death rate for the 13 countries. The ME indicates whether forecast mortality is higher or lower than the observed mortality.

Note the following highlights in Table 6.5 and the explanations given below the highlights:

- Six of the eight models yield negative overall ME for the countries. From the fitting period (1948-1994) to the forecasting period (1995-2009), males experience an increasing ROMD, while females experience a slightly decreasing ROME (Table 6.3). The combining of females and males gives a forecast log death rate generally above the observed log death rate in 1995-2009.
- The JWT model produces a small ME (0.019) average across countries, as well as a small standard deviation of ME (0.066) across countries. This indicates values
 ME – close to zero in most countries, rather than from a combination of large negative and large positive MEs, as can be observed in Table 6.5. The JWT model performs well in most countries.

Models		Individual	1	Joint					
Models	WT	LC	HU	SJLC	JLC	TLC	PLC	JWT	
AUS	-0.025	-0.062	-0.058	-0.068	-0.070	-0.057	-0.067	-0.028	
CAN	0.022	-0.042	-0.036	-0.047	-0.065	-0.033	-0.051	0.031	
ENW	0.034	-0.008	0.002	-0.022	0.026	0.006	-0.020	0.018	
FIN	0.114	0.059	0.050	0.036	0.106	0.052	0.038	0.114	
FRA	0.009	-0.069	-0.067	-0.087	-0.071	-0.062	-0.084	0.007	
ITA	0.007	-0.066	-0.085	-0.084	0.026	-0.084	-0.084	0.012	
ESP	0.076	-0.001	-0.111	-0.013	0.094	-0.010	-0.015	0.077	
SWE	0.043	0.007	-0.069	-0.005	0.013	0.016	-0.009	0.042	
CHE	-0.086	-0.146	-0.134	-0.152	-0.115	-0.140	-0.153	-0.096	
USA	-0.012	-0.049	-0.044	-0.052	-0.045	-0.041	-0.052	0.000	
JPN	0.271	0.196	0.012	0.177	0.279	0.166	0.187	0.139	
DNK	-0.084	-0.164	-0.137	-0.158	-0.214	-0.127	-0.170	-0.074	
NLD	-0.002	-0.042	-0.050	-0.052	-0.056	-0.039	-0.060	0.005	
Mean	0.028	-0.030	-0.056	-0.040	-0.007	-0.027	-0.042	0.019	
SD	0.092	0.090	0.055	0.085	0.122	0.079	0.089	0.066	

Table 6.5: ME of the forecast log death rate of each of the 13 countries

Table 6.6: Ranking of ME of the 13 countries

Models		Individua	1			Joint		
	WT	LC	HU	SJLC	JLC	TLC	PLC	JWT
AUS	1	5	4	7	8	3	6	2
CAN	1	5	4	6	8	3	7	2
ENW	8	3	1	6	7	2	5	4
FIN	7	5	3	1	6	4	2	8
FRA	2	5	4	8	6	3	7	1
ITA	1	4	8	5	3	6	7	2
ESP	5	1	8	3	7	2	4	6
SWE	7	2	8	1	4	5	3	6
CHE	1	6	4	7	3	5	8	2
USA	2	6	4	8	5	3	7	1
JPN	7	6	1	4	8	3	5	2
DNK	2	6	4	5	8	3	7	1
NLD	1	4	5	6	7	3	8	2
Mean	4	5	8	6	1	3	7	2
SD	7	6	1	4	8	3	5	2

3) The eight models all produce positive ME for Japan. Japanese females and males experienced the largest difference of ROMD between the fitting and forecasting periods as displayed in Table 6.3. The forecast log death rate is below the observed log death rate in 1995-2009.

General overestimation of mortality for the countries

The eight models yield negative overall ME averaged across countries. From the fitting period (1948-1994) to the forecasting period (1995-2009), the observed ROMD of females decreases from 0.025 to 0.024, while that of males increases from 0.017 to 0.030 (Table 6.3). Comparing the 0.001 reduction with the 0.013 increase, males dominate the trends of mortality, and thus the eight models overestimate mortality in 1995-2009.

However, the WT and JWT models are different from the other six models. The WT model always produces increasing ROMD as described in Section 6.2.2 and thus results in a larger underestimation for females but a smaller overestimation for males than the other six models when forecasting mortality. The JWT model uses a time trend to adjust the drift of z–scores. However, the time trend converges to zero in the long term, and the JWT model may be similar to the WT model, yielding increasing ROMD. Therefore, the WT and JWT models yield positive ME.

Small projection bias of the JWT model

The JWT model produces a small overall ME (0.019) across the 13 countries, where "small" means close to zero. The small standard deviation (0.066) indicates that the overall ME does not result from a combination of large positive and large negative MEs of individual populations. The JWT model performs well across most countries.

The JLC model yields an ME closest to zero (-0.007), but with the largest standard deviation (0.122) of the eight models. The small overall ME of the JLC model is due to a combination of large positive and large negative MEs of individual countries. The JLC model is not appropriate for the "across sex" joint forecasting for general populations.

Underestimation of mortality for Japan

The eight models yield a positive overall ME for Japan. From the fitting period (1948-1994) to the forecasting period (1995-2009), the observed ROMD of Japanese females decreases from 0.044 to 0.023, while that of males increases from 0.034 to 0.024 (Table 6.3). Both females and males in Japan have the largest reduction of ROMD among the 13 countries. The large change of ROMD in Japan results in the underestimation of mortality for Japan.

6.3.3 Overall forecast accuracy – MAE

Tables 6.7, 6.8 and 6.9 display the overall MAE, the ranking of MAE and the CMAE for the 13 countries.

Models		Individual	l			Joint		
Models	WT	LC	HU	SJLC	JLC	TLC	PLC	JWT
AUS	0.148	0.147	0.141	0.143	0.147	0.143	0.143	0.145
CAN	0.116	0.115	0.118	0.105	0.109	0.106	0.106	0.106
ENW	0.122	0.110	0.115	0.103	0.148	0.110	0.102	0.113
FIN	0.245	0.218	0.217	0.206	0.240	0.213	0.207	0.244
FRA	0.111	0.136	0.132	0.132	0.121	0.130	0.133	0.098
ITA	0.126	0.145	0.140	0.137	0.138	0.138	0.140	0.117
ESP	0.161	0.146	0.156	0.133	0.160	0.134	0.135	0.156
SWE	0.196	0.209	0.215	0.203	0.203	0.204	0.203	0.188
CHE	0.202	0.233	0.231	0.229	0.203	0.228	0.229	0.195
USA	0.091	0.099	0.099	0.092	0.085	0.092	0.092	0.082
JPN	0.287	0.221	0.111	0.206	0.316	0.199	0.212	0.165
DNK	0.204	0.251	0.246	0.246	0.275	0.233	0.250	0.199
NLD	0.127	0.140	0.142	0.136	0.137	0.133	0.138	0.120
Mean	0.164	0.167	0.159	0.159	0.176	0.159	0.161	0.148
SD	0.058	0.052	0.050	0.052	0.068	0.049	0.053	0.048

Table 6.7: MAE of the forecast log death rate of the 13 countries

Models		Individua	ıl			Joint		
	WT	LC	HU	SJLC	JLC	TLC	PLC	JWT
AUS	8	6	1	4	7	2	3	5
CAN	7	6	8	1	5	3	4	2
ENW	7	3	6	2	8	4	1	5
FIN	8	5	4	1	6	3	2	7
FRA	2	8	5	6	3	4	7	1
ITA	2	8	6	3	5	4	7	1
ESP	8	4	5	1	7	2	3	6
SWE	2	7	8	5	4	6	3	1
CHE	2	8	7	5	3	4	6	1
USA	3	7	8	5	2	6	4	1
JPN	7	6	1	4	8	3	5	2
DNK	2	7	4	5	8	3	6	1
NLD	2	7	8	4	5	3	6	1
Mean	6	7	3	4	8	2	5	1
SD	7	5	3	4	8	2	6	1

Table 6.8: Ranking of MAE of the 13 countries

Table 6.9: CMAE of the forecast log death rate of each of the 13 countries

Models		Individua	ıl			Joint		
	WT	LC	HU	SJLC	JLC	TLC	PLC	JWT
AUS	1.19	0.00	-3.73	-2.46	0.15	-2.94	-2.75	-0.25
CAN	-1.92	0.00	2.77	-9.09	-5.35	-7.95	-7.58	-10.12
ENW	4.37	0.00	4.71	-6.25	34.42	0.47	-7.55	2.95
FIN	6.85	0.00	-0.47	-5.61	10.41	-2.04	-5.15	6.03
FRA	-19.49	0.00	-3.48	-3.45	-11.57	-4.79	-2.50	-26.95
ITA	-13.19	0.00	-3.18	-5.40	-4.42	-4.96	-3.01	-19.84
ESP	0.67	0.00	6.87	-8.29	10.19	-7.90	-7.36	-1.53
SWE	-7.10	0.00	2.98	-2.57	-2.62	-2.35	-2.80	-9.97
CHE	-10.72	0.00	-0.68	-1.69	-12.77	-2.04	-1.65	-12.80
USA	-6.46	0.00	0.27	-6.93	-13.88	-6.52	-7.11	-15.22
JPN	10.79	0.00	-49.75	-6.83	43.32	-10.07	-3.80	-33.41
DNK	-16.90	0.00	-1.95	-1.95	9.57	-7.00	-0.29	4.91
NLD	-9.67	0.00	1.10	-3.41	-2.72	-5.56	-1.98	-14.70
Mean	-4.78	0.00	-4.88	-4.55	5.26	-4.90	-3.66	-9.73
SD	-0.00	0.00	-2.74	-0.25	31.65	-4.53	1.42	5.98

Note the following highlights in Tables 6.7, 6.8 and 6.9, and the explanations given below the highlights.

- The JWT model has the lowest overall MAE (0.148) and is ranked first in seven of the 13 countries. This provides further evidence that the JWT model performs the best among the eight models, by allowing a flexible ROMD.
- 2) The HU model performs better than the other seven models for Japan. However, the good performance can not be generalised to other populations. The HU model is appropriate for specific populations and may not be suitable for general application.
- 3) The JLC model is ranked last according to the overall MAE (0.176). The SJLC model excludes individual time trends and performs better than the JLC model. The JLC model may not be appropriate for short-term projection "across sex".
- 4) The average of MAE across joint models and across the 13 countries is slightly smaller than that of individual models (0.161 vs. 0.163). Using the *t*-test to evaluate the significance of the difference between individual and joint models (see Section 5.9), the joint models are not significantly better than the individual models.

Further evidence – reliable projection of the JWT model

The JWT model has the lowest overall MAE (0.148) and is ranked first in seven of the 13 countries. The advantages of the JWT model as described in Section 6.2.2 are also appropriate for this application.

Inappropriate general projection of the HU model

The HU model performs much better than the other seven models for Japan (Table 6.9). Japanese females and males experience a decrease in ROMD over time as described in Section 6.3.2. The HU model uses the ARIMA process for most of the individual populations, which results in a smaller decline of time trends in the forecasting period than in the fitting period. As a result, the HU model produces a smaller ROMD in the forecasting period than in the fitting period. This forecast happens to be appropriate for the variation in mortality of Japan. This also accounts for the relatively good performance of the HU model for females in Table 6.2, and the poor performance for males. The HU model does not perform consistently well in different scenarios.

Unreliable short-term projection of the JLC model

The JLC model yields the largest overall MAE (0.176) among the eight models and may be inappropriate for short-term projection. The MAE of 0.176 is produced using the AR(1) process to forecast k_{ti} . Using the ARIMA(1,1,0) process to forecast k_{ti} , the JLC model has an overall MAE of 0.160, which is much smaller than overall MAE using the AR(1) process. The forecast methods of k_{ti} significantly influence the forecast accuracy. Excluding k_{ti} , the JLC model becomes the SJLC model which has an overall MAE of 0.159, which is smaller than that of the JLC model using either AR(1) or ARIMA(1,1,0) processes. Therefore, the k_{ti} sequences reduce the short-term forecast accuracy of the JLC model, regardless of the forecasting time series used.

Insignificant differences between joint and individual models

From Table 6.7, the average MAE across the 13 countries is similar between the individual models (WT, LC and HU) and the joint models (SJLC, JLC, TLC, PLC and JWT).² Using a *t*-test, the p-value (see Section 5.9) is 0.478. Therefore, the difference between individual and joint models is not significant in this analysis.

²The average of MAE is 0.1633 for individual models and 0.1605 for joint models.

Table 6.10 combines Tables 6.2 and 6.7 and illustrates that the joint models perform relatively better in the evaluations of "across country" than in those "across sex". This difference in performance of joint models is due to the differences in mortality across populations. The standard deviations of mortality across females and males (discussed in Section 6.3.5) are generally larger than those across the 13 countries (Section 6.2.4). The larger standard deviations indicate less similarities of mortalities across populations, and thus reduce the forecast accuracy of the joint models.

Table 6.10: Overall MAE of the forecast log death rate for evaluations of "across country" (Section 6.2.2) and "across sex" (Section 6.3.3)

Models		Individua	1	Joint				
WIGUEIS	WT	LC	HU	SJLC	JLC	TLC	PLC	JWT
By country	0.164	0.167	0.159	0.154	0.169	0.155	0.151	0.145
By sex	0.164	0.167	0.159	0.159	0.176	0.159	0.161	0.148

6.3.4 Forecast accuracy over time

This section shows the trends in ME and MAE over the forecasting period and provides more information to evaluate the models. Sections 6.3.2 and 6.3.3 described the overall performance in the 15-year projection, which does not reflect the performance in the longer term. An analysis of forecast accuracy over time complements this shortcoming by describing the trends of ME and MAE over time, allowing us to anticipate performance over a longer forecasting period.

The ME of the forecast log death rate over time is introduced in Section 5.8.1 and denoted by $\varepsilon_{\cdot ti}$, reflecting the difference between observed and forecast log death rate of population *i* over *t*. Positive ME indicates that the observed log death rate is higher than the forecast log death rate. Negative ME indicates that the observed log death rate is lower than the forecast log death rate. An $\varepsilon_{\cdot ti}$ close to zero means the difference between observed and forecast mortality is small.

The MAE of the forecast log death rate over time is introduced in Section 5.8.2 and denoted by $\tilde{\varepsilon}_{\cdot ti}$, reflecting the absolute difference between the observed and forecast log death rate of population *i* over *t*. Slopes of the fitted straight lines of the curves are defined as the average difference between $\tilde{\varepsilon}_{\cdot ti}$ and $\tilde{\varepsilon}_{\cdot,t-1,i}$, reflecting the trend of MAE in the forecasting period. The trend of MAE is helpful to assess a model's performance in the long-term projection. For example, the forecasting period is from 1995 to 2009 in this research and the trend of MAE, or the slope of sequence $\tilde{\varepsilon}_{\cdot ti}$ over *t*, can be used to give some indication of the patterns of MAE after 2009.

Let $\varepsilon_{\cdot t}$ ($\tilde{\varepsilon}_{\cdot t}$) denote the average $\varepsilon_{\cdot ti}$ ($\tilde{\varepsilon}_{\cdot ti}$) across i – in this case, females and males. Figure 6.10 displays the average $\varepsilon_{\cdot t}$ (ME, left panel) and the average $\tilde{\varepsilon}_{\cdot t}$ (MAE, right panel) across the 13 countries. For example, the WT model yields an increasing $\tilde{\varepsilon}_{\cdot t}$ of the 13 countries over time with a slope 1.08 ($\times 10^{-2}$).



Note:

- The ME and MAE over time here are average $\varepsilon_{\cdot t}$ and $\tilde{\varepsilon}_{\cdot t}$ across the 13 countries.
- The numerics in the left panel denote the slopes of the fitted straight lines of the curves. The unit is 10⁻².

Figure 6.10: ME and MAE over time of the forecast log death rate

Note the following highlights in Figure 6.10 and the explanations given below the highlights:

- The WT and JWT models produce increasingly ME over time while the other six models yield increasingly negative ME over time. The WT and JWT models use the z-score to forecast mortality. When transforming the z-score to the log death rate, the constant increase in z-scores results in an accelerating decline in the log death rate as described in Section 3.2.1.
- 2) The JWT model has the smallest slope (0.81) in MAE. Since the JWT model permits a flexible ROMD over the forecasting period, the JWT model is better able to capture the path of mortality decline than the other seven models. As a result, the JWT model keeps the advantage of projection during the forecasting horizon and may maintain this advantage in the longer term.
- 3) The performance of forecast accuracy across the eight models is more similar in the "across sex" analysis than in the "across country" analysis. The joint models perform better in forecast accuracy in the "across country" analysis than in the "across sex" analysis, and thus the difference between individual and joint models is smaller in the "across sex" analysis.

Differences between the LC family and WT family

The LC family (LC, HU, SJLC, JLC, TLC and PLC) and WT family (WT and JWT) produce different directions of ME over time as can be seen in the left panel of Figure 6.10, due to these two families using different mortality scales. The models of the LC family use $\log(m)$ or $\log_i(q)$ as mortality scales and apply a constant ROMD to forecast mortality.

Table 6.3 illustrates that females experience a slight decrease in ROMD but males

experience a large increase in ROMD. As a result, the models of the LC family almost always overestimate mortality in the forecasting period. However, the models of the WT family use z–scores which may result in an increasing ROMD as described in Section 6.3.2. Therefore, the models of the WT family underestimate mortality.

Further evidence – reliable projection of the JWT model

The JWT model has the smallest slope (0.81) in the MAE over time. The flexible ROMD over time enables the JWT model to perform the best in the joint forecast of "across country" as discussed in Section 6.2.2. The best performance in the joint forecast of "across sex" further supports the advantages of the JWT model. During the forecasting horizon, the JWT model holds the lowest MAE which can be maintained in the longer-term projection.

Smaller difference in performance across the models in the "across sex" analysis than in the "across country" analysis

As described in Section 6.3.3, the difference in forecast accuracy between individual and joint models is smaller in the "across sex" analysis than in the "across country" analysis. This finding reflects the overall performance in the limited forecasting horizon.

The standard deviation of the slopes is $0.21 (\times 10^{-2})$ for females and $0.18 (\times 10^{-2})$ for males in the "across country" analysis and $0.10 (\times 10^{-2})$ in the "across sex" analysis. The slopes are displayed in Figures 6.6 and 6.10. A small difference of the slopes indicates that the models may experience little difference in the long-term projection. Therefore, in the long-term projection, the difference across the eight models is smaller for the "across sex" analysis than for the "across country" analysis.

6.3.5 Convergence properties of mortality between females and males within a country

This section illustrates the forecast differences in mortality across females and males over the forecasting period. Throughout the world, mortalities of females and males in a country are generally converging (Mathers et al., 2001; Moser et al., 2005). The difference in forecast mortality between females and males is therefore expected to be constant or reducing over time. In this section, the differences in mortality between females and males are measured by the standard deviation of the forecast log death rate across populations – in this case, females and males, ϑ_{xt} , as introduced in Section 5.10.

As described in Section 3.3, individual models permit populations to have individual rates of mortality decline, hence allowing for divergence. Joint models, in contrast, constrain populations to share common mortality trends and do not permit divergence. If a model yields horizontal or decreasing ϑ_{xt} over t, the model is verified to control the divergence in forecast mortality.

Figure 6.11 displays the standard deviation of the log death rate of females and males in Italy as an example. Other countries display similar patterns. The randomly selected ages are 0, 10, 20, 30, 49, 59, 69, 79 and 89. Every panel displays the ϑ_{xt} of a model. In the fitting period (1948-1994), the ϑ_{xt} is calculated using the observed log death rate. The eight models have the same patterns of ϑ_{xt} in this period. In the forecasting period (1995-2044), ϑ_{xt} is calculated using the forecast log death rate by a given model. Since convergence properties of mortality of females and males require a long-term assessment, the forecasting period is 1995-2044 in this evaluation.



Note:

- From 1948 to 1994, standard deviation is calculated using observed log death rate.
- From 1995 to 2044, standard deviation is calculated using forecast log death rate.
- Plotted curves are standard deviation at ages 0, 10, 20, 30, 49, 59, 69, 79 and 89.

Figure 6.11: Standard deviation of log death rate across females and males in Italy

Note the following highlights in Figure 6.11 and the explanations given below the highlights.

- In the fitting period, standard deviations increase before the 1970s-1980s and decrease after the 1980s. The patterns of standard deviation are consistent with the conclusions of Liu et al. (2012). Differences in mortality between females and males have generally been converging since the 1980s in developed countries.
- 2) In the forecasting period, the WT, LC, HU and TLC models yield increasing standard deviations in the forecast rates of mortality over time. For some ages, the LC and HU models yield decreasing standard deviations. The decrease in standard deviation is due to the larger ROMD of males and the smaller ROMD of females. Individual models (WT, LC and HU) and the TLC model produce apparently unreliable results as their forecasts are likely to diverge, which is inconsistent with preconceptions about the future trajectory of mortality by sex.
- 3) In the forecasting period, the SJLC, PLC and JWT models produce constant standard deviation over time. The SJLC, PLC and JWT models appear more reliable for mortality forecasting since they inhibit divergence in forecast mortality across populations. The JLC model produces, in the short term, and for differing age groups, differing directions in the standard deviation over time. The JLC model is reasonable in long-term projection, since this model permits divergence in mortality across populations in short-term projection and keeps a constant difference across populations in long-term projection.

Convergence of mortality between females and males

The differences in mortality between females and males can be assessed by using different mortality scales, for example, male-to-female ratio of death rate (Li, 2013), standard deviation in life expectancy (Li and Lee, 2005) or difference in life expectancy (Liu et al., 2012). This research applies the standard deviation of log death rate across populations.

In the fitting period, standard deviations increase before the 1970s-1980s and decrease after the 1980s (Figure 6.11). The patterns of standard deviation are consistent with the conclusions of Liu et al. (2012), indicating that the measure of standard deviation is suitable to analyse the relationships of mortality between females and males.

Since the 1980s, females and males have convergent mortality in many developed countries. Table 6.3 illustrates that the ROMD is generally getting smaller for females and larger for males. However, the change of the ROMD may trend to zero and the differences of mortality between females may be constant in the long term as described in Section 2.3.2. As a result, mortality (the log death rate) of females may remain lower than that of males, in spite of the ROMD being currently larger for males than for females.

Unreliable projection of the TLC model and the individual models

The TLC model and the individual models (WT, LC and HU) have similar patterns of standard deviation to those displayed in Figures 6.7 and 6.8. Increasing standard deviations indicate divergence of mortality between females and males. The WT and TLC models yield increasing standard deviation for all ages and therefore the WT and TLC models are apparently unreliable for long-term projection.

The LC and HU models yield different directions of standard deviations for some ages. Figure 6.12 displays the observed and forecast log death rate using the HU model for ages 0 and 49 in Italy. Since males and females have different ROMD, the two curves cross over in the two panels. The mortalities of age 0 converge between females and males and then diverge. This type of cross over is unrealistic and can be avoided by the use of joint models.



Figure 6.12: Observed and forecast log death rate for ages 0 and 49 of Italian females and males using the HU model

Reliable projection of the SJLC, JLC, PLC and JWT models

As in Section 6.2.4, the SJLC, PLC and JWT models yield constant standard deviations over time. As k_{ti} sequences converge at zero in 2009 as displayed in Figure 6.13, the JLC model yields constant standard deviations over time. These four models produce non-divergent projections between females and males.



Figure 6.13: Estimates and forecasts of k_{ti} with the JLC model for Italy

6.4 Conclusion of the fits and forecasts

This chapter displays the finds of the WT, LC, HU, SJLC, JLC, TLC, PLC and JWT models. These eight models are applied to fit and forecast the mortality of females and males of 13 countries. Since the 13 countries have not been specifically selected, the findings related to the eight models should be able to be generalised to other mortality data sets. Additionally, the eight models have not been specially selected, and thus the performance of these eight models should also be broadly representative of the performance of general individual and joint models.

Joint models (SJLC, JLC, TLC, PLC and JWT) are applied to jointly fit and forecast mortality combining the 13 countries by a single sex – "across country" (Section 6.2), and to jointly fit and forecast mortality combining females and males within each of the 13 countries – "across sex" (Section 6.3). Both the "across country" and "across sex" analyses illustrate the eight models' goodness of fit and forecasting performance.

Conclusion one: Models with more parameters generally provide a better fit to mortality data (Sections 6.2.1 and 6.3.1). In both evaluations, the LC, HU and JLC models, which have the most parameters, gave better goodness of fit than the other five models. A large number of parameters generally enables a model to quantitatively describe the properties of mortality data well. However, a model with few parameters may miss some important features of mortality data. Therefore, models with more parameters generally have better fit than those with fewer parameters.

Although this research does not apply the models to analyse mortality, a model that fits better will provide a more appropriate basis for analysis. The analysis of mortality is influenced by the factors involved in a model. In this research, the HU and LC models use only individual factors, while the JLC and TLC models involve both individual and common factors. As a result, the LC model has the best goodness of fit, but can only be applied to individual populations. The JLC and TLC models – ranked second and fourth in terms of fit – are able to describe both individual populations and common features across the member populations. Therefore, if a model involves individual factors, it is appropriate to analyse the mortality of individual populations. If a model involves both individual and common factors, it is appropriate to analyse relationships of mortality across multiple populations.

Conclusion two: Joint models generally have better forecasting performance than individual models. Specifically, in this study, the joint models (SJLC, JLC, TLC, PLC and JWT) have better forecast accuracy than the individual models (WT, LC and HU). Furthermore, the individual models (WT, LC and HU) produce increasing standard deviations of forecast mortality across populations over time, but the joint models (SJLC, JLC, PLC and JWT) yield constant standard deviations (Sections 6.2.4 and 6.3.5). The different performance of projection between individual and joint models is due to the properties of the models. Firstly, joint models aggregate similar populations to jointly fit and forecast mortality. Aggregating data enables joint models to grasp more information than individual models, resulting in better forecast accuracy. Secondly, joint models are more realistic than individual models. Mortalities of different populations are increasingly similar, since the 1950s as described in Section 2.3. However, individual models allow populations to hold individual mortality trends over time, and thus generally yield increasing differences in mortality across populations in the long-term projection.

Conclusion three: Simple models have better forecasting performance than complicated models. In this research, the LC, HU and JLC models have the most parameters, but their forecasting performance is unexpectedly worse than the other five models. The JWT model has much fewer parameters, but better forecasting performance than the LC,

HU and JLC models. The better performance of simple models may be due to two factors. Firstly, if a model uses a large number of parameters, the errors may be cumulatively large when forecasting these parameters. Secondly, forecasts of time trends significantly influence the forecast of mortality (Lee and Miller, 2001). If a model involves more than two time trends, these time trends may be in conflict in the long-term projection and the conflict may result in poor forecast accuracy.

Conclusion four: The JWT model is generally reliable for both short-term and long-term projections. This research proposes the JWT model and illustrates its advantages in forecasting. In the short-term projection (1995-2009), the JWT model yields the best forecast accuracy in both the "across country" and "across sex" analyses. This advantage may be extended to a longer-run forecast, since the JWT model has the smallest slope of MAE over time of the eight models (Sections 6.2.3 and 6.3.4). In the long-term projection (1995-2044), the JWT model yields constant differences of mortality across populations. The JWT model yields the highest forecast accuracy and non-divergence over time, and thus is the most reliable in both short-term and long-term forecasts.

The JWT model is more practical than existing mortality models. Mortality trends are uncertain over time and the rate of mortality decline (ROMD) varies over time as described in Section 2.4. Existing mortality models assume that the ROMD is constant over time, but the JWT model permits the ROMD to vary over time. As a result, the JWT model is consistent with the variation of mortality over time, and thus is more practical than existing individual and joint mortality models. In addition, the JWT model possesses the advantages of joint models and simple models. Therefore, in projection, the JWT model performs the best of the eight models. In this research, the other seven models (WT, LC, HU, SJLC, JLC, TLC and PLC) and the mortality data sets (females and males in the 13 countries) are randomly selected. Therefore, the advantages of the JWT model can be generalised to other mortality data sets and compared.

However, the JWT model has its limitations. The "across country" and "across sex" analyses demonstrate that these limitations do not influence the JWT model's performance, at least in the selected data sets. However, these limitations are potential risks in application to other data sets.

Firstly, transformed from the survival probability, the z–scores do not directly reflect mortality data. Hence, a model using z–scores may not be easily understood. However, forecast z–scores can be transformed into commonly used and understood mortality measures, for example, the log death rates.

Secondly, the survival probabilities are the product of the one year death rates across ages, and thus correlated across ages. Modelling highly correlated variables is not usually satisfactory. However, in this case, the dependent survival probability can be converted back to the independent one year death rate. Therefore, modelling the correlated z–scores is acceptable in this scenario.

Thirdly, z-scores varying with age-specific factors (a_x) may theoretically result in crossover of z-scores. This risk does not significantly influence the application of the JWT model. The a_x is estimated using non-crossing z-scores, and the estimates of z-scores are expected to be non-crossing. Considering the initial differences of z-scores (z_{xni}) , the forecast z-scores do not appear to cross over by age in practice.

Conclusion five: The PLC and SJLC models are also reliable for projection, but do not perform as well as the JWT model. On one hand, for forecast accuracy, the SJLC and PLC models are ranked second and third in the evaluation of "across country" (Section 6.2.2) and ranked fifth and sixth in the evaluation of "across sex" (Section 6.3.3). Since the SJLC and PLC models assume that the ROMD varies constantly over time, these

two models do not have the same flexibility as the JWT model. Therefore, the PLC and SJLC models have worse forecast accuracy than the JWT model. On the other hand, for divergence control, the SJLC and PLC models yield constant differences in mortality across populations over time (Sections 6.2.4 and 6.3.5). Since the SJLC and PLC models use one time trend to forecast mortality, member populations share the same time trend and have the same patterns of mortality decline in the forecasting horizon. As a result, the differences of mortality across populations are constant over time. Consequently, the SJLC and PLC models possess the advantages of both joint models and simple models, and are thus reliable for projection but are less flexible than the JWT model.

Conclusion six: The JLC model is not reliable for short-term projection, but may be appropriate for long-term projection. Firstly, in the short-term projection (1995-2009) for forecast accuracy, the JLC model performs the worst in both "across country" and "across sex" analyses (Sections 6.2.2, 6.3.2 and 6.3.3). The poor performance may be due to the inappropriate forecast methods of the individual time trends (k_{ti}), but not the JLC model itself. Secondly, in the long-term projection (1995-2044), the JLC model is able to constrain the divergence when forecasting mortality. The JLC model yields constant standard deviations of the forecast log death rate across populations over the forecasting horizon (Sections 6.2.4 and 6.3.5). Therefore, in spite of the reliable long-term projection, the JLC model may produce poor forecast accuracy in the short term.

Conclusion seven: The TLC model is appropriate for the short-term projection, but may be unreliable for the long-term projection. In the projection of 1995-2009, for forecast accuracy, the TLC model is ranked fourth in the evaluation of "across country" (Section 6.2.2) and ranked second and fourth in the evaluation of "across sex" (Section 6.3.3). According to the rankings in this research, the TLC model has high accuracy in short-term projection. However, in the projection of 1995-2044, the TLC model produces diver-

gence. The TLC model yields increasing standard deviations of the forecast log death rate across populations over the forecasting horizon. Therefore, the TLC model is unreliable when forecasting mortality in the long run.

Conclusion eight: The WT, LC and HU models may be unreliable to forecast mortality of a group of similar populations. First of all, these three models cause divergence of mortality across populations (Sections 6.2.4 and 6.3.5) because, as individual models, they permit different populations to hold individual mortality trends over time. These individual mortality trends result in increasing differences in mortality across populations in the long term, referred to as divergence. Furthermore, the LC model yields poor forecast accuracy, being ranked seventh in both analyses. The WT and HU models produce good forecast accuracy for either females or males in the "across country" analysis (Sections 6.2.2) but are inappropriate for general projection as described in Section 6.3.3. Consequently, the WT, LC and HU models are unreliable for projection.

In conclusion, the JWT model has a simple form, incorporates similar populations to jointly forecast mortality and allows rate of mortality decline (ROMD) to vary over time. These three features enable the JWT model to perform the best in both short-term and long-term forecasting. The other models may perform well according to fit or forecast, but offer fewer advantages compared with the JWT model.

Chapter 7

Discussion and Further Research Directions

Mortality modelling – using mathematical and statistical approaches – increases demographers' understanding of mortality variation and patterns. Raw mortality data comprise a mass of numbers and these numbers are difficult to comprehend and understand without the aid of a model. Thus, mortality models aim to describe the levels and variation in mortality across ages, years and populations. For example, the LC model uses age and time factors to describe the variation of mortality across ages and over time. These parameters are estimated using mathematical and statistical techniques and constraints. By using a limited number of parameters, mortality models simplify the expression of mortality variation. Demographers can more easily understand the patterns of mortality via a mortality model, which is generally difficult from a direct inspection of raw mortality.

In addition, mortality models provide an intuitive and consistent method to forecast mortality. For example, the LC model forecasts mortality by forecasting the time factors, where the time factors can be viewed as a time series process of random walk with drift. By using statistical methods to forecast, uncertainty of the projection is incorporated and the risk of projection can be analysed. Hence, mortality models provide convenient and reliable forecasts. In addition, mortality models can be applied to foresee the risks of industries that are operating products related to mortality.

This research firstly provides an overview of variations in mortality, illustrating how mortality rates vary over time (Section 2.2 and Section 2.4) and how the variations become increasingly similar (Section 2.3). These two features of the variation of mortality demonstrate that mortality models that use a constant rate of mortality decline to forecast and that use separate populations are generally inappropriate and/or fail to exploit common features of mortality across populations. However, there are no existing mortality models which satisfy both these conditions.

For this purpose, this research develops a new joint model – the JWT model (Chapter 4). The JWT model incorporates three features: jointly fitting and forecasting mortality, allowing a flexible rate of mortality decline and simplicity. These three features are expected to enable the JWT model to perform well when forecasting mortality.

The JWT model assumes that populations share the same increase of z-scores over time, subject to common age and time variations, but permit populations to possess individual initial rates – the jump-off z-scores. The differences of z-scores across populations are constant and are equal to the differences of the jump-off z-scores. The constant differences of z-scores result in decreased differences of survival probability across populations as detailed in Section 4.4. Therefore, the JWT model captures the convergence properties of survival probabilities across populations. The JWT model constrains mortality to be non-divergent across populations in long-term projections (Sections 6.2.4 and 6.3.5).

By using a time trend to describe the changes of z–scores over time, the JWT model permits z–scores to vary flexibly over time in the short term. Since the time trend is

stationary, the z–scores vary constantly over time in the long term. The flexibility enables the JWT model to yield good forecast accuracy in both short-term projections (Sections 6.2.2, 6.3.2 and 6.3.3) and long-term projections (Sections 6.2.3, 6.2.4, 6.3.4 and 6.3.5).

The JWT model is simple in application. The JWT model uses z-scores and the variation of mortality is linear over time under the scales of z-scores. Therefore, the JWT model uses a linear form to model the linear variation of z-scores. The JWT model provides simple estimation and forecasting (Section 4.3).

To evaluate the performance of the JWT model, this research compares the JWT model with seven other existing commonly used individual models (WT, LC and HU) and joint models (SJLC, JLC, TLC and PLC). These eight models are applied to fit and forecast the mortality of 13 countries. The joint models are applied to jointly fit and forecast mortalities "across country" (Section 6.2) and "across sex" (Section 6.3) separately. Both analyses illustrate the goodness of fit and forecasting performance of the models.

Goodness of fit reflects the match between mortality data and the models in the fitting period. Using a large number of parameters generally enhances the describability of mortality data. Therefore, the models with more parameters normally perform better than those with fewer parameters. The LC, HU and JLC models use more parameters and thus produce better BIC than the five other models. However, using additional parameters may result in over-parameterizations and may produce poor projection.

Forecasting performance consists of forecast accuracy and convergence properties of mortality across populations. Forecast accuracy illustrates the differences between the observed and forecast mortality and small differences indicate good forecast accuracy. Convergence properties of mortality across populations illustrate the differences of forecast mortality across populations, using the standard deviation of the forecast log death rate across populations. Due to the increasing similarities in mortality across populations, the standard deviations are expected to decrease or to be constant over time.

In the evaluation of forecasting performance, the JWT model performs the best of the eight models. The JWT model possesses three main features which result in good performance in both short-term and long-term projections. Allowing a flexible rate of mortality decline over time enables the JWT model to grasp the trends of mortality decline. Aggregating similar populations enables the JWT model to incorporate the relationships of mortality across populations. Having a simple form helps the JWT model to reduce the risk of over-parameterization. Therefore, the JWT model yields the highest forecast accuracy and a constant difference of mortality during the forecasting horizon. The other joint models (SJLC, JLC, TLC and PLC) generally have better forecast accuracy and divergence control than the individual models (WT, LC and HU).

In the evaluation of convergence properties of mortality across populations, the joint models (JWT, SJLC, JLC and PLC) yield constant differences of forecast log death rates across populations. However, the individual models (WT, LC and HU) and the TLC model produce increasing differences of forecast log death rates across populations. Therefore, the joint models generally are able to control the divergence when forecasting mortality, but individual models produce divergence.

While these models all perform differently and may be appropriate for different applications, the JWT model is the most appropriate for short-term and long-term projections, followed by the PLC and SJLC models. The remaining models may only be appropriate for long-term (JLC) or short-term (TLC) projection or for analysing mortality of individual populations (LC and HU) or a group of populations (JLC and TLC).

The evaluations carried out in this research may not be comprehensive. The poor performance of a model may not be due to the model itself but due to an inappropriate application. The forecast methods of the time trends may not be appropriate for the JLC model, which may have contributed to the weak forecast accuracy. The analysis is concerned only with the fit and forecast accuracy of models, not with their ability to describe and assist the understanding of the properties of mortality. Furthermore, the analysis of relationships across populations using joint models is not considered. Finally, the mortality data sets of the 13 developed countries may not completely reflect the performance of the models for other data sets, specifically of developing countries.

The JWT model has the risk of permitting z–scores to cross over ages as described in Section 4.5. This risk does not significantly influence the performance of the JWT model in this research, but may be inappropriate in other applications. A further study can generalise the JWT model to avoid the risk of crossing z–scores over ages.

Acknowledgement of the above mentioned shortcomings points to the need for further research to address the following areas. Firstly, the existing models can be improved or extended to match the advantages of the JWT model. In particular, the methods used to forecast can be enhanced to better reflect actual mortality trends. Secondly, since females and males within a country are similar and countries are similar by given sex, a combination of females and males across the 13 countries may also be appropriate for joint fitting and forecasting. More mortality data sets can be applied to evaluate the models performance, for example, data sets of developing countries. Thirdly, the JWT model can be applied to analyse the properties of mortality of multiple populations. In this way, a comprehensive application of the JWT model would be demonstrated.

References

- Alho, J., Spencer, B., 1990. Error models for official mortality forecasts. Journal of the American Statistical Association, 609–616.
- Bell, W., 1992. Arima and principal component models in forecasting age-specific fertility.
- Bell, W., 1997. Comparing and assessing time series methods for forecasting age-specific fertility and mortality rates. Journal Of Official Statistics-Stockholm 13, 279–304.
- Bell, W., Monsell, B., 1991. Using principal components in time series modeling and forecasting of age-specific mortality rates.
- Biatat, V. D., Currie, I. D., 2010. Joint models for classification and comparison of mortality in different countries. In: Proceedings of 25th International Workshop on Statistical Modelling, Glasgow. pp. 89–94.
- Booth, H., 2006. Demographic forecasting: 1980 to 2005 in review. International Journal of Forecasting 22 (3), 547–581.
- Booth, H., Hyndman, R. J., Tickle, L., De Jong, P., 2006. Lee-Carter mortality forecasting: a multi-country comparison of variants and extensions. Working Paper-Monash University Department Of Econometrics and Business Statistics 13.

- Booth, H., Maindonald, J., Smith, L., 2002a. Age-time interactions in mortality projection: Applying Lee-Carter to Australia. Working Papers in Demography (85).
- Booth, H., Maindonald, J., Smith, L., 2002b. Applying Lee-Carter under conditions of variable mortality decline. Population Studies 56 (3), 325–336.
- Booth, H., Tickle, L., 2008. Mortality modelling and forecasting: A review of methods. Annals of Actuarial Science 3 (1-2), 3–43.
- Brass, W., et al., 1971. On the scale of mortality. Biological aspects of demography, 69– 110.
- Brillinger, D., 1986. A biometrics invited paper with discussion: the natural variability of vital rates and associated statistics. Biometrics, 693–734.
- Brouhns, N., Denuit, M., Vermunt, J. K., 2002. A Poisson log-bilinear regression approach to the construction of projected lifetables. Insurance: Mathematics and Economics 31 (3), 373–393.
- Cairns, A., Blake, D., Dowd, K., Coughlan, G., Khalaf-Allah, M., 2011a. Bayesian stochastic mortality modelling for two populations. ASTIN Bulletin-Actuarial Studies in non LifeInsurance 41 (1), 29.
- Cairns, A. J., Blake, D., Dowd, K., Coughlan, G. D., Epstein, D., Khalaf-Allah, M., 2011b. Mortality density forecasts: An analysis of six stochastic mortality models. Insurance: Mathematics and Economics 48 (3), 355–367.
- Cairns, A. J., Blake, D., Dowd, K., Coughlan, G. D., Epstein, D., Ong, A., Balevich, I., 2009. A quantitative comparison of stochastic mortality models using data from England and Wales and the United States. North American Actuarial Journal 13 (1), 1–35.

- Cairns, A. J., Dowd, K., Blake, D., Coughlan, G. D., 2014. Longevity hedge effectiveness: A decomposition. Quantitative Finance 14 (2), 217–235.
- Cairns, A. J. G., Blake, D., Dowd, K., 2006. A two-factor model for stochastic mortality with parameter uncertainty: Theory and calibration. The Journal of Risk and Insurance 73 (4), 687–718.
- Cairns, A. J. G., Blake, D. P., Dowd, K., 2008. Modelling and management of mortality risk: a review. Scandinavian Actuarial Journal 2, 79–113.
- Carriere, J. F., 1992. Parametric models for life tables. Transactions of the Society of Actuaries 44, 77–99.
- Carter, L., Lee, R., 1992. Modeling and forecasting us sex differentials in mortality. International Journal of Forecasting 8 (3), 393–411.
- Carter, L., Prskawetz, A., 2000. Examining structural shifts in mortality using the Lee-Carter method. Methoden und Ziele Septemper, 39–54.
- Chan, W., Li, S., Cheung, S., 2008. Testing deterministic versus stochastic trends in the Lee-Carter mortality indexes and its implications for projecting mortality improvements at advanced ages. Living to 100, 7–9.
- Chan, W.-S., Li, J. S.-H., Li, J., 2014. The cbd mortality indexes: modeling and applications. North American Actuarial Journal 18 (1), 38–58.
- Coale, A., 1996. Age patterns and time sequence of mortality in national populations with the highest expectation of life at birth. Population and Development Review, 127–135.
- Coale, A., Demeny, P., Vaughan, B., 1983. Regional model life tables and stable populations. Vol. 41. Academic Press New York.

- Cockerham, W., 1997. The social determinants of the decline of life expectancy in Russia and Eastern Europe: a lifestyle explanation. Journal of Health and Social Behavior, 117–130.
- Crimmins, E., Finch, C., 2006. Infection, inflammation, height, and longevity. Proceedings of the National Academy of Sciences of the United States of America 103 (2), 498–503.
- Crimmins, E., Preston, S., Cohen, B., 2011. Difference between life expectancy in the United States and other high-income countries.
- Currie, I., 2006. Smoothing and forecasting mortality rates with p-splines. DP Heriot Watt University.
- Currie, I. D., Durban, M., Eilers, P. H. C., 2004. Smoothing and forecasting mortality rates. Statistical Modelling 4 (4), 279.
- Cutler, D., Meara, E., 2001. Changes in the age distribution of mortality over the 20th century. Tech. rep., National Bureau of Economic Research.
- Cutler, D. M., Kadiyala, S., 1999. The economics of better health: the case of cardiovascular disease. New York (NY): The Lasker Foundation.
- Czado, C., Delwarde, A., Denuit, M., 2005. Bayesian poisson log-bilinear mortality projections. Insurance: Mathematics and Economics 36 (3), 260–284.
- D'Amato, V., Piscopo, G., Russolilo, M., 2011. The mortality of the Italian population: Smoothing techniques on the Lee–Carter model. The Annals of Applied Statistics 5(2A), 705–724.
- De Jong, P., Marshall, C., 2007. Mortality projection based on the Wang Transform. Astin Bulletin 37 (1), 149.

- De Jong, P., Tickle, L., 2006. Extending Lee–Carter mortality forecasting. Mathematical Population Studies 13 (1), 1–18.
- Deaton, A., Paxson, C., 2001. Mortality, education, income, and inequality among american cohorts.
- Debón, A., Martínez-Ruiz, F., Montes, F., 2012. Temporal evolution of mortality indicators: Application to spanish data. North American Actuarial Journal 16 (3), 364–377.
- Debón, A., Montes, F., Martínez-Ruiz, F., 2011. Statistical methods to compare mortality for a group with non-divergent populations: an application to Spanish regions. European Actuarial Journal 1 (2), 291–308.
- Debón, A., Montes, F., Puig, F., 2008. Modelling and forecasting mortality in spain. European Journal of Operational Research 189 (3), 624–637.
- Delwarde, A., Denuit, M., Eilers, P., 2007. Smoothing the Lee-Carter and Poisson logbilinear models for mortality forecasting: A penalized log-likelihood approach. Statistical Modelling 7 (1), 29.
- Denison, R., 2011. Past evolutionary tradeoffs represent opportunities for crop genetic improvement and increased human lifespan. Evolutionary Applications 4 (2), 216–224.
- Dorrington, R., 2006. Demographic Impact of HIV/AIDS in South Africa: National and Provinsial Indicators for 2006. Centre for Actuarial Research.
- Dowd, K., Cairns, A., Blake, D., Coughlan, G., Epstein, D., Khalaf-Allah, M., 2010. Evaluating the goodness of fit of stochastic mortality models. Insurance: Mathematics and Economics 47 (3), 255–265.
- Elo, I., Preston, S., 1992. Effects of early-life conditions on adult mortality: a review. Population Index, 186–212.

- Finch, C., 2010. Evolution of the human lifespan and diseases of aging: Roles of infection, inflammation, and nutrition. Proceedings of the National Academy of Sciences 107 (suppl 1), 1718–1724.
- Floyd, S., Marston, M., Baisley, K., Wringe, A., Herbst, K., Chihana, M., Kasamba, I., Bärnighausen, T., Urassa, M., French, N., et al., 2012. The effect of antiretroviral therapy provision on all-cause, AIDS and non-AIDS mortality at the population level–a comparative analysis of data from four settings in Southern and East Africa. Tropical Medicine & International Health 17 (8), e84–e93.
- Forfar, D., McCutcheon, J., Wilkie, A., 1988. On graduation by mathematical formula. Journal of the Institute of Actuaries 115 (01), 1–149.
- Forfar, D., Smith, D., 1985. The changing shape of english life tables. Transactions of the Faculty of Actuaries 40, 98–134.
- Gero, D., 2010. The East–West health divide in Europe: growing and shifting eastwards.
- Girosi, F., King, G., 2007. Understanding the lee-carter mortality forecasting method.
- Glei, D., Horiuchi, S., 2007. The narrowing sex differential in life expectancy in highincome populations: effects of differences in the age pattern of mortality. Population Studies 61 (2), 141–159.
- Gogola, J., 2014. Stochastic Mortality Models. Application to CR Mortality Data. In: Mathematical and Statistical Methods for Actuarial Sciences and Finance. Springer, pp. 113–116.
- Gompertz, B., 1825. On the nature of the function expressive of the law of human mortality, and on a new mode of determining the value of life contingencies. Philosophical Transactions of the Royal Society of London 115, 513–583.

- Hannerz, H., 1999. Methodology and applications of a new law of mortality. Department of Statistics, Lund University.
- Hannerz, H., 2001a. An extension of relational methods in mortality estimation. Demographic Research 4 (10), 337–368.
- Hannerz, H., 2001b. Presentation and derivation of a five-parameter survival function intended to model mortality in modern female populations. Scandinavian Actuarial Journal 2001 (2), 176–187.
- Hári, N., De Waegenaere, A., Melenberg, B., Nijman, T., 2008. Longevity risk in portfolios of pension annuities. Insurance: Mathematics and Economics 42 (2), 505–519.
- Hatzopoulos, P., Haberman, S., 2013. Common mortality modelling and coherent forecasts. an empirical analysis of worldwide mortality data. Insurance: Mathematics and Economics.
- Heligman, L., Pollard, J. H., 1980. The age pattern of mortality. Journal of the Institute of Actuaries 107 (1), 49–80.
- Hosegood, V., Vanneste, A., Timæus, I., 2004. Levels and causes of adult mortality in rural south africa: the impact of aids. Aids 18 (4), 663–671.
- Hosseinpoor, A., Harper, S., Lee, J., Lynch, J., Mathers, C., Abou-Zahr, C., 2012. International shortfall inequality in life expectancy in women and in men, 1950-2010. Bulletin of the World Health Organization 90 (8), 588–594.
- Hyndman, R., Ullah, S., et al., 2007. Robust forecasting of mortality and fertility rates:
 A functional data approach. Computational Statistics & Data Analysis 51 (10), 4942–4956.
- Hyndman, R. J., Booth, H., Yasmeen, F., 2013. Coherent mortality forecasting: the
product-ratio method with functional time series models. Demography 50 (1), 261–283.

- Jahn, A., Floyd, S., Crampin, A., Mwaungulu, F., Mvula, H., Munthali, F., McGrath, N., Mwafilaso, J., Mwinuka, V., Mangongo, B., et al., 2008. Population-level effect of hiv on adult mortality and early evidence of reversal after introduction of antiretroviral therapy in malawi. The Lancet 371 (9624), 1603–1611.
- Jarner, S., Kryger, E., 2009. Modelling adult mortality in small populations: The SAINT model. ASTIN Bulletin.
- Jarner, S., Kryger, E., Dengsøe, C., 2008. The evolution of death rates and life expectancy in Denmark. Scandinavian Actuarial Journal 2008 (2-3), 147–173.
- Keyfitz, N., 1982. Can knowledge improve forecasts? Population and Development Review, 729–751.
- Kiers, H., Kinderen, A., 2003. A fast method for choosing the numbers of components in Tucker3 analysis. British Journal of Mathematical and Statistical Psychology 56 (1), 119–125.
- Klenk, J., Rapp, K., Büchele, G., Keil, U., Weiland, S., 2007. Increasing life expectancy in Germany: quantitative contributions from changes in age- and disease-specific mortality. The European Journal of Public Health 17 (6), 587–592.
- Kogan, N., Tucker, J., Porter, M., 2011. Extending the human life span: an exploratory study of proand anti-longevity attitudes. The International Journal of Aging and Human Development 73 (1), 1–25.
- Koissi, M., Shapiro, A., Högnäs, G., 2006. Evaluating and extending the Lee-Carter model for mortality forecasting: Bootstrap confidence interval. Insurance: Mathematics and Economics 38 (1), 1–20.

- Lee, R., 2000. The Lee-Carter method for forecasting mortality, with various extensions and applications. North american actuarial journal 4 (1), 80–93.
- Lee, R., Miller, T., 2001. Evaluating the performance of the Lee-Carter method for forecasting mortality. Demography 38 (4), 537–549.
- Lee, R. D., Carter, L. R., 1992. Modeling and forecasting US mortality. Journal of the American Statistical Association 87 (419), 659–671.
- Leon, D., 2011. Trends in european life expectancy: a salutary view. International Journal of Epidemiology 40 (2), 271–277.
- Li, J., 2013. A Poisson common factor model for projecting mortality and life expectancy jointly for females and males. Population Studies 67 (1), 111–126.
- Li, J. S.-H., Hardy, M. R., 2011. Measuring basis risk in longevity hedges. North American Actuarial Journal 15 (2), 177–200.
- Li, N., Lee, R., 2005. Coherent mortality forecasts for a group of populations: An extension of the Lee-Carter method. Demography 42 (3), 575–594.
- Li, N., Lee, R., Tuljapurkar, S., 2004. Using the lee-carter method to forecast mortality for populations with limited data. International Statistical Review/Revue Internationale de Statistique, 19–36.
- Li, S.-H., Hardy, M. R., Tan, K. S., 2006. Uncertainty in mortality forecasting: an extension to the classical Lee-Carter approach. University of Waterloo.
- Liu, Y., Arai, A., Kanda, K., Lee, R., Glasser, J., Tamashiro, H., 2012. Gender gaps in life expectancy: generalized trends and negative associations with development indices in OECD countries. The European Journal of Public Health.
- Liu, Y., Arai, A., Obayashi, Y., Kanda, K., Boostrom, E., Lee, R. B., Tamashiro, H., 2013. Trends of gender gaps in life expectancy in Japan, 1947–2010: associations with gender

mortality ratio and a social development index. Geriatrics & gerontology international 13 (3), 792–797.

- Lundström, H., Qvist, J., 2004. Mortality forecasting and trend shifts: An application of the Lee–Carter model to Swedish mortality data. International Statistical Review 72 (1), 37–50.
- Mackenbach, J., 2012. Convergence and divergence of life expectancy in Europe: a centennial view. European Journal of Epidemiology.
- Mathers, C., Sadana, R., Salomon, J., Murray, C., Lopez, A., 2001. Healthy life expectancy in 191 countries, 1999. The Lancet 357 (9269), 1685–1691.
- McCartney, G., Mahmood, L., Leyland, A., Batty, G., Hunt, K., 2011. Contribution of smoking-related and alcohol-related deaths to the gender gap in mortality: evidence from 30 European countries. Tobacco Control 20 (2), 166–168.
- McCullagh, P., Nelder, J. A., 1989. Generalized linear models.
- McMichael, A., McKee, M., Shkolnikov, V., Valkonen, T., 2004. Mortality trends and setbacks: global convergence or divergence? The Lancet 363 (9415), 1155–1159.
- Men, T., Brennan, P., Boffetta, P., Zaridze, D., 2003. Russian mortality trends for 1991-2001: analysis by cause and region. Bmj 327 (7421), 964.
- Meslé, F., Vallin, J., et al., 2002. Mortality in Europe: the divergence between east and west. Population (English Edition) 57 (1), 157–197.
- Moser, K., Shkolnikov, V., Leon, D., 2005. World mortality 1950-2000: divergence replaces convergence from the late 1980s. Bulletin of the World Health Organization 83 (3), 202–209.

- Murray, C., Rosenfeld, L., Lim, S., Andrews, K., Foreman, K., Haring, D., Fullman, N., Naghavi, M., Lozano, R., Lopez, A., 2012. Global malaria mortality between 1980 and 2010: a systematic analysis. The Lancet 379 (9814), 413–431.
- Nolte, E., Shkolnikov, V., McKee, M., 2000a. Changing mortality patterns in East and West Germany and Poland. I: Long term trends (1960–1997). Journal of Epidemiology and Community Health 54 (12), 890–898.
- Nolte, E., Shkolnikov, V., McKee, M., 2000b. Changing mortality patterns in East and West Germany and Poland. II: Short-term trends during transition and in the 1990s. Journal of Epidemiology and Community Health 54 (12), 899–906.
- Plat, R., 2009a. On stochastic mortality modeling. Insurance Mathematics and Economics 45, 393–404.
- Plat, R., 2009b. Stochastic portfolio specific mortality and the quantification of mortality basis risk. Insurance: Mathematics and Economics 45 (1), 123–132.
- Preston, S., Heuveline, P., Guillot, M., 2001. Demography: measuring and modeling population processes. Population and Development Review 27, 365.
- Preston, S., et al., 1976. Mortality patterns in national populations. With special reference to recorded causes of death. Academic Press, Inc., Ltd., 24/28 Oval Road, London, NWI.
- Ramsay, J. O., Silverman, B. W., 2005. Functional data analysis (2nd edition). Springer, New York.
- Renshaw, A., Haberman, S., 2003a. Lee-Carter mortality forecasting: a parallel generalized linear modelling approach for England and Wales mortality projections. Applied statistics 52 (1), 119–137.

- Renshaw, A., Haberman, S., 2003b. On the forecasting of mortality reduction factors. Insurance: Mathematics and Economics 32 (3), 379–401.
- Renshaw, A. E., Haberman, S., 2003c. Lee-Carter mortality forecasting with age-specific enhancement. Insurance Mathematics and Economics 33 (2), 255–272.
- Renshaw, A. E., Haberman, S., 2006. A cohort-based extension to the Lee-Carter model for mortality reduction factors. Insurance Mathematics and Economics 38 (3), 556– 570.
- Richel, T., 2003. Will human life expectancy quadruple in the next hundred years? Sixty gerontologists say public debate on life extension is necessary. Journal of anti-aging medicine 6 (4), 309–314.
- Rogers, A., Planck, F., 1983. Model: A General Program for Estimating Parametrized Model Schedules of Fertility, Mortality, Migration, and Marital and Labor Force Status Transitions. IIASA, International Institute for Applied Systems Analysis.
- Russolillo, M., Giordano, G., Haberman, S., 2011. Extending the Lee–Carter model: a three-way decomposition. Scandinavian Actuarial Journal 2011 (2), 96–117.
- Saito, M., Kondo, N., Kondo, K., Ojima, T., Hirai, H., 2012. Gender differences on the impacts of social exclusion on mortality among older Japanese: AGES cohort study. Social Science & Medicine.
- Shang, H., Hyndman, R., Booth, H., 2010. A comparison of ten principal component methods for forecasting mortality rates. Monash Econometrics and Business Statistics Working Papers.
- Shkolnikov, V., Cornia, G., Leon, D., Meslé, F., 1998. Causes of the Russian mortality crisis: evidence and interpretations. World Development 26 (11), 1995–2011.

- Siler, W., 1983. Parameters of mortality in human populations with widely varying life spans. Statistics in Medicine 2 (3), 373–380.
- Tabeau, E., 2002. A review of demographic forecasting models for mortality. In: Forecasting Mortality in Developed Countries. Springer, pp. 1–32.
- Tickle, L., Booth, H., 2014. The longevity prospects of Australian seniors: An evaluation of forecast method and outcome. Asia-Pacific Journal of Risk and Insurance 8(2), 149– 327.
- Trovato, F., Heyen, N., 2006. A varied pattern of change of the sex differential in survival in the G7 countries. Journal of Biosocial Science 38 (3), 391.
- Tucker, L., 1966. Some mathematical notes on three-mode factor analysis. Psychometrika 31 (3), 279–311.
- Tuljapurkar, S., Boe, C., 1998. Mortality change and forecasting: how much and how little do we know? North American Actuarial Journal 2, 13–47.
- Tuljapurkar, S., Li, N., Boe, C., 2000. A universal pattern of mortality decline in the G7 countries. Nature 405 (6788), 789–792.
- Turner, H., Firth, D., 2012. Generalized nonlinear models in r: An overview of the gnm package.(r package version 1.0-7).

URL http://go.warwick.ac.uk/gnm

- Urassa, M., Boerma, J., Isingo, R., Ngalula, J., Ng'weshemi, J., Mwaluko, G., Zaba, B., 2001. The impact of hiv/aids on mortality and household mobility in rural tanzania. Aids 15 (15), 2017–2023.
- Vaupel, J., Zhang, Z., van Raalte, A., 2011. Life expectancy and disparity: an international comparison of life table data. BMJ Open 1 (1).

- Villegas, A. M., Haberman, S., 2014. On the modeling and forecasting of socioeconomic mortality differentials: An application to deprivation and mortality in england. North American Actuarial Journal 18 (1), 168–193.
- Waldron, H., 2005. Literature review of long-term mortality projections. Social Security Bulletin 66, 16.
- Wang, D., Lu, P., 2005. Modelling and forecasting mortality distributions in England andWales using the Lee–Carter model. Journal of Applied Statistics 32 (9), 873–885.
- Wang, S., 2000. A class of distortion operators for pricing financial and insurance risks. Journal of Risk and Insurance 67 (1), 15–36.
- White, K., 2002. Longevity advances in high-income countries, 1955–96. Population and Development Review 28 (1), 59–76.
- Wilmoth, J., 1998. Is the pace of japanese mortality decline converging toward international trends? Population and Development Review, 593–600.
- Wilmoth, J., Andreev, K., Jdanov, D., Glei, D., 2007. Methods protocol for the human mortality database. University of California, Berkeley, and Max Planck Institute for Demographic Research, Rostock. URL: http://mortality.org [version 31/05/2007].
- Wilmoth, J. R., 1993. Computational methods for fitting and extrapolating the lee-carter model of mortality change. Tech. rep., Technical report, Department of Demography, University of California, Berkeley.
- Wilson, C., 2001. On the scale of global demographic convergence 1950–2000. Population and Development Review 27 (1), 155–171.
- Wilson, C., 2011. Understanding global demographic convergence since 1950. Population and Development Review 37 (2), 375–388.

Wood, S., 2000. Modelling and smoothing parameter estimation with multiple quadratic penalties. Journal of the Royal Statistical Society: Series B (Statistical Methodology) 62 (2), 413–428.

Appendix A

Estimates of the JWT model

	Estimate	Std. Error	t value	Pr(> t)
Age(0)	0.01782709	0.00090909	19.60992584	0.00000000
Age(1)	0.01841896	0.00091381	20.15617020	0.00000000
Age(2)	0.01863512	0.00091430	20.38173366	0.00000000
Age(3)	0.01878886	0.00091458	20.54377661	0.00000000
Age(4)	0.01883511	0.00091477	20.58994717	0.00000000
Age(5)	0.01888840	0.00091493	20.64471498	0.00000000
Age(6)	0.01888866	0.00091505	20.64213159	0.00000000
Age(7)	0.01889913	0.00091516	20.65107134	0.00000000
Age(8)	0.01889693	0.00091526	20.64647991	0.00000000
Age(9)	0.01888071	0.00091535	20.62684215	0.00000000
Age(10)	0.01885134	0.00091543	20.59298056	0.00000000
Age(11)	0.01881550	0.00091550	20.55215337	0.00000000
Age(12)	0.01879698	0.00091557	20.53026837	0.00000000
Age(13)	0.01875865	0.00091565	20.48666308	0.00000000

Table A.1: Estimates of the Parameters of the JWT model usingGLM (across countries)

175

Age(14)	0.01872694	0.00091573	20.45019948	0.00000000
Age(15)	0.01870919	0.00091583	20.42872331	0.00000000
Age(16)	0.01864974	0.00091593	20.36142910	0.00000000
Age(17)	0.01859902	0.00091606	20.30331402	0.00000000
Age(18)	0.01854863	0.00091620	20.24527742	0.00000000
Age(19)	0.01849138	0.00091635	20.17946163	0.00000000
Age(20)	0.01843792	0.00091651	20.11763967	0.00000000
Age(21)	0.01841050	0.00091667	20.08414402	0.00000000
Age(22)	0.01842139	0.00091684	20.09235069	0.00000000
Age(23)	0.01844503	0.00091701	20.11432291	0.00000000
Age(24)	0.01845066	0.00091719	20.11652716	0.00000000
Age(25)	0.01843071	0.00091737	20.09072831	0.00000000
Age(26)	0.01841372	0.00091757	20.06798248	0.00000000
Age(27)	0.01839028	0.00091777	20.03806589	0.00000000
Age(28)	0.01835083	0.00091797	19.99056238	0.00000000
Age(29)	0.01829769	0.00091819	19.92791541	0.00000000
Age(30)	0.01828072	0.00091842	19.90449259	0.00000000
Age(31)	0.01819528	0.00091866	19.80631377	0.00000000
Age(32)	0.01814251	0.00091891	19.74353033	0.00000000
Age(33)	0.01806722	0.00091918	19.65586191	0.00000000
Age(34)	0.01801479	0.00091946	19.59273473	0.00000000
Age(35)	0.01792207	0.00091977	19.48548331	0.00000000
Age(36)	0.01783472	0.00092009	19.38361140	0.00000000
Age(37)	0.01772493	0.00092044	19.25694270	0.00000000
Age(38)	0.01763032	0.00092082	19.14635607	0.00000000
Age(39)	0.01753217	0.00092123	19.03134093	0.00000000
Age(40)	0.01744354	0.00092166	18.92611933	0.00000000
Age(41)	0.01727108	0.00092214	18.72927958	0.00000000
Age(42)	0.01716848	0.00092265	18.60772091	0.00000000

Age(43)	0.01702592	0.00092322	18.44189167	0.00000000
Age(44)	0.01689263	0.00092384	18.28532737	0.00000000
Age(45)	0.01672731	0.00092450	18.09328954	0.00000000
Age(46)	0.01655716	0.00092524	17.89507454	0.00000000
Age(47)	0.01638085	0.00092604	17.68920775	0.00000000
Age(48)	0.01622354	0.00092692	17.50270764	0.00000000
Age(49)	0.01606822	0.00092789	17.31700537	0.00000000
Age(50)	0.01591322	0.00092894	17.13051420	0.00000000
Age(51)	0.01575807	0.00093009	16.94243276	0.00000000
Age(52)	0.01562982	0.00093133	16.78220997	0.00000000
Age(53)	0.01551967	0.00093271	16.63934366	0.00000000
Age(54)	0.01540839	0.00093419	16.49392378	0.00000000
Age(55)	0.01529377	0.00093581	16.34284754	0.00000000
Age(56)	0.01518375	0.00093756	16.19503641	0.00000000
Age(57)	0.01506394	0.00093948	16.03433545	0.00000000
Age(58)	0.01499987	0.00094157	15.93065339	0.00000000
Age(59)	0.01494371	0.00094390	15.83191273	0.00000000
Age(60)	0.01489833	0.00094644	15.74148573	0.00000000
Age(61)	0.01483997	0.00094928	15.63293210	0.00000000
Age(62)	0.01483479	0.00095234	15.57717988	0.00000000
Age(63)	0.01479993	0.00095581	15.48418252	0.00000000
Age(64)	0.01483464	0.00095964	15.45848399	0.00000000
Age(65)	0.01491487	0.00096391	15.47326435	0.00000000
Age(66)	0.01494627	0.00096869	15.42929811	0.00000000
Age(67)	0.01502214	0.00097398	15.42351850	0.00000000
Age(68)	0.01513506	0.00097993	15.44508148	0.00000000
Age(69)	0.01526349	0.00098663	15.47036622	0.00000000
Age(70)	0.01548891	0.00099415	15.58000610	0.00000000
Age(71)	0.01568484	0.00100280	15.64098671	0.00000000

Age(72)	0.01595791	0.00101241	15.76237281	0.00000000
Age(73)	0.01625121	0.00102359	15.87670071	0.00000000
Age(74)	0.01650701	0.00103629	15.92887892	0.00000000
Age(75)	0.01691740	0.00105084	16.09889832	0.00000000
Age(76)	0.01727187	0.00106754	16.17907200	0.00000000
Age(77)	0.01762345	0.00108680	16.21594509	0.00000000
Age(78)	0.01808525	0.00110886	16.30975104	0.00000000
Age(79)	0.01844446	0.00113457	16.25675971	0.00000000
Age(80)	0.01885443	0.00116432	16.19345755	0.00000000
Age(81)	0.01922706	0.00119904	16.03542281	0.00000000
Age(82)	0.01965776	0.00123886	15.86765934	0.00000000
Age(83)	0.02006941	0.00128629	15.60259948	0.00000000
Age(84)	0.02050895	0.00134204	15.28192372	0.00000000
Age(85)	0.02094997	0.00140846	14.87442885	0.00000000
Age(86)	0.02141591	0.00148683	14.40377445	0.00000000
Age(87)	0.02187095	0.00158085	13.83493114	0.00000000
Age(88)	0.02233385	0.00169277	13.19364094	0.00000000
Age(89)	0.02272675	0.00182711	12.43864519	0.00000000
k(1950)	0.03029481	0.00081518	37.16333620	0.00000000
k(1951)	-0.00137056	0.00081459	-1.68251698	0.09247403
k(1952)	0.02411362	0.00081194	29.69863660	0.00000000
k(1953)	0.00683748	0.00081108	8.43008448	0.00000000
k(1954)	0.02233795	0.00080894	27.61371977	0.00000000
k(1955)	-0.00173513	0.00080815	-2.14703888	0.03179433
k(1956)	0.00070204	0.00080784	0.86902729	0.38483581
k(1957)	-0.01635338	0.00080763	-20.24868184	0.00000000
k(1958)	0.02549078	0.00080550	31.64610481	0.00000000
k(1959)	-0.00857766	0.00080504	-10.65495522	0.00000000
k(1960)	0.00260798	0.00080449	3.24179370	0.00118847

k(1961)	-0.00169768	0.00080358	-2.11264594	0.03463536
k(1962)	-0.01312844	0.00080376	-16.33369455	0.00000000
k(1963)	-0.00332151	0.00080329	-4.13487428	0.00003556
k(1964)	0.00788397	0.00080207	9.82956629	0.00000000
k(1965)	-0.00544567	0.00080188	-6.79109606	0.00000000
k(1966)	-0.00260173	0.00080128	-3.24695475	0.00116713
k(1967)	-0.00225069	0.00080058	-2.81133844	0.00493525
k(1968)	-0.01235102	0.00080060	-15.42726510	0.00000000
k(1969)	-0.01022997	0.00080045	-12.78032603	0.00000000
k(1970)	0.00244362	0.00079953	3.05632269	0.00224172
k(1971)	-0.00512598	0.00079898	-6.41563275	0.00000000
k(1972)	0.00512201	0.00079825	6.41653291	0.00000000
k(1973)	-0.00304730	0.00079774	-3.81991696	0.00013364
k(1974)	0.00236982	0.00079707	2.97316386	0.00294869
k(1975)	0.00368710	0.00079637	4.62988752	0.00000367
k(1976)	0.00052510	0.00079590	0.65976007	0.50941044
k(1977)	0.00842238	0.00079481	10.59666649	0.00000000
k(1978)	-0.00469463	0.00079443	-5.90941826	0.00000000
k(1979)	-0.00255675	0.00079385	-3.22068257	0.00127956
k(1980)	-0.00328128	0.00079354	-4.13500919	0.00003554
k(1981)	0.00430583	0.00079297	5.43001044	0.0000006
k(1982)	-0.00468133	0.00079246	-5.90734795	0.00000000
k(1983)	-0.01168069	0.00079227	-14.74338593	0.00000000
k(1984)	0.00491215	0.00079159	6.20538609	0.00000000
k(1985)	-0.01706375	0.00079166	-21.55436251	0.00000000
k(1986)	-0.00597011	0.00079125	-7.54519892	0.00000000
k(1987)	-0.00496561	0.00079073	-6.27979733	0.00000000
k(1988)	-0.01435672	0.00079064	-18.15831986	0.00000000
k(1989)	-0.00821320	0.00079024	-10.39325229	0.00000000

k(1990)	-0.00882158	0.00078998	-11.16679292	0.00000000
k(1991)	-0.00445558	0.00078962	-5.64268119	0.00000002
k(1992)	-0.00118262	0.00078924	-1.49843286	0.13402628
k(1993)	-0.00756635	0.00078917	-9.58776185	0.00000000
k(1994)	-0.00001708	0.00078860	-0.02166364	0.98271634

Appendix B

"Across country" analysis, separately for females and males

Madala		Individua	1			Joint		
widdels	WT	LC	HU	SJLC	JLC	TLC	PLC	JWT
AUS	0.017	-0.024	-0.024	0.035	0.090	-0.019	0.010	-0.014
CAN	0.092	0.014	0.033	0.074	0.120	0.032	0.049	0.025
ENW	0.076	0.031	0.049	0.101	0.152	0.034	0.076	0.052
FIN	0.152	0.090	0.084	0.064	0.112	0.084	0.039	0.015
FRA	0.075	-0.009	-0.004	0.028	0.057	-0.009	0.004	-0.021
ITA	0.062	-0.019	-0.041	-0.046	0.013	-0.025	-0.071	-0.095
ESP	0.134	0.064	-0.080	-0.000	0.040	0.055	-0.025	-0.049
SWE	0.104	0.052	-0.031	0.114	0.172	0.065	0.089	0.065
CHE	-0.006	-0.077	-0.061	-0.035	0.016	-0.072	-0.060	-0.084
USA	0.058	0.018	0.027	0.117	0.223	0.029	0.092	0.068
JPN	0.324	0.258	0.057	0.074	-0.004	0.225	0.050	0.025
DNK	-0.076	-0.157	-0.100	-0.021	0.103	-0.121	-0.046	-0.070
NLD	0.048	-0.011	-0.001	0.086	0.190	0.016	0.061	0.037
Mean	0.081	0.018	-0.007	0.045	0.099	0.023	0.021	-0.004
SD	0.094	0.096	0.056	0.056	0.072	0.083	0.056	0.056

Table B.1: ME of the forecast log death rate for females of the 13 countries

Table B.2: ME of the forecast log death rate for males of the 13 countries

Models		Individual	[Joint		
widdels	WT	LC	HU	SJLC	JLC	TLC	PLC	JWT
AUS	-0.066	-0.101	-0.092	-0.075	-0.053	-0.100	-0.085	-0.045
CAN	-0.048	-0.098	-0.104	-0.076	-0.031	-0.093	-0.086	-0.047
ENW	-0.008	-0.047	-0.046	-0.015	0.025	-0.051	-0.025	0.015
FIN	0.076	0.028	0.015	-0.027	-0.014	-0.003	-0.037	0.003
FRA	-0.058	-0.129	-0.129	-0.096	-0.053	-0.132	-0.106	-0.066
ITA	-0.049	-0.113	-0.130	-0.128	-0.110	-0.117	-0.139	-0.099
ESP	0.018	-0.066	-0.142	-0.087	-0.006	-0.060	-0.098	-0.058
SWE	-0.019	-0.038	-0.106	-0.026	0.038	-0.052	-0.036	0.004
CHE	-0.166	-0.216	-0.208	-0.194	-0.111	-0.214	-0.204	-0.164
USA	-0.081	-0.116	-0.115	-0.052	0.045	-0.110	-0.062	-0.022
JPN	0.219	0.134	-0.033	0.002	-0.075	0.110	-0.008	0.032
DNK	-0.092	-0.172	-0.173	-0.056	0.057	-0.131	-0.066	-0.026
NLD	-0.053	-0.074	-0.099	-0.033	0.032	-0.073	-0.044	-0.004
Mean	-0.025	-0.077	-0.105	-0.066	-0.020	-0.079	-0.077	-0.037
SD	0.093	0.088	0.059	0.053	0.058	0.076	0.053	0.053

Models		Individua	1			Joint		
WIDdels	WT	LC	HU	SJLC	JLC	TLC	PLC	JWT
AUS	3	5	6	7	8	4	1	2
CAN	7	1	4	6	8	3	5	2
ENW	5	1	3	7	8	2	6	4
FIN	8	6	4	3	7	5	2	1
FRA	8	4	2	6	7	3	1	5
ITA	6	2	4	5	1	3	7	8
ESP	8	6	7	1	3	5	2	4
SWE	6	2	1	7	8	4	5	3
CHE	1	7	5	3	2	6	4	8
USA	4	1	2	7	8	3	6	5
JPN	8	7	4	5	1	6	3	2
DNK	4	8	5	1	6	7	2	3
NLD	5	2	1	7	8	3	6	4
Mean	7	3	2	6	8	5	4	1
SD	7	8	1	3	5	6	2	3

Table B.3: Ranking of ME for females of the 13 countries

Table B.4: Ranking of ME for males of the 13 countries

Madala		Individua	1			Joint		
Models	WT	LC	HU	SJLC	JLC	TLC	PLC	JWT
AUS	3	8	6	4	2	7	5	1
CAN	3	7	8	4	1	6	5	2
ENW	1	7	6	3	4	8	5	2
FIN	8	6	4	5	3	1	7	2
FRA	2	6	7	4	1	8	5	3
ITA	1	4	7	6	3	5	8	2
ESP	2	5	8	6	1	4	7	3
SWE	2	6	8	3	5	7	4	1
CHE	3	8	6	4	1	7	5	2
USA	5	8	7	3	2	6	4	1
JPN	8	7	4	1	5	6	2	3
DNK	5	7	8	2	3	6	4	1
NLD	5	7	8	3	2	6	4	1
Mean	2	6	8	4	1	7	5	3
SD	8	7	5	2	4	6	1	2

Note: Mean and SD refer to the ranking of Mean and SD.

Models		Individual	1			Joint		
widdels	WT	LC	HU	SJLC	JLC	TLC	PLC	JWT
AUS	0.149	0.134	0.131	0.149	0.181	0.139	0.140	0.140
CAN	0.133	0.092	0.100	0.116	0.148	0.096	0.103	0.098
ENW	0.124	0.108	0.114	0.133	0.168	0.094	0.114	0.106
FIN	0.275	0.227	0.226	0.219	0.265	0.223	0.209	0.205
FRA	0.121	0.111	0.111	0.098	0.108	0.097	0.094	0.095
ITA	0.140	0.110	0.113	0.113	0.145	0.103	0.119	0.136
ESP	0.183	0.143	0.140	0.123	0.136	0.135	0.119	0.128
SWE	0.200	0.195	0.197	0.210	0.238	0.190	0.200	0.197
CHE	0.194	0.212	0.213	0.187	0.195	0.194	0.190	0.199
USA	0.079	0.068	0.070	0.121	0.224	0.062	0.101	0.084
JPN	0.342	0.286	0.111	0.146	0.110	0.243	0.128	0.118
DNK	0.207	0.245	0.236	0.208	0.243	0.221	0.208	0.215
NLD	0.132	0.124	0.123	0.144	0.215	0.121	0.133	0.125
Mean	0.175	0.158	0.145	0.151	0.183	0.148	0.143	0.142
SD	0.071	0.067	0.054	0.041	0.052	0.059	0.043	0.046

Table B.5: MAE of the forecast log death rate for females of the 13 countries

Table B.6: MAE of the forecast log death rate for males of the 13 countries

Models		Individual	1			Joint		
widdels	WT	LC	HU	SJLC	JLC	TLC	PLC	JWT
AUS	0.147	0.160	0.151	0.161	0.191	0.165	0.165	0.149
CAN	0.098	0.138	0.136	0.127	0.118	0.128	0.130	0.110
ENW	0.119	0.112	0.116	0.114	0.129	0.114	0.116	0.120
FIN	0.215	0.208	0.208	0.200	0.201	0.198	0.201	0.200
FRA	0.101	0.162	0.152	0.142	0.135	0.151	0.143	0.115
ITA	0.111	0.179	0.168	0.167	0.175	0.155	0.173	0.146
ESP	0.138	0.148	0.172	0.153	0.135	0.142	0.154	0.143
SWE	0.192	0.222	0.233	0.198	0.180	0.202	0.202	0.204
CHE	0.211	0.254	0.250	0.239	0.207	0.249	0.245	0.219
USA	0.103	0.130	0.128	0.109	0.124	0.120	0.108	0.098
JPN	0.231	0.156	0.111	0.092	0.106	0.139	0.093	0.096
DNK	0.201	0.256	0.256	0.195	0.182	0.221	0.198	0.192
NLD	0.123	0.157	0.161	0.133	0.129	0.139	0.136	0.125
Mean	0.153	0.176	0.172	0.156	0.155	0.163	0.159	0.147
SD	0.050	0.046	0.049	0.043	0.035	0.042	0.044	0.043

Models		Individua	ıl	Joint					
widdels	WT	LC	HU	SJLC	JLC	TLC	PLC	JWT	
AUS	6	2	1	7	8	3	5	4	
CAN	7	1	4	6	8	2	5	3	
ENW	6	3	5	7	8	1	4	2	
FIN	8	6	5	3	7	4	2	1	
FRA	8	7	6	4	5	3	1	2	
ITA	7	2	3	4	8	1	5	6	
ESP	8	7	6	2	5	4	1	3	
SWE	6	2	4	7	8	1	5	3	
CHE	4	7	8	1	5	3	2	6	
USA	4	2	3	7	8	1	6	5	
JPN	8	7	2	5	1	6	4	3	
DNK	1	8	6	2	7	5	3	4	
NLD	5	3	2	7	8	1	6	4	
Mean	7	6	3	5	8	4	2	1	
SD	8	7	5	1	4	6	2	3	

Table B.7: Ranking of MAE for females of the 13 countries

Table B.8: Ranking of MAE for males of the 13 countries

Models		Individua	.1	Joint					
Models	WT	LC	HU	SJLC	JLC	TLC	PLC	JWT	
AUS	1	4	3	5	8	7	6	2	
CAN	1	8	7	4	3	5	6	2	
ENW	6	1	5	3	8	2	4	7	
FIN	8	7	6	3	4	1	5	2	
FRA	1	8	7	4	3	6	5	2	
ITA	1	8	5	4	7	3	6	2	
ESP	2	5	8	6	1	3	7	4	
SWE	2	7	8	3	1	5	4	6	
CHE	2	8	7	4	1	6	5	3	
USA	2	8	7	4	6	5	3	1	
JPN	8	7	5	1	4	6	2	3	
DNK	5	7	8	3	1	6	4	2	
NLD	1	7	8	4	3	6	5	2	
Mean	2	8	7	4	3	6	5	1	
SD	8	6	7	3	1	2	5	4	

Note: Mean and SD refer to the ranking of Mean and SD.

Models	Individual			Joint					
widdels	WT	LC	HU	SJLC	JLC	TLC	PLC	JWT	
AUS	10.87	0.00	-2.04	11.18	35.26	3.69	4.63	4.45	
CAN	45.04	0.00	9.02	25.62	60.84	4.29	11.96	6.63	
ENW	15.04	0.00	5.88	23.54	55.81	-12.65	5.45	-2.22	
FIN	21.01	0.00	-0.68	-3.45	16.61	-2.00	-8.04	-9.68	
FRA	8.47	0.00	-0.03	-12.41	-3.09	-12.47	-15.24	-14.33	
ITA	27.28	0.00	2.21	2.87	31.80	-6.60	8.14	23.69	
ESP	28.24	0.00	-2.40	-13.66	-5.08	-5.84	-16.71	-10.73	
SWE	2.40	0.00	1.01	7.37	21.62	-2.80	2.35	0.71	
CHE	-8.52	0.00	0.36	-11.94	-8.15	-8.64	-10.41	-6.13	
USA	16.96	0.00	2.81	79.02	229.80	-8.47	48.54	23.24	
JPN	19.68	0.00	-61.31	-48.71	-61.60	-14.78	-55.15	-58.58	
DNK	-15.83	0.00	-3.99	-15.20	-1.12	-9.83	-15.15	-12.45	
NLD	6.39	0.00	-0.49	16.51	73.89	-1.96	7.75	1.21	
Mean	10.86	0.00	-8.35	-4.25	15.53	-6.67	-9.57	-10.21	
SD	5.01	0.00	-20.17	-39.10	-23.09	-11.74	-36.30	-31.63	

Table B.9: CMAE based on the LC model for females of the 13 countries

Table B.10: CMAE based on the LC model for males of the 13 countries

Madala		Individua	ıl	Joint					
Models	WT	LC	HU	SJLC	JLC	TLC	PLC	JWT	
AUS	-7.82	0.00	-5.15	0.97	19.81	3.13	3.07	-6.71	
CAN	-28.85	0.00	-1.39	-8.14	-14.78	-7.84	-5.87	-20.34	
ENW	6.65	0.00	3.59	2.30	15.28	1.93	3.57	7.48	
FIN	3.34	0.00	-0.24	-3.84	-3.77	-4.90	-3.40	-4.20	
FRA	-37.59	0.00	-5.86	-12.34	-16.22	-6.64	-11.74	-28.90	
ITA	-37.90	0.00	-6.49	-7.06	-2.19	-13.48	-3.77	-18.47	
ESP	-6.81	0.00	15.81	3.17	-8.99	-4.07	3.87	-3.69	
SWE	-13.77	0.00	4.71	-11.05	-18.78	-9.10	-9.20	-8.23	
CHE	-16.96	0.00	-1.55	-5.86	-18.49	-2.11	-3.43	-13.93	
USA	-20.32	0.00	-1.06	-15.93	-4.56	-7.48	-17.05	-24.77	
JPN	48.28	0.00	-28.59	-40.72	-31.88	-11.15	-40.41	-38.51	
DNK	-21.23	0.00	0.01	-23.75	-28.99	-13.47	-22.57	-24.71	
NLD	-21.73	0.00	2.35	-15.28	-18.15	-11.70	-13.69	-20.62	
Mean	-12.71	0.00	-1.76	-11.02	-11.84	-7.03	-9.63	-16.04	
SD	7.76	0.00	7.06	-7.15	-23.62	-9.28	-3.98	-7.04	

Note: Mean and SD refer to the CMAE of Mean and SD according to Table B.5 and B.6

Appendix C

"Across sex" analysis, separately for each country

Models		Individual		Joint					
Models	WT	LC	HU	SJLC	JLC	TLC	PLC	JWT	
AUS	0.109	0.076	0.073	0.083	0.077	0.085	0.108	0.114	
CAN	0.136	0.089	0.079	0.109	0.092	0.109	0.145	0.114	
ENW	0.227	0.125	0.129	0.157	0.125	0.185	0.244	0.181	
FIN	0.095	0.075	0.074	0.084	0.076	0.084	0.115	0.103	
FRA	0.367	0.146	0.148	0.228	0.171	0.216	0.407	0.468	
ITA	0.492	0.168	0.179	0.298	0.228	0.314	0.454	0.642	
ESP	0.375	0.192	0.315	0.271	0.159	0.340	0.337	0.443	
SWE	0.098	0.071	0.074	0.093	0.078	0.086	0.103	0.111	
CHE	0.090	0.069	0.072	0.079	0.075	0.075	0.095	0.100	
USA	0.630	0.204	0.239	0.350	0.237	0.346	0.735	0.400	
JPN	0.855	0.342	0.300	0.585	0.315	0.569	0.664	0.713	
DNK	0.099	0.068	0.073	0.079	0.076	0.080	0.086	0.083	
NLD	0.127	0.082	0.085	0.120	0.087	0.113	0.146	0.150	
Mean	0.285	0.131	0.142	0.195	0.138	0.200	0.280	0.279	
SD	0.246	0.080	0.090	0.150	0.079	0.152	0.224	0.225	

Table C.1: Log-likelihood of each of the 13 countries

Models	Individual			Joint					
Models	WT	LC	HU	SJLC	JLC	TLC	PLC	JWT	
AUS	0.109	0.081	0.082	0.089	0.082	0.087	0.110	0.115	
CAN	0.136	0.093	0.087	0.115	0.097	0.112	0.147	0.116	
ENW	0.227	0.129	0.135	0.162	0.130	0.188	0.246	0.183	
FIN	0.095	0.079	0.080	0.089	0.081	0.087	0.117	0.104	
FRA	0.367	0.151	0.155	0.233	0.177	0.219	0.409	0.469	
ITA	0.492	0.172	0.185	0.303	0.233	0.317	0.456	0.644	
ESP	0.375	0.196	0.319	0.276	0.164	0.343	0.339	0.444	
SWE	0.098	0.075	0.078	0.098	0.084	0.089	0.106	0.112	
CHE	0.090	0.073	0.076	0.085	0.080	0.078	0.097	0.101	
USA	0.630	0.208	0.243	0.355	0.242	0.349	0.737	0.401	
JPN	0.855	0.346	0.305	0.590	0.321	0.572	0.666	0.714	
DNK	0.099	0.072	0.077	0.085	0.082	0.082	0.088	0.084	
NLD	0.127	0.086	0.089	0.125	0.092	0.116	0.148	0.151	
Mean	0.285	0.135	0.147	0.200	0.143	0.203	0.282	0.280	
SD	0.246	0.080	0.089	0.150	0.079	0.152	0.224	0.225	

Table C.2: BIC of each of the 13 countries

Table C.3: Ranking of BIC

Models	Individual			Joint					
Models	WT	LC	HU	SJLC	JLC	TLC	PLC	JWT	
AUS	6	1	2	5	3	4	7	8	
CAN	7	2	1	5	3	4	8	6	
ENW	7	1	3	4	2	6	8	5	
FIN	6	1	2	5	3	4	8	7	
FRA	6	1	2	5	3	4	7	8	
ITA	7	1	2	4	3	5	6	8	
ESP	7	2	4	3	1	6	5	8	
SWE	6	1	2	5	3	4	7	8	
CHE	6	1	2	5	4	3	7	8	
USA	7	1	3	5	2	4	8	6	
JPN	8	3	1	5	2	4	6	7	
DNK	8	1	2	6	3	4	7	5	
NLD	6	1	2	5	3	4	7	8	
Mean	8	1	3	4	2	5	7	6	
SD	8	2	3	4	1	5	6	7	