UNIVERSITÀ
DEGLI STUDI
DI PADOVA

Sede Amministrativa: Università degli Studi di Padova

Dipartimento di Scienze Statistiche
SCUOLA DI DOTTORATO DI RICERCA IN SCIENZE STATISTICHE
CICLO XXVI

Forecasting Mortality in Related Populations Using Lee-Carter Type Models

Direttore della Scuola: Ch.ma Prof.ssa Monica Chiogna

Supervisore: Ch.mo Prof. Nicola Torelli

Co-supervisori: Ch.mo Prof. Steven Haberman
Dott. Pietro Millossovich
Ch.mo Prof. Ermanno Pitacco

Dottorando: Ivan Luciano Danesi

31 Gennaio 2014

Acknowledgements

Many people have helped and supported me during these past few years and I have learned valuable lessons from each of them.

I am grateful for the guidance of my advisor Professor Torelli. He helped me to choose a field appropriate to my background, he constantly reminded me of the aims that I had to pursue and he encouraged me when I needed it the most. He has been a great mentor.

Professor Pitacco introduced me to the idea of forecasting mortality, and he always gave me supportive advice that helped me to see the topic from the right perspective.

I am also grateful to Professor Haberman for giving me the opportunity to work with him at the Sir John Cass Business School in London. He provided me with invaluable help in deciding on a fruitful research direction.

I am extremely indebted to Dr. Millossovich who spent a lot of time with me and patiently followed my research, step by step. From Dr. Millossovich, I learned about programming, the research process and the importance of being reflective and precise.

I would like to thanks Dr. Kaucic, with whom I had several interesting discussions in Trieste. One of these discussions evolved into a project last year.

Best wishes to my Ph.D. fellows, Akram, Erlis, Lorenzo, Luca, Md Abud Darda, Roberta and Shireen. Thank you all for the laughs and challenges we shared over the years. In Padova, other than first-class courses, I found friends. Regarding this, I should mention Tonio, my Wednesday guest.

Finally, special thanks to my family and my girlfriend. Their love and support have made all of this possible.

Padua, Ivan Luciano Danesi

24 January 2014

Abstract

Some aspects of modern society are planned according to the values of expected future mortality rates. Due to the relevance of this issue, several approaches for treating this problem have been proposed. Among them, one of the most influential is the Lee-Carter model.

The aim of this thesis is to forecast mortality rates of related populations. In order to do this, some models based on Lee-Carter approach are considered; the models are applied to central death rates and to mortality improvement rates. Firstly, the models are discussed in a qualitative way. Secondly, the models are evaluated on a real dataset and their ability in fitting the data and forecasting are compared. The results highlight strengths and weaknesses of the different approaches.

A further discussion relates applications of the models on mortality improvement rates. More specifically, the hypothesis of constant variance of the parametric structure is discussed and the impact of changes in this assumption is investigated by means of an application on a real dataset.

Sommario

Alcuni aspetti della società moderna sono pianificati tenendo conto dei valori futuri attesi dei livelli di mortalità. La rilevanza di tali questioni ha determinato un forte interesse riguardo ai modelli per la previsione dei tassi di martalità futuri, e a fatto sì che negli anni numerosi approcci siano stati proposti per trattare questo problema. Tra questi, il modello di Lee-Carter presentato nel 1992 è senza dubbio uno dei più influenti.

Nella presente tesi, a partire dal modello di Lee-Carter si considera il problema di previsione dei tassi di mortalità per più popolazioni che presentano caratteristiche in comune. A tal fine vengono proposti diversi modelli, alcuni dei quali sono applicati ai tassi centrali di mortalità, mentre altri sugli incrementi di questi ultimi. Innanzitutto i modelli proposti sono analizzati e confrontati in modo qualitativo. Successivamente i modelli sono applicati a dati reali e sono comparati riguardo le capacità di adattamento e previsiva. I risultati evidenziano punti di forza e di debolezza dei modelli considerati.

Infine, con riguardo ai modelli applicati agli incrementi dei tassi centrali di mortalità, viene analizzata l'ipotesi di varianza costante della struttura parametrica. L'impatto di cambiamenti in questa assunzione viene analizzato mediante un'ulteriore applicazione a dati reali.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Life expectancy at birth has changed during the history in response to human life changes and technological improvements. Among early humans this figure was very different from that observed nowadays and ranged between 20 and 30 years. Around 1750, life expectancy at birth in the more developed countries was still around 30-40 years. The improvements began to increase more rapidly starting by the end of the 19th century: it raised from 40-45 years to 60-65 around 1950, up to over 70 years in the first years of the 21st century. Much of these changes have happened in the last 150 years and there is no evidence that improvements in longevity are slowing down. These changes in longevity are not uniform with respect to ages: the age specific mortality levels structure changed its shape in response to the evolution of society. An example of this is the reduction of mortality at early ages due to improvements in medicine.

The level of mortality influence several aspects of our society. As a matter of fact, the private and the public retirement systems, as well as other components of the social security system, are planned and modified according to the values assumed by mortality rates. In this context, it is important to know and forecast mortality for any age and calendar year. Among the applications, one of the more notable examples is related to correctly pricing and reserving for life insurance products.

Due to the relevance of this issue, several models have been proposed for forecasting future mortality, and the literature is now wide. Nonetheless, none of these approaches is considered uniformly better than the others. In the recent years, interest on such models has grown and this is one of the hottest topic in actuarial research. The present dissertation is aimed at studying developments on one of

the most relevant models, introduced by Lee and Carter (1992). More specifically, in this work the model is generalized to take into account multiple populations; then the model variants are applied both on central death rates and on mortality improvement rates.

A short review on the topic is given in Section 1.1. In Section 1.2 the summary and the main contributions of the thesis are outlined.

1.1 Overview

In the previous section it is mentioned the fact that the structure of mortality rates is changing, and that this process became faster in the last 100-150 years. Observing a survival function (defined in Section 2.3.1), two main phenomena, the so called rectangularization and expansion, emerge. The first one, rectangularization, refers to the fact that deaths tend to concentrate around the upper limit age, and the second one, expansion, means that this boundary tends to increase with time. There are different views about what should be expected for the future; some authors (Olshansky *et al.*, 2005) argue that life expectancy might level off or decline, while others (Oeppen and Vaupel, 2002) claim that there could not be a limit to human life.

In actuarial sciences, one of the fields where this issue is strongly discussed, the risk of underestimating mortality is called longevity risk. An insurance company cannot consider too much prudential levels of mortality, otherwise they would be not able to offer competitive prices. On the other hand, underestimating the longevity of the customers could influence the solvency of the insurance company.

In population studies, many models were proposed for modelling and projecting demographic quantities (see Booth (2006) for a review of the modern demographic projection models). Regarding the forecasting of mortality, in many fields, such as the actuarial one, the level of mortality is computed for every age and for every calendar year of interest, then these values are ordered in a mortality matrix: this is the data here considered.

Among the models used for projecting mortality rates, one of the most influential is the one proposed in Lee and Carter (1992). This approach has received, in the last years, great deal of attention and has been extended in several directions. Some of

the extensions focus on different error structure (Brouhns *et al.*, 2002), addition of extra time factors (Renshaw and Haberman, 2003b) and the introduction of the cohort effect (Renshaw and Haberman, 2006).

In this work the Lee-Carter model is taken as starting point for mortality modelling. The reasons for choosing this model are related to its main characteristics. First of all it has a simple structure and do not incorporate any knowledge on the studied phenomenon. Therefore without particular intensive computations and without taking into account biological, environmental or other factors, it is possible to obtain a set of parameters which describe the observed phenomenon. This simplicity is not a drawback of the model: should be borne in mind that the aim of the study is, in fact, to obtain a good projection regarding the future evolution of mortality, not to analyse the observed values. The Lee-Carter model catch the general trend of a mortality table and permits to project it into the future.

This thesis focuses on a specific issue related to forecasting mortality rates, that is forecasting the mortality of more than one population at the same time. It is then necessary that the different populations share some common characteristics, such as similar socio-economic conditions, climate and geographical environment, common genetic traits between the individuals, as well as other close connections. Another important requirement is that these conditions should be expected to continue in the future.

When a group of linked populations is observed, it is expected that similarities and differences among them would be reflected even on mortality. In such a group of populations the mortality should be considered linked but not equal, therefore a specific way of forecasting mortality is needed.

In Li and Lee (2005) the importance of forecasting mortality jointly for related populations was highlighted and, subsequently, other approaches for dealing with multiple population data were proposed, as presented in the next chapter.

1.2 Summary and main contributions

The main contributions of this thesis regard extensions of the Lee-Carter model in order to apply it to multiple populations datasets. The proposed approaches to the problem consider common and/or specific factors, in order to give a series of

models with different complexity, applied to the dataset or a transformed version of it, the relative mortality improvement rates.

A short introduction to mortality rates, as well as some notions about forecasting mortality, are outlined in Chapter 2. In the same chapter, a presentation of the Lee-Carter model, with the review of the extensions that are relevant for this work is provided.

In Chapter 3 the ten models used for the analysis are presented. The models are explained giving attention to similar approaches present in the literature and the relations between the models themselves. The models are then applied to regional Italian mortality data in Chapter 4.

Chapter 5 discusses the introduction of an additional parameter in the case of forecasting mortality improvement rates. In fact, in the model presented in Chapter 3, the transformed mortality rates are modelled as realizations of independent normal random variables with constant variance. The hypothesis of constant variance can be seen as too much restrictive, therefore the variance is allowed to vary with respect to the age. The extended models are then applied to a set of Nordic Countries.

Chapter 6 contains some concluding remarks.

Chapter 2

Introduction to mortality models

2.1 Introduction

The aim of this chapter is to introduce the problem of forecasting mortality. In particular, the mortality phenomenon of human populations is studied using life tables: a definition is given in Section 2.2 while Section 2.3 discusses some basic functions which are useful for better understanding the mortality models. In Section 2.4 the concept of projected life tables is briefly revised and a possible classification of existing models is presented. One of the most influential approach which deals with mortality rates, the Lee-Carter model, is presented in Section 2.5, together with some extensions. In Section 2.6 the possibility of forecasting mortality improvements is presented, with details about the most relevant proposals. Section 2.7 deals with the notion of multiple population forecasts, with focus on extensions of the Lee-carter model. In Section 2.8 a brief discussion is provided.

2.2 Life tables

The life (or mortality) tables are an appropriate tool for conveniently study the mortality phenomenon. A life table, for a specific population, is defined as a decreasing sequence $l_0, l_1, \ldots, l_\omega$ where l_x represent the estimated number of people alive at age x. The age takes the values $x = 0, 1, \ldots, \omega$, where ω is the upper limit age, *i.e.* the age such that $l_\omega > 0$ and $l_{\omega+1} = 0$.

The decreasing sequence $l_0, l_1, \ldots, l_\omega$ represents the number of people alive at every age from an initial group of l_0 individuals aged 0 in a specific year.

2.2.1 Cohort life table

The most intuitive approach for obtaining a life table consists in two steps. At first, an initial group of l_0 individuals in a specific year t is considered: this constitutes the cohort of people born in the year t. Subsequently, this cohort is observed longitudinally across the years. The object of observation is, year by year, the actual number of individuals from the selected cohort alive at year $t + x$, with $x = 1, 2, \ldots, \omega$.

A life table $l_0, l_1, \ldots, l_\omega$ obtained with this procedure is called a cohort life table.

2.2.2 Period life table

There is another method for obtaining a mortality table, which does not require to observe a cohort of individuals across years until its extinction. The first step consists in estimating the q_x, that is the probability of an individual age x dying within one year, and is defined as

$$q_x = P[T_x < 1]$$

where T_x is a random variable representing the remaining life for a person age x. Once values of the age-specific probability of deaths q_x are estimated for $x = 0, 1, \ldots, \omega$, the life table is computed recursively using the formula

$$l_{x+1} = l_x(1 - q_x)$$

starting by an assigned value for l_0 (usually is fixed equal to $100,000$). A life table obtained in this way is called a period life table and represents the number of survivors out of a hypothetical cohort composed by l_0 individuals. The procedure for the construction of a period life tables requires the observation of the mortality phenomenon only for the chosen period, that can be one or a few years. This approach is based on the assumption that the mortality pattern does not change over time.

2.3 Death probabilities

2.3.1 Basic functions

The random variable T_x is considered again, but it is studied in a continuous context: ages and time are now assumed to take any number, not just integers. The survival function $S(t)$ is defined as

$$S(t) = P[T_0 > t]$$

for $t \geq 0$. A similar function of t is the distribution function of T_0, defined as $F_0(t) = P[T_0 \leq t]$ for $t \geq 0$. Obviously, $F_0(t) = 1 - S(t)$.

The force of mortality μ_x is defined by

$$\mu_x = \lim_{t \to 0} \frac{P[T_x \leq t]}{t}$$

and represents the instantaneous rate of mortality at a given age x. The force of mortality is strictly connected with the survival function, in fact μ_x can be written in terms of $S(x)$. Since

$$P[T_x < t] = \frac{P[x < T_0 \leq x + t]}{P[T_0 > x]} = P[T_0 \leq x + t | T_0 > x] = \frac{F_0(x+t) - F_0(x)}{S(x)}$$

therefore

$$\begin{aligned} \mu_x &= \lim_{t \to 0} \frac{F_0(x+t) - F_0(x)}{tS(x)} = \frac{1}{S(x)} \lim_{t \to 0} \frac{F_0(x+t) - F_0(x)}{t} \\ &= \frac{1}{S(x)} \frac{\partial}{\partial x} F_0(x) = -\frac{1}{S(x)} \frac{\partial}{\partial x} S(x). \end{aligned}$$

It follows that $S(x) = \exp\{-\int_0^x \mu_z \, \mathrm{d}z\}$.

The central death rate summarises the force of mortality over a given interval and is denoted by m_x for a given age x. The definition of m_x is

$$m_x = \frac{\int_0^1 S(x+u)\mu_{x+u} \, \mathrm{d}u}{\int_0^1 S(x+u) \, \mathrm{d}u} \tag{2.1}$$

that is the weighted mean of μ_z over the interval $(x, x+1)$, where the weighting function is the probability of being alive at age $x+u$ (*i.e.* the survival function). The formula (2.1) can be rewritten in a simplified way, since

$$\int_0^1 S(x+u)\mu_{x+u}\,\mathrm{d}u = S(x) - S(x+1).$$

Furthermore, as $\int_0^1 S(x+u)\,\mathrm{d}u$ is often approximated using the trapezoidal rule, (2.1) becomes

$$m_x \simeq \frac{S(x) - S(x+1)}{(S(x) + S(x+1))/2}$$

which should be used when only some data are available, for example when only a life table is available. In fact, the quantities defined above regards a time-continuous phenomenon, but are often in practice treated in a discrete context.

A model used for describing mortality age patterns in terms of parametric functions, *e.g.* μ_x or $S(x)$, is called a mortality law.

2.3.2 Approximations and estimating procedure

Approximation methods are widely used in actuarial practice in order to obtain the survival function for all real ages x starting by a life table. The assumption adopted here is the piece-wise constant force of mortality within each integer age band. This means that, for every age x and every value t such that $0 \leq t < 1$, we have $\mu_{x+t} = \mu_{(x)}$, where $\mu_{(x)}$ is the force of mortality corresponding to age x. Due to this assumption, from the definition of central death rate, it follows that

$$m_x = \mu_x.$$

So far, the calendar year has been kept fixed. In the study of mortality data through time, more than one calendar year is considered, hence a new notation should be used. From now on we indicate with the index t the calendar year and with $m_{x,t}$ the central death rate relative to age x in the calendar year t. The central death rates, and consequently the corresponding forces of mortality, are estimated with the crude death rate which is, for age x and year t, defined as

$$\hat{m}_{x,t} = \frac{D_{x,t}}{\mathrm{ETR}_{x,t}},$$

where $D_{x,t}$ is the number of deaths recorded at age x last birthday during the calendar year t and $\text{ETR}_{x,t}$ is the exposure to risk for age x and year t. The exposure to risk is a measure of individuals with a certain age x in the calendar year t. This number can be seen as the average of the individuals aged x over the selected calendar year. Clearly, this average is adjusted by the length of time the individuals actually remains in the population.

Once the central death rates has been estimated, the values of $q_{x,t}$ should be obtained for the construction of the life table. Central death rates and death probabilities are usually very close to one another in value and, under the assumption of piece-wise constant force of mortality, are approximately related with the formula

$$q_{x,t} \simeq 1 - \exp\{-m_{x,t}\}.$$

For more details on life tables and mortality estimation procedures see Pitacco *et al.* (2009).

2.4 Forecasting mortality

2.4.1 Projected life table

Due to the fact that human mortality has strongly declined over the last decades, period life tables cannot be used in actuarial practice for all the life insurance products. As mentioned before, such life tables are constructed under the assumption that mortality phenomenon is time constant. For this reasons, the use of period life table in actuarial practice is restricted to short or medium term insurance products, *i.e.* applied to a time interval of 5 or 10 years. However, many life insurance products, as life annuities and pension plans, require to consider longer time intervals. The life tables used in this case should be constructed in order to incorporate the experienced mortality trend, with the aim to anticipate its future evolution. Such a life table is called a projected life table and it is based on forecast mortality rates.

2.4.2 Classification of forecasting methods

The problem of forecasting mortality has been widely discussed, and it is still considered a hot topic among actuaries and demographers. There are several different mortality models in the literature, and there is more than one way to classify these methods (Tabeau *et al.*, 2001). The classification shortly summarized below is the one used in Booth (2006), where three main approaches to forecast demographic processes are indicated.

The first approach refers to extrapolative methods, which are based on the assumption that trends observed in the past are likely to remain constant into the future. Such an approach does not incorporate into the analysis any knowledge about the studied phenomenon. An example of extrapolative methods is the direct application of univariate ARIMA models (Box and Jenkins, 1976) to demographic time series in order to obtain forecasts.
The second one is the expectation approach, and it is based on demographic developments that are considered more likely to occur. The expectation about future trends are usually decided using the intuition or the informed judgement of experts in the considered field. Some demographic projections performed using the expectation approach can be found in Lutz (1996).
The third approach collects the theory-based structural modelling involving exogenous variables. These models are constructed in order to explain demographic quantities using the relationship between the index object of interest and other variables. Many of these models start by collecting informations about the phenomena, then they are estimated using regression. A review of structural models for population projections by a socio-economic perspective is in Sanderson (1998). In some cases, the classification of a particular model into the three approaches presented may be ambiguous.

Between the extrapolative methods applied to mortality data, one in particular is receiving great deal of attention due to its good predictive power and its simple structure: the Lee-Carter model (Lee and Carter, 1992).

2.5 Lee-Carter model and principal extensions

2.5.1 The original Lee-Carter formulation

The original formulation of the Lee-Carter model, presented in Lee and Carter (1992), is

$$\log m_{x,t} = a_x + b_x k_t + \varepsilon_{x,t}. \tag{2.2}$$

The logarithm of the central death rates $m_{x,t}$ is specified as a function of x, a_x, which is the general mortality shape across age, a bilinear term $b_x k_t$ and an error term $\varepsilon_{x,t} \sim N(0, \sigma^2)$ which reflects the age-specific variability that is not captured by the model. The bilinear term is composed by k_t, an index of the level of mortality across time, and b_x, the age specific response to variations in the time index. A model written in this way is overparametrised, therefore Lee and Carter introduced two additional constraints to determine a unique solution, *i.e.*

$$\sum_x b_x = 1 \text{ and } \sum_t k_t = 0. \tag{2.3}$$

This model does not describe mortality age patterns in terms of parametric functions, therefore cannot be considered a mortality law. In fact, the dependence on age is non parametric, as it is given by the sequences a_x and b_x. Another characteristic that should be noticed is that there are no observable variables between the independent variables. Due to this aspect, the Lee-Carter model cannot be estimated by simple regression. Lee and Carter proposed to estimate the parameters with two steps, at first the $\hat{a}_x$ are computed, then the bilinear term $\hat{b}_x \hat{k}_t$.

The values of $\hat{a}_x$ are the averages over time of the $\log m_{x,t}$: this follows by considering the sum over time of (2.2), then applying the second of the constraints in (2.3). Once the $\hat{a}_x$ are computed, they are subtracted from the corresponding logarithm of the death rates, and results are placed in a matrix, with ages on rows and years on columns, where the generic element is the centered log death rate

$$\log \hat{m}_{x,t} = \log m_{x,t} - \hat{a}_x. \tag{2.4}$$

The estimated values for the bilinear term $b_x k_t$ are the two first terms of the singular value decomposition of the centered log death rates matrix. A solution obtained in this way does not satisfy (2.3), thus the following operations are performed

1. obtain $\bar{k}$ as the average of $\hat{k}_t$;
2. compute $B = \sum_x \hat{b}_x$;
3. replace $\hat{a}_x$ with $\hat{a}_x + \hat{b}_x \bar{k}$;
4. substitute $\hat{k}_t$ with $(\hat{k}_t - \bar{k})B$;
5. replace $\hat{b}_x$ with $\hat{b}_x / B$.

Once the parameters are estimated, the time varying coefficients k_t are modelled as an ARIMA process. The time series of the k_t is the only part of the model that should be forecast in order to obtain the future mortality rates. Lee and Carter observed that in most cases a random walk with drift can be appropriate for modelling the time varying coefficients.

2.5.2 The Poisson assumption

In the specification of the model described in the previous section the random errors are homoscedastic, which is commonly a strong and often unrealistic hypothesis. In order to solve this problem, Brouhns *et al.* (2002) proposed a modification of the Lee-Carter model assuming that the number of deaths is a realisation of a Poisson random variable:

$$D_{x,t} \sim \text{Poisson}\left(\text{ETR}_{x,t}\, \mu_{x,t}\right),$$

where $\text{ETR}_{x,t}$ is the central number of exposure to risk and

$$\log \mu_{x,t} = \alpha_x + \beta_x k_t, \quad \sum_x b_x = 1, \ \sum_t k_t = 0$$

which has the form of the Lee-Carter model.

The procedure for estimating the parameters proposed in Lee and Carter (1992) is no longer valid. The approach proposed by Brouhns *et al.* (2002) consists in maximising the log-likelihood function

$$l = \sum_{x,t} \{D_{x,t}(\alpha_x + \beta_x k_t) - \text{ETR}_{x,t} \exp(\alpha_x + \beta_x k_t)\} + \text{constant}. \tag{2.5}$$

Optimisation is done with an iterative procedure, such as the Newton-Raphson method. The starting points for the algorithm are 0 for $\hat{\alpha}_x$ and $\hat{k}_t$, whereas $\hat{\beta}_x$ are set equal to 1. If j denotes the generic iteration, the parameter θ at the $(j+1)$-th iteration is obtained using the updating scheme

$$\hat{\theta}^{(j+1)} = \hat{\theta}^{(j)} - \frac{\partial l^{(j)}/\partial\theta}{\partial^2 l^{(j)}/\partial^2\theta^2}$$

where $l^{(j)}$ is (2.5) in which the values of the parameters are the ones obtained in iteration j. This updating scheme applied to (2.5) is formalized in the following three formulas

1. $\hat{\alpha}_x^{(j+1)} = \hat{\alpha}_x^{(j)} - \dfrac{\sum_t (D_{x,t} - \mathrm{ETR}_{x,t}\exp(\hat{\alpha}_x^{(j)} + \hat{\beta}_x^{(j)}\hat{k}_t^{(j)}))}{\sum_t -\mathrm{ETR}_{x,t}\exp(\hat{\alpha}_x^{(j)} + \hat{\beta}_x^{(j)}\hat{k}_t^{(j)})}$
2. $\hat{k}_t^{(j+1)} = \hat{k}_t^{(j)} - \dfrac{\sum_x (D_{x,t} - \mathrm{ETR}_{x,t}\exp(\hat{\alpha}_x^{(j+1)} + \hat{\beta}_x^{(j)}\hat{k}_t^{(j)}))\hat{\beta}_x^{(j)}}{\sum_x -\mathrm{ETR}_{x,t}\exp(\hat{\alpha}_x^{(j+1)} + \hat{\beta}_x^{(j)}\hat{k}_t^{(j)})(\hat{\beta}_x^{(j)})^2}$
3. $\hat{\beta}_x^{(j+1)} = \hat{\beta}_x^{(j)} - \dfrac{\sum_t (D_{x,t} - \mathrm{ETR}_{x,t}\exp(\hat{\alpha}_x^{(j+1)} + \hat{\beta}_x^{(j)}\hat{k}_t^{(j+1)}))\hat{k}_x^{(j+1)}}{\sum_t -\mathrm{ETR}_{x,t}\exp(\hat{\alpha}_x^{(j+1)} + \hat{\beta}_x^{(j)}\hat{k}_t^{(j+1)})(\hat{k}_x^{(j+1)})^2}.$

The updating scheme is repeated until the variation of the log-likelihood function into successive iterations is below a given threshold. At the end of the estimating procedure the identifiability constraints are applied using the steps described in the previous section.
The time index k_t is then forecast as in the Lee-Carter model.

2.5.3 Other relevant extensions

Due to its diffusion, the Lee-Carter model has been a hot research topic over the last twenty years. What follows are some of the most relevant proposed extensions.

Evaluating more complex model structures

When mortality data are described using parsimonious models as the Lee-Carter one, a significant part of the variance remains not explained. In order to improve the approximation of the model to the phenomenon, some generalizations of the Lee-Carter model were proposed.

In the original version of Lee-Carter model, parameters are obtained as the first terms of singular value decomposition (SVD). The idea of considering the second and possibly higher terms of SVD was introduced in Booth *et al.* (2002a). In that work a Lee-Carter model with n bilinear components, *i.e.*

$$\log m_{x,t} = a_x + \sum_{i=1}^{n} b_x^{(i)} k_t^{(i)},$$

is taken as a starting point for further analysis. The first bilinear term is obtained by applying SVD to the data, then the authors suggest an estimating procedure which incorporates the bilinear components one by one. At every step, the parameters are adjusted, then a new SVD is performed.

In Renshaw and Haberman (2003c) the quality of modelling mortality data with particular attention to the residuals using Lee-Carter models is discussed. This analysis not only shows that the first SVD component fails in capturing some important aspects of the data, but also highlights the presence of noteworthy residual patterns in the second SVD vectors. A further discussion on Lee-Carter models with more than one bilinear component can be found in Renshaw and Haberman (2003b), where the first two sets of SVD vectors are used. A different perspective is adopted in Renshaw and Haberman (2003a), where the Lee-Carter model is estimated using two or more terms of the SVD, then the time-varying coefficients are forecast using multivariate time series.

The idea of generalising Lee-Carter model incorporating other terms is stressed in Hyndman and Ullah (2007), where some examples of models based on this intuition are listed. These variants are designed in order to threat differences between groups, or to incorporate more complex cases where several parameters are included, as

$$y_{t,j}(x) = \mu_j(x) + \sum_{k=1}^{K} \beta_{t,k}\phi_k(x) + \sum_{l=1}^{L} \gamma_{t,j,l}\psi_{l,j}(x) + e_{t,j}(x)$$

where $y_{t,j}(x)$ is used to define the central death rates. By choosing the parameters appropriately, the previous general formula includes a wide number of different model designs.

Smoothing the parameters

One strengths of the Lee-Carter model is its simplicity. As a matter of fact, in the original formulation, once the parameters has been estimated, the forecast values are obtained by modelling just one time series. However, this aspect can be a problem in some cases. In fact, parameters irregularities are magnified when future values are computed. Whereas it is reasonable that the time-varying coefficients k_t present an irregular path, it is intuitive to think that the a_x and b_x should have a smooth shape. The idea of smoothing the age-dependent parameters is used in Renshaw and Haberman (2003d). In that work the mortality is expressed in terms of reduction factors RF, defined by the equation

$$\mu_{x,t} = \mu_{x,0} RF_{x,t}$$

for all $t \geq 0$ such that $RF_{x,0} = 1$ for every x, and $0 < RF_{x,t} \leq 1$ for every x when $t > 0$. Once estimated the reduction factors, the authors applied the Lee-Carter model to forecast these values, but they modified appropriately the original approach and incorporate a smoothing of a_x and b_x. This approach is used also in Renshaw and Haberman (2003c), where different approaches are compared.

In De Jong and Tickle (2006) the notion of smoothing is applied again in this context and a smooth version of the Lee-Carter model was proposed. The smooth version is here obtained using some terms of SVD applied to a generalised version of the approach proposed in Lee and Carter (1992). The aim in this case is to ensure smoothness in the age direction.

Another application of smoothing idea in this context can be found in Delwarde *et al.* (2007). In that work the attention is on the coefficients b_x, and the proposed approach assures values that are smooth. This is applied to the classical Lee-Carter model and the Poisson log-bilinear model (Brouhns *et al.*, 2002). The choice of the optimal value for the smoothing parameter is performed using cross validation.

The cohort effect

The cohort effect modelled with an additional parameter of Lee-Carter model was introduced in Renshaw and Haberman (2006). In that work, the authors consider

$$\log m_{x,t} = a_x + b_x^{(0)} \iota_{t-x} + b_x^{(1)} k_t$$

which is the Lee-Carter classical model formulation with a new term: $b_x^{(0)}\iota_{t-x}$. The parameter ι_{t-x} depends on the specific age of birth $t-x$. In Renshaw and Haberman (2006) this approach is applied using different choices of the error distribution. The results evidence how the introduction of cohort effect reduces significantly the presence of systematic behaviours in the residuals. This reduction is more evident if the residuals are analysed with respect to years of birth.

2.6 Mortality improvement rates

Modelling the central death rates or the force of mortality is not the only way to forecast future mortality. An alternative approach proposed in the literature is to model the improvements in mortality rates, rather than the rates themselves.

One example of this approach was used in Group Annuity Valuation Table Task Force (1995), where a procedure used for obtaining a new mortality table is described. In that work, the use of observed trends in mortality improvement rates for obtaining the future mortality values is considered appropriate. In fact, given $q_{x,t}$, the value of $q_{x,t+n}$ is obtained with the relation

$$q_{x,t+n} = q_{x,t}(1 - AA_x)^n$$

where AA_x is the annual improvement factor in the mortality rate for age x.

2.6.1 The volatility of the improvements

With a different perspective, the mortality improvement rates were used in Willets (2004), where the cohort effect in U.K. is analysed using three-dimensional block graphs of the improvement rates, as in Willets (1999). In the former work the representation regards the average mortality improvement rates by 5 years age groups and by 5 calendar years groups: each represented value is the mean of 25 mortality improvement rates. Furthermore, Willets (2004) gives some other remarks using the values of the improvements, but always taking into account the average of grouped data.
Some similarities can be seen in Baxter (2007), where the existence of a minimum level of improvement in mortality rates for future years is discussed. The rate of

improvements in mortality

$$1 - \frac{q_{x,t}}{q_{x,t-1}}$$

is studied in its historic trends. Furthermore, it is argued that an appropriate average of the past improvements can be used for actuarial projections.

Averaging these modified version of mortality values can help in the analysis: improvements in mortality are usually much more volatile rather than the rates themselves. In fact, if we apply the ratio between one mortality rate and that of the previous year, considering real data, we would obtain quite unstable results. Another work where this problem is also discussed is Richards *et al.* (2005) where the mortality improvement rates are defined as

$$\Delta m_{x,t} = 1 - \frac{m_{x,t}}{m_{x,t-1}}.$$

Subsequently, the authors removed the effects of random variations by smoothing with a moving average.

2.6.2 Recent developments

A different approach was proposed in Haberman and Renshaw (2012), where a more stable formulation for mortality improvement rates was introduced. The idea is to create values that do not need to be smoothed in a second stage. The transformation applied to the central death rates is

$$z_{x,t} = 2\frac{1 - m_{x,t}/m_{x,t-1}}{1 + m_{x,t}/m_{x,t-1}}.$$

The authors refer to this version of mortality improvement rates as scaled or relative, due to its structure that reduces the magnitude of the extreme values (further remarks on this formula are in Section 3.2.2). The aim of Haberman and Renshaw (2012) is to model and forecast the values of $z_{x,t}$ from an extrapolative perspective. That work is extended in order to include the cohort effect in Haberman and Renshaw (2013).

2.7 Multiple populations mortality models

The approaches for modelling and forecasting mortality catch the general mortality trend of a population. However, there are cases when the observed population is divided into subpopulations, or it is the combination of more than one population. In spite of models such as Lee-Carter, which are single population oriented, in the above mentioned cases two or more populations must be dealt with. One simple solution consists in merging the considered populations (or subpopulations) and to study them as an aggregate. This could be useful in terms of stability, since larger population usually implies less variability in the mortality process. However, if someone chooses to follow that way, there is a relevant waste of informations about the mortality process. The alternative is to separately study the mortality of all considered populations. In this way it is possible to observe and study differences in the mortality patterns.

2.7.1 Common and specific factors

Convergence and coherence

In Wilson (2001) an important concept was documented: a global convergence in mortality levels. This should be taken in consideration for multiple populations analysis. Clearly, in a medium-short time horizon it is not possible to observe a global convergence. Conversely, a similar mortality path can be expected between populations which shares similar characteristics. An example where this is valid could be given in neighbouring countries with similar economies and welfare systems. These similarities are obviously stronger if we consider regions (or other subpopulations) within a country. Starting with this idea, Li and Lee (2005) shows the importance of avoiding that the difference of general mortality levels in the forecasts for related population increase across the years. The authors propose an extension of Lee-Carter model that considers more than one population and guarantee non-divergent forecasts of life expectancy in the long run, and they called such an analysis a coherent forecast.

In Li and Lee (2005) the model assumes the form of

$$\log m_{x,t,i} = a_{x,i} + B_x K_t + b_{x,i} k_{t,i} + \varepsilon_{x,t,i}, \tag{2.6}$$

where the index i refers to the i-th population. We can observe, in (2.6), the presence of two bilinear terms: a common factor $B_x K_t$, that assures a long-term convergence in mortality, and a specific factor $b_{x,i}k_{t,i}$, which allows for short or medium term differences. The authors show that the convergence is guaranteed if the time varying coefficients of the specific terms tend toward a constant value, hence $k_{t,i}$ should has a null long-term mean. This is to assure the coherence of the forecasts.

Relative approach

As in Li and Lee (2005), the mortality rates are studied following a relative perspective in Villegas and Haberman (2014). In particular, the mortality modelling of subpopulations within a larger population is investigated. The authors specify the larger population mortality with an age-period-cohort Lee-Carter model

$$\log m'_{x,t} = \alpha'_x + \beta'_x k'_t + \gamma'_{t-x},$$

then mortality of the subpopulation i as

$$\log m_{x,t,i} = \log m'_{x,t} + \alpha_{x,i} + \beta_x k_{t,i}.$$

Considering the entire population and a part of it as a subpopulation assures to have consistency between them. The mortality data of the larger population are usually more accurate and allow to introduce a cohort effect, that is reasonably in common for socio-economic subpopulations within a country. Furthermore, the data of the larger population are usually available for a longer years interval. In their work, the authors suggest a way to consider these additional data into the analysis, in order to obtain a more precise estimation of the long-run mortality trend.

2.7.2 Other extensions of Lee-Carter model

One of the first extensions of the Lee-Carter model regarding more than one population was introduced in Carter and Lee (1992), where the U.S. population is divided into two groups: male and females. In that work three alternatives are presented in order to study the two subpopulations together. The first is the

straightforward separate application of the Lee-Carter model, with the possibility of searching for dependence between the two time varying coefficients after the estimation procedure. The second approach proposed is to estimate a single vector k_t which could be appropriate for both the populations. The third approach consists in modelling the data as a co-integrated process. However, in the paper Lee and Carter discuss with more emphasis the first approach, and use the other two approaches for results comparison or for making further considerations.

Subsequently, other extensions of the Lee-Carter model for treating multiple populations were proposed. The philosophy that connect some of these approaches is the idea of adding terms to the original formulation of Lee-Carter model. The aim is to introduce terms that allow to reproduce differences and similarities among several populations instead of adding terms in order to explain more accurately the data of a single population, as seen in most of the approaches presented in Section 2.5.3. Many of these proposals can be seen as particular cases of the general form

$$\log m_{x,t,i} = \sum_{l=1}^{L} \beta_{x,i}^{(l)} k_t^{(l)} \gamma_i^{(l)} \tag{2.7}$$

where i is the considered population. Formula (2.7) is similar to the general formulation used in Cairns *et al.* (2009).

Common factor model

Considering a common factor is a first extension of the Lee-Carter model which permits to deal with more than one population. It can be obtained by choosing in (2.7) a formulation with two terms, *i.e.* $L = 2$, with $k_t^{(1)} = \gamma_i^{(1)} = \gamma_i^{(2)} = 1$ and $\beta_{x,i}^{(2)} = \beta_x$. Accordingly, the formula

$$\log m_{x,t,i} = a_{x,i} + b_x k_t + \varepsilon_{x,t,i} \tag{2.8}$$

where $a_{x,i}$ is the general level of mortality of the i-th population is obtained. Conversely, the bilinear component is common to all the groups. It follows that this approach considers a common evolution of the mortality, which is added to a population-specific mortality level. This approach was used as a starting point in Li and Lee (2005). In fact, in their paper, the authors start by using the formula (2.8), then they add the specific terms and obtained the formulation described in

(2.6). This model is also discussed in Li and Hardy (2011), where it is compared with other three approaches. In this last work, formula (2.8) is extended in a second step as in (2.6) in order to add population-specific terms. The extended version of this formula is also known as augmented common factor model (Li and Hardy, 2011).

Common time-varying coefficient

Another perspective for dealing with the problem of coherent forecasting can be obtained with a specification which is slightly different with respect to (2.8). It consists in

$$\log m_{x,t,i} = a_{x,i} + b_{x,i}k_t + \varepsilon_{x,t,i}, \tag{2.9}$$

where $b_{x,i}$ substitutes b_x. This means that, with this specification, only the time varying coefficient k_t is common to all the populations. This model was introduced as one of the possible implementations of Lee-Carter model in Carter and Lee (1992). Furthermore, specification (2.9) is one of the four considered variants of Lee-Carter model compared in Li and Hardy (2011).

In Wilmoth and Valkonen (2001) the authors considered Finnish mortality data partitioned by social group, and use a generalised Lee-Carter model to deal with it. The generalisation consists in a Lee-Carter model with C covariates, each with one or more categories. For instance, considering two covariates, denoted by c_1 and c_2, the approach described in Wilmoth and Valkonen (2001) can be written as

$$\log m_{x,t,\tau_1,\tau_2} = \lambda^{(0)} + \lambda^{(1)}_{\tau_1} + \lambda^{(2)}_{\tau_2} + \alpha^{(0)}_x + \alpha^{(1)}_{x,\tau_1} + \alpha^{(2)}_{x,\tau_2} + (\beta^{(0)}_x + \beta^{(1)}_{x,\tau_1} + \beta^{(2)}_{x,\tau_2})k_t$$

where

- τ_1 and τ_2 are respectively the categories of the covariates c_1 and c_2;
- $\lambda^{(0)}$ is the overall level of mortality, with an adjustment of $\lambda^{(c)}_\tau$, for the τ-th category of the c-th factor;
- $\alpha^{(0)}$ is the typical age pattern of mortality, with an adjustment of $\alpha^{(c)}_\tau$, for the τ-th category of the c-th factor;

- $\beta^{(0)}$ is the age pattern of mortality decline, with an adjustment of $\beta_\tau^{(c)}$, for the τ-th category of the c-th factor;
- k_t represents the time pattern of mortality.

Note that the time-varying component k_t is assumed to be constant across social categories.

Stratified Lee-Carter model

The classical Lee-Carter model is extended in Butt and Haberman (2009) in order to include an additional covariate. It can be obtained by (2.7) in the same way as in the common factor Lee-Carter models, *i.e.* choosing a formulation with $L = 2$ and $k_t^{(1)} = \gamma_i^{(1)} = \gamma_i^{(2)} = 1$ and $\beta_{x,i}^{(2)} = \beta_x$. The proposed variation aims to quantify the differences in mortality of populations subgroups. Unlike the common factor Lee-Carter models, the general shape term a_x term of (2.2) is seen here as

$$a_{x,i} = a_x + a_i.$$

With this modification applied to (2.8), the authors define the stratified Lee-Carter model as

$$\log m_{x,t,i} = a_x + a_i + b_x k_t + \varepsilon_{x,t,i}. \tag{2.10}$$

The additional parameter a_i is the relative difference between the age-specific mortality profiles of the subpopulation i with respect to the generic a_x. The stratified Lee-Carter model has similarities with the approach seen in Li and Lee (2005), due to its structure with a common factor plus specific terms.

A similar approach can be seen in Currie (2009a) and Currie (2010). The specification of this model is similar to (2.10), with the addition of another term. The Author proposed

$$\log m_{x,t,i} = a_x + a_i^0 + a_i^1 x + b_x k_t + \varepsilon_{x,t,i}, \tag{2.11}$$

where $a_i^0 + a_i^1 x$ is a linear adjustment term which depends on the considered population. This population-specific linear adjustment determines the population mortality levels. The idea of levels in mortality is the same as in Butt and Haberman (2009), but in (2.11) the levels are defined proportionally to the age,

rather than with an additive constant. Equation (2.11) is a variant of the Piggy-back model, presented in Currie (2009b) and Currie (2009c), that was designed to make estimation and forecast which are based on an existing Lee-Carter model.

Three-way Lee-Carter model

The so called three-way Lee-Carter model was introduced in Russolillo *et al.* (2011). Its formulation is

$$\log m_{x,t,i} = a_{x,i} + b_x k_t \gamma_i, \tag{2.12}$$

which is (2.7) with $L = 2$, $k_t^{(1)} = \gamma_i^{(1)} = 1$ and $\beta_{x,i}^{(2)} = \beta_x^{(2)}$ for all i. The interpretation of the parameters in (2.12) is the classical one for $a_{x,i}$, b_x and k_t. The term γ_i is the factor associated to the i-th population. This factor influence the mortality levels in a multiplicative way. In Russolillo *et al.* (2011) the authors consider the mean-centered death rates, as defined in (2.4), then estimate the parameters using singular value decomposition, as in Lee and Carter (1992). Conversely, a generalised SVD should be used here, due to the three dimensions of the considered data, and the SVD first components gives (2.12).

2.7.3 Focus on the time varying coefficient

In this Section we want to emphasise the possibility of forecasting the populations separately, each with its own time index. In a second stage the correlation between the populations is introduced from modelling jointly the time varying coefficients. Lee-Carter model is not the only starting point in this approach. In particular there are two important proposals, the first is the so called gravity model and the second is based on a Bayesian perspective. These two approaches, as well as some applications to the Lee-Carter model, are presented in the sequel.

Lee-Carter models with co-integrated time indices

As mentioned before, Carter and Lee (1992) introduced some possible extensions of their model in order to consider more than one population. One of these methods considers the time varying coefficients as a co-integrated process. This follows from

the intuition that there are common components across the considered k_t series. An example of this is when time series share the same trend. The authors refer to the relationship between the series not as direct and explicit, but it is assumed that all the series respond similarly to unknown and exogenous forces. In order to obtain forecast values that move together in the long-term, Carter and Lee (1992) suggested to use co-integrated time series, rather than multivariate structural time series. This means that, when looking at two populations, the mortality evolution is determined by just one unobserved driving force, instead of two.

The Lee-Carter model with co-integrated time indices is also studied in Li and Hardy (2011), among other models. The authors explain the importance of testing for co-integration before proceeding with this method. In Li and Hardy (2011) the co-integration is proved for the considered dataset using both a statistic test and a graphical analysis.

The co-integration analysis of the time indices applied to the Lee-Carter model is also performed in Yang and Wang (2013), where further a vector error correction model (VECM) is applied to mortality forecast. Another work where the VECM approach is studied is Zhou *et al.* (2013).

The gravity model

The so called gravity model was introduced in Dowd *et al.* (2011). The main idea is to create a model inspired to the gravitational force. The authors distinguish between two cases: two populations with similar size (such as males and females within a country), and one population much larger than the other. In Dowd *et al.* (2011) the focus is on this last case, following the intuition that the larger population exerts a pull on the smaller one. Two mortality datasets, coming from a larger population, denoted by (1), and a smaller population, denoted by (2), are considered. The data are modelled with the following age-period-cohort model

$$\log m_{x,t}^{(1)} = \beta_x^{(1)} + n_a^{-1} k_t^{(1)} + n_a^{-1} \gamma_{t-x}^{(1)} \tag{2.13}$$

$$\log m_{x,t}^{(2)} = \beta_x^{(2)} + n_a^{-1} k_t^{(2)} + n_a^{-1} \gamma_{t-x}^{(2)} \tag{2.14}$$

where, for $i = 1, 2$, can be seen

- $k_t^{(i)}$, the time dependent parameter;

- $\beta_x^{(i)}$, the age dependent parameter;
- $\gamma_{t-x}^{(i)}$, the cohort effect;
- n_a^{-1}, the number of ages in the sample data used to estimate the parameters.

The gravity effect depends on the way of modelling $k_t^{(i)}$ and $\gamma_{t-x}^{(i)}$. The time varying coefficients of the two populations are modelled by

$$\begin{aligned} k_t^{(1)} &= k_{t-1}^{(1)} + \mu^{(1)} + C^{(11)} Z_t^{(1)} \\ k_t^{(2)} &= k_{t-1}^{(2)} + \phi(k_{t-1}^{(1)} - k_{t-1}^{(2)}) + \mu^{(2)} + C^{(21)} Z_t^{(1)} + C^{(22)} Z_t^{(2)} \end{aligned}$$

where, for $i = 1, 2$,

- $\mu^{(i)}$ is a constant drift term;
- $\phi^{(k)} \geq 0$ is the gravity parameter;
- $C = \begin{bmatrix} C^{(1,1)} & C^{(1,2)} \\ C^{(2,1)} & C^{(2,2)} \end{bmatrix}$ is the two by two dimensions correlation matrix;
- $Z_t^{(i)}$ is an independent error distributed as a standard normal random variable.

Notice that the time varying coefficient of the larger population $k_t^{(1)}$ is described as a random walk with drift. Its path influences the behaviour of the time index of the smaller population $k_t^{(2)}$, proportionally to ϕ. The cohort time index $\gamma_{t-x}^{(i)}$ is specified similarly.

Introducing a Bayesian perspective

A version of the age-period-cohort model in (2.13) and (2.14) is used as a starting point in Cairns *et al.* (2011). As in the gravity model, in that work two cases are discussed: one where one population is dominant, and the other where two equal populations are considered. The model is then estimated in a Bayesian framework, in order to allow for different trends in the short run, but parallel improvements in the long run. The authors show how it is possible to obtain the parameters of the mortality models jointly for the two populations. These parameters should determine mortality forecast values in short and long term, according to opportune

characteristics in short and long term. The estimation procedure is based on the Bayesian Markov chain Monte Carlo approach. The approach presented in Cairns *et al.* (2011) can help in dealing with small populations, where usually there is a strong volatility, or in presence of missing values.

2.7.4 Relative models

Another perspective adopted for forecasting multiple populations mortality is given by relative models. A model is called relative if the mortality of the target population is specified as difference with respect to the reference one. Examples of this in the context of Lee-Carter type models, are the Li and Lee (2005), Li and Hardy (2011) and the Piggyback model. However, there are examples of relative models which are not defined as extensions of the Lee-Carter structure. Some of the most influential models are presented in the following subsections.

The SAINT model

Jarner and Kryger (2011) proposed an approach for the robust forecasting of small population mortality. The main idea consist in estimating the long term trend of a large population at first. In a second stage the small population mortality level is specified in terms of deviation from the long term trend obtained for the large population. The mortality levels estimated in large populations are usually less volatile and more regular than those estimated in smaller populations. With this perspective, death counts of the large population, denoted by l, are modelled as independent Poisson random variables, thus

$$D^l_{x,t} \sim \text{Poisson}(\hat{\mu}^l_{x,t} E^l_{x,t})$$

where $\hat{\mu}^l_{x,t}$ is the force of mortality. It should be noticed that, if $\hat{\mu}^l_{x,t}$ is defined through a Lee-Carter model, this specification is the same as that in Brouhns *et al.* (2002). However, other parametric structures for $\hat{\mu}^l_{x,t}$ can be considered.
Subsequently, death counts of the small population, denoted by s, are modelled as

$$D^s_{x,t} \sim \text{Poisson}(\hat{\mu}^s_{x,t} E^s_{x,t})$$

with

$$\hat{\mu}^s_{x,t} = \hat{\mu}^l_{x,t} \exp(y'_t r_x).$$

The bilinear term $y'_t r_x$ is a measure of the difference in mortality between the large population and the small one. This spread is defined in a way which allows different parametric specifications, such as the Lee-Carter one. This method, based on the spread between the mortality data, is called SAINT (Spread Adjusted InterNational Trend).

A similar approach is used in Jarner and Møller (2013), where longevity risk is considered. In that work the authors first estimate a mortality benchmark then, given the benchmark, each company estimates its own specific mortality.

Plat relative model

Similarly to the SAINT model, in Plat (2009b) a relative approach is proposed. The aim is to quantify the mortality level for a specific insurance portfolio. It is clear that the size of an insurance portfolio is much smaller than the entire population. The author suggests to use a stochastic model which target the insurance portfolio, and to combine it with the mortality process of the entire population. This last process drives the mortality rates evolution of the specific portfolio.
The object of analysis is the quantity $P_{x,t}$, defined by

$$P_{x,t} = \frac{q^s_{x,t}}{q^l_{x,t}}$$

which is the ratio between the portfolio specific mortality rate $q^s_{x,t}$ and the country population mortality rate $q^l_{x,t}$, for every age x and year t. As the difference between the two populations is expected to reduce at higher ages, the value of $P_{x,t}$ should approach 1 for x close to the upper limit ω.
The proposed model for $P_{x,t}$ is

$$P_{x,t} = 1 + \sum_{i=1}^{n} b^{(i)}_x k^{(i)}_t + \varepsilon_{x,t},$$

where n is the number of bilinear components $b_x^i k_t^i$ considered in the model and $\varepsilon_{x,t}$ is the error term. In order to ensure that $P_{x,t}$ approaches 1, the constraint

$$\sum_{i=1}^{n} b_x^{(i)} k_\omega^{(i)} = 0$$

is added. The authors explain how the model can be set in different ways, depending on the characteristics of the data.

The sum of two models

Another approach for studying the mortality of a small population with respect to the changes in a larger population can be found in Wan *et al.* (2013). The proposed specification consists in the combination of two different models. In fact, the authors consider

$$\log m_{x,t,i} = \beta_x + k_t^1 + k_t^2(x - \hat{x}) + \gamma_{t-x} + a_{x,i} + \sum_{j=1}^{m} b_{x,i}^{(j)} k_{t,i}^{(j)}. \tag{2.15}$$

The first part of this formula, *i.e.*

$$\beta_x + k_t^1 + k_t^2(x - \hat{x}) + \gamma_{t-x},$$

is the Plat age-period-cohort model (Plat, 2009a) and it is used to model the mortality of the larger population. It is composed by

- $\hat{x}$, the mean age in the sample range;
- β_x, the general mortality shape by age;
- k_t^1, the changes in the level of mortality for all ages with respect to the year;
- k_t^2, which allows the changes in mortality to vary by ages;
- γ_{t-x}, the cohort effect.

The second part of (2.15), that is

$$a_{x,i} + \sum_{j=1}^{m} b_{x,i}^{(j)} k_{t,i}^{(j)},$$

is a Lee-Carter model with a generalised number of bilinear terms (Booth *et al.*, 2002a), and it is used for modelling the spread between the large and the small populations. Clearly, formula (2.15) requires additional constraints, some for the Plat model part and some for the Lee-Carter model part, summarised in Wan *et al.* (2013).

The product ratio method

In Hyndman *et al.* (2013), the authors study the mortality of two populations, but not directly in terms of mortality rates. Instead, two other quantities are defined. Considering two populations, *e.g.* males and females, denoted respectively by M and F, the defined quantities are the products

$$p_{x,t} = \sqrt{m^M_{x,t} m^F_{x,t}} \tag{2.16}$$

and the ratios

$$r_{x,t} = \sqrt{m^M_{x,t} / m^F_{x,t}}, \tag{2.17}$$

where $m^i_{x,t}$ are the smoothed mortality rates of the i-th population, $i = M, F$. The quantities $p_{x,t}$ and $r_{x,t}$ are then modelled by a Lee-Carter type model, as proposed in Hyndman and Ullah (2007), that is

$$\log p_{x,t} = \mu^p_x + \sum_{k=1}^{K} \beta^{(k)}_t \phi^{(k)}_x + \varepsilon^p_{x,t} \tag{2.18}$$

$$\log r_{x,t} = \mu^r_x + \sum_{l=1}^{L} \beta^{(l)}_t \phi^{(l)}_x + \varepsilon^r_{x,t}, \tag{2.19}$$

where K and L are the number of considered bilinear terms. Formulas (2.16) and (2.17) can be generalised in order to deal with more than two populations. The corresponding versions of (2.16) and (2.17) are then

$$p_{x,t} = \left[f^{(1)}_{x,t} f^{(2)}_{x,t} \dots f^{(I)}_{x,t} \right]^{1/I}$$

for the products and

$$r^{(i)}_{x,t} = f^{(i)}_{x,t} / p_{x,t}$$

for the ratios, where the index $i = 1, \ldots, I$ refers to the considered population. The parametric structures (2.18) and (2.19) are opportunely modified in order to deal with I populations, as shown in Hyndman *et al.* (2013).

Generalised linear model approach

Hatzopoulos and Haberman (2013) used a method for multiple populations mortality forecasting which is based on generalised linear modelling. The authors select a pool of countries characterised by a common pattern of mortality dynamics. The selection is performed using cluster analysis.
Once the countries with similar mortality dynamics has been selected, death counts for all the populations are specified using a generalised linear model. A system of weights is introduced in order to avoid that the larger populations dominate the overall mortality trend. In this way the authors give equal weight to the mortality dynamic of each country.

Generalised linear models in a multi-population framework are also used in Biatat and Currie (2010). In that paper, mortality tables are smoothed, then classified and compared in terms of their distance from a reference table.

2.7.5 Forecasting mortality using mixed mortality data

The problem of small population mortality forecasting is analysed in Ahcan *et al.* (2014) with a perspective that differs from the approaches seen so far. In fact, in that work a method is developed to deal with data which present problems, such as missing data or a very high volatility. The idea is here to replicate the mortality of the small population using mortality data of similar countries. In Ahcan *et al.* (2014) neighbouring countries are used as reference populations. New data are obtained by mixing the observed mortality of the small population with that of the reference populations.

2.8 Discussion

This chapter outlines how the mortality phenomenon can be summarised in order to being conveniently used in practice. However, the mortality is a dynamic phenomenon, thus the described synthetic quantities cannot be applied on wide time intervals.

The consequence of this problem is the need of projected mortality tables, in order to anticipate the future trends. Several models were proposed in the literature in order to obtain accurate forecasts. In particular, in this thesis the focus in on the Lee-Carter model and its extensions, that produce good quality forecasts with reasonable and simple model structures.

A specific aspect of forecasting mortality is here investigated: forecasting mortality of more than one population. Some of the most influential models are reviewed in this chapter. Many of these approaches use Lee-Carter model as a starting point or in some step of the method. In these works, new parameters are added to the Lee-Carter model in order to catch some further characteristics of the data. Nevertheless, it is desirable that a model for coherent forecast is as parsimonious as possible. In fact, theoretic Lee-Carter generalisations in the number of parameters, as proposed in Hyndman and Ullah (2007), have been used mostly in the more concise forms possible.

Chapter 3

Generalized Lee-Carter type models for multiple populations

3.1 Introduction

This chapter outlines the models here considered. That are versions or extensions of the one proposed by Lee and Carter. The focus is on possible approaches that allow to forecast multiple populations mortality data at the same time, by considering two parametric structures.

Section 3.2 gives a presentation of the two selected parametric structures, and these specifications are generalized in Section 3.3. Once defined the generalised parametric structures, the models are defined in Section 3.4. In Section 3.5 each model is briefly discussed and, in the same section, the link between the presented models and similar approaches is evaluated. Section 3.6 describes the forecast procedure used.

3.2 The two selected parametric structures

3.2.1 Central mortality rates

The Lee-Carter model is here considered in the version proposed in Brouhns *et al.* (2002), already introduced in Subsection 2.5.2. It is worth noting that in its

original formulation (Brouhns *et al.*, 2002), the target of the Lee-Carter model with Poisson error structure is the force of mortality $\mu_{x,t}$, while here the mortality rates $m_{x,t}$ are used (since $\mu_{x,t} = m_{x,t}$ due to the assumption in Subsection 2.3.2). With this change of notation, the number of deaths $D_{x,t}$ is described by

$$D_{x,t} \sim \text{Poisson}\left(\text{ETR}_{x,t}\, m_{x,t}\right)$$

where $\text{ETR}_{x,t}$ is the central number of exposed to risk and

$$\log m_{x,t} = \alpha_x + \beta_x k_t, \quad \sum_x b_x = 1, \ \sum_t k_t = 0. \tag{3.1}$$

3.2.2 Mortality improvement rates

The alternative parametric structure considered derives from modelling the improvements in mortality rates, rather than the rates themselves. Among the approaches presented in Section 2.6, the chosen formulation is the one introduced in Haberman and Renshaw (2012), which defines improvement rates as

$$z_{x,t} = 2\frac{1 - m_{x,t}/m_{x,t-1}}{1 + m_{x,t}/m_{x,t-1}}. \tag{3.2}$$

Formula (3.2) can be seen as the ratio between the incremental mortality improvements $(m_{x,t-1} - m_{x,t})$ and the average of the two adjacent mortality rates $(m_{x,t} + m_{x,t-1})/2$. The values of $z_{x,t}$ are modelled as realizations of independent Gaussian random variables $Z_{x,t}$ assuming constant dispersion and mean $\eta_{x,t}$, hence

$$Z_{x,t} \sim N(\eta_{x,t}, \sigma^2).$$

The new first moment predictor structure

For the expected values $\eta_{x,t}$, the predictor structure is

$$\eta_{x,t} = \beta_x k_t, \quad \sum_x \beta_x = 1. \tag{3.3}$$

Clearly, the coefficients k_t are not the same as in (3.1). (3.3) can be obtained as a derivation of (3.1) (Haberman and Renshaw, 2012). In fact, deriving (3.1)

$$\frac{1}{m_{x,t}}\frac{\partial m_{x,t}}{\partial t} = \beta_x \frac{\partial k_t}{\partial t}$$

and, with the redefinition $\frac{\partial k_t}{\partial t} \to k_t$

$$\frac{1}{m_{x,t}}\frac{\partial m_{x,t}}{\partial t} = \beta_x k_t.$$

The left hand side of this last formula can be approximated in a context of discrete time as $z_{x,t}$.

The reason of this particular formulation of mortality improvement rates

The computation of the mortality improvement rates is performed using the formula (3.2) instead of the more simple and intuitive rate $m_{x,t}/m_{x,t-1}$, in order to reduce the impact of extreme values. In fact, it is reported in Subsection 2.6.1 how some authors just apply the ratio between one mortality rate and the previous year's one, but then the effects of random variations need to be removed by smoothing. An example of this is the approach proposed by Richards *et al.* (2005) (see Subsection 2.6.1).

Here the choice is not to smooth in a second stage, but rather to use a formulation that generates improvement rates where the size of the outliers is reduced.

3.3 The generalized parametric structures

3.3.1 Notation for multiple population

From now on, the index $i = 1, \ldots, I$ denotes subpopulation i among the I populations under study. For each i, it is assumed that the following data are available: for ages $x = x_1, \ldots, x_k$ and (consecutive) calendar years $t = t_1, \ldots, t_n$

- $D^i_{x,t}$, the number of deaths last birthday x in year t

- $\text{ETR}^i_{x,t}$, the central exposure at age x in year t.

3.3.2 Central mortality rates

The next step consists in computing central death rates

$$m^i_{x,t} = \frac{D^i_{x,t}}{\text{ETR}^i_{x,t}}$$

and year-on-year improvement rates

$$z^i_{x,t} = \frac{1 - m^i_{x,t}/m^i_{x,t-1}}{1 + m^i_{x,t}/m^i_{x,t-1}}$$

using the definitions outlined in Subsections 2.3.2 and 3.2.2, and considering the notation introduced in Section 3.3.1.

In this context, the Brouhns *et al.* (2002) version of the Lee-Carter model specified in (3.1), models the number of deaths as Poisson random variables

$$D^i_{x,t} \sim \text{Poisson}(\text{ETR}^i_{x,t}\, m^i_{x,t}),$$

independent across ages, years and subpopulations. The mean of these variables is modelled linearly through a number of time factors, according to

$$\log m^i_{x,t} = \alpha^i_x + \sum_{j=1}^{L} \beta^i_{x,j} k_{t,j} \tag{3.4}$$

where L is the number of considered factors. This expression is in spirit similar to those found in Booth *et al.* (2002b) and Hyndman and Ullah (2007).

3.3.3 Mortality improvement rates

When modelling improvement rates, it is assumed that the $z^i_{x,t}$ are realizations of Gaussian random variables

$$Z^i_{x,t} \sim N(\eta^i_{x,t}, \sigma^2_i),$$

independent across ages, years and subpopulations. Note that the variance is allowed to vary between populations. The mean of these variable is expressed by

generalising (3.3) in a form similar to (3.4), that is

$$\eta^i_{x,t} = \sum_{j=1}^{L} \beta^i_{x,j} k_{t,j}. \tag{3.5}$$

3.3.4 The philosophy of the models

The aim of (3.4)-(3.5) is to consider a general framework allowing for different levels of complexities and interactions in and between the subpopulations. The number L of factors will usually be driven by the number of populations and the chosen degree of complexity. Some particular cases of (3.4) and (3.5) are considered, in order to make estimation feasible and ease comparison between the models. More precisely, five specifications of (3.4) are selected. Subsequently, the counterparts of these five models are defined in terms of mortality improvement rates (3.5). The considered models are ten in total.

3.4 The proposed models

The considered models will be defined in Subsections 3.4.1 and 3.4.2. The names of the models are composed by two parts. The first part is a capital letter which refers to the parametric structure:

- P for the models that target the central mortality rates (P is for the Poisson random variable used for modelling the number of deaths);
- M for the ones which model the mortality improvement rates (M is for MIR, the abbreviation used in Haberman and Renshaw (2013) for mortality improvement rates).

The second part is a word which recall the type of particular case of the selected generalized parametric structure: double, common, simple, division and one.

3.4.1 The P models

The five specifications of (3.4) are listed below.

1. P-double:
$$\log m^i_{x,t} = \alpha^i_x + \beta^i_{x,1} k^i_{t,1} + \beta^i_{x,2} k^i_{t,2}$$
with the identifiability constraints $\sum_t k^i_{t,1} = 0$, $\sum_x \beta^i_{x,1} = 1$, $\sum_t k^i_{t,2} = 0$, $\sum_x \beta^i_{x,2} = 1$ and $\sum_x \beta^i_{x,1}\beta^i_{x,2} = 0$ for all i.

2. P-common:
$$\log m^i_{x,t} = \alpha^i_x + \beta^i_{x,1} k_{t,1} + \beta^i_{x,2} k^i_{t,2}$$
with the identifiability constraints $\sum_t k_t = 0$, $\sum_x \beta^i_{x,1} = 1$, $\sum_t k^i_t = 0$, $\sum_x \beta^i_{x,2} = 1$ and $\sum_x \beta_x \beta^i_x = 0$ for all i.

3. P-simple:
$$\log m^i_{x,t} = \alpha^i_x + \beta^i_x k^i_t$$
with the identifiability constraints $\sum_t k^i_t = 0$, $\sum_x \beta^i_x = 1$ for all i.

4. P-division:
$$\log m^i_{x,t} = \alpha^i_x + \beta^i_x k^i_t$$
with $k^i_t = k^{(h)}_t$ for $i \in J_h$, where $J_1, \ldots, J_{I'}$ is a partition of $\{1, \ldots, I\}$; the identifiability constraints are $\sum_t k^h_t = 0$ and $\sum_{i \in J_h, x} \beta^i_x = |J_h|$ for $h = 1, \ldots, I'$. Here $|J|$ is the cardinality of the set J.

5. P-one:
$$\log m^i_{x,t} = \alpha^i_x + \beta^i_x k_t$$
with the identifiability constraints $\sum_t k_t = 0$ and $\sum_{i,x} \beta^i_x = I$.

3.4.2 The M models

The five particular cases of (3.5) are defined below.

6. M-double:
$$\eta^i_{x,t} = \beta^i_{x,1} k^i_{t,1} + \beta^i_{x,2} k^i_{t,2}$$
with the identifiability constraints $\sum_x \beta^i_{x,1} = 1$, $\sum_x \beta^i_{x,2} = 1$ and $\sum_x \beta^i_{x,1}\beta^i_{x,2} = 0$ for all i.

7. M-common:
$$\eta^i_{x,t} = \beta^i_{x,1} k^i_{t,1} + \beta^i_{x,2} k^i_{t,2}$$

with the identifiability constraints $\sum_x \beta^i_{x,1} = 1$, $\sum_x \beta^i_{x,2} = 1$ and $\sum_x \beta_x \beta^i_x = 0$ for all i.

8. M-simple:
$$\eta^i_{x,t} = \beta^i_x k^i_t$$
with the identifiability constraint $\sum_x \beta^i_x = 1$ for all i.

9. M-division:
$$\eta^i_{x,t} = \beta^i_x k^i_t$$
with $k^i_t = k^{(h)}_t$ for $i \in J_h$, where $J_1, \ldots, J_{I'}$ is a partition of $\{1, \ldots, I\}$; the identifiability constraints are $\sum_{i \in J_h, x} \beta^i_x = |J_h|$ for $h = 1, \ldots, I'$.

10. M-one:
$$\eta^i_{x,t} = \beta^i_x k_t$$
with the identifiability constraint $\sum_{i,x} \beta^i_x = I$.

3.5 Discussion about the models

3.5.1 P-double and M-double

The models (1) and (6), called respectively P-double and M-double, are inspired by Renshaw and Haberman (2003b), where a Lee-Carter model with two bilinear components is considered. P-double is a Lee-Carter model with the error structure of (3.1) with a second bilinear component. Analogously, M-double is the parametric structure used for modelling $\eta^i_{x,t}$ with an additional bilinear component. The models are then estimated separately for every population. The aim of these parametric structures is to provide a good approximation to the data, using more parameters than in the other considered approaches.

P-double and M-double models will be the starting point for further reduction of the number of parameters. Furthermore, these general structures are important in order to compare the quality of the forecast and its descriptive capacity.

3.5.2 P-common and M-common

Models (2) and (7), called respectively P-common and M-common, are very similar to the models marked with *double*, but with the introduction of a common factor. As a matter of fact, the time index in the first bilinear component is in common for all of the considered populations. Nevertheless, the model still has two bilinear components, allowing for a good approximation to the data. However, the number of parameters is reduced. These are the first models, between the presented ones, that allow for common factors in this step of the analysis.

This model is inspired by Li and Lee (2005) which used a common factor estimated on all the considered populations and, in a second stage, a second bilinear component was estimated in a specific way for every population. This model structure allow to explain the general trend with the first component and the specific mortality values in terms of distance from this general component. The intuition behind these formulations is that mortality has a general trend for related populations, and hence the dissimilarities can be explained with a second component that merges the specific characteristics as well as the random variations.

3.5.3 P-simple and M-simple

Models (3) and (8), called respectively P-simple and M-simple, are obtained from the previous ones by removing the common factors. P-simple and M-simple present one bilinear component and correspond to the original formulations of the models, as presented in (3.1) and (3.3). Like in the models marked with *double*, the parameters are now estimated at this step without considering interactions between the populations. P-simple and M-simple are the complete form of the models considering only one bilinear component.

3.5.4 P-division and M-division

Models (4) and (9), called respectively P-division and M-division, are similar to the models marked with *simple*. The case that is investigated regards a number I of populations. The intuition behind these *division* approaches is that in the I populations data, there is a number of different mortality trends which is less than or equal to I. If I is large enough, it is likely that some of the mortality trends

would be very similar and create clusters of similar paths. In this case it is possible to define a model with a number of time-varying coefficients I' lower than the number of populations I. This allow to reduce the number of parameters without a strong reduction in the descriptive capacity of the model. These formulations have some common time varying coefficients, one for each cluster of populations.

3.5.5 P-one and M-one

Models (5) and (10), called respectively P-one and M-one, follow the same idea of the models marked with *division*, but assume that there is a unique cluster. In fact, the hypothesis is that only one time varying coefficient exists, and hence all the populations share the time pattern of mortality and differ only in the age varying coefficients. The number of parameters is the smallest among the models presented.

3.5.6 Complexity

Both set of models (1)-(5) and (6)-(10) are presented in decreasing order of complexity and number of parameters (see Table 3.1).

Models (2)-(5) and (6)-(10) can be seen as particular cases of the more general forms P-double and M-double. Recall that P-common model can be obtained from P-double model by assuming

$$\beta^i_{x,1}k^i_{t,1} = \beta^*_{x,1}k^*_{t,1} \quad \forall i = 1, \ldots, I.$$

P-simple model derives by P-common if $\beta^*_{x,1} = 0$. In the same way, P-division model can be obtained from the P-simple model by just requiring some of the time varying coefficients to have the same value with respect to the similar populations. Note that the P-one model is an P-division model when all the time varying coefficients have the same values. The same idea applies to M-models.

It follows that each model includes the next one by restricting some of the parameters. In other words (1)-(5) is a complete sequence of nested models, and the same applies to (6)-(10).

	model	number of time factors	number of parameters	number of constraints
1.	P-double	$2I$	$I(3k+2n)$	$5I$
2.	P-common	$I+1$	$k(2I+1)+n(I+1)$	$3I+2$
3.	P-simple	I	$I(2k+n)$	$2I$
4.	P-division	I'	$2kI+nI'$	$2I'$
5.	P-one	1	$2kI+n$	2
6.	M-double	$2I$	$2I(k+n-1)$	$3I$
7.	M-common	$I+1$	$(I+1)(k+n-1)$	$1+2I$
8.	M-simple	I	$I(k+n-1)$	I
9.	M-division	I'	$kI+(n-1)I'$	I'
10.	M-one	1	$kI+(n-1)$	1

TABLE 3.1: Number of time factors (H), parameters (d) and constraints for the ten models.

3.5.7 The identifiability constraints

P models

Due to its formulation, P-simple needs the identifiability constraints defined for the classical version of the model (Lee and Carter, 1992). When the time-varying coefficients are reduced, as in P-division and in P-one, it is no longer possible to use the same constraints.
In P-division the number k_t vectors coincides with the number of partitions. Every one of this time-varying coefficients have the constraint regarding its sum, that should be equal to 0. Regarding the constraints on the β_x^i, they are outlined in order to consider jointly all the parameters within the same partition. The identifiability constraints in P-one are designed following the same idea than in P-division, as if the considered populations were all in the same partition.

A different approach should be followed in order to understand the identifiability constraints in the models with two bilinear components, *i.e.* P-double and P-common. In P-double five constraints are present for every population. Two of them regard the first bilinear component, one on $\beta_{x,1}^i$ parameters and one on $k_{t,1}^i$ parameters. Other two constraints interest the second bilinear component, one on $\beta_{x,2}^i$ parameters and one on $k_{t,2}^i$ parameters. These four constraints follow the same logic than the P-simple constraints, just replaced for the second bilinear term. The fifth constraint, which is $\sum_x \beta_{x,1}^i \beta_{x,2}^i = 0$ for every i, links the two bilinear terms. This is imposed in order to assure the orthogonality between the estimated values

of $\beta^i_{x,1}k^i_{t,1}$ and $\beta^i_{x,2}k^i_{t,2}$. The exigence of imposing this fifth constraint follows by the estimation procedure. In fact, the estimated parameters are orthogonal if they are obtained using SVD, as in Renshaw and Haberman (2003b).
The identifiability constraints used for P-common are an opportune variant of the P-double ones.

M models

The identifiability constraints used for M models are the opportune adaptation of the ones presented above for the P models. The number of constraints is reduced with respect to P models. This is due to the fact that the term α_x is no longer present. Considering the M-simple model, it follows that just one constraint on the β^i_x parameters is enough for identify the bilinear term of the i-th population. The constraints for the other M models are derived using this concept and the guidelines applied to the P models.

3.6 Forecast procedure

The next step is to model the time varying coefficients as a time series and proceed with forecast. The time series selected are a random walk with drift for the P-models and an auto regressive time series of order 1 for the M-models. These time series are the ones used in Lee and Carter (1992) and Haberman and Renshaw (2012). Note that, except for P-one and M-one, all the models require a multidimensional time series as they include multiple time indices (see Table 3.1). Therefore, the univariate random walk with drift

$$k_t = a + k_{t-1} + \varepsilon_t \quad \text{with} \quad \varepsilon_t \sim N(0, \sigma^2)$$

where a is the trend, is used for P-one, while for the other P-models, the multivariate random walw with drift

$$\begin{bmatrix} k^1_t \\ \vdots \\ k^H_t \end{bmatrix} = \begin{bmatrix} a^1 \\ \vdots \\ a^H \end{bmatrix} + \begin{bmatrix} k^1_{t-1} \\ \vdots \\ k^H_{t-1} \end{bmatrix} + \begin{bmatrix} \varepsilon^1_t \\ \vdots \\ \varepsilon^H_t \end{bmatrix} \quad \text{with} \quad \begin{bmatrix} \varepsilon^1_t \\ \vdots \\ \varepsilon^H_t \end{bmatrix} \sim N_H(0, \Sigma_H)$$

is used, where H is the number of time indices. Conversely, an the auto regressive time series of order one

$$k_t = a + \rho k_{t-1} + \varepsilon_t \quad \text{with} \quad \varepsilon_t \sim N(0, \sigma^2)$$

where ρ is the multiplicative constant, is used for M-one, while for other M-models, the corresponding multidimensional version

$$\begin{bmatrix} k_t^1 \\ \vdots \\ k_t^H \end{bmatrix} = \begin{bmatrix} a^1 \\ \vdots \\ a^H \end{bmatrix} + \begin{bmatrix} \rho_{11} & \dots & \rho_{1H} \\ \vdots & & \vdots \\ \rho_{H1} & \dots & \rho_{HH} \end{bmatrix} \begin{bmatrix} k_{t-1}^1 \\ \vdots \\ k_{t-1}^H \end{bmatrix} + \begin{bmatrix} \varepsilon_t^1 \\ \vdots \\ \varepsilon_t^H \end{bmatrix} \quad \text{with} \quad \begin{bmatrix} \varepsilon_t^1 \\ \vdots \\ \varepsilon_t^H \end{bmatrix} \sim N_H(0, \Sigma_H)$$

is considered.

Once the forecast of the time varying coefficients has been completed, the values of $\hat{m}_{x,t}^i$ for P-models can be computed. For the M-models, applied to mortality improvement rates, a further step is needed in order to get the mortality rates using the forecast values of $\hat{z}_{x,t}^i$. The forecast data $\hat{m}_{x,t}^i$ are obtained by applying iteratively the formula

$$\hat{m}_{x,t+j}^i = \hat{m}_{x,t+j-1}^i \frac{(2 - z_{x,t+j}^i)}{(2 + z_{x,t+j}^i)}, \quad j = 1, 2, 3, \dots \tag{3.6}$$

starting with $m_{x,t}^{i*}$, an adjusted value of $m_{x,n}^i$, the last column of the observed matrix $m_{x,t}^i$. The adjustment adopted consists in computing the mean of the last three observed values of $m_{x,t}^i$ and use the mean of the last two observed $z_{x,t}^i$ for obtaining $m_{x,t}^{i*}$. This procedure is done in order to diminish the influence of the last observed value in the construction of the forecast values, thus have results less affected by outliers.

3.7 Discussion

The Lee-Carter model is taken as starting point for forecasting mortality in the case of multiple populations data. The model is set in order to target the count of deaths (Brouhns *et al.*, 2002), and it is written in (3.4). This is the Lee-Carter model with generalised number of bilinear components. Five approaches are

obtained as particular cases of (3.4), with the aim of considering the interactions between the populations in different ways. These five models are inspired by the approaches present in literature (reviewed in Chapter 2), and are selected with two main characteristics:

- every model should have notable aspects;
- they should be as parsimonious as possible.

A modified version of these five models is also considered, where the object of the modelling is the mortality improvement rates, as defined in Haberman and Renshaw (2012). This is done in order to evaluate an alternative equivalent approach to mortality forecasting.

These ten methods are an attempt to consider different multiple populations Lee-Carter type models. The analysis of differences and similarities across them, as well as some results (in the next Chapter), can help to better understand some aspects of this problem. The aim is to discuss the characteristics that a Lee-Carter type model should have for being adequate for multiple populations mortality analysis.

Chapter 4

Application of generalized Lee-Carter models to Italian regions

4.1 Introduction

In Chapter 3 ten models have been introduced. All of those models can be applied to multiple population mortality data. However, each model has strengths and drawbacks, and this makes the selection of an appropriate model for an application to real populations a not simple problem.

In this chapter the ten models are applied to a multiple population mortality dataset. The aim is to make considerations about the characteristics of the models observing the performance of them. The performance of them is compared in terms of quality of fitting and forecast capacity.

In Section 4.2 the dataset considered for the application is introduced. Section 4.3 presents some further specifications of the models, with regard to the estimation and the forecast procedures. Section 4.4 and 4.5 outline the quantities considered for the model selections and introduce specific comments for all the models. Section 4.6 of this chapter presents some overall conclusions.

4.2 The application

4.2.1 The dataset

The considered data are the mortality rates of Italian regions. Italy is divided into 18 regions out of the official 20, since two regions (Val d'Aosta and Molise) are too small to be kept alone. These two last regions are merged with one of their neighbours.

The regions of a country are clearly related: such populations share some common characteristics. However, it is also true that Italian regions can be very different, either economically as well as along other dimensions, and this may be reflected in the mortality experience: so that the considered populations should be treated as linked but not equal. In this application $I = 18$, and the index $i = 1, \ldots, 18$ is used for indicating the regions, along this order: Piemonte-Valle d'Aosta, Lombardia, Trentino-Alto Adige, Veneto, Friuli-Venezia Giulia, Liguria, Emilia-Romagna, Toscana, Umbria, Marche, Lazio, Abruzzo-Molise, Campania, Puglia, Basilicata, Calabria, Sicilia, Sardegna. The geographical areas can be seen in Figure 4.1.

FIGURE 4.1: Italy divided in the considered 18 areas.

The data cover a span of 35 years, from 1974 to 2008[1]. In the analysis the focus is on the male mortality data for the age interval 20-89. The ages over 89 years are excluded due to the exiguous number of subjects involved, especially for smaller regions. Conversely, the ages 0-19, are excluded because of in this age interval there is a significant presence of empty values of $D_{x,t}$ and high variance of the mortality phenomenon, especially for smaller regions. In the Lee-Carter model, the problem of null values of $m_{x,t}$ can be solved in many ways. An example of this is the introduction of weights in the estimation procedure, as in Haberman and Renshaw (2012). That weights assume value 0 when there are empty or omitted data in $m_{x,t}$ matrix, and 1 elsewhere. By way of contrast, the high value of variance at lower ages could create problems due to the hypothesis of constant variance considered in the models with mortality improvement rates (the possibility of age-varying variance is studied in the next chapter).

The choice of the ages 20 to 89 is not uncommon. In fact, this age interval is adopted in many actuarial applications since the lower ages are not relevant for most insurance products. Regarding the higher ages, in practice it is often preferred to reconstruct the corresponding values with some extrapolative procedures. In this case also there are algorithms that can be used to obtain high-age values, such as the approaches proposed by Coale and Kisker (1990) and Haberman and Renshaw (2009).

As a first step, some graphical exploration analyses of the data are performed. Due to the number of populations and the huge variety of possible graphs, these plots are here omitted, with the exception of Figure 4.2, where four plots are shown, representing the evolution of $\log m^i_{x,t}$ for fixed ages for three out of the 18 populations (Lombardia, Lazio and Sicilia) with respect to the period 1974 to 2008. These plots confirm that the evolution of mortality follows similar patterns for the different populations (the phenomenon is less evident at lower ages, since the mortality rates are so low that random variations have more influence).

4.2.2 The number of groups

A further step must be done before estimating the models: the classification in P-division and M-division models. This classification is done in order to define

[1]The data were provided by Istat (`www.istat.it`).

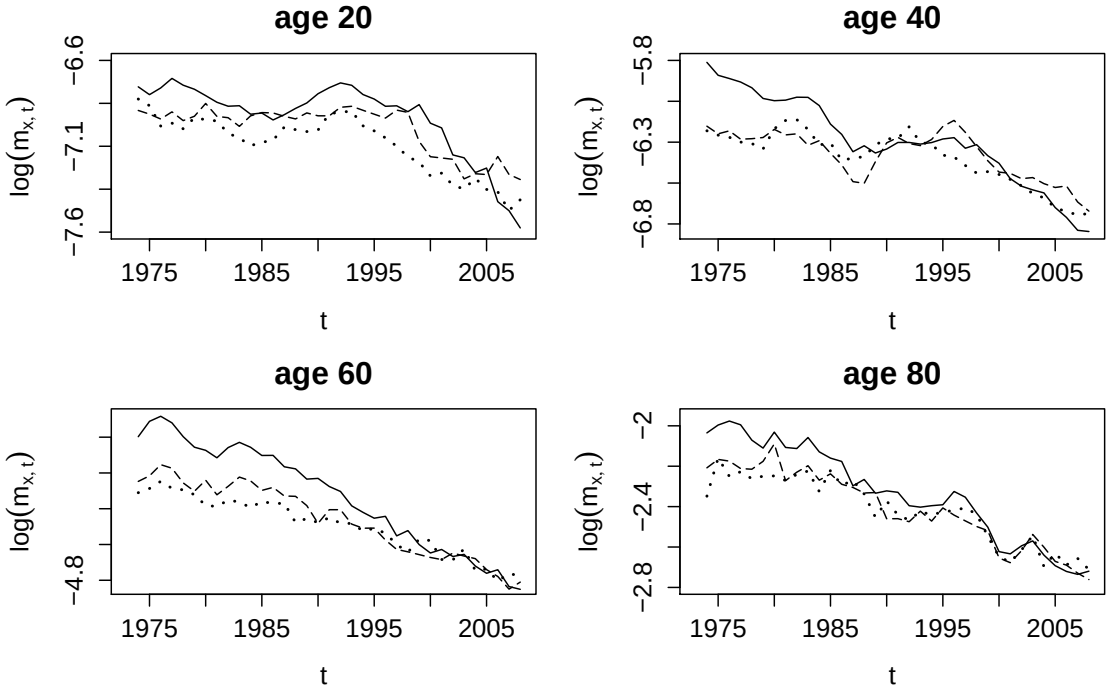

FIGURE 4.2: Evolution of $\log m^i_{x,t}$ for fixed ages of Lombardia (solid line), Lazio (dashed line) and Sicilia (dotted line).

groups of populations with similar mortality characteristics. The procedure for creating these groups here is inspired by the assignment of *model* life tables (as, for example, in the United Nations (1982) classification). The procedure consists in computing some indices regarding the target populations, then dividing the populations according to the values of the indices. More precisely, here the groups are obtained in terms of similarity with respect to life expectancy at birth, whose values are taken from Minelli *et al.* (2012). In this procedure only contiguous regions are allowed to merge together. The resulting group number is five. The two models become

- P-division: $D^i_{x,t} \sim \text{Poisson}(\text{ETR}^i_{x,t}\, m^i_{x,t})$, where

$$\log m^i_{x,t} = \alpha^i_x + \beta^i_x k^i_t$$

with $k^i_t = k^{(1)}_t$ for $i \in I_1 = \{1,2,3,4,5,6\}$, $k^i_t = k^{(2)}_t$ for $i \in I_2 = \{7,8,9\}$ and $k^i_t = k^{(3)}_t$ for $i \in I_3 = \{10,11,12\}$, $k^i_t = k^{(4)}_t$ for $i \in I_4 = \{13,18\}$, $k^i_t = k^{(5)}_t$ for $i \in I_5 = \{14,15,16,17\}$; the identifiability constraints are $\sum_t k^{(1)}_t = 0$ and $\sum_{i \in I_1,x} \beta^i_x = 1$, $\sum_t k^{(2)}_t = 0$ and $\sum_{i \in I_2,x} \beta^i_x = 1$, $\sum_t k^{(3)}_t = 0$ and $\sum_{i \in I_3,x} \beta^i_x = 1$, $\sum_t k^{(4)}_t = 0$ and $\sum_{i \in I_4,x} \beta^i_x = 1$, $\sum_t k^{(5)}_t = 0$ and $\sum_{i \in I_5,x} \beta^i_x = 1$;

- M-division: $Z^i_{x,t} \sim N(\eta^i_{x,t}, \sigma^2_i)$,, where

$$\eta^i_{x,t} = \beta^i_x k^i_t$$

with $k^i_t = k^{(1)}_t$ for $i \in I_1 = \{1,2,3,4,5,6\}$, $k^i_t = k^{(2)}_t$ for $i \in I_2 = \{7,8,9\}$ and $k^i_t = k^{(3)}_t$ for $i \in I_3 = \{10,11,12\}$, $k^i_t = k^{(4)}_t$ for $i \in I_4 = \{13,18\}$, $k^i_t = k^{(5)}_t$ for $i \in I_5 = \{14,15,16,17\}$; the identifiability constraints are $\sum_{i \in I_1,x} \beta^i_x = 1$, $\sum_{i \in I_2,x} \beta^i_x = 1$, $\sum_{i \in I_3,x} \beta^i_x = 1$, $\sum_{i \in I_4,x} \beta^i_x = 1$ and $\sum_{i \in I_5,x} \beta^i_x = 1$.

The groups for the models P-division and M-division can be obtained in other ways, that could be considering different variables rather than life expectancy (*e.g.* consider some aspects of the behaviour of the time varying coefficient k_t, estimated with a classical Lee-Carter model on each population) or using different statistics techniques (*e.g.* cluster analysis).

4.3 Specifications about the method

4.3.1 Estimation procedure

The models are estimated considering only the first 25 years of observed data, from 1974 to 1998 (the remaining 10 years, from 1999 to 2008, are used in order to assess the quality of the resulting forecasts).

The parameters of the ten models are estimated by maximum likelihood using iterative numerical procedures. Regarding the starting point for the optimisation procedure, the values used were set using the following criteria:

- for P-models, the starting points assuming the normal distribution for the errors and using the procedure described in Lee and Carter (1992);
- for M-models, the starting points are obtained using the first (or the first two) components of the singular value decomposition of the improvement rates matrix;
- in all the models, in the presence of common factors, the chosen starting points (for the common parameters) are those computed on the larger populations, since the latter are likely to be more stable and reliable.

In order to reduce the possibility that the algorithm converges to a local maxima, the optimisation procedure is repeated with modified starting values and is interrupted when there are no further improvements in the optimised likelihood.

4.3.2 Forecast procedure

There is a number of time series involved in this application, namely:

- two sets of eighteen time varying coefficients in the P-double and M-double models;
- one set of eighteen plus one time varying coefficients in the P-common and M-common models;
- eighteen time varying coefficients in the P-simple and M-simple models;
- five in the P-division and M-division models.

For these cases, one or more multivariate time series need to be used. In the other cases, the P-one and M-one models, a single time series is involved. The time series used in this application are those described in Section 3.6.

Regarding the starting adjusted value $m^*_{x,t}$ from which the iterative procedure described in Section 3.6 starts, in this application it is equal to

$$m^*_{x,25} = \frac{m_{x,23} + m_{x,24} + m_{x,25}}{3} \frac{(2 - z_{x,23-24})}{(2 + z_{x,23-24})},$$

where $z_{x,23-24}$ is the average between $z_{x,23}$ and $z_{x,24}$. Recall that the values corresponding to $z_{x,24}$ are the most recent in-sample data, due to the transformation from mortality rates to mortality improvement rates.

4.4 Model selection

4.4.1 Goodness of fit indices based on information criteria

The indices described above are composed of two parts: the log-likelihood and a function of the number of parameters. In such an index the number of parameters

has the opposite sign with respect to the log-likelihood, since the goal is to consider as the best model the one with the higher log-likelihood but fewer parameters. Due to the fact that the number of parameters is used as a penalisation of the log-likelihood, it is possible to refer to these indices as penalised log-likelihood indices.

The two most commonly used indices for the goodness of fit based on the penalised log-likelihood indices are (see Burnham and Anderson (2004))

- $AIC = 2d - 2\ell$ with d the dimension of the parametrised prediction structure;
- $BIC = d\log(g) - 2\ell$ with g the numbers of data;

where

$$\ell = \sum_i \sum_x \sum_t (D^i_{x,t}(\log \hat{m}^i_{x,t}) - \mathrm{ETR}^i_{x,t}\hat{m}^i_{x,t})$$

for P-models, defined up to an additive constant independently of the chosen model, and

$$\ell = -\frac{1}{2}\sum_i \sum_x \sum_t \left\{ \log(2\pi\hat{\sigma}_i^2) + \frac{(z^i_{x,t} - \hat{\eta}^i_{x,t})^2}{\hat{\sigma}_i^2} \right\}$$

for M-models.

When the value of d is large relative to the number of data g, the index AIC_c, an adjusted version of the Akaike information criterion, defined by

$$\mathrm{AIC}_c = 2d + \frac{2d(d+1)}{g-d-1} - 2\ell$$

can be considered (Burnham and Anderson, 2004). AIC_c can be used when $g/d < 40$. Clearly the value of AIC_c converges to AIC as g gets large relative to d.

The best models are those with smaller values of the indices. That indices cannot be compared across the two main model structures. In fact, the values can be used to compare the five models applied to mortality rates and, separately, the five models applied to mortality improvement rates. For this reason, the results are presented separately for the first five and the other models.

Considering penalised log-likelihood indices, the absolute values are less important than the relative values of the indices themselves and their ranked order. This is due to the fact that the values of the penalised log likelihood indices contain

arbitrary constants and are affected by the dimension of the data. In order to make the values easy to interpret and to highlight the ranked order, the difference Δ between the target index and its minimum value is computed. It follows that $\Delta = 0$ for the best model and $\Delta > 0$ for the others, selecting as the best model the one with smaller values of these differences.

The values of AIC, AIC_c and BIC are presented in Table 4.1, together with other related quantities.

	P-double	P-common	P-simple	P-division	P-one
d	4590	3009	2934	2635	2543
ℓ	-28332454	-28332976	-28334463	-28334716	-28335099
g	31500	31500	31500	31500	31500
AIC	56674088	56671969	56674794	56674702	56675285
Δ-AIC	2119	0	2825	2733	3315
rank-AIC	2	1	4	3	5
g/d	7	10	11	12	12
AIC_c	56675654	56672605	56675397	56675183	56675731
Δ-AIC_c	3049	0	2792	2578	3127
rank-AIC_c	4	1	3	2	5
BIC	56712450	56697118	56699316	56696725	56696538
Δ-BIC	15912	579	2777	186	0
rank-BIC	5	3	4	2	1

	M-double	M-common	M-simple	M-division	M-one
d	3330	1749	1674	1375	1283
ℓ	42056	39903	38760	37058	36617
g	30240	30240	30240	30240	30240
AIC	-77453	-76309	-74173	-71366	-70668
Δ-AIC	0	1144	3280	6086	6785
rank-AIC	1	2	3	4	5
g/d	9	17	18	22	24
AIC_c	-76628	-76094	-73977	-71235	-70554
Δ-AIC_c	0	534	2652	5393	6074
rank-AIC_c	1	2	3	4	5
BIC	-49757	-61763	-60250	-59931	-59997
Δ-BIC	12005	0	1512	1832	1765
rank-BIC	5	1	2	4	3

TABLE 4.1: Dimension of the parametrised prediction structure (d), likelihood of the model (ℓ), dimension of the data (g), value of g/d, AIC, AIC_c and BIC (and its Δ and its rank) of the ten models (when applicable, the values are rounded to the integer)

4.4.2 The likelihood-ratio test

As the ten models can be seen as two sets of nested models (see Subsection 3.5.6), it is possible to use the likelihood ratio test to select the best model. The null hypothesis is that the restricted model is correct, and the alternative hypothesis is in favour of the more general one (in Cairns *et al.* (2009) there is an example of

likelihood ratio test used in this framework). The test statistic LR is equal to

$$LR = 2(\ell_G - \ell_R),$$

where ℓ_G is the likelihood of the general model and ℓ_R the likelihood of the restricted model. Under the null hypothesis, LR could be approximated by a chi square random variable with $v = d_G - d_R$ degrees of freedom, where d_G and d_R are the dimensions of the parametrised prediction structure of the general and the restricted models, respectively. The null hypothesis is rejected if LR is too large, therefore

$$LR > \chi^2_{v,0.95}$$

where $\chi_{v,0.95}$ is the 95th percentile of a chi square random variable with v degrees of freedom, corresponding to the significance level 0.05.

In Table 4.2 are considered all the possible combinations between general and restricted models, 20 in total, with the value of v, the value of the test statistic, the critical value and the p-value, this latter obtained as

$$p = 1 - Prob(\chi^2_v > 2(\ell_G - \ell_R)).$$

Restricted	General	LR	v	$\chi^2_{v,0.95}$	p-value
P-common	P-double	1 043	1 581	1 675	1
P-simple	P-double	4 018	1 656	1 752	<0.001
P-division	P-double	4 524	1 955	2 059	<0.001
P-one	P-double	5 291	2 047	2 153	<0.001
P-simple	P-common	2 975	75	96.2	<0.001
P-division	P-common	3 481	374	420	<0.001
P-one	P-common	4 247	466	517	<0.001
P-division	P-simple	506	299	340	<0.001
P-one	P-simple	1 273	391	438	<0.001
P-one	P-division	766	92	115	<0.001
M-common	M-double	4 306	1 581	1 675	<0.001
M-simple	M-double	6 592	1 656	1 752	<0.001
M-division	M-double	9 996	1 955	2 059	<0.001
M-one	M-double	10 879	2 047	2 153	<0.001
M-simple	M-common	2 286	75	96.2	<0.001
M-division	M-common	5 690	374	420	<0.001
M-one	M-common	6 573	466	517	<0.001
M-division	M-simple	3 405	299	340	<0.001
M-one	M-simple	4 287	391	438	<0.001
M-one	M-division	883	92	115	<0.001

TABLE 4.2: Likelihood ratio test for all the possible combinations of general and restricted models

4.4.3 Mean absolute percentage errors

The goodness of fit or forecast of a model can be measured, for instance, with the Mean Absolute Percentage Error (MAPE). This index applied to in-sample data is defined as

$$\text{MAPE}^i = \frac{1}{n_1 k} \sum_{x,t} \left| \frac{m^i_{x,t} - \hat{m}^i_{x,t}}{m^i_{x,t}} \right| \tag{4.1}$$

where $n_1 \leq n$ (in this case $n_1 = 25$ and $k = 70$), $m^i_{x,t}$ are the observed values and $\hat{m}^i_{x,t}$ the values predicted by using the model. The MAPE for out-of-sample data can be easily derived by (4.1).

The values of MAPE obtained for in-sample fitting are summarized in Table 4.3. The fitted values $\hat{m}_{x,t}$ in the case of M-models are obtained iteratively starting from the last year of the considered data. The formula that should be used for this procedure can be derived by (3.6), that is the equation used for the construction of the predicted data.

This MAPE is also applied, for the ten models, to the out-of-sample forecast computed with a time horizon of 10 years. The values of MAPE for every model and for every population are shown in Table 4.4.

The values of MAPE seem high, but it should be noted that the underlying phenomena have a high volatility, so there is a level of error that cannot be avoided. This error can be observed both in the in-sample and in the out-of-sample analysis.

The MAPE for the in-sample analysis is always lower than the MAPE of out-of-sample analysis. This is due to the structure of the forecast data. If the observed pattern of mortality evolves in an unusual way for even a few years, this impacts markedly on the values of MAPE.

4.4.4 Graphical analysis

Due to the nature of the phenomenon, graphical analysis is a powerful tool for analysing the models and comparing them. A graphical analysis of the residual plots can be useful for investigating if the models are able to describe the general shape of the data and to capture any systematic patterns. The residual plots are constructed by plotting the scaled residuals with respect to age, year and cohort.

The scaled residuals are obtained with

$$r^i_{x,t} = \frac{D^i_{x,t} - ETR^i_{x,t}\hat{m}^i_{x,t}}{\sqrt{ETR^i_{x,t}\hat{m}^i_{x,t}}}$$

for P-models, where $\hat{m}^i_{x,t}$ are the estimated data, and

$$r^i_{x,t} = \frac{z^i_{x,t} - \hat{\eta}^i_{x,t}}{\sqrt{\hat{\sigma}^2_i}}, \text{ with } \hat{\sigma_i}^2 = \sum_{x,t} \frac{(z^i_{x,t} - \hat{\eta}^i_{x,t})^2}{\nu}$$

for M-models, where ν is the dimension of the dataset, *i.e.* number of years multiplied by the number of ages. It should be noted that the scaled residuals are computed with respect to the target quantity in the optimisation procedure: the number of deaths $D^i_{x,t}$ in the first case and the mortality improvement rates in the second one.

The value of d, *i.e.* the dimension of the parametrised prediction structure, refers to the single population, therefore is the number of parameters that influence $m^i_{x,t}$ (or $z^i_{x,t}$ in the other case).

The aim of this analysis is to check if the residuals are randomly distributed above and below the horizontal line representing the value 0. Additionally, the presence of any regular patterns in the residual plot should be checked: its presence would suggest that the model has not captured the general evolution of the underlying phenomena.

The residual plots for the ten models for one region (Lombardia) are presented in 4.3 and 4.4 (the others are not reported here).

4.4.5 Actuarial application

Since the presented models are used to evaluate the general trend of mortality, an index which takes into account several years of forecast values would be a more appropriate way for comparing the predictive capacity of the models. An actuarial index, the truncated expected residual lifetime computed along cohort trajectories, is considered for evaluating the quality of the forecasts. The computation is performed considering a time horizon of 10 years in order not to introduce a mortality extrapolation at higher ages.

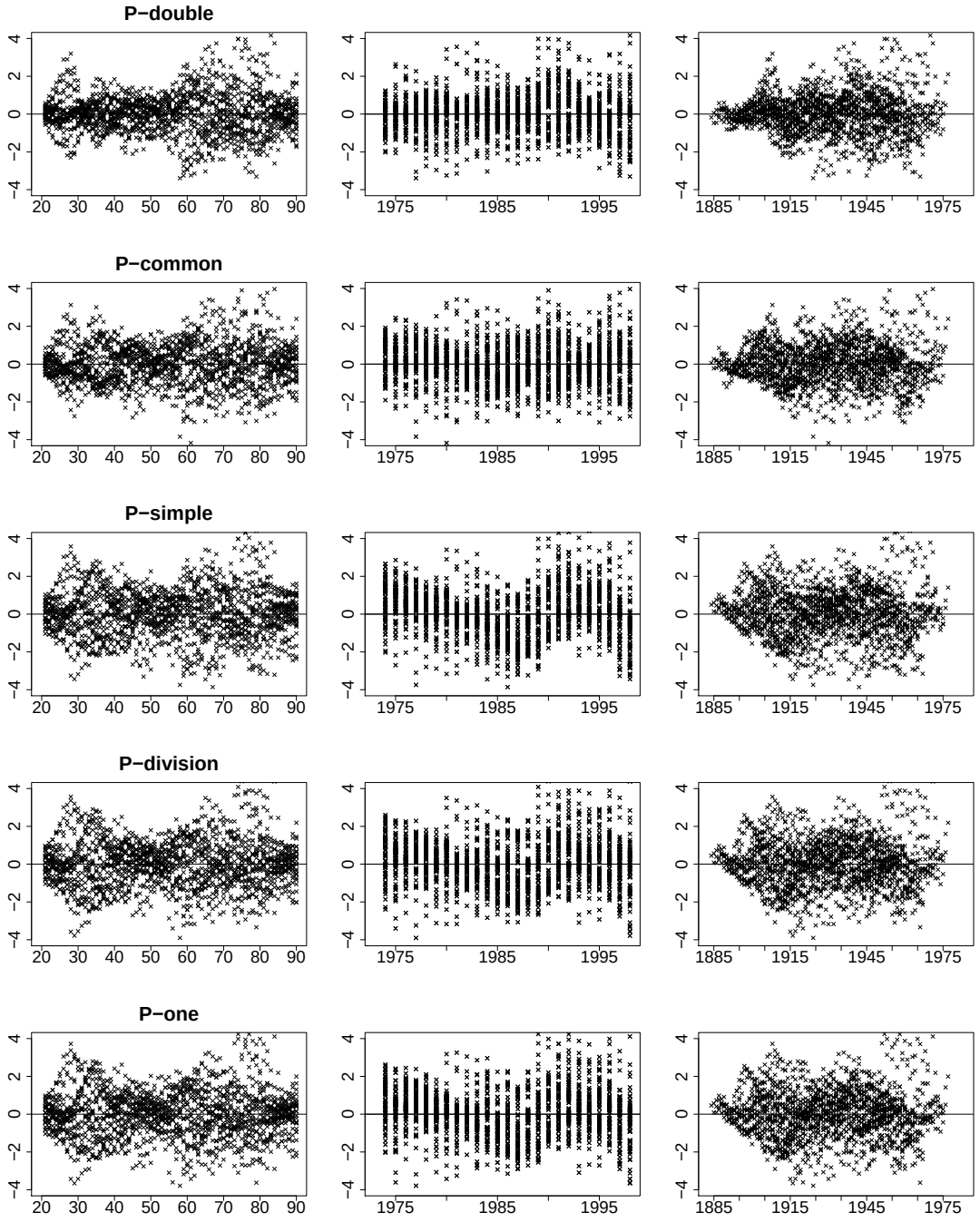

FIGURE 4.3: Age, year and cohort residual plots for P-models - population (2), Lombardia

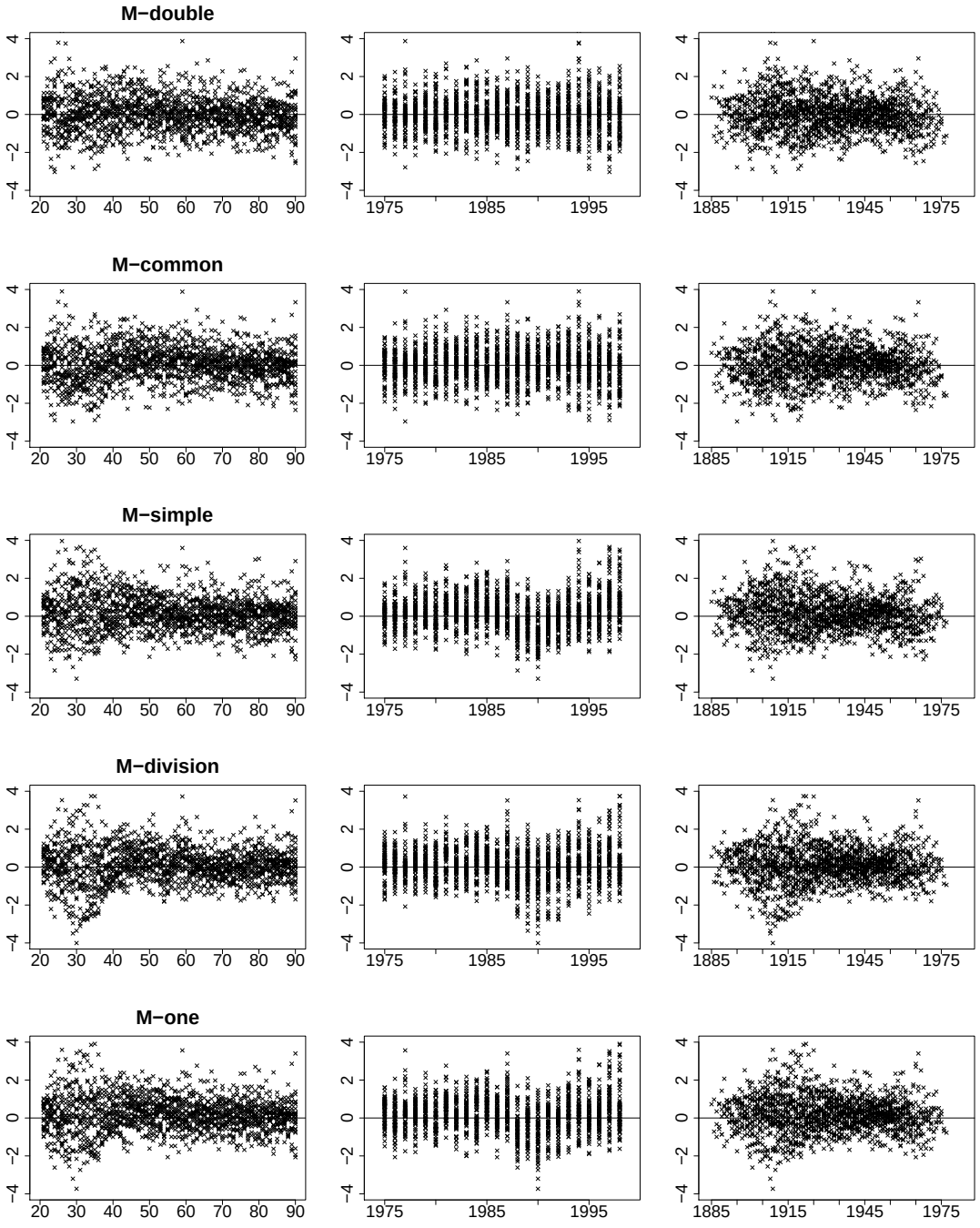

FIGURE 4.4: Age, year and cohort residual plots for M-models - population (2), Lombardia

Since the probability of death $q^i_{x,t}$ are now needed, they are evaluated applying the transformation

$$q^i_{x,t} \approx 1 - \exp(-m^i_{x,t}).$$

The expected residual lifetime truncated after 10 years for population i, denoted by $e^i_{x:\overline{10}|}$, is computed by

$$e^i_{x:\overline{10}|} = \frac{\sum_{j=1}^{10} l^i_{x+j}(t_n + j)\{1 - \frac{1}{2}q^i_{x+j,t_n+j}\}}{l^i_x(t_n)},$$
$$l^i_{x+1}(t_n + 1) = (1 - q^i_{x,t})l^i_x(t_n),$$

where t_n is the most recent time period for which data are available and $l^i_x(t_n)$ are set to an arbitrary value (Haberman and Renshaw, 2013). This index is computed for all the regions for ages 60, 70 and 80. The results are summarised in Table 4.5, Table 4.6 and Table 4.7, for ages 60, 70 and 80, respectively. In these tables, other than the observed values and the values obtained by the models with its percentage error, means and standard deviations of the errors are reported.

4.5 Discussion of the results

In this section, taking into account all the diagnostic tools and test introduced, the alternative models are compared and discussed. The indices are shown in the tables 4.1-4.7 for all models and populations (where appropriate). As explained before, the indices based penalised log-likelihood are clearly used to compare only models that share the same error structure.

4.5.1 P-double model

This is the model configuration with the highest number of parameters among those here considered.

According to the likelihood-ratio test this model should be preferred over the models nested in it, with the exception of P-common. As for the penalised log-likelihood indices, the model is ranked second best after P-common according to AIC. On the other hand, when looking at corrected AIC and BIC, this model is

REGION	P-double	P-common	P-simple	P-division	P-one	M-double	M-common	M-simple	M-division	M-one
1	4.04	4.51	5.21	5.3	5.3	9	9.15	10.41	11.38	13.54
2	4.18	4.84	5.94	5.9	5.9	9.49	12.66	13.11	17.35	14.04
3	5.84	6.56	6.74	6.94	6.99	16.59	18.25	18.18	14.22	20.4
4	3.78	4.63	5.2	5.36	5.37	10.15	10.33	12.07	13.02	15.21
5	5.32	6.16	6.05	6.5	6.58	12.19	15.62	13.36	18.19	16.45
6	4.86	5.59	6.4	6.83	6.75	14.18	10.51	15.24	28.11	14.24
7	4.65	5.12	6.32	6.29	6.31	11.29	11.2	19.17	19.82	13.91
8	4.4	4.71	5.21	5.3	5.3	12.48	9.79	15.64	17.32	11.96
9	6.94	7.5	7.69	7.76	8.2	16.47	11.62	17.43	17.05	15.63
10	4.98	5.67	6.12	6.19	6.28	11.93	9.33	13.13	10.41	12.28
11	4.02	4.31	4.99	5.22	5.3	7.39	8.04	9.67	12.83	10.72
12	5.38	5.95	5.8	5.97	5.96	15.61	12.57	16	11.78	13.26
13	3.45	4.05	4.08	4.1	4.41	9.67	8.74	10.26	10.33	10.81
14	4	4.75	5.42	5.46	5.51	9.35	8.15	13.07	12.61	10.56
15	7.42	8.51	8.01	8.28	8.92	13.9	12.38	14.49	15.68	17.79
16	4.43	5.57	5.51	5.66	5.75	13.19	12.1	13.66	12.84	14.26
17	3.76	4.45	4.93	5.02	5.09	8.06	11.31	10.28	10.77	12.73
18	4.88	5.47	6.35	7.05	6.94	10.81	9.85	12.71	14.2	14.33
mean	4.8	5.46	5.89	6.06	6.16	11.76	11.2	13.77	14.88	14.01
st. dev.	1.05	1.13	0.94	1.01	1.09	2.72	2.51	2.7	4.25	2.45

TABLE 4.3: MAPE of fitted data with respect to observed data

REGION	P-double	P-common	P-simple	P-division	P-one	M-double	M-common	M-simple	M-division	M-one
1	16.09	14.85	18.47	19.55	27.38	18.9	33.37	19.19	19.55	21.85
2	25.9	28.27	37.75	35.98	41.55	20.24	35.27	30.3	25.25	29.62
3	14.86	13.68	16.3	18.31	32.01	17.61	35.12	20.53	19.32	24.21
4	16.06	13.64	15.91	16.58	29.16	17.28	32.32	19.43	17.3	22.52
5	16.44	15.94	16.38	14.08	26.85	17.94	34.35	18.89	25.77	23.42
6	31.84	35.93	46.42	45.14	43.39	24.71	42.45	33.79	55.36	35.56
7	23.22	27.78	34.83	34.73	37.03	23.44	37.65	30.28	29.73	29.09
8	15.94	14.49	19.28	19.63	25.62	16.84	32.85	20.59	21.52	18.41
9	17.16	15.46	19.14	16.69	21.26	22.43	37.66	24.96	24.34	22.04
10	14.2	14.72	18.42	17.67	22.81	17.1	33.9	16.84	18.06	17.22
11	19.56	19.02	24.98	25.19	27.49	16.36	34.04	17.69	21.54	22.1
12	8.81	9.78	8.84	10.21	15.12	13.91	29.7	14.06	10.9	12.32
13	11.24	9.6	12.36	12.82	16.87	11.87	29.73	12.21	12.39	12.19
14	14.33	15.65	20.53	19.35	23.29	16.94	33.37	21.04	17.78	17.89
15	18.47	16.93	18.58	18.65	21.88	19.91	34.99	19.89	20.1	19.9
16	15.6	14.17	18.44	21.47	25.57	17.08	32.46	17.07	17.08	16.52
17	12.89	16.58	21.83	22.53	25.41	12.87	31.44	14.76	14.97	13.59
18	20.59	19.92	25.43	24.87	26.07	20.21	34.04	19.82	18.35	19.69
mean	17.4	17.58	21.88	21.86	27.15	18.09	34.15	20.63	21.63	21.01
st. dev.	5.26	6.53	9.02	8.56	7.27	3.3	2.91	5.63	9.36	5.94

TABLE 4.4: MAPE of forecast data with respect to observed data

REGION	Observed	P-double		P-common		P-simple		P-division		P-one		M-double		M-common		M-simple		M-division		M-one	
1	9.32	9.31	0.15	9.29	0.31	9.31	0.15	9.3	0.28	9.24	0.93	9.33	0.07	9.35	0.28	9.33	0.08	9.33	0.06	9.31	0.14
2	9.26	9.24	0.21	9.27	0.09	9.23	0.35	9.27	0.02	9.19	0.81	9.3	0.44	9.33	0.68	9.27	0.08	9.31	0.45	9.28	0.2
3	9.32	9.31	0.16	9.31	0.16	9.32	0.04	9.31	0.21	9.24	0.95	9.34	0.21	9.35	0.23	9.33	0.09	9.33	0.05	9.31	0.18
4	9.29	9.27	0.21	9.27	0.28	9.27	0.29	9.27	0.28	9.2	0.98	9.33	0.38	9.36	0.68	9.32	0.33	9.35	0.55	9.31	0.16
5	9.25	9.23	0.17	9.2	0.54	9.2	0.57	9.22	0.28	9.15	1.06	9.29	0.48	9.33	0.83	9.31	0.61	9.27	0.22	9.28	0.37
6	9.32	9.31	0.14	9.32	0	9.32	0.04	9.3	0.23	9.24	0.84	9.3	0.21	9.33	0.04	9.29	0.31	9.31	0.13	9.29	0.36
7	9.36	9.35	0.14	9.34	0.19	9.34	0.18	9.34	0.17	9.3	0.65	9.4	0.45	9.41	0.5	9.36	0	9.36	0.01	9.4	0.38
8	9.37	9.35	0.16	9.36	0.08	9.35	0.15	9.35	0.19	9.31	0.6	9.39	0.26	9.42	0.62	9.37	0.07	9.37	0.08	9.4	0.39
9	9.37	9.36	0.17	9.35	0.2	9.35	0.24	9.37	0.01	9.34	0.4	9.38	0.1	9.4	0.33	9.37	0	9.37	0.01	9.41	0.42
10	9.42	9.41	0.13	9.4	0.23	9.4	0.22	9.42	0.09	9.38	0.42	9.41	0.19	9.43	0.12	9.41	0.13	9.42	0.02	9.42	0
11	9.32	9.3	0.16	9.31	0.13	9.3	0.18	9.29	0.28	9.26	0.61	9.35	0.4	9.36	0.49	9.35	0.39	9.33	0.17	9.33	0.16
12	9.35	9.34	0.08	9.34	0.07	9.34	0.11	9.32	0.28	9.3	0.54	9.32	0.35	9.36	0.06	9.32	0.33	9.34	0.09	9.34	0.14
13	9.2	9.18	0.15	9.18	0.22	9.18	0.18	9.18	0.22	9.15	0.54	9.19	0.02	9.23	0.34	9.19	0.03	9.2	0.02	9.21	0.17
14	9.37	9.35	0.15	9.34	0.26	9.34	0.29	9.35	0.17	9.32	0.5	9.38	0.09	9.4	0.33	9.34	0.26	9.36	0.14	9.38	0.08
15	9.33	9.32	0.1	9.33	0.05	9.33	0.04	9.33	0.1	9.31	0.14	9.29	0.42	9.32	0.05	9.29	0.35	9.29	0.36	9.31	0.2
16	9.35	9.33	0.16	9.34	0.14	9.34	0.09	9.32	0.3	9.3	0.56	9.33	0.15	9.38	0.31	9.33	0.16	9.35	0.05	9.35	0
17	9.34	9.33	0.13	9.31	0.34	9.3	0.38	9.3	0.43	9.27	0.69	9.35	0.1	9.37	0.32	9.34	0.05	9.33	0.05	9.36	0.28
18	9.29	9.27	0.22	9.27	0.21	9.27	0.26	9.28	0.15	9.26	0.33	9.29	0.01	9.32	0.3	9.29	0.02	9.29	0.02	9.31	0.16
mean			0.15		0.19		0.21		0.21		0.64		0.24		0.36		0.18		0.14		0.21
st.dev.			0.04		0.13		0.14		0.11		0.25		0.16		0.23		0.17		0.16		0.13

TABLE 4.5: Expected residual lifetime truncated after 10 years for age 60 (each couple of columns refers to a model: on the left one the estimated index and on the right one the percentage error with respect to the observed value)

REGION	Observed	P-double		P-common		P-simple		P-division		P-one		M-double		M-common		M-simple		M-division		M-one	
1	8.36	8.35	0.11	8.34	0.26	8.35	0.15	8.32	0.46	8.21	1.84	8.29	0.84	8.3	0.77	8.3	0.77	8.3	0.74	8.27	1.1
2	8.23	8.21	0.21	8.24	0.09	8.19	0.51	8.25	0.19	8.11	1.48	8.28	0.62	8.31	0.99	8.21	0.27	8.26	0.31	8.23	0.07
3	8.42	8.41	0.03	8.42	0.09	8.43	0.21	8.41	0.04	8.27	1.68	8.38	0.46	8.41	0.11	8.39	0.27	8.42	0.01	8.36	0.61
4	8.34	8.33	0.11	8.3	0.54	8.32	0.21	8.33	0.18	8.2	1.72	8.31	0.33	8.35	0.12	8.32	0.26	8.32	0.27	8.28	0.78
5	8.26	8.25	0.17	8.18	0.99	8.18	0.93	8.23	0.38	8.1	1.88	8.24	0.19	8.31	0.65	8.27	0.17	8.23	0.35	8.23	0.38
6	8.39	8.38	0.11	8.39	0.05	8.38	0.07	8.35	0.47	8.23	1.9	8.28	1.33	8.34	0.59	8.28	1.31	8.34	0.64	8.3	1.06
7	8.48	8.48	0.07	8.46	0.32	8.47	0.19	8.46	0.24	8.36	1.44	8.48	0.1	8.43	0.59	8.35	1.62	8.34	1.63	8.44	0.57
8	8.42	8.41	0.17	8.4	0.2	8.42	0.01	8.42	0.06	8.33	1.09	8.31	1.31	8.4	0.21	8.31	1.33	8.31	1.35	8.39	0.37
9	8.46	8.45	0.09	8.43	0.38	8.43	0.32	8.49	0.29	8.39	0.79	8.32	1.65	8.39	0.82	8.29	2	8.3	1.95	8.38	1
10	8.55	8.53	0.15	8.51	0.37	8.51	0.41	8.53	0.17	8.46	1	8.45	1.18	8.51	0.45	8.45	1.11	8.49	0.62	8.49	0.6
11	8.28	8.26	0.23	8.25	0.38	8.25	0.36	8.24	0.49	8.18	1.27	8.22	0.78	8.24	0.58	8.22	0.76	8.16	1.49	8.21	0.95
12	8.51	8.52	0.03	8.51	0.04	8.52	0.03	8.46	0.62	8.4	1.37	8.43	0.97	8.5	0.21	8.43	1	8.49	0.28	8.46	0.62
13	8.1	8.08	0.23	8.07	0.4	8.08	0.17	8.08	0.28	8.02	1.01	8.09	0.1	8.15	0.63	8.09	0.07	8.1	0.06	8.13	0.42
14	8.4	8.39	0.2	8.39	0.21	8.38	0.3	8.4	0.05	8.33	0.91	8.35	0.59	8.42	0.17	8.32	1	8.38	0.24	8.38	0.32
15	8.38	8.36	0.2	8.44	0.71	8.45	0.83	8.46	0.91	8.4	0.24	8.42	0.49	8.48	1.16	8.42	0.44	8.42	0.46	8.49	1.32
16	8.41	8.39	0.23	8.39	0.15	8.4	0.09	8.36	0.5	8.31	1.15	8.33	0.88	8.43	0.24	8.33	0.91	8.38	0.35	8.39	0.23
17	8.32	8.3	0.21	8.29	0.31	8.29	0.31	8.29	0.4	8.22	1.16	8.26	0.68	8.32	0.08	8.27	0.55	8.27	0.57	8.32	0.04
18	8.31	8.29	0.25	8.29	0.25	8.28	0.3	8.31	0.05	8.27	0.5	8.36	0.59	8.41	1.16	8.36	0.62	8.36	0.64	8.4	1.06
mean			0.16		0.32		0.3		0.32		1.25		0.73		0.53		0.8		0.66		0.64
st.dev.			0.07		0.24		0.25		0.23		0.47		0.44		0.36		0.53		0.56		0.38

TABLE 4.6: Expected residual lifetime truncated after 10 years for age 70 (each couple of columns refers to a model: on the left one the estimated index and on the right one the percentage error with respect to the observed value)

REGION	Observed	P-double		P-common		P-simple		P-division		P-one		M-double		M-common		M-simple		M-division		M-one	
1	6.36	6.35	0.16	6.31	0.81	6.34	0.27	6.3	0.93	6.14	3.45	6.18	2.77	6.12	3.7	6.21	2.31	6.23	2.05	6.19	2.63
2	6.29	6.28	0.11	6.29	0.12	6.2	1.31	6.3	0.22	6.11	2.88	6.17	1.9	6.21	1.2	6.06	3.64	6.06	3.54	6.1	3.02
3	6.51	6.49	0.33	6.47	0.55	6.51	0.03	6.48	0.48	6.28	3.51	6.21	4.64	6.38	1.92	6.27	3.73	6.4	1.66	6.27	3.67
4	6.45	6.45	0.04	6.44	0.05	6.45	0.01	6.44	0.19	6.25	3.13	6.24	3.29	6.28	2.63	6.31	2.21	6.3	2.34	6.25	3.12
5	6.27	6.24	0.45	6.19	1.29	6.22	0.79	6.26	0.12	6.09	2.89	6.19	1.25	6.26	0.1	6.23	0.7	6.18	1.49	6.12	2.44
6	6.4	6.38	0.16	6.4	0.13	6.39	0.04	6.35	0.78	6.19	3.25	6.33	0.95	6.33	1.07	6.32	1.11	6.44	0.74	6.35	0.65
7	6.51	6.5	0.24	6.52	0.11	6.55	0.56	6.54	0.47	6.39	1.84	6.35	2.44	6.31	3.09	6.19	4.88	6.19	4.93	6.35	2.47
8	6.41	6.38	0.57	6.39	0.3	6.4	0.28	6.39	0.38	6.27	2.18	6.24	2.73	6.37	0.66	6.2	3.27	6.2	3.26	6.3	1.77
9	6.35	6.3	0.79	6.26	1.51	6.26	1.55	6.32	0.46	6.21	2.23	6.25	1.62	6.33	0.4	6.21	2.23	6.21	2.23	6.32	0.53
10	6.47	6.44	0.5	6.45	0.26	6.46	0.13	6.49	0.36	6.39	1.16	6.28	2.88	6.35	1.75	6.3	2.63	6.35	1.8	6.38	1.35
11	6.26	6.23	0.37	6.25	0.14	6.25	0.16	6.23	0.47	6.13	1.93	6.21	0.8	6.23	0.37	6.21	0.66	6.19	1.07	6.2	0.83
12	6.54	6.51	0.38	6.47	1.03	6.49	0.71	6.42	1.88	6.34	3.1	6.33	3.22	6.46	1.23	6.35	2.96	6.46	1.26	6.46	1.14
13	6.14	6.11	0.35	6.07	1.07	6.12	0.34	6.1	0.6	6.01	2.11	5.99	2.45	6.04	1.49	5.99	2.37	6.01	2.02	6.06	1.23
14	6.31	6.28	0.51	6.29	0.3	6.3	0.27	6.33	0.29	6.23	1.28	6.17	2.23	6.23	1.26	6.11	3.24	6.09	3.49	6.24	1.16
15	6.49	6.47	0.32	6.29	3.01	6.31	2.78	6.32	2.57	6.26	3.53	6.32	2.54	6.42	1.05	6.32	2.62	6.37	1.79	6.28	3.23
16	6.36	6.31	0.68	6.32	0.53	6.34	0.24	6.29	1.04	6.23	2.06	6.22	2.13	6.36	0.02	6.22	2.13	6.29	1	6.32	0.63
17	6.12	6.07	0.74	6.14	0.36	6.13	0.26	6.11	0.04	6.04	1.33	6.04	1.25	6.12	0.03	6.07	0.74	6.04	1.18	6.14	0.46
18	6.47	6.46	0.2	6.42	0.79	6.42	0.85	6.47	0.01	6.41	0.98	6.43	0.66	6.52	0.66	6.43	0.68	6.49	0.24	6.52	0.79
mean			0.38		0.69		0.59		0.63		2.38		2.21		1.26		2.34		2.01		1.73
st.dev.			0.22		0.73		0.7		0.66		0.86		1.02		1.05		1.21		1.17		1.08

TABLE 4.7: Expected residual lifetime truncated after 10 years for age 80 (each couple of columns refers to a model: on the left one the estimated index and on the right one the percentage error with respect to the observed value)

the worst in the class of models with the Poisson error structure, due to the very high number of parameters.

The average MAPE for both in-sample fit and out-of-sample forecast for this model is the best across all models, with a relatively low dispersion, and a similar conclusion holds when looking at annuity values. The residual plots suggest that this model can adequately describe the data, although it cannot catch the variability across all the ages. No systematic pattern with respect to year is apparent, although for some of the considered populations there seems to be the evidence of a cohort effect which is not explained by the model.

4.5.2 P-common model

This is the second model in terms of number of parameters. The reduction of the total likelihood makes this model preferable to both the P-double and the other nested models, according to likelihood ratio test, with a test value much higher than the critical value.

The AIC and its corrected version single out this model as the best one, while the BIC ranks this third due to the heavy penalization for the number of parameters. In-sample and out-of-sample performance are only slightly worse than the P-double model, while annuity values cannot be reproduced well, as percentage errors are the highest among Poisson models.

The residual plots has no evidence of systematic deviations, although residuals are higher due to the reduction of the complexity of the parametric structure. Some evidence of an unexplained cohort effect is present for this model too.

4.5.3 P-simple model

This is the first model with just one bilinear component and can be seen as a sort of benchmark since it is essentially a Lee-Carter model applied to each subpopulation.

The likelihood ratio test indicates that this model is preferable over its special cases. Regarding the penalised log-likelihood indices, P-simple is the fourth ranked

model out of five (third if the corrected version of AIC is used). In-sample and out-of-sample performance distinctly worsen when one of the two bilinear components is dropped. This is less evident when considering the residual lifetime which is, on average, similar to P-double for ages 60 and 70 and even better for age 80.

The standardised residuals plots show more dispersion with respect to the models with two bilinear components: clearly, this model is simpler and cannot capture the variance as well as more complex model can do.

4.5.4 P-division model

This model is strongly influenced by the choice of the partition of the set of populations. With the 5 considered groups, this model has a reduced number of parameters but a high likelihood value. In fact, this model is indicated as the second best choice by the adjusted AIC and the BIC, better than the model P-simple.

Looking at forecast values, this model has more or less the same MAPE value of P-simple, with reduced variance. The standardised residuals display a significant variability on the quality of the results. In fact, the models capture the mortality shape of some populations much better than others.

4.5.5 P-one model

The P-one model, with his single time varying coefficient, has the smallest number of parameters among the Poisson error models. The decrease in likelihood is not counterbalanced by a reduction in the number of parameters, since the likelihood ratio test rejects this model with a large difference between the test value and the critical value.

Unsurprisingly, the BIC index shows this as the preferred model, while it is the worst one, and second worst according to AIC and its corrected version. This performance reduction is not strong if it is considered the reconstruction of the theoretical data. On the contrary, this model does not perform well with respect to forecast data compared to the other Poisson based models, considering both the MAPE and truncated expected residual lifetime.

As expected, the residuals show the weaknesses of such a simple model in capturing specific mortality behaviour.

4.5.6 M-double model

M-double is the first model applied to mortality improvement rates. The number of parameters is the maximum within M-models, but it is strongly reduced compared to P-double. This model is indicated as the preferable one considering AIC and its adjusted version, and the worst one accordingly to BIC.

Regarding the fit of the observed data, this model does not perform well compared to the ones commented above. This is due to the fact that the historical data obtained here are reconstructed starting by the last observation, therefore the historical shape could be too much smoothed. Regarding the forecast data, the same considerations are valid. The values of truncated expected residual lifetime are better than in P-one model, the less accurate models between the Poisson structured data.

The graphical analysis of the residuals suggest that M-double is able to capture the mortality shape without significant systematic deviations.

4.5.7 M-common model

M-common model differs in this analysis from M-double due a reduction in the number of parameters and in likelihood levels that seems not to reduce the quality of the forecast.

This model is the second best according to AIC (and its adjusted version) and the best for BIC. Considering the penalised log-likelihood indices, this model would be the best choice. This model gives back the best result in the study of theoretical data and has the more accurate actuarial indices, between the models applied to mortality improvement rates. M-double is better than M-common only in the MAPE of forecast data and for one actuarial index out of the three computed here.

The results of graphical analysis do not provide evidence of significant systematic behaviours.

4.5.8 M-simple model

M-simple seems to be the second best choice, after M-common, considering the penalised log-likelihood. In fact this model is second or third ranked in AIC and BIC. The MAPE of the forecast data is only slightly higher than MAPE for M-common. The drawback of this model is higher in the reconstruction of historical data. Considering the computed actuarial indices M-simple is not particularly accurate compared with the other models.

Observing the graphical representation of standardised residuals, for almost all the populations this model seems not to create systematic errors.

4.5.9 M-division model

M-division is ranked in fourth position with respect to the penalised log-likelihood indices, and this is a worse performance compared to its counterpart, P-division.

The results of this model are not good if the fitted data or the forecast data are observed. In contrast, this model has a quality of the actuarial indices which is at the same level of the others and for one of the observed ages (60 years) the truncated expected residual lifetime is the best of all the models.

The results of graphical analysis shows some systematic behaviours of standardised residuals, especially with respect to age and years.

4.5.10 M-one model

M-one has the smallest number of parameters of the models considered, but this simplification leads to a drawback in terms of the likelihood level, and therefore the AIC rank this as the fifth model and BIC the third among the mortality improvement rate versions.

The MAPE of the theoretical and forecast errors are similar to the ones of the other mortality improvement rate models. The computed actuarial indices are quite accurate: Mone is the second-third best model in this analysis.

The graphical analysis shows again some systematic behaviours of standardised residuals, especially with respect to age and years and the presence of more outlier values.

4.6 Conclusions

From the indices computed on the models, the best performance is observed in the ones with Poisson error structure with higher number of parameters. A good choice could be the P-common which has a reduced number of parameters, good level of likelihood, not significant behaviour of the errors and predictive capacity.

Reducing the number of parameters in the Poisson error structure models, it is possible to observe that there is a drawback in the performance of the models. The drawback is less evident in the models applied to mortality improvement rates. The overall performance of such models is comparable to the Poisson ones.

Summing up, it is possible to conclude that, in a study where the predictive capacity is the aim, P-common model is the best choice between the proposed models. A good alternative could be M-division or M-one when the aim is to consider

- a simpler version of the model;
- a stronger common structure;
- it is preferable not considering the number of exposure to risk (sometimes this data are not available).

The standardised residuals with respect to age and year for the 18 populations are reported in Figures 4.5 and 4.6 for P-common model and in Figures 4.7 and 4.8 for M-one model.

Some further considerations are presented below.

Small different populations are considered

One important aspect of the application which should be recalled is that the selected populations are 18 regions within a country. These regions are far from being homogeneous in the exposure to risk amount.

All of the considered populations can be considered 'small populations': this means that a significant variability of the parameters cannot be avoided. However, the dimension of the populations is taken into account during the estimation procedure. This is true for P-models since in the log-likelihood function which is optimised there is the number of exposure to risk and the number of deaths. In M-models this is due to the variance of the normal distribution: that value is usually higher for smaller populations, therefore the influence of the small populations is lower with respect to the larger ones.

Parsimony

The penalised log-likelihood indices indicate as preferable, most of the time, the model which is the best in terms of likelihood value or the most parsimonious (in the number of parameters). In fact the preferable models according to AIC are the most complex models. On the other hand, observing BIC, the preferable model is usually the simpler. The adjusted version of AIC it seems to be in the middle between these two criteria.

Starting by these considerations, when a model does not follows these rules and have better rank than expected, it is likely a model with remarkable characteristics. Such a model is probably the best choice considering at the same time approximation and parsimony.

The common models, *i.e.* P-common and M-common, seem to be the preferable model designs considering both AIC and BIC.

Not only one best model

In the analysis performed above there are some graphical analysis and some indices designed for understanding the quality of the forecast, and others for understanding if the model can capture the historical behaviour of the phenomenon.

Based on the results, there is not a model which is the best in all of these analysis. Therefore, the choice between the approach should be done taking in account the strength and weakness of the models, according to the purpose for which the chosen model should be used.

The risk of a wrong partition

As highlighted by the indices, models P-division and M-divison could be a good choice for forecasting mortality.

However, further attention should be paid to choosing the groups, and to exploring different ways for classifying the populations. This is evident from the graphical analysis, where it can be seen that the mortality behaviour of some populations is explained by the model much better than for other populations. Such a phenomenon could be justified as differences in population volatility or with a non efficient partition.

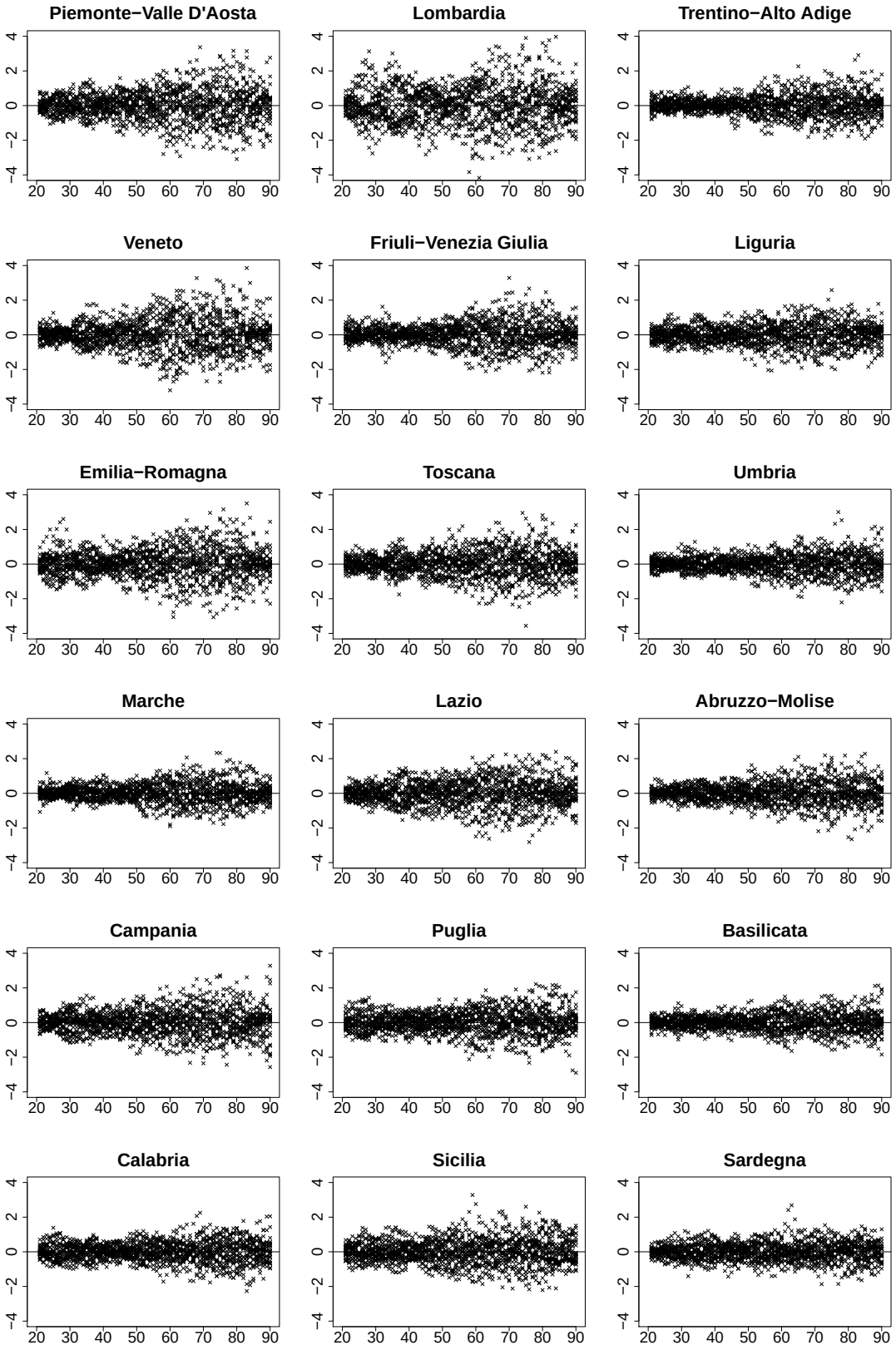

FIGURE 4.5: Standardised residuals with respect to age for the 18 populations - Model P-common

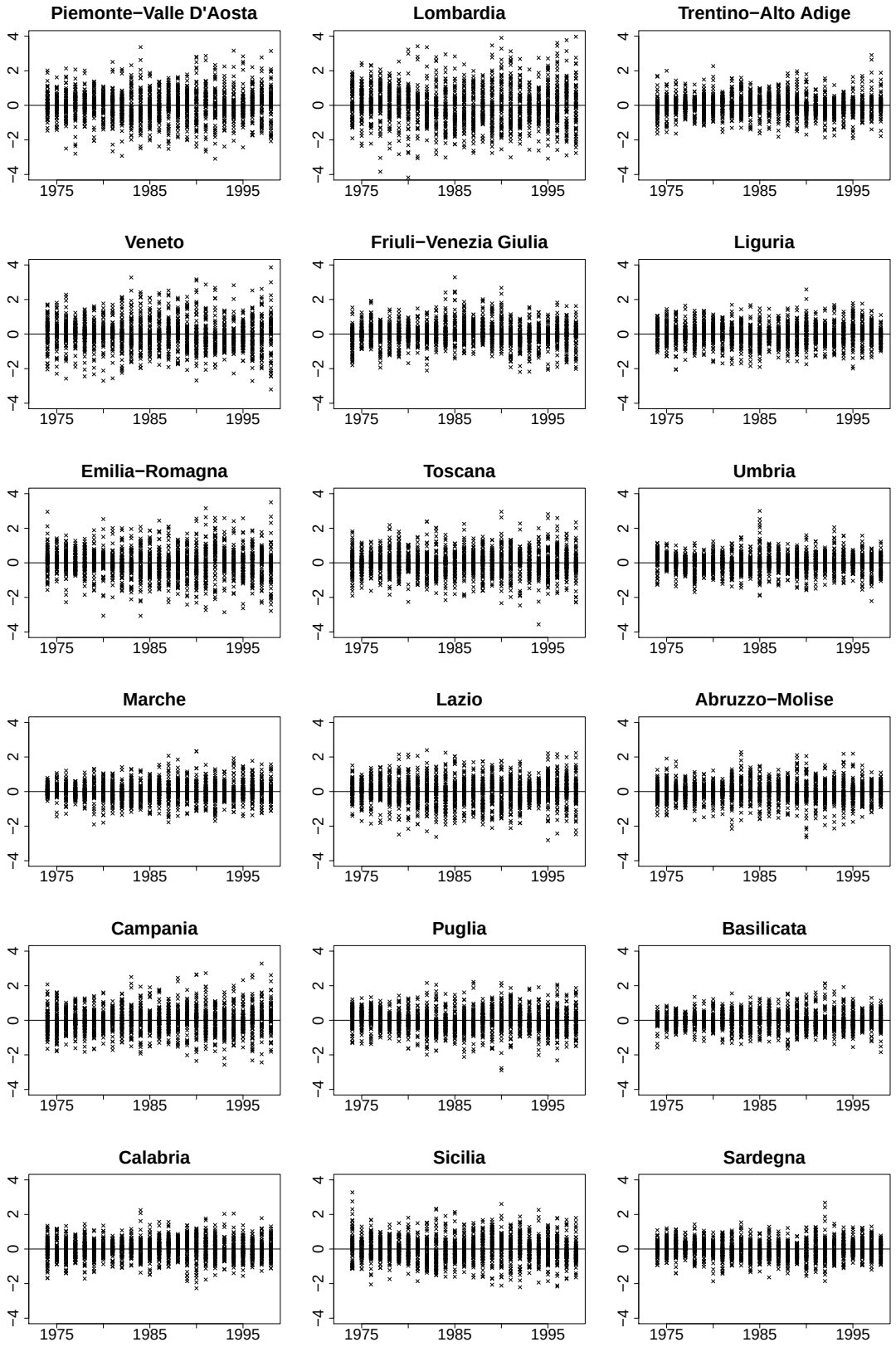

FIGURE 4.6: Standardised residuals with respect to year for the 18 populations - Model P-common

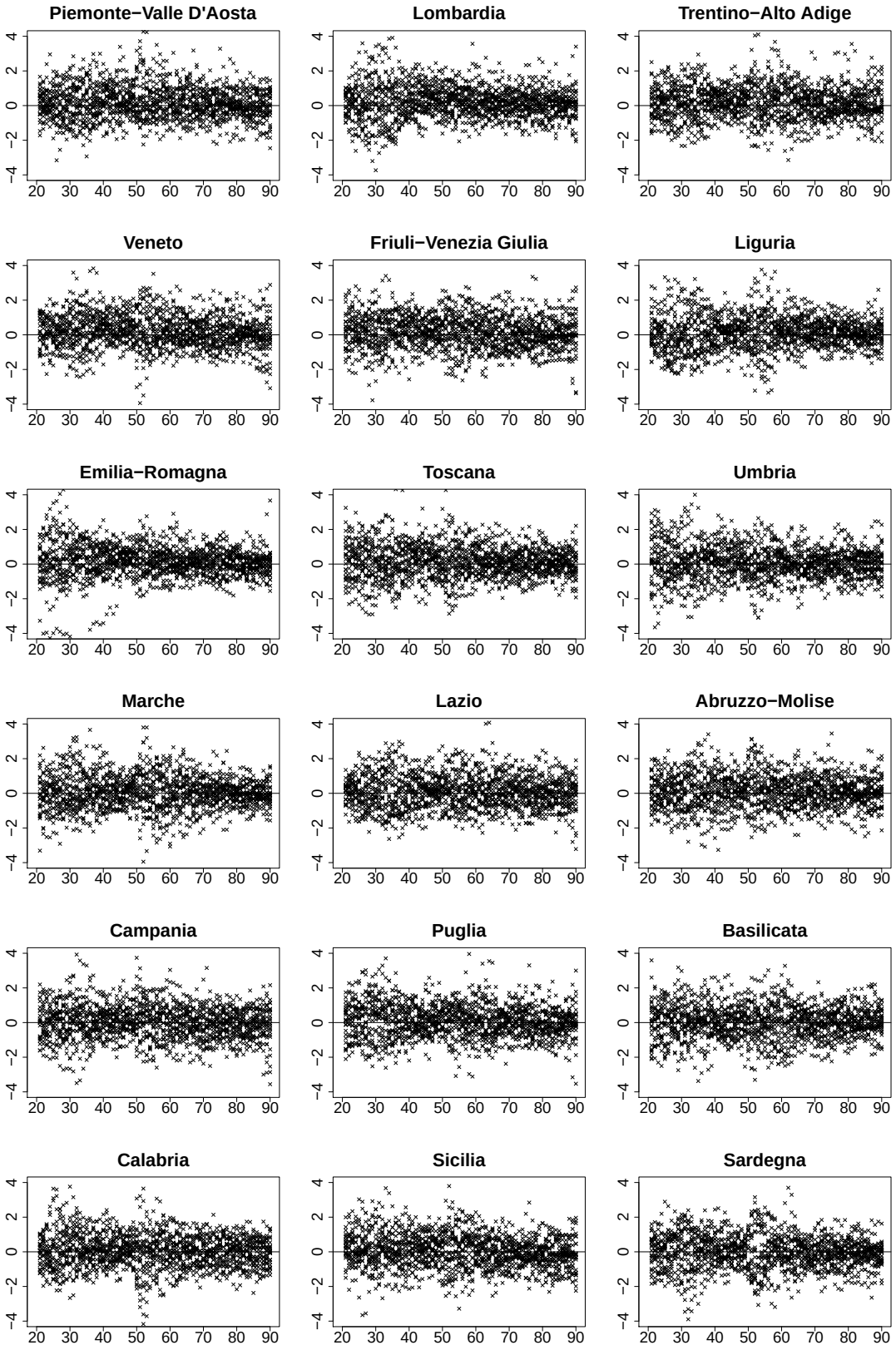

FIGURE 4.7: Standardised residuals with respect to age for the 18 populations - Model M-one

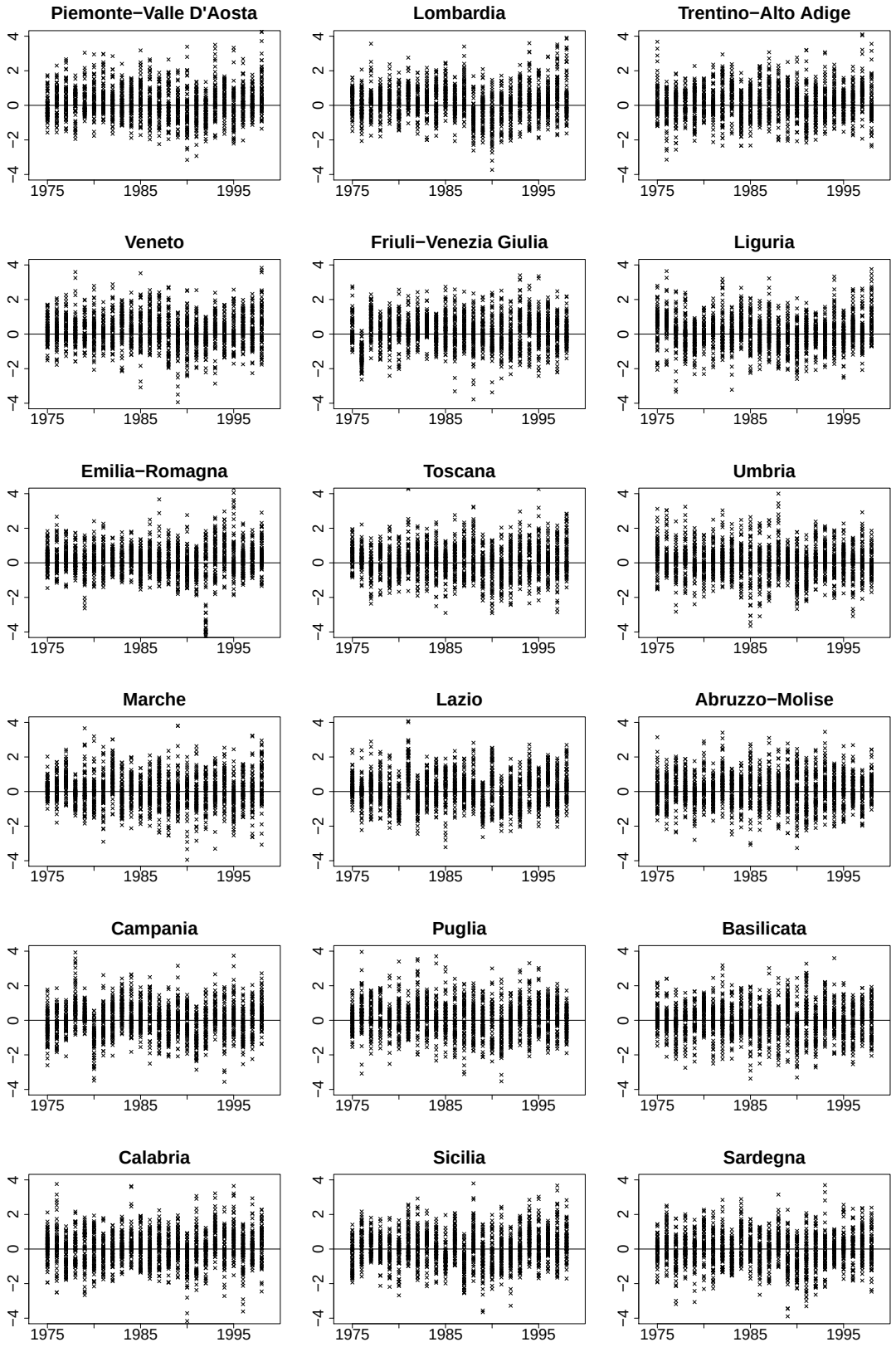

FIGURE 4.8: Standardised residuals with respect to year for the 18 populations - Model M-one

Chapter 5

Forecasting mortality improvement rates for related populations with non-costant variance

5.1 Introduction

Some of the models considered in the previous chapters target mortality improvement rates. This parametric structure is based on the strong assumption that the variance of the mortality improvement rates is constant. In this chapter this assumption is dropped.

In Section 5.2 the new parametric structure is presented. The modified models are defined in Section 5.3. The estimation procedure which is needed for the new formulation is outlined in Section 5.4. Section 5.5 reports the tools and the results of an application performed on Nordic Countries mortality data. In Section 5.6 there are some concluding remarks.

5.2 The parametric structure

In Section 3.3.3 the mortality improvement rates $z_{x,t}$ are modelled as realizations of independent Gaussian random variables $Z_{x,t}$ assuming constant dispersion, hence

$$Z^i_{x,t} \sim N(\eta^i_{x,t}, \sigma^2_i), \tag{5.1}$$

with variance σ^2_i that is only population-dependent. The hypothesis of constant dispersion could be restrictive. As a matter of fact the variance of $Z^i_{x,t}$ is not constant over ages and years.

The heterogeneity of the variance in (5.1) is introduced by the weights $\phi^i_{x,t}$ (Haberman and Renshaw, 2012). Such weights modify the population-specific variance σ^2_i allowing it to vary with ages and years. These weights are population-specific. The model specified in (5.1) becomes

$$Z^i_{x,t} \sim N(\eta^i_{x,t}, \phi^i_{x,t}\sigma^2_i). \tag{5.2}$$

The first moment predictor structure is still specified by $\eta^i_{x,t} = \beta^i_x k^i_t$, or by its generalised version

$$\eta^i_{x,t} = \sum_{j=1}^{L} \beta^i_{x,j} k_{t,j} \tag{5.3}$$

as in (3.5). Like in Chapter 3, some special cases of (5.3) are considered. The aim here is to stress the impact of $\phi^i_{x,t}$ on (5.2).

In the models described in the following section one bilinear component is considered, hence $L = 1$ in (5.3).

5.3 The selected models

5.3.1 Constant variance

The first two models are defined assuming $\phi^i_{x,t} = 1$ for every x, t and i. It follows that (5.2) becomes equal to (5.1).

Two models are selected by (5.3).

1. M-simple:
$$\eta^i_{x,t} = \beta^i_x k^i_t$$
with the identifiability constraint $\sum_x \beta^i_x = 1$ for all i.

2. M-one:
$$\eta^i_{x,t} = \beta^i_x k_t$$
with the identifiability constraint $\sum_{i,x} \beta^i_x = I$.

These two models are analogous to the ones defined in Chapter 3. The first moment predictor structure do not change also in the following four models presented below.

5.3.2 Population-age-specific variance

The third and the fourth models are defined by setting $\phi^i_{x,t}$ equal to ϕ^i_x. It means that the variance σ^2_i is assumed constant across time. This implies that (5.2) becomes
$$Z^i_{x,t} \sim N(\eta^i_{x,t}, \phi^i_x \sigma^2_i).$$
The first moment predictor structure is not changed with respect to the previous case, then the following two models are defined.

3. M-ϕ-simple:
$$\eta^i_{x,t} = \beta^i_x k^i_t$$
with the identifiability constraint $\sum_x \beta^i_x = 1$ for all i.

4. M-ϕ-one:
$$\eta^i_{x,t} = \beta^i_x k_t$$
with the identifiability constraint $\sum_{i,x} \beta^i_x = I$.

5.3.3 Age-specific variance

The fifth and the sixth models are obtained by a further restriction on $\phi^i_{x,t}$. In fact these coefficients are set equal to ϕ_x. It follows that the variances $\phi_x \sigma^2_i$ for $i = 1, \ldots, I$ are

- different for each population;
- constant across time;
- have the same age-dependency structure.

Formula (5.2) becomes

$$Z^i_{x,t} \sim N(\eta^i_{x,t}, \phi_x \sigma_i^2)$$

then the following two models are defined (cϕ is for 'common ϕ').

5. M-cϕ-simple:

$$\eta^i_{x,t} = \beta^i_x k^i_t$$

with the identifiability constraint $\sum_x \beta^i_x = 1$ for all i.

6. M-cϕ-one:

$$\eta^i_{x,t} = \beta^i_x k_t$$

with the identifiability constraint $\sum_{i,x} \beta^i_x = I$.

5.3.4 The philosophy of the models

The six models are designed in order to be very similar to each others, and differ only for a single detail. This is because they are oriented to evaluate whether the introduction of the parameter ϕ_x improves the ability of the model to fit the data. This parameter is clearly age-dependent. When more than one related population is considered, the differences in the nature of the parameter ϕ^i_x need to be explored.

Regarding the first and the second models, they are here reported in order to create the benchmark with respect to compare the models with non constant variance. Among the models proposed in Chapter 3, the models M-simple and M-one are considered relevant since

- M-simple is the mortality improvement rate counterpart of the classical Lee-Carter model;
- in Chapter 4, M-one is indicated as a good choice in presence of multiple populations data, due to its parsimony and good performance.

The third and the fourth models are the first two models with a different variance of $Z^i_{x,t}$ distribution. In fact, this becomes equal to $\phi^i_x \sigma^2_i$. The meaning of this is to allow the variance varying with respect to the ages and populations.

The fifth and the sixth models are defined in the middle of the other two perspectives. In fact, the variance of the distribution of $Z^i_{x,t}$ is not constant. However, the general level of variance defined for every population is modified with respect to the vector of parameters ϕ_x, which is in common for all the populations.

5.4 The estimation procedure

Regarding the first and the second models, the estimation procedure is equivalent to the one presented in Chapter 4. It consists in maximising the log-likelihood function

$$\ell = -\frac{1}{2} \sum_i \sum_x \sum_t \left\{ \log(2\pi \hat{\phi}^i_{x,t} \hat{\sigma}^2_i) + \frac{(z^i_{x,t} - \hat{\eta}^i_{x,t})^2}{\hat{\phi}^i_{x,t} \hat{\sigma}^2_i} \right\} \tag{5.4}$$

where $\hat{\phi}^i_{x,t}$ is equal to 1 for every x, t and i.

For the other models, a two-stages iterative estimation procedure is adopted, as presented in Haberman and Renshaw (2012). Initially the weights $\hat{\phi}^i_{x,t}$ are set equal to 1.

The first stage consists in maximising (5.4).

The second stage consists in obtaining the values of $\hat{\phi}^i_{x,t}$. This is done by assuming that the squared residuals

$$r^2_{x,t,i} = (z^i_{x,t} - \hat{\eta}^i_{x,t})^2$$

are realisations of independent gamma random variables. In practice, these values are computed by minimising the model deviance

$$Dev = 2 \sum_{x,t,i} \left[\frac{r^2_{x,t,i} - \hat{\phi}^i_{x,t}}{\hat{\phi}^i_{x,t}} - \log \left(\frac{r^2_{x,t,i}}{\hat{\phi}^i_{x,t}} \right) \right]$$

under the condition that $\hat{\phi}^i_{x,t} > 0$.

These two stages are iterated until the variations of the log-likelihood and deviance functions into successive iterations are below a given threshold.

5.5 Application to Nordic Countries mortality data

5.5.1 The dataset

A different dataset than the one used in Chapter 4 is now considered. The data are now the mortality rates of four Nordic Countries (in Figure 5.1): Denmark (DK), Norway (N), Sweden (SE) and Finland (FI)[1]. It follows that $I = 4$, and $i = 1, \ldots, 4$. These four countries share common traits in their respective societies, therefore can be considered as related populations.

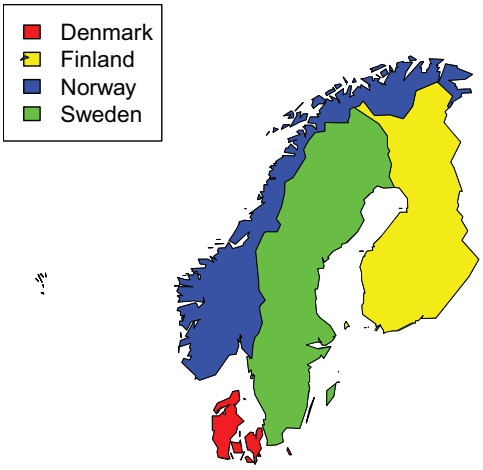

FIGURE 5.1: Nordic Countries

The models defined in Section 5.3 are applied to female mortality data for the ages 20-89 and for the years 1965-1994. The data for the years 1995-2009 are used to check the quality of the forecast.

The time series used for the forecast are those considered in Chapter 3.

5.5.2 The results

Goodness of fit

Three indicators for the goodness of fit based on the penalised log-likelihood indices are considered, described in Section 4.4.1. The difference with the formulas in

[1]The data were downloaded from Human Mortality Database (`www.mortality.org`).

Chapter 4 are in the log-likelihood function, that take in account the weights $\hat{\phi}^i_{x,t}$, *i.e.* (5.4). The values of the penalised log-likelihood indices are summarised in Table 5.1.

	M-simple	M-one	M-ϕ-simple	M-ϕ-one	M-cϕ-simple	M-cϕ-one
d	392	308	392	308	392	308
ℓ	10381	10271	14044	13794	13542	13357
g	8120	8120	8120	8120	8120	8120
AIC	-19978	-19926	-27304	-26972	-26300	-26098
Δ-AIC	7326	7378	0	332	1003	1206
rank-AIC	5	6	1	2	3	4
g/d	21	26	21	26	21	26
AIC_c	-19938	-19902	-27264	-26947	-26261	-26074
Δ-AIC_c	7326	7362	0	317	1003	1191
rank-AIC_c	5	6	1	2	3	4
BIC	-17233	-17770	-24559	-24815	-23556	-23941
Δ-BIC	7582	7046	256	0	1259	874
rank-BIC	6	5	2	1	4	3

TABLE 5.1: Dimension of the parametrised prediction structure (d), likelihood of the model (ℓ), dimension of the data (g), value of g/d, AIC, AIC_c and BIC (and its Δ and its rank) of the ten models (when applicable, the values are rounded to the integer)

The scaled residuals are obtained considering the values of $\hat{\phi}^i_{x,t}$, and are now defined as

$$r^i_{x,t} = \frac{z^i_{x,t} - \hat{\eta}^i_{x,t}}{\sqrt{\hat{\sigma}^2_i \hat{\phi}^i_{x,t}}}, \text{ with } \hat{\sigma_i}^2 = \sum_{x,t} \frac{(z^i_{x,t} - \hat{\eta}^i_{x,t})^2}{\hat{\phi}^i_{x,t} \nu}$$

where ν is the size of the dataset. For the Finnish population, the residual plots with respect to age, year and cohort of the six models are presented in 5.2 (the others are not reported here).

Forecast

The quality of the forecasts is evaluated using the mean absolute percentage error and an actuarial index, the truncated expected residual lifetime. These two values are used in the form presented in Chapter 4.

The MAPE is applied to the forecast computed using the six models with a time horizon of 15 years and compared with the observed data. The values of MAPE for every model and for every population are shown in Table 5.2.

The truncated expected residual lifetime is computed for all the populations for ages 55, 65 and 75. This index is obtained considering a time horizon of 15 years. The results are summarised in Table 5.3.

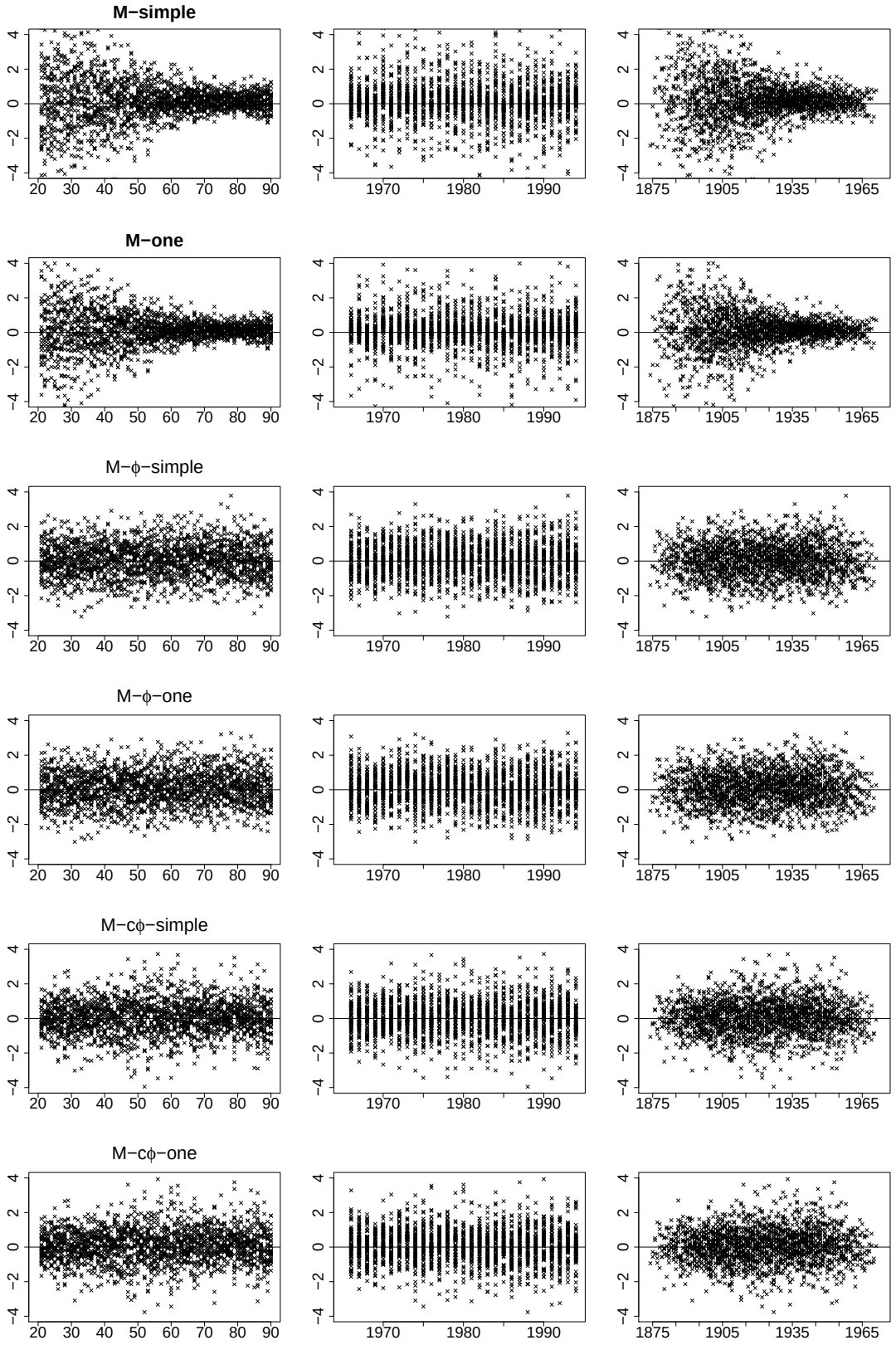

FIGURE 5.2: Age, year and cohort residual plots - population (2), Finland

POP	M-simple	M-one	M-ϕ-simple	M-ϕ-one	M-cϕ-simple	M-cϕ-one
DK	29.74	29.34	29.07	30.71	29.07	30.8
FI	26.39	26.35	25.11	24.9	25.2	24.89
N	23.27	23.3	23.82	23.62	23.44	23.63
SE	23.73	23.52	21.51	21.36	21.62	21.51
mean	25.78	25.63	24.87	25.15	24.83	25.21
st. dev.	2.58	2.46	2.74	3.45	2.76	3.45

TABLE 5.2: MAPE of forecast data with respect to observed data

	Obs	M-simple		M-one		M-ϕ-simple		M-ϕ-one		M-cϕ-simple		M-cϕ-one	
$e_{55:\overline{15}\rceil}$													
DK	13.97	13.95	0.15	13.95	0.13	13.97	0.01	13.97	0.02	13.97	0.02	13.96	0.05
FI	14.39	14.38	0.05	14.38	0.05	14.42	0.21	14.41	0.13	14.42	0.19	14.41	0.11
N	14.32	14.31	0.05	14.31	0.06	14.32	0.04	14.33	0.08	14.32	0.02	14.33	0.06
SE	14.36	14.35	0.03	14.35	0.03	14.38	0.14	14.37	0.09	14.38	0.13	14.37	0.08
mean			0.07		0.07		0.1		0.08		0.09		0.07
st.dev.			0.05		0.04		0.09		0.05		0.08		0.03
$e_{65:\overline{15}\rceil}$													
DK	12.62	12.6	0.12	12.61	0.06	12.66	0.38	12.67	0.46	12.66	0.33	12.67	0.4
FI	13.23	13.19	0.32	13.19	0.29	13.32	0.65	13.28	0.4	13.3	0.53	13.28	0.36
N	13.29	13.27	0.16	13.27	0.18	13.31	0.19	13.32	0.26	13.3	0.06	13.32	0.23
SE	13.42	13.41	0.12	13.41	0.1	13.5	0.58	13.49	0.46	13.5	0.55	13.47	0.38
mean			0.18		0.16		0.45		0.4		0.37		0.34
st.dev.			0.1		0.1		0.21		0.09		0.23		0.08
$e_{75:\overline{15}\rceil}$													
DK	9.95	9.94	0.18	9.94	0.1	10.05	0.99	10.08	1.24	10.04	0.87	10.06	1.07
FI	10.02	9.98	0.41	9.98	0.39	10.27	2.51	10.18	1.58	10.23	2.08	10.17	1.49
N	10.45	10.42	0.24	10.42	0.31	10.52	0.7	10.54	0.89	10.48	0.25	10.53	0.75
SE	10.66	10.64	0.13	10.65	0.07	10.84	1.74	10.8	1.39	10.83	1.67	10.78	1.18
mean			0.24		0.22		1.49		1.27		1.22		1.12
st.dev.			0.12		0.16		0.81		0.29		0.82		0.31

TABLE 5.3: Expected residual lifetimes truncated after 15 years for age 55, 65 and 75 (each couple of columns refers to a model: on the left one the estimated index and on the right one the percentage error with respect to the observed value)

5.6 Discussion

The results of the application presented in the previous section both reject and support the introduction of the the weights $\phi^i_{x,t}$. In fact, setting non constant variance for the mortality improvement rates, this do not lead to universally better results.

The penalised log-likelihood indices indicate as preferable the models containing the weights $\phi^i_{x,t}$. In particular, accordingly to Table 5.1 the M-ϕ-simple and M-ϕ-one models are preferable. This is clearly due to the improvements in the log-likelihood value, while the dimension of the first moment predictor structure is unchanged. This last value does not changes because of it is referred to the dimension of the first moment predictor structure (5.3).

Besides observing the likelihood, the higher accuracy of the models with weights $\phi^i_{x,t}$ in fitting the in-sample data can be seen by observing Figure 5.2. The residual plots highlight how the first two models cannot catch the differences in variance of $z^i_{x,t}$, in particular with respect to age and cohort. Conversely, it seems that models with non constant variance do not have significant systematic behaviours in the residuals.

Due to its ability in describing data with non constant variance, models that include weights $\phi^i_{x,t}$ could be applied to data with a larger age interval than 20-89.

Regarding the quality of the forecast, the results need to be observed more critically. The mean of MAPE applied to the forecast have slightly lower values in third to sixth models. It is also true that in these latter models the variance of the MAPE values is higher.

Despite the previous remarks, in the expected residual lifetime the M-simple and M-one models are the more accurate. Observing again the expected residual lifetime, the M-ϕ-simple and M-ϕ-one models, that were considered so far as the best choice, are here the worst.

It seems by these latter commented results that the models catching more accurately the pattern of the data cannot be conveniently used for the forecast. It is also true that the computation of the expected residual lifetimes involves 15 years forecast data at the same time: the errors in the expected residual lifetimes could be attributed to just a few values of the forecast.

Clearly, these results could also follow by the choice of the data, that consist into a small amount of populations with few millions of individuals each.

Chapter 6

Concluding remarks

In this thesis some models are evaluated in order to investigate the mortality phenomena. The aim is to consider a reasonable number of different approaches for facing with multiple population mortality data.

Starting by Lee-Carter model, ten different approaches for studying multiple population mortality data are considered: five applied to central death rates and five to mortality improvement rates. Regarding the models applied to mortality improvement rates, it is then discussed the introduction of a further parameter.

The methods are evaluated by analysing their performance regarding two characteristics: (i) the goodness in approximating the in-sample mortality data and (ii) the ability of anticipating the out-of-sample mortality data. Observing the results of the two applications performed, it seems that it is not possible to identify one model as the best in all the considered analysis. Conversely, it appears appropriate to think this problem as target oriented, thus to choose the approach which can better pursue the goal of the study.

With reference to the applications here considered, the models with Poisson error structure and two bilinear components are probably the most performing approaches. However, the models applied to mortality improvement rates are good alternatives, and allow to consider more parsimonious model structures. In these latter cases, when it is important to have a good approximation to the in-sample data, it is not opportune to introduce the assumption of constant variance of the mortality improvement rates.

Bibliography

Ahcan, A., D. Medved, A. Olivieri, and E. Pitacco (2014). Forecasting mortality for small populations by mixing mortality data. *Insurance: Mathematics and Economics 54*, 12 – 27.

Baxter, S. (2007). Should projections of mortality improvements be subject to a minimum value? *British Actuarial Journal 13*, 375–464.

Biatat, V. D. and I. D. Currie (2010). Joint models for classication and comparison of mortality in different countries. *Proceedings of International Workshop on Statistical Modelling 2010*.

Booth, H. (2006). Demographic forecasting: 1980 to 2005 in review. *International Journal of Forecasting 22*(3), 547–581.

Booth, H., J. Maindonald, and L. Smith (2002a). Age-time interactions in mortality projection: Applying Lee-Carter to Australia. Available at: https://digitalcollections.anu.edu.au/bitstream/1885/41457/4/85.pdf.

Booth, H., J. Maindonald, and L. Smith (2002b). Applying Lee-Carter under conditions of variable mortality decline. *Population Studies 56*(3), 325–336.

Box, G. E. P. and G. M. Jenkins (1976). *Time series analysis : forecasting and control (Revised Edition ed.)*. Prentice Hall, Englewood Cliff, N.J.

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

Burnham, P. and R. Anderson (2004). Multimodel inference: Understanding AIC and BIC in model selection. *Sociological Methods & Research 33*(2), 261–304.

Butt, Z. and S. Haberman (2009). Ilc: A collection of R functions for fitting a class of Lee-Carter mortality models using iterative fitting algorithms. Available at: http://bunhill.city.ac.uk/research/actprojects.nsf/All/DFA6C422FF690625802576C10051EAB5?opendocument.

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

Cairns, A. J. G., D. Blake, K. Dowd, G. D. Coughlan, and M. Khalaf-Allah (2011). Bayesian stochastic mortality modelling for two populations. *ASTIN Bulletin 41*, 29–59.

Carter, L. and R. D. Lee (1992). Modeling and forecasting US sex differentials in mortality. *International Journal of Forecasting 8*, 393–411.

Coale, A. J. and E. E. Kisker (1990). Defects in data on old-age mortality in the United States: new procedures for calculating mortality schedules and life tables at the highest ages. *Asian and Pacific Population Forum 4*, 1–31.

Currie, I. (2009a). Adjusting for bias in mortality forecasts. Available at: http://www.macs.hw.ac.uk/~iain/research/talks/Currie_Longevity_4up.pdf.

Currie, I. (2009b). Basis risk in mortality projections. Available at: http://www.longevitas.co.uk/site/informationmatrix/basisriskinmortalityprojections.html.

Currie, I. (2009c). Forecasting with limited portfolio data. Available at: http://www.longevitas.co.uk/site/informationmatrix/piggybackmodelsorforecastingwithlimitedportfoliodata.html.

Currie, I. (2010). Everything counts in large amounts. Available at: http://www.longevitas.co.uk/site/informationmatrix/everythingcountsinlargeamounts.html.

De Jong, P. and L. Tickle (2006). Extending the Lee-Carter model of mortality projection. *Mathematical Population Studies 13*, 1–18.

Delwarde, A., M. Denuit, and P. Eilers (2007). Smoothing the LeeCarter and Poisson log-bilinear models for mortality forecasting: A penalized log-likelihood approach. *Statistical Modelling 7*(1), 29–48.

Dowd, K., A. Cairns, D. Blake, G. Coughlan, and M. Khalaf-Allah (2011). A gravity model of mortality rates for two related populations. *North American Actuarial Journal 15*(2), 334–356.

Group Annuity Valuation Table Task Force (1995). 1994 Group annuity mortality table and 1994 Group annuity reserving table. Technical report.

Haberman, S. and A. Renshaw (2009). On age-period-cohort mortality rate projections. *Insurance: Mathematics and Economics 45*(2), 255–270.

Haberman, S. and A. Renshaw (2012). Parametric mortality improvement rate modelling and projecting. *Insurance: Mathematics and Economics 50*(3), 309–333.

Haberman, S. and A. Renshaw (2013). Modelling and projecting mortality improvement rates using a cohort perpsective. *Insurance: Mathematics and Economics 53*(1), 150–168.

Hatzopoulos, P. and S. Haberman (2013). Common mortality modeling and coherent forecasts. an empirical analysis of worldwide mortality data. *Insurance: Mathematics and Economics 52*(2), 320–337.

Hyndman, R. and S. M. Ullah (2007). Robust forecasting of mortality and fertility rates: A functional data approach. *Computational Statistics & Data Analysis 51*(10), 4942–4956.

Hyndman, R. J., H. Booth, and F. Yasmeen (2013). Coherent mortality forecasting: The product-ratio method with functional time series models. *Demography 50*(1), 261–283.

Jarner, S. F. and E. M. Kryger (2011). Modelling adult mortality in small populations: The saint model. *ASTIN Bulletin 41*(2), 377–418.

Jarner, S. J. and T. Møller (2013). A partial internal model for longevity risk. Available at: http://www.actuaries.org/lyon2013/papers/LIFE_Jarner_Moeller.pdf.

Lee, R. D. and L. R. Carter (1992). Modelling and forecasting U.S. mortality. *Journal of the American Statistical Association 87*(14), 659–675.

Li, J. S. and M. R. Hardy (2011). Measuring basis risk in longevity hedges. *North American Actuarial Journal 15*(2), 177–200.

Li, N. and R. D. Lee (2005). Coherent mortality forecasts for a group of populations: an extension of the Lee Carter method. *Demography* (42), 575–594.

Lutz, W. (1996). *The future population of the world: what can we assume today? (Revised Edition ed.)*. Earthscan, London.

Minelli, G., V. Manno, S. M. D'Ottavi, M. Masocco, G. Rago, M. Vichi, L. Frova, S. Marchetti, M. Demaria, and S. Conti (2012). La mortalità in Italia nell'anno 2009. *Rapporti ISTISAN* (12/15).

Oeppen, J. and J. W. Vaupel (2002). Broken limits to life expectancy. *Science 296*(5570), 1029–1031.

Olshansky, S. J., D. J. Passaro, R. C. Hershow, J. Layden, B. A. Carnes, J. Brody, L. Hayflick, R. N. Butler, D. B. Allison, and D. S. Ludwig (2005). A potential decline in life expectancy in the United States in the 21st century. *New England Journal of Medicine 352*(11), 1138–1145.

Pitacco, E., M. Denuit, S. Haberman, and A. Olivieri (2009). *Modelling Longevity Dynamics for Pensions and Annnuity Business*. Oxford University Press.

Plat, R. (2009a). On stochastic mortality modeling. *Insurance: Mathematics and Economics 45*(3), 393–404.

Plat, R. (2009b). Stochastic portfolio specific mortality and the quantification of mortality basis risk. *Insurance: Mathematics and Economics 45*(1), 123–132.

Renshaw, A. and S. Haberman (2003a). Lee-Carter mortality forecasting incorporating bivariate time series. Available at: http://openaccess.city.ac.uk/2287/.

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

Renshaw, A. E. and S. Haberman (2003c). Lee-Carter mortality forecasting, a parallel generalized linear modelling approach for England & Wales mortality projections. *Applied Statistics 52*, 119–137.

Renshaw, A. E. and S. Haberman (2003d). On the forecasting of mortality reduction factors. *Insurance: Mathematics and Economics 32*(3), 379–401.

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

Richards, S. J., J. G. Kirkby, and I. D. Currie (2005). The importance of year of birth in two-dimensional mortality data. *British Actuarial Journal 12*(1), 5–61.

Russolillo, M., G. Giordano, and S. Haberman (2011). Extending the LeeCarter model: a three-way decomposition. *Scandinavian Actuarial Journal 2011*(2), 96–117.

Sanderson, W. C. (1998). Knowledge can improve forecasts: A review of selected socioeconomic population projection models. *Population and Development Review 24*, 88–117.

Tabeau, E., A. van den Berg Jeths, and C. Heathcote (Eds.) (2001). *Forecasting mortality in developed countries*. Kluwer Academic Publishers.

United Nations (1982). *Model Life Tables for Developing Countries*. United Nations publication.

Villegas, A. M. and S. Haberman (2014). On the modelling and forecasting of socioeconomic mortality differentials: An application to deprivation and mortality in england. *North American Actuarial Journal*. To appear.

Wan, C., L. Bertschi, and Y. Yang (2013). Coherent mortality forecasting for small populations: an application to Swiss mortality data. Available at: http://www.actuaries.org/lyon2013/papers/AFIR_Wan_Bertschi_Yang.pdf.

Willets, R. C. (1999). Mortality in the next millennium. *Paper presented to the Staple Inn Actuarial Society*.

Willets, R. C. (2004). The cohort effect: insights and explanations. *British Actuarial Journal 10*, 833–877.

Wilmoth, J. R. and T. Valkonen (2001). A parametric representation of mortality differentials over age and time. Fifth seminar of the EAPS Working Group on Differentials in Health, Morbidity and Mortality in Europe. Available at: http://demog.berkeley.edu/~jrw/Papers/mortdiff2.pdf.

Wilson, C. (2001). On the scale of global demographic convergence 1950-2000. *Population and Development Review 27*(1), 155–172.

Yang, S. S. and C. Wang (2013). Pricing and securitization of multi-country longevity risk with mortality dependence. *Insurance: Mathematics and Economics 52*(2), 157–169.

Zhou, R., Y. Wangb, K. Kaufholda, J. S. H. Lib, and K. S. Tanb (2013). Modeling mortality of multiple populations with vector error correction models: Applications to Solvency II. Presented at LIFE Colloquium 2013. Available at: http://www.actuaries.org/lyon2013/papers/LIFE_Zhou_Wang_Kaufhol_Li_Tan.pdf.

Ivan Luciano Danesi

CURRICULUM VITAE

Contact Information

University of Padova
Department of Statistics
via Cesare Battisti, 241-243
35121 Padova. Italy.

Tel. +39 349 1813062
e-mail: danesi@stat.unipd.it
ivandanesi@yahoo.it

Current Position

Since January 2011; (expected completion: February 2014)
PhD Student in Statistical Sciences, University of Padova.

Thesis title: Forecasting Mortality in Related Populations Using Lee-Carter Type Models
Supervisor: Prof. Nicola Torelli
Co-supervisors: Prof. Steven Haberman, Dr. Pietro Millossovich, Prof. Ermanno Pitacco

Research interests

- Projected life tables
- Longevity risk
- Actuarial science
- Gaussian process regression

Education

October 2007 – October 2010
Master (*laurea specialistica*) degree in Statistics, Economics and Actuarial Science.

Università Cattolica del Sacro Cuore,
Faculty of Economics/Banking, Finance and Insurance Sciences
Title of dissertation: "Credit risk in Solvency II: regulation and calculation models" (in Italian)
Supervisor: Prof. Nino Savelli
Final mark: 108/110

October 2004 – October 2007
Bachelor degree (*laurea triennale*) in Statistics, Economics and Actuarial Science.

Università Cattolica del Sacro Cuore,
Faculty of Economics/Banking, Finance and Insurance Sciences
Title of dissertation: "Elements of Statistics Decision Theory" (in Italian)
Supervisor: Prof. Diego Zappa
Final mark: 110/110 cum laude

Visiting periods

September 2012 – December 2012
CASS Business School, City University
London, United Kingdom.
Supervisor: Prof. Steven Haberman

Work experience

September 2010 – December 2010
Università Cattolica del Sacro Cuore.
Faculty tutor (Faculty of Banking, Finance and Insurance Sciences).

June 2007 – December 2010
IRSA, Institute for research, consulting and training for insurance companies.
Collaborator (programming and testing).

Awards and Scholarship

2011
PhD scholarship (University of Padova).

2013
Best poster presentation at GSP conference.

2013
CKER Travel Grant, USA Society of Actuaries (SOA).

Computer skills

- Operative System: Windows
- Programming: R, Visual Basic
- Other skills: Latex, Windows Office.

Language skills

Italian: native; English: good.

Publications

Papers in conference proceedings

Danesi, I.L., Haberman, S., Millossovich, P. (2013). Mortality forecasting for related populations using Lee-Carter type models. *Proceedings of the 28th International Workshop on Statistical Modelling* **2**, 551–554, (ISBN 978-88-96251-49-2).

Danesi, I.L., Kaucic, M., Torelli, N. (2013). An application of Kriging to Italian mortality rates. In S.Co. 2013, Milan, September 9-11, USB stick (ISBN 97888-6493-019-0).

Abstracts in conference proceedings

Danesi, I.L., Haberman, S., Millossovich, P. (2014). Forecasting Mortality in Related Populations Using Lee-Carter Type Models: A Comparison. *ARCH 2014.1.*

Danesi, I.L., Haberman, S., Millossovich, P. (2013). Forecasting Mortality in Related Populations Using Lee-Carter Type Models: A Comparison. In 17th International Congress on Insurance Mathematics and Economics (IME), Copenhagen, July 1-3.

Conference presentations

Danesi, I.L. (2013). Forecasting mortality for related sub-population: an application to Italian regional tables. (poster presentation) *GSP*, Bressanone, Italy, 06.02.2013 – 08.02.2013.

Danesi, I.L., Haberman, S., Millossovich, P. (2013). Mortality forecasting for related populations using Lee-Carter type models (poster presentation) *IWSM*, Palermo, Italy, 08.07.2013 – 12.07.2013.

Danesi, I.L., Haberman, S., Millossovich, P. (2013). Forecasting Mortality in Related Populations Using Lee-Carter Type Models: A Comparison (oral presentation) *ARC*, Philadelphia, USA, 01.08.2013 – 03.08.2013.

Danesi, I.L., Kaucic, M., Torelli, N. (2013). An application of Kriging to Italian mortality rates (poster presentation) *S.Co.*, Milano, Italy, 09.09.2013 – 11.09.2013.

Teaching experience

October 2013 – December 2013
Metodi Statistici per la Finanza e le Assicurazioni
MSc in Scienze Statistiche, Attuariali ed Economiche
R laboratory, 12 hours
Università Cattolica del Sacro Cuore
Instructor: Prof. Diego Zappa

October 2013 – December 2013
Statistica II
MSc in Scienze Statistiche, Attuariali ed Economiche
Exercises, 10 hours
Università Cattolica del Sacro Cuore
Instructor: Dr. Diego Attilio Mancuso