



Eidgenössische Technische Hochschule Zürich  
Swiss Federal Institute of Technology Zurich

Swiss Re



ETH ZURICH

Master Thesis

---

Construction and Comparison of Mortality  
Tables Based on Different Techniques

---

Author

Gaby BINDER

Supervisors

Prof. Dr. Mario V. WÜTHRICH

Dr. Klemens BINSWANGER

February 26, 2014



## Abstract

Human longevity has undergone a remarkable increase during history. Therefore, it is crucial for all life insurance companies to have good mortality rate models at their disposal. Our goal is to compare different models to each other in order to choose optimal ones for the construction of mortality tables. We select the following models for this task:

- five common mortality rate models including the generalized Nolfi model,
- generalized linear models,
- the model of Lee and Carter with two variations,
- the Cairns-Blake-Dowd model, and
- three adaptations of the generalized Nolfi model.

We describe these models, we discuss their calibration and we explain how they can be used for the prediction of future mortality. For this purpose, all these models are fitted to data from the Swiss population from 1912 to 2010, for men and women separately using regressions and iterative algorithms. The mortality rates are then extrapolated to higher ages and forecasted to the future. For those models which include a time-dependent process, the forecasting is done using a random walk with drift. To be able to quantify the quality of each model, we define concrete measures based on which the model selection is carried out. After comparing all models to each other, we come to the final conclusion that an adaptation of the generalized Nolfi model with a time-dependent base mortality rate fits the given data best and yields from today's point of view realistic forecasts for the future. So in the end, we are able to construct a mortality table with this model at hand.

# Contents

<b>1</b>	<b>Introduction.....</b>	<b>6</b>
1.1	Overview.....	6
1.2	Frequently Used Notations.....	7
<b>I</b>	<b>Methods to Model Mortality Rates</b>	<b>9</b>
<b>2</b>	<b>Methods to Analyse Data .....</b>	<b>10</b>
2.1	Clearing Method by Whittaker-Henderson .....	10
2.2	Neumann Test.....	11
2.3	Cox and Stuart Test .....	13
<b>3</b>	<b>Common Mortality Rate Models.....</b>	<b>16</b>
3.1	Nolfi Model.....	16
3.2	Generalized Nolfi Model .....	16
3.3	Marginal Mortality Rate Model .....	16
3.4	Extrapolation Model.....	16
3.5	Combined Model.....	17
3.6	Sensitivity Analysis .....	17
<b>4</b>	<b>Generalized Linear Models.....</b>	<b>20</b>
4.1	Generalized Linear Models.....	20
4.2	The Models.....	21
4.3	Fitting Procedure.....	21
4.4	Overdispersion.....	23
<b>5</b>	<b>Models of Lee and Carter .....</b>	<b>25</b>
5.1	The Models.....	25
5.1.1	Classical Model .....	25
5.1.2	Generalized Model .....	25
5.1.3	Extended Model .....	26
5.2	Fitting Procedure.....	26
5.2.1	Classical Model .....	26
5.2.2	Generalized Model .....	30
5.2.3	Extended Model .....	30
<b>6</b>	<b>Cairns-Blake-Dowd Model.....</b>	<b>32</b>
6.1	The Model.....	32
6.2	Fitting Procedure.....	33
<b>7</b>	<b>Adapted Nolfi Models .....</b>	<b>34</b>
7.1	The Models.....	34
7.2	Fitting Procedure.....	35
<b>8</b>	<b>Extrapolating the Data .....</b>	<b>37</b>

*Contents*

<b>9</b>	<b>Forecasting Mortality Rates</b> .....	<b>39</b>
9.1	Generalized Linear Models .....	39
9.2	Models of Lee and Carter .....	39
9.3	Cairns-Blake-Dowd Model .....	40
9.4	Adapted Nolfi Models .....	43
<b>10</b>	<b>Mortality Rates of Order 1</b> .....	<b>45</b>
10.1	Residual Risk for one Period of Time .....	45
10.2	Model and Parameter Risk for one Period of Time .....	49
<b>11</b>	<b>Measures for Longevity</b> .....	<b>50</b>
11.1	Life Expectancy .....	50
11.2	Pricing Annuities .....	51
<b>II</b>	<b>Construction and Comparison of Mortality Tables</b>	<b>53</b>
<b>12</b>	<b>Description of the Data</b> .....	<b>54</b>
<b>13</b>	<b>Application of Common Mortality Rate Models</b> .....	<b>57</b>
13.1	Generalized Nolfi Model .....	57
13.2	Marginal Mortality Rate Model .....	63
13.3	Extrapolation Model.....	67
<b>14</b>	<b>Application of Generalized Linear Models</b> .....	<b>73</b>
<b>15</b>	<b>Application of the Models of Lee and Carter</b> .....	<b>79</b>
15.1	Classical Model.....	79
15.2	Extended Model .....	86
<b>16</b>	<b>Application of the Cairns-Blake-Dowd Model</b> .....	<b>92</b>
<b>17</b>	<b>Application of Adapted Nolfi Models</b> .....	<b>98</b>
17.1	Model 1.....	99
17.2	Model 2.....	104
17.3	Model 3.....	108
<b>18</b>	<b>Comparison of Mortality Rate Models</b> .....	<b>114</b>
18.1	Age Restrictions .....	114
18.2	Quality of the Estimators .....	115
18.3	Quality of the Forecasts .....	116
18.4	Measures for Longevity .....	118
18.5	Construction of a Mortality Table .....	121
<b>19</b>	<b>Conclusion</b> .....	<b>126</b>
<b>A</b>	<b>Appendix</b> .....	<b>131</b>
A.1	Notations .....	131

# 1 Introduction

## 1.1 Overview

People live longer today than they did in the past. Due to medical and industrial development it was possible to nearly eliminate child mortality. For middle-aged and old people, improved technologies helped cure diseases which were lethal years ago. For the very old generation, however, the improvement is not as significant since deaths caused by the symptoms of old age cannot yet be eliminated. This matter was discussed by the Swiss Federal Statistical Office (2009a) in their release concerning longevity in Switzerland. They showed that since 1876, the life expectancy at birth of men and women in Switzerland increased each year by 3 - 5 months on average. Thus, the life expectancy has nearly doubled since then. This fact and low birth rates caused the population structure to change over the years from a pyramid-like form to a more urn shaped one. This development is not only present in Switzerland but in almost all industrial countries.

The longevity of society poses major long-term problems for life insurance companies and pension funds which hold lots of annuities. As people started putting more emphasis on the increased life time of the population, the companies were facing a new challenge in pricing annuities. Due to the recent financial crisis, the importance of models which accurately describe the change of the mortality rates of insured people became even more evident as investment returns do no longer make up for underestimated longevity issues. In addition, these models have to provide reasonable forecasts for the future since we assume that longevity will increase even further. Unfortunately, the speed of this increase is not known so insurers are facing longevity risk which they cannot simply eliminate through diversification. To this day, there does not exist one commonly accepted model. In fact, each firm chooses one according to its own individual criteria and beliefs. Moreover, there is no unique best choice for a company as they may pursue different goals.

In this thesis, we discuss various existing models and show their advantages as well as their disadvantages by applying them to real data. We aim to show a possible approach of how a life insurance company can select a mortality table by using the mortality rate model which yields the best results for the insurance company's objectives. We will apply the derived theory to data from the Swiss population from 1912 to 2010 to give one explicit example of how to construct a mortality table. It is important to note that the resulting model may not be the best choice for every ethnological group. However, all techniques we describe can easily be adapted to other datasets.

A very simple model has been introduced in 1959 by P. Nolfi. Various more complex models have been invented since then. Some of them even include an explicit method of how to make forecasts for the future. We focus here on five different groups of models. We will discuss some commonly used models to which also the Nolfi model belongs, generalized linear models, the models of Lee and Carter (1992), the Cairns-Blake-Dowd model (2006) that was used to analyse the mortality rates of men in England and Wales and finally three adaptations of the generalized Nolfi model that were designed by Lüthy et al. (2001) to construct a new annuity

## 1 Introduction

product. This is only a small selection from the variety of existing models. However, these five groups form a good sample since each of them uses other notions to tackle the task of modeling mortality rates. For all these models, we explain how to find parameter estimators using different types of regressions as well as iterative algorithms. These parameter estimators are needed to estimate mortality rates in the past as well as in the future. The forecasting techniques, however, depend on the model. If it contains no evolving time-dependence, the forecasting can be done by inserting any desired year into the model equation. Otherwise, we model the time-varying process via a random walk with drift. The estimated and forecasted mortality rates are then used to derive concrete measures which quantify the longevity of a society. They serve as indicators about its past and future development.

This thesis is split up into two parts. In the first part, starting in Chapter 2, we give a general discussion about data analysis. Afterwards, we introduce in Chapters 3 to 7 the mortality rate models we are going to analyse. In the two subsequent chapters, we extend the data with respect to age and show how the rates can be forecasted to the future. Chapter 10 contains a short discussion about mortality rates which take care of the residual risk for one period of time. We conclude the first part with Chapter 11 where we introduce three measures for longevity. In the second part, we apply all these methods to data from the Swiss population described in Chapter 12 for men and women separately. We finish the thesis with comparing all models to each other and explicitly constructing a mortality table with the most preferable model in Chapter 18.

### 1.2 Frequently Used Notations

Before we start the main part of this thesis, we give a short description of the most important notations we will use. This should serve as a reference for the basic concepts of this work.

We will have a look at discrete, two-dimensional models where the mortality rate depends on the age  $x \in \mathbb{N}$ , which takes values inside the set  $\{x_0, x_0 + 1, \dots, x_0 + \omega\}$ , and on the year  $t \in \mathbb{N}$ , which takes values from  $t_0$  to  $t_0 + \tau$  this age was measured in. It is important to give a more explicit definition of how the age of a person is measured because this is not obvious in discrete time. Since we will have a look at data from the Swiss population in the second part of this thesis, we apply one of the following two definitions of age given by the Swiss Federal Statistical Office (2013a):

- |                    |  |
|--------------------|--|
| Alter, erfülltes:  | Das Alter beim letzten Geburtstag, d.h. die Anzahl der vollendeten Lebensjahre.  |
| Alter, erreichtes: | Das im betrachteten Kalenderjahr erreichte Alter, d.h. die Differenz zwischen der Jahreszahl des betreffenden Kalenderjahres und der Jahreszahl des Geburtsjahres. |

The data we will use for modeling were collected according to the second definition. This means that the age  $x$  of a person is equal to the year  $t$  under consideration minus the year of birth.

We denote by  $D_{x,t}$  the number of deaths of persons with age  $x$  in year  $t$  and by  $D_t = \sum_x D_{x,t}$  the total number of deaths in that year. If both these quantities are not known they need to be

## 1 Introduction

modeled by random variables. We denote by  $(\Omega, \mathcal{F}, \mathbb{P})$  a probability space with  $(\Omega, \mathcal{F})$  assumed to be sufficiently rich and  $\mathbb{P}$  being the probability measure under which  $(D_{x,t})_{x,t}$  are observed. This is often called the real world probability measure. The expectation with respect to  $\mathbb{P}$  is denoted by  $\mathbb{E}$ . We additionally choose a filtration  $\mathbb{F} = (\mathcal{F}_t)_t$  such that  $\mathcal{F}_t$  contains all information available up to time  $t$ . The filtration must satisfy  $\{\emptyset, \Omega\} = \mathcal{F}_{t_0} \subset \dots \subset \mathcal{F}_{t_0+\tau} \subset \mathcal{F}$ . This means, we actually work on a filtered probability space  $(\Omega, \mathcal{F}, \mathbb{P}, \mathbb{F})$ . We assume that  $D_{x,t}$  is  $\mathcal{F}_t$ -measurable for all  $x$  and its realizations are denoted by  $d_{x,t} \in \mathbb{N}$ . We further let  $l_{x,t} \in \mathbb{N}$  be the number of persons of age  $x$  that are alive in the beginning of year  $t$ .

These concepts are needed to define the most important quantity in this thesis which is the annual mortality rate

$$Q_{x,t} = \frac{D_{x,t}}{l_{x,t}} \sim F_\theta,$$

where  $F_\theta$  is a parametric distribution function depending on the year  $t$ . The first models we will introduce are deterministic and therefore only model the expected annual mortality rate  $q_{x,t} = \mathbb{E}[Q_{x,t}] \in [0, 1]$ . Later on, we will work with the rate  $q_{x,t}^\Theta = \mathbb{E}[Q_{x,t} \mid \Theta_t] \in [0, 1]$ , where  $\Theta_t$  is an  $\mathcal{F}_t$ -measurable stochastic process, which already contains some a priori knowledge about the true mortality rate. We will also have a look at mortality rates of order 1, denoted by  $q_{x,t}^1 \in [0, 1]$ , which take care of the residual risk for one period of time.

Another crucial factor is the life expectancy  $e_{x,t} \in \mathbb{R}_+$  of a person of age  $x$  in year  $t$ . This is the number of years we expect this person to have left until he dies. For the definition of the life expectancy, we need the probability that a person of age  $x$  in year  $t$  survives to age  $x + j$ . We denote this by  $p_{x,t,j} \in [0, 1]$ . This is connected with the remaining life time of a person of age  $x$  in year  $t$ , denoted by  $T_{x,t}$ . It must hold that  $\mathbb{P}[T_{x,t} > j] = p_{x,t,j}$ . With this, we define  $e_{x,t} = \sum_{j=1}^{\infty} j \mathbb{P}[T_{x,t} = j]$ .

Throughout the whole thesis, a  $\hat{x}$  indicates that this is an estimator whereas a  $\tilde{x}$  describes a forecasted value. In Chapter 2, however,  $\tilde{x}$  also stands for adjusted data. It will always be clear, though, what we mean. For a complete list of notations, we refer the reader to Appendix A.1.

Now that we know the most basic notations, we start Part I of this thesis with a general discussion about data analysis. We will have a look at data adjustments that always need to be done when working with raw data.



## Part I

# Methods to Model Mortality Rates

## 2 Methods to Analyse Data

Since we are working with real data in the second part of this thesis, we should know what needs to be done in advance to analyse these raw data in a reasonable way. First, one needs to smoothen them to even out the most extreme fluctuations. This can be done with a clearing method, for example the one by Whittaker-Henderson explained in Marguerite (2011). What is also good to know is whether there is a trend in the data and if there is one, if it is increasing or decreasing. Therefore, we start these notes by showing some methods of how to approach these two tasks.

### 2.1 Clearing Method by Whittaker-Henderson

We first introduce a clearing method that can be applied in all fields of statistics. Clearing methods are used to smoothen raw data. This is needed if they fluctuate too much or if they contain jumps or errors. There are three groups of clearing methods. We will only have a look at one method from the group of so-called mechanical clearing methods. In this group, the error between the raw and the adjusted data is minimized for a range of data points simultaneously.

We denote by  $(x_i)_{i=1}^n$  the raw data and by  $(\tilde{x}_i)_{i=1}^n$  the adjusted data. Marguerite (2011) explains that for this clearing method, one needs to choose appropriate measures to quantify the influence of the adjustment as well as the smoothness of the adjusted values and make them both small simultaneously. To be able to define these measures, we need to know what the discrete derivative is.

**Definition 2.1.** For a series of observations  $(x_i)_i$  the **discrete derivative**  $\Delta^v$  of order  $v \in \mathbb{N}$  is defined as

$$\Delta^v x_i = \sum_{j=0}^v (-1)^j \binom{v}{j} x_{i+j}. \quad (2.1)$$

**Example 2.2.** For  $v = 1$ , this means  $\Delta x_i = x_i - x_{i+1}$ , and for  $v = 2$ , we get  $\Delta^2 x_i = x_i - 2x_{i+1} + x_{i+2}$ . This could also be calculated via  $\Delta^2 x_i = \Delta(\Delta x_i) = \Delta x_i - \Delta x_{i+1} = x_i - 2x_{i+1} + x_{i+2}$ . ■

Using this definition, we specify the two desired measures. We quantify the influence of the adjustment by the weighted  $l^2$ -distance

$$\sum_{i=1}^n w_i (\tilde{x}_i - x_i)^2 \quad (2.2)$$

for weights  $w_i \in [0, 1]$  that sum up to 1, and the smoothness of the adjusted values by

$$\sum_{i=1}^{n-v} (\Delta^v \tilde{x}_i)^2, \quad (2.3)$$

where  $v \in \mathbb{N}$  is a constant describing how smooth we want the adjusted data to be. This leads to the following optimization problem:

## 2 Methods to Analyse Data

$$\sum_{i=1}^n w_i (\tilde{x}_i - x_i)^2 + g \sum_{i=1}^{n-v} (\Delta^v \tilde{x}_i)^2 = \text{minimal}, \quad (2.4)$$

where  $g \geq 0$  is a constant determining the trade-off between the fit (2.2) and the smoothness (2.3).

It is easy to solve this problem using linear algebra. We denote by  $\mathbf{x} \in \mathbb{R}^n$  and  $\tilde{\mathbf{x}} \in \mathbb{R}^n$  the vectors of the raw and adjusted data, respectively, by  $W \in \mathbb{R}^{n \times n}$  the diagonal matrix containing the weights  $(w_i)_i$  and by  $S \in \mathbb{R}^{(n-v) \times n}$  the matrix describing the discrete derivative, which means that  $\Delta^v x_i = (S\mathbf{x})_i$  holds. In other words, in the  $i^{\text{th}}$  row of  $S$ , the first  $i-1$  and the last  $n-v-i$  values are equal to 0 whereas the remaining  $v+1$  values are given by  $(-1)^j \binom{v}{j}$  for  $j = 0, \dots, v$ . The optimization problem can then be written as

$$(\tilde{\mathbf{x}} - \mathbf{x})^\top W (\tilde{\mathbf{x}} - \mathbf{x}) + g (S\tilde{\mathbf{x}})^\top (S\tilde{\mathbf{x}}) = \text{minimal}. \quad (2.5)$$

Since  $(W + g S^\top S)$  is non-singular and positive definite, it is invertible. The solution to this problem is thus equal to

$$\tilde{\mathbf{x}} = (W + g S^\top S)^{-1} W \mathbf{x}. \quad (2.6)$$

These are now the values that are used for any application instead of the originally measured ones. How  $g$  and  $v$  should be chosen depends on the smoothness one wants to achieve. How this choice influences the adjusted data is shown in the following example.

**Example 2.3.** We have a look at the sensitivity of this clearing method to the parameters  $g$  and  $v$ . Hence, we simulate 50 data points from the  $t$ -distribution with 4 degrees of freedom. We have a look at the adjusted data for  $g = 1, 3, 5, 7, 9$  and  $v = 1, 2, 4, 6, 8, 10$ . The resulting adjusted data  $\tilde{\mathbf{x}}$  are shown in Figure 2.1. We observe that they do not change much for different values of  $g$  but a change in  $v$  has a large impact on the shape of the adjusted curve. The smaller  $v$  is, the smoother the curve gets but unfortunately, the approximation of the original data is worse in this case. This follows from the fact that  $v$  determines the number of terms in a Taylor-like approximation since we calculate the  $v^{\text{th}}$  discrete derivative in the measure of smoothness.  $v = 1$ , for example, provides a first-order approximation. ■

## 2.2 Neumann Test

Next, we have a look at two different techniques to test for a trend in the data. The first one can only be used to test if there is a trend in the data or not. We define the null-hypothesis such that it neglects the existence of a trend.

The Neumann test measures the dispersion of successive observations  $x_1, \dots, x_n$ . To do so, we define

$$d^2 = \frac{1}{n-1} \sum_{i=1}^{n-1} (x_i - x_{i+1})^2, \quad (2.7)$$

$$s^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2, \quad (2.8)$$

## 2 Methods to Analyse Data

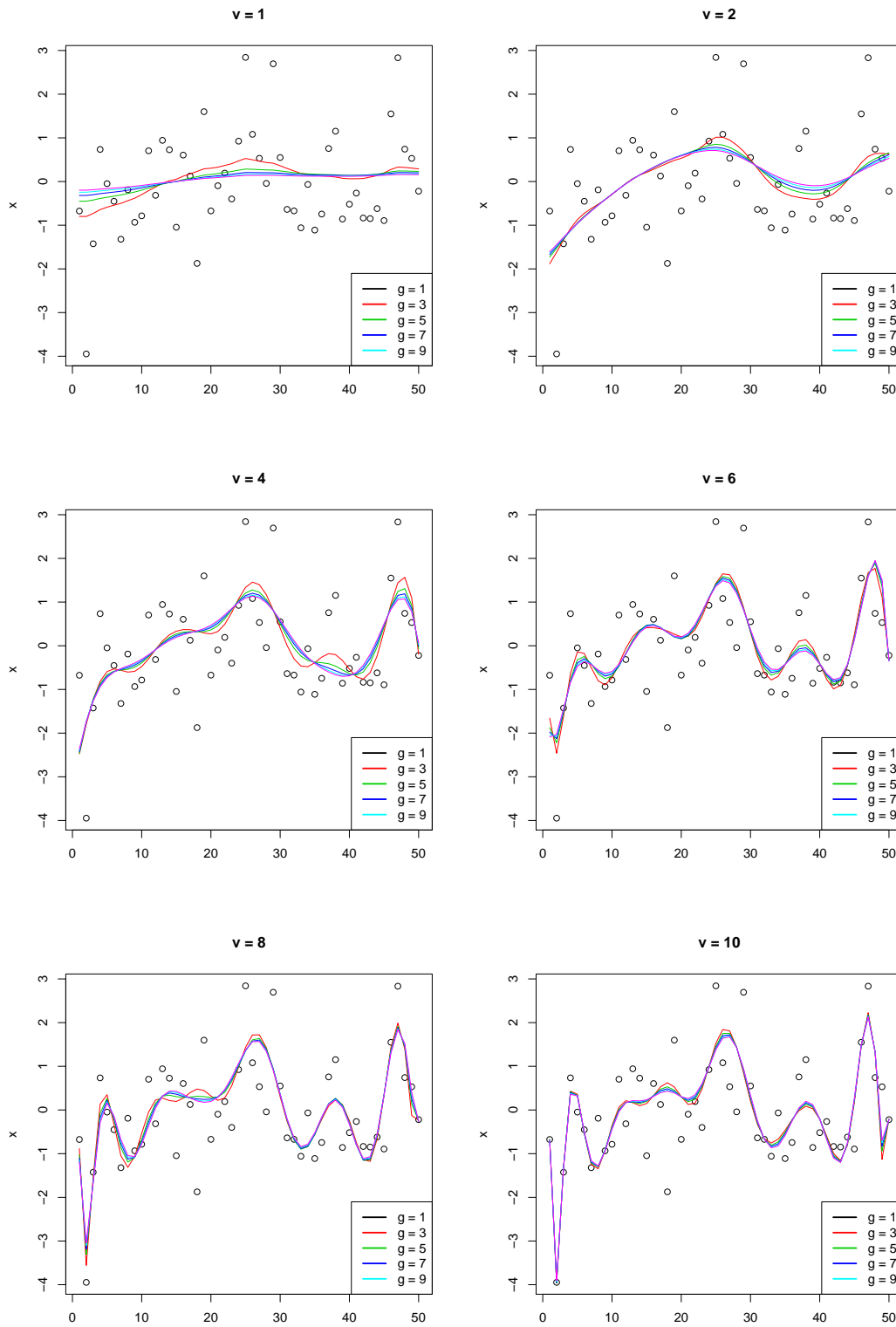


Figure 2.1: Adjusted data  $\tilde{x}$  for different values of  $g$  and  $v$

where  $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$ . Sachs and Hedderich (2006) claim that if the successive data are uncorrelated and have the same mean, it must hold that  $\frac{d^2}{s^2} \approx 2$ . As soon as there is a trend in the data, we will get  $\frac{d^2}{s^2} < 2$  because in this case, successive observations are closer to each other than to their sample mean  $\bar{x}$ . Why the critical value is equal to 2 follows from the next proposition.

**Proposition 2.4.** *For uncorrelated random variables  $(X_i)_{i=1}^n$  with  $\mathbb{E}[X_i] = \mu$  and  $\text{Var}(X_i) = \sigma^2 < \infty$ , it holds that  $\mathbb{E}[d^2] = 2 \mathbb{E}[s^2]$ .*

*Proof.* To prove this, we need the following:

$$\begin{aligned} \mathbb{E}[\bar{X}] &= \frac{1}{n} \sum_{i=1}^n \mathbb{E}[X_i] = \mathbb{E}[X_i] = \mu, \\ \text{Var}(\bar{X}) &= \frac{1}{n^2} \text{Var}\left(\sum_{i=1}^n X_i\right) = \frac{1}{n^2} \sum_{i=1}^n \text{Var}(X_i) = \frac{\sigma^2}{n}. \end{aligned}$$

This is used to calculate  $\mathbb{E}[d^2]$  and  $\mathbb{E}[s^2]$ .

$$\begin{aligned} \mathbb{E}[d^2] &= \frac{1}{n-1} \sum_{i=1}^{n-1} \mathbb{E}[(X_i - X_{i+1})^2] = 2 \mathbb{E}[X_i^2] - 2 \mathbb{E}[X_i]^2 = 2 \sigma^2, \\ \mathbb{E}[s^2] &= \frac{1}{n-1} \mathbb{E}\left[\sum_{i=1}^n (X_i - \mu + \mu - \bar{X})^2\right] \\ &= \frac{1}{n-1} \left( \sum_{i=1}^n \mathbb{E}[(X_i - \mu)^2] - 2 \mathbb{E}\left[\sum_{i=1}^n (X_i - \mu)(\bar{X} - \mu)\right] + n \mathbb{E}[(\bar{X} - \mu)^2] \right) \\ &= \frac{1}{n-1} \left( n \text{Var}(X_i) - 2n \mathbb{E}[(\bar{X} - \mu)^2] + n \mathbb{E}[(\bar{X} - \mu)^2] \right) \\ &= \frac{1}{n-1} \left( n \text{Var}(X_i) - n \text{Var}(\bar{X}) \right) = \frac{n}{n-1} \left( \sigma^2 - \frac{\sigma^2}{n} \right) = \sigma^2. \end{aligned}$$

■

This means that for uncorrelated random variables with identical mean, we expect  $d^2$  to be twice as large as  $s^2$  which explains the critical value of 2.

For the significance level  $\alpha \in (0, 1)$ , we reject the null-hypothesis if  $\frac{d^2}{s^2}$  is smaller or equal to the  $\alpha$ -quantile of the distribution of  $\frac{d^2}{s^2}$  which depends on  $n$ . These values can be found in Table 7.18 on page 361 in Sachs and Hedderich (2006). For  $n$  large, they can be approximated by

$$2 - 2 \frac{\Phi^{-1}(\alpha)}{\sqrt{n+1}}. \tag{2.9}$$

### 2.3 Cox and Stuart Test

The second test can be used to perform a two-sided test to show the existence of a trend in the data, or two one-sided tests to show if there is an increasing or a decreasing trend. The null-hypothesis is then chosen accordingly but always neglecting the existence of a trend.

## 2 Methods to Analyse Data

At the beginning of the Cox and Stuart test, the  $n$  successive observations are divided into 3 groups. It is important that the first and the last group have equal size. We calculate the difference between the  $k^{\text{th}}$  element of the last group and the  $k^{\text{th}}$  element of the first group. If the result is negative, we set the trend for this pair equal to  $-1$ . Otherwise, we set it equal to  $+1$ . Sachs and Hedderich (2006) claim that the sum  $S$  over all  $+1$  and  $-1$ , respectively, depending on the test one is performing, is approximately  $\mathcal{N}(\frac{n}{6}, \frac{n}{12})$ -distributed. We use this fact to apply the Central Limit Theorem (CLT) and get that

$$\hat{z} = \frac{|S - \frac{n}{6}|}{\sqrt{\frac{n}{12}}} \quad (2.10)$$

is approximately standard normally distributed. For  $n < 30$ , we need to adjust  $\hat{z}$  using the continuity correction of Yates (1934). In this case, we set

$$\hat{z} = \frac{|S - \frac{n}{6}| - \frac{1}{2}}{\sqrt{\frac{n}{12}}}. \quad (2.11)$$

For the one-sided tests, we reject the null-hypothesis if  $\hat{z}$  is larger than the  $\alpha$ -quantile of the standard normal distribution for the significance level  $\alpha \in (0, 1)$ . For the two-sided test, we reject it if  $\hat{z}$  is larger than the  $\frac{\alpha+1}{2}$ -quantile.

**Example 2.5.** To improve the understanding of these two tests, we apply them to a simple dataset. We consider the following 17 observations:

$$\{1, 2, 5, 7, 1, 5, 8, 9, 4, 5, 7, 8, 3, 3, 5, 6, 8\}.$$

We choose the significance level  $\alpha = 0.95$  for both of them to test the null-hypothesis

$$H_0 : \text{There is no trend in the data.} \quad (2.12)$$

In the Neumann Test, we get  $d^2 = 8.8125$  and  $s^2 = 6.3603$ . Together with (2.9), this yields

$$\frac{d^2}{s^2} = 1.3855 > 1.2246 = 2 - 2 \frac{\Phi^{-1}(0.95)}{\sqrt{18}}.$$

This implies that we cannot reject the null-hypothesis.

<i>Last group</i>	8	3	3	5	6	8
<i>First group</i>	1	2	5	7	1	5
<i>Trend</i>	+1	+1	-1	-1	+1	+1

Table 2.2: Cox and Stuart test

For the Cox and Stuart test, we divide the observations into 3 groups and calculate the differences between the elements in the first and the last group. The 3 groups are given by  $\{1, 2, 5, 7, 1, 5\}$ ,  $\{8, 9, 4, 5, 7\}$  and  $\{8, 3, 3, 5, 6, 8\}$ . The results are shown in Table 2.2. Since we are performing a two-sided test and 4 out of 6 signs are positive, we get  $S = 2$  and

$$\hat{z} = \frac{|2 - \frac{17}{6}| - \frac{1}{2}}{\sqrt{\frac{17}{12}}} = 0.2801 < 1.9600 = \Phi^{-1}(0.975).$$

## *2 Methods to Analyse Data*

So we again cannot reject the null-hypothesis and get the same result as in the Neumann test. We conclude that there might not be a trend in the data. ■

There exist different clearing methods as well as tests for the existence of a trend. We have only shown a small selection to get a feeling for the importance of these issues. With this remark, we finish the chapter on general data analysis and start the introduction of various models for the mortality rates.

### 3 Common Mortality Rate Models

After the data are adjusted and the testing for a trend is done, we start introducing different methods to model the mortality rate of a person. As a first step, we have a look at a few simple but commonly used ones that can be found in Koller (2003). This should serve as an indicator of the variety of models that are known to this day.

#### 3.1 Nolfi Model

The Nolfi model was introduced by P. Nolfi in 1959 and is given by

$$q_{x,t} = q_{x,t_0} \exp(-\lambda_x (t - t_0)), \quad \text{where} \quad \lambda_x = -\frac{\log(0.5)}{\max(40, x)} > 0. \quad (3.1)$$

$\lambda_x \in \mathbb{R}_+$  is defined such that its denominator is the period of time after which the mortality rate of a person of age  $x$  will be halved. Note that we need to fix the initial year  $t_0$ . Hence, the model highly depends on the choice of  $q_{x,t_0} \in [0, 1]$ .

#### 3.2 Generalized Nolfi Model

This model is a generalization of the Nolfi model (3.1). It looks exactly the same, namely

$$q_{x,t} = q_{x,t_0} \exp(-\lambda_x (t - t_0)), \quad (3.2)$$

but the difference is that  $\lambda_x > 0$  is estimated directly from the data.  $\lambda_x$  can again be interpreted to contain information about the half-life of a person of age  $x$ .

#### 3.3 Marginal Mortality Rate Model

We assume that

$$q_{x,t} = q_{x,t_0} (\alpha_x + (1 - \alpha_x) \exp(-\lambda_x (t - t_0))) \quad (3.3)$$

converges to a marginal mortality rate  $\alpha_x q_{x,t_0}$  for  $\lambda_x > 0$  when  $t$  goes to infinity. This rate is not observable and therefore needs to be estimated as well. This model is a weighted mean between the constant model  $q_{x,t_0}$  and the generalized Nolfi model (3.2), where  $\alpha_x \in [0, 1]$  defines the weight that is assigned to  $q_{x,t_0}$ .

#### 3.4 Extrapolation Model

We assume that the mortality rate decays super- or sub-exponentially which can be modeled by

$$q_{x,t} = q_{x,t_0} \exp(-\lambda_x (t - t_0)^{c_x}), \quad (3.4)$$

where  $c_x > 1$  creates a super-exponential decay and  $c_x < 1$  causes the rate to decrease sub-exponentially. Unfortunately, the estimation of  $c_x \in \mathbb{R}_+$  highly depends on  $x$  and is thus highly non-trivial, see for example the right-hand side of Figure 3.2 below.



### 3.5 Combined Model

We can combine all the models from above and get

$$q_{x,t} = q_{x,t_0} (\alpha_x + (1 - \alpha_x) \exp(-\lambda_x (t - t_0)^{c_x})). \quad (3.5)$$

Due to its general form, however, the estimation of the parameters becomes rather difficult.

### 3.6 Sensitivity Analysis

We study the sensitivity of these models to the choice of parameters. For this, we calculate the logit-transform  $g : [0, 1] \rightarrow \mathbb{R}$ , given by

$$g(x) = \log\left(\frac{x}{1-x}\right), \quad (3.6)$$

of the mortality rates for the first four models above by plugging in different values for the parameters. We choose  $x$  from 0 to 95, so all parameters are sequences of length 96. We choose them as realistic as possible. Since we will estimate them in the second part, we choose for each parameter a linear approximation of these values. This means

$$\lambda_x = 0.0416 - 0.0004 x, \quad (3.7)$$

$$\alpha_x = 0.0092 - 0.0001 x, \quad (3.8)$$

$$c_x = 0.6259 + 0.0136 x. \quad (3.9)$$

We compare the mortality logit-rates calculated with these parameters to those where we perturb them by  $\pm 10\%$ . We choose for  $q_{x,t_0}$  the adjusted mortality rates (2.6) from the clearing method by Whittaker-Henderson of women in 1912 with equidistant weights,  $v = 2$  and  $g = 1.5$ . For more details on the data see Chapter 12. Lastly, we choose  $t - t_0 = 20$ .

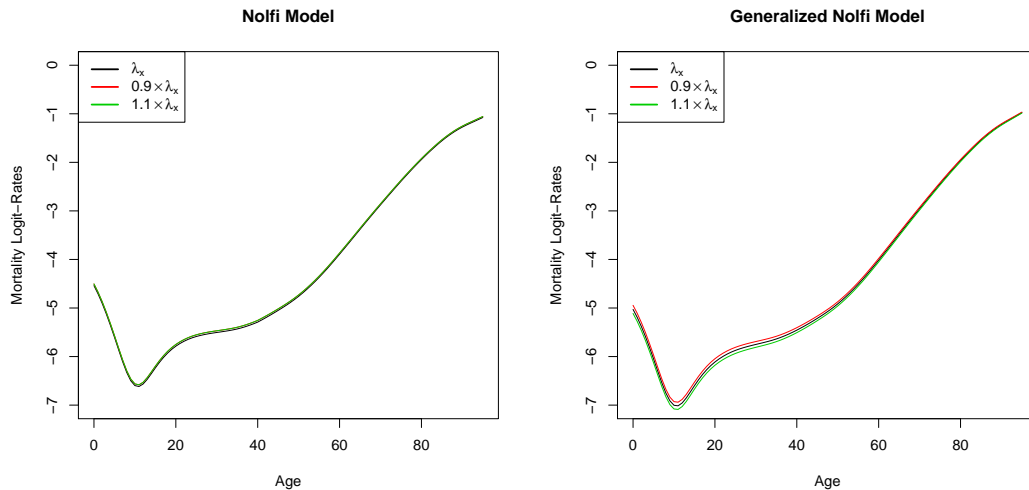


Figure 3.1: (LHS) Nolfi model, (RHS) generalized Nolfi model

### 3 Common Mortality Rate Models

For the Nolfi model (3.1),  $\lambda_x$  can be calculated directly. We see in Figure 3.1 that a change of 10% has nearly no impact at all on the mortality curve. The generalized Nolfi model (3.2), on the other hand, was calculated with the values (3.7) for  $\lambda_x$ . We can see on the right-hand side of the figure that  $\lambda_x$  has an impact on the mortality rate since the perturbed curves lay a bit above and below the original one. However, the difference is only very small.

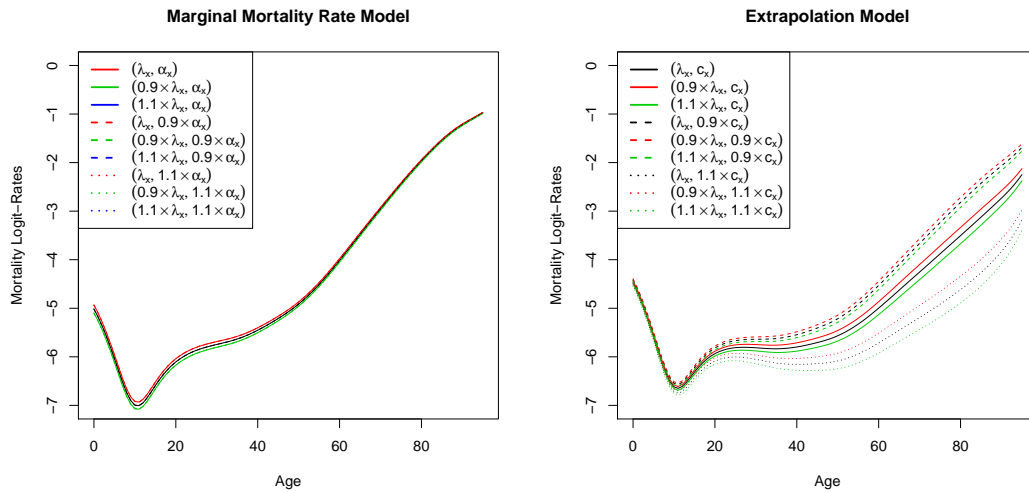


Figure 3.2: (LHS) Marginal mortality rate model, (RHS) extrapolation model

The marginal mortality rate model (3.3), shown in Figure 3.2, is rather unaffected by the choice of both parameters.  $\lambda_x$ , however, has a slightly larger effect on the rate. This parameter again shifts the mortality curve to the top or the bottom. For the extrapolation model (3.4), a change in  $\lambda_x$  affects the logit-rate in a similar way as it did in the generalized Nolfi model, but the effect is intensified by  $c_x$ . For  $c_x > 1$ , the difference between the three curves is most distinct but it is clearly visible for all ages. Also  $c_x$  itself affects the shape of the mortality curve. When it increases, the logit-rates become smaller. This means it is very important to estimate this parameter correctly since the model looks completely different rather quickly.

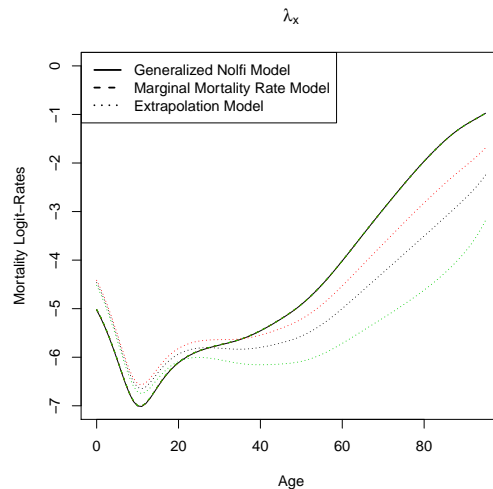


Figure 3.3: Comparison of the models

### 3 Common Mortality Rate Models

We compare the four models to each other directly. Therefore, we fix  $\lambda_x$  to be equal to (3.7) and only let the other parameters vary. The results are shown in Figure 3.3. We observe that the marginal mortality rate model and the generalized Nolfi model nearly coincide for all values of  $\alpha_x$ . The extrapolation model, on the other hand, looks different to both of them for nearly all ages and is generally more flat than the other two. For ages up to 30, the mortality logit-rates are higher than for the other models whereas the ones for old people are lower. This is the age where  $c_x$  gets larger than 1. All results, of course, depend on the choice of  $q_{x,t_0}$ .

In conclusion, we say that a change in  $\lambda_x$  seems to have a slight impact on the mortality curve whereas one in  $\alpha_x$  nearly does not influence it at all. This may also be because we chose it rather close to zero. A slight change in  $c_x$ , on the other hand, completely changes the curve. This means, even though these models are simple they are rather sensitive to parameter changes. Therefore, we need other models in order to get a better understanding of the data. This is one reason why we discuss more complex models in the subsequent chapters.

## 4 Generalized Linear Models

The first group of models we will discuss in detail is rather straightforward. These models are called generalized linear models and depend on the choice of a link function. Thus, the group is rather large since we can choose from a variety of link functions and in addition plug in different predictors. This group of models was described in the notes of Koller (2003) where most results in this chapter are coming from.

### 4.1 Generalized Linear Models

Generalized linear models (GLMs) are not only used in modeling mortality rates but in various fields of statistics. Therefore, we first give an introduction of the general concept.

The idea of GLMs is based on linear regression. We have a random vector  $\mathbf{Y} \in \mathbb{R}^n$  which we want to explain by a design matrix  $X \in \mathbb{R}^{n \times k}$ .  $x_{i0} = 1$  is assumed for all  $i = 1, \dots, n$ . If we assume this relation to be linear we are exactly in the case of linear regression. This means that there exists a parameter vector  $\boldsymbol{\theta} \in \mathbb{R}^k$  and *i.i.d.* error terms  $\varepsilon_i \sim \mathcal{N}(0, \sigma^2)$ , such that  $Y_i = \theta_0 + \sum_{j=1}^{k-1} x_{ij}\theta_j + \varepsilon_i$ . The main difficulty is to estimate the parameter vector  $\boldsymbol{\theta} \in \mathbb{R}^k$ . This can be done using the least squares method where  $\sum_{i=1}^n \varepsilon_i^2$  is minimized for given observations  $\mathbf{Y}$ . This yields that the estimator  $\hat{\boldsymbol{\theta}} \in \mathbb{R}^k$  must satisfy

$$(X^\top X) \hat{\boldsymbol{\theta}} = X^\top \mathbf{Y}. \quad (4.1)$$

$(X^\top X)^{-1}$  only exists if  $X$  is of full rank since this implies that  $X^\top X$  is positive definite. We denote by  $\hat{Y}_i = \mathbf{x}_i^\top \hat{\boldsymbol{\theta}}$  the estimated values of  $Y_i$  and by  $r_i = Y_i - \hat{Y}_i$  the residuals. However, we generally only want those predictors in the model that describe the variability of  $\mathbf{Y}$  best at a minimal number of parameters. Therefore, we compare a model with  $\hat{\boldsymbol{\theta}}_1 \in \mathbb{R}^k$  to a smaller one with  $\hat{\boldsymbol{\theta}}_2 \in \mathbb{R}^l$ , where  $l < k$ , which is contained in the other model. We can apply an  $F$ -test to test the null-hypothesis

$$H_0 : \hat{\theta}_{l+1} = \dots = \hat{\theta}_k = 0. \quad (4.2)$$

We define the sum of squares  $SS(\hat{\boldsymbol{\theta}}) = \sum_{i=1}^n (Y_i - \hat{Y}_i)^2$ . Under  $H_0$ , the test statistic

$$\frac{n-k}{k-l} \frac{SS(\hat{\boldsymbol{\theta}}_2) - SS(\hat{\boldsymbol{\theta}}_1)}{SS(\hat{\boldsymbol{\theta}}_1)} \quad (4.3)$$

is  $F_{k-l, n-k}$  distributed. If it is larger than the significance level of this particular  $F$ -distribution, the null-hypothesis is rejected. For more information on linear regression, we refer the reader to Stahel (1999).

The assumptions of the linear regression model are not always satisfied. For example in our framework, they do not hold true. To show this, we denote the number of persons alive at age  $x$  in the beginning of some year  $t$  by  $l_{x,t} \in \mathbb{N}$  and the number of persons who die within that year by  $D_{x,t}$ . We may assume that for a fixed year, the number of deaths for age  $x$  in a portfolio are *i.i.d.* This implies

$$D_{x,t} \sim \text{Binom}(l_{x,t}, q_{x,t}), \quad (4.4)$$

## 4 Generalized Linear Models

where  $q_{x,t} \in [0, 1]$  denotes the expected annual mortality rate of age  $x$  in year  $t$ . We denote by  $Q_{x,t} = \frac{D_{x,t}}{l_{x,t}}$  the annual mortality rate. We observe that

$$\mathbb{E}[Q_{x,t}] = \frac{\mathbb{E}[D_{x,t}]}{l_{x,t}} = q_{x,t}. \quad (4.5)$$

If we want the expected mortality rate to be linear in  $x$  and  $t$ , we would set  $\mathbb{E}[Q_{x,t}] = \theta_1 + \theta_2 x + \theta_3 t$ . However, this is not possible since the left-hand side takes only values in  $[0, 1]$  whereas the right-hand side may achieve any value in  $\theta_1 + \theta_2 \mathbb{N} + \theta_3 \mathbb{N}$ . Hence, we relax the assumption of linearity by choosing a monotone and continuously differentiable link function  $g : [0, 1] \rightarrow \mathbb{R}$ , such that

$$g(\mathbb{E}[Q_{x,t}]) = g(q_{x,t}) = \theta_1 + \theta_2 x + \theta_3 t. \quad (4.6)$$

In theory, every monotone and continuously differentiable function could be chosen for  $g$  but in our case, the logit-link  $g : [0, 1] \rightarrow \mathbb{R}$  given by (3.6) is a good choice since it makes calculations easy and is appropriate in many mortality rate modeling examples. For more details on generalized linear models, we refer to Johnson and Wichern (2007).

### 4.2 The Models

We now model the expected mortality rate  $q_{x,t}$  with such a generalized linear model. We have already chosen the link-function to be the logit-link which implies that

$$g^{-1}(y) = \frac{e^y}{1 + e^y}. \quad (4.7)$$

What is left is the choice of the predictor. We will have a look at two different predictors. We choose the linear one given by  $\theta_1 + \theta_2 x + \theta_3 t$  and a more general one, namely  $\theta_1 + \theta_2 x + \theta_3 t + \theta_4 x t + \theta_5 x^2 t$ . After combining (4.6) and (4.7), we obtain the models

$$q_{x,t} = g^{-1}(\theta_1 + \theta_2 x + \theta_3 t) = \frac{\exp(\theta_1 + \theta_2 x + \theta_3 t)}{1 + \exp(\theta_1 + \theta_2 x + \theta_3 t)}, \quad (4.8)$$

$$q_{x,t} = g^{-1}(\theta_1 + \theta_2 x + \theta_3 t + \theta_4 x t + \theta_5 x^2 t) = \frac{\exp(\theta_1 + \theta_2 x + \theta_3 t + \theta_4 x t + \theta_5 x^2 t)}{1 + \exp(\theta_1 + \theta_2 x + \theta_3 t + \theta_4 x t + \theta_5 x^2 t)}. \quad (4.9)$$

We observe that if the predictor is very small, the denominator in (4.9) is approximately equal to 1 and thus,

$$q_{x,t} \approx e^{\theta_1 + \theta_2 x} e^{(\theta_3 + \theta_4 x + \theta_5 x^2)t}.$$

This is exactly equal to the generalized Nolfi model (3.2) with  $t_0 = 0$ ,  $\lambda_x = -(\theta_3 + \theta_4 x + \theta_5 x^2)$  and  $q_{x,0} = e^{\theta_1 + \theta_2 x}$ .

### 4.3 Fitting Procedure

We want to use these models to estimate mortality rates. Hence, we need to estimate the parameters of the models. This is done with maximum likelihood estimation (MLE). For simplicity, we only show it for the model with the linear predictor. We assume that we have

#### 4 Generalized Linear Models

independent Binomial observations  $(d_{x,t})_{x,t} \in \mathbb{N}^{(\omega+1) \times (\tau+1)}$ . The log-likelihood function for  $\boldsymbol{\theta} \in \mathbb{R}^3$  is then given by

$$\begin{aligned} l(\boldsymbol{\theta}) &= \log \left( \prod_{x,t} \mathbb{P} \left[ Q_{x,t} = \frac{d_{x,t}}{l_{x,t}} \right] \right) = \log \left( \prod_{x,t} \mathbb{P} [D_{x,t} = d_{x,t}] \right) \\ &= \sum_{x,t} \log \left( \binom{l_{x,t}}{d_{x,t}} q_{x,t}^{d_{x,t}} (1 - q_{x,t})^{l_{x,t} - d_{x,t}} \right) \\ &\propto \sum_{x,t} d_{x,t} \log(q_{x,t}) + \sum_{x,t} (l_{x,t} - d_{x,t}) \log(1 - q_{x,t}). \end{aligned} \quad (4.10)$$

When we plug (4.8) into the log-likelihood function (4.10) and differentiate with respect to  $\theta_k$  for all  $k = 1, 2, 3$ , we can set this equal to 0 to maximize  $l(\boldsymbol{\theta})$ . We obtain the equations

$$\begin{aligned} \frac{\partial l(\boldsymbol{\theta})}{\partial \theta_1} &= \sum_{x,t} \left( d_{x,t} - l_{x,t} \frac{\exp(\theta_1 + \theta_2 x + \theta_3 t)}{1 + \exp(\theta_1 + \theta_2 x + \theta_3 t)} \right) = 0, \\ \frac{\partial l(\boldsymbol{\theta})}{\partial \theta_2} &= \sum_{x,t} \left( d_{x,t} - l_{x,t} \frac{\exp(\theta_1 + \theta_2 x + \theta_3 t)}{1 + \exp(\theta_1 + \theta_2 x + \theta_3 t)} \right) x = 0, \\ \frac{\partial l(\boldsymbol{\theta})}{\partial \theta_3} &= \sum_{x,t} \left( d_{x,t} - l_{x,t} \frac{\exp(\theta_1 + \theta_2 x + \theta_3 t)}{1 + \exp(\theta_1 + \theta_2 x + \theta_3 t)} \right) t = 0. \end{aligned}$$

Unlike for linear regression, this system of equations can only be solved numerically, for instance using the Newton-Raphson algorithm or Fisher's scoring method. The solution  $\hat{\boldsymbol{\theta}} \in \mathbb{R}^3$  to this system is called the maximum likelihood estimator of  $\boldsymbol{\theta} \in \mathbb{R}^3$ . By plugging it back into (4.8), we get estimated mortality rates.

To be able to analyse the quality of the resulting fit, we need to know how the residuals look like and define a measure for the quality. There are different ways of defining residuals but we only introduce the deviance residuals. For these, we need to define the scaled deviance

$$D(\hat{\boldsymbol{\theta}}) = 2 l^M - 2 l(\hat{\boldsymbol{\theta}}), \quad (4.11)$$

where  $l^M \in \mathbb{R}_+$  is the maximal log-likelihood achieved at the point  $\frac{d_{x,t}}{l_{x,t}}$  and therefore given by

$$l^M = \sum_{x,t} \left( \log \left( \frac{l_{x,t}}{d_{x,t}} \right) + d_{x,t} \log(d_{x,t}) + (l_{x,t} - d_{x,t}) \log(l_{x,t} - d_{x,t}) - l_{x,t} \log(l_{x,t}) \right), \quad (4.12)$$

and  $l(\hat{\boldsymbol{\theta}})$  is the log-likelihood of the estimated model for  $\hat{\boldsymbol{\theta}} \in \mathbb{R}^k$ . We call  $\text{dev}_{x,t} \in \mathbb{R}_+$  the part of the deviance resulting from observation  $d_{x,t}$ , which means that  $D(\hat{\boldsymbol{\theta}}) = \sum_{x,t} \text{dev}_{x,t}$ . It follows from (4.11) and (4.12) that for the binomial distribution,

$$\text{dev}_{x,t} = 2 \left( d_{x,t} \log \left( \frac{d_{x,t}}{l_{x,t} \hat{q}_{x,t}} \right) + (l_{x,t} - d_{x,t}) \log \left( \frac{l_{x,t} - d_{x,t}}{l_{x,t} (1 - \hat{q}_{x,t})} \right) \right). \quad (4.13)$$

The deviance residuals are then defined as

$$r_{x,t}^D = \text{sgn}(q_{x,t} - \hat{q}_{x,t}) \sqrt{\text{dev}_{x,t}}. \quad (4.14)$$

## 4 Generalized Linear Models

As a measure of the quality of the model, we take Akaike's information criterion

$$AIC = 2 df - 2l(\hat{\theta}), \quad (4.15)$$

where  $df = (\omega + 1)(\tau + 1) - k$  are the degrees of freedom. This should be as small as possible. However, we could also do an asymptotic  $F$ -test like we did for linear regression where we compare the scaled deviance of two models to each other instead of their total sum of squares. We again compare a model with  $\hat{\theta}_1 \in \mathbb{R}^k$  to a smaller one with  $\hat{\theta}_2 \in \mathbb{R}^l$  and  $l < k$  contained in the other model. As before, we test if the null-hypothesis (4.2) can be rejected. Under  $H_0$ , the test statistic

$$\frac{df}{k-l} \frac{D(\hat{\theta}_2) - D(\hat{\theta}_1)}{D(\hat{\theta}_1)} \quad (4.16)$$

has approximately an  $F_{k-l, df}$ -distribution. Therefore, we reject the null-hypothesis if it is larger than the significance level of this specific  $F$ -distribution. This gives us an additional instrument to compare different models to each other and possibly reduce the number of parameters. In particular, we can test model (4.8) against (4.9).

### 4.4 Overdispersion

We need to be careful with the assumption that  $D_{x,t} \sim \text{Binom}(l_{x,t}, q_{x,t})$ . It is possible that the variance of the observed data is much higher than the one for the binomial distribution. The ratio between the observed variance and the one of the binomial distribution with the same parameters is called overdispersion. We show here where this phenomenon may come from.

We set  $X \sim \text{Binom}(n, p)$  the number of defaults in a portfolio of  $n$  policies. It holds that

$$\mathbb{E} \left[ \frac{X}{n} \right] = p \quad \text{and} \quad \text{Var} \left( \frac{X}{n} \right) = \frac{p(1-p)}{n}.$$

It is more realistic, however, to assume that a person holds more than one policy at a time. We denote by  $n_i \in \mathbb{N}$  the number of persons holding  $i$  policies and by  $X_i$  the number of deaths in group  $i$ . It is easy to see that  $n = \sum_i i n_i$ ,  $X = \sum_i i X_i$  and  $X_i \sim \text{Binom}(n_i, p)$ . This implies that

$$\mathbb{E} \left[ \frac{X}{n} \right] = p \quad \text{and} \quad \text{Var} \left( \frac{X}{n} \right) = \underbrace{\frac{\sum_i i^2 n_i}{n}}_{\phi} \frac{p(1-p)}{n},$$

where  $\phi \geq 1$  is the overdispersion. Note that  $\phi < 1$  would denote underdispersion. Overdispersion can also be caused by dependence and parameter uncertainty. An example for this would be the mixed Poisson distribution.

To see if overdispersion is present in the data, one has to estimate  $\phi$ . This can be done in two ways which are both given in Wüthrich (2013). We can either use Pearson's residuals to define

$$\hat{\phi}_P = \frac{1}{df} \sum_{x,t} \frac{(d_{x,t} - l_{x,t} \hat{q}_{x,t})^2}{l_{x,t} \hat{q}_{x,t} (1 - \hat{q}_{x,t})}, \quad (4.17)$$

#### 4 Generalized Linear Models

where  $df = (\omega + 1)(\tau + 1) - k$  are the degrees of freedom. Or, we can use the scaled deviance and estimate  $\phi$  by

$$\hat{\phi}_D = \frac{D(\hat{\boldsymbol{\theta}})}{df}. \quad (4.18)$$

These are the most important facts we need to fit observed mortality rates with a generalized linear model. Thus, we finish the chapter and continue with the next group of models. They will not be linear in time anymore but will model the change of mortality over the years with a stochastic process. This is more realistic since the assumption of linearity is rather restrictive.



## 5 Models of Lee and Carter

In this group of models, we model the annual mortality rate of a portfolio on the basis of the model introduced by Lee and Carter (1992). The classical model is two-dimensional and describes the change of mortality over time by a time-varying index. There also exist two variations of this model which additionally include the generation of a person and some other general factor, respectively. They were discussed in Butt and Haberman (2010).

### 5.1 The Models

#### 5.1.1 Classical Model

First, we have a look at the classical model introduced by Lee and Carter (1992). The model for the mortality log-rate of a person of age  $x$  in year  $t$  looks as follows:

$$\log(Q_{x,t}) = a_x + b_x k_t + \varepsilon_{x,t}, \quad (5.1)$$

where  $a_x \in \mathbb{R}$  and  $b_x \in \mathbb{R}_+$  are age-specific constants,  $k_t \in \mathbb{R}$  is a time-varying index describing the level of mortality and  $\varepsilon_{x,t} \in \mathbb{R}$  is an error term specifying distortions in  $Q_{x,t}$  that cannot be explained by the systemic part of the model. We assume that  $\mathbb{E}[\varepsilon_{x,t}] = 0$  and  $\text{Var}(\varepsilon_{x,t}) = \sigma^2$ , which is reasonable since  $\varepsilon_{x,t}$  are the errors of mortality log-rates which vary less than the rates themselves. Because we want the mortality rates to decrease over the years,  $(k_t)_t$  should have a decreasing trend.  $\exp(a_x)$  describes the general shape of the mortality rate for a fixed year  $t$  and  $b_x$  describes how quickly the mortality rate decreases for changes in  $k_t$ .  $b_x$  has to be positive such that the mortality rate goes to 0 for any  $x$  as  $k_t$  goes to minus infinity.

Butt and Haberman (2010) show that this model can also be interpreted as a GLM with the log-link  $g(x) = \log(x)$ . We take here the log-link instead of the logit-link to make sure that the resulting GLM fits into the Lee and Carter framework. This means, we assume a Poisson distribution for  $D_{x,t}$  with mean  $l_{x,t} \exp(a_x + b_x k_t)$ . Thus, we can write in terms of the expected number of deaths

$$g(\mathbb{E}[D_{x,t}]) = a_x + b_x k_t + \log(l_{x,t}). \quad (5.2)$$

The concept behind this model is different to the classical one since the expected number of deaths are fitted and not the mortality log-rates. Nevertheless, it is needed for the iterative fitting procedure we will introduce in the next section.

#### 5.1.2 Generalized Model

In the generalized model of Lee and Carter discussed in the paper of Butt and Haberman (2010), we have an additional term that depends on the year of birth  $t - x$  of a person of age  $x$  in year  $t$ . This is often called the cohort effect. The mortality log-rate is modeled by

$$\log(Q_{x,t}) = a_x + b_x k_t + c_x v_{t-x} + \varepsilon_{x,t}, \quad (5.3)$$

## 5 Models of Lee and Carter

where  $a_x$ ,  $b_x$ ,  $k_t$  and  $\varepsilon_{x,t}$  are as in the classical model and  $v_{t-x} \in \mathbb{R}$  describes the cohort effect. Hence,  $c_x \mathbb{R}_+$  is also non-negative, just like  $b_x$ . We see that the classical model of Lee and Carter can be found by setting  $c_x = 0$  for all  $x$ . This model is equivalent to the GLM

$$g(\mathbb{E}[D_{x,t}]) = a_x + b_x k_t + c_x v_{t-x} + \log(l_{x,t}), \quad (5.4)$$

where  $g$  and  $D_{x,t}$  are as in (5.2). This form will be needed in the fitting procedure.

### 5.1.3 Extended Model

The idea of the extended model, which is also found in Butt and Haberman (2010), is that the mortality rate may also depend on an additional variable  $w \in \mathbb{N}$ , taking values from  $w_0$  to  $w_0 + \nu$ , different from age and year. In general, one chooses a geographical, economical or sociological factor. The number of deaths, the population size and the mortality rate now depend on the three indices  $(x, t, w)$ , where  $w$  describes the group each person belongs to. The model is then given by

$$\log(Q_{x,t,w}) = a_x + b_x k_t + d_w + \varepsilon_{x,t,w}, \quad (5.5)$$

where  $a_x$ ,  $b_x$ ,  $k_t$  and  $\varepsilon_{x,t,w}$  are as in the classical model and  $d_w \in \mathbb{R}$  measures the differences between the groups defined by the choice of  $w$ . We could also make  $b_x$  and  $k_t$  dependent on  $w$  but this would yield a more complex model so we assume that the effect of the variable  $w$  is only via  $d_w$ . We again express this as a GLM with a log-link and a Poisson distribution for the number of deaths. This means,

$$g(\mathbb{E}[D_{x,t,w}]) = a_x + b_x k_t + d_w + \log(l_{x,t,w}). \quad (5.6)$$

This form will again be needed for finding parameter estimators.

## 5.2 Fitting Procedure

We want to fit these models to real data. This is done for each of them separately. For the classical model, we introduce three different techniques to get parameter estimators whereas for the other two, we give only one possibility.

### 5.2.1 Classical Model

Unfortunately, the classical model of Lee and Carter is under-determined. For any solution  $\mathbf{a} \in \mathbb{R}^{\omega+1}$ ,  $\mathbf{b} \in \mathbb{R}_+^{\omega+1}$ ,  $\mathbf{k} \in \mathbb{R}^{\tau+1}$  and a constant  $c \in \mathbb{R} \setminus \{0\}$ , also  $(\mathbf{a} - c\mathbf{b}, \mathbf{b}, \mathbf{k} + c)$  and  $(\mathbf{a}, c\mathbf{b}, \frac{\mathbf{k}}{c})$  are solutions to the same least squares problem. This means,  $\mathbf{a}$  is only unique up to an additive constant,  $\mathbf{b}$  only up to a multiplicative constant and  $\mathbf{k}$  only up to linear transformations. Nevertheless, we want to get a unique solution so we normalize the parameters in a reasonable way. We choose the normalization

$$\sum_x b_x = 1 \quad \text{and} \quad \sum_t k_t = 0. \quad (5.7)$$

With this,  $(b_x)_x$  can be interpreted as weights for  $k_t$  and it implies that  $k_t$  is zero on average. We show now three methods to estimate the model parameters.

## 5 Models of Lee and Carter

The first two methods were introduced by Lee and Carter (1992). We assume that we know all realizations of  $D_{x,t}$  up to year  $t$  and denote them by  $d_{x,t} \in \mathbb{N}$ . After summing the observed mortality log-rates over  $t$ , we choose

$$\hat{a}_x = \frac{1}{\tau + 1} \sum_t \log \left( \frac{d_{x,t}}{l_{x,t}} \right). \quad (5.8)$$

The estimation of  $k_t$  and  $b_x$ , however, can be done in two ways. We could either sum over  $x$  as well and find that

$$\hat{k}_t = \sum_x \left( \log \left( \frac{d_{x,t}}{l_{x,t}} \right) - \hat{a}_x \right), \quad (5.9)$$

even though the error terms need not sum to 0 over age in general. It then remains to estimate  $b_x$ . This is done by first normalizing the mortality log-rates by subtracting  $\hat{a}_x$  and then performing a regression without a constant term on  $\hat{k}_t$  for each age  $x$ . This means, we minimize

$$\sum_t \left( \log \left( \frac{d_{x,t}}{l_{x,t}} \right) - \hat{a}_x - \hat{b}_x \hat{k}_t \right)^2, \quad (5.10)$$

with respect to  $\hat{b}_x$  for all  $x$  separately. This is the easiest version to get the estimators but we had to assume that the  $\varepsilon_{x,t}$  approximately sum to 0 over  $x$  which does not hold true necessarily.

The alternative method is to estimate  $b_x$  and  $k_t$  simultaneously. This can be done by optimizing

$$\sum_{x,t} \left( \log \left( \frac{d_{x,t}}{l_{x,t}} \right) - \hat{a}_x - \hat{b}_x \hat{k}_t \right)^2 = \text{minimal}, \quad (5.11)$$

with respect to  $\hat{b}_x$  and  $\hat{k}_t$ . We cannot solve this with ordinary regression techniques since we have too many parameters that need to be estimated at once. Therefore, we apply the Singular Value Decomposition (SVD) method.

We first show how the SVD for a matrix  $A \in \mathbb{R}^{m \times n}$  for  $m \geq n$  looks like. We call the non-zero square roots of the eigenvalues from  $A^\top A$  the singular values of the matrix  $A$  and denote them by  $\sigma_1 \geq \dots \geq \sigma_n \geq 0$ . Golub and Reinsch (1970) show that we can decompose  $A$  in the following way:

$$A = UDV^\top, \quad (5.12)$$

where  $U \in \mathbb{R}^{m \times m}$  and  $V \in \mathbb{R}^{n \times n}$  contain the orthonormal eigenvectors of  $AA^\top$  and  $A^\top A$ , respectively, and it holds that  $U^\top U = I$  and  $V^\top V = I$ . These vectors are called the left and right singular vectors of  $A$ .  $D \in \mathbb{R}^{m \times n}$  can be written as  $D = (\sigma_1 e_1, \dots, \sigma_r e_r, 0, \dots, 0)$ , where  $r \in \mathbb{N}$  is the number of positive singular values and  $e_i \in \mathbb{R}^m$  denotes the  $i^{\text{th}}$  unit vector.

In our case,  $A$  contains the normalized mortality log-rates and thus,  $A \in \mathbb{R}^{(\omega+1) \times (\tau+1)}$ . In general, we have more ages we are interested in than years of observations which means that  $\omega \geq \tau$ . So we really are in the case described above and can apply the SVD method. We decompose  $A$  as in (5.12) and obtain  $A = UDV^\top = (\hat{b}_x \hat{k}_t + \varepsilon_{x,t})_{x,t}$ . We take the first left singular vector of  $A$ , which is given by the first column of  $U$ , to be equal to  $(\hat{b}_x)_x$  and the first right singular vector, which is equal to the first column of  $V$ , times the largest singular value of  $A$  to be equal to  $(\hat{k}_t)_t$ . We show with a simple example why this is a reasonable choice of estimators.

**Example 5.1.** We take  $A \in \mathbb{R}^{3 \times 2}$  and decompose

$$A = (u_1 \ u_2 \ u_3) \begin{pmatrix} \sigma_1 & 0 \\ 0 & \sigma_2 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} v_1^\top \\ v_2^\top \end{pmatrix} = (u_1 \ u_2 \ u_3) \begin{pmatrix} \sigma_1 v_1^\top \\ \sigma_2 v_2^\top \\ 0 \end{pmatrix} = \sigma_1 u_1 v_1^\top + \sigma_2 u_2 v_2^\top.$$

This has to be equal to  $(\hat{b}_x \hat{k}_t + \varepsilon_{x,t})_{x,t}$ . Since  $\sigma_2$  is smaller than  $\sigma_1$ , we set the second term equal to  $(\varepsilon_{x,t})_{x,t}$ . It then remains that  $(\hat{b}_x)_x = u_1$  and  $(\hat{k}_t)_t = \sigma_1 v_1$ . This can similarly be done for any values of  $m$  and  $n$  because the first term always dominates since  $\sigma_1$  is largest. So in any case, we choose the first term to be equal to  $(\hat{b}_x \hat{k}_t)_{x,t}$  and the rest is then simply the error.  $\blacksquare$

For both methods, we are left to check whether the normalizations (5.7) are satisfied. Since  $\hat{a}_x$  was fixed beforehand, the normalization for  $k_t$  already holds true. So it is left to make sure that  $\sum_x b_x = 1$ . This is easily done by normalizing each  $\hat{b}_x$  by dividing it by their sum and multiplying each  $\hat{k}_t$  by the same constant. This transformation then achieves the desired normalization.

In the last method, which is explained in Butt and Haberman (2010), we fit model (5.2) to the data by using properties from Section 4.3. Our aim is to choose the estimators such that the scaled deviance is minimized. We do this since due to (4.11) this means that the log-likelihood function is maximized. Since we assume that  $D_{x,t}$  is Poisson distributed, the scaled deviance is given by

$$D(\hat{\theta}) = \sum_{x,t} \text{dev}_{x,t} = \sum_{x,t} 2 \left( d_{x,t} \log \left( \frac{d_{x,t}}{l_{x,t} \hat{q}_{x,t}} \right) + l_{x,t} \hat{q}_{x,t} - d_{x,t} \right), \quad (5.13)$$

where  $\hat{\theta} \in \mathbb{R}^{2\omega+\tau+3}$  and  $\hat{q}_{x,t} = \exp(\hat{a}_x + \hat{b}_x \hat{k}_t)$ . We minimize this for each parameter  $\hat{\theta} \in \{\hat{a}_x, \hat{b}_x, \hat{k}_t\}$  for all  $x$  and  $t$  separately. Hence, when applying the Newton-Raphson method, we need to update the parameters each step according to

$$u(\hat{\theta}) = \hat{\theta} - \left( \frac{\partial D(\hat{\theta})}{\partial \hat{\theta}} \right) \left( \frac{\partial^2 D(\hat{\theta})}{\partial \hat{\theta}^2} \right)^{-1}, \quad (5.14)$$

where all remaining parameters are fixed. We find that

$$\frac{\partial D(\hat{\theta})}{\partial \hat{\theta}} = \sum \frac{\partial \text{dev}_{x,t}}{\partial \hat{\theta}} = \sum 2 \frac{\partial \hat{q}_{x,t}}{\partial \hat{\theta}} \left( l_{x,t} - \frac{d_{x,t}}{\hat{q}_{x,t}} \right) = \sum 2a (l_{x,t} \hat{q}_{x,t} - d_{x,t}), \quad (5.15)$$

since it holds that

$$\frac{\partial \hat{q}_{x,t}}{\partial \hat{\theta}} = a \hat{q}_{x,t} = \begin{cases} \hat{q}_{x,t}, & \hat{\theta} = \hat{a}_x \\ \hat{k}_t \hat{q}_{x,t}, & \hat{\theta} = \hat{b}_x \\ \hat{b}_x \hat{q}_{x,t}, & \hat{\theta} = \hat{k}_t \end{cases} \quad \text{for } a = \begin{cases} 1, & \hat{\theta} = \hat{a}_x \\ \hat{k}_t, & \hat{\theta} = \hat{b}_x \\ \hat{b}_x, & \hat{\theta} = \hat{k}_t \end{cases}.$$

We sum either over  $x$  or  $t$ , depending on the parameter  $\hat{\theta}$ . We can use (5.15) to calculate the second derivative

$$\frac{\partial^2 D(\hat{\theta})}{\partial \hat{\theta}^2} = \sum 2a^2 l_{x,t} \hat{q}_{x,t}. \quad (5.16)$$

Plugging everything back into (5.14) finally yields

$$u(\hat{\theta}) = \hat{\theta} + \frac{\sum a (d_{x,t} - l_{x,t} \hat{q}_{x,t})}{\sum a^2 l_{x,t} \hat{q}_{x,t}}. \quad (5.17)$$

Thus, we do the following steps:

**Algorithm 5.2.**

1.) Choose initial values: For all  $x$  and  $t$ , we take

$$\hat{a}_x = \frac{1}{\tau + 1} \sum_t \log \left( \frac{d_{x,t}}{l_{x,t}} \right), \quad \hat{b}_x = \frac{1}{\omega + 1} \quad \text{and} \quad \hat{k}_t = 0.$$

Calculate  $\hat{q}_{x,t}$  and  $D(\hat{\theta})$  with these parameters.

2.) Update parameters: Each parameter is updated for all  $x$  and  $t$  separately.

- $\hat{a}_x = \hat{a}_x + \frac{\sum_t (d_{x,t} - l_{x,t} \hat{q}_{x,t})}{\sum_t l_{x,t} \hat{q}_{x,t}}$ .
- $\hat{k}_t = \hat{k}_t + \frac{\sum_x \hat{b}_x (d_{x,t} - l_{x,t} \hat{q}_{x,t})}{\sum_x \hat{b}_x^2 l_{x,t} \hat{q}_{x,t}}$ , and afterwards set  $\hat{k}_t = \hat{k}_t - \frac{1}{\tau + 1} \sum_{t^*} \hat{k}_{t^*}$ .
- $\hat{b}_x = \hat{b}_x + \frac{\sum_t \hat{k}_t (d_{x,t} - l_{x,t} \hat{q}_{x,t})}{\sum_t \hat{k}_t^2 l_{x,t} \hat{q}_{x,t}}$ .

Calculate  $\hat{q}_{x,t}$  after each parameter was updated completely and use it to update the next one.

3.) Check convergence: Calculate  $D(\hat{\theta})$  and repeat Step 2 until  $|D_{old} - D_{new}| < \delta$  for some  $\delta > 0$ . If this does not converge, change the initial values.

4.) Normalize the parameters: Set  $\hat{b}_x = \hat{b}_x \left( \sum_x \hat{b}_x \right)^{-1}$  and  $\hat{k}_t = \hat{k}_t \left( \sum_x \hat{b}_x \right)$ .

This was the final method for estimating  $a_x$ ,  $b_x$  and  $k_t$  we wanted to show for the classical model of Lee and Carter. After a remark, we go on and have a look at the generalized model.

We could also use the classical model of Lee and Carter for another purpose, namely to construct a mortality table that produces the exact number  $d_t \in \mathbb{N}$  of observed deaths per year for given population sizes  $(l_{x,t})_x$ . This means, we could look for  $k_t$  such that

$$d_t = \sum_x l_{x,t} \exp(a_x + b_x k_t). \quad (5.18)$$

This can only be solved with an iterative search and is useful in two different scenarios. First, when plugging the least squares estimators  $\hat{a}_x$ ,  $\hat{b}_x$  and  $\hat{k}_t$  into the model, this will generally not yield the true number of deaths per year since we have optimized over the mortality log-rates and not the rates themselves. Therefore, we could fix  $\hat{a}_x$  and  $\hat{b}_x$  and re-estimate  $k_t$  with the procedure above. The estimated number of deaths will then coincide with the observed one. Second, it might happen that the size of the population and the total number of deaths are known for some years but not the number of deaths per age in each year. With this technique, the model can still be fitted.

### 5.2.2 Generalized Model

The fitting procedure for the generalized model is an iterative one given in Butt and Haberman (2010). We first need to choose a normalization for the parameters to get a unique solution. Analogous to the last section, we take

$$\sum_x b_x = \sum_x c_x = 1 \quad \text{and} \quad k_{t_0} = v_{t_0-x} = 0. \quad (5.19)$$

For this model, we have  $\hat{\theta} \in \{\hat{a}_x, \hat{b}_x, \hat{k}_t, \hat{c}_x, \hat{v}_{t-x}\}$ . We find that  $a = \hat{v}_{t-x}$  for  $\hat{\theta} = \hat{c}_x$  and  $a = \hat{c}_x$  for  $\hat{\theta} = \hat{v}_{t-x}$ . In contrast to the last section,  $\hat{a}_x = \frac{1}{\tau+1} \sum_t \log\left(\frac{d_{x,t}}{l_{x,t}}\right)$  is fixed for all  $x$ . For the remaining parameters, we do the following:

**Algorithm 5.3.**

1.) *Choose initial values:* For all  $x$  and  $t$ , we choose  $\hat{b}_x = \hat{c}_x = \frac{1}{\omega+1}$  and for  $\hat{k}_t$  and  $\hat{v}_{t-x}$ , we take the estimators that result from fitting  $\log\left(\frac{d_{x,t}}{l_{x,t}}\right) - \hat{a}_x = k_t + v_{t-x}$ . Calculate  $\hat{q}_{x,t}$  and  $D(\hat{\theta})$  with these parameters.

2.) *Update parameters:* Each parameter is updated for all  $x$  and  $t$  separately.

- $\hat{k}_t = \hat{k}_t + \frac{\sum_x \hat{b}_x (d_{x,t} - l_{x,t} \hat{q}_{x,t})}{\sum_x \hat{b}_x^2 l_{x,t} \hat{q}_{x,t}}$ , and afterwards set  $\hat{k}_t = \hat{k}_t - \hat{k}_{t_0}$ .
- $\hat{b}_x = \hat{b}_x + \frac{\sum_t \hat{k}_t (d_{x,t} - l_{x,t} \hat{q}_{x,t})}{\sum_t \hat{k}_t^2 l_{x,t} \hat{q}_{x,t}}$ .
- $\hat{v}_{t-x} = \hat{v}_{t-x} + \frac{\sum_x \hat{c}_x (d_{x,t} - l_{x,t} \hat{q}_{x,t})}{\sum_x \hat{c}_x^2 l_{x,t} \hat{q}_{x,t}}$ , and afterwards set  $\hat{v}_{t-x} = \hat{v}_{t-x} - \hat{v}_{t_0-x}$ .
- $\hat{c}_x = \hat{c}_x + \frac{\sum_t \hat{v}_{t-x} (d_{x,t} - l_{x,t} \hat{q}_{x,t})}{\sum_t \hat{v}_{t-x}^2 l_{x,t} \hat{q}_{x,t}}$ .

Calculate  $\hat{q}_{x,t}$  after each parameter was updated completely and use it to update the next one.

3.) *Check convergence:* Calculate  $D(\hat{\theta})$  and repeat Step 2 until  $|D_{old} - D_{new}| < \delta$  for some  $\delta > 0$ . If this does not converge, change the initial values.

4.) *Normalize the parameters:* Set  $\hat{b}_x = \hat{b}_x \left(\sum_x \hat{b}_x\right)^{-1}$ ,  $\hat{k}_t = \hat{k}_t \left(\sum_x \hat{b}_x\right)$ ,  $\hat{c}_x = \hat{c}_x \left(\sum_x \hat{c}_x\right)^{-1}$  and  $\hat{v}_{t-x} = \hat{v}_{t-x} \left(\sum_x \hat{c}_x\right)$ .

These final estimators are the ones we choose to estimate the mortality rates. Hence, we can go on and do the same thing for the extended model.

### 5.2.3 Extended Model

We fit the parameters of the extended model with the same iterative procedure as before. Since we now have three explanatory variables, we need to adjust the procedure a bit. We observe that  $\hat{\theta} \in \{\hat{a}_x, \hat{b}_x, \hat{k}_t, \hat{d}_w\}$  and  $a = 1$  for  $\hat{\theta} = \hat{d}_w$ . Additionally, we always need to sum over two variables at once. The normalizations, however, are the same as in the classical model adding  $d_{w_0} = 0$ . This leads to the following iterative procedure:

**Algorithm 5.4.**

1.) Choose initial values: For all  $x$  and  $t$ , we take

$$\hat{a}_x = \frac{1}{(\tau + 1)(\nu + 1)} \sum_{t,w} \log \left( \frac{d_{x,t,w}}{l_{x,t,w}} \right), \quad \hat{b}_x = \frac{1}{\omega + 1} \quad \text{and} \quad \hat{k}_t = \hat{d}_w = 0.$$

Calculate  $\hat{q}_{x,t,w}$  and  $D(\hat{\theta})$  with these parameters.

2.) Update parameters: Each parameter is updated for all  $x$  and  $t$  separately.

- $\hat{a}_x = \hat{a}_x + \frac{\sum_{t,w} (d_{x,t,w} - l_{x,t,w} \hat{q}_{x,t,w})}{\sum_{t,w} l_{x,t,w} \hat{q}_{x,t,w}}$ .
- $\hat{k}_t = \hat{k}_t + \frac{\sum_{x,w} \hat{b}_x (d_{x,t,w} - l_{x,t,w} \hat{q}_{x,t,w})}{\sum_{x,w} \hat{b}_x^2 l_{x,t,w} \hat{q}_{x,t,w}}$ , and afterwards set  $\hat{k}_t = \hat{k}_t - \frac{1}{\tau+1} \sum_{t^*} \hat{k}_{t^*}$ .
- $\hat{b}_x = \hat{b}_x + \frac{\sum_{t,w} \hat{k}_t (d_{x,t,w} - l_{x,t,w} \hat{q}_{x,t,w})}{\sum_{t,w} \hat{k}_t^2 l_{x,t,w} \hat{q}_{x,t,w}}$ .
- $\hat{d}_w = \hat{d}_w + \frac{\sum_{x,t} (d_{x,t,w} - l_{x,t,w} \hat{q}_{x,t,w})}{\sum_{x,t} l_{x,t,w} \hat{q}_{x,t,w}}$ , and afterwards set  $\hat{d}_w = \hat{d}_w - \hat{d}_{w_0}$ .

Calculate  $\hat{q}_{x,t}$  after each parameter was updated completely and use it to update the next one.

3.) Check convergence: Calculate  $D(\hat{\theta})$  and repeat Step 2 until  $|D_{old} - D_{new}| < \delta$  for some  $\delta > 0$ . If this does not converge, change the initial values.

4.) Normalize the parameters: Set  $\hat{b}_x = \hat{b}_x \left( \sum_x \hat{b}_x \right)^{-1}$  and  $\hat{k}_t = \hat{k}_t \left( \sum_x \hat{b}_x \right)$ .

We then choose the normalized estimators found in the last iteration step for all further calculations.

With the classical model of Lee and Carter, we are able to model the change of mortality over time by only one index. Additionally, it separates the change in age and year which is reasonable since we do not expect the behaviour in age to change highly over the years. In the generalized model, we added a process which describes the cohort effect. In the extended one, we even added a new variable according to which the data are measured. For all of them, though, we assumed that the mortality rate simply increases exponentially. This assumption is weakened in the next model.

## 6 Cairns-Blake-Dowd Model

The next model for the mortality rate we introduce here is a two-factor model explained in detail in the paper by Cairns, Blake and Dowd (2006). This model is a mixture of the two models we have discussed in Chapter 4 and 5. The two factors are not constant over time but will be modeled as stochastic processes like the time-varying index  $(k_t)_t$  in the models of Lee and Carter. On the other hand, the model has a form that is similar to the one we used for the generalized linear models.

### 6.1 The Model

Cairns, Blake and Dowd (2006) take up the idea of Lee and Carter (1992) to model the change of mortality over the years with a stochastic process. They adapted this idea and suggested to model the mortality rate as follows:

$$q_{x,t}^{\Theta} = \mathbb{E}[Q_{x,t} \mid \Theta_t] = \frac{\exp(\Theta_{t,1} + x \Theta_{t,2})}{1 + \exp(\Theta_{t,1} + x \Theta_{t,2})}, \quad (6.1)$$

where  $\Theta_t = (\Theta_{t,1}, \Theta_{t,2})^\top \in \mathbb{R}^2$  is a two-dimensional stochastic process measurable at time  $t$ . We have changed the original model slightly by exchanging the process  $\Theta_{t,1}$  by  $\Theta_{t,1} + t \Theta_{t,2}$ . The model is defined such that the first factor affects all ages in the same way whereas the second one is stronger for higher ages than for lower ones. We find that  $\mathbb{E}[q_{x,t}^{\Theta}] = \mathbb{E}[Q_{x,t}] = q_{x,t}$ . However, this model is only applicable for high ages, for instance  $x \geq 40$ . Otherwise, more than two factors would be needed.

Since this model includes uncertainty, we are exposed to different kinds of risks. The most important one is the mortality risk which characterizes the uncertainty about future mortality rates. This risk can be split up into two parts: short-term mortality risk and longevity risk. The first one corresponds to the risk of having much higher or lower mortality rates over a short period of time due to some big event. For instance, the Spanish flu in 1918 was such an event. After this period, however, we expect the rates to go back to the trend. The longevity risk, on the other hand, describes the uncertainty about the future trend of mortality. The model we have just introduced is used for modeling longevity risk rather than short-term mortality risk.

The reason for considering a stochastic model for the mortality rates is that it has been observed in the past that the longevity of a society has increased rather unpredictably and not at a constant pace. Hence, it is more realistic to assume the mortality rates to change over time like a stochastic process rather than at a fixed rate. On the other hand, we also do not know the true trend and therefore, a stochastic model reflects our modeling uncertainty.

As already mentioned, this model is related to both the generalized linear model with a linear predictor as well as the models of Lee and Carter. As in (4.8), it is assumed that the mortality rates decay according to the inverse logit function. This is not the case in (5.1) where we have chosen an exponential decay. In contrast to (4.8), though, we do not have



constant parameters  $\theta_1$ ,  $\theta_2$  and  $\theta_3$  anymore. We assume them to be stochastic processes just like Lee and Carter (1992) already suggested by introducing the time-varying index  $k_t$ . Therefore, this model contains characteristics from both these models.

## 6.2 Fitting Procedure

To fit this model to real data, we need to estimate the unobservable factors  $\Theta_{t,1}$  and  $\Theta_{t,2}$  in each year  $t$ . This is done with a non-linear least squares regression on (6.1) since this model is not linear in the predictors. There are several ways of solving this regression problem. The easiest one is to fix  $t$  and transform the model so that it becomes linear. It is easy to see that we can rewrite (6.1) as

$$\log\left(\frac{q_{x,t}^\Theta}{1 - q_{x,t}^\Theta}\right) = \Theta_{t,1} + x \Theta_{t,2}, \quad (6.2)$$

which is now linear in  $\Theta_{t,1}$  and  $\Theta_{t,2}$ . Thanks to this transformation, we are able to estimate  $\Theta_{t,1}$  and  $\Theta_{t,2}$  for each  $t$  using ordinary linear least squares regression if we replace  $(q_{x,t}^\Theta)_x$  by the observations  $\left(\frac{d_{x,t}}{l_{x,t}}\right)_x$ . An alternative would be to fit each year a GLM with a logit-link (3.6). How this works was already explained in Section 4.3 and will therefore not be repeated.

When we plug the estimators  $\hat{\Theta}_{t,1}$  and  $\hat{\Theta}_{t,2}$  back into (6.1), we get estimators for  $q_{x,t}^\Theta$  in each year. Since we want to get a realistic model, we expect  $\hat{\Theta}_{t,1}$  to be decreasing in  $t$  on average because this implies that also the mortality rate decreases over time on average.  $\hat{\Theta}_{t,2}$  on the other hand should on average be increasing in  $t$  because the mortality rate curve generally becomes steeper over time. This is based on the fact that mortality decreases more for low ages than for high ones. The difference in the sign of the slopes of the factors is the reason why both of them are needed. Otherwise, it would not be possible for the model to account for these two different effects. If we in addition require  $q_{x,t}^\Theta$  to be increasing in  $x$ , we need  $\hat{\Theta}_{t,2} > 0$ . Note that in order to use this model for prediction we need to extrapolate the process  $(\Theta_t)_t$  into the future. This is discussed in Chapter 9 below.

The Cairns-Blake-Dowd model is more general than the ones we have seen before since it involves two stochastic processes. Unfortunately, it is only applicable for ages above some high threshold. This is not true anymore for the group of models we will introduce in the next chapter. They are all adaptations from a simple model which we have already discussed.

## 7 Adapted Nolfi Models

This chapter is about a group of models introduced by Lüthy et al. (2001). All models are basically equal to the generalized Nolfi model (3.2) but they are adapted such that they can account more appropriately for the change of mortality over the years. The change with respect to age, however, is assumed to be the same as in the generalized Nolfi model.

### 7.1 The Models

Since Lüthy et al. (2001) suggest to take the generalized Nolfi model as a basis for constructing new models, we first recall the most important properties of this model. The expected mortality rate was modeled by

$$q_{x,t} = q_{x,t_0} \exp(-\lambda_x (t - t_0)),$$

where  $\lambda_x > 0$  are age-specific constants and  $t_0$  is fixed in the beginning for each model. When  $\lambda_x$  is known, we can plug in any value for  $t$  to get mortality rates for the past or forecasted rates for the future since the model is deterministic. The reason why they suggest to take this model to construct other models from is its simple form and its common use.

The first adjustment we do is to change  $\lambda_x$  each year depending on the observed mortality rates. One way to achieve this is by making  $\lambda_x$  time-dependent. However, this is rather complex so we simply multiply it each year by  $\alpha_{t,2} \in \mathbb{R}_+$ , where  $\alpha_{t_0,2} = 1$  is assumed. We stick to the notation  $q_{x,t}^\ominus$  from the last chapter to be consistent even though the stochastic process is now called  $(\alpha_{t,2})_t$ . The new model is then given by

$$\text{Model 1: } q_{x,t}^\ominus = q_{x,t_0} \exp(-\lambda_x \alpha_{t,2} (t - t_0)), \quad (7.1)$$

where  $(\alpha_{t,2})_t$  is a positive stochastic process taking on a new value each year. Another possibility would be not to adjust the age-specific constants but the base mortality rate  $q_{x,t_0}$  with a constant  $\alpha_{t,1} \in \mathbb{R}_+$  for all  $t$ . This leads to

$$\text{Model 2: } q_{x,t}^\ominus = \alpha_{t,1} q_{x,t_0} \exp(-\lambda_x (t - t_0)), \quad (7.2)$$

where  $(\alpha_{t,1})_t$  is a positive stochastic process with  $\alpha_{t_0,1} = 1$ . Here, we need to keep in mind that we need  $q_{x,t}^\ominus \leq 1$ . This means, if we do not want to restrict  $\alpha_{t,1}$  from above, we can only apply this model to ages below some high threshold, for example 95. Lastly, we could also adjust both  $\lambda_x$  and  $q_{x,t_0}$ . This means

$$\text{Model 3: } q_{x,t}^\ominus = \alpha_{t,1} q_{x,t_0} \exp(-\lambda_x \alpha_{t,2} (t - t_0)), \quad (7.3)$$

where  $(\alpha_{t,1})_t$  and  $(\alpha_{t,2})_t$  satisfy the properties from above.

Note that the generalized Nolfi model as well as the first adaption of it can both be interpreted as a classical model of Lee and Carter. For the generalized Nolfi model, we find that

$$\log(q_{x,t}) = \underbrace{\log(q_{x,t_0})}_{a_x} - \underbrace{\lambda_x}_{-b_x} \underbrace{(t - t_0)}_{k_t},$$

## 7 Adapted Nolfi Models

as in (5.1). For Model 1, we similarly get

$$\log(q_{x,t}^\Theta) = \underbrace{\log(q_{x,t_0})}_{a_x} - \underbrace{\lambda_x}_{-b_x} \underbrace{\alpha_{t,2} (t - t_0)}_{k_t}.$$

This means,  $k_t$  has a specific form. For the other two models, however, this is not true since they depend on two time-varying processes.

All three methods may yield a reasonable model for the mortality rate. It will depend on the data which one performs best. To be able to compare the three on the basis of data, we need to know how to estimate their parameters.

### 7.2 Fitting Procedure

We want to fit these models to real data. Since  $\lambda_x$  was assumed to be the same as in the generalized Nolfi model, it can be estimated by applying linear regression to (3.2). This means, we use the observations  $\left(\frac{d_{x,t}}{l_{x,t}}\right)_t$  and fit  $\lambda_x$  over the years for each  $x$ . Since the solution depends on the choice of  $q_{x,t_0}$ , we first smoothen the observed rates in year  $t_0$  with the clearing method by Whittaker-Henderson from Section 2.1 to improve the resulting fit. Then, we minimize

$$\sum_t \left( \log \left( \frac{d_{x,t}}{l_{x,t} q_{x,t_0}} \right) + \hat{\lambda}_x (t - t_0) \right)^2. \quad (7.4)$$

Afterwards, we need to check if  $\hat{\lambda}_x > 0$  is satisfied. The resulting  $\hat{\lambda}_x$  are then used in all the following steps.

The estimation of  $\alpha_{t,1}$  and  $\alpha_{t,2}$  can be done rather easily. On first sight, we would guess that a non-linear regression is needed since the models are not linear in the parameters. However, they can easily be transformed to be linear so we can apply ordinary least squares regressions. We thus rewrite the model equations (7.1), (7.2) and (7.3) to

$$\log \left( \frac{d_{x,t}}{l_{x,t} q_{x,t_0}} \right) = -\alpha_{t,2} \hat{\lambda}_x (t - t_0), \quad (7.5)$$

$$\frac{d_{x,t}}{l_{x,t} q_{x,t_0}} = \alpha_{t,1} \exp \left( -\hat{\lambda}_x (t - t_0) \right), \quad (7.6)$$

$$\log \left( \frac{d_{x,t}}{l_{x,t} q_{x,t_0}} \right) = \log(\alpha_{t,1}) - \alpha_{t,2} \hat{\lambda}_x (t - t_0). \quad (7.7)$$

For each fixed  $x$ , we apply a least squares regression to this with the restrictions that  $\alpha_{t,1}, \alpha_{t,2} > 0$  for all  $t$ . For the first model, we could also use the fitting procedures introduced in Section 5.2 since it is a special case of the classical model of Lee and Carter. When we now plug the resulting estimators for  $(\lambda_x)_x$ ,  $(\alpha_{t,1})_t$  and  $(\alpha_{t,2})_t$  back into the model equations, we get the estimated mortality rates we were looking for.

In all the procedures above, we only used data from one year. If we have more years of data available, say  $k \in \mathbb{N}$ , we could also use all of them to find the estimators. For this, we

## 7 Adapted Nolfi Models

assume that the estimated probability that a person of age  $x$  in year  $t - k$  survives to year  $t$  is equal to the one that was observed. This means

$$\prod_{j=0}^k (1 - q_{x+j,t-k+j}^{\ominus}) = 1 - \frac{1}{l_{x,t-k}} \sum_{j=0}^k d_{x+j,t-k+j}. \quad (7.8)$$

If we replace  $q_{x+j,t-k+j}^{\ominus}$  with the corresponding model, we can solve this for  $\alpha_{t,1}$  and  $\alpha_{t,2}$  using an iterative search like the Newton-Raphson method to get the estimators.

With this chapter, we end the introduction of mortality rate models. We now go on and discuss some issues that concern all of them. In the following two chapters, for example, we show how to get mortality rates for any age and any year we want even though the number of available data is always limited.

## 8 Extrapolating the Data

Until now, we have introduced various ways of modeling the mortality rates of a group of people and we have shown how to fit observed data with these models. The next step is to have a closer look at the data themselves. We have already discussed the fact that usually, we do not have a lot of years of data available. How one can handle this problem, however, is discussed in the next chapter.

Here, we focus on another problem, namely that the data are also bounded with respect to age. This means, we have data only for fixed ages  $x$  from  $x_0$  to  $x_0 + \omega$ . The first problem is that maybe  $x_0 \neq 0$ . However, this is not an issue here since we are mostly interested in people with an age of 65 and higher. The problem we are approaching is that due to the longevity of a society, there are more and more people achieving ages where we may not have sufficiently many data available. This is in fact a big problem since we need to account for those as well. We now show a method to overcome this issue.

We extend the data to ages above  $x_0 + \omega$  by applying a method introduced by Denuit and Goderniaux (2005). They claim that the mortality rates increase more slowly for very old ages than for lower ones. The method takes advantage of this fact by performing a constrained quadratic regression on the mortality rates. We assume that the mortality rates follow a log-quadratic regression model given by

$$\log(q_{x,t}^{\ominus}) = \theta_{t,1} + \theta_{t,2} x + \theta_{t,3} x^2, \quad (8.1)$$

for parameters  $(\theta_{t,1}, \theta_{t,2}, \theta_{t,3}) \in \mathbb{R}^3$ . Denuit and Goderniaux (2005) additionally assume that  $q_{130,t}^{\ominus} = 1$  and  $(q_{130,t}^{\ominus})' = 0$  for all  $t$ , where the derivative is with respect to  $x$ . The first assumption tells us that we do not expect anyone to survive to an age above 130. This is not a universal choice but since the highest age ever achieved was 123, it seems reasonable. The second assumption is needed to make sure that the mortality increase is slowing down when approaching the age 130. These two assumptions are equivalent to  $\log(q_{130,t}^{\ominus}) = 0$  and  $\log(q_{130,t}^{\ominus})' = 0$ . This means

$$\begin{aligned} \theta_{t,1} + 130 \theta_{t,2} + 130^2 \theta_{t,3} &= 0, \\ \theta_{t,2} + 260 \theta_{t,3} &= 0. \end{aligned}$$

The range of solutions to these equations is given by  $(130^2 \theta_t, -260 \theta_t, \theta_t)$  where we have exchanged  $\theta_{t,3}$  by  $\theta_t$ . When we plug the parameters back into (8.1), we get

$$\log(q_{x,t}^{\ominus}) = \theta_t (130 - x)^2, \quad (8.2)$$

We can estimate  $\theta_t$  in each year by performing a linear regression onto  $\log(q_{x,t}^{\ominus})$  but we only use ages 75 and higher for the fitting procedure. The full dataset for ages 0 to 130 is then achieved by taking the original mortality rates from 0 to 85 and replacing the others with the estimated mortality rates

$$\hat{q}_{x,t} = \exp(\hat{\theta}_t (130 - x)^2). \quad (8.3)$$

## 8 Extrapolating the Data

**Example 8.1.** We could also do the same thing where the maximal age  $y \in \mathbb{N}$  is not equal to 130. Thus, we want to know how this choice influences the extension of the data. Hence, we fit

$$\log(q_{x,t}^\ominus) = \theta_t (y - x)^2$$

for  $y \in \{120, 121, \dots, 140\}$ . To analyse the dependency of the extension on  $y$ , we use data from men in Switzerland between 1912 and 2010 to estimate  $(\theta_t)_t$ . Then, we compare the different models to each other on the basis of the estimated mortality rates in 2010. More details on the data can be found in Chapter 12.

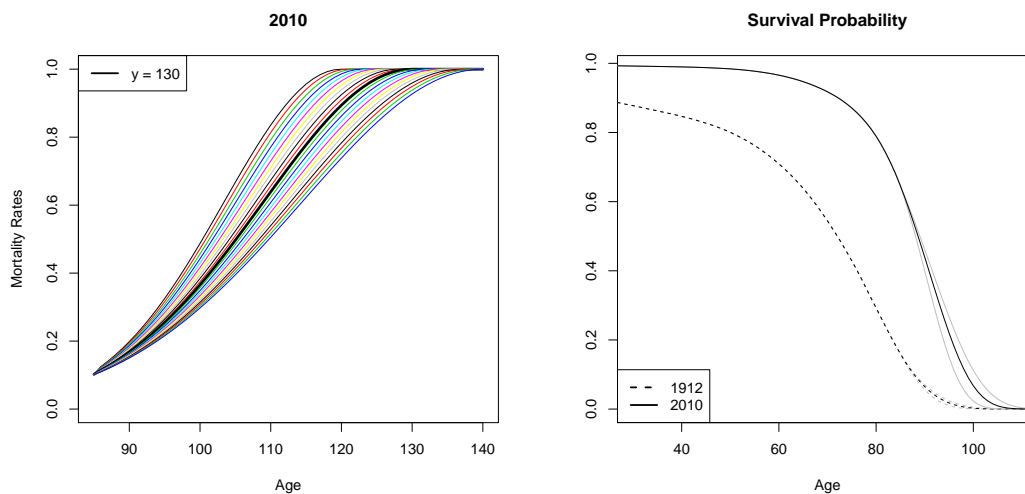


Figure 8.1: (LHS) Mortality rates in 2010,  
(RHS) survival probability  $p_{0,t,j}$  from the second adapted Nolfi model

On the left-hand side of Figure 8.1, we see the extended mortality rates. They are drawn such that  $y$  increases from left to right. The rates are rather different for ages above 100. The maximal deviation between the mortality rates for  $y = 130$  and any other choice of  $y$  is achieved for the model with  $y = 120$ , given by approximately +20% at age 111. To see if this difference has an impact on the results later on, we calculate the survival probability  $p_{0,t,j}$ , see (11.1) below, for these data with the second adapted Nolfi model. On the right-hand side of the figure, the grey lines correspond to the calculated survival probability with  $y = 120$  and 140 and the black ones are for  $y = 130$ . In this picture, we are mostly interested in the age the probability starts decreasing much faster. In both years, the three lines still coincide at that point. Hence, we conclude that the choice of  $y$  is not too important for this picture but we will see that it influences other quantities where the mortality rates of people with an age of 90 and higher are needed. ■

From Example 8.1, we conclude that the extension is sensitive in the choice of the maximal age and we need to carefully analyse its influence on annuities and other life insurance products. This was only one way to extended the available data to ages above  $x_0 + \omega$ . There exist other techniques but this one is rather easy and we will see later on that it yields good results. Therefore, we go on and show how one can extend the data to years in the future.

## 9 Forecasting Mortality Rates

Now that we know how to extend the data to higher ages, we want to show how to find mortality rates for the future. After having calculated the forecasted rates, it is easy to also get forecasts for other objects that only depend on the mortality rate.

### 9.1 Generalized Linear Models

The generalized linear models do not depend on a time-varying index. Since we have shown for model (4.9) how to find the parameter estimator  $\hat{\theta} \in \mathbb{R}^5$ , we can simply insert  $t > 0$  into

$$\tilde{q}_{x,t+s-1} = \frac{\exp\left(\hat{\theta}_1 + \hat{\theta}_2 x + \hat{\theta}_3(t+s-1) + \hat{\theta}_4 x(t+s-1) + \hat{\theta}_5 x^2(t+s-1)\right)}{1 + \exp\left(\hat{\theta}_1 + \hat{\theta}_2 x + \hat{\theta}_3(t+s-1) + \hat{\theta}_4 x(t+s-1) + \hat{\theta}_5 x^2(t+s-1)\right)} \quad (9.1)$$

to get forecasted expected mortality rates.  $s \in \mathbb{N}$  from now on always denotes the beginning of the current year. This procedure works analogously for any predictor and for all models of Chapter 3. For the other three groups of models, the forecasting procedure is not so easy since they include a more evolved time-dependency. We will, however, apply the same method for all of them.

### 9.2 Models of Lee and Carter

In the classical and in the extended model of Lee and Carter, the time-varying index is the only parameter which depends on  $t$ . Therefore, we need to forecast  $k_t$  to get forecasted mortality rates while keeping  $\hat{a}_x$ ,  $\hat{b}_x$  and  $\hat{d}_w$  fixed, where  $\hat{d}_w = 0$  for the classical model. Lee and Carter (1992) suggest to model  $(k_t)_t$  as a random walk with drift. This means we model

$$k_{t+1} = k_t + \mu + c Z_t, \quad (9.2)$$

where  $\mu \in \mathbb{R}$  is the drift parameter,  $c \in \mathbb{R}$  is the standard deviation and  $(Z_t)_t$  are *i.i.d.* error terms with a standard normal distribution. Since in the fitting procedure we have already estimated  $k_t$  for the past, we can use  $(\hat{k}_t)_t$  to get estimators for  $\mu$  and  $c$ . With these, we can calculate the forecast  $\tilde{k}_{t+s-1}$  for any  $t > 0$  according to

$$\tilde{k}_{t+s-1} | \mathcal{F}_s = \hat{k}_{s-1} + t \hat{\mu} + \hat{c} \sum_{u=s}^{t+s-1} Z_u \quad (9.3)$$

for *i.i.d.* random variables  $Z_u \sim \mathcal{N}(0, 1)$ . This implies that

$$\mathbb{E}\left[\tilde{k}_{t+s-1} | \mathcal{F}_s\right] = \hat{k}_{s-1} + t \hat{\mu} \quad \text{and} \quad \text{Var}\left(\tilde{k}_{t+s-1} | \mathcal{F}_s\right) = \hat{c}^2 t.$$

This can now be used to define the forecasted mortality rates in the classical as well as in the extended model by

## 9 Forecasting Mortality Rates

$$\begin{aligned}\tilde{q}_{x,t+s-1} &= \mathbb{E} \left[ \exp \left( \hat{a}_x + \hat{d}_w + \hat{b}_x \tilde{k}_{t+s-1} \right) \mid \mathcal{F}_s \right] \\ &= \exp \left( \hat{a}_x + \hat{b}_x (\hat{k}_{s-1} + t \hat{\mu}) + \frac{1}{2} \hat{b}_x^2 \hat{c}^2 t + \hat{d}_w \right).\end{aligned}\tag{9.4}$$

For the generalized model, we also need to forecast the cohort effect  $v_{t-x}$ . This is done simultaneously with  $k_t$ . We define  $\Theta_t = (k_t, v_{t-x})^\top$  and model

$$\Theta_{t+1} = \Theta_t + \boldsymbol{\mu} + C \mathbf{Z}_t\tag{9.5}$$

where  $\boldsymbol{\mu} \in \mathbb{R}^2$  is the drift parameter vector,  $C \in \mathbb{R}^{2 \times 2}$  is an upper triangular matrix and the error terms  $\mathbf{Z}_t$  are *i.i.d.* and bivariate standard normally distributed. Here, we need to keep in mind that we can use  $\hat{v}_{t+s-1-x}$  as a forecast when  $t + s - 1 - x \leq t_0 + \tau$  for  $t > 0$ . The procedure for finding estimators for  $\boldsymbol{\mu}$  and  $C$  is explained in the section below since it is the same as in the Cairns-Blake-Dowd model.

### 9.3 Cairns-Blake-Dowd Model

In the Cairns-Blake-Dowd model, we have to forecast  $\Theta_t = (\Theta_{t,1}, \Theta_{t,2})^\top \in \mathbb{R}^2$ . Cairns, Blake and Dowd (2006) suggest to model this with a two-dimensional random walk with drift. This means

$$\Theta_{t+1} = \Theta_t + \boldsymbol{\mu} + C \mathbf{Z}_t,\tag{9.6}$$

where  $\boldsymbol{\mu} \in \mathbb{R}^2$  contains the drift parameters,  $C \in \mathbb{R}^{2 \times 2}$  is an upper triangular matrix and  $(\mathbf{Z}_t)_t$  are *i.i.d.* with a two-dimensional standard normal distribution. We define  $V = C^\top C \in \mathbb{R}^{2 \times 2}$  which means that  $C$  is the unique Cholesky factor of  $V$ . This matrix is symmetric and positive definite since all elements of  $C$  are real-valued and the diagonal elements are positive. This altogether means that  $\Theta_{t+1} - \Theta_t$  has a two-dimensional normal distribution with mean  $\boldsymbol{\mu}$  and covariance matrix  $V$  for all  $t$ .

$(\Theta_{t,1})_t$  and  $(\Theta_{t,2})_t$  are in general correlated since they both depend on the second component of  $\mathbf{Z}_t$  unless  $C$  is diagonal. If one wants the two processes in the model to be independent, this is easily done by applying the orthogonal decomposition theorem. This theorem says that for any subspace  $W \subset \mathbb{R}^n$  and  $\mathbf{v} \in \mathbb{R}^n$ , there exist vectors  $\mathbf{w} \in W$  and  $\mathbf{z} \in W^\perp$  such that  $\mathbf{v} = \mathbf{w} + \mathbf{z}$ . We use the theorem by choosing  $W$  to be the subspace that is spanned by  $(\Theta_{t,2})_t$  and  $\mathbf{v} = (\Theta_{t,1})_t$ . The theorem then implies that there exists  $(\Theta_{t,1}^*)_t \in W^\perp$  such that for all  $t$

$$\Theta_{t,1} = \Theta_{t,1}^* + u \Theta_{t,2}\tag{9.7}$$

for some  $u \in \mathbb{R}$ . It is left to find  $u$ . If  $(\Theta_{t,1}^*)_t$  and  $(\Theta_{t,2})_t$  are orthogonal, the new Cholesky factor  $C^*$  needs to be diagonal. If we set  $C = (C_{ij})_{i,j=1,2}$ , we see that the off diagonal element of  $C^*$  is given by  $C_{12} - u C_{22}$ . Therefore, we must choose  $u = \frac{C_{12}}{C_{22}}$  which implies that  $\Theta_{t,1}^*$  is well-defined since  $C_{22} > 0$ .

To get forecasted values for  $\Theta_t$ , we need to know  $\boldsymbol{\mu}$  and  $V$ . There exist two ways to achieve this. The first possibility is to calculate estimators for  $\boldsymbol{\mu}$  and  $V$  similar as in the last section and plug them into (9.6). This means, we assume that there is no parameter uncertainty. The alternative is to include parameter uncertainty. This is reasonable since the estimators will never be exactly correct due to the bounded number of data. In this case, we simulate  $\boldsymbol{\mu}$



## 9 Forecasting Mortality Rates

and  $V$  in a Bayesian way. We first have to choose a prior distribution for  $\boldsymbol{\mu}$  and  $V$ . Since  $V$  is positive definite, we get that  $\det(V) > 0$ . Cairns, Blake and Dowd (2006) suggest to take Jeffrey's prior

$$p(\boldsymbol{\mu}, V) \propto \det(V)^{-3/2}. \quad (9.8)$$

We additionally choose  $n$  observations  $\mathbf{B}_t = \boldsymbol{\Theta}_{t+1} - \boldsymbol{\Theta}_t \in \mathbb{R}^2$ . Because of (9.6), this means

$$\mathbf{B}_t \mid \boldsymbol{\mu}, V \sim \mathcal{N}_2(\boldsymbol{\mu}, V) \quad (9.9)$$

for all  $t$ . We define  $B = (\mathbf{B}_1, \dots, \mathbf{B}_n) \in \mathbb{R}^{2 \times n}$ . We need the following definition by Sawyer (2007) to get the posterior distributions for  $\boldsymbol{\mu}$  and  $V$ .

**Definition 9.1.** We choose  $n$  *i.i.d.* random variables  $\mathbf{Y}_i \sim \mathcal{N}_m(\mathbf{0}, \Sigma)$  where  $\Sigma \in \mathbb{R}^{m \times m}$  is a positive definite covariance matrix. Then, the matrix  $W = \sum_{i=1}^n \mathbf{Y}_i \mathbf{Y}_i^\top \in \mathbb{R}^{m \times m}$  has a **Wishart distribution** with  $n$  degrees of freedom and scale matrix  $\Sigma$ . This is denoted by  $\mathcal{W}(\Sigma, n)$ .

This distribution has the following nice property. If we take some matrix  $A \in \mathbb{R}^{k \times m}$  and  $W \sim \mathcal{W}(\Sigma, n)$ , it holds that  $AWA^\top \sim \mathcal{W}(A\Sigma A^\top, n)$ . From this follows that if  $n \geq m$ ,  $W$  is positive definite. In this case, the density for  $W$  is given by

$$f(w, \Sigma, n) = \frac{\det(w)^{(n-m-1)/2} \exp\left(-\frac{1}{2} \operatorname{tr}(w \Sigma^{-1})\right)}{2^{nm/2} \pi^{m(m-1)/4} \det(\Sigma)^{n/2} \prod_{i=1}^m \Gamma\left(\frac{n+1-i}{2}\right)}, \quad (9.10)$$

where  $\Gamma$  denotes the Gamma function. Since the density of a two-dimensional normal distribution is equal to

$$\varphi_2(\mathbf{B}_t \mid \boldsymbol{\mu}, V) = \frac{1}{2\pi \det(V)^{1/2}} \exp\left(-\frac{1}{2} (\mathbf{B}_t - \boldsymbol{\mu})^\top V^{-1} (\mathbf{B}_t - \boldsymbol{\mu})\right), \quad (9.11)$$

we find that the posterior density must satisfy

$$p(\boldsymbol{\mu}, V^{-1} \mid B) \propto p(\boldsymbol{\mu}, V^{-1}) \prod_{t=1}^n \varphi_2(\mathbf{B}_t \mid \boldsymbol{\mu}, V^{-1}). \quad (9.12)$$

$V$  is positive definite and therefore invertible. Hence,  $\varphi_2(\mathbf{B}_t \mid \boldsymbol{\mu}, V^{-1}) = \varphi_2(\mathbf{B}_t \mid \boldsymbol{\mu}, V)$ . The calculations in the paper by Sawyer (2007) lead to the following result.

**Proposition 9.2.** *The posterior distributions of  $\boldsymbol{\mu}$  and  $V^{-1}$  are given by*

$$V^{-1} \mid B \sim \mathcal{W}\left(\frac{1}{n} \bar{V}^{-1}, n-1\right), \quad (9.13)$$

$$\boldsymbol{\mu} \mid V, B \sim \mathcal{N}_2\left(\bar{\boldsymbol{\mu}}, \frac{1}{n} V\right), \quad (9.14)$$

where

$$\bar{\boldsymbol{\mu}} = \frac{1}{n} \sum_{t=1}^n \mathbf{B}_t \quad \text{and} \quad \bar{V} = \frac{1}{n} \sum_{t=1}^n (\mathbf{B}_t - \bar{\boldsymbol{\mu}})(\mathbf{B}_t - \bar{\boldsymbol{\mu}})^\top.$$

## 9 Forecasting Mortality Rates

*Proof.* We calculate the posterior density by hand.

$$\begin{aligned}
p(\boldsymbol{\mu}, V^{-1} | B) &\propto p(\boldsymbol{\mu}, V^{-1}) \prod_{t=1}^n \varphi_2(\mathbf{B}_t | \boldsymbol{\mu}, V) \\
&\propto \det(V^{-1})^{-3/2} \frac{1}{\det(V)^{n/2}} \exp\left(\underbrace{-\frac{1}{2} \sum_{t=1}^n (\mathbf{B}_t - \boldsymbol{\mu})^\top V^{-1} (\mathbf{B}_t - \boldsymbol{\mu})}_{=-\frac{1}{2} \text{tr}(V^{-1}(n\bar{V} + n(\boldsymbol{\mu} - \bar{\boldsymbol{\mu}})(\boldsymbol{\mu} - \bar{\boldsymbol{\mu}})^\top))}\right) \\
&\propto \det(V^{-1})^{(n-4)/2} \exp\left(-\frac{1}{2} \text{tr}(nV^{-1}\bar{V})\right) \frac{1}{\det(V)^{1/2}} \exp\left(-\frac{n}{2} (\boldsymbol{\mu} - \bar{\boldsymbol{\mu}})^\top V^{-1} (\boldsymbol{\mu} - \bar{\boldsymbol{\mu}})\right),
\end{aligned}$$

where in the last step we used that  $\text{tr}(V^{-1}(\boldsymbol{\mu} - \bar{\boldsymbol{\mu}})(\boldsymbol{\mu} - \bar{\boldsymbol{\mu}})^\top) = (\boldsymbol{\mu} - \bar{\boldsymbol{\mu}})^\top V^{-1} (\boldsymbol{\mu} - \bar{\boldsymbol{\mu}})$ . The first part in this equation corresponds to the density of the  $\mathcal{W}\left(\frac{1}{n}\bar{V}^{-1}, n-1\right)$  distribution and the second part, when we assume  $V$  is given, is proportional to the density of a  $\mathcal{N}_2\left(\bar{\boldsymbol{\mu}}, \frac{1}{n}V\right)$  distribution which is exactly what we claimed.  $\blacksquare$

Thanks to this proposition, we are now able to simulate  $\boldsymbol{\mu}$  and  $V$ . For  $V$ , this is done in the following way:

**Algorithm 9.3.**

- 1.) Choose an upper triangular matrix  $S \in \mathbb{R}^{2 \times 2}$  such that  $SS^\top = \frac{1}{n}\bar{V}^{-1}$ .
- 2.) Simulate  $n-1$  i.i.d. vectors  $\mathbf{Y}_i = S\mathbf{Z}_i \in \mathbb{R}^2$ , where  $\mathbf{Z}_i \sim \mathcal{N}_2(\mathbf{0}, I)$ .
- 3.)  $Y = \sum_{i=1}^{n-1} \mathbf{Y}_i \mathbf{Y}_i^\top$  has a  $\mathcal{W}\left(\frac{1}{n}\bar{V}^{-1}, n-1\right)$  distribution.

We then simply set  $V^{-1} = Y$ . This exists since  $V$  is positive definite. As a last step, we need to simulate  $\boldsymbol{\mu}$  from a bivariate normal distribution with mean  $\bar{\boldsymbol{\mu}}$  and covariance matrix  $\frac{1}{n}V$  which is easy since we have already simulated  $V$ .

This was the second possibility of how to find estimators for  $\boldsymbol{\mu}$  and  $V$ . Both methods can now be used to calculate forecasts for  $\Theta_t$ . For  $t > 0$ , we get

$$\tilde{\Theta}_{t+s-1} | \mathcal{F}_s = \hat{\Theta}_{s-1} + t \hat{\boldsymbol{\mu}} + \hat{C} \sum_{u=s}^{t+s-1} \mathbf{Z}_u, \tag{9.15}$$

where  $\mathbf{Z}_u \sim \mathcal{N}_2(\mathbf{0}, I)$  are i.i.d.. This yields

$$\mathbb{E} \left[ \tilde{\Theta}_{t+s-1} | \mathcal{F}_s \right] = \hat{\Theta}_{s-1} + t \hat{\boldsymbol{\mu}} \quad \text{and} \quad \text{Cov} \left( \tilde{\Theta}_{t+s-1} | \mathcal{F}_s \right) = t \hat{V}.$$

In this case, we cannot find a nice expression for the forecasted mortality rates. Therefore, we have to simulate  $q_{x,t+s-1}^\Theta$  to be able to calculate  $\tilde{q}_{x,t+s-1} = \mathbb{E}[q_{x,t+s-1}^\Theta | \mathcal{F}_s]$ .

## 9.4 Adapted Nolfi Models

Lüthy et al. (2001) did not give an explicit method of how one should forecast the mortality rates in the adapted Nolfi models. Hence, we simply apply the same method as in the Cairns-Blake-Dowd model because the rates also depend on two stochastic processes. This means, we forecast  $\boldsymbol{\alpha}_t = (\alpha_{t,1}, \alpha_{t,2})^\top \in \mathbb{R}_+^2$  in all models with a two-dimensional random walk with drift given by

$$\boldsymbol{\alpha}_{t+1} = \boldsymbol{\alpha}_t + \boldsymbol{\mu} + C\mathbf{Z}_t, \quad (9.16)$$

where  $\boldsymbol{\mu} \in \mathbb{R}^2$  is the drift parameter vector,  $C \in \mathbb{R}^{2 \times 2}$  is an upper triangular matrix and  $(\mathbf{Z}_t)_t$  are *i.i.d.* with a two-dimensional standard normal distribution. Since for the first two models, one of the components is equal to 1 for all  $t$ , we simply set the  $i^{\text{th}}$  row of  $\boldsymbol{\mu}$  and  $C$  equal to 0 which then leads to a one-dimensional random walk with drift equivalent to (9.2). The estimation of  $\boldsymbol{\mu}$  and  $C$  are done according to the last section. What remains, though, is to check in each step that the resulting forecasts are positive.

The forecasted mortality rates for the first two models are found analogously as in Section 9.2. If we define  $\hat{C} = (\hat{C}_{ij})_{i,j=1,2}$  we see that for  $t > 0$ ,

$$\tilde{\alpha}_{t+s-1,1} | \mathcal{F}_s = \hat{\alpha}_{s-1,1} + t \hat{\mu}_1 + \hat{C}_{11} \sum_{u=s}^{t+s-1} Z_{u,1} + \hat{C}_{12} \sum_{u=s}^{t+s-1} Z_{u,2}, \quad (9.17)$$

$$\tilde{\alpha}_{t+s-1,2} | \mathcal{F}_s = \hat{\alpha}_{s-1,2} + t \hat{\mu}_2 + \hat{C}_{22} \sum_{u=s}^{t+s-1} Z_{u,2}. \quad (9.18)$$

To make notation easier, we set  $a = -\hat{\lambda}_x(t+s-1-t_0)$  since this is a constant for fixed  $x$ ,  $t$  and  $s$ . For the models (7.1) and (7.2), we define

$$\begin{aligned} \text{Model 1: } \tilde{q}_{x,t+s-1} &= \mathbb{E}[q_{x,t+s-1}^\ominus | \mathcal{F}_s] = q_{x,t_0} \mathbb{E}[\exp(a \tilde{\alpha}_{t+s-1,2}) | \mathcal{F}_s] \\ &= q_{x,t_0} \exp\left(a(\hat{\alpha}_{s-1,2} + t \hat{\mu}_2) + \frac{1}{2} a^2 \hat{C}_{22}^2 t\right), \end{aligned} \quad (9.19)$$

$$\begin{aligned} \text{Model 2: } \tilde{q}_{x,t+s-1} &= \mathbb{E}[q_{x,t+s-1}^\ominus | \mathcal{F}_s] = q_{x,t_0} \exp(a) \mathbb{E}[\tilde{\alpha}_{t+s-1,1} | \mathcal{F}_s] \\ &= q_{x,t_0} (\hat{\alpha}_{s-1,1} + t \hat{\mu}_1) \exp(a). \end{aligned} \quad (9.20)$$

For model (7.3), more calculations are needed. First, we define  $\alpha_{t,1}^* = \alpha_{t,1}^* + \frac{\hat{C}_{12}}{\hat{C}_{22}} \alpha_{t,2}$  due to (9.7) for all  $t$  to get two independent processes. Next, we find that

$$\tilde{\alpha}_{t+s-1,1}^* | \mathcal{F}_s = \hat{\alpha}_{s-1,1} + t \hat{\mu}_1 - \frac{\hat{C}_{12}}{\hat{C}_{22}} (\hat{\alpha}_{s-1,2} + t \hat{\mu}_2) + \hat{C}_{11} \sum_{u=s}^{t+s-1} Z_{u,1}. \quad (9.21)$$

With this, we are able to calculate the forecasted mortality rates.

$$\begin{aligned} \tilde{q}_{x,t+s-1} &= \mathbb{E}[q_{x,t+s-1}^\ominus | \mathcal{F}_s] = q_{x,t_0} \mathbb{E}[\tilde{\alpha}_{t+s-1,1}^* \exp(a \tilde{\alpha}_{t+s-1,2}) | \mathcal{F}_s] \\ &= q_{x,t_0} \left( \frac{\hat{C}_{12}}{\hat{C}_{22}} \mathbb{E}[\tilde{\alpha}_{t+s-1,2} \exp(a \tilde{\alpha}_{t+s-1,2}) | \mathcal{F}_s] \right. \\ &\quad \left. + \underbrace{\mathbb{E}[\tilde{\alpha}_{t+s-1,1}^* | \mathcal{F}_s] \mathbb{E}[\exp(a \tilde{\alpha}_{t+s-1,2}) | \mathcal{F}_s]} \right) \\ &= \left( \hat{\alpha}_{s-1,1} + t \hat{\mu}_1 - \frac{\hat{C}_{12}}{\hat{C}_{22}} (\hat{\alpha}_{s-1,2} + t \hat{\mu}_2) \right) \exp\left(a(\hat{\alpha}_{s-1,2} + t \hat{\mu}_2) + \frac{1}{2} a^2 \hat{C}_{22}^2 t\right) \end{aligned}$$

## 9 Forecasting Mortality Rates

It remains to calculate the first conditional expectation. We do this in an intermediate step.

$$\begin{aligned}
 \mathbb{E}[\tilde{\alpha}_{t+s-1,2} \exp(a \tilde{\alpha}_{t+s-1,2}) \mid \mathcal{F}_s] &= \frac{1}{\hat{C}_{22} \sqrt{2\pi t}} \int_{-\infty}^{\infty} x e^{ax} \exp\left(-\frac{1}{2} \frac{(x - \hat{\alpha}_{s-1,2} - t \hat{\mu}_2)^2}{\hat{C}_{22}^2 t}\right) dx \\
 &= \exp\left(a (\hat{\alpha}_{s-1,2} + t \hat{\mu}_2) + \frac{1}{2} a^2 \hat{C}_{22}^2 t\right) \\
 &\quad \underbrace{\frac{1}{\hat{C}_{22} \sqrt{2\pi t}} \int_{-\infty}^{\infty} x \exp\left(-\frac{1}{2} \frac{(x - \hat{\alpha}_{s-1,2} - t \hat{\mu}_2 - t a \hat{C}_{22}^2)^2}{\hat{C}_{22}^2 t}\right) dx}_{= \hat{\alpha}_{s-1,2} + t \hat{\mu}_2 + t a \hat{C}_{22}^2}.
 \end{aligned}$$

Finally, when putting everything together, we get

$$\tilde{q}_{x,t+s-1} = q_{x,t_0} \left( \hat{\alpha}_{s-1,1} + t \hat{\mu}_1 + t a \hat{C}_{12} \hat{C}_{22} \right) \exp\left(a (\hat{\alpha}_{s-1,2} + t \hat{\mu}_2) + \frac{1}{2} a^2 \hat{C}_{22}^2 t\right). \quad (9.22)$$

We have explained how one can find mortality rates for the future in all models we have introduced. To see how good these procedures are, one can have a look at the paper by Dowd et al. (2010). They performed various backtests to compare the quality of the forecasts for the classical model of Lee and Carter to the one of the Cairns-Blake-Dowd model. They concluded that the procedures work rather nicely. Hence, we finish the discussion of extrapolating the data with respect to age and time and go on to show how one can find mortality rates that are more cautious than the ones we have discussed so far.

# 10 Mortality Rates of Order 1

After all models have been introduced, fitted and forecasted, we need to keep in mind that there is risk involved in all these procedures. The sources of risk, however, are different. We quickly go over the most important ones here.

## Definition 10.1.

- (a) All models depend on the annual mortality rate  $Q_{x,t}$ . Since this is a random variable, we are facing **residual risk**. This means that even if we have chosen the correct model, the number of deaths still fluctuates randomly.
- (b) We have chosen  $Q_{x,t} \sim F$  so we risk having chosen the distribution  $F$  inappropriately. This is called **model risk**.
- (c) We assumed that  $F = F_\theta$  is parametric. Hence, we need to find an estimator for  $\theta$ . It is possible that even though we have chosen  $F$  correctly, we estimate  $\theta$  wrongly which still yields a false distribution. We call this **parameter risk**.
- (d) Finally, it is possible that  $\theta$  is in fact a stochastic process  $(\theta_t)_t$  which generates a trend in the data. Since we need to model this trend, we are facing **trend risk**.

Because of these different sources of risk, we would like to control the mean squared error of prediction of the forecasted mortality rates. However, we restrict ourselves to only considering one period of time since controlling more periods at once would require much more work. We find that the mean squared error of prediction for year  $s$  is given by

$$\begin{aligned} \text{mse}_{\sum_x Q_{x,s} | \mathcal{F}_s} \left( \sum_x \tilde{q}_{x,s} \mid \mathcal{F}_s \right) &= \mathbb{E} \left[ \left( \sum_x Q_{x,s} - \sum_x \tilde{q}_{x,s} \right)^2 \mid \mathcal{F}_s \right] \\ &= \text{Var} \left( \sum_x Q_{x,s} \mid \mathcal{F}_s \right) + \left( \sum_x (\mathbb{E}[Q_{x,s} \mid \mathcal{F}_s] - \tilde{q}_{x,s}) \right)^2, \end{aligned} \quad (10.1)$$

since  $\tilde{q}_{x,s}$  is an  $\mathcal{F}_s$ -measurable estimator for the conditional mean  $\mathbb{E}[Q_{x,s} \mid \mathcal{F}_s]$ . The first term corresponds to the residual risk since it measures the variability of the true rates in year  $s$ . The second term contains the model, parameter and trend risk because it measures the error between the expected and the estimated mortality rate. We try to control these sources of risk separately.

## 10.1 Residual Risk for one Period of Time

This section is split up into two parts. First, we quantify the residual risk for a fixed age  $x$  and afterwards, we do the same for the whole portfolio. Koller (2003) suggests to calculate a mortality loading  $m_s \in \mathbb{R}_+$  such that the probability that the forecasted mortality rate is smaller than the true one plus this loading is as large as possible. In other words, we try to find  $m_s$  such that

$$\mathbb{P}[\tilde{q}_{x,s} \leq Q_{x,s} + m_s] = \mathbb{P}[D_{x,s} \geq l_{x,s} (\tilde{q}_{x,s} - m_s)] \geq \alpha \quad (10.2)$$

## 10 Mortality Rates of Order 1

for some  $\alpha \in (0, 1)$ . By subtracting the mortality loading from the forecasted mortality rate, we assume that the person lives longer than he actually does. We could also add  $m_s$  to assume that a person dies earlier if we want to model the event of death. To be able to calculate (10.2), we need the approximation given in the following proposition.

**Proposition 10.2.** *For any random variable  $X \sim \text{Binom}(n, p)$  with  $n \in \mathbb{N}$  and  $p \in [0, 1]$ , it holds that*

$$S_n = \frac{X - np}{\sqrt{np(1-p)}} \xrightarrow{d} \mathcal{N}(0, 1). \quad (10.3)$$

*Proof.* The moment generating function of  $X$  is given by  $M_X(t) = (1 - p + pe^t)^n$ . For a standard normally distributed random variable  $Y$ , this function is equal to  $M_Y(t) = \exp\left(\frac{t^2}{2}\right)$ . We want to show that  $\lim_{n \rightarrow \infty} \log(M_{S_n}(t)) = \log(M_Y(t))$  since the uniqueness of the moment generating function then implies that  $S_n \xrightarrow{d} Y$ . We start by finding an expression for  $\log(M_{S_n}(t))$ .

$$\begin{aligned} \log(M_{S_n}(t)) &= \log\left(\mathbb{E}\left[\exp\left(t \frac{X - np}{\sqrt{np(1-p)}}\right)\right]\right) \\ &= -\frac{npt}{\sqrt{np(1-p)}} + \log\left(M_X\left(\frac{t}{\sqrt{np(1-p)}}\right)\right) \\ &= n\left(\log\left(1 - p + p \exp\left(\frac{t}{\sqrt{np(1-p)}}\right)\right) - \frac{pt}{\sqrt{np(1-p)}}\right). \end{aligned}$$

This means, we need to calculate

$$\lim_{n \rightarrow \infty} \log(M_{S_n}(t)) = \lim_{n \rightarrow \infty} \frac{\log\left(1 - p + p \exp\left(\frac{t}{\sqrt{np(1-p)}}\right)\right) - \frac{pt}{\sqrt{np(1-p)}}}{n^{-1}}.$$

For  $n \rightarrow \infty$  the nominator as well as the denominator go to zero. Hence, we can apply l'Hôpital's rule and get

$$\begin{aligned} \lim_{n \rightarrow \infty} \log(M_{S_n}(t)) &= \lim_{n \rightarrow \infty} \frac{t\sqrt{p(1-p)} \frac{\sqrt{n}\left(\exp\left(\frac{t}{\sqrt{np(1-p)}}\right) - 1\right)}{1 - p + p \exp\left(\frac{t}{\sqrt{np(1-p)}}\right)}}{2} \\ &= \frac{t\sqrt{p(1-p)}}{2} \lim_{n \rightarrow \infty} \frac{\exp\left(\frac{t}{\sqrt{np(1-p)}}\right) - 1}{n^{-1/2}\left(1 - p + p \exp\left(\frac{t}{\sqrt{np(1-p)}}\right)\right)}. \end{aligned}$$

We again find that the nominator as well as the denominator go to zero for  $n \rightarrow \infty$ . So we apply l'Hôpital's rule a second time and find

$$\begin{aligned} &\lim_{n \rightarrow \infty} \log(M_{S_n}(t)) \\ &= \frac{t\sqrt{p(1-p)}}{2} \lim_{n \rightarrow \infty} \frac{t}{\sqrt{p(1-p)}} \frac{\exp\left(\frac{t}{\sqrt{np(1-p)}}\right)}{1 - p + p \exp\left(\frac{t}{\sqrt{np(1-p)}}\right) + \frac{t}{\sqrt{np(1-p)}} \exp\left(\frac{t}{\sqrt{np(1-p)}}\right)} \\ &= \frac{t^2}{2} = \log(M_Y(t)). \end{aligned}$$

This is what we wanted to show so the proof is done. ■

## 10 Mortality Rates of Order 1

Hence, we approximate  $D_{x,s}$  by  $D_{x,s}^* \sim \mathcal{N}(l_{x,s} \tilde{q}_{x,s}, l_{x,s} \tilde{q}_{x,s}(1 - \tilde{q}_{x,s}))$ . Stahel (1999) argues that this approximation is good if  $l_{x,s} \tilde{q}_{x,s}(1 - \tilde{q}_{x,s}) \geq 9$ . This yields

$$\begin{aligned} \mathbb{P}[D_{x,s} \geq l_{x,s} (\tilde{q}_{x,s} - m_s)] &\approx \mathbb{P} \left[ \frac{D_{x,s}^* - l_{x,s} \tilde{q}_{x,s}}{\sqrt{l_{x,s} \tilde{q}_{x,s}(1 - \tilde{q}_{x,s})}} \geq -\frac{l_{x,s} m_s}{\sqrt{l_{x,s} \tilde{q}_{x,s}(1 - \tilde{q}_{x,s})}} \right] \\ &= \Phi \left( \frac{l_{x,s} m_s}{\sqrt{l_{x,s} \tilde{q}_{x,s}(1 - \tilde{q}_{x,s})}} \right) \geq \alpha, \end{aligned}$$

which holds true for

$$m_s = \Phi^{-1}(\alpha) \sqrt{\frac{\tilde{q}_{x,s}(1 - \tilde{q}_{x,s})}{l_{x,s}}}. \quad (10.4)$$

We want to know how  $m_s$  behaves for changes in the portfolio size or the mortality rate. To see this, we give a short example.

**Example 10.3.** We choose  $\tilde{q}_{x,s} = 0.01s$  and  $l_{x,s} = 1000s$  for  $s$  from 1 to 100. We first fix  $l_{x,s} = 5'000$  and vary  $\tilde{q}_{x,s}$ . Then, we set  $\tilde{q}_{x,s} = 0.3$  and use the whole series  $(l_{x,s})_s$  for the calculation. To interpret the plots in Figure 10.1 properly, we have a look at the slope of  $m_s$  when we change either  $\tilde{q}_{x,s}$  or  $l_{x,s}$  by a small amount. We get

$$\begin{aligned} \frac{\partial m_s}{\partial l_{x,s}} &= \frac{-\Phi^{-1}(\alpha) \sqrt{\tilde{q}_{x,s}(1 - \tilde{q}_{x,s})}}{2} \frac{1}{\sqrt{l_{x,s}^3}}, \\ \frac{\partial m_s}{\partial \tilde{q}_{x,s}} &= \frac{\Phi^{-1}(\alpha)}{2\sqrt{l_{x,s}}} \frac{1 - 2\tilde{q}_{x,s}}{\sqrt{\tilde{q}_{x,s}(1 - \tilde{q}_{x,s})}}. \end{aligned}$$

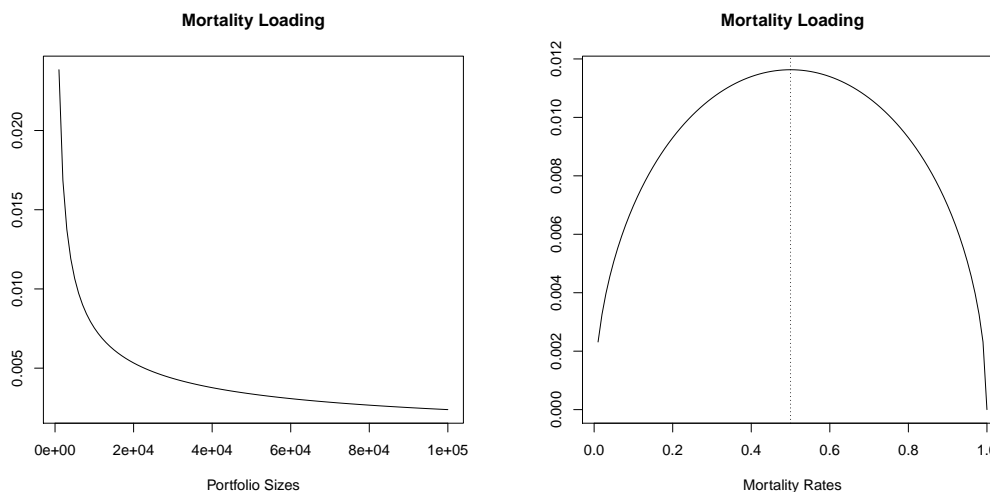


Figure 10.1: Mortality loading  $m_s$ : (LHS) with respect to  $l_{x,s}$ , (RHS) with respect to  $\tilde{q}_{x,s}$

We see that the derivative with respect to the portfolio size is always negative. Therefore,  $m_s$  decreases as  $l_{x,s}$  increases. This is also observed on the left-hand side of the figure. The derivative with respect to  $\tilde{q}_{x,s}$ , on the other hand, has not always the same sign. It is positive for rates  $\tilde{q}_{x,s} < 0.5$  and negative for  $\tilde{q}_{x,s} > 0.5$ . Therefore, the mortality loading increases for small rates and decreases again for rates above 0.5. This is clearly visible in the plot on the right-hand side of the figure. ■

## 10 Mortality Rates of Order 1

Generally, we are interested in mortality loadings for the whole portfolio and not only for a fixed age. The procedure is similar to the one we have just discussed. We denote the true and approximated total number of deaths in year  $s$  by

$$D_s = \sum_x D_{x,s} \quad \text{and} \quad D_s^* = \sum_x D_{x,s}^*. \quad (10.5)$$

In contrast to the former section, the mortality loading  $m_{x,s} \in \mathbb{R}_+$  now also depends on age. We define the mortality rate of order 1 by

$$q_{x,s}^1 = \tilde{q}_{x,s} - m_{x,s}. \quad (10.6)$$

We require that  $\mathbb{P}[D_{x,s} \geq l_{x,s} q_{x,s}^1] \geq \beta$  holds true for  $\beta \in (0, 1)$  for all  $x$  in a simultaneous sense, meaning for the whole portfolio. Hence, we require

$$\begin{aligned} \mathbb{P}[D_s \geq \sum_x l_{x,s} q_{x,s}^1] &\approx \mathbb{P}\left[\frac{D_s^* - \sum_x l_{x,s} \tilde{q}_{x,s}}{\sqrt{\text{Var}(D_s^*)}} \geq -\frac{\sum_x l_{x,s} m_{x,s}}{\sqrt{\text{Var}(D_s^*)}}\right] \\ &= \Phi\left(\frac{\sum_x l_{x,s} m_{x,s}}{\sqrt{\text{Var}(D_s^*)}}\right) \geq \alpha \end{aligned}$$

for some  $\alpha \in (0, 1)$ . Hence, it must hold that  $\sum_x l_{x,s} m_{x,s} \geq \Phi^{-1}(\alpha) \sqrt{\text{Var}(D_s^*)}$  for all  $t$ . To get a unique solution for  $m_{x,s}$ , we simplify the problem and assume that  $\mathbb{P}[D_{x,s} \geq l_{x,s} q_{x,s}^1]$  is constant for all  $x$ . Of course, it would also be possible to make another assumption here. This yields

$$\begin{aligned} \mathbb{P}[D_{x,s} \geq l_{x,s} q_{x,s}^1] &\approx \mathbb{P}\left[\frac{D_{x,s}^* - l_{x,s} \tilde{q}_{x,s}}{\sqrt{l_{x,s} \tilde{q}_{x,s} (1 - \tilde{q}_{x,s})}} \geq -\frac{l_{x,s} m_{x,s}}{\sqrt{l_{x,s} \tilde{q}_{x,s} (1 - \tilde{q}_{x,s})}}\right] \\ &= \Phi\left(\frac{l_{x,s} m_{x,s}}{\sqrt{l_{x,s} \tilde{q}_{x,s} (1 - \tilde{q}_{x,s})}}\right) \geq \beta, \end{aligned}$$

so  $l_{x,s} m_{x,s} \geq \Phi^{-1}(\beta) \sqrt{l_{x,s} \tilde{q}_{x,s} (1 - \tilde{q}_{x,s})}$  must hold. If we combine these two results and use the fact that  $D_{x,s}^*$  are independent for  $s$  fixed, we see that

$$\Phi^{-1}(\beta) = \Phi^{-1}(\alpha) \frac{\sqrt{\sum_z \text{Var}(D_{z,s}^*)}}{\sum_z \sqrt{l_{z,s} \tilde{q}_{z,s} (1 - \tilde{q}_{z,s})}}$$

satisfies both conditions. This finally leads to

$$m_{x,s} = \Phi^{-1}(\alpha) \sqrt{\frac{\tilde{q}_{x,s} (1 - \tilde{q}_{x,s})}{l_{x,s}} \frac{\sqrt{\sum_z l_{z,s} \tilde{q}_{z,s} (1 - \tilde{q}_{z,s})}}{\sum_z \sqrt{l_{z,s} \tilde{q}_{z,s} (1 - \tilde{q}_{z,s})}}}, \quad (10.7)$$

and therefore

$$q_{x,s}^1 = \tilde{q}_{x,s} - \Phi^{-1}(\alpha) \sqrt{\frac{\tilde{q}_{x,s} (1 - \tilde{q}_{x,s})}{l_{x,s}} \frac{\sqrt{\sum_z l_{z,s} \tilde{q}_{z,s} (1 - \tilde{q}_{z,s})}}{\sum_z \sqrt{l_{z,s} \tilde{q}_{z,s} (1 - \tilde{q}_{z,s})}}}. \quad (10.8)$$

In this section, we have discussed how the residual risk can be estimated. Hence, it is left to have a look at the model and parameter risk of the forecasting and estimation procedures.



## 10.2 Model and Parameter Risk for one Period of Time

We need to estimate  $(\sum_x (\mathbb{E}[Q_{x,s} | \mathcal{F}_s] - \tilde{q}_{x,s}))^2$  to quantify the model, parameter and trend uncertainty. For simplicity, we will omit the trend risk here. We find that  $\mathbb{E}[\tilde{q}_{x,s}] = \mathbb{E}[q_{x,s}^\ominus] = q_{x,s}$  if  $\tilde{q}_{x,s}$  is an unbiased estimator. Hence, we take

$$\text{Var} \left( \sum_x \tilde{q}_{x,s} \mid \mathcal{F}_s \right) \quad (10.9)$$

as an estimator for the uncertainty since we are interested in the variance of  $\tilde{q}_{x,s}$  around  $q_{x,s}$ . However, this variance cannot be calculated directly which is why we estimate it by  $\text{Var} \left( \sum_x q_{x,s}^\ominus \mid \mathcal{F}_s \right)$ . This can now easily be evaluated using the results from Chapter 9. As an example, we show it for the classical model of Lee and Carter.

**Example 10.4.** We define  $q_{x,s}^\ominus = \exp(\hat{a}_x + \hat{b}_x \tilde{k}_s)$ . In Chapter 9, we have assumed that  $\tilde{k}_s | \mathcal{F}_s \sim \mathcal{N}(\hat{k}_{s-1} + \mu, c^2)$ . Since we are ignoring the trend risk,  $\mu$  and  $c$  are fixed. We find

$$\text{Var} \left( \sum_x q_{x,s}^\ominus \mid \mathcal{F}_s \right) = \sum_x \text{Var} \left( q_{x,s}^\ominus \mid \mathcal{F}_s \right) + 2 \sum_{x \neq z} \text{Cov} \left( q_{x,s}^\ominus, q_{z,s}^\ominus \mid \mathcal{F}_s \right).$$

We calculate these two terms individually and get the following:

$$\begin{aligned} \text{Var} \left( q_{x,s}^\ominus \mid \mathcal{F}_s \right) &= \exp \left( 2 \hat{a}_x + 2 \hat{b}_x (\hat{k}_{s-1} + \mu) + \hat{b}_x^2 c^2 \right) \left( \exp \left( \hat{b}_x^2 c^2 \right) - 1 \right), \\ \text{Cov} \left( q_{x,s}^\ominus, q_{z,s}^\ominus \mid \mathcal{F}_s \right) &= \exp \left( \hat{a}_x + \hat{a}_z + (\hat{b}_x + \hat{b}_z)(\hat{k}_{s-1} + \mu) + \frac{1}{2}(\hat{b}_x^2 + \hat{b}_z^2)c^2 \right) \\ &\quad \left( \exp \left( \hat{b}_x \hat{b}_z c^2 \right) - 1 \right). \end{aligned}$$

So we have estimated the model and parameter uncertainty for year  $s$  in the classical model of Lee and Carter. ■

We now know everything we need to be able to apply these models to real data. This means, we know how to estimate the parameters to get mortality rates for the past and how to make forecasts for the future. Before we actually do this, we define in the next chapter some measures for longevity with which we can compare the mortality rate models to each other.

# 11 Measures for Longevity

We want to know how much the longevity of a society increases over the years. Therefore, we need to define measures with which this is calculable. First, we clarify what we mean by longevity.

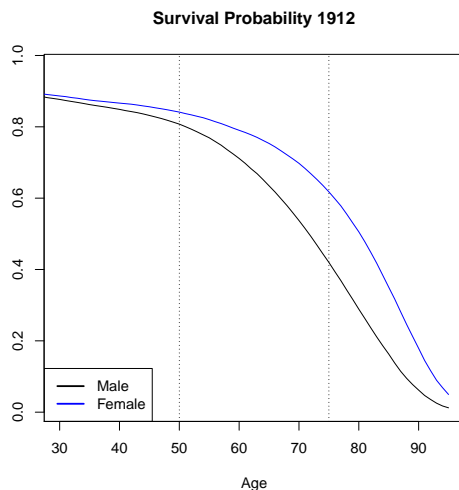


Figure 11.1: Survival probability  $p_{0,t,j}$

In Figure 11.1, we show the survival probability  $p_{0,t,j}$  of a person born in Switzerland in 1912, see (11.1) below. We are interested in the age where the kink in this curve, which is marked by the dotted lines, happens. This is the range where a miscalculation of mortality has a huge impact on values of life insurance portfolios. Note that longevity means that this kink will move to the right and it is crucial to know at which speed. The goal of this chapter is to find measures that give us a hint at which age the mortality rates start increasing rapidly. One measure for this is the survival probability we have just discussed. Another straightforward measure is the life expectancy since this tells us the number of years a person is expected to live on. We can also price mortality-linked securities and check how sensitive their prices are to changes in mortality rates. In this context, we will have a look at annuities.

## 11.1 Life Expectancy

The life expectancy  $e_{x,t} \in \mathbb{R}_+$  of a person of age  $x$  in year  $t$  denotes the number of years we expect this person to live on when we know he already survived to age  $x$ . In other words, we anticipate that this person dies on average at an age of  $x + e_{x,t}$ . For its definition, we need the probability  $p_{x,t,j} \in [0, 1]$  for this person to survive to age  $x + j$ . This means

$$p_{x,t,j} = (1 - q_{x,t})(1 - q_{x+1,t+1}) \cdots (1 - q_{x+j,t+j}) = \prod_{i=0}^j (1 - q_{x+i,t+i}). \quad (11.1)$$

This can also be defined in terms of the remaining life time  $T_{x,t}$  of a person aged  $x$  in year  $t$ . It must hold that

## 11 Measures for Longevity

$$\mathbb{P}[T_{x,t} > j] = p_{x,t,j}. \quad (11.2)$$

The life expectancy is then equal to the expected value of the remaining life time. Since  $T_{x,t}$  is a discrete random variable, we get

$$e_{x,t} = \mathbb{E}[T_{x,t}] = \sum_{j=1}^{\infty} j \mathbb{P}[T_{x,t} = j] = \sum_{j=1}^{\infty} j p_{x,t,j-1} q_{x+j,t+j}, \quad (11.3)$$

where we used that  $\mathbb{P}[T_{x,t} = j]$  is equal to the probability that the person survives to age  $x + j - 1$  and dies in year  $t + j$  at age  $x + j$ . We are mostly interested in the life expectancies for people aged 30 and 65 since longevity is visible in the tail of the mortality curve and not in the beginning where we often have untypical variability.

We go on and discuss how to measure longevity by pricing mortality-linked securities, in particular annuities.

### 11.2 Pricing Annuities

An annuity is a financial instrument that can be purchased by a person of age  $x$  for a fixed amount of money. Starting after the first year, this person receives each year predefined payments until a fixed maturity  $M \in \mathbb{N}$  or his time of death. These payments can either be fixed or variable.

We assume here that the payments are variable. This means that someone of age  $x$  receives in year  $t$  a predefined amount of capital  $C(t) \in \mathbb{R}_+$  if he is still alive by then. We additionally say that the interest rate  $r \in [0, 1)$  is constant over time. The present value of this annuity is then given by

$$PV(x, t) = \sum_{j=1}^M \frac{C(t+j)}{(1+r)^j} p_{x,t,j}. \quad (11.4)$$

This means, we add up the discounted payments times the probability that the person is still alive at age  $x + j$ . For simplicity, we may set  $C(t) = 1$  for all  $t$  indicating a fixed payment.

As we have seen, the price of an annuity is rather easily calculated. We could also exchange in (11.3) as well as in (11.4) the mortality and survival rates by the ones of order 1 and check how this influences the price. This ends the first part of this thesis. We now start applying the models to real data.



## Part II

# Construction and Comparison of Mortality Tables

## 12 Description of the Data

The aim of the second part of this thesis is to apply all models and methods introduced in Part I to real data. We start with a short description of the data we will use for this analysis.

The data we have at our disposal are stored in 4 tables containing the annual number of inhabitants and the annual number of deaths of the Swiss population per age and year both for men and women separately. The number of inhabitants was measured in the beginning of the year and the number of deaths at the end of the year. These two numbers are available for ages  $x \in \{0, 1, \dots, 99\}$  and years  $t \in \{1912, 1913, \dots, 2010\}$ . This means, we have a total of 9'900 data points per gender. The data were collected by the Human Mortality Database and the Swiss Federal Statistical Office and are already roughly adjusted for data errors.

Unfortunately, it happened in some of the earlier years that the observed mortality rates for some age groups were equal to 0. On the other hand, when the population size was small, it also happened that the mortality rates were equal to 1. Since both these cases are not reasonable, we decided to only apply the data for ages from 0 to 95 since this is the range where these two extrema did not occur. Nevertheless, a few rates for small ages were also equal to 0. Since for us, the critical rates are only those above age 65, these rates are not too important and so we decided to set them equal to 0.1% and adjusted the number of deaths accordingly. Thus, we are now left with 19'008 data points in total. How they are distributed for men and women is visualized in the following levelplots.

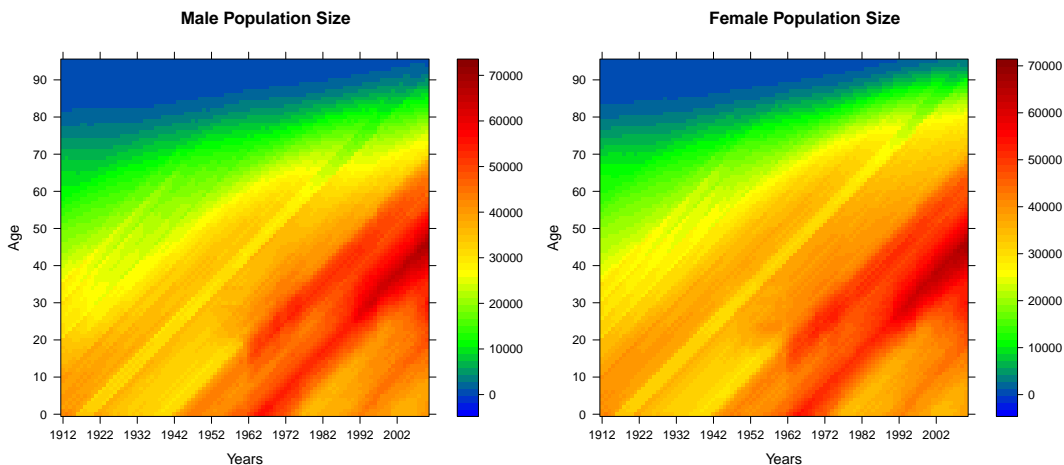


Figure 12.1: Population size  $l_{x,t}$

Figure 12.1 shows how the population size of men and women in Switzerland is distributed. We see that more people achieved an age above 60 in 2010 than in 1912 because of the triangular blue area in the top left corner. We can see the years of the baby-boom generation starting in 1962 where more babies were born than in any other year. This is the reason why there is a growing number of people between ages 30 and 50 in the years after 1990. But the boom ended rapidly around 1970 when the birth control pill was invented. This changed

## 12 Description of the Data

the age structure of society from a population pyramid in 1912, where the population size decreased for an increasing age, to a more urn shaped form in 2010, which means that there are fewer people born than having an age from 30 to 50.

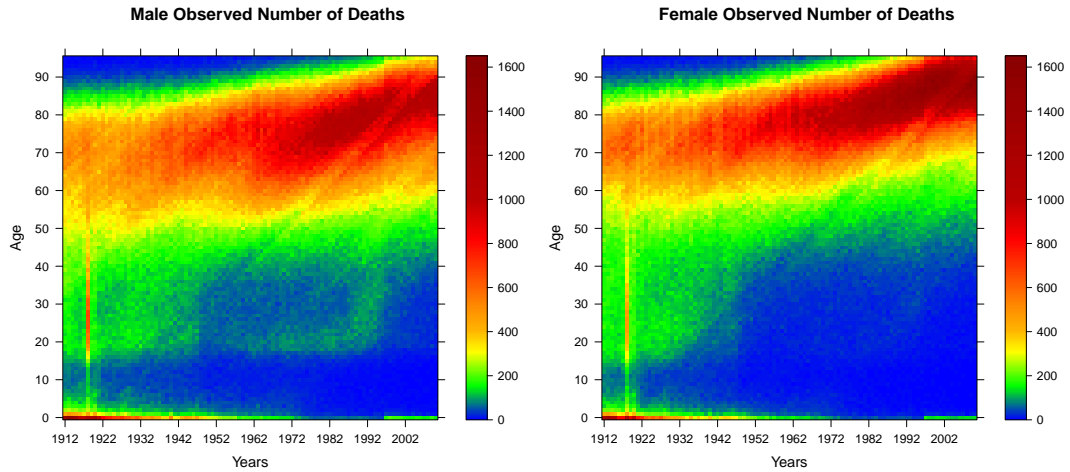


Figure 12.2: Observed number of deaths  $d_{x,t}$

Figure 12.2 gives an insight into the observed number of deaths per year. It is again visible that people lived longer in 2010 than in 1912 due to the diagonal shift of the red area on top. What is most surprising, though, is the red peak in 1918 that indicates a much higher death rate than in any other year. This is caused by the Spanish flu that caused many deaths in that year. We also see that the small green area in the lower left corner nearly disappeared after 1950 which means that the infant death rate decreased significantly.

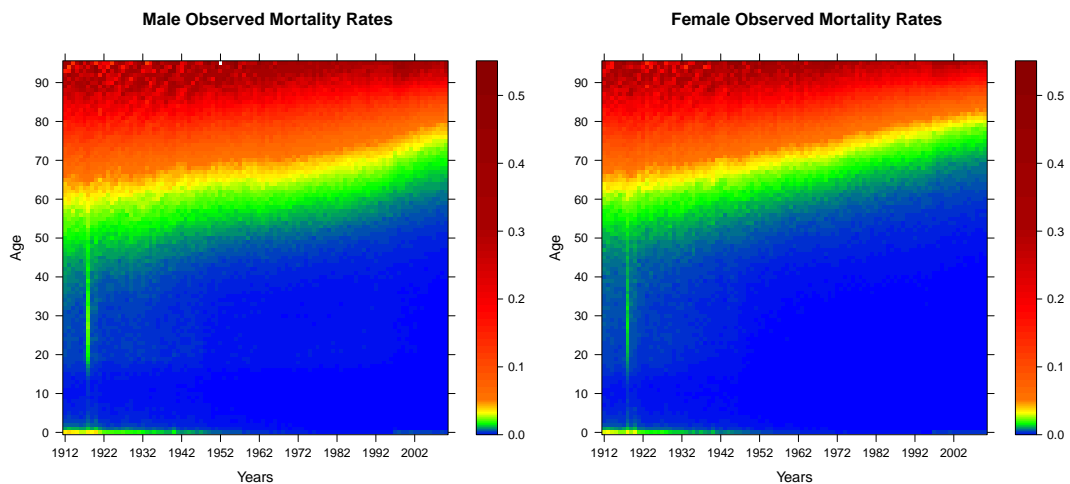


Figure 12.3: Observed mortality rates  $\frac{d_{x,t}}{l_{x,t}}$

Figure 12.3 shows the observed mortality rates for men and women and in Figure 12.4, we see the logit-transforms thereof. In both figures, we observe the increasing age of both men and women. The Spanish flu and the decreasing infant mortality rates are again clearly visible.

## 12 Description of the Data

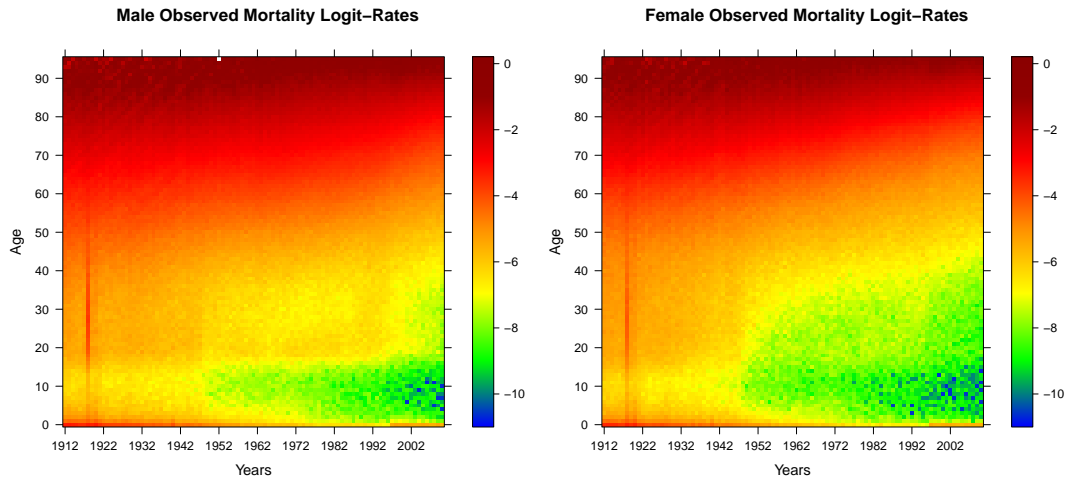


Figure 12.4: Observed mortality logit-rates

What was not observed in the other plots, however, is the rather large blue and green area in the lower right corner of Figure 12.4 indicating a very low mortality rate for ages 0 to 10 in the years after 1950. This might be caused by the improved hygiene and new medical innovations.

From now on, all calculations will be done for men and women separately and all models will be fitted for all ages and all years even though we sometimes compare the estimators of groups of years or ages to each other. In those cases, the aggregation of the results will be carried out only after the estimation.



# 13 Application of Common Mortality Rate Models

In the following data analysis, we will always do the same steps for each model. We first fit the model to the data to get estimators for the parameters. With these, we calculate the estimated mortality rates and their logit-transforms. After having a look at the residuals from this procedure, we go on and calculate forecasted mortality rates. Finally, we extend the mortality rates with respect to age to show how the survival probability, the life expectancy and the price of an annuity change over the years.

In the first application, we have a look at the common mortality rate models introduced in Chapter 3. We will not discuss the Nolfi model since the generalized Nolfi model is already an improved version of it. The combined model will also be left out because we think that it is enough to discuss models with two age-specific constants since an additional one would not improve much.

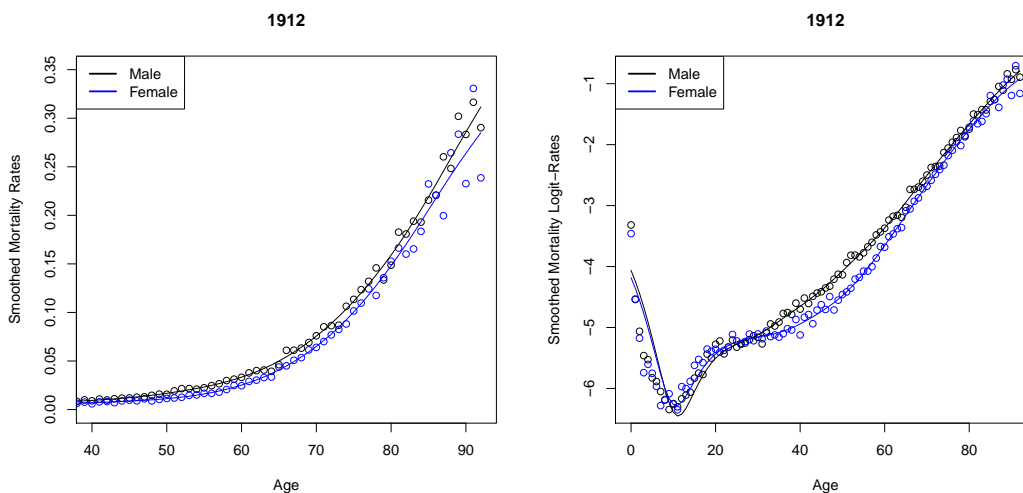


Figure 13.1: Smoothed mortality rates in 1912

Throughout this whole chapter, we take  $t_0 = 1912$ . For  $q_{x,t_0}$ , we choose the smoothed mortality rates in year  $t_0$  from the clearing method by Whittaker-Henderson explained in Section 2.1. For constant weights, we see from Figure 13.1 that the choices  $v = 2$  and  $g = 1.5$  yield appropriate adjusted rates, at least for ages above 20.

## 13.1 Generalized Nolfi Model

The generalized Nolfi model only depends on one age-specific constant and the expected mortality rates are assumed to grow exponentially. This means

$$q_{x,t} = q_{x,t_0} \exp(-\lambda_x (t - t_0)). \tag{13.1}$$

**Fitting Procedure**

The first step we do is to estimate  $\lambda_x$  with an ordinary linear regression for all  $x$ . Unfortunately, we find that  $\hat{\lambda}_x$  is negative for  $x \geq 93$  which is not allowed. Hence, we can only use the data up to age 92 in this application.

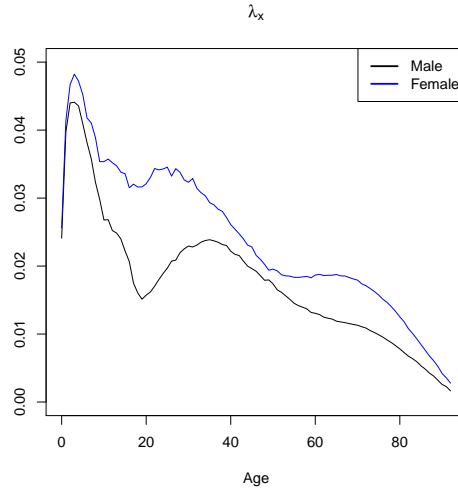


Figure 13.2:  $\hat{\lambda}_x$

In Figure 13.2, we see the estimator  $\hat{\lambda}_x$  for the male and female data. As we would expect, its trend is decreasing since this means, the mortality rates generally increase with respect to age. The only rather long deviation from the trend is found for men between the ages 20 and 40. This might be caused by car driving. We additionally observe that the estimator for the females is always larger than the one for the male data. This is clear, though, since women are expected to live longer than men over the whole range of ages.

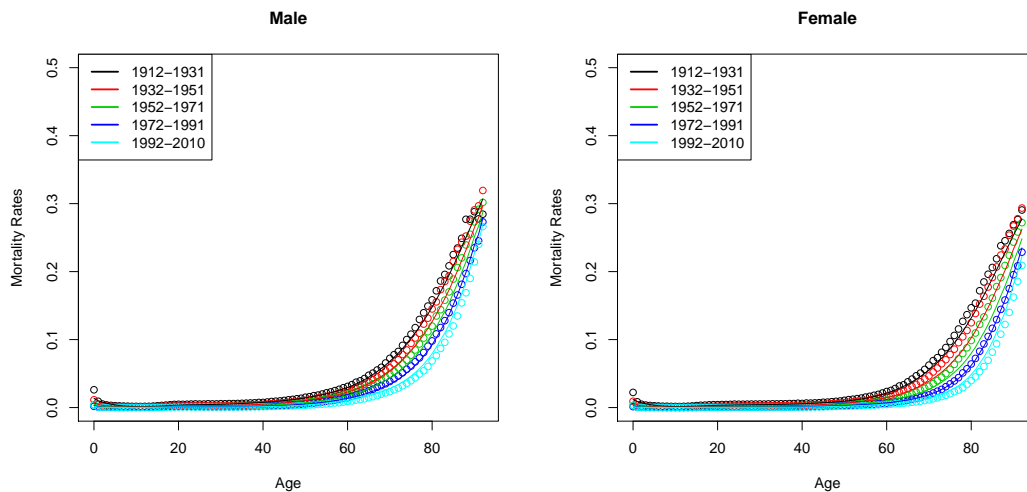


Figure 13.3: Observed (dots) and estimated (lines) mortality rates

We can now use  $(\hat{\lambda}_x)_x$  to estimate the mortality rates from 1912 to 2010. In Figure 13.3, we grouped the data into 5 groups, where each but the last one contain the data of 20 successive

### 13 Application of Common Mortality Rate Models

years, and calculated the average rates for each of them. We observe in this plot that the estimated mortality curves have a nice shape but we already see that they differ slightly from the observed ones.

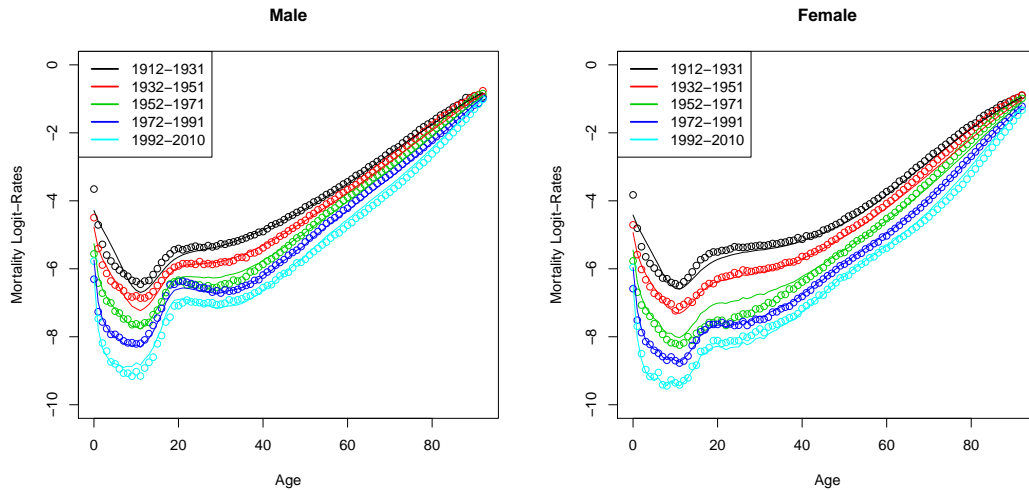


Figure 13.4: Observed (dots) and estimated (lines) mortality logit-rates

Next, we calculate the mortality logit-rates to see the difference between the observed and the estimated rates even better. We find that for ages above 40, we cannot see much from this plot. For ages 0 to 40, we clearly observe some differences between the two rates, especially for the earliest three year groups. The deviance for ages 0 to 20 might be caused by the smoothing of  $q_{x,t_0}$ . Overall, the general shape of the curve is again replicated nicely.

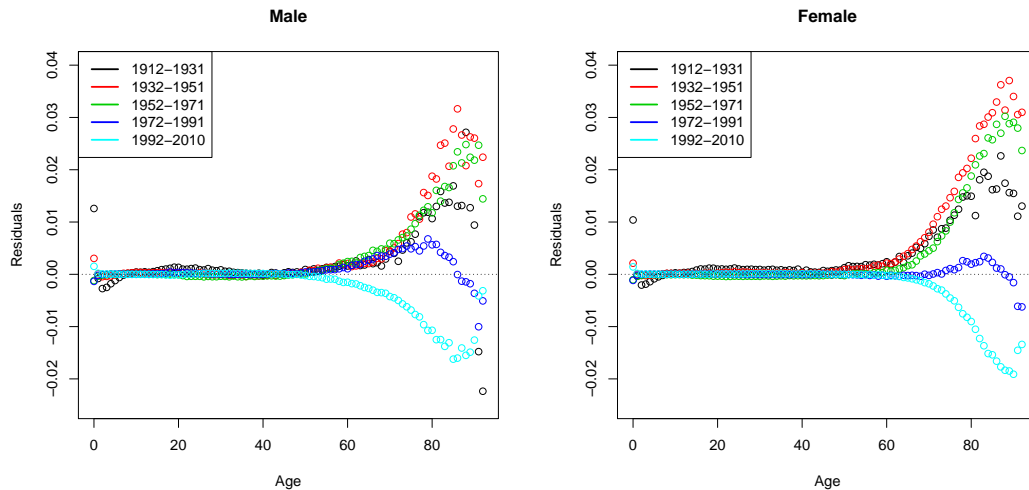


Figure 13.5: Residuals

A better indicator for the quality of the fit are the residuals of the estimated mortality rates. They are calculated by subtracting the estimated rates from the observed ones. The resulting values are shown in Figure 13.5 for the same 5 groups as above. The residuals are positively biased for all but the last group so they do not have an expected value of zero.

Hence, the rates are generally underestimated. They are correlated with age for the whole range of ages. The residuals are close to zero for ages below 80 and only increase or decrease afterwards to maximally +3.7% and -2.2%, respectively, where the maxima are achieved in the first two year groups. This is not too large so we conclude that the fit is fine.

**Forecasting and Extrapolating**

The next thing we want to know is if this model can be used to make realistic forecasts. Since the time-dependence in (13.1) is only via  $t$ , we can simply insert values  $t > 2010$  to get forecasted mortality rates. The results for year 2060 are shown on the left-hand side of Figure 13.6. When we compare the forecasted logit-rates to the ones we have estimated for 2010, we find that they are lower 50 years later. Over the whole range of ages, the two curves look similar but are steeper in 2060.

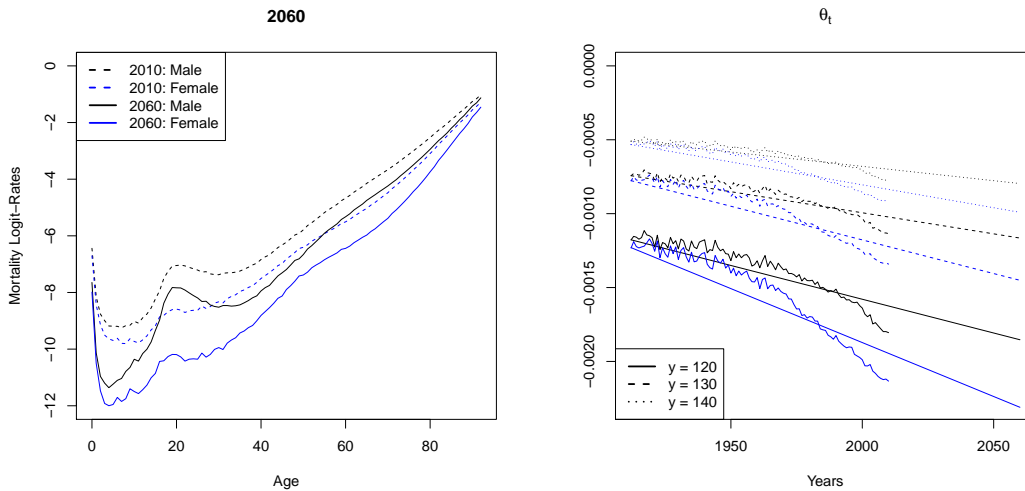


Figure 13.6: (LHS) Forecasted mortality logit-rates for year 2060, (RHS)  $\hat{\theta}_t$  from the extrapolation

The forecasted rates are needed to calculate the three measures for longevity. To calculate them according to Chapter 11, we first have to extend the mortality rates to a final age  $y$ . This is done by applying the method introduced in Chapter 8. We choose three different values for  $y$ , namely  $y = 120, 130$  and  $140$ , to see how it influences the results. We additionally extend the observed rates as a comparison. In the extrapolation process, we have to estimate  $\theta_t$  for each year separately. We see  $(\hat{\theta}_t)_t$  for the observed, estimated and forecasted mortality rates on the right-hand side of Figure 13.6. Unfortunately, we find that the two parameters in the past do not really match. We see this since the estimated and the observed parameters have different slopes which is also the reason why the curve in the future seems to be too flat. Since this does not depend on the choice of  $y$ , it could imply that we significantly underestimate longevity with this model. It is thus important to keep this plot in mind for all further analysis. We could try to solve this problem by extrapolating  $\theta_t$  from the observed rates itself with a random walk with drift or a similar process. Some smoothing of the resulting forecasted mortality rates would then be needed afterwards. However, this goes beyond the scope of this thesis.

**Measures for Longevity**

Nevertheless, we use the extended mortality rates to calculate the survival probability  $p_{0,t,j}$  given by (11.1). This is the probability that a person born in year  $t$  survives to the age  $j$ . We see in Figure 13.7 how this evolved from 1912 to 2010. We have calculated  $p_{0,t,j}$  for all three extended datasets. Luckily, the choice of  $y$  does not change the probability much. In this plot, we want to know at which age the kink in the curve happens. We see that it shifted in the past by about 15 years to the right. We also observe that the probability is close to 1 much longer in 2010. In 1912, it already starts decreasing for ages below 30 which is one of the reasons why the curve in 2010 is steeper after the kink than it was in 1912.

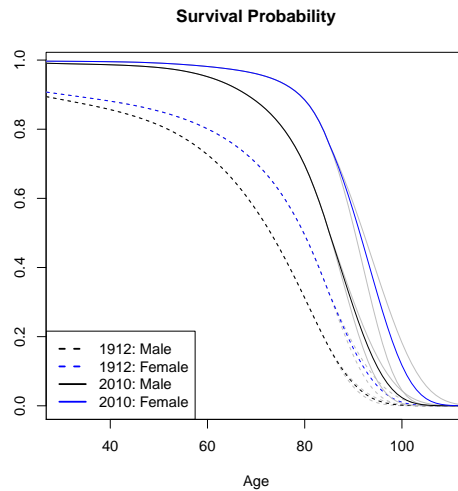


Figure 13.7: Survival probability  $p_{0,t,j}$  for  $t = 1912$  and 2010

A second indicator of the longevity of a society is the life expectancy  $e_{x,t}$ . We only calculate it for  $x = 30$  and 65 and from 1912 to 2060. We see in Figure 13.8 that the life expectancy for women is higher than the one for men and the curve is slightly steeper. The grey lines correspond to the values calculated from the extended datasets with  $y = 120$  and 140 whereas the coloured ones are for  $y = 130$ . We see that  $y$  has only a small influence on  $e_{x,t}$  and it is largest for the highest value of  $t$ . Hence, we have decided to show the explicit values from the past in Table 13.9 only for  $y = 130$ . We also observe here that the life expectancy increased a lot in these 99 years for both genders.

		1912	1920	1930	1940	1950	1960	1970	1980	1990	2000	2010
30	Male	37.8	39.3	41.1	42.7	44.2	45.7	47.0	48.2	49.3	50.4	51.4
	Female	42.4	44.2	46.4	48.3	50.1	51.7	53.2	54.5	55.8	56.9	57.9
65	Male	10.0	10.6	11.3	12.0	12.7	13.4	14.1	14.8	15.5	16.2	16.8
	Female	11.2	12.2	13.3	14.4	15.5	16.6	17.7	18.7	19.6	20.6	21.5

Table 13.9: Life expectancy at age 30 and 65 for  $y = 130$

Lastly, we calculate the price of an annuity (11.4) that a buyer of age 65 has to pay. We again use all three sets of extended data to show how  $y$  influences this price. We assume that the maturity is infinity so the annuity expires when the person dies. We choose three differ-

### 13 Application of Common Mortality Rate Models

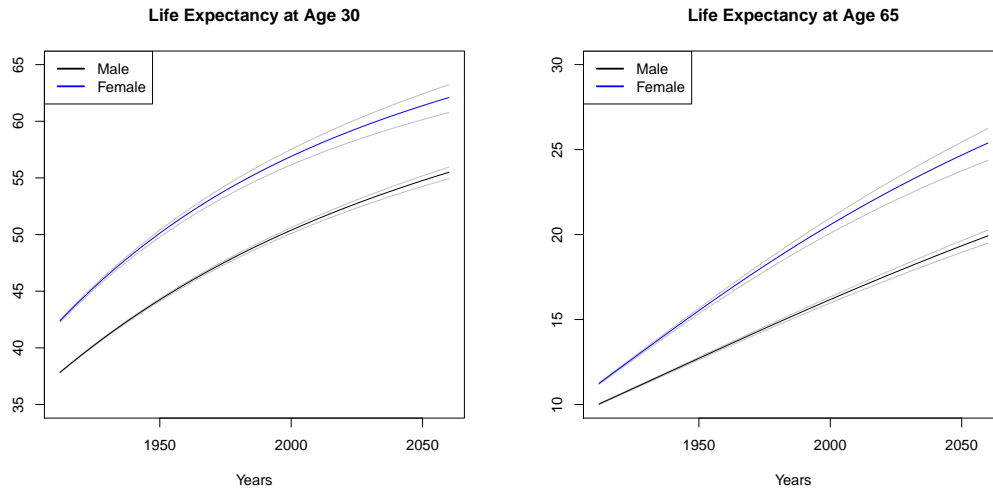


Figure 13.8: Life expectancy at age 30 and 65

ent values for the interest rate, namely  $r = 0\%$ ,  $1\%$  and  $5\%$ , to see how big its role is here. From Figure 13.10, we find that  $y$  only influences the price for small interest rates and the difference between the two grey curves is largest for women. It even increases as  $t$  increases. The interest rate has a much larger influence on the price, though. It does not change the general shape of the price curve but its position which means that the price decreases as  $r$  increases. Hence, the choice of  $r$  is more important than the one of  $y$ . As expected, the price of this annuity with  $r = 0\%$  coincides with the life expectancy of a person of age 65.

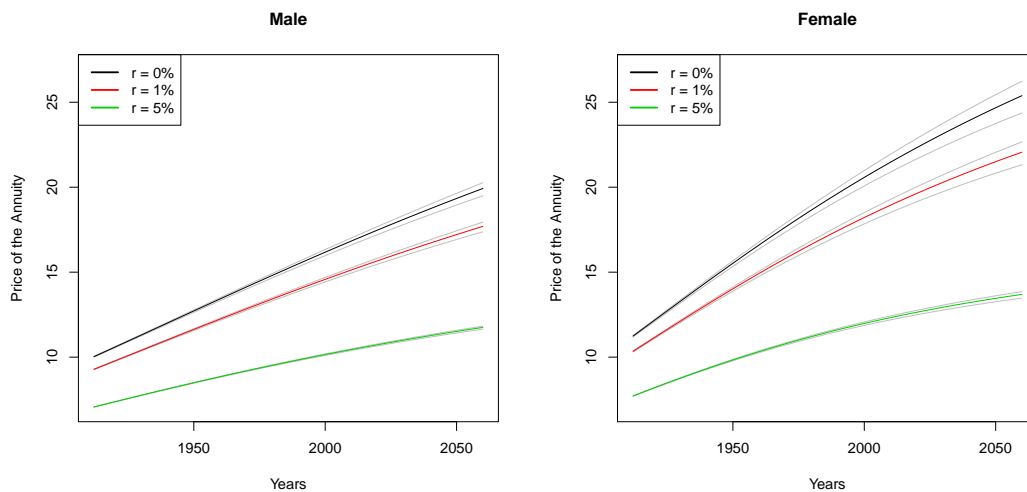


Figure 13.10: Price of the annuity for a person of age 65

As a conclusion, we can say that this model works rather well in the past. The estimated mortality rates are very easily achieved and yield reasonable results. The only downside is that it predicts too high mortality rates for the future. We try to improve this by applying the marginal mortality rate model to the same data in the next section since it uses an additional age-specific constant.

### 13.2 Marginal Mortality Rate Model

There are two age-specific constants in the marginal mortality rate model. Additionally to  $\lambda_x$ , there are  $(\alpha_x)_x$  which act as weights between the constant rate  $q_{x,t_0}$  and the generalized Nolfi model. The model equation is thus the following:

$$q_{x,t} = q_{x,t_0} (\alpha_x + (1 - \alpha_x) \exp(-\lambda_x (t - t_0))). \tag{13.2}$$

#### Fitting Procedure

We estimate both parameters simultaneously with a non-linear least squares regression where we add the constraints  $\lambda_x > 0$  and  $\alpha_x \in [0, 1]$ . In a first run, we find that  $\hat{\lambda}_x < 0$  for ages 93 and higher so we restrict the data to ages up to 92 in this section.

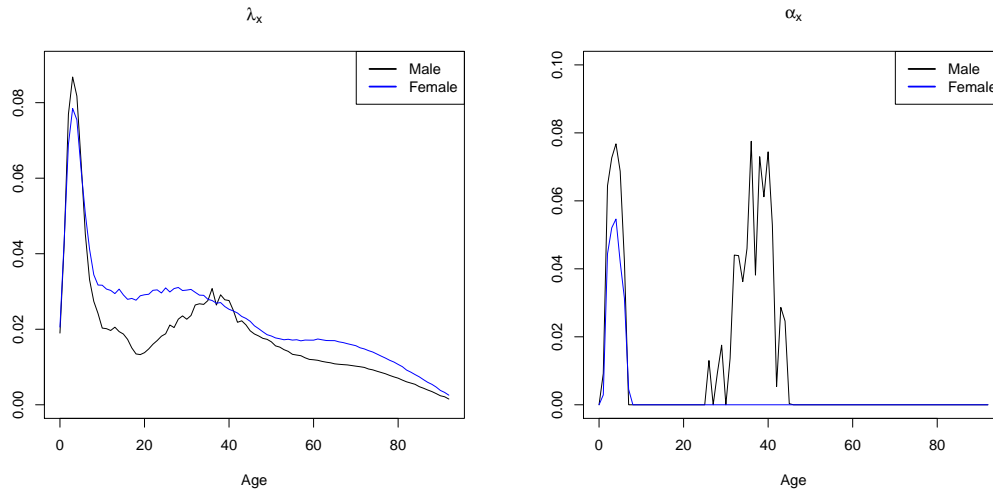


Figure 13.11: (LHS)  $\hat{\lambda}_x$ , (RHS)  $\hat{\alpha}_x$

In Figure 13.11, we see the estimated parameters. We find that  $\hat{\lambda}_x$  looks similar as in the generalized Nolfi model as a result of  $\hat{\alpha}_x$  being equal to zero for many  $x$  and small otherwise.

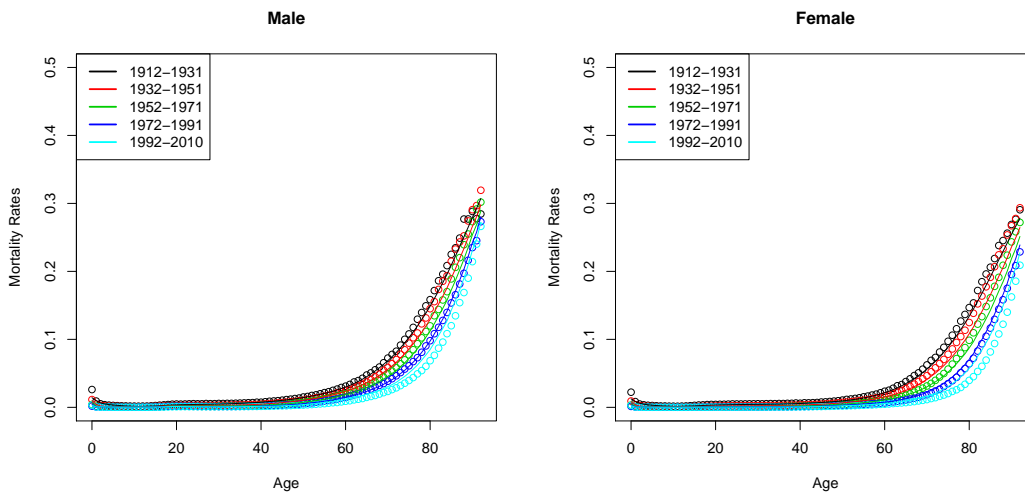


Figure 13.12: Observed (dots) and estimated (lines) mortality rates

### 13 Application of Common Mortality Rate Models

We again find that  $\hat{\lambda}_x$  is decreasing for both genders with one exception for men between 20 and 40. Since  $\hat{\alpha}_x$  is often equal to zero, it does not follow a linear trend but it indicates that the generalized Nolfi model is mostly a better fit to the data than the base mortality rate  $q_{x,t_0}$ .

We use the obtained parameter estimators to calculate the mortality rates. They are plotted in Figure 13.12 for the 5 groups introduced in the last section. The estimated rates lay on a nice curve but we already see that they are rather far away from the observed ones for ages above 70.

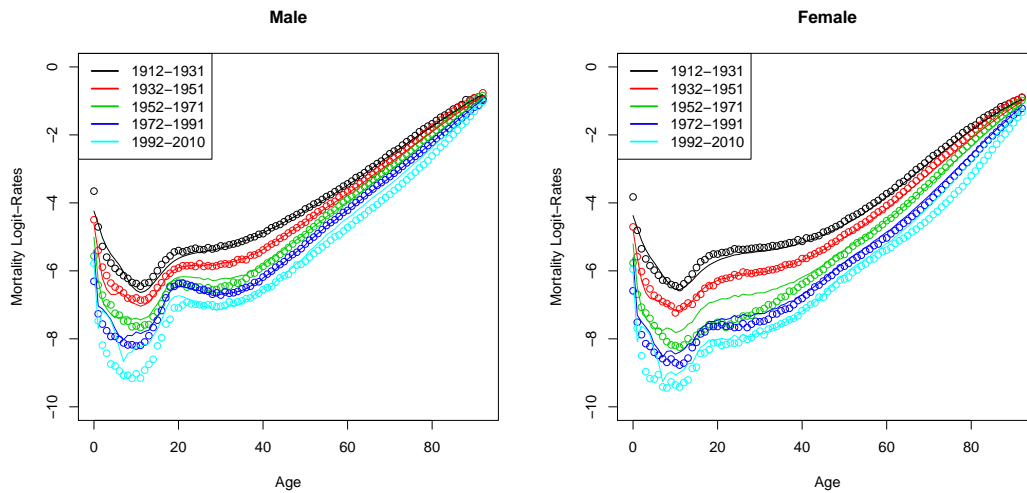


Figure 13.13: Observed (dots) and estimated (lines) mortality logit-rates

We want to know if this also holds true for low ages. Hence, we logit-transform the rates and plot them in Figure 13.13. We also see in this figure that the fit is not that good since the estimated logit-rates differ from the observed ones. We especially see this for ages until 20. Since the rates for these ages are still small, this is not too problematic, though.

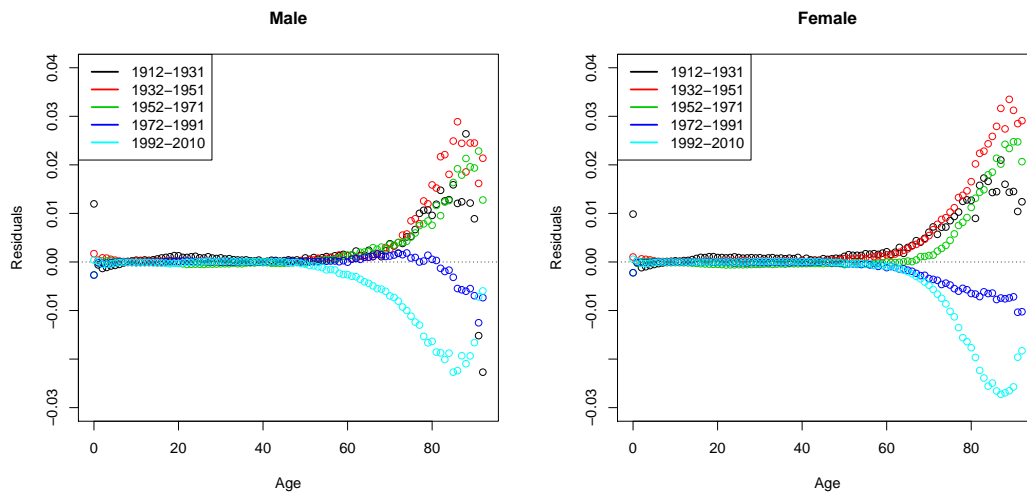


Figure 13.14: Residuals



We have a look at the residuals to see how good the fit with this model really is. They are shown in Figure 13.14 for the 5 year groups. We see that they are not zero everywhere for the low ages but at least they are very close with one exception in the earliest years. They only start increasing after the age of 60. We find that the residuals for the earliest years are positive after this point which means that the rates are underestimated. However, it is exactly the other way around for the last two year groups. This could mean that we will also overestimate the rates in the future. The residuals mostly stay between  $\pm 3\%$  with only very few exceptions. Thus, the fit is still fine.

### Forecasting and Extrapolating

Next, we would like to know how the mortality rates look like in the future. We can again simply insert any value for  $t$  into (13.2) to get forecasted mortality rates. The results for  $t = 2060$  are shown on the left-hand side of Figure 13.15. We observe that the rates are highly fluctuating and have another shape than they did in 2010. This may be caused by the rather varying  $(\hat{\alpha}_x)_x$ . Nevertheless, they are always smaller than in 2010 for ages above 20 even though the distance is not constant for all  $x$ . We will see only afterwards how reasonable these forecasts really are.

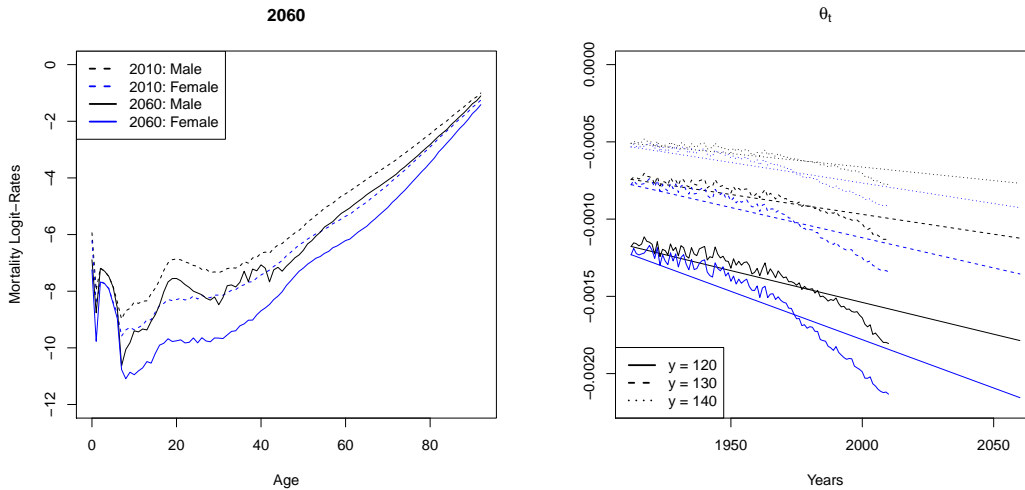


Figure 13.15: (LHS) Forecasted mortality logit-rates for year 2060, (RHS)  $\hat{\theta}_t$  from the extrapolation

The next thing we do is to extend the estimated and the forecasted mortality rates to ages above 92. We do this three times for the maximal ages  $y = 120, 130$  and  $140$ . The estimated parameters  $(\theta_t)_t$  from the extrapolation procedures are shown on the right-hand side of Figure 13.15 additionally to the ones found with the observed rates. We see that in all three cases and for both genders, they form nice processes which unfortunately highly differ in the past from the ones calculated with the observed rates. Hence, the extrapolation seems to have worked but like in the generalized Nolfi model, it is questionable how realistic the estimated and forecasted rates really are since the slope is again rather far away from the true one. This might again cause problems when trying to measure longevity.

**Measures for Longevity**

Nonetheless, we use these new datasets to calculate the survival probability  $p_{0,t,j}$  to have a first measure for longevity. In Figure 13.16, we observe that the difference between the three datasets is only visible for ages above 90 where the kink we are interested in already happened. Thus, the choice of  $y$  is not important for this plot. We see that the kink shifted from 1912 to 2010 by about 15 years to the right. This and the fact that the child death rate is nearly zero in 2010 are the reasons why the two curves look so different.

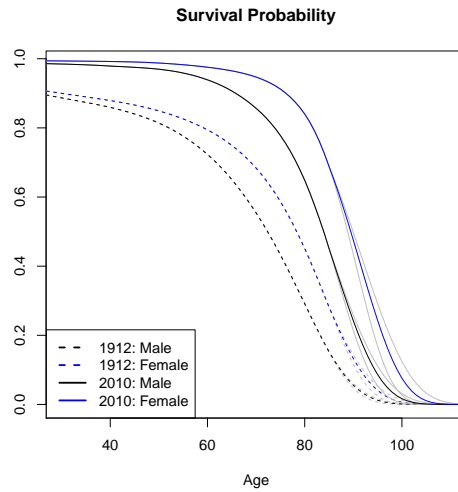


Figure 13.16: Survival probability  $p_{0,t,j}$  for  $t = 1912$  and 2010

The next thing we do is to calculate the life expectancy for a person of age 30 or 65. The results, shown in Figure 13.17, indicate that the life expectancy is increasing over the years and higher for women than for men. The curves are nice and a bit steeper for women, especially for those which are already 65 years old. We have used all three datasets for the calculation. The grey lines in the figure are the ones for  $y = 120$  and 140. We see that  $y$  has a

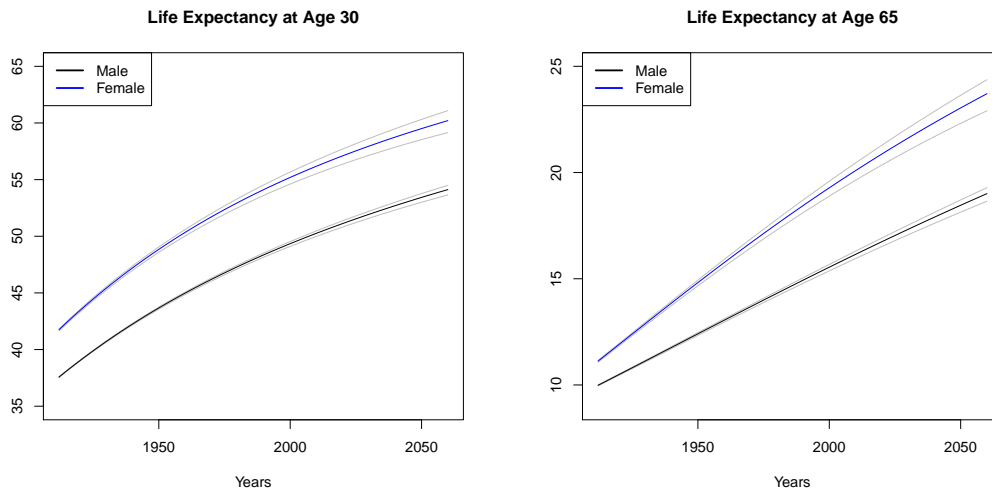


Figure 13.17: Life expectancy at age 30 and 65

### 13 Application of Common Mortality Rate Models

rather insignificant effect on  $e_{x,t}$  since the difference between the three curves is rather small everywhere. Thus, we choose  $y = 130$  to fill Table 13.18 with some explicit values from the past. From this table, we recognize the same characteristics of the life expectancy we have already seen in the two figures.

		1912	1920	1930	1940	1950	1960	1970	1980	1990	2000	2010
30	Male	37.6	39.0	40.7	42.3	43.7	45.0	46.2	47.3	48.4	49.3	50.3
	Female	41.7	43.5	45.4	47.2	48.8	50.3	51.7	53.0	54.1	55.2	56.2
65	Male	10.0	10.5	11.1	11.8	12.4	13.0	13.7	14.3	14.9	15.5	16.1
	Female	11.1	11.9	12.9	13.9	14.8	15.8	16.7	17.6	18.4	19.3	20.1

Table 13.18: Life expectancy at age 30 and 65 for  $y = 130$

As a last step, we calculate the price of the annuity introduced in the last section. We have chosen three interest rates, namely  $r = 0\%$ ,  $1\%$  and  $5\%$ . In Figure 13.19, we see that the choice of  $y$  and  $r$  influence the price in the same way as they did in the last section where the influence of  $y$  is largest for women with the smallest interest rate. The prices for women are higher than the ones for men and they form nice curves which are steeper for smaller values of  $r$ .

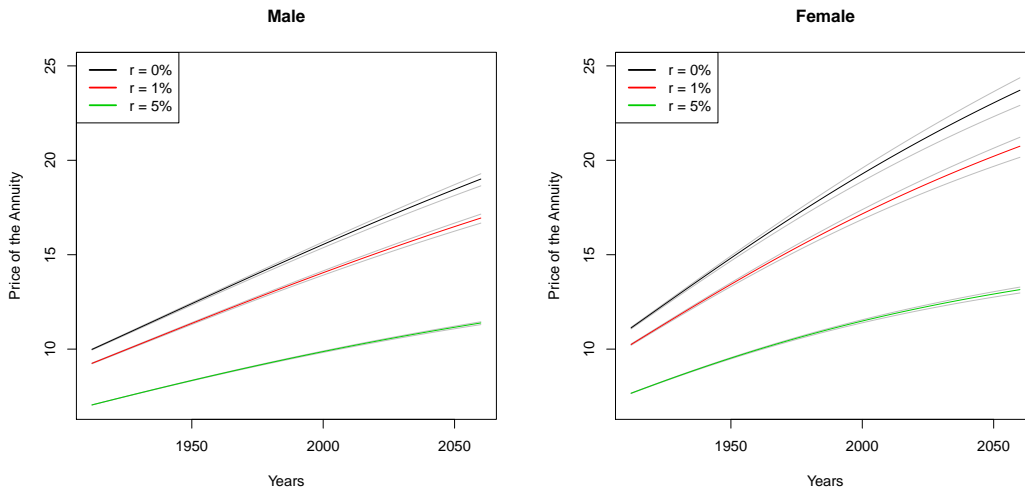


Figure 13.19: Price of the annuity for a person of age 65

We have seen that the estimator of  $\alpha_x$  is close to zero for all  $x$ . This means, we would guess that the results from this model are very close to the ones from the generalized Nolfi model. However, this is not completely true. In general, it is even slightly worse than the last one since the forecasted rates are less convincing. We now want to know if the extrapolation model yields similarly good results for the past and if it might provide better predictions.

### 13.3 Extrapolation Model

The extrapolation model contains two age-specific constants. It is assumed that the expected mortality rate grows super- or sub-exponentially. This means

$$q_{x,t} = q_{x,t_0} \exp(-\lambda_x (t - t_0)^{c_x}). \quad (13.3)$$

**Fitting Procedure**

The estimation of  $\lambda_x$  and  $c_x$  is done with a non-linear least squares regression like in the last section where we need to make sure that  $\lambda_x$  is positive. We find that  $\hat{\lambda}_x < 0$  for  $x \geq 92$  so we only work with the data for  $x \leq 91$  from now on.

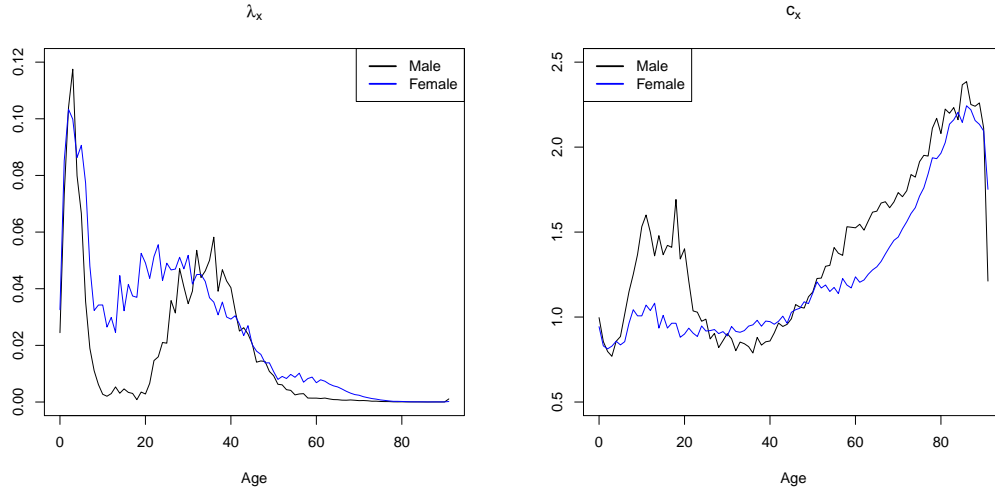


Figure 13.20: (LHS)  $\hat{\lambda}_x$ , (RHS)  $\hat{c}_x$

The resulting estimators  $\hat{\lambda}_x$  and  $\hat{c}_x$  are shown in Figure 13.20. We observe that for this model,  $\hat{\lambda}_x$  looks clearly different as it did in the last two sections. It is still mostly decreasing but much more varying.  $\hat{c}_x$  is also fluctuating but it has an increasing trend after age 30. After the age 46, it is always larger than 1 which means that the mortality rates increase super-exponentially for those ages. We could apply the clearing method by Whittaker-Henderson from Section 2.1 to the estimators  $(\hat{\lambda}_x)_x$  and  $(\hat{c}_x)_x$  if they are too volatile. However, we will use the original estimators in this section since the results are only slightly affected by their volatility.

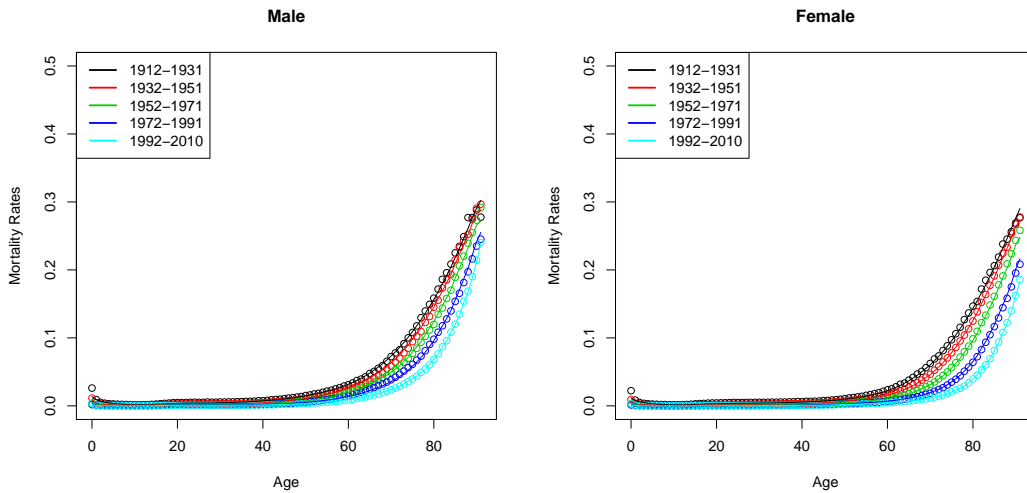


Figure 13.21: Observed (dots) and estimated (lines) mortality rates

### 13 Application of Common Mortality Rate Models

When inserting these estimators into the model equation (13.3), we get the estimated mortality rates shown in Figure 13.21. We grouped the data into the same 5 groups as before. From a first glance, the rates look rather appropriate. They even look better than in the last two sections.

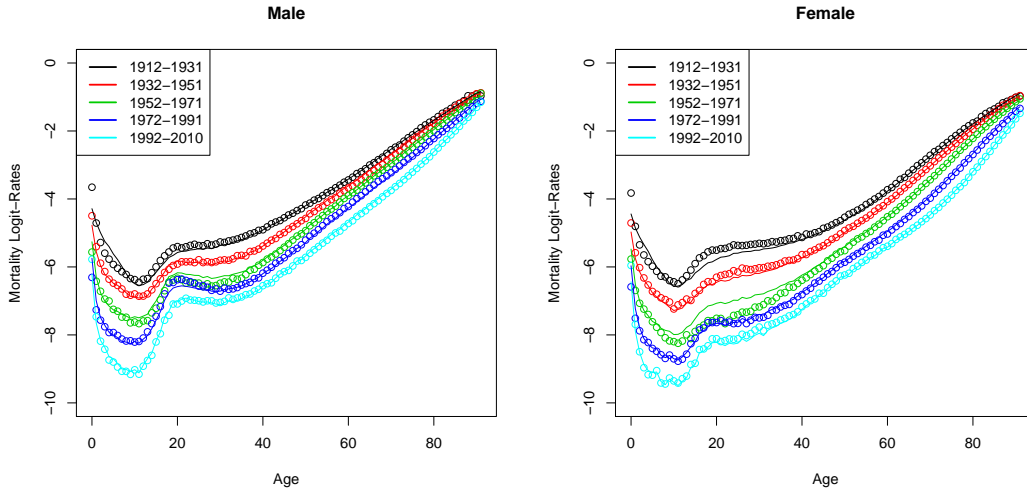


Figure 13.22: Observed (dots) and estimated (lines) mortality logit-rates

The good quality of this model is also visible when we logit-transform the estimated mortality rates. In Figure 13.22, the difference between the observed and the estimated rates is very small. The bumps for ages below 20 are captured very well which implies that the model also fits the data nicely for small ages.

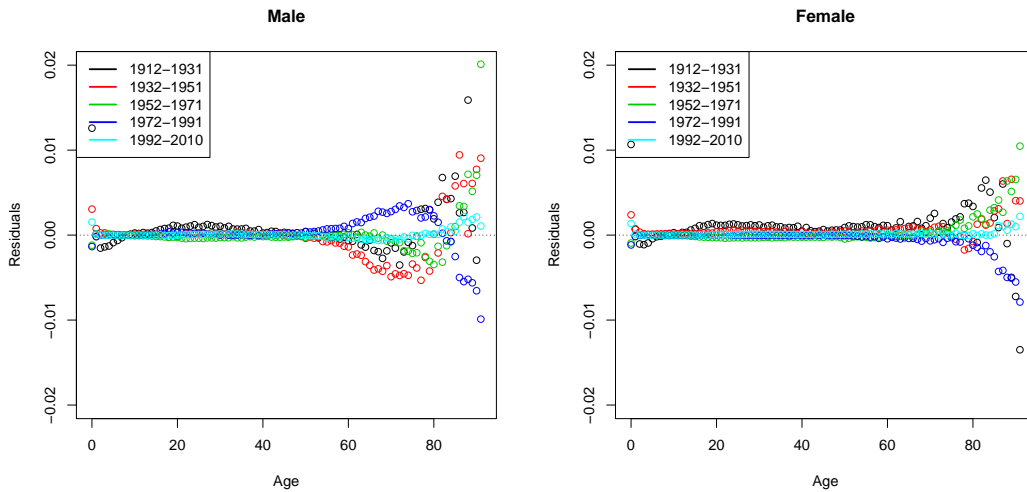


Figure 13.23: Residuals

We see this again in Figure 13.23 where the residuals are shown. Until age 80, they are very close to zero with a few exceptions for very small ages. Afterwards, they start increasing but always stay between  $\pm 2\%$  which is a very narrow range. Thus, we conclude that the model fits the data very well even though  $\hat{\lambda}_x$  and  $\hat{c}_x$  are more varying than expected.

**Forecasting and Extrapolating**

It is very easy to get mortality rates for the future with this model since it only depends on the year via  $t$ . We can plug any value  $t > 2010$  into (13.3) to get forecasted mortality rates for the future. On the left-hand side of Figure 13.24, we see the ones we get for year 2060. They are clearly lower than the ones we estimated for 2010 which is a good thing. Unfortunately, they vary for all ages and have a rather different shape than in 2010. This may be caused by the rather varying values of  $\hat{\lambda}_x$  and  $\hat{c}_x$ .

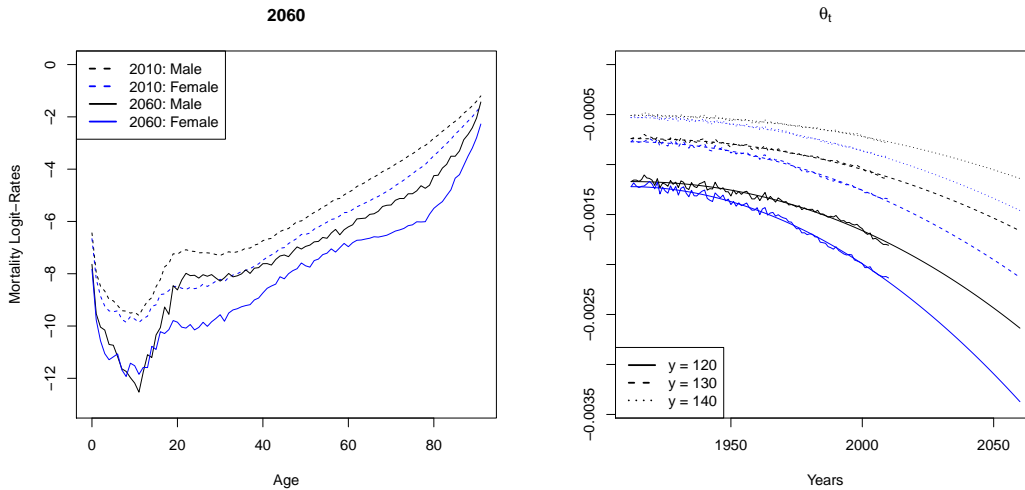


Figure 13.24: (LHS) Forecasted mortality logit-rates for year 2060, (RHS)  $\hat{\theta}_t$  from the extrapolation

We now use the estimated and forecasted mortality rates for the calculation of other quantities. The first thing we need to do, though, is to extend the observed, estimated and forecasted rates to a maximal age of  $y = 120, 130$  and  $140$ . The estimators  $(\hat{\theta}_t)_t$  we get from this procedures are shown on the right-hand side of Figure 13.24. We see that they do not form a straight line anymore but one that is slightly bent. Therefore, it coincides much better with the estimators found with the observed rates. From this group of models, this one yields the most realistic estimators and forecasts. It captures the slope of  $\theta_t$  very nicely and projects it in the same way to the future. However, we guess that the slope is probably slightly too steep so longevity might be overestimated.

**Measures for Longevity**

We now use all three extended datasets to calculate the survival probability  $p_{0,t,j}$  for the years  $t = 1912$  and  $2010$ . In Figure 13.25, we see that  $y$  influences this probability because the kink in 2010 will not happen until an age of about 95 and for this age, the influence of  $y$  is clearly existing. Here, the difference between these 99 years is much larger than in the last two models. The figure also indicates that nearly all people that are 30 years old in 2010 will achieve an age of 95 which is rather high.

Another way to show that people really live longer today than 99 years ago is by calculating the life expectancies at age 30 and 65. From Figure 13.26, it is clear that  $y$  has a huge influence on  $e_{x,t}$  after year 2000. In 2060, the spread between the grey curve for  $y = 120$

### 13 Application of Common Mortality Rate Models

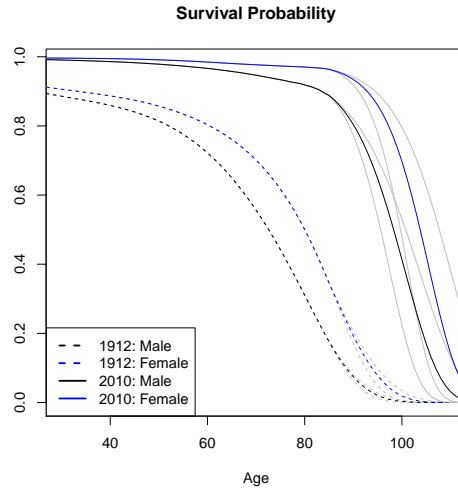


Figure 13.25: Survival probability  $p_{0,t,j}$  for  $t = 1912$  and 2010

and the one for  $y = 140$  are nearly 10 years for both genders if  $x = 30$ . The life expectancy increases nearly linearly over the years and is larger for women than for men. However, the slope of the curve is steeper for the most recent years and decreases again in the future. This is also visible in Table 13.27 where we have chosen to show some values from the past for the dataset with  $y = 130$ .

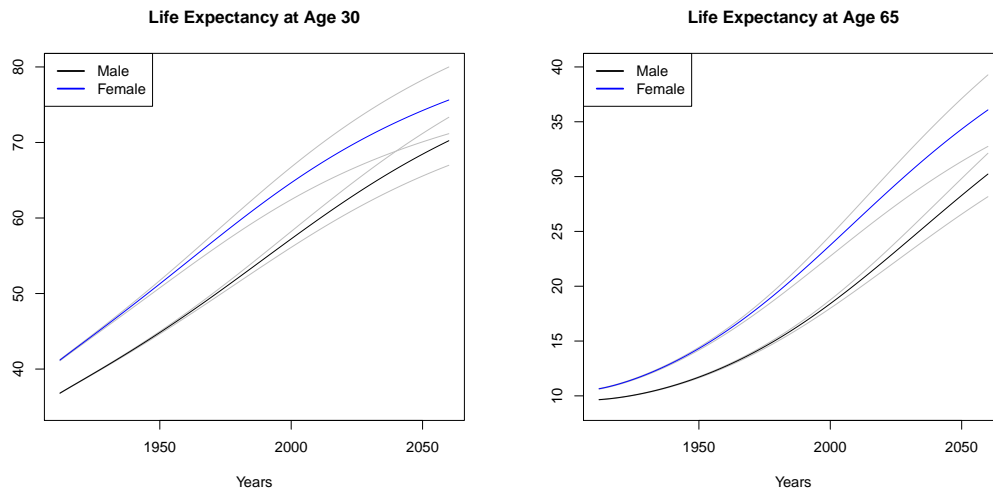


Figure 13.26: Life expectancy at age 30 and 65

		1912	1920	1930	1940	1950	1960	1970	1980	1990	2000	2010
30	Male	36.8	38.4	40.5	42.6	44.8	47.2	49.6	52.1	54.7	57.2	59.7
	Female	41.2	43.3	45.9	48.6	51.3	54.1	56.8	59.6	62.2	64.7	67.0
65	Male	9.7	9.9	10.3	10.9	11.7	12.7	13.8	15.2	16.7	18.4	20.3
	Female	10.6	11.1	12.0	13.0	14.3	15.8	17.6	19.5	21.6	23.7	26.0

Table 13.27: Life expectancy at age 30 and 65 for  $y = 130$

### 13 Application of Common Mortality Rate Models

We again observe this development when calculating the price of the annuity we have introduced earlier. We recall that the maturity was equal to infinity and we have chosen three interest rates for its calculation, namely  $r = 0\%$ ,  $1\%$  and  $5\%$ . We find that the price is higher for women than for men. We observe that after 1950, the prices start increasing faster than before and the slope of the curve decreases again in the future. Unfortunately, the influence of  $y$  is larger than in the last two models. An increasing  $r$  at least changes the price in the same way as before, namely it shifts the price curve downwards.

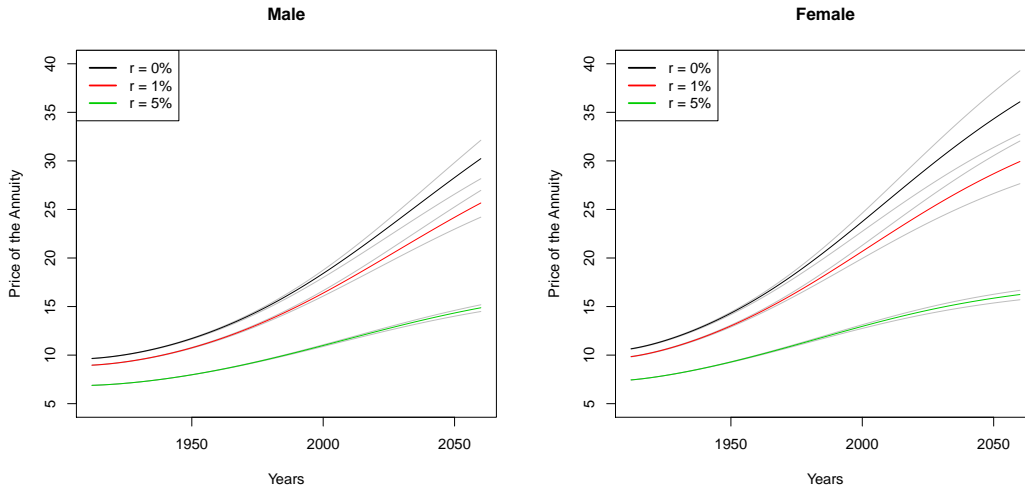


Figure 13.28: Price of the annuity for a person of age 65

As a conclusion for all three models we have analysed in this chapter, we say that they fit the data for the past very well. The downside is that the rates for the future of the first two models are not convincing. For the extrapolation model, however, the predictions seem realistic but we guess that longevity is probably overestimated since it is by far the most conservative model. We hope to overcome this problem by applying more evolved models in the next chapters. The first one we will have a look at has a more complex model structure even though the forecasting procedure is still the same as for these three models.



# 14 Application of Generalized Linear Models

We start with the first more complex group of models, namely the generalized linear models that were introduced in Chapter 4. We recall that we have defined

$$q_{x,t} = g^{-1}(\theta_1 + \theta_2 x + \theta_3 t), \tag{14.1}$$

$$q_{x,t} = g^{-1}(\theta_1 + \theta_2 x + \theta_3 t + \theta_4 x t + \theta_5 x^2 t), \tag{14.2}$$

where we have chosen  $g$  to be the logit link (3.6). This means that  $g^{-1}$  is given by (4.7).

## Fitting Procedure

We fit both models and afterwards discuss which one may be the better choice. The parameter estimator vector  $\hat{\theta}_1 \in \mathbb{R}^3$  for model (14.1) is given by

	MALE:				FEMALE:			
	Estimate	Std. Error	z value	Pr(> z )	Estimate	Std. Error	z value	Pr(> z )
(Intercept)	2.007e+01	4.401e-02	456.1	<2e-16 ***	2.693e+01	4.568e-02	589.5	<2e-16 ***
x	7.879e-02	3.928e-05	2005.8	<2e-16 ***	8.673e-02	4.385e-05	1977.6	<2e-16 ***
t	-1.456e-02	2.264e-05	-643.3	<2e-16 ***	-1.850e-02	2.366e-05	-782.1	<2e-16 ***
---								
	Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1							

We see that all parameters are highly significant and therefore need to stay in the model. The value of the AIC (4.15) is equal to 618'060 for men and 727'400 for women. As a comparison, we fit model (14.2) to the male and female data. The estimator  $\hat{\theta}_2 \in \mathbb{R}^5$  then looks as follows:

	MALE:				FEMALE:			
	Estimate	Std. Error	z value	Pr(> z )	Estimate	Std. Error	z value	Pr(> z )
(Intercept)	4.860e+01	1.525e-01	318.6	<2e-16 ***	6.506e+01	1.794e-01	362.7	<2e-16 ***
x	-3.344e-01	2.336e-03	-143.1	<2e-16 ***	-3.974e-01	2.558e-03	-155.4	<2e-16 ***
t	-2.822e-02	7.870e-05	-358.6	<2e-16 ***	-3.674e-02	9.283e-05	-395.8	<2e-16 ***
xt	1.716e-04	1.225e-06	140.1	<2e-16 ***	1.965e-04	1.349e-06	145.6	<2e-16 ***
x2t	3.607e-07	6.977e-10	517.0	<2e-16 ***	4.450e-07	7.300e-10	609.5	<2e-16 ***
---								
	Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1							

Even though  $\hat{\theta}_5$  is rather small in both cases, all predictors are highly significant. The value of the AIC is now equal to 245'176 for the male data and 193'969 for the female data. It is smaller than in the first model for both genders so we prefer this one. To check if this is sensible, we perform an  $F$ -test to test the null-hypothesis

$$H_0 : \theta_4 = \theta_5 = 0.$$

Hence, we need to calculate the test statistic (4.16). For men, we get  $D(\hat{\theta}_1) = 553'468$  and  $D(\hat{\theta}_2) = 180'581$ . Since  $df = 9499$  and the 99%-quantile of the  $F_{2,9499}$ -distribution is given by 4.6, we get that the test statistic is equal to

$$\frac{9499}{2} \frac{553'468 - 180'581}{180'581} = 9807.4 > 4.6.$$

## 14 Application of Generalized Linear Models

Equivalently, we get for the female data

$$\frac{9499}{2} \frac{665'003 - 131'568}{131'568} = 19'256.6 > 4.6.$$

Therefore, we clearly reject the null-hypothesis and only analyse model (14.2). The estimated mortality rates are thus given by

$$\text{Male: } \hat{q}_{x,t} = g^{-1}(48.60 - 0.3344x - 0.0282t + 0.0002xt + 3.6 \cdot 10^{-7}x^2t), \quad (14.3)$$

$$\text{Female: } \hat{q}_{x,t} = g^{-1}(65.06 - 0.3974x - 0.0367t + 0.0002xt + 4.5 \cdot 10^{-7}x^2t). \quad (14.4)$$

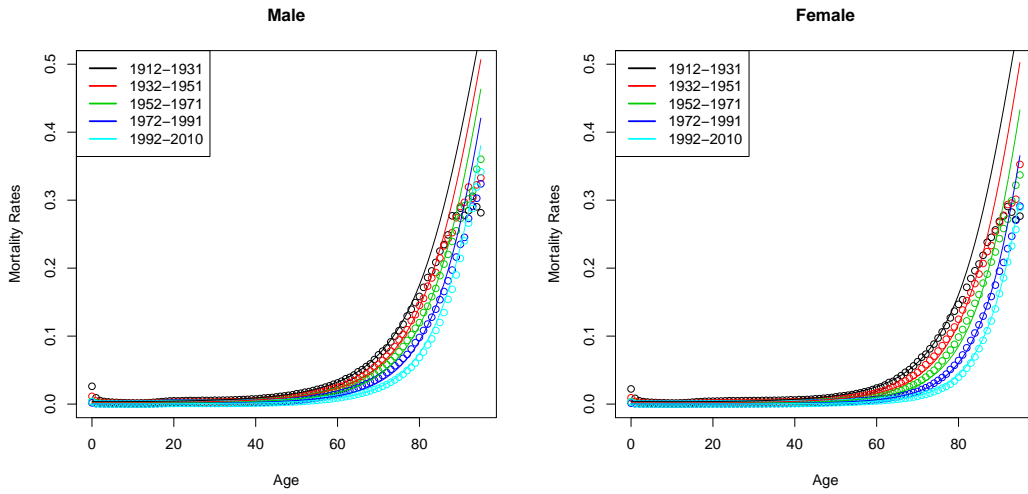


Figure 14.1: Observed (dots) and estimated (lines) mortality rates

In Figure 14.1, the estimated mortality rates for men and women using models (14.3) and (14.4) are shown. We grouped the data into the 5 groups that were introduced in the last chapter. We see that the fit for ages above 80 is unsuitable for all 5 year groups. The mortality rates are highly overestimated because they increase too quickly.

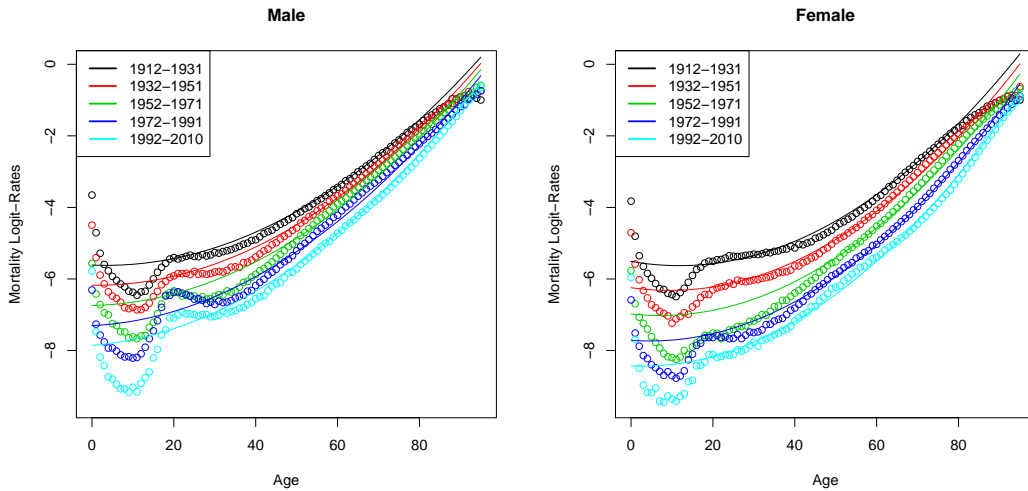


Figure 14.2: Observed (dots) and estimated (lines) mortality logit-rates

## 14 Application of Generalized Linear Models

Since the estimated mortality rates are very close to zero for ages up to 60, we logit-transform the rates and plot the mortality logit-rates in Figure 14.2 to have a clearer view on the estimated rates for these ages. For ages 0 to 40, we find that the estimated rates differ from the observed ones. The model cannot account for the bumps in the curves for these ages but only fits a line that is slightly bent. This indicates that the model might not be a good fit for people with an age below 40.

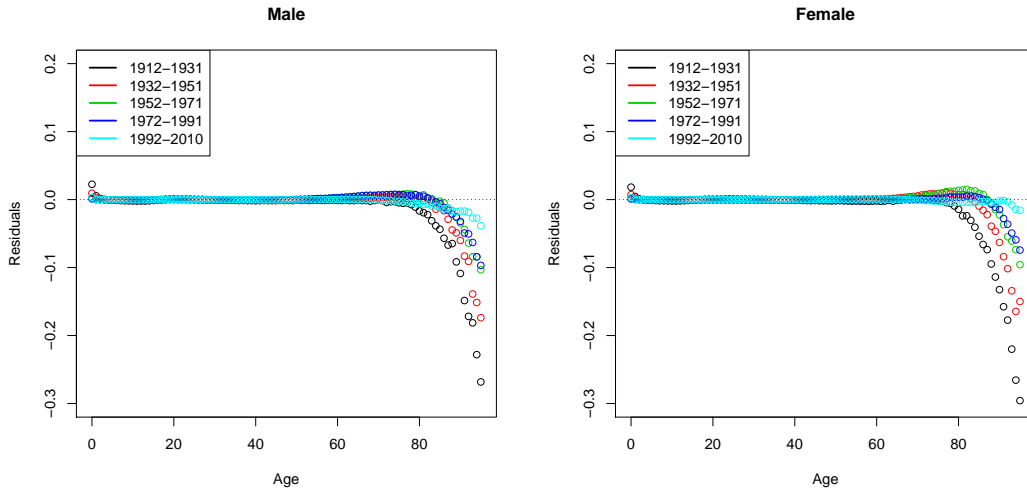


Figure 14.3: Residuals

To get a better understanding of the quality of the model, we plot the residuals for the 5 groups in Figure 14.3. We observe that the residuals are highly negative for ages above 80, especially for the earliest years. Furthermore, we see that the residuals are not independent but are rather highly correlated with age. It is clear that the residuals do not have an expected value of zero but are negatively biased. This means, we generally overestimate the mortality rates with this model. Therefore, we are still under the impression that this model does not fit the data well.

### Forecasting and Extrapolating

As a next step, we want to get forecasted mortality rates. We calculate the rates up to year 2060 which means we calculate 50-year forecasts. As discussed in Chapter 9, this is done by inserting  $t > 2010$  into (14.3) and (14.4). On the left-hand side of Figure 14.4, we see the estimated mortality logit-rates in 2010 and the forecasted ones for 2060. The forecasted rates in 2060 are smaller than the ones estimated for 2010. This means, we expect people in 50 years to live longer than today.

We use the estimated and forecasted mortality rates to extend the mortality rates to ages up to  $y = 120, 130$  and  $140$ . On the right-hand side of Figure 14.4, we see how  $(\theta_t)_t$  was estimated during the extrapolation procedure. We also show how it looks like for the observed mortality rates. We find that the estimated parameters are close to the observed ones for the years after 1950 which implies that the slope in the future is realistic. Hence, this model yields good forecasts even though the estimators before that year are not as accurate.

## 14 Application of Generalized Linear Models

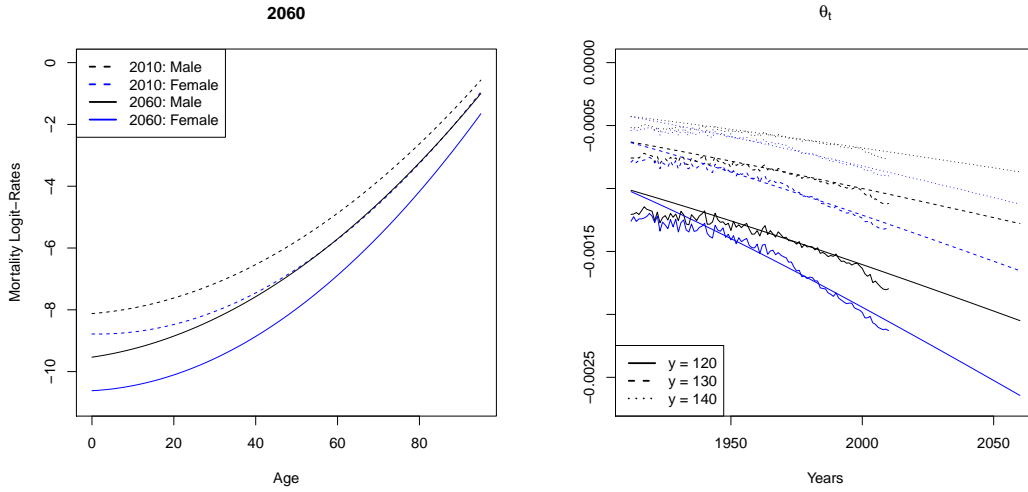


Figure 14.4: (LHS) Forecasted mortality logit-rates for year 2060, (RHS)  $\hat{\theta}_t$  from the extrapolation

### Measures for Longevity

To get a better feeling for longevity, we show the change of the survival probability  $p_{0,t,j}$  between the years  $t = 1912$  and 2010 using the extended data for all values of  $y$ . In Figure 14.5, the grey lines correspond to the datasets with  $y = 120$  and 140 and the coloured ones are for  $y = 130$ . We observe that  $y$  only influences the probability after an age of about 95 which is good since at this age, the kink already happened in both years. Hence,  $y$  does not play a crucial role for this analysis. We see that in 1912, the kink was at an age of about 60 for men and 65 for women whereas in 2010, the estimated survival probability starts decreasing rapidly between ages 75 and 80. This means, the curve shifted in the last 99 years by about 15 years to the right. What also happened is that the probability that small children die nearly vanished until 2010. This is the reason why the curve in 2010 lays above the one in 1912 everywhere.

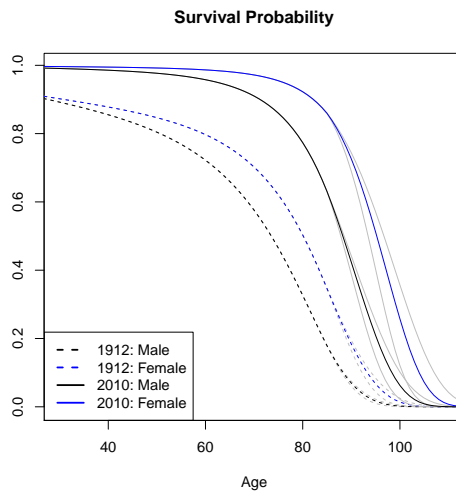


Figure 14.5: Survival probability  $p_{0,t,j}$  for  $t = 1912$  and 2010

## 14 Application of Generalized Linear Models

Next, we use the extended data to estimate the life expectancy (11.3) for a person of age  $x$  in year  $t$ . We only calculate this for  $x = 30$  and  $65$ . In Figure 14.6, the calculated life expectancies are shown. The coloured lines correspond to the ones with  $y = 130$  whereas the grey lines are for  $y = 120$  and  $140$ . We see that the influence of  $y$  is rather small in the past but we clearly see a difference between the three values of  $y$  afterwards. All curves are increasing and the life expectancy of the women is always larger than the one of the men. The curves are nearly linear where the slope decreases slightly as  $t$  increases. In Table 14.7, we only show the life expectancies for  $y = 130$  in some selected years from the past. We recognize here the same behaviour of  $e_{x,t}$  as in the figure above.

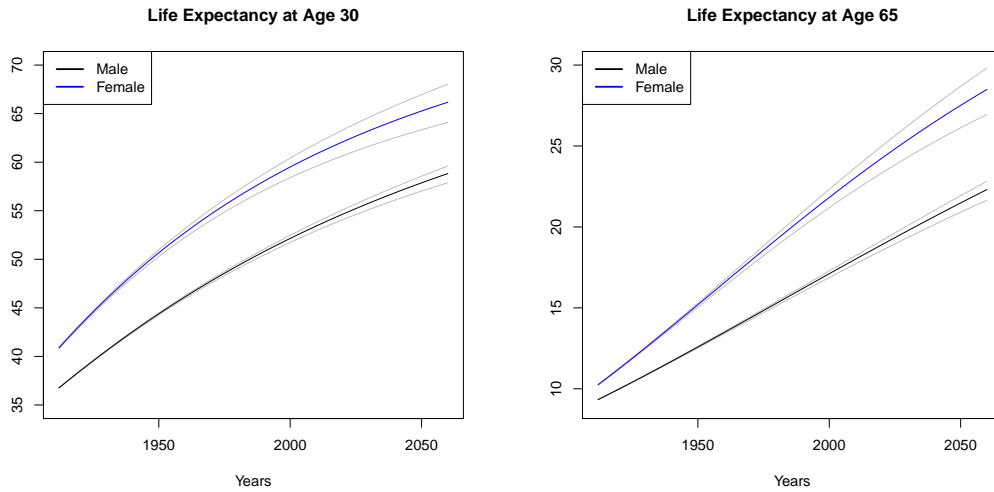


Figure 14.6: Life expectancy at age 30 and 65

		1912	1920	1930	1940	1950	1960	1970	1980	1990	2000	2010
30	Male	36.8	38.5	40.6	42.5	44.4	46.1	47.8	49.3	50.8	52.1	53.4
	Female	40.9	43.2	45.9	48.4	50.7	52.8	54.7	56.4	58.0	59.5	60.9
65	Male	9.3	10.0	10.8	11.7	12.6	13.5	14.4	15.3	16.2	17.1	18.0
	Female	10.2	11.2	12.5	13.9	15.2	16.5	17.9	19.2	20.5	21.8	23.1

Table 14.7: Life expectancy at age 30 and 65 for  $y = 130$

The same development is visible when calculating the price of an annuity sold to a person of age 65 when the interest rate is assumed to be constant and equal to  $r = 0\%$ ,  $1\%$  and  $5\%$ , respectively. In Figure 14.8, the coloured lines correspond to the extended data with  $y = 130$  and the grey ones are the prices for  $y = 120$  and  $140$ . We see that  $y$  influences the price in a similar intensity as it did for the life expectancy. The interest rate, on the other hand, has a much larger impact on it. The larger  $r$  is, the cheaper the annuity gets. The general shape of the price curve is at least only slightly influenced by an increasing  $r$  since it only reduces the slope of the price curve. We observe that the price of the annuity increases over the years which is due to the fact that people live longer today than 99 years ago and that they are expected to live even longer 50 years from now.

## 14 Application of Generalized Linear Models

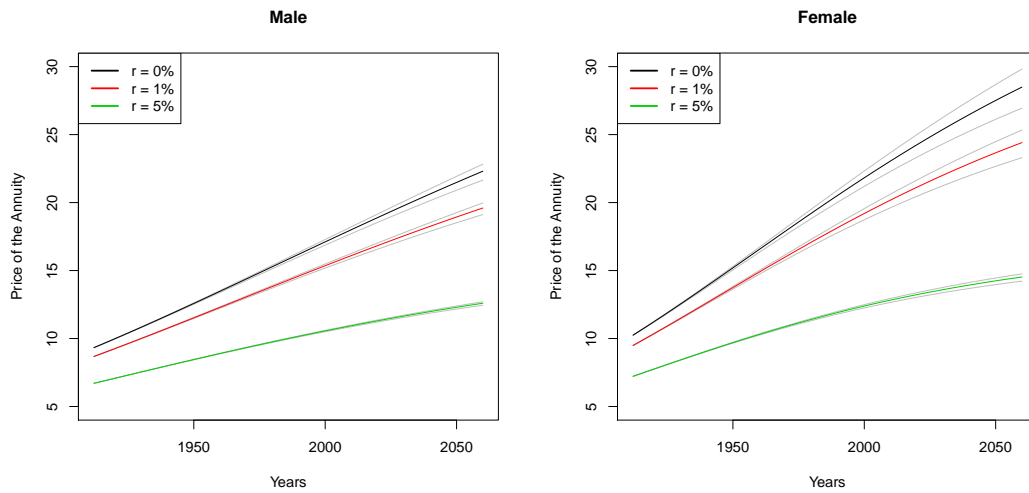


Figure 14.8: Price of the annuity for a person of age 65

We have seen how the survival probability, the life expectancy and the price of an annuity behave over the years for one specific generalized linear model. However, to be able to say something significant about the longevity of society, we need to compare these results to those calculated with other models. Therefore, we go on and apply the models of Lee and Carter to the same data.

# 15 Application of the Models of Lee and Carter

In this application, we are going to analyse the models of Lee and Carter introduced in Chapter 5. We will not have a look at the generalized model since we would need a lot more data to be able to calculate the cohort effect accurately. Therefore, we restrict ourselves to the classical and the extended model and treat them both individually.

## 15.1 Classical Model

We recall that in the classical model of Lee and Carter, we assumed that the mortality log-rate was given by

$$\log(Q_{x,t}) = a_x + b_x k_t + \varepsilon_{x,t}. \quad (15.1)$$

### Fitting Procedure

We start with the estimation of  $a_x$ ,  $b_x$  and  $k_t$  for all  $x$  and  $t$ . We have seen three methods to do so. We use all of them to see if they give different estimators and if so, how much they differ from each other. In the end, we will choose the best pair for all further analysis. The Newton-Raphson Algorithm 5.2 did converge for  $\delta = 0.1$  after 12 iterations for the male data whereas for the female data, 16 runs were needed. Unfortunately, we find that  $\hat{b}_x < 0$  for  $x = 94$  and  $95$  which is not allowed. Therefore, we only use the data for ages 0 to 93 in this application.

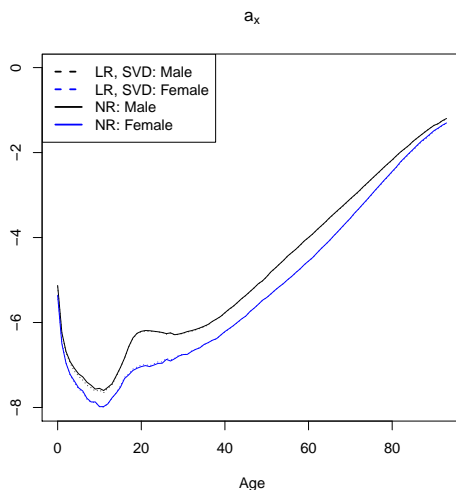


Figure 15.1:  $\hat{a}_x$

The estimator of  $a_x$  from the linear regression (LR) and the one from the SVD method (SVD) are the same. In Figure 15.1, we see the estimated parameter  $\hat{a}_x$  for men and women for these two and for the Newton-Raphson method (NR). The two estimators are very similar. When ignoring the fluctuations for ages below 20, both curves are increasing with age. This

15 Application of the Models of Lee and Carter

is what we expect since it indicates that the mortality rate generally increases with age as well. The negative slope for ages 0 to 10 may be caused by the fact that babies die more often than children who already survived a few years.

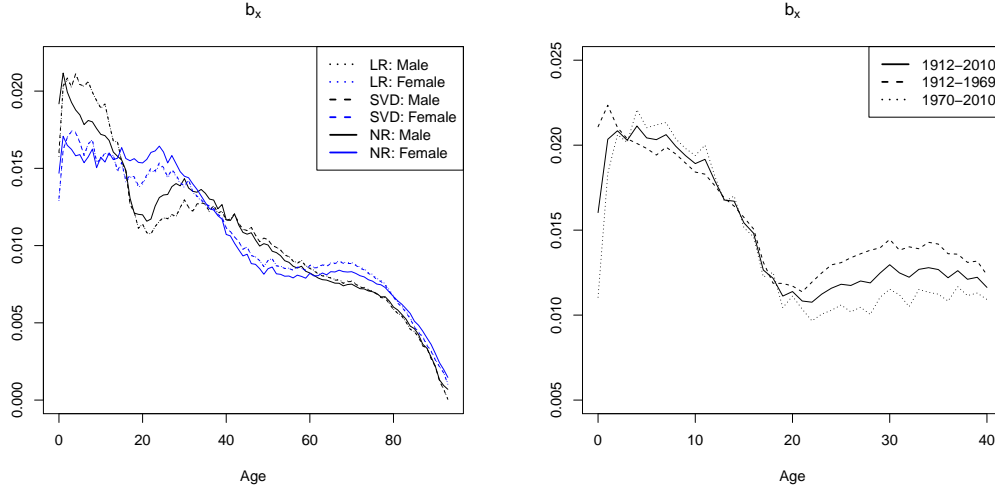


Figure 15.2:  $\hat{b}_x$ : (LHS) 1912 - 2010, (RHS) 1912 - 1969 and 1970 - 2010

The estimator  $(\hat{b}_x)_x$ , shown in Figure 15.2, is rather small for all three methods. The estimators from the first two methods are very close and decreasing whereas the one from the Newton-Raphson method is slightly different. For ages between 10 and 20, the curve for the male data is rather steep in all three cases. We guess that this is caused by car driving so we split the data into two groups and re-estimate  $\hat{b}_x$  for both of them with the SVD method. First, we only use data until 1969 and afterwards only those from 1970 to 2010. We observe that for both groups,  $(\hat{b}_x)_x$  looks similar to the curve over the whole range of years even though the bump is more pronounced after 1970. This indicates that the bump might be caused by car driving but not exclusively.

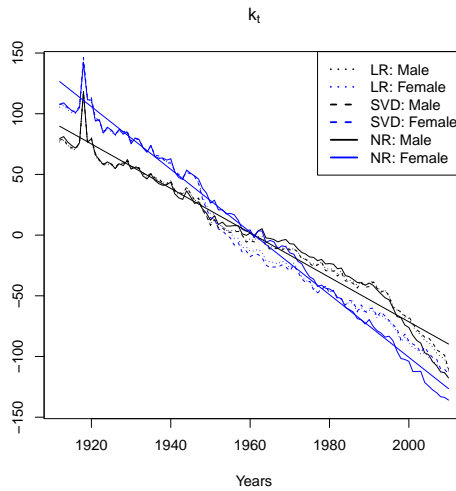


Figure 15.3:  $\hat{k}_t$



The estimator of the time-varying index  $(k_t)_t$ , shown in Figure 15.3, has a decreasing trend for all three methods which indicates that the mortality rates decrease over the years. We observe that the first two methods still approximately yield the same estimator whereas the estimator from the iterative method is rather different, especially in the years after 1950. The big peak in 1918, which is clearly visible for both genders, is captured by all methods and can be explained by the Spanish flu. The index generally varies a bit and therefore, it is not too easy to find its true trend. Nevertheless, the straight lines signify the trend of the estimators from the Newton-Raphson method which will be estimated later on. We see that it is very accurate

Since all three pairs of estimators are similar, we only work with the estimators obtained with the Newton-Raphson method for the remainder of this section. We choose this method since it minimizes the scaled deviance which means that the log-likelihood function is maximized.

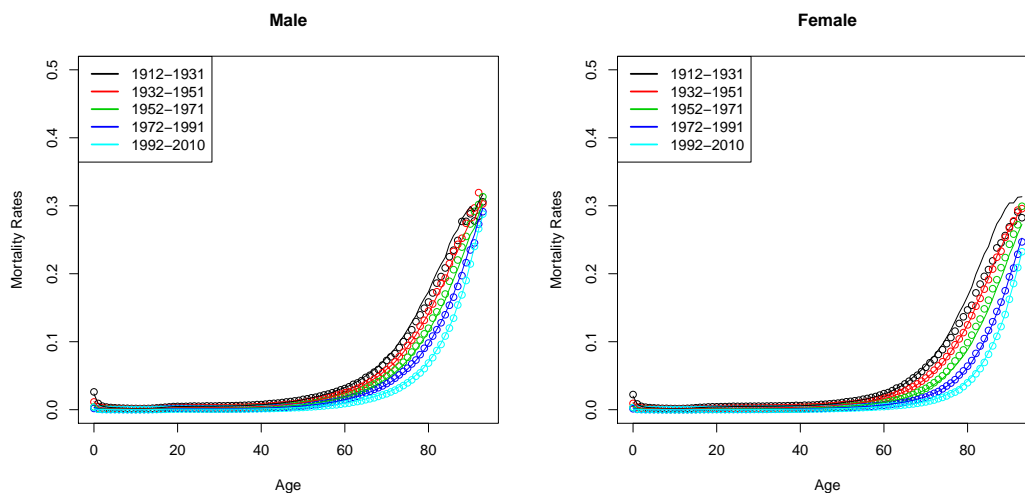


Figure 15.4: Observed (dots) and estimated (lines) mortality rates

We are now able to calculate the estimated mortality rates by plugging all estimators into the model equation (15.1). We do this for the same 5 year groups as before where all but the last one contain 20 successive years. The results are shown in Figure 15.4. We observe that the estimation is very accurate. One can only see small difference between the estimated and observed mortality rates with the naked eye.

Next, we do a logit-transformation on the mortality rates to have a clearer view on the estimated rates of the lower ages. The mortality logit-rates are plotted in Figure 15.5. We observe that they differ a bit from the observed ones for ages below 40. Nevertheless, we get the feeling that the model is rather accurate. However, we get no sound conclusion without analysing the residuals.

Figure 15.6 shows the residuals of the estimated mortality rates. We observe that after age 70, the estimated values differ from the observed ones but the difference between the rates lays between  $\pm 3\%$  for most of them. The residuals are correlated and their expectations are

## 15 Application of the Models of Lee and Carter

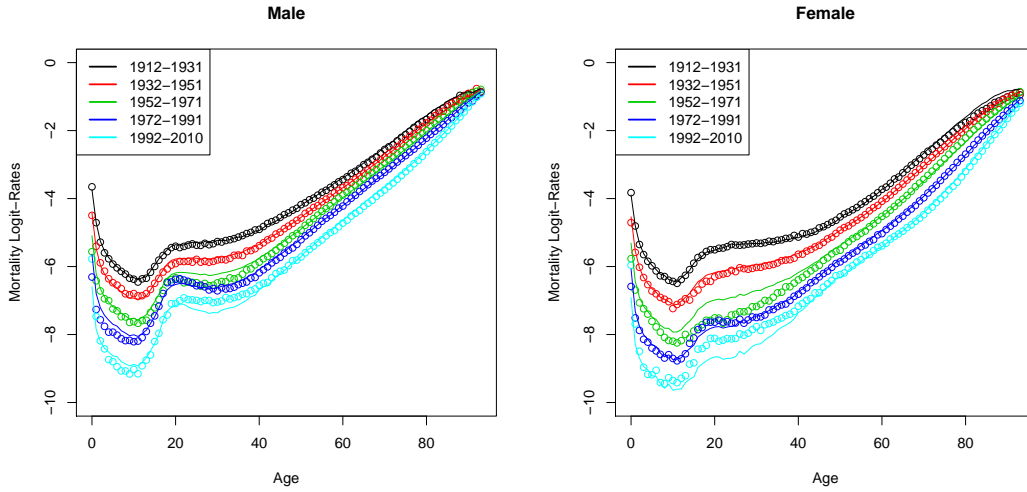


Figure 15.5: Observed (dots) and estimated (lines) mortality logit-rates

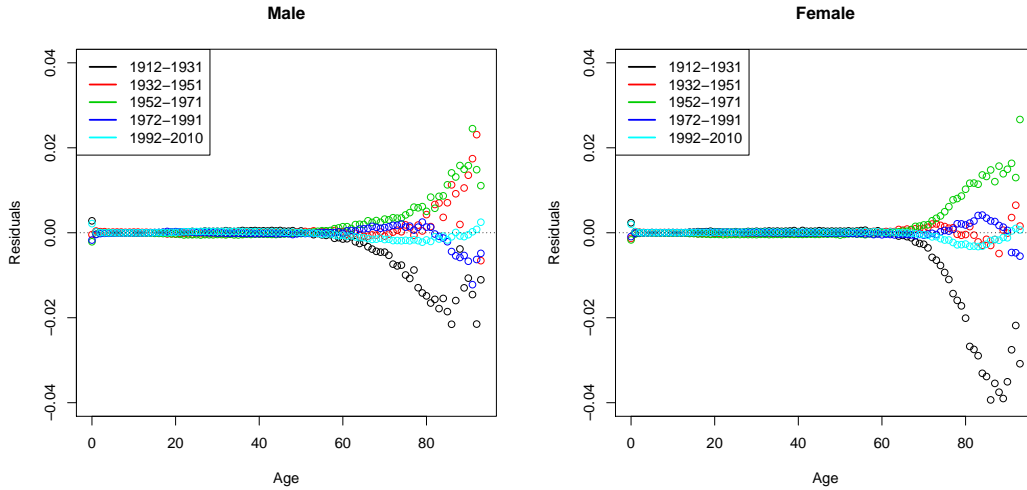


Figure 15.6: Residuals

not equal to zero. What stands out, though, is that for the three year groups from 1932 to 1991, we generally underestimate the mortality rates whereas for the earliest and the latest years, it is precisely the other way around. Luckily, the residuals are still close to zero after 1992 so this does not necessarily mean that it will remain this way in the future.

### Forecasting and Extrapolating

As a next step, we want to know how the mortality rates might look like in the future. Hence, we forecast the time-varying index  $(k_t)_t$  to the future according to (9.2). We use all available years for the estimation of the drift whereas for the variability, we only take the years after 1960 since we see in Figure 15.7 that the variance is nearly constant afterwards. This yields the following two models:

$$\text{Male: } k_{t+1} = k_t - 1.8312 + 2.6401 Z_t, \quad (15.2)$$

$$\text{Female: } k_{t+1} = k_t - 2.5815 + 3.2250 Z_t, \quad (15.3)$$

15 Application of the Models of Lee and Carter

for  $(Z_t)_t$  being *i.i.d.* standard normal random variables. We find that  $\hat{\mu} < 0$  for both genders which implies that the index is decreasing. The standard deviation is not large compared to the drift so the index will not vary too much in the future.

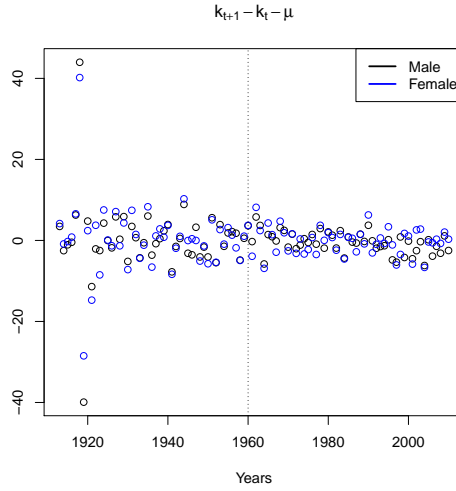


Figure 15.7:  $\hat{k}_{t+1} - \hat{k}_t - \hat{\mu}$

With these models, we are able to calculate mortality rate forecasts by inserting  $\hat{\mu}$  and  $\hat{c}$  into (9.4). On the left-hand side of Figure 15.8, we show the estimated mortality logit-rates in 2010 as well as the forecasted ones for year 2060. We find that the mortality curve is steeper in the future which is what we expect since it means that people generally will live longer in the future and mortality will increase faster for old people.

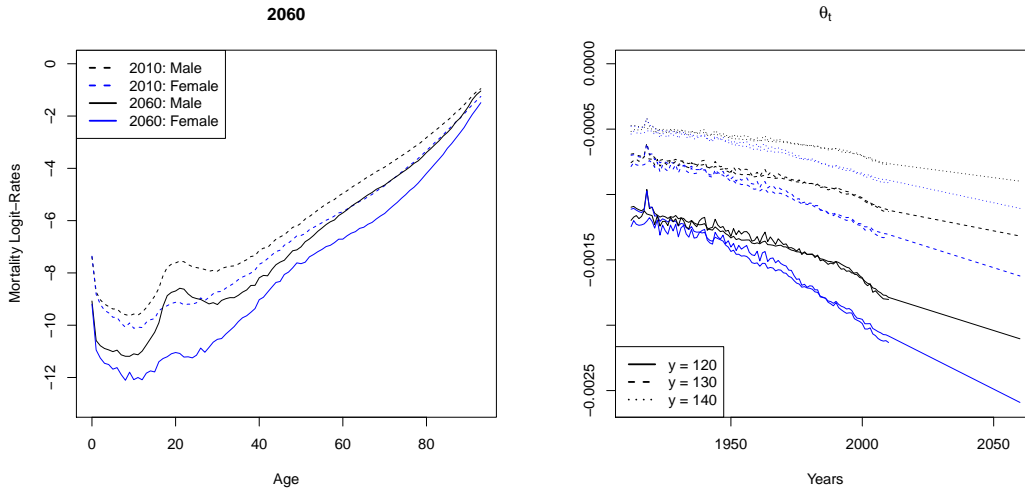


Figure 15.8: (LHS) Forecasted mortality logit-rates for year 2060, (RHS)  $\hat{\theta}_t$  from the extrapolation

Before we can calculate the three measures for longevity, we have to extrapolate the mortality rates according to Chapter 8. We do this for the maximal ages  $y = 120, 130$  and  $140$ . The estimators  $(\hat{\theta}_t)_t$  obtained from the estimated, observed and forecasted mortality rates are

shown on the right-hand side of Figure 15.8. The parameter estimators from the estimated and observed rates are very close and it seems like the overall trend has been captured nicely by the forecasted rates. However, it is possible that the slope should be even steeper, especially for the males. Nevertheless, the model seems to work nicely in the past and additionally makes realistic predictions for the future.

### Measures for Longevity

We now calculate the survival probability  $p_{0,t,j}$  for the years 1912 and 2010 using all three extended datasets to show how it evolved over the last 99 years. In Figure 15.9, it is observable that  $y$  only influences the probability for people of an age above 90 so its choice is not too important here. We find that people in 2010 live longer than those in 1912. It starts already for the lowest ages since the child death rate is very small in 2010. Additionally, it holds true for high ages since the age where the survival probability starts decreasing rapidly is much later in 2010 than it was 99 years earlier. What we detect as well is that the steepness of the curves increases over the years. This means that even though people live longer, the probability that someone dies increases much faster after the kink. We have already recognized this development on the left-hand side of Figure 15.8.

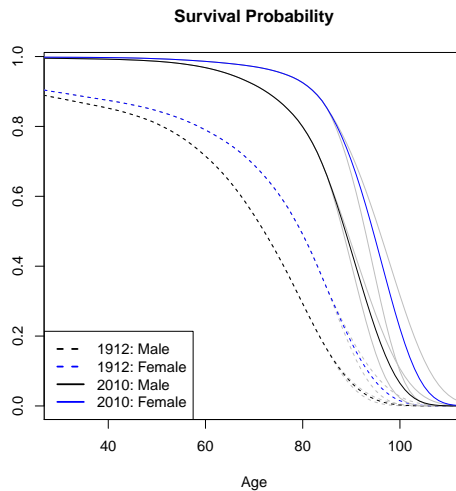


Figure 15.9: Survival probability  $p_{0,t,j}$  for  $t = 1912$  and  $2010$

Another indicator for the longevity of a society is the life expectancy of a person of age 30 or 65. We observe in Figure 15.10 that its behaviour for men and women is similar since the shape of the curves are similar in both cases. However, they do not coincide since the life expectancy for women is always higher than the one for men. We see that the life expectancy only clearly depends on the choice of  $y$  in future years. This is most visible for the life expectancy of a 30 year old woman. This means, the more we go into the future, the more the grey curves, which correspond to  $y = 120$  and  $140$ , and the coloured ones for  $y = 130$  differ from each other. Since the difference in the past is not as large, we only show some values of  $e_{x,t}$  for  $y = 130$  in Table 15.11.

An additional way to see how mortality changes with time is by calculating the price of the annuity we have already seen above. We have chosen three different interest rates, namely  $r = 0\%$ ,  $1\%$  and  $5\%$ . We have used all three datasets with  $y = 120$ ,  $130$  and  $140$  for plotting

## 15 Application of the Models of Lee and Carter

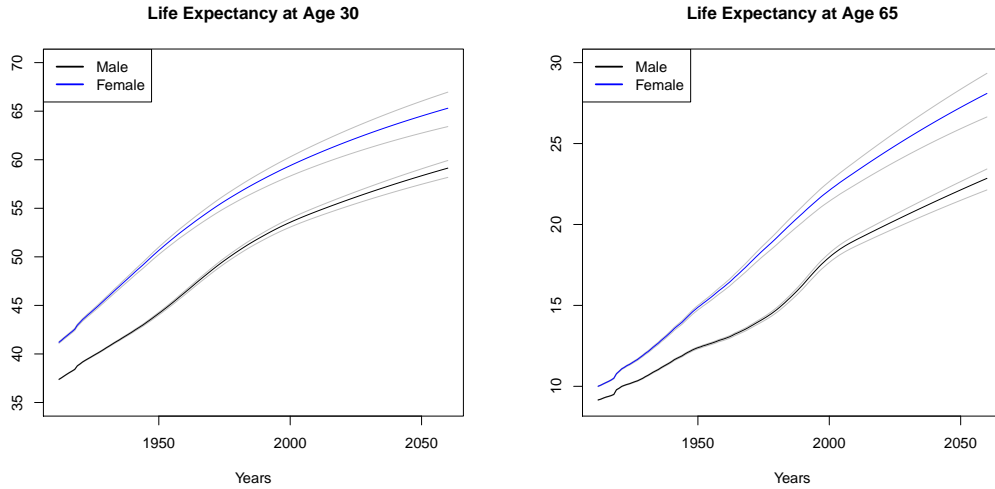


Figure 15.10: Life expectancy at age 30 and 65

		1912	1920	1930	1940	1950	1960	1970	1980	1990	2000	2010
30	Male	37.4	39.0	40.6	42.3	44.2	46.3	48.6	50.6	52.2	53.6	54.7
	Female	41.2	43.2	45.7	48.2	50.7	52.9	54.8	56.5	58.1	59.4	60.6
65	Male	9.2	9.9	10.6	11.5	12.4	12.9	13.7	14.7	16.3	18.0	19.0
	Female	10.0	10.9	12.0	13.4	14.9	16.2	17.7	19.2	20.7	22.1	23.3

Table 15.11: Life expectancy at age 30 and 65 for  $y = 130$

the price. The coloured lines in Figure 15.12 are the prices for  $y = 130$  and the grey ones are for the other two maximal ages. We see that  $y$  has only an effect on the price in the future which is largest for the smallest interest rate. In all cases, the prices are increasing over the years and higher for women than for men since we expect them to live longer. The curves are not that nice anymore, though. There are two small kinks in 1918 and around 2000. Furthermore, the figure shows that an increasing interest rate decreases the price of the annuity.

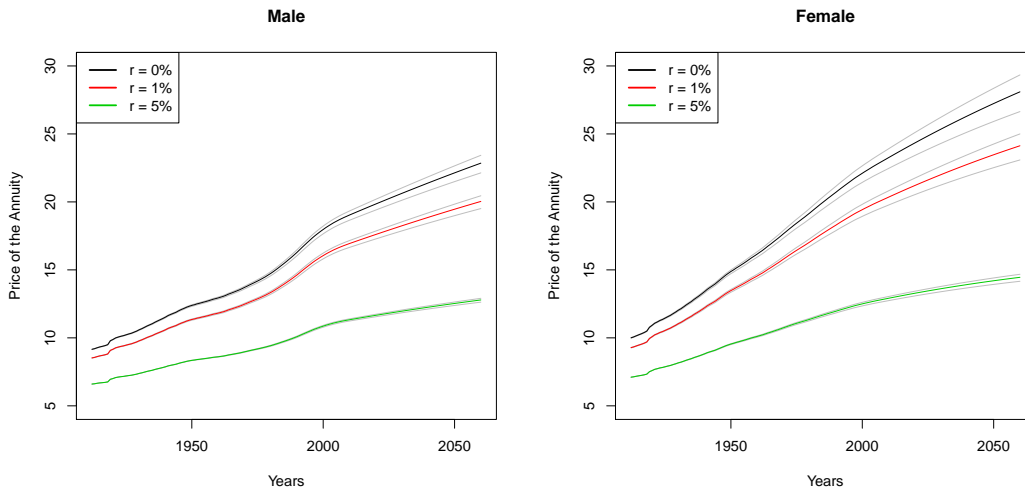


Figure 15.12: Price of the annuity for a person of age 65

In summary, we say that this model fits the data in the past very well and makes realistic forecasts for the future. Hence, we go on and have a look at the extended model. It will be interesting to see how much the results differ from the ones found in this section.

## 15.2 Extended Model

We recall the extended model of Lee and Carter given in Chapter 5. We modeled

$$\log(Q_{x,t,w}) = a_x + b_x k_t + d_w + \varepsilon_{x,t,w}. \tag{15.4}$$

Since the only additional information we have on the data is the gender, we choose  $w = 1$  if a person is male and  $w = 2$  otherwise even though we know that a constant spread of the mortality rates is not realistic for all ages.

### Fitting Procedure

We have only seen one way to estimate the parameters of this model, namely the Newton-Raphson method described in Algorithm 5.4. Thus, the estimation of  $a_x$ ,  $b_x$ ,  $k_t$  and  $d_w$  is done simultaneously. For  $\delta = 0.1$ , we have to do 16 iterations before the algorithm converges.

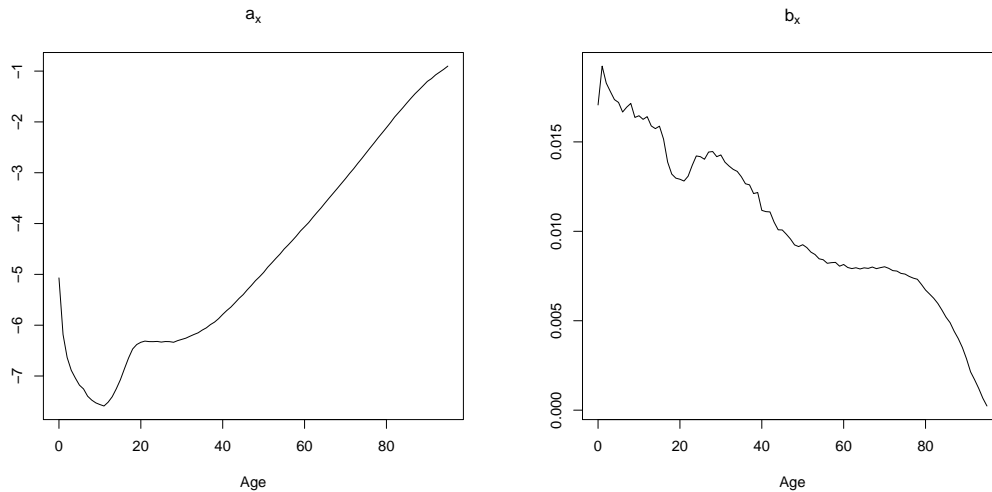


Figure 15.13: (LHS)  $\hat{a}_x$ , (RHS)  $\hat{b}_x$

In Figure 15.13, we see the estimators for  $a_x$  and  $b_x$ . They look similar to the classical model even though there is only one parameter for both genders.  $(\hat{a}_x)_x$  is increasing after the age of 20 and  $(\hat{b}_x)_x$  has a decreasing trend. The bump we recognized for  $\hat{b}_x$  in the classical model for the male data is still visible but not as pronounced anymore.

Figure 15.14 shows the estimator of  $k_t$ . The time-varying index is again decreasing and has a big peak in 1918 due to the Spanish flu. We also show here its trend which will be calculated later on. We observe that it is very accurate. For the gender factor  $d_w$ , we get that  $d_1 = 0$  and  $d_2 = -0.3831$ . As expected,  $d_2 < 0$  since the mortality rates for women are generally lower than for men.

15 Application of the Models of Lee and Carter

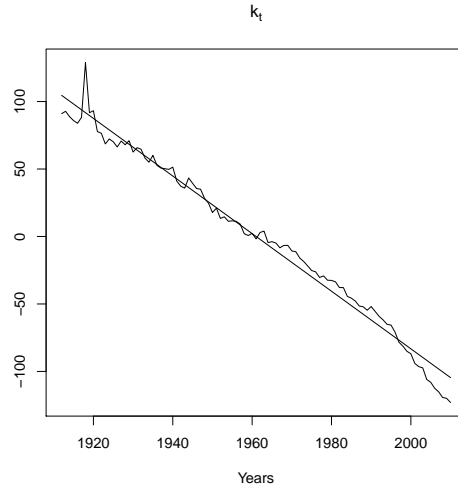


Figure 15.14:  $\hat{k}_t$

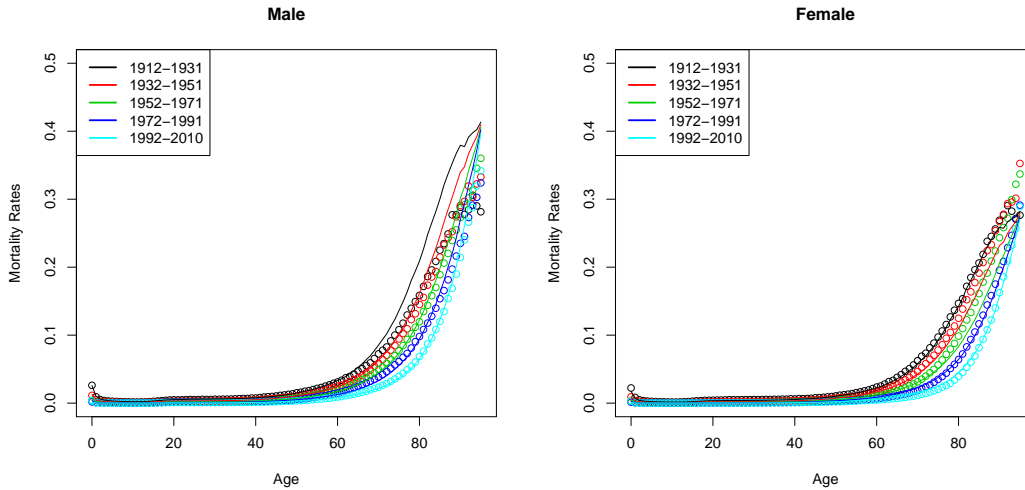


Figure 15.15: Observed (dots) and estimated (lines) mortality rates

From now on, we show everything for men and women separately. As a first step, we use the estimators calculated above to get estimated mortality rates in the past. In Figure 15.15, we see the observed and estimated mortality rates for the same 5 year groups as in the classical model. We already observe that for men, the estimated rates are generally too high whereas the ones for women are too low. This is most pronounced for the highest ages. From this figure, it is already clear that the difference between the mortality rates for men and women is not only caused by a constant factor as it is assumed in the model.

This fact is also visible when plotting the logit-transformations of the mortality rates in Figure 15.16. The estimated logit-rates are much lower than the observed ones but the difference for the smallest ages is not as large. This means that for these ages, the constant spread is more reasonable than for the older ones.

Next, we have a look at the residuals from this estimation procedure. As anticipated due to the mortality rate plots, the residuals for men are mostly negative whereas the ones for women

## 15 Application of the Models of Lee and Carter

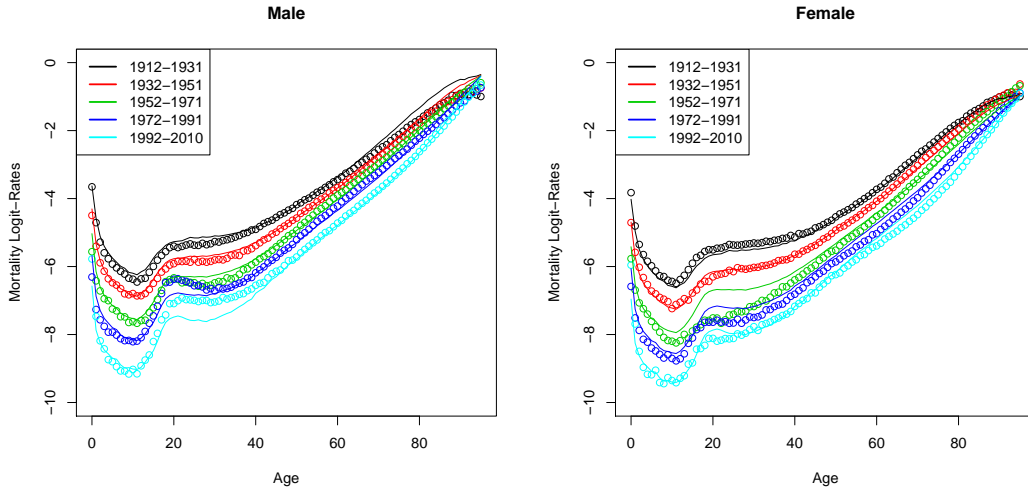


Figure 15.16: Observed (dots) and estimated (lines) mortality logit-rates

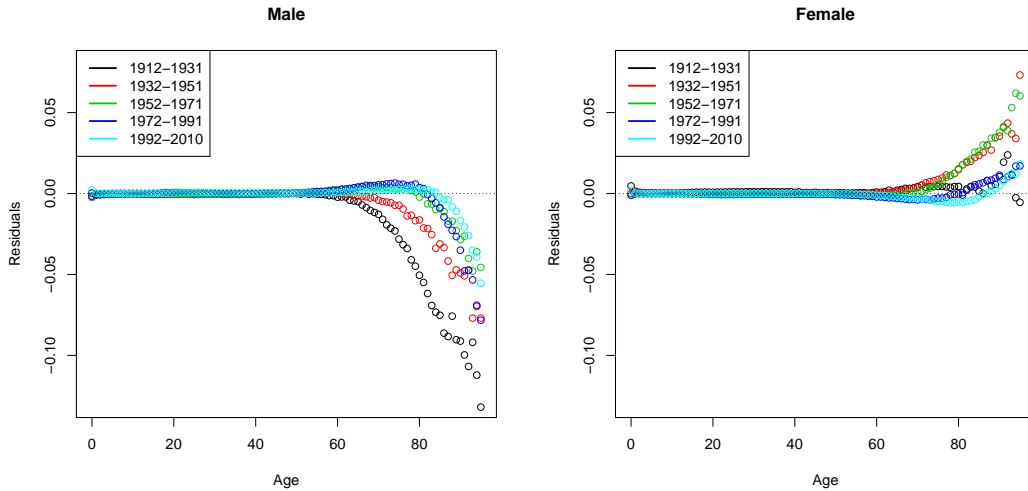


Figure 15.17: Residuals

are positive. This means that for men, the rates are mostly overestimated and the ones for women are underestimated. This discrepancy is caused by the constant spread  $d_w$ . Hence, this model will never work perfectly for both genders at once. The residuals are very far away from zero for ages above 70. In the maximum, which is achieved for men, the estimated rates are 13% too high. Therefore, we conclude that the model does not yield a good fit for the data.

### Forecasting and Extrapolating

Nevertheless, we want to find mortality rates in the future. Since the model only depends on time via  $k_t$ , we just need to forecast this index by modeling it as a random walk with drift. We estimate  $\mu$  by using all values  $(\hat{k}_t)_t$  from the past whereas for the estimation of  $c$ , we only use the years after 1960 because we see in Figure 15.18 that the variability is untypically large in the earlier years. This yields

$$k_{t+1} = k_t - 2.1342 + 2.6905 Z_t, \quad (15.5)$$



15 Application of the Models of Lee and Carter

for *i.i.d.* standard normally distributed random variables  $(Z_t)_t$ . We observe that  $\hat{\mu} < 0$  which means that  $k_t$  is decreasing. The standard deviation has nearly the same size as the drift so the values are not expected to vary too much.

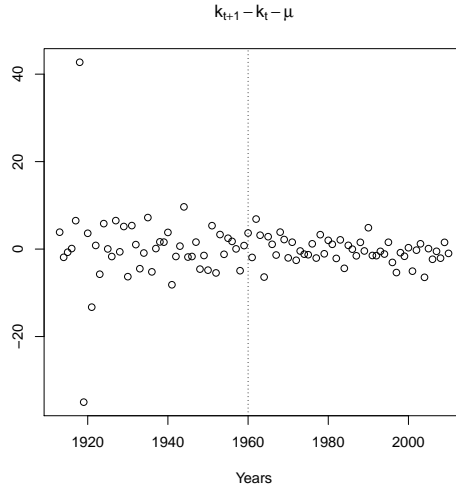


Figure 15.18:  $\hat{k}_{t+1} - \hat{k}_t - \hat{\mu}$

We now calculate the forecasted mortality rates in the same way as in the classical model. The resulting mortality logit-rates are shown on the left-hand side of Figure 15.19. We see that the forecasted rates in 2060 are smaller than the ones estimated for 2010 and that they are perfectly parallel due to the constant spread for all ages. This is what we expect so the forecasting procedure seems to be reasonable.

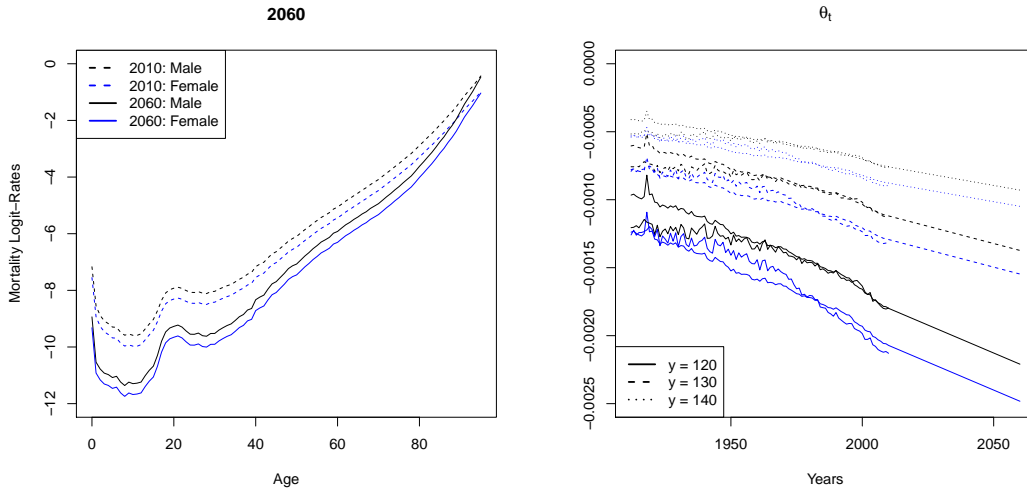


Figure 15.19: (LHS) Forecasted mortality logit-rates for year 2060, (RHS)  $\hat{\theta}_t$  from the extrapolation

As a next step, we use all results we have found so far to extend the mortality rates to ages up to  $y = 120, 130$  and  $140$ . On the right-hand side of Figure 15.19, we show the estimators of  $(\theta_t)_t$  from the extrapolation of the estimated and forecasted mortality rates where we added

the ones we find with the observed mortality rates. The estimated and observed parameters are rather close together after 1980 but not for the earlier years. Nevertheless, the trend in the future seems to be realistic which implies that the forecasted mortality rates are sensible.

**Measures for Longevity**

We now calculate  $p_{0,t,j}$  for  $t = 1912$  and  $2010$  and all values of  $y$ . In Figure 15.20, the coloured lines are the ones for  $y = 130$ . We see a small difference in the probabilities for ages above 95 for the three values of  $y$  but luckily, the kink we are most interest in happens earlier in both years. This means, the choice of  $y$  is not too important for this analysis. We observe that the difference between the two years is rather large, though. The shape of the curves in 1912 does not look like the one in 2010 since it is more flat until age 60 and starts decreasing much later. This implies that the survival probability for young people is higher and it decreases faster for the old ones.

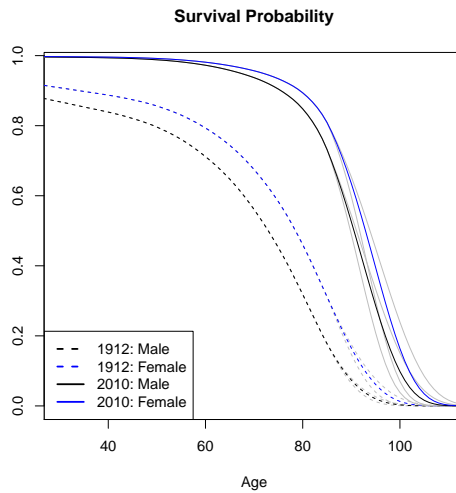


Figure 15.20: Survival probability  $p_{0,t,j}$  for  $t = 1912$  and  $2010$

We calculate the life expectancy for people of age 30 and 65 according to (11.3) as well. All results are plotted in Figure 15.21 where the grey lines are the life expectancies for  $y = 120$  and  $140$  and the coloured ones are for  $y = 130$ . We see that the choice of  $y$  almost exclusively influences the results in the future. For all values of  $y$ , the life expectancy is increasing with age and higher for women than for men. The curves are not completely smooth but we observe that they are parallel for the two genders because of the constant spread. In Table 15.22, we show some explicit values of  $e_{x,t}$  in the past for  $y = 130$  where we recognize the same characteristics as in the figure above.

		1912	1920	1930	1940	1950	1960	1970	1980	1990	2000	2010
30	Male	36.9	38.7	40.8	42.9	45.2	47.5	49.8	51.7	53.4	54.9	56.1
	Female	41.7	43.4	45.4	47.4	49.5	51.6	53.6	55.3	56.7	58.0	59.1
65	Male	8.4	9.2	10.0	11.1	12.2	13.1	14.2	15.4	17.0	18.6	19.8
	Female	10.9	11.8	12.7	13.9	15.0	15.9	17.1	18.4	20.0	21.4	22.5

Table 15.22: Life expectancy at age 30 and 65 for  $y = 130$

## 15 Application of the Models of Lee and Carter

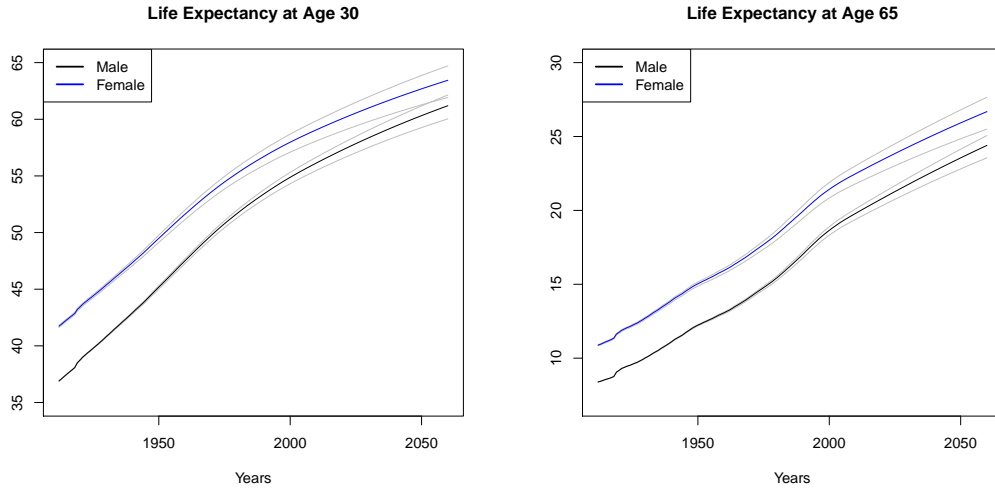


Figure 15.21: Life expectancy at age 30 and 65

Finally, we calculate the price of the same annuity as in the last section for all three extended datasets to have an additional measure for longevity. We assume that the interest rate is constant and equal to  $r = 0\%$ ,  $1\%$  and  $5\%$ . The results, shown in Figure 15.23, indicate that the price is increasing over the years. The difference between the three values of  $y$  is only visible in the future for low interest rates and the curves are higher for women than for men. The difference with respect to the interest rate is the same as in the classical model. The curves are shifted if  $r$  changes and the price is highest for the smallest interest rate. The curves have again two small kinks in 1918 and around 2000 just like for the classical model.

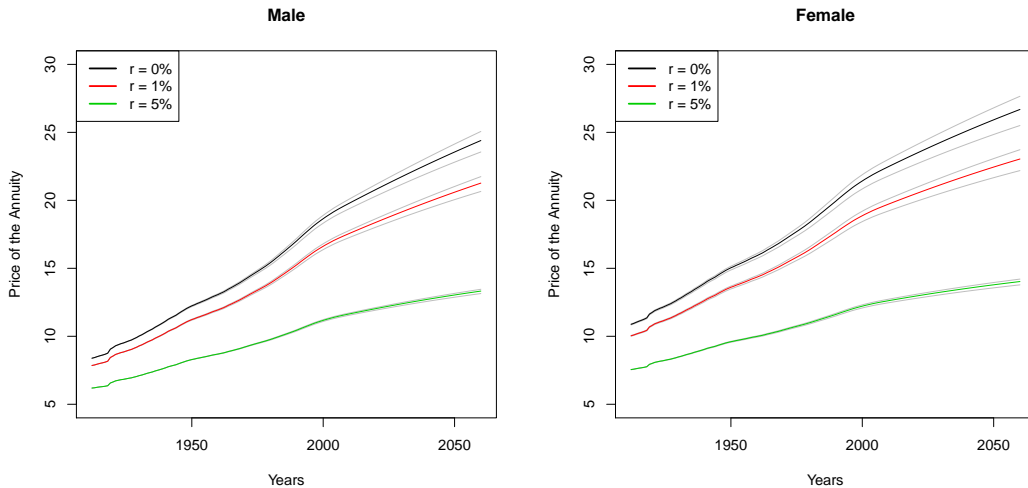


Figure 15.23: Price of the annuity for a person of age 65

We conclude that the classical model of Lee and Carter yields good estimators for the past and reasonable forecasts for the future. The extended model, on the other hand, is not a good fit for the data since  $d_w$  is constant for all ages and years which is not realistic. With these remarks, we finish this application and go on with showing how good the Cairns-Blake-Dowd model fits the same data.

# 16 Application of the Cairns-Blake-Dowd Model

The next model we are going to analyse is the Cairns-Blake-Dowd model. In Chapter 6, we introduced the following model for the mortality rate that already contains some a priori knowledge about the mortality curve:

$$q_{x,t}^{\Theta} = \frac{\exp(\Theta_{t,1} + x \Theta_{t,2})}{1 + \exp(\Theta_{t,1} + x \Theta_{t,2})} \quad (16.1)$$

for two stochastic processes  $(\Theta_{t,1})_t$  and  $(\Theta_{t,2})_t$ . Since we know that it is only sensible for high ages we only apply it to ages above 39. We have chosen this boundary since the observed mortality logit-rates are close to linear for these ages.

## Fitting Procedure

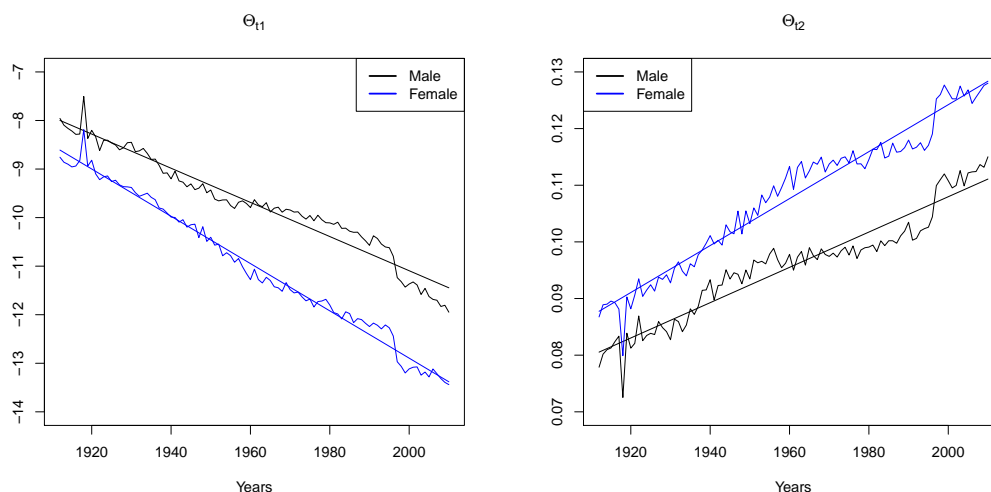


Figure 16.1: (LHS)  $\hat{\Theta}_{t,1}$ , (RHS)  $\hat{\Theta}_{t,2}$

We start with estimating  $\Theta_{t,1}$  and  $\Theta_{t,2}$  for all years separately using a GLM with a logit-link. In Figure 16.1, we see the parameter estimators for both genders. We observe that when neglecting the fluctuations,  $(\hat{\Theta}_{t,1})_t$  is decreasing in  $t$  and negative whereas  $(\hat{\Theta}_{t,2})_t$  is increasing in  $t$  and positive. This is exactly what we expect in a realistic model since it implies that the mortality rate decreases for a fixed age and the mortality curve becomes steeper over the years. The big peaks in 1918 are caused by the Spanish flu just like for  $(k_t)_t$  in the last chapter. We also plot the trends of the two processes that will be estimated later on. Both of them are rather accurate.

We use the values  $(\hat{\Theta}_{t,1})_t$  and  $(\hat{\Theta}_{t,2})_t$  to get estimated mortality rates by plugging them into the model equation (16.1). In Figure 16.2, we see the observed and estimated mortality rates for the 5 year groups we already had a look at in the earlier chapters. We observe that the estimated mortality rates are mostly lower than the observed ones for the ages from 70 to 90. Afterwards, this changes for the first three year groups.

## 16 Application of the Cairns-Blake-Dowd Model

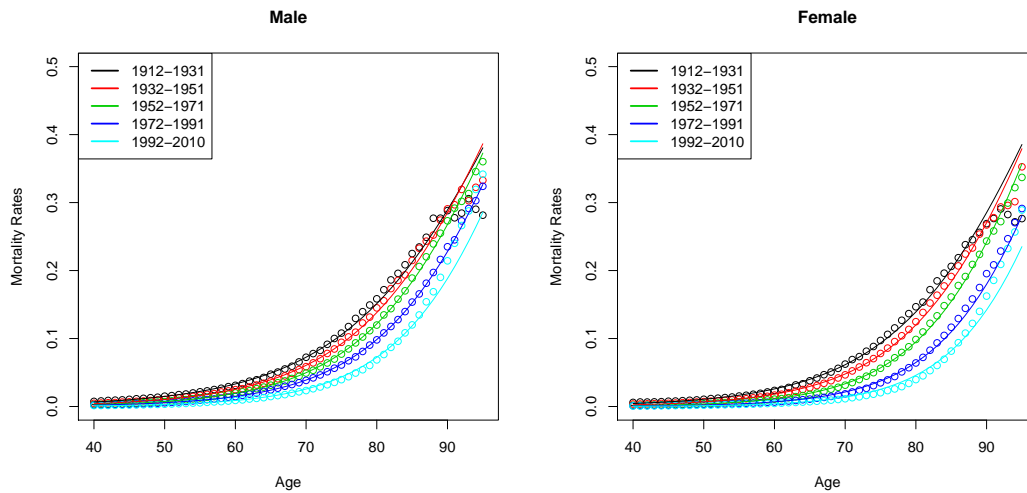


Figure 16.2: Observed (dots) and estimated (lines) mortality rates

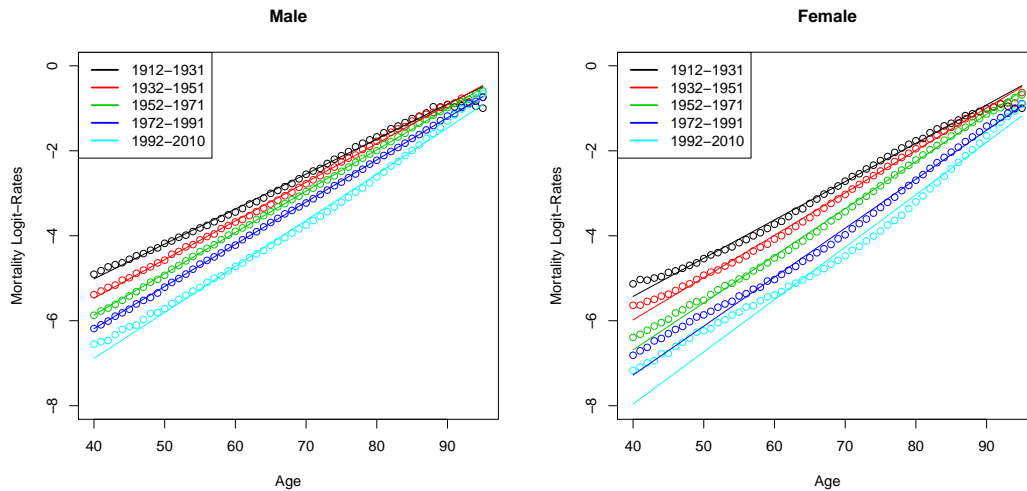


Figure 16.3: Observed (dots) and estimated (lines) mortality logit-rates

To get an even clearer view on the mortality rates, we logit-transform them and plot the resulting rates in Figure 16.3. We observe here that the fit for men between ages 40 and 60 is better than the one for women. Due to the specific form of the model, the fitted values lay on a straight line. Since the observed logit-rates are nowhere perfectly linear, it is never possible for this model to replicate them precisely. This means that more factors might be needed to get more accurate estimators.

We want to know more about the quality of the estimated mortality rates. Therefore, we have a look at their residuals in Figure 16.4. We see that up to age 70, the residuals are close to zero. Between the ages 70 and 90, they increase slightly but are still not too large. However, the estimated mortality rates differ highly from the observed ones above age 90. The residuals are both positive and negative where both signs occur for all groups. From this plot, we conclude that the model fits the data well for ages from 40 to 90 but the fit is rather bad outside this range.

## 16 Application of the Cairns-Blake-Dowd Model

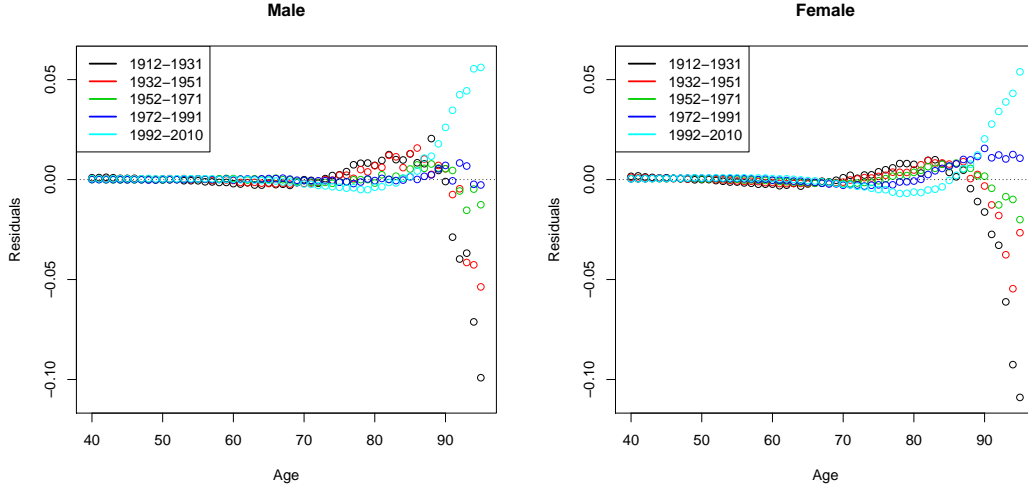


Figure 16.4: Residuals

### Forecasting and Extrapolating

After having estimated the mortality rates in the past, we want to know how they will behave in the future. Therefore, we forecast  $\Theta_t = (\Theta_{t,1}, \Theta_{t,2})^\top$  using (9.6). For the estimation of the drift, we use all 99 years but for the covariance, we only choose the years from 1965 to 2010 because we observe in Figure 16.5 that the variability is more stable for these years. This yields the following two-dimensional random walks with drift:

$$\text{Male: } \Theta_{t+1} = \Theta_t + \begin{pmatrix} -0.0352 \\ 0.0003 \end{pmatrix} + \begin{pmatrix} 0.1032 & -0.0015 \\ 0 & 0.0003 \end{pmatrix} \mathbf{Z}_t, \quad (16.2)$$

$$\text{Female: } \Theta_{t+1} = \Theta_t + \begin{pmatrix} -0.0487 \\ 0.0004 \end{pmatrix} + \begin{pmatrix} 0.1193 & -0.0016 \\ 0 & 0.0003 \end{pmatrix} \mathbf{Z}_t, \quad (16.3)$$

for *i.i.d.* random vectors  $(\mathbf{Z}_t)_t$  with a bivariate standard normal distribution. As anticipated in a realistic model,  $\hat{\mu}_1 < 0$  and  $\hat{\mu}_2 > 0$  for both genders. We see that the estimated drift and

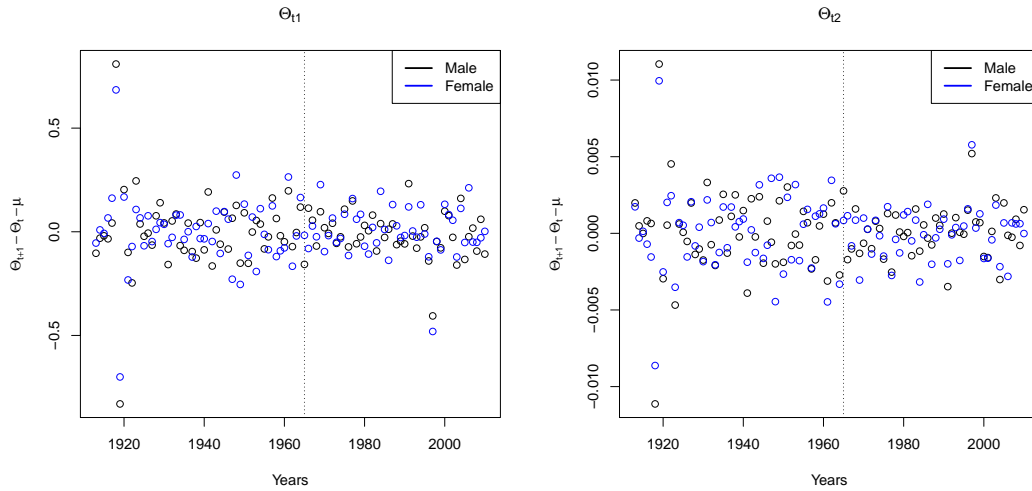


Figure 16.5: (LHS)  $\hat{\Theta}_{t+1,1} - \hat{\Theta}_{t,1} - \hat{\mu}_1$ , (RHS)  $\hat{\Theta}_{t+1,2} - \hat{\Theta}_{t,2} - \hat{\mu}_2$

covariance for men and women are similar which means that the processes are alike. We find that  $(\Theta_{t,1})_t$  and  $(\Theta_{t,2})_t$  are highly negatively correlated for both genders. We have already observed all these characteristics in Figure 16.1.

Next, we use the estimators  $\hat{\mu}$  and  $\hat{C}$  to find forecasted mortality rates. We observe on the left-hand side of Figure 16.6 that the mortality logit-rates in 2060 are smaller than the ones estimated for 2010 and that the slope of the logit-rate curve is larger. This means that people are expected to live longer 50 years from now.

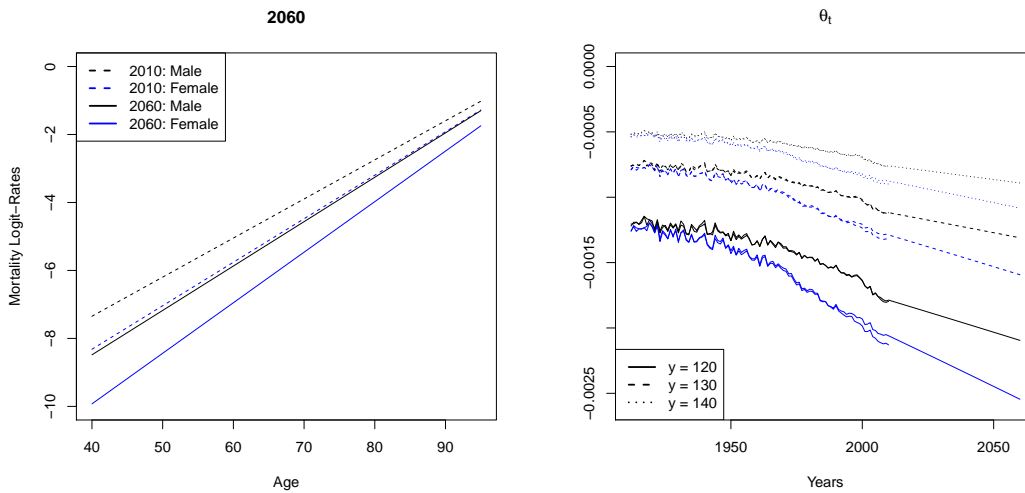


Figure 16.6: (LHS) Forecasted mortality logit-rates for year 2060, (RHS)  $\hat{\theta}_t$  from the extrapolation

The estimated and forecasted mortality rates are now extrapolated to maximal ages  $y = 120, 130$  and  $140$ . We see the estimators of  $(\theta_t)_t$  from the extrapolation on the right-hand side of Figure 16.6 for the observed, estimated and forecasted mortality rates. The resulting parameters for the observed and estimated ones are very close together for all three values of  $y$ . This implies that the model yields accurate estimators for the past. Since it also captures the slope of  $(\theta_t)_t$  very nicely, the forecasts are reasonable as well. We also would like to extrapolate the estimated mortality rates to age 30. Since the model cannot be applied to those ages directly, we have to construct them ourselves. Since the mortality rates are rather small for ages from 30 to 39, it will not be too important how this is done. We decide to estimate the slope of the observed mortality logit-rate curve for those ages in each year and use the mean slope to extrapolate the logit-rates linearly from 40 to 30. This is reasonable since the logit-rate curve is close to linear for those ages but it has a smaller slope than for ages above 40.

### Measures for Longevity

Due to the two extrapolations, we now have estimated and forecasted mortality rates for people with an age of 30 and higher at our disposal. However, this is still not enough to calculate the survival probability  $p_{0,t,j}$ . Thus, we directly go on to calculating the life expectancies at ages 30 and 65 for all three extended datasets. The results are shown in Figure 16.7 where the coloured lines are for  $y = 130$  and the grey lines for  $y = 120$  and  $140$ . We

## 16 Application of the Cairns-Blake-Dowd Model

observe that the life expectancy is increasing over the years and the difference between men and women is clearly present. We find that the effect of the choice of  $y$  is largest for women in future years. We find that the curves are rather nice with one exception around 2000 where we see a small kink. Some explicit values in the past for  $y = 130$  are written in Table 16.8. We see here again that there is a significant difference between the 99 years from 1912 to 2010.

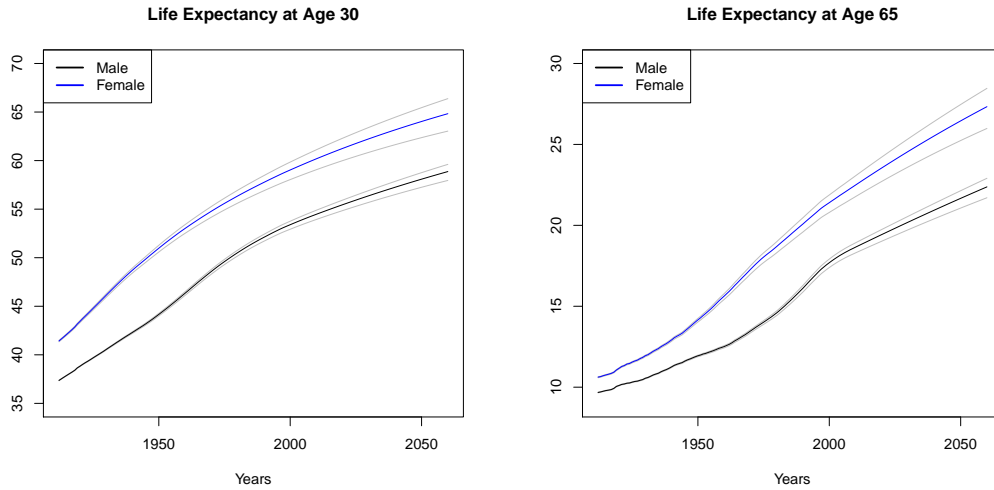


Figure 16.7: Life expectancy at age 30 and 65

		1912	1920	1930	1940	1950	1960	1970	1980	1990	2000	2010
30	Male	37.4	38.8	40.6	42.3	44.2	46.3	48.6	50.5	52.1	53.4	54.5
	Female	41.4	43.4	46.1	48.7	51.0	53.0	54.8	56.4	57.8	59.0	60.2
65	Male	9.7	10.1	10.6	11.3	11.9	12.5	13.5	14.6	16.1	17.7	18.7
	Female	10.6	11.2	11.9	12.9	14.2	15.6	17.3	18.7	20.1	21.4	22.5

Table 16.8: Life expectancy at age 30 and 65 for  $y = 130$

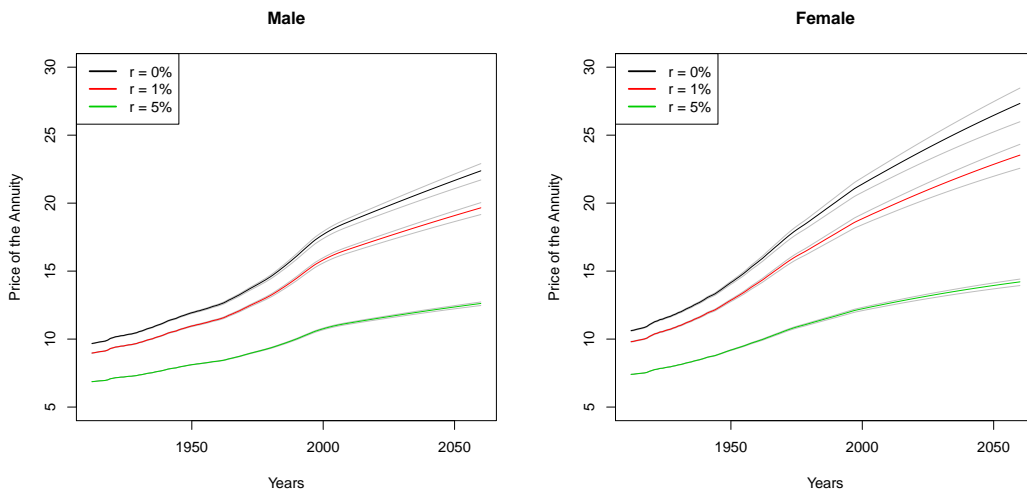


Figure 16.9: Price of the annuity for a person of age 65



## 16 Application of the Cairns-Blake-Dowd Model

Lastly, we calculate the price of the same annuity as in the last three chapters. We recall that we plugged  $r = 0\%$ ,  $1\%$  and  $5\%$  into (11.4). The colours in Figure 16.9 are the same as in Figure 16.7. We observe that  $y$  does not influence the price very much in the past but in the future, we see a clear difference between the three curves. The largest deviation is found for women with an interest rate of  $0\%$ . The prices have an increasing trend due to the higher life expectancy of a person of age 65 in 2060 against one in 1912. They differ in the same way with respect to  $r$  as we have already seen in the last chapters which means that an increasing interest rate reduces the price.

To sum up, we say that the model yields good results in the past as well as in the future for people with ages from 30 to 90. However, we were not able to calculate the survival probability  $p_{0,t,j}$  since we cannot reasonably estimate the mortality rates for ages below 30. Nevertheless, we continue and have a look at the last group of models which will hopefully produce similarly good estimators and forecasts as the ones we have found in this chapter.

# 17 Application of Adapted Nolfi Models

To come to an end with the applications, we are lastly discussing the adapted Nolfi models from Chapter 7. We have introduced the following three models:

$$\text{Model 1: } q_{x,t}^{\ominus} = q_{x,t_0} \exp(-\lambda_x \alpha_{t,2} (t - t_0)), \quad (17.1)$$

$$\text{Model 2: } q_{x,t}^{\ominus} = \alpha_{t,1} q_{x,t_0} \exp(-\lambda_x (t - t_0)), \quad (17.2)$$

$$\text{Model 3: } q_{x,t}^{\ominus} = \alpha_{t,1} q_{x,t_0} \exp(-\lambda_x \alpha_{t,2} (t - t_0)). \quad (17.3)$$

## Fitting Procedure

As a first step, we estimate the parameters. The first one we need is  $\hat{\lambda}_x$  since this is the same in all three models. To do so, we fix  $t_0 = 1912$  and choose for  $q_{x,t_0}$  the adjusted mortality rates (2.6) in 1912 we get with the clearing method by Whittaker-Henderson with equal weights,  $v = 2$  and  $g = 1.5$ , see Figure 13.1 for more details. Unfortunately, we find that  $\hat{\lambda}_{93} < 0$  which is not allowed. Therefore, we only work with the data up to age 92 in this chapter. The estimation of  $\lambda_x$  is carried out according to (7.4) for all ages  $x \leq 92$ . Since this was already done in Section 13.1, we only recall the results in Figure 17.1.

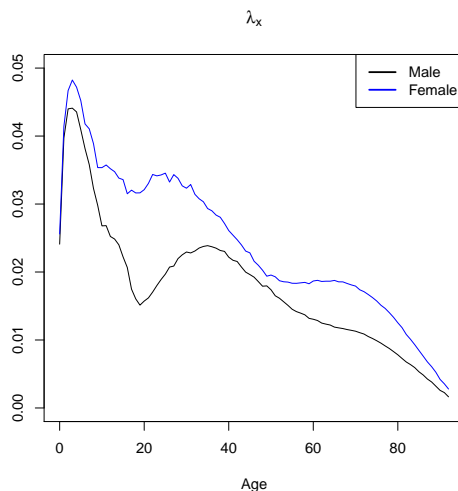


Figure 17.1:  $\hat{\lambda}_x$

The estimation of the parameters  $\alpha_{t,1}$  and  $\alpha_{t,2}$  is done separately for each model and each year where we use the fitting procedures introduced in Section 7.2. However, we apply the Newton-Raphson method from the classical model of Lee and Carter for Model 1 because the results in Chapter 15 were that convincing. For the initial value  $\alpha_{t,2} = 1$  and  $\delta = 0.1$ , we have to repeat the algorithm 6 times for men and 7 times for women. In Model 1 and 3, we set  $\hat{\alpha}_{t,2} = 0.0001$  for  $t$  around 1918 since the positivity would have been violated otherwise. All parameter estimators are shown in Figure 17.2. We observe that both estimators take rather untypical values for the years up to 1930 but  $\hat{\alpha}_{t,2}$  at least looks similar for the later years for both models. Unfortunately, it is not clear if these two parameters really follow a linear trend. Therefore, it may be challenging to forecast them to the future. All further analysis will now be done for the three models separately.

## 17 Application of Adapted Nolfi Models

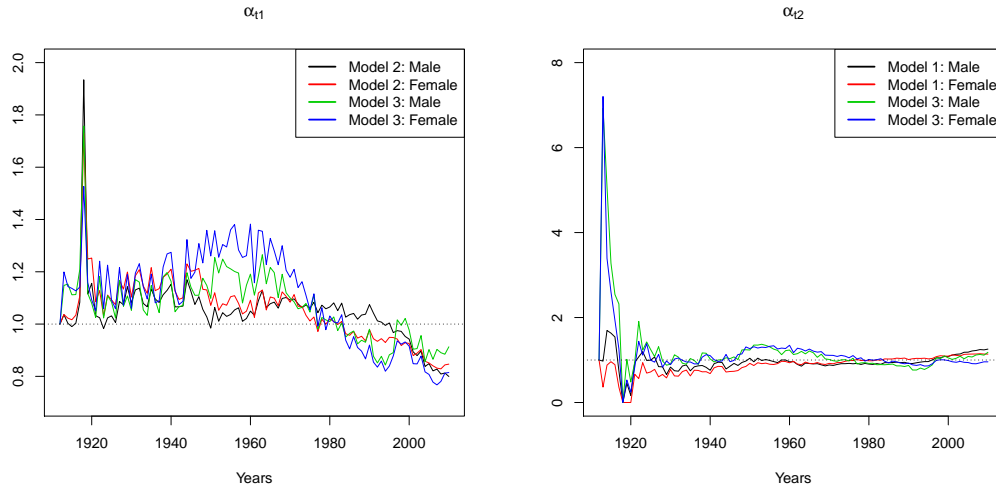


Figure 17.2: (LHS)  $\hat{\alpha}_{t,1}$ , (RHS)  $\hat{\alpha}_{t,2}$

### 17.1 Model 1

To get estimated mortality rates for Model 1, we plug the estimators  $(\hat{\alpha}_{t,2})_t$  into the model equation (17.1). In Figure 17.3, we find that until 1971, the estimated mortality rates above age 80 are below the observed ones. Afterwards, it seems to be the other way around. Nevertheless, the curves seem reasonable.

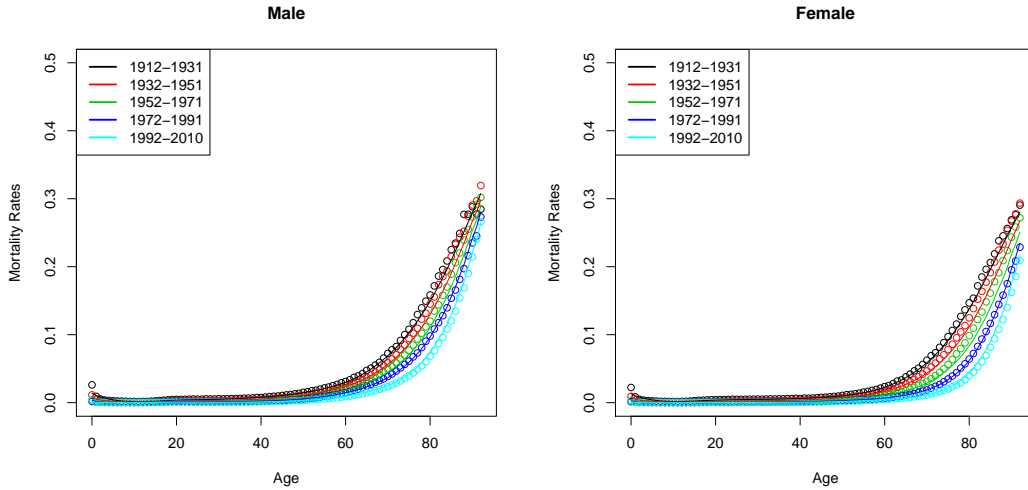


Figure 17.3: Observed (dots) and estimated (lines) mortality rates

To see the quality of the fit even better, we logit-transform the mortality rates and plot them in Figure 17.4. We see from these two plots that the fit is rather nice. The kinks for ages below 20 are captured even though they are a bit too steep in the earlier years. However, this is probably caused by the adjustment of  $q_{x,t_0}$ .

To come to a final conclusion, we have a look at the residuals for these estimators. We observe in Figure 17.5 that they are mostly close to zero for ages below 70. Afterwards, they

## 17 Application of Adapted Nolfi Models

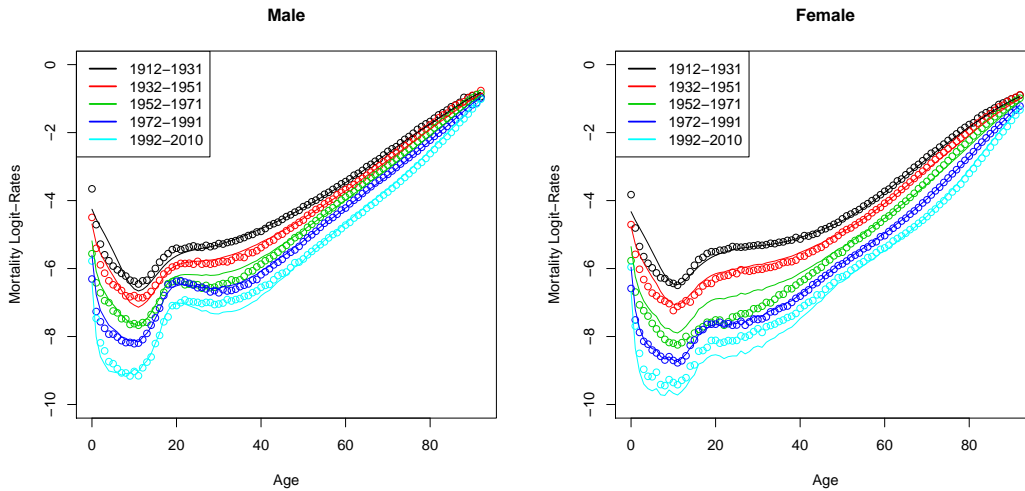


Figure 17.4: Observed (dots) and estimated (lines) mortality logit-rates

increase to a maximum of +2.9%. We additionally see that the residuals are neither always positive nor always negative. What we do find, however, is that for the last 19 years, the mortality rates are overestimated which could mean that it remains this way in the future. Nevertheless, we conclude that the model performs rather nicely in the past.

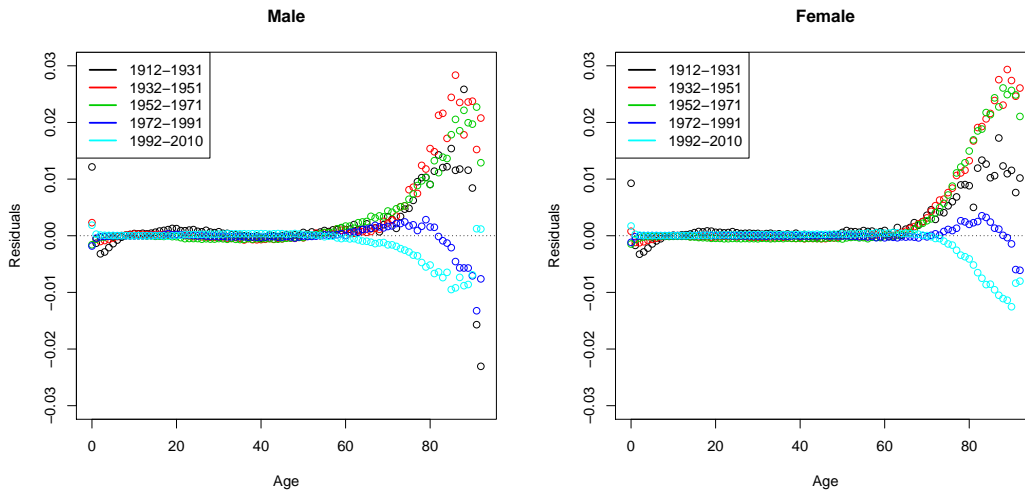


Figure 17.5: Residuals

### Forecasting and Extrapolating

To see if this is also true in the future, we need to forecast the mortality rates. Since model (17.1) only depends on time via  $\alpha_{t,2}$ , it is enough to forecast this process with a one-dimensional random walk with drift.  $\mu$  is estimated using all values  $(\hat{\alpha}_{t,2})_t$  from the past whereas the estimation of  $c$  only uses the years from 1965 to 2010 to get more reasonable estimators, see Figure 17.6. This yields the following two models:

$$\text{Male: } \alpha_{t+1,2} = \alpha_{t,2} + 0.0020 + 0.0188 Z_t, \quad (17.4)$$

$$\text{Female: } \alpha_{t+1,2} = \alpha_{t,2} + 0.0061 + 0.0177 Z_t, \quad (17.5)$$

## 17 Application of Adapted Nolfi Models

where  $(Z_t)_t$  are *i.i.d.* standard normal random variables. We observe that  $\hat{\mu}$  is rather small and positive for both genders which means that  $(\alpha_{t,2})_t$  is slowly increasing. This implies that it remains positive for all years which is what we need. Since the standard deviation is small as well, the values will not vary too much. We have already seen this behaviour on the right-hand side of Figure 17.2.

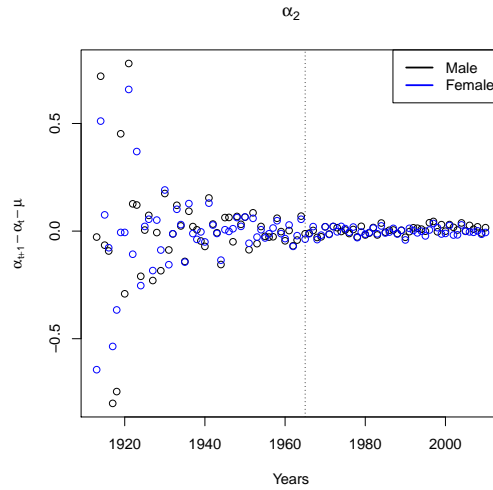


Figure 17.6:  $\hat{\alpha}_{t+1,2} - \hat{\alpha}_{t,2} - \hat{\mu}$

We are now able to calculate the forecasted mortality rates according to (9.19). We show the forecasted logit-rates we get for year 2060 on the left-hand side of Figure 17.7 and compare them to the ones we estimated in 2010. We find that the forecasted rates are always smaller than the estimated ones and that the two curves are similar to each other. What is untypical, though, is that the mortality rates for women are predicted to decrease more than the ones for men. This means, the difference between the two genders will increase during the next 50 years. Nonetheless, the forecasted rates seem to be realistic.

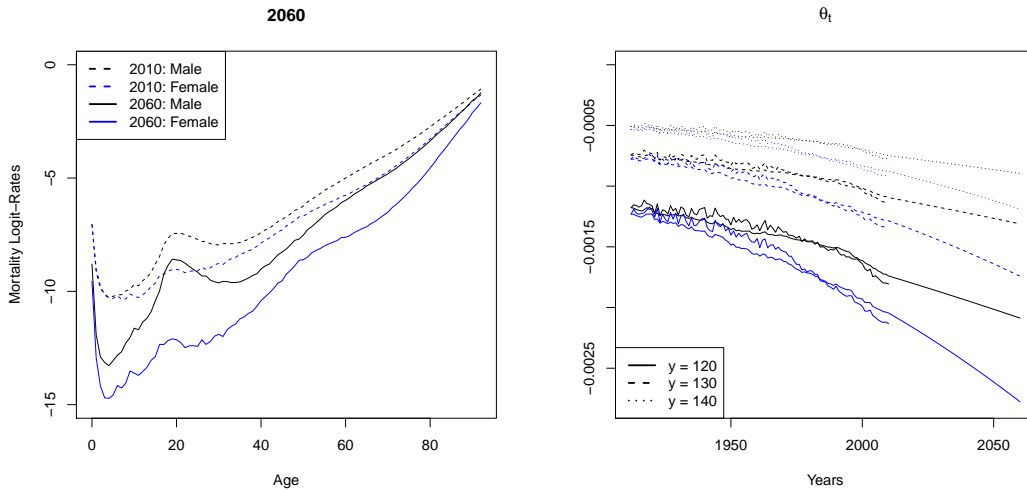


Figure 17.7: (LHS) Forecasted mortality logit-rates for year 2060,  
(RHS)  $\hat{\theta}_t$  from the extrapolation

To be able to calculate the three measures for longevity, we have to extend the estimated and forecasted mortality rates up to ages  $y = 120, 130$  and  $140$ . On the right-hand side of Figure 17.7, we see the estimators of  $(\theta_t)_t$  from the extrapolation procedure for the observed, estimated and forecasted mortality rates. We find that the observed and estimated parameters do not coincide perfectly but they are very close to each other. Furthermore, the trend in the future seems to be predicted in a reasonable way even though it might be too flat for men so longevity will perhaps be underestimated for them.

### Measures for Longevity

In Figure 17.8, we show the survival probability  $p_{0,t,j}$  calculated with the extended data for  $y = 120, 130$  and  $140$ . However, the choice of  $y$  influences the survival probability only for ages above 95. This is nice since it is after the age where the probability starts decreasing rapidly and this is the most important information we get from this plot. We see here that there clearly changed a lot in the years from 1912 to 2010. For example, child mortality nearly vanishes completely and the age where mortality starts increasing more quickly differs by about 20 years. This is the reason why the curves in 2010 are much steeper than 99 years earlier.

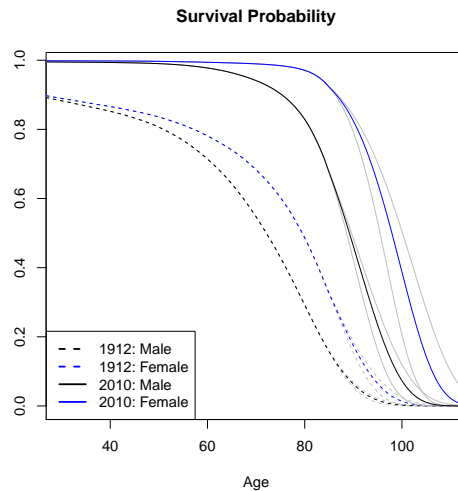


Figure 17.8: Survival probability  $p_{0,t,j}$  for  $t = 1912$  and  $2010$

Additionally to the survival probability, we calculate the life expectancies for people of age 30 and 65 and show them in Figure 17.9. We do this for all three extended datasets. The influence of the choice of  $y$  is negligible for men but is rather large for the females since we see here a clear difference between the coloured curves for  $y = 130$  and the grey ones for  $y = 120$  and  $140$  for high values of  $t$ . We observe that the curves are increasing but do not have a constant slope. In addition, we find that there are more years remaining to live for women than for men. This is also clear from Table 17.10 where we show some explicit values of  $e_{x,t}$  from the past but only for  $y = 130$ .

As a last step, we calculate the price of the annuity we introduced earlier. We insert  $r = 0\%, 1\%$  and  $5\%$  into (11.4) for the years from 1912 to 2060. By looking at Figure 17.11, we observe that the influence of the choice of  $y$  is small for the past but cannot be neglected

## 17 Application of Adapted Nolfi Models

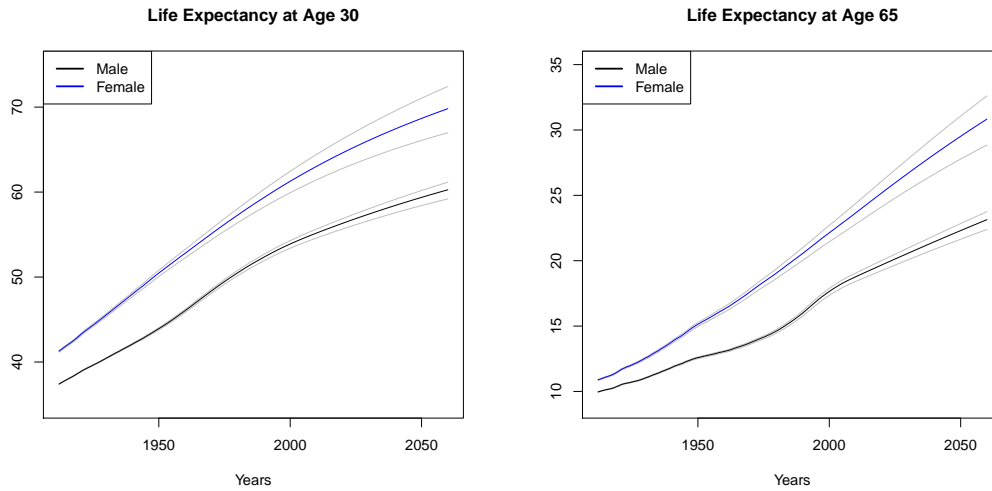


Figure 17.9: Life expectancy at age 30 and 65

		1912	1920	1930	1940	1950	1960	1970	1980	1990	2000	2010
30	Male	37.4	38.8	40.5	42.1	43.9	46.0	48.4	50.5	52.3	53.9	55.2
	Female	41.3	43.2	45.6	48.0	50.5	52.8	55.1	57.3	59.4	61.3	63.0
65	Male	10.0	10.4	11.0	11.8	12.6	13.1	13.7	14.6	16.0	17.6	18.8
	Female	10.9	11.6	12.6	13.8	15.1	16.3	17.6	19.1	20.6	22.1	23.6

Table 17.10: Life expectancy at age 30 and 65 for  $y = 130$

in the future. The prices lay on a rather nice, increasing curve with a small kink around 2000. The three values of  $r$  clearly separate the prices which means that the price decreases for larger interest rates. Thus, it is important to choose the interest rate correctly.

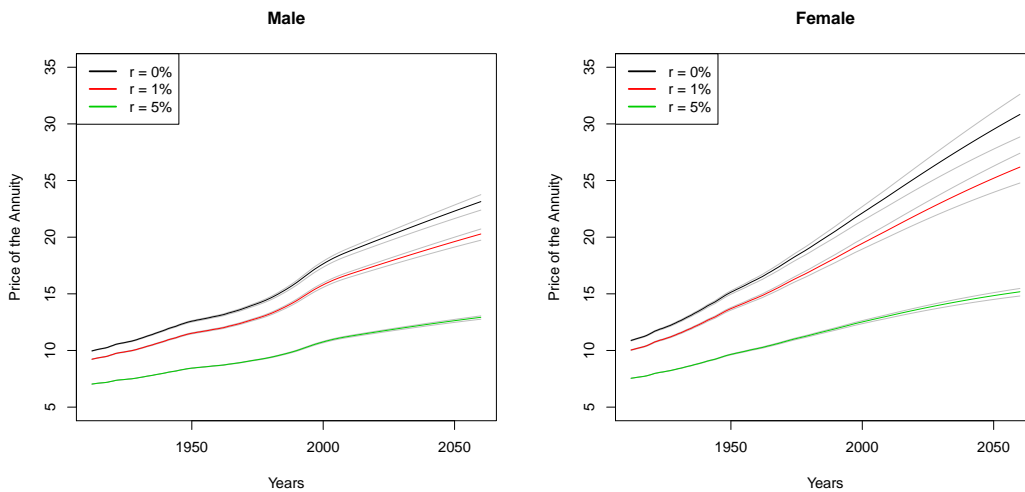


Figure 17.11: Price of the annuity for a person of age 65

We conclude that the model works nicely in the past and that we get realistic predictions for the future. Therefore, we go on and have a look at how this model behaves if we do not alter  $\lambda_x$  for different years but instead let  $q_{x,t_0}$  change over time.

## 17.2 Model 2

For Model 2, we estimate the mortality rates shown in Figure 17.12 by plugging the values  $(\hat{\alpha}_{t,1})_t$  into (17.2). We observe that the rates are rather nicely replicated by the model. We see some differences between the rates, however, but they seem to be small.

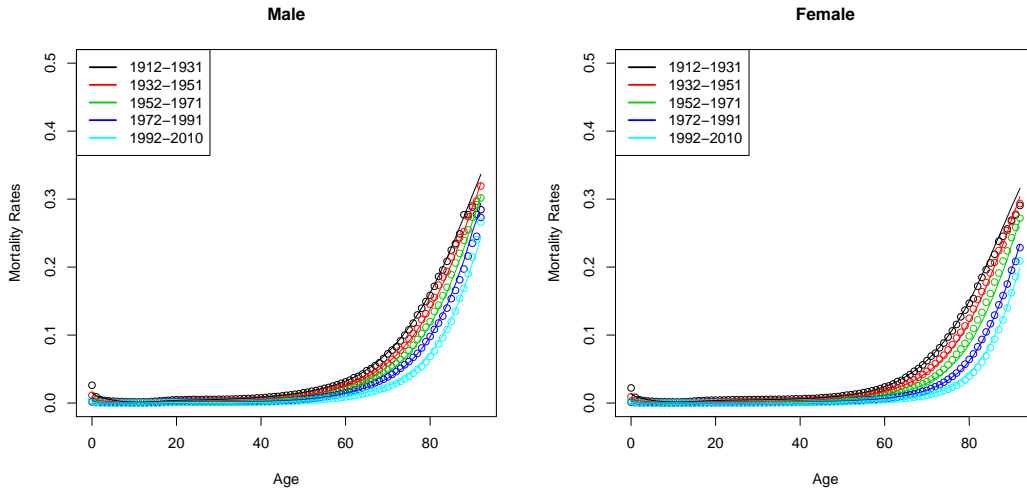


Figure 17.12: Observed (dots) and estimated (lines) mortality rates

To see how good the model performs for rates below 40, we logit-transform the mortality rates and have a look at the mortality logit-rates in Figure 17.13. The bumps for ages below 20 are captured very well with some exceptions in the earliest three year groups. For ages between 20 and 40, there is a clear differences between the observed and the estimated logit-rates for some groups. Nonetheless, the rates are nice since the overall shape of the curves are captured quite well.

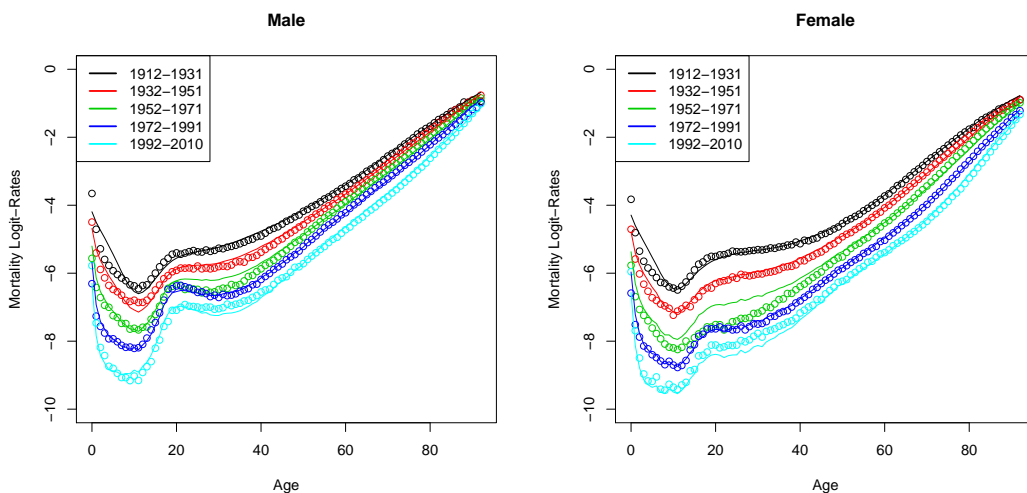


Figure 17.13: Observed (dots) and estimated (lines) mortality logit-rates

The residuals, shown in Figure 17.14, are close to zero for ages up to 80. Afterwards, they are farther away from 0 but generally stay between  $\pm 3\%$  with two exceptions in the earliest



## 17 Application of Adapted Nolfi Models

years for the male data. What is interesting, though, is that the residuals are not always positive or always negative because both signs occur for all year groups. Therefore, the expected value of the residuals is closer to zero but they are still correlated with age. From all these plots, we conclude that the model is appropriate for finding estimated mortality rates.

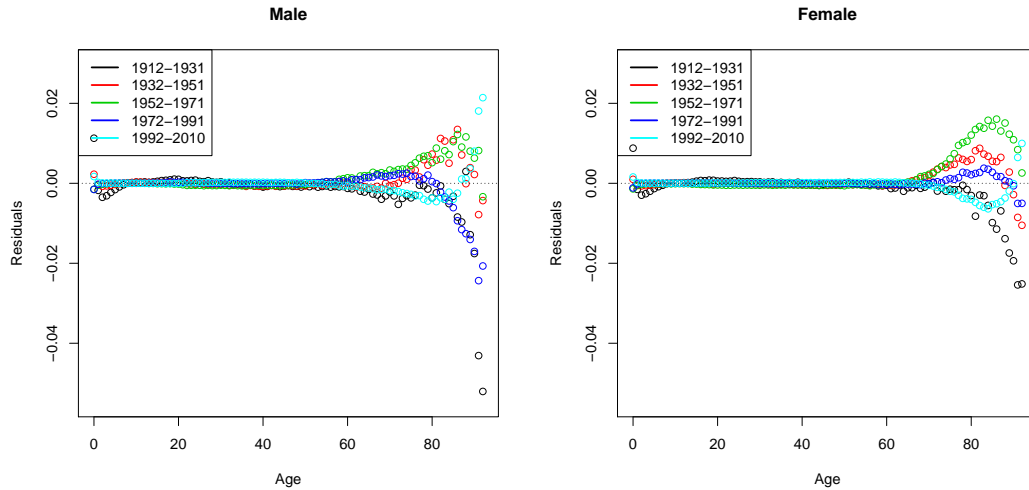


Figure 17.14: Residuals

### Forecasting and Extrapolating

The quality of a model also depends on its ability to forecast the mortality rates to the future. This is done as described in Section 9.4 by modeling  $\alpha_{t,1}$  as a one-dimensional random walk with drift. We only use the estimators from 1960 to 2010 to estimate the volatility to get a more realistic model, see Figure 17.15. For the drift, however, we use all available information. This yields the following models:

$$\text{Male: } \alpha_{t+1,1} = \alpha_{t,1} - 0.0021 + 0.0261 Z_t, \quad (17.6)$$

$$\text{Female: } \alpha_{t+1,1} = \alpha_{t,1} - 0.0032 + 0.0316 Z_t, \quad (17.7)$$

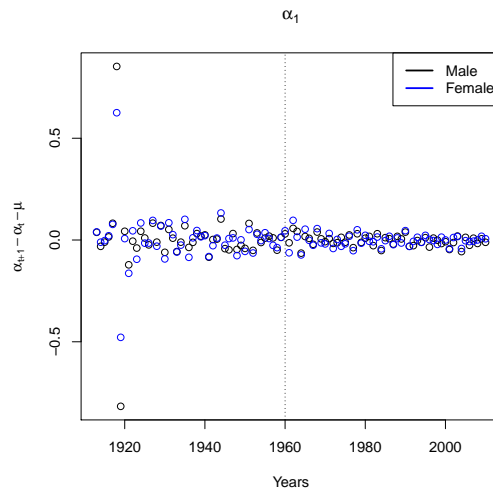


Figure 17.15:  $\hat{\alpha}_{t+1,1} - \hat{\alpha}_{t,1} - \hat{\mu}$

for *i.i.d.* standard normally distributed  $(Z_t)_t$ . We need to make sure that  $\alpha_{t,1}$  is always positive which does not hold true necessarily for all  $t$  since  $\hat{\mu} < 0$ . We find that  $\hat{\mu}$  is small which means that we expect  $(\alpha_{t,1})_t$  to decrease very slowly. The standard deviation is also rather low so we do not expect  $(\alpha_{t,1})_t$  to vary that much. We have already seen these characteristics on the left-hand side of Figure 17.2.

These two processes are now used to get forecasted mortality rates up to year 2060 by inserting the estimators  $\hat{\mu}$  and  $\hat{c}$  into (9.20). The results shown on the left-hand side of Figure 17.16 look similar to the ones from the last section. The rates are expected to be smaller 50 years from now but the mortality curve is slightly steeper for most ages. Therefore, we decide that it is reasonable to use these rates for all further calculations.

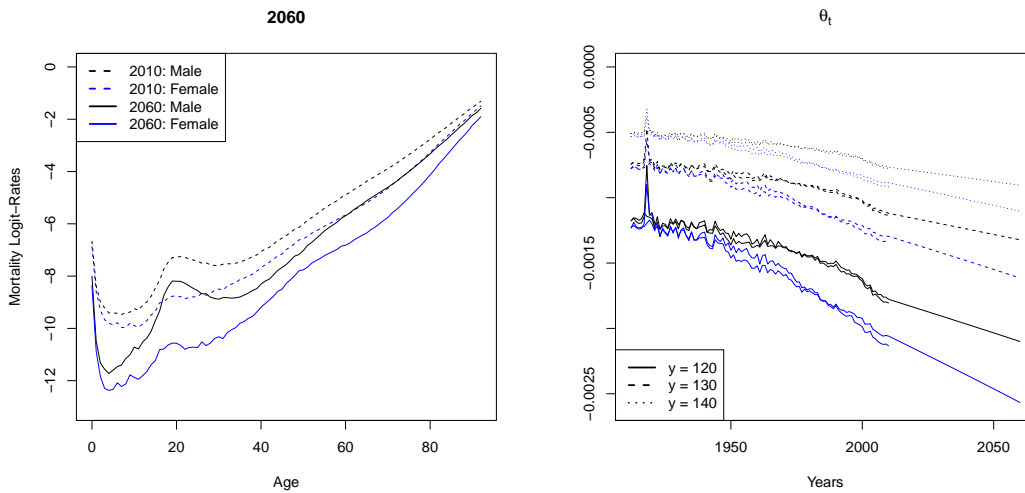


Figure 17.16: (LHS) Forecasted mortality logit-rates for year 2060, (RHS)  $\hat{\theta}_t$  from the extrapolation

The last step is to calculate the three measures for longevity we have already seen for Model 1. Thus, we need to extend the mortality rates to ages up to  $y = 120, 130$  and  $140$ . The resulting estimators of  $(\theta_t)_t$  are shown on the right-hand side of Figure 17.16 including the ones we get for the observed rates. We find that the parameters from the observed and estimated rates are rather close together. The forecasted ones capture the trend nicely even though it could even be slightly steeper, especially for the males. Nevertheless, we conclude that this model yields good mortality rate estimators in the past as well as in the future.

### Measures for Longevity

The extended rates for all three values of  $y$  are now used to calculate the survival probability  $p_{0,t,j}$  for both men and women separately. This is the probability that someone born in 1912 and 2010, respectively, achieves the age  $j$ . In Figure 17.17, we see that this probability is much higher in 2010 than it was in 1912 for all  $j$  which clearly shows that people live longer today than they did 99 years ago. Luckily, the influence of the choice of  $y$  is rather small because the coloured and grey curves are very close to each other until the age 95. Since we are mostly interested in persons with an age below 95, it makes nearly no difference how  $y$  is chosen.

## 17 Application of Adapted Nolfi Models

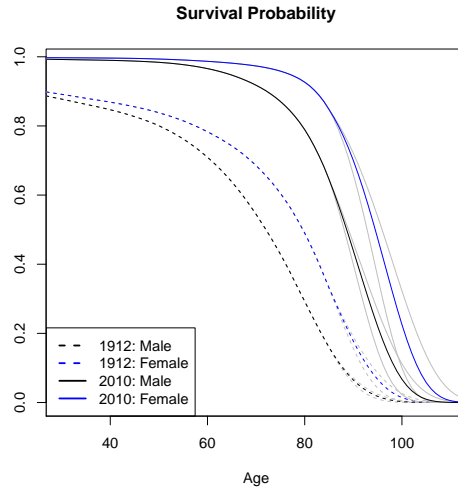


Figure 17.17: Survival probability  $p_{0,t,j}$  for  $t = 1912$  and 2010

Another indicator of the longevity of a society is the life expectancy at age 30 and 65. We show it in Figure 17.18 for all values of  $y$ . We see that  $y$  clearly influences the results since there is an obvious difference between the three curves for future years. Nevertheless, they all have a nice shape with the exception of a kink in 1918 which is more pronounced for  $x = 65$ . The life expectancy is increasing everywhere and the curves for the female data are always above the ones for the males. This is also clear from Table 17.19 where we show the values of  $e_{x,t}$  in some selected years in the past for  $y = 130$ .

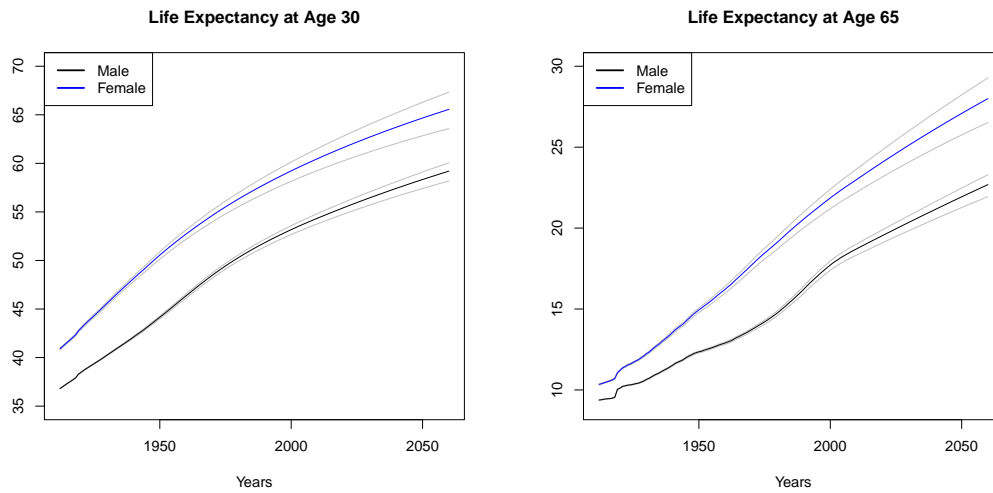


Figure 17.18: Life expectancy at age 30 and 65

Finally, we have a look at the price of the annuity where the interest rate is assumed to be constant equal to  $r = 0\%$ ,  $1\%$  and  $5\%$ , respectively. We again use all three extended datasets for its calculation. However, we see in Figure 17.20 that the choice of  $y$  only influences the price for small interest rates. For  $r = 5\%$ , the three curves are nearly identical in all years. All curves have a small kink in 1918 caused by the Spanish flu. For all other years, however, they are rather smooth. We observe that the price for women is higher than the one for men

## 17 Application of Adapted Nolfi Models

		1912	1920	1930	1940	1950	1960	1970	1980	1990	2000	2010
30	Male	36.8	38.5	40.3	42.1	44.2	46.4	48.5	50.3	51.8	53.2	54.4
	Female	40.9	43.0	45.6	48.1	50.6	52.7	54.6	56.3	57.9	59.2	60.5
65	Male	9.4	10.1	10.6	11.5	12.4	12.9	13.7	14.8	16.2	17.7	18.7
	Female	10.3	11.2	12.2	13.6	14.9	16.2	17.7	19.1	20.6	21.9	23.0

Table 17.19: Life expectancy at age 30 and 65 for  $y = 130$

for all values of  $r$ . Furthermore, we find that the interest rate has a huge impact on the price since the price curve is shifted downwards when it increases. We thus observe that the choice of  $r$  is more important than the one of  $y$ .

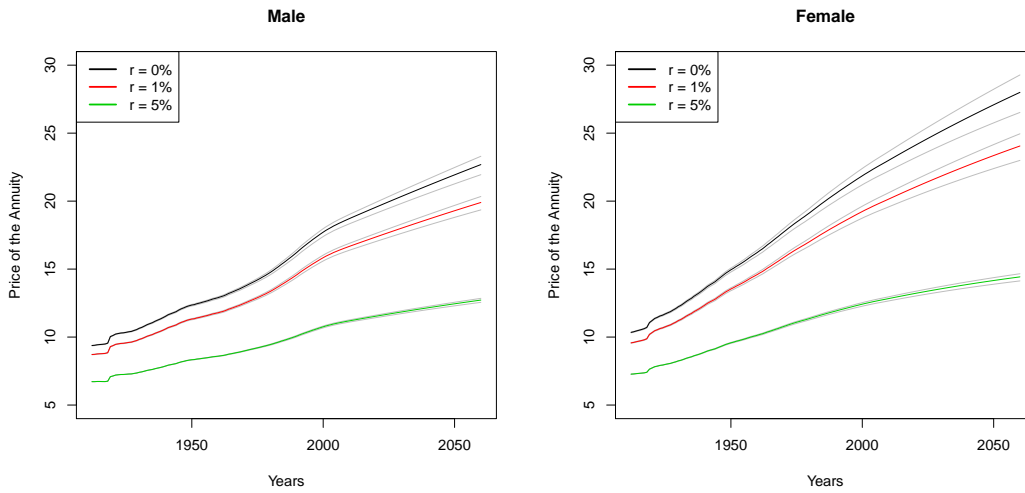


Figure 17.20: Price of the annuity for a person of age 65

In summary, we have found good estimators for the mortality rates in the past and have shown a reasonable way to forecast them to the future. Both procedures provided realistic results. If the last two models can further be improved when varying both the basic mortality rate as well as  $\lambda_x$  is discussed in the next section.

### 17.3 Model 3

Lastly, we repeat everything for Model 3. This now depends on both parameters  $\alpha_{t,1}$  and  $\alpha_{t,2}$  and therefore, we expect the results to look similar to the models above. The mortality rates that we get after the fitting procedure is done are shown in Figure 17.21. The rates form nice curves which are close to the observed ones. For ages above 90, however, we clearly see some differences but for the other ages, the fit is fine.

Next, we logit-transform the estimated mortality rates. In Figure 17.22, we observe that the model captures the fluctuations of the mortality rates nicely for ages below 20. Nevertheless, the estimated logit-rates are not always equal to the observed ones but the differences are very small. Thus, we find that the fit is also accurate for ages below 40.

## 17 Application of Adapted Nolfi Models

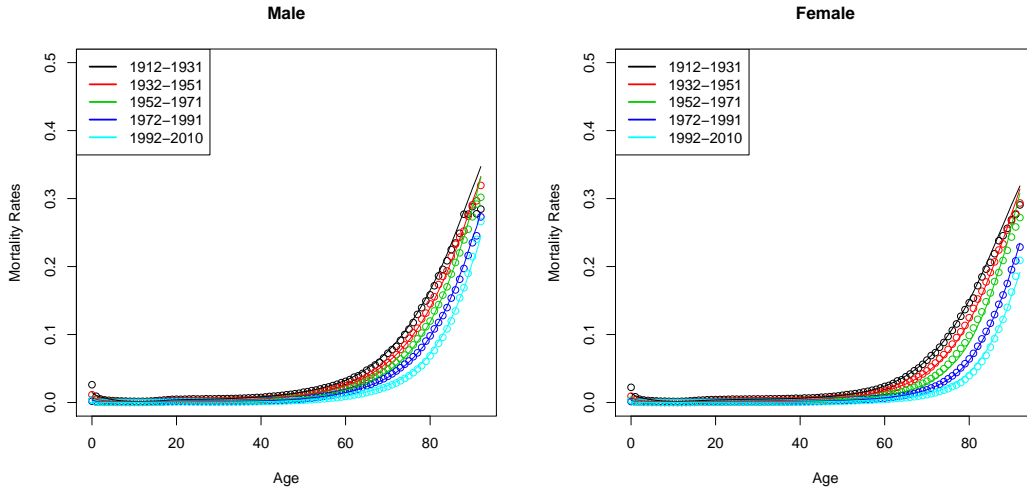


Figure 17.21: Observed (dots) and estimated (lines) mortality rates

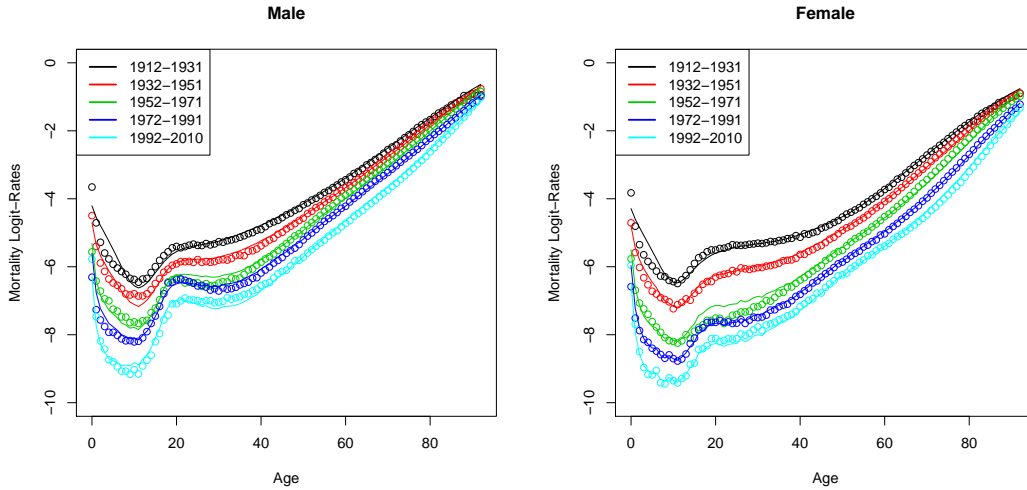


Figure 17.22: Observed (dots) and estimated (lines) mortality logit-rates

To have an additional measure for the quality of the estimation procedure, we plot the residuals in Figure 17.23. They are close to zero for ages up to 80 and they are mostly inside the range of  $\pm 2\%$ . They are again positive as well as negative for all year groups and correlated with age. Hence, this model produces a fit that is similar to the ones found with the two models above.

### Forecasting and Extrapolating

As a next step, we are interested in predicting the future. Hence, we forecast  $\alpha_t = (\alpha_{t,1}, \alpha_{t,2})^\top$ . This is a two-dimensional process so we model it with a two-dimensional random walk with drift. For the estimation of  $C$ , we only use the estimators from 1965 to 2010 since we observe in Figure 17.24 that the variability of  $\alpha_t$  is untypically large for the earlier years. We use the years after 1919 for the estimation of the drift of  $\alpha_{t,2}$ , because the earliest years falsify the drift, whereas we use all years for the one of  $\alpha_{t,1}$ . We then get

## 17 Application of Adapted Nolfi Models

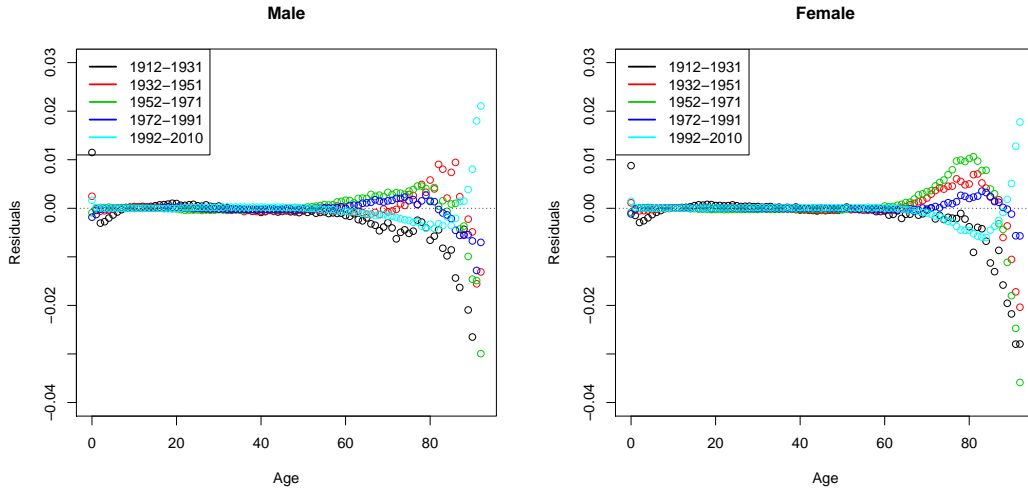


Figure 17.23: Residuals

$$\text{Male: } \boldsymbol{\alpha}_{t+1} = \boldsymbol{\alpha}_t + \begin{pmatrix} -0.0029 \\ -0.0019 \end{pmatrix} + \begin{pmatrix} 0.0515 & 0.0510 \\ 0 & 0.0221 \end{pmatrix} \mathbf{Z}_t, \quad (17.8)$$

$$\text{Female: } \boldsymbol{\alpha}_{t+1} = \boldsymbol{\alpha}_t + \begin{pmatrix} -0.0040 \\ -0.0009 \end{pmatrix} + \begin{pmatrix} 0.0485 & 0.0317 \\ 0 & 0.0150 \end{pmatrix} \mathbf{Z}_t, \quad (17.9)$$

where  $(\mathbf{Z}_t)_t$  are *i.i.d.* with a bivariate normal distribution. We always need to check if both components are positive because we find that  $\hat{\boldsymbol{\mu}} < 0$  which contradicts the results from Model 1. Since  $\hat{\boldsymbol{\mu}}$  is rather small, both parameters are slowly decreasing. Furthermore, the standard deviation of both of them is large compared to their mean. All these characteristics were already visible in Figure 17.2.

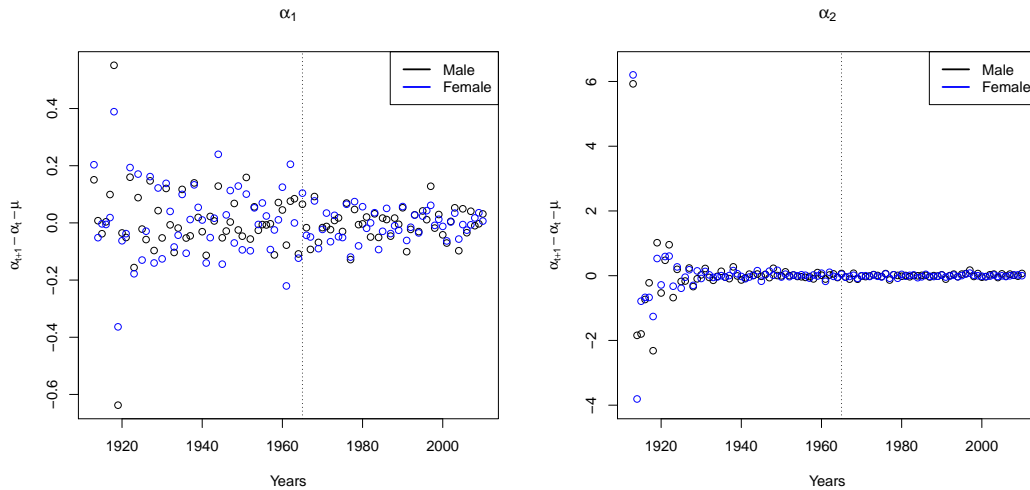


Figure 17.24: (LHS)  $\hat{\alpha}_{t+1,1} - \hat{\alpha}_{t,1} - \hat{\mu}_1$ , (RHS)  $\hat{\alpha}_{t+1,2} - \hat{\alpha}_{t,2} - \hat{\mu}_2$

Unfortunately, we are only able to forecast the mortality rates to year 2087 since for  $t > 77$ ,  $(\hat{\alpha}_{2010,1} + t \hat{\mu}_1 + t \hat{\lambda}_x(t + 98) \hat{C}_{12} \hat{C}_{22})$  in (9.22) becomes negative for the male data. This is caused by the fact that  $(\hat{\alpha}_{t,1})_t$  and  $(\hat{\alpha}_{t,2})_t$  are positively correlated. The mortality logit-rates

for year 2060 are shown on the left-hand side of Figure 17.25. We see that they are smaller than in 2010 for both genders. The curves look similar in both years but are shifted to the bottom. Thus, we still conclude that the forecasting procedure works even though the results are not reasonable for all years.

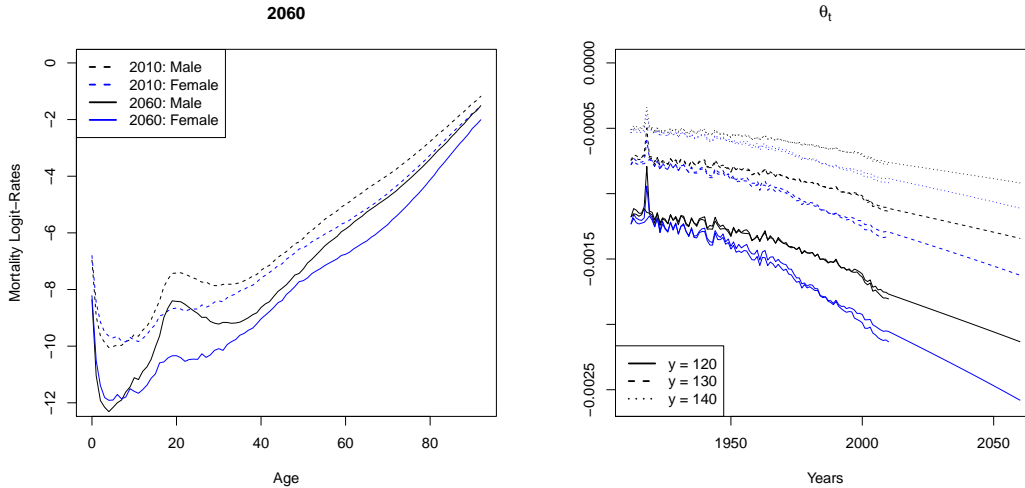


Figure 17.25: (LHS) Forecasted mortality logit-rates for year 2060, (RHS)  $\hat{\theta}_t$  from the extrapolation

We can only extend the mortality rates to ages  $y = 120, 130$  and  $140$  until year 2087. The estimators of  $(\theta_t)_t$  for  $t$  from 1912 to 2060 are shown on the right-hand side of Figure 17.25 for the observed, estimated and forecasted mortality rates. The parameters for the observed and estimated rates are rather close together and the trend for the future looks quite well. This means, the forecasted rates for the 50 years we show here seem to be realistic.

### Measures for Longevity

Since we cannot forecast the mortality rates to years after 2087, we cannot calculate the survival probability  $p_{0,t,j}$  in 2010. Hence, we only show it for 1912 and 1947 in Figure 17.26 for all three values of  $y$ . Fortunately, the choice of  $y$  is unimportant for this plot since it does not influence the age where there is a kink in the probability curve. We observe that the kink shifted in the first 35 years by about 10 years to the right. This is the reason why the curves in 2047 are steeper than the ones in 1912.

We can also only calculate the life expectancy at age 30 and 65 for a limited number of years because we do not have that many forecasted mortality rates available. We plot the remaining values in Figure 17.27. What stands out here is the small kink in 1918 which is caused by the Spanish flu. We have calculated  $e_{x,t}$  with all three extended datasets. The grey curves in the figure correspond to  $y = 120$  and  $140$  whereas the coloured ones are for  $y = 130$ . We observe that we nearly do not see any difference between the three values of  $y$  for all 75 years. The life expectancy is increasing over time and the slope for women is a bit larger than the one for men. Additionally,  $e_{x,t}$  is always larger for the women. We see this again in Table 17.28 where we show some explicit values of the life expectancy in the past

## 17 Application of Adapted Nolfi Models

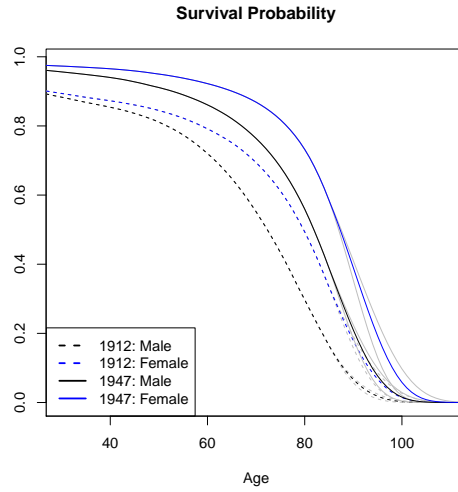


Figure 17.26: Survival probability  $p_{0,t,j}$  for  $t = 1912$  and 1947

but only for the extended rates with  $y = 130$ . The table shows more clearly how much  $e_{x,t}$  increased in the 75 years from 1912 to 2087.



Figure 17.27: Life expectancy at age 30 and 65

		1912	1920	1930	1940	1950	1960	1970	1980	1987
30	Male	36.8	38.5	40.4	42.4	44.3	46.3	48.5	50.6	51.9
	Female	40.9	43.2	45.9	48.4	50.7	52.7	54.5	56.2	57.3
65	Male	9.3	10.1	10.6	11.4	12.3	12.9	13.8	14.7	15.6
	Female	10.3	11.2	12.2	13.4	15.0	16.4	17.8	19.1	20.0

Table 17.28: Life expectancy at age 30 and 65 for  $y = 130$

Lastly, we calculate the price of the annuity with an interest rate of  $r = 0\%$ ,  $1\%$  and  $5\%$ , respectively. We see in Figure 17.29 that the choice of  $y$  does not influence the price in the past that much since the three curves are very close together. What stands out, though, is



## 17 Application of Adapted Nolfi Models

the kink in 1918 which was already visible on the right-hand side of Figure 17.27. All three curves are generally increasing and shifted to the top as  $r$  decreases. A decreasing interest rate also raises the steepness of the price curve.

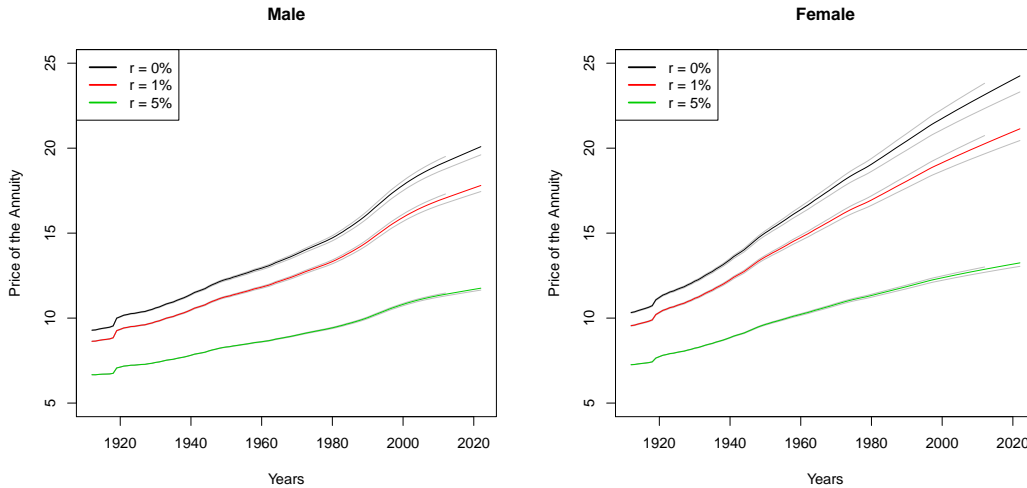


Figure 17.29: Price of the annuity for a person of age 65

To conclude, we say that the estimated mortality rates in the past are rather convincing and easily achieved with all three models. The residuals are small until age 80 and do not increase too much afterwards. Unfortunately, the forecasting procedure of Model 3 did not yield reasonable results. This was unexpected since it worked fine for the other two. This might be caused by the negative slope of  $\hat{\alpha}_{t,2}$  or it may indicate that the model is over-parametrized, meaning it has too many degrees of freedom. Nevertheless, the forecasting procedure itself is still a good choice. With this, we end the last application and start comparing the models to each other to find the best one in the end.

# 18 Comparison of Mortality Rate Models

The aim of this chapter is to find a model which fits the data of the Swiss population from 1912 to 2010 best. We compare all models we have discussed on the basis of different characteristics and the model that is preferred will then depend on what one is looking for. However, we will not include the extended model of Lee and Carter in this analysis since it was clear from the beginning that the choice of  $d_w$  was not realistic.

We will analyse the following characteristics. We start with discussing restrictions with respect to age which sometimes imply a loss of data. Next, we examine the estimation quality by looking at the residuals. Then, we calculate the forecasted mortality logit-rates for year 2060 as well as the estimator of  $(\theta_t)_t$  from the extrapolation procedure to see if the forecasted mortality rates are realistic. Finally, we show the three measures for longevity and compare them, if possible, to the ones given by the Swiss Federal Statistical Office (2009b). In the end, we draw a conclusion and construct a mortality table with our preferred model. All abbreviations which are used in this chapter are given in Table 18.1.

---

<b>SFS</b>	Swiss Federal Statistical Office (2009b)
<b>OBS</b>	Observed mortality rates
<b>GNO</b>	Generalized Nolfi model
<b>MM</b>	Marginal mortality rate model
<b>EX</b>	Extrapolation model
<b>GLM</b>	Generalized linear model
<b>LC</b>	Classical model of Lee and Carter
<b>CBD</b>	Cairns-Blake-Dowd model
<b>NO1</b>	First adapted Nolfi model
<b>NO2</b>	Second adapted Nolfi model
<b>NO3</b>	Third adapted Nolfi model

---

Table 18.1: Abbreviations

## 18.1 Age Restrictions

We start this analysis with discussing which models have any age restrictions that correspond to a potential loss of information. There are two things we need to consider. It is either possible that we cannot use the data of the smallest ages or it could happen that we need to omit data for ages above some high threshold.

We have seen that the Cairns-Blake-Dowd model is only sensible for high ages. Therefore, we could not apply it to the data of people with an age up to 29 although the lower boundary could be even chosen to be higher. This is not an issue if we are only interested in people of age 30 and higher anyway. However, if we want the mortality rates for children and young adults as well, this model is not appropriate since more factors would be needed to account for them, too. All other models, though, are applicable to small ages without causing any problems. The other issue is that we might not be able to apply a model to high ages. We have seen this for the classical model of Lee and Carter, where we need  $b_x > 0$ , and for the

common mortality rate models and the adapted Nolfi models, where we have the condition that  $\lambda_x > 0$ . Both requirements were violated in some cases but we maximally lost 4 age groups because of this. Since the extended data are used for most calculations, the original rates above age 85 are left out anyway. Hence, these restrictions are not too important even though they require additional work.

If someone wants a model with no restrictions with respect to age, the common mortality rate models as well as the generalized linear model are the best choices. All other models have some restrictions which fortunately did not cause too many troubles in our analysis. How appropriate the estimators for the remaining age groups are is the topic we are going to analyse next.

## 18.2 Quality of the Estimators

We are looking for a model which replicates the past as closely as possible. The best indicator for the quality of the estimated mortality rates are the residuals. We calculate them for ages between 0 and 91 and then choose two year groups to see if the estimation quality depends on time. The first group goes from 1912 to 1961 and the second one from 1962 to 2010. The average residuals for these two groups are shown in Figures 18.2.1 and 18.2.2 but only those between  $\pm 4\%$  since otherwise, they are too large anyway.

When comparing the residuals for these two year groups, we actually see that to a certain extent they do depend on the year of measurement. The residuals for the second group are mostly closer to zero so all models are more appropriate for the last 50 years. We observe that until an age of 60, all models have residuals which are very close to zero with some exceptions for very small ages. This means that until then, they do not differ very much from each other. Hence, the performance of a model only depends on the quality of its estimators above this threshold. The first thing we recognize here is that the generalized linear model performs worst because its residuals are highly negative in all four plots. In the first 50 years, they even go down to more than -10%. What we can see as well is that the extrapolation model, the classical model of Lee and Carter and the second adapted Nolfi model have residuals which are closest to zero for both genders. The third adapted Nolfi model also works rather nicely even though its residuals are farther away from zero. The remaining models are harder to compare. In the first year group, we see that the marginal mortality rate model as well as the generalized and the first adapted Nolfi model perform more or less equally well. Their residuals are mostly positive and smaller than +3%. Furthermore, we observe in the second year group that the Cairns-Blake-Dowd model has larger residuals than the others.

By only looking at the residuals, we find that the extrapolation model, the classical model of Lee and Carter and the second adapted Nolfi model are the most preferable models since their residuals are closest to zero. They are followed by the third adapted Nolfi model which still performs nicely. The remaining models need to be treated more carefully. Their estimated mortality rates are sometimes too far away from the truly observed rates. However, one often wants a model that can also be used to make realistic predictions for the future. Whether these three models are the best choice for this task as well is shown in the following section.

## 18 Comparison of Mortality Rate Models

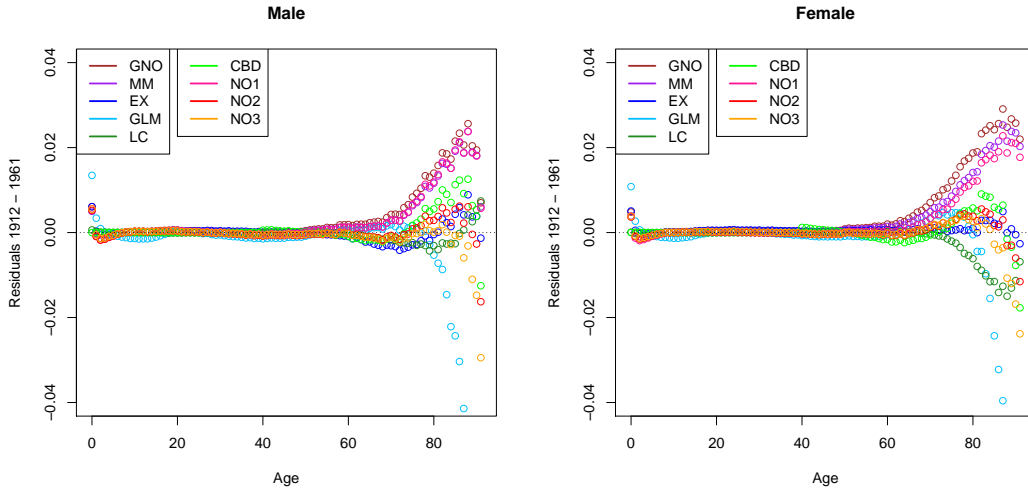


Figure 18.2.1: Residuals: 1912 - 1961

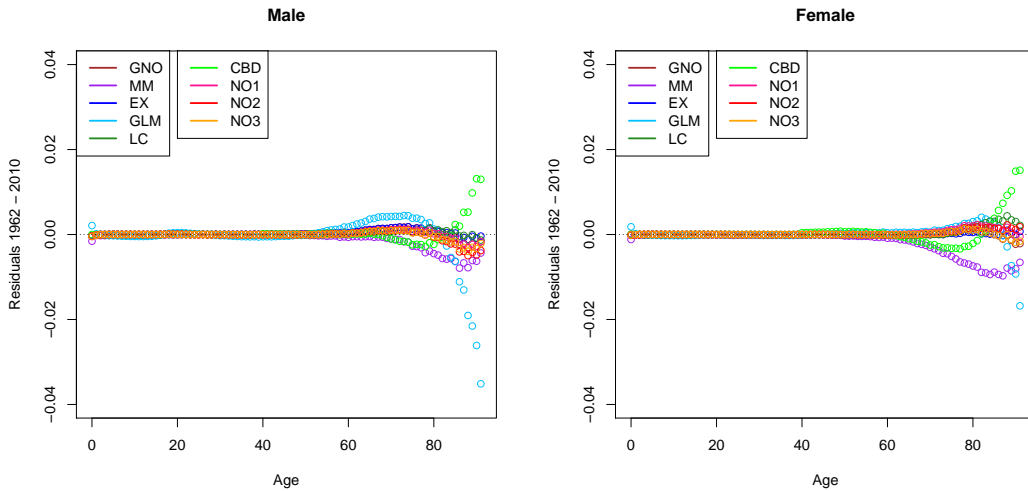


Figure 18.2.2: Residuals: 1962 - 2010

### 18.3 Quality of the Forecasts

As a next step, we compare the quality of the forecasts to each other. However, it is difficult to do so since we do not know the true mortality rates for future years. Nonetheless, we analyse the following two quantities. First, we calculate the mortality logit-rates in 2060 and discuss how the predictions differ from each other for ages above 40. Secondly, we have a look at the slope of  $(\hat{\theta}_t)_t$  from the extrapolation procedure to see if its trend in the future is realistic. We extend the mortality rates until the maximal age  $y = 130$ . We have explained in Chapter 9 how to find forecasted mortality rates for all models. Unfortunately, we could not calculate that many forecasts with the third adapted Nolfi model. However, this does not necessarily mean that it will not work for any dataset even though it is possible that it will again cause some problems. We still include it in this analysis but the conclusions need to be treated more carefully.

## 18 Comparison of Mortality Rate Models

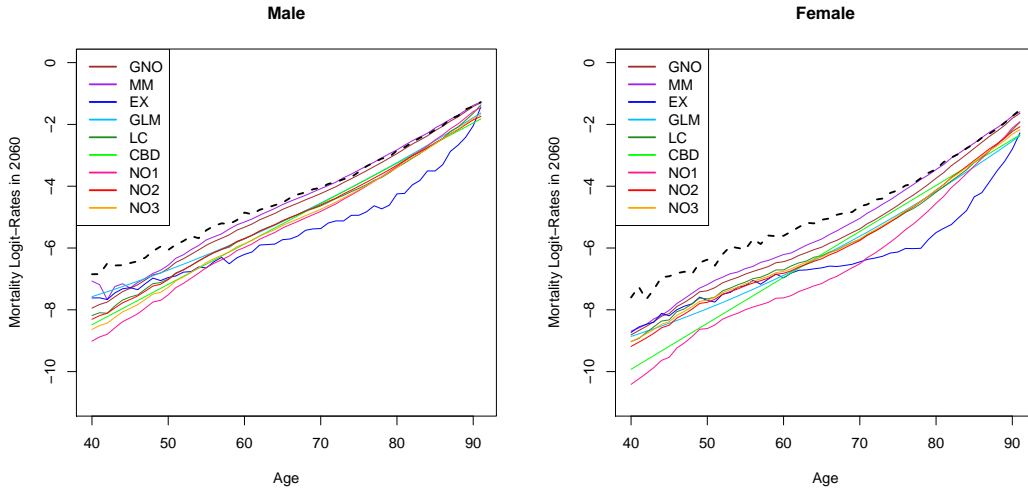


Figure 18.3: Observed mortality logit-rates in 2010 (dashed) and forecasted logit-rates for year 2060 (solid)

We have seen that the mortality rates decreased in a rather constant manner over the last few years for a fixed age. Therefore, the shape of the logit-rate curve should look similar in 2060 as it did in 2010 but the rates should be lower everywhere. Additionally, an insurance company with many annuities is generally more interested in a slightly conservative model since this makes it less problematic when the true mortality rates are lower than expected. After taking this into account, we find that the generalized Nolfi model and the marginal mortality rate model do not yield good forecasts since we see in Figure 18.3 that the curves for these two models are too close to the ones observed in 2010. All remaining models, though, yield forecasts which might be true. What we also observe is that the most conservative ones are the extrapolation model and the first adapted Nolfi model since they forecast the lowest mortality logit-rates for ages above 50.

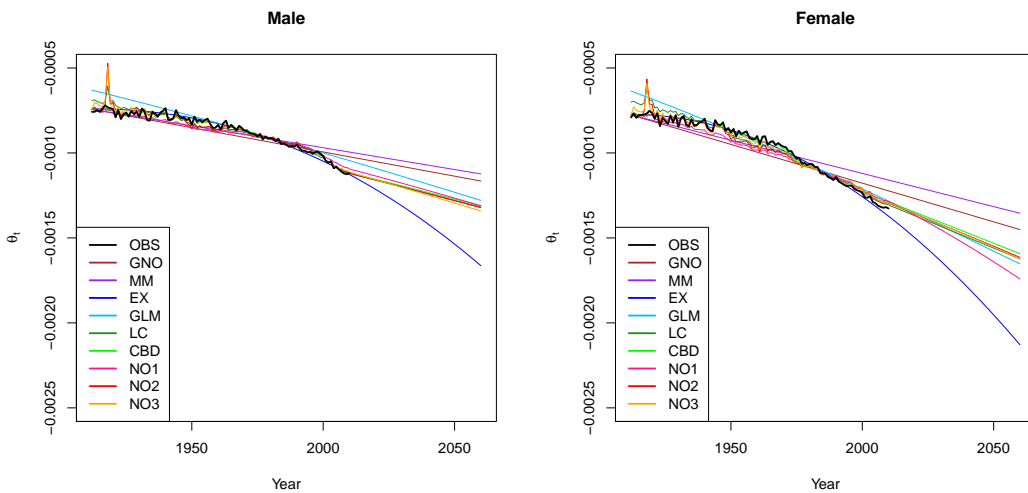


Figure 18.4:  $\hat{\theta}_t$  from the extrapolation with  $y = 130$

In the process of extending the mortality rates to higher ages, we have to estimate the parameter  $\theta_t$  in all years. We show its estimators obtained with the models as well as with the observed rates in Figure 18.4. A realistic model should accurately replicate the observed slope and nicely forecast it to the future. We again find that the generalized Nolfi model and the marginal mortality rate model yield slopes which are farthest away from the truly observed one. All remaining predictions are in a realistic range and cannot be eliminated. We observe that their slopes are rather close together aside from the extrapolation model so they are likely to predict a similar increase in longevity.

If someone wants realistic and not too aggressive forecasted mortality rates, we recommend that he should eliminate the generalized Nolfi model and the marginal mortality rate model because their forecasted rates are probably too high. Additionally, he needs to be careful with the third adapted Nolfi model since it did not work well in our analysis. All other models can reasonably be chosen for this task. However, we need to keep in mind that we only compared them on the basis of their 50-year forecast as well as the estimation of  $(\theta_t)_t$ . How good and realistic the forecasted mortality rates really are will be better visible in the next section where we calculate the three measures for longevity.

### 18.4 Measures for Longevity

From now on, we always work with the extrapolated mortality rates until the maximal age  $y = 130$ . However, due to the previous three sections, we will not analyse the generalized Nolfi model, the marginal mortality rate model, the generalized linear model, the Cairns-Blake-Dowd model and the third adapted Nolfi model any further since their performances were not good enough from our point of view.

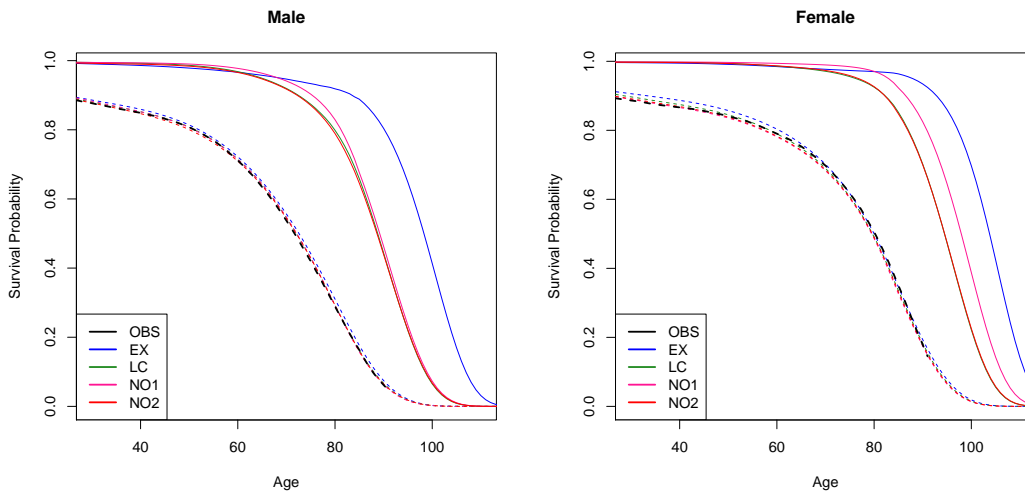


Figure 18.5: Survival probability  $p_{0,t,j}$  for  $t = 1912$  (dashed) and 2010 (solid)

The first measure we show here is the survival probability  $p_{0,t,j}$  given by (11.1) for  $t = 1912$  and 2010. In 1912, all models are rather close together even though we see some small differences to the probabilities we get for the observed mortality rates. The divergence between the models increases heavily in the following 99 years which clearly affects the kink in this

## 18 Comparison of Mortality Rate Models

curve. Hence, it is hard to say how large the survival probability will most likely be in 2010 because the predicted values of the models differ that much from each other. The difference is even larger for women. However, the figure signifies that the extrapolation model probably suggests too high survival rates.

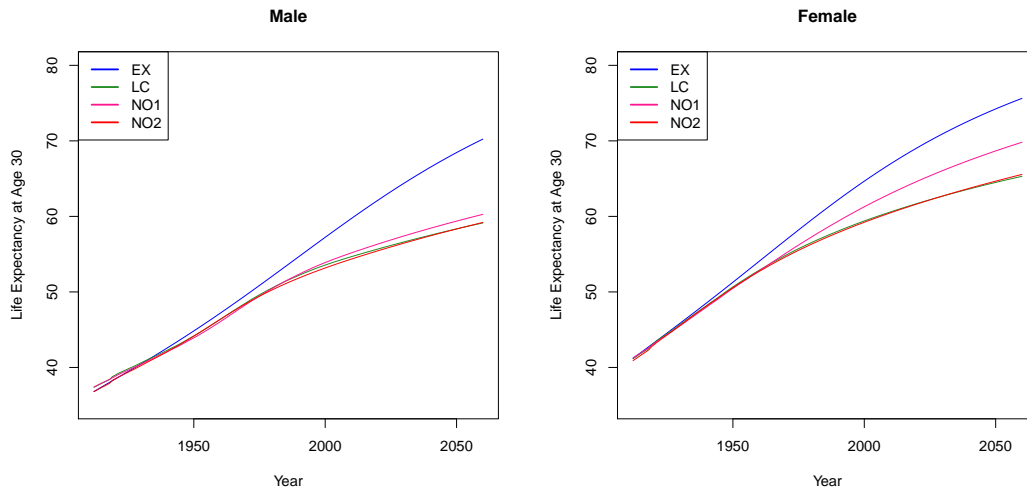


Figure 18.6.1: Life expectancy at age 30

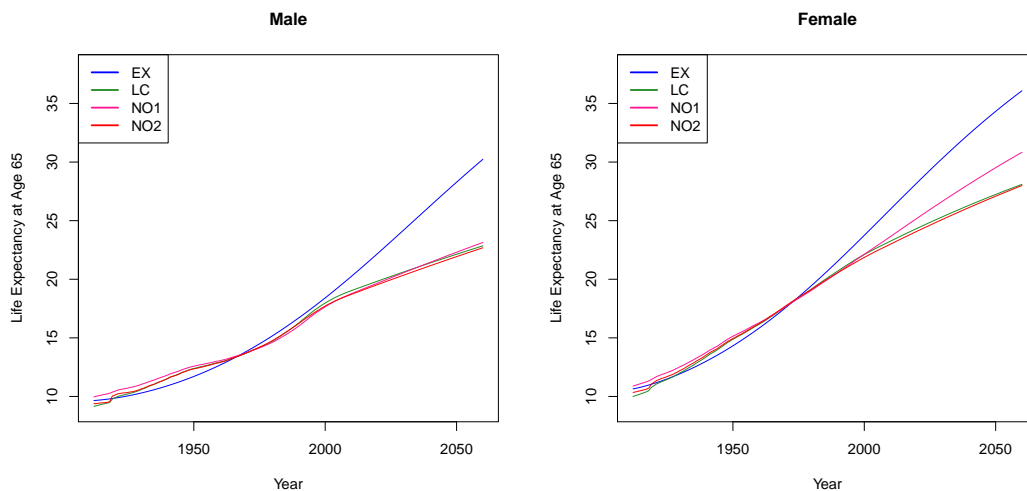


Figure 18.6.2: Life expectancy at age 65

The next measure we are discussing in Figures 18.6.1 and 18.6.2 is the life expectancy  $e_{x,t}$ , see (11.3), at age 30 and 65 from 1912 until 2060. We see that the values until about 1950 are rather close together. Afterwards, we observe that the forecasted life expectancies vary highly between the models. The model which clearly stands out is the extrapolation model since it predicts much higher life expectancies than any other model. This again suggests that it might be too conservative. The life expectancies for the remaining three models are in a rather close range until 2060.

## 18 Comparison of Mortality Rate Models

		SFS	EX	LC	NO1	NO2
30	Male	56.6	59.7	54.7	55.2	54.4
	Female	60.5	67.0	60.6	63.0	60.5
65	Male	20.5	20.3	19.0	18.8	18.7
	Female	24.0	26.0	23.3	23.6	23.0

Table 18.7: Life expectancy at age 30 and 65 in 2010

To see it even better that the model really does have an impact on the value of the life expectancy, we show  $e_{30,2010}$  and  $e_{65,2010}$  for all models in Table 18.7 and compare them to the ones given by the Swiss Federal Statistical Office (2009b). We find that the life expectancies given by the Swiss Federal Statistical Office are very close to the ones calculated with the models. Here, the range of possible values for the life expectancy of a person of age 30 is given by 54.4 to 59.7 years for men and 60.5 to 67 years for women. At age 65, the spread is a bit smaller. For men, it is 18.7 to 20.3 years and for women, we get 23 to 26 years. Hence, the maximal difference between the life expectancies is 6.5 years for the 30 year old women which is quite a lot. The model which estimates the largest life expectancies is the extrapolation model.

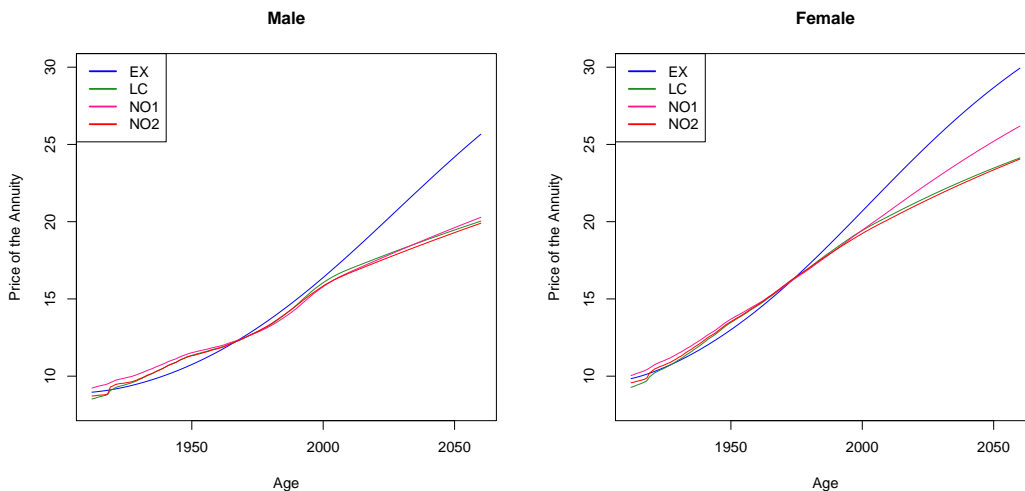


Figure 18.8: Price of the annuity for a person of age 65 when  $r = 1\%$

We lastly compare the price of the annuity defined in (11.4) that a person of age 65 has to pay when we assume that the interest rate is constant over time and equal to 1%. The results in Figure 18.8 look similar to the ones in Figure 18.6.2. Until 2000, we observe that the models yield prices which are rather close together. For the years thereafter, though, we find that the range of possible outcomes is rather large. What we can say, however, is that the extrapolation model predicts the highest prices whereas the forecasts of the other models are very similar.

We show the price of the annuity in Table 18.9 for year 2010 to compare them to the life expectancies from Table 18.7. We recognize that the prices calculated with these four models lay inside the range from 16.7 to 17.9 monetary units for men against one from 20.1 to 22.4



## 18 Comparison of Mortality Rate Models

	EX	LC	NO1	NO2
<i>Male</i>	17.9	16.9	16.7	16.7
<i>Female</i>	22.4	20.3	20.7	20.1

Table 18.9: Price of the annuity for a person of age 65 in 2010 when  $r = 1\%$

monetary units for women. Thus, the spread for women is 2.3 units which is not too much in comparison to the price itself. It is again the extrapolation model which suggests the highest prices.

It is hard to find one model which is a best choice for all three measures we have discussed in this section because the results of all but the extrapolation model are very similar. However, we have seen that the extrapolation model is most conservative and that its predicted mortality rates might be too low even though they seemed reasonable at first.

### 18.5 Construction of a Mortality Table

Due to the last section, we are left with three models for a final discussion. We find that the classical model of Lee and Carter and the first and second adapted Nolfi models yield nearly the same results in the past. In the future, there are some differences between them but their predictions are generally still close. Note that in Section 7.1 we have seen that the first adapted Nolfi model is a special case of the classical model of Lee and Carter which is why their results are so similar.

After all these considerations, we come to the final conclusion that the second adapted Nolfi model has the best performance for all the characteristics we have discussed in this chapter since it fits the data in the past slightly better than the two alternatives when only considering the residuals. Therefore, it is the model we prefer. However, it is possible that its forecasted mortality rates might be slightly too high. Perhaps a mixture between its predictions and the ones from the extrapolation model or the first adapted Nolfi model would be more accurate.

With this, we can finally construct the mortality table we were looking for in Table 18.11 with the second adapted Nolfi model up to an age of 107. Since we would like to compare the estimated mortality rates to the ones given by the Swiss Federal Statistical Office (2013b), we only show the average rates from 1998 to 2003. We additionally calculate the estimated mortality rates of order 1 for one period of time for the whole portfolio. We fix the confidence level  $\alpha = 0.99$  and calculate  $q_{x,t}^1$  for all  $x$  and  $t$  according to (10.8). Since we need the observed number of deaths for these calculations, we can only do it for ages up to 95.

Before we actually show the mortality table, we recall in Figures 18.10.1 to 18.10.6 the most important figures we have found for the second adapted Nolfi model and compare them, if possible, to the results we get with the observed mortality rates. We additionally have a closer look at the life expectancy in 2010. We calculate  $e_{x,2010}$  for  $x$  from 30 to 100 and with it the expected time of death for the same ages. We now have shown how to explicitly construct a mortality table and how one can decide which model should be chosen for the construction. Since this was the aim of the thesis, our task is completed and we can finish the analysis.

## 18 Comparison of Mortality Rate Models

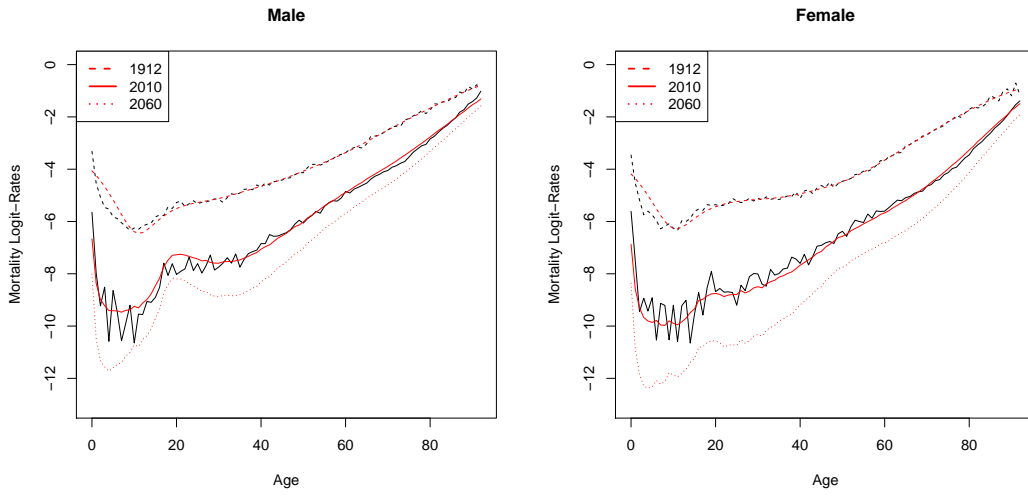


Figure 18.10.1: Observed (black) and estimated (red) mortality logit-rates

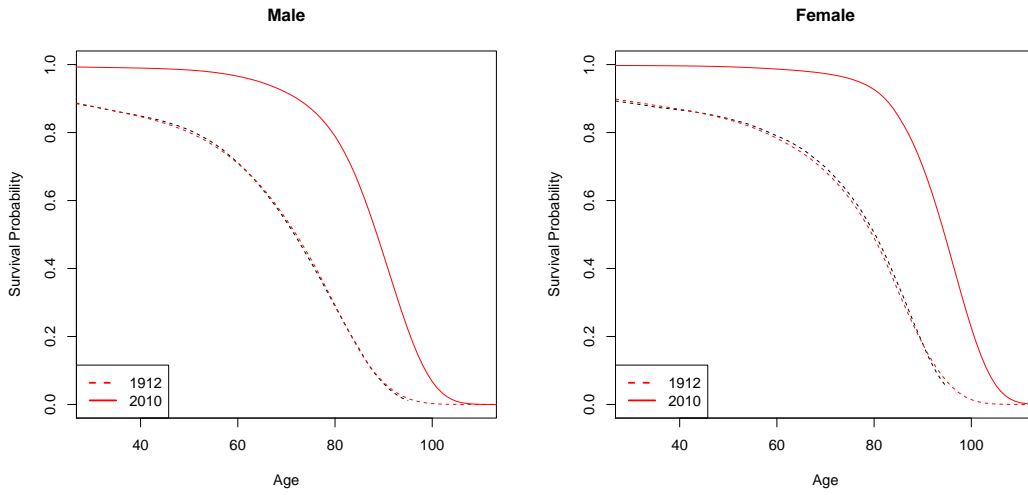


Figure 18.10.2: Observed (black) and estimated (red) survival probability  $p_{0,t,j}$

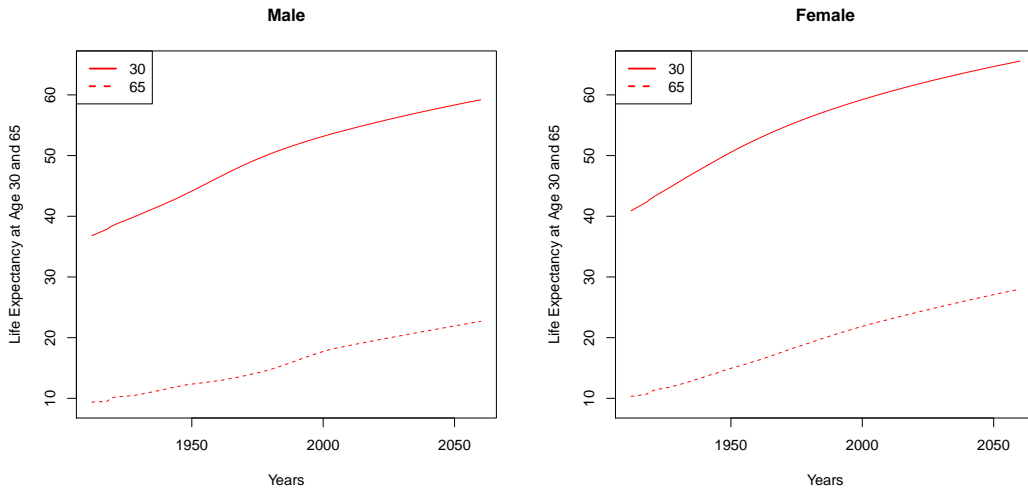


Figure 18.10.3: Life expectancy at age 30 and 65

## 18 Comparison of Mortality Rate Models

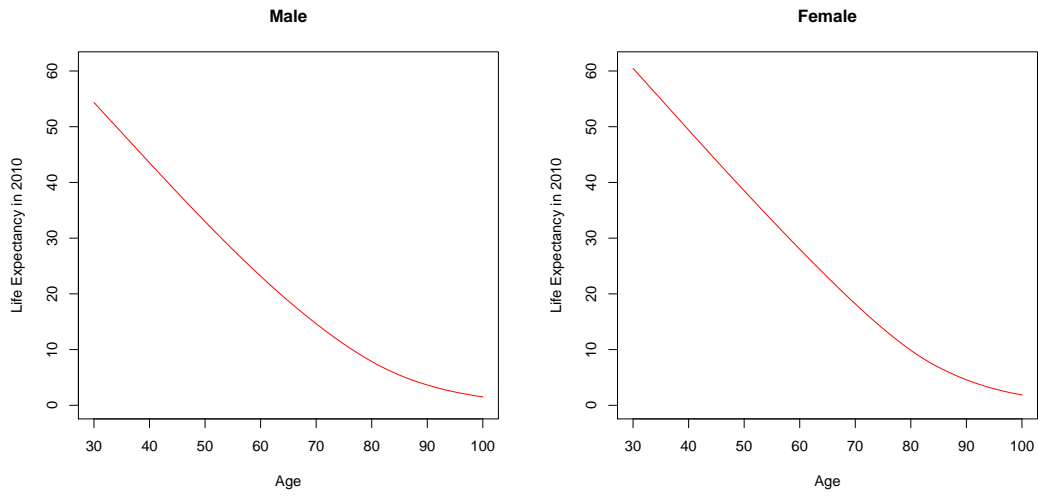


Figure 18.10.4: Life expectancy in 2010

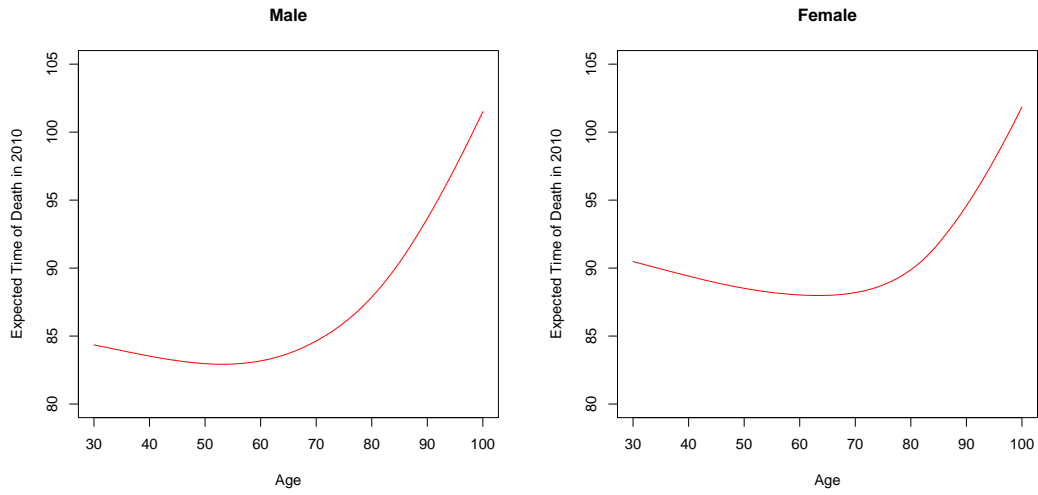


Figure 18.10.5: Expected time of death in 2010

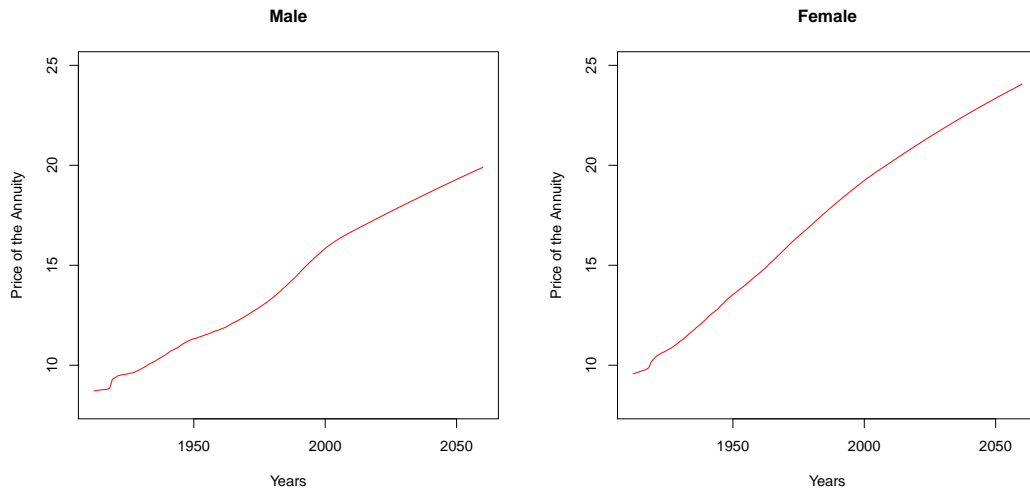


Figure 18.10.6: Price of the annuity for a person of age 65 when  $r = 1\%$

18 Comparison of Mortality Rate Models

Age	Male			Female		
	SFS	$q_x$	$q_x^1$	SFS	$q_x$	$q_x^1$
0	0.0053	0.0019	0.0018	0.0042	0.0014	0.0014
1	0.0004	0.0004	0.0004	0.0004	0.0003	0.0003
2	0.0003	0.0002	0.0002	0.0002	0.0002	0.0001
3	0.0002	0.0002	0.0002	0.0002	0.0001	0.0001
4	0.0002	0.0001	0.0001	0.0001	0.0001	0.0001
5	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001
6	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001
7	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001
8	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001
9	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001
10	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001
11	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001
12	0.0002	0.0002	0.0001	0.0001	0.0001	0.0001
13	0.0002	0.0002	0.0002	0.0001	0.0001	0.0001
14	0.0002	0.0002	0.0002	0.0001	0.0001	0.0001
15	0.0003	0.0003	0.0003	0.0002	0.0001	0.0001
16	0.0004	0.0004	0.0004	0.0002	0.0002	0.0002
17	0.0005	0.0006	0.0006	0.0003	0.0002	0.0002
18	0.0007	0.0008	0.0007	0.0003	0.0002	0.0002
19	0.0008	0.0009	0.0009	0.0003	0.0002	0.0002
20	0.0010	0.0009	0.0009	0.0003	0.0002	0.0002
21	0.0010	0.0010	0.0009	0.0003	0.0002	0.0002
22	0.0010	0.0009	0.0009	0.0003	0.0002	0.0002
23	0.0010	0.0009	0.0009	0.0003	0.0002	0.0002
24	0.0010	0.0009	0.0008	0.0003	0.0002	0.0002
25	0.0009	0.0008	0.0008	0.0003	0.0002	0.0002
26	0.0009	0.0008	0.0007	0.0003	0.0003	0.0002
27	0.0009	0.0008	0.0008	0.0003	0.0002	0.0002
28	0.0009	0.0007	0.0007	0.0004	0.0003	0.0002
29	0.0009	0.0007	0.0007	0.0004	0.0003	0.0003
30	0.0009	0.0007	0.0007	0.0004	0.0003	0.0003
31	0.0009	0.0008	0.0007	0.0004	0.0003	0.0003
32	0.0009	0.0008	0.0007	0.0004	0.0003	0.0003
33	0.0010	0.0008	0.0007	0.0005	0.0004	0.0003
34	0.0010	0.0008	0.0008	0.0005	0.0004	0.0004
35	0.0010	0.0008	0.0008	0.0005	0.0004	0.0004
36	0.0011	0.0009	0.0008	0.0006	0.0004	0.0004
37	0.0012	0.0009	0.0009	0.0006	0.0005	0.0005
38	0.0012	0.0010	0.0010	0.0007	0.0005	0.0005
39	0.0013	0.0011	0.0010	0.0007	0.0006	0.0005
40	0.0014	0.0012	0.0012	0.0008	0.0006	0.0006
41	0.0016	0.0013	0.0013	0.0009	0.0007	0.0007
42	0.0017	0.0014	0.0014	0.0010	0.0008	0.0007
43	0.0019	0.0016	0.0016	0.0011	0.0009	0.0008
44	0.0020	0.0018	0.0018	0.0012	0.0010	0.0009
45	0.0022	0.0020	0.0019	0.0013	0.0010	0.0010
46	0.0024	0.0022	0.0021	0.0014	0.0012	0.0012
47	0.0027	0.0025	0.0024	0.0015	0.0013	0.0013
48	0.0029	0.0028	0.0027	0.0017	0.0015	0.0015
49	0.0032	0.0030	0.0029	0.0019	0.0017	0.0017
50	0.0035	0.0033	0.0032	0.0021	0.0018	0.0017
51	0.0039	0.0038	0.0038	0.0023	0.0020	0.0019
52	0.0043	0.0042	0.0041	0.0025	0.0022	0.0022
53	0.0047	0.0047	0.0046	0.0027	0.0024	0.0023

18 Comparison of Mortality Rate Models

Age	Male			Female		
	SFS	$q_x$	$q_x^1$	SFS	$q_x$	$q_x^1$
54	0.0051	0.0053	0.0052	0.0030	0.0026	0.0025
55	0.0057	0.0059	0.0058	0.0032	0.0029	0.0028
56	0.0062	0.0066	0.0065	0.0035	0.0031	0.0031
57	0.0068	0.0072	0.0071	0.0038	0.0034	0.0033
58	0.0075	0.0079	0.0078	0.0041	0.0037	0.0036
59	0.0083	0.0089	0.0088	0.0044	0.0041	0.0040
60	0.0091	0.0097	0.0096	0.0048	0.0044	0.0043
61	0.0100	0.0107	0.0105	0.0052	0.0048	0.0047
62	0.0110	0.0119	0.0118	0.0056	0.0053	0.0052
63	0.0121	0.0131	0.0129	0.0061	0.0059	0.0057
64	0.0133	0.0144	0.0142	0.0066	0.0065	0.0063
65	0.0146	0.0161	0.0159	0.0072	0.0070	0.0069
66	0.0161	0.0177	0.0175	0.0079	0.0079	0.0078
67	0.0177	0.0194	0.0192	0.0086	0.0087	0.0085
68	0.0195	0.0214	0.0211	0.0095	0.0097	0.0096
69	0.0216	0.0234	0.0232	0.0104	0.0109	0.0107
70	0.0238	0.0257	0.0254	0.0116	0.0121	0.0119
71	0.0264	0.0284	0.0281	0.0129	0.0139	0.0137
72	0.0292	0.0313	0.0310	0.0144	0.0155	0.0153
73	0.0324	0.0349	0.0346	0.0161	0.0176	0.0174
74	0.0360	0.0387	0.0383	0.0182	0.0199	0.0197
75	0.0401	0.0429	0.0426	0.0206	0.0226	0.0224
76	0.0446	0.0477	0.0473	0.0235	0.0260	0.0257
77	0.0498	0.0532	0.0527	0.0268	0.0293	0.0291
78	0.0555	0.0590	0.0586	0.0306	0.0335	0.0332
79	0.0618	0.0659	0.0654	0.0351	0.0384	0.0381
80	0.0689	0.0735	0.0729	0.0403	0.0443	0.0440
81	0.0767	0.0824	0.0818	0.0464	0.0507	0.0503
82	0.0853	0.0921	0.0914	0.0533	0.0593	0.0589
83	0.0947	0.1016	0.1008	0.0613	0.0674	0.0669
84	0.1049	0.1124	0.1115	0.0704	0.0771	0.0765
85	0.1160	0.1261	0.1251	0.0806	0.0882	0.0876
86	0.1281	0.1361	0.1350	0.0922	0.0939	0.0932
87	0.1411	0.1488	0.1476	0.1051	0.1044	0.1037
88	0.1552	0.1624	0.1611	0.1194	0.1158	0.1150
89	0.1704	0.1769	0.1754	0.1349	0.1282	0.1273
90	0.1868	0.1923	0.1905	0.1517	0.1415	0.1405
91	0.2044	0.2086	0.2065	0.1695	0.1558	0.1547
92	0.2232	0.2259	0.2233	0.1881	0.1712	0.1699
93	0.2435	0.2440	0.2410	0.2072	0.1876	0.1861
94	0.2651	0.2630	0.2593	0.2263	0.2052	0.2033
95	0.2880	0.2830	0.2784	0.2454	0.2238	0.2215
96	0.3120	0.3039		0.2644	0.2434	
97	0.3368	0.3256		0.2833	0.2642	
98	0.3621	0.3481		0.3023	0.2861	
99	0.3873	0.3715		0.3214	0.3089	
100	0.4121	0.3956		0.3408	0.3329	
101	0.4358	0.4204		0.3609	0.3577	
102	0.4577	0.4458		0.3819	0.3836	
103	0.4772	0.4718		0.4044	0.4102	
104	0.4935	0.4983		0.4288	0.4377	
105	0.5059	0.5251		0.4558	0.4658	
106	0.5137	0.5523		0.4861	0.4946	
107	0.5680	0.5797		0.5207	0.5238	

Table 18.11: Average mortality rates between 1998 and 2003

## 19 Conclusion

The goal of this thesis is to show how one can construct a mortality table with the model which fits the available data best. However, the quality of each model depends on individual goals. We have studied various methods to model the mortality rates of a person with a certain age in a certain year. The common goals of these models are the measurement of the increase in longevity in the past as well as predicting its behaviour in the future. Most specialists agree that the life time of people will continue to increase over the years but there are different beliefs at which speed.

We tackled this issue by comparing several models to each other and analysing the results with respect to accuracy of the mortality rates in the past as well as in the future. Additionally, we calculated three measures for longevity, namely the survival probability of a person born in a certain year, the life expectancy at age 30 and 65 and the price of one particular annuity that is sold to a person of age 65. In the end, we identified the model which fits the data for men and women in Switzerland best from 1912 to 2010. We selected the second adaption of the generalized Nolfi model which includes a time-dependent base mortality rate. According to our chosen criteria, this model turned out to be most suitable to build a mortality table since its estimators as well as forecasts were very convincing from today's point of view. While comparing the models, we had to eliminate two potential choices very early since one was not realistic from the start due to insufficient information about the data and the other one did not yield reasonable mortality rate forecasts. Therefore, we decided not to investigate them further since no proper conclusion could be given.

With this particular mortality rate model, we were able to get an idea about the development of longevity in the past and we could make reasonable predictions about its future behaviour. We found out that if a person is born in a certain year, there exists an age after which his survival probability starts decreasing much faster than before. This results in a kink when plotting this probability. We have seen that from 1912 to 2010, this specific age increased by about 20 years for both genders. This suggests that nearly all people which are born in 2010 will survive at least until the age 75 for the males and 85 for the females when ignoring the atypical behaviour for ages below 30 which cannot accurately be replicated with any model. We expect that the survival probability will increase even further for a fixed age in future years.

The tendency to generally longer life is also apparent when analysing the life expectancy of people of age 30 and 65. We find that from 1912 to 2010, it increased by 18 years for a 30 year old man and by 20 years for a woman of the same age. For people of age 65, it increased by 9 and 13 years, respectively. For the next 50 years, we predict that people will be expected to live again 5 more years when they have already achieved the age 30. When they are 65 years old, the males will live 4 more years and the females 5 years which implies that their expected time of death will be at the age of 87.7 and 93, respectively. When additionally taking into account that the probability of surviving from age 30 to 65 will be approximately 0.97, it is possible that nearly everyone of age 30 achieves such a high age. For people with an age below 30, this does not need to be true necessarily since the model cannot accurately account for events like accidents, sickness or suicides.

## 19 Conclusion

These results are of crucial importance for life insurers and pension funds. They sell annuities at a price depending on the survival probability of the buyer. Many people buying such an annuity are 65 years old and receive their pension until they die. We have seen that the price of this particular annuity increased in the past by 8 monetary units for men and 11 for women, when the interest rate was assumed to be constant over time and equal to 1%. Due to the higher life time of people in 2060, this is expected to increase even further by again 3 and 4 more units, respectively. Since the annuity is paid out until the person dies, the life expectancy at age 65 plays a crucial role here. In general, the annuity is constructed such that the insurer is expected to face losses from it. Hence, calculating the mortality rates with a model which replicates reality and future behaviour as closely as possible is very important. Otherwise, the insurance company will face high risks if life expectancy increases more than expected.

We would like to emphasize that the problem we tackled in this thesis is very up-to-date and will still be important in the future as longevity predictions have always been and will continue to be changing over time. Hence, the choice of the model has a crucial impact on estimation and prediction. The key finding is that there exist many methods for modeling mortality rates. All these models work well for some purposes and are less appropriate for others. Therefore, it is crucial to know which criteria we need a model to meet and to derive which model is best to achieve these goals.

In this thesis, we compared ten models to each other. It would be interesting to extend the investigation to other models as well. One could repeat everything we did for these models and compare them to each other and to the ones we analysed here. Additionally, one could apply them to more than one dataset. The comparison of the models would then be independent of the choice of data which might yield further insight into the problem. However, this field of study is not limited to analysing and comparing models. Based on additional factors, for instance using sociological or geographical information, there is also a huge scope for new models. Furthermore, there exist other options for forecasting the mortality rates to the future. We have used a procedure which assumes that the change of mortality in the future is similar to the one in the past. As this does not necessarily represent reality, it would be interesting to follow a different path. Hence, this research area is clearly not yet saturated but there are many questions that still can be answered. We encourage further research in this topic as it is crucial for insurance companies to gain further insight into the development of longevity by finding models that can handle this problem efficiently.

# Bibliography

- [1] Butt, Z. and S. Haberman. 2010. *ilc: A Collection of R Functions for Fitting a Class of Lee-Carter Mortality Models using Iterative Fitting Algorithms*. Faculty of Actuarial Science and Insurance, City University London. Actuarial Research Paper No. 190.
- [2] Cairns, A.J.G., D. Blake and K. Dowd. 2006. *A Two-Factor Model for Stochastic Mortality with Parameter Uncertainty: Theory and Calibration*. The Journal of Risk and Insurance. 73:4, 687-718.
- [3] Denuit, M. and A.-C. Goderniaux. 2005. *Closing and Projecting Life Tables using Log-Linear Models*. Mitteilungen der Schweizerischen Aktuarvereinigung. 1:2005, 29-49.
- [4] Dowd K., A.J.G. Cairns, D. Blake, G.D. Coughlan, D. Epstein and M. Khalaf-Allah. 2010. *Backtesting Stochastic Mortality Models: An Ex Post Evaluation of Multiperiod-Ahead Density Forecasts*. North American Actuarial Journal. 14:3, 281-298.
- [5] Golub, G.H. and C. Reinsch. 1970. *Singular Value Decomposition and Least Squares Solutions*. Numerische Mathematik. 14:5, 403-420.
- [6] Johnson, R.A. and D.W. Wichern. 2007. *Applied Multivariate Statistical Analysis*. Pearson Education.
- [7] Koller, M. 2003. *Lebensversicherungsmathematik*. 17.10.2013.  
[http://www.math.ethz.ch/finance/Skripte/LV12\\_v060\\_151003.pdf](http://www.math.ethz.ch/finance/Skripte/LV12_v060_151003.pdf).
- [8] Lee, R.D. and L.R. Carter. 1992. *Modeling and Forecasting U.S. Mortality*. Journal of the American Statistical Association. 87:419, 659-671.
- [9] Lüthy, H., P.L. Keller, K. Binswanger and B. Gmür. 2001. *Adaptive Algorithmic Annuities*. Mitteilungen der Schweizerischen Aktuarvereinigung. 2:2001, 123-138.
- [10] Marguerite, C. 2011. *Methodik von Sterblichkeitsuntersuchungen*. 22.10.2013.  
[http://www.fam.tuwien.ac.at/~sgerhold/pub\\_files/theses/marguerite.pdf](http://www.fam.tuwien.ac.at/~sgerhold/pub_files/theses/marguerite.pdf).
- [11] Nolfi, P. 1959. *Die Berücksichtigung der Sterblichkeitsverbesserung in der Rentenversicherung nach der Optimalmethode der Spieltheorie*. Mitteilungen der Vereinigung Schweizerischer Versicherungsmathematiker. 59, 29-48.
- [12] Sachs, L. and J. Hedderich. 2006. *Angewandte Statistik. Methodensammlung mit R*. Springer.
- [13] Sawyer S. 2007. *Wishart Distributions and Inverse-Wishart Sampling*. 17.10.2013.  
<http://www.math.wustl.edu/~sawyer/hmhandouts/Wishart.pdf>.
- [14] Stahel, W. A. 1999. *Statistische Datenanalyse. Eine Einführung für Naturwissenschaftler*. Vieweg.
- [15] Swiss Federal Statistical Office. 2009a. *Die Zukunft der Langlebigkeit in der Schweiz*. Statistik der Schweiz.



## Bibliography

- [16] Swiss Federal Statistical Office. 2009b. *La mortalité par génération en Suisse*. Statistique Vaud.
- [17] Swiss Federal Statistical Office. 2013a. *Definitionen*. 17.10.2013.  
<http://www.bfs.admin.ch/bfs/portal/de/index/themen/01/11/def.html>.
- [18] Swiss Federal Statistical Office. 2013b. *Bevölkerungsbewegung - Analysen: Sterbetafeln für die Schweiz 1998/2003*. 30.12.2013.  
[http://www.bfs.admin.ch/bfs/portal/de/index/themen/01/06/blank/dos/la\\_mortalite\\_en\\_suisse/tab101.html](http://www.bfs.admin.ch/bfs/portal/de/index/themen/01/06/blank/dos/la_mortalite_en_suisse/tab101.html).
- [19] Wüthrich, M.V. 2013. *Non-Life Insurance: Mathematics & Statistics*. 23.10.2013.  
[http://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=2319328](http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2319328).
- [20] Yates, F. 1934. *Contingency Tables Involving Small Numbers and the  $\chi^2$  Test*. Supplement to the Journal of the Royal Statistical Society. 1:2, 217-235.

# Acknowledgments

I thank Prof. Dr. Mario V. Wüthrich and Dr. Klemens Binswanger for their help in preparing this thesis. I also say thanks to Robert Gantner and Nadja Ulrich for correcting parts of it. Additionally, I thank my family for their moral and financial support.

Gaby Binder, 26.02.2014

# A Appendix

## A.1 Notations

Variable	Description
$a_x, b_x, c_x$	Age-specific constants in the models of Lee and Carter
AIC	Akaike's information criterion
$\alpha, \beta$	Significance levels
$\alpha_x$	Age-specific constant in the marginal mortality rate model
$\boldsymbol{\alpha}_t$	Stochastic process in the adapted Nolfi models
$c$	Standard deviation
$C$	Cholesky factor of the covariance matrix
$c_x$	Age-specific constant in the extrapolation model
$C(t)$	Cash-flow of an annuity
$D_t$	Total number of deaths
$d_t$	Realization of $D_t$
$D_{x,t}$	Number of deaths
$d_{x,t}$	Realization of $D_{x,t}$
$D_{x,t}^*$	Normal approximation of $D_{x,t}$
$d_w$	Additional factor in the extended model of Lee and Carter
$D(\hat{\boldsymbol{\theta}})$	Scaled deviance
$\text{dev}_{x,t}$	Deviance from $d_{x,t}$
$df$	Degrees of freedom
$\Delta^v$	Discrete derivative of order $v$
$e_{x,t}$	Life expectancy
$\varepsilon_{x,t}$	Error term in the models of Lee and Carter
$g : [0, 1] \rightarrow \mathbb{R}$	Link function
$H_0$	Null-hypothesis
$k_t$	Time-varying index in the models of Lee and Carter
$l(\boldsymbol{\theta})$	Log-likelihood function
$l_{x,t}$	Number of living people
$\lambda_x$	Age-specific constant in the Nolfi models
$M$	Maturity
$m_s, m_{x,s}$	Mortality loading
msep	Mean squared error of prediction
$\boldsymbol{\mu}$	Drift parameter
$n_i$	Number of persons holding $i$ policies
$\omega + 1$	Number of ages
$(\Omega, \mathcal{F}, \mathbb{P}, \mathbb{F})$	Filtered probability space
$p_{x,t,j}$	Survival probability

A Appendix

Variable	Description
$PV(x, t)$	Price of an annuity
$\phi$	Overdispersion
$Q_{x,t}$	Annual mortality rate
$q_{x,t}$	Expected mortality rate
$\tilde{q}_{x,t}$	Forecasted mortality rate
$\hat{q}_{x,t}$	Estimated mortality rate
$q_{x,t}^1$	Expected mortality rate of order 1
$q_{x,t}^{\Theta}$	Mortality rate including a priori knowledge
$r$	Interest Rate
$r_{x,t}^D$	Deviance residuals
$s$	Beginning of the current year
$SS(\hat{\theta})$	Sum of squares
$\sigma_i$	Singular values of a matrix
$\sigma^2$	Variance of $\varepsilon_{x,t}$
$t$	Year
$\theta$	Parameter vector
$\theta_t$	Stochastic process for extrapolating the data
$\Theta_t$	Stochastic process in the Cairns-Blake-Dowd model
$\tau + 1$	Number of years
$T_{x,t}$	Remaining life time
$V$	Covariance matrix
$v_{t-x}$	Cohort effect in the generalized model of Lee and Carter
$w_i$	Weights
$x$	Age
$y$	Maximal age to which the data are extrapolated
$\mathbf{Z}_t$	Normally distributed random vector