Delft University of Technology
Faculty of Electrical Engineering, Mathematics and Computer Science
Delft Institute of Applied Mathematics

Longevity Risk
and the Consequences for the Actuary

A thesis submitted to the
Delft Institute of Applied Mathematics
in partial fulfillment of the requirements

for the degree

MASTER OF SCIENCE
in
APPLIED MATHEMATICS

by

DAVID PLOMP

Amsterdam, the Netherlands
April 2009

MSc THESIS APPLIED MATHEMATICS

"Longevity Risk and the Consequences for the Actuary"

DAVID PLOMP

Delft University of Technology

Daily supervisors	**Responsible professor**
Dr. H.P Lopuhaä	Prof. dr. F.M Dekking
Dr. C. Kraaikamp	

Other thesis committee members

Ir. C. Vos AAG	Drs. A.N. Jager AAG

April 2009 Amsterdam, the Netherlands

Contents

Introduction

The motivation for this thesis

This thesis is written in order to obtain a Master degree in Applied Mathematics at the Technical University of Delft. The research has been done at Hewitt Associates, which is a global human resources, outsourcing and consulting firm. I worked at the office in Amsterdam for a period of one year starting in March 2008. More specifically, I did research at the department of Retirement and Financial Management (RFM). This department deals with financial risks regarding retirement, and consults pension funds about these risks. The scientific component of this discipline is called *actuarial science*. Someone who earned a degree in actuarial science is called an *actuary*.
Nowadays people on average life longer than any period before in history. In the actuarial world this phenomenon is called *longevity* which literally means long life span. Although longevity is a great achievement for humanity, it poses a risk to pension funds. In order to avoid financial setbacks in the future, actuaries need to anticipate on longevity. Hewitt wants to have more insight into this risk of longevity. Therefore the core question of this study has been: What are the financial consequences of longevity for pension funds?

The method of this thesis

In this thesis, current mortality and its development in the future is investigated. Mostly we used *death rates* to measure mortality. A death rate of a given group of people is defined as the percentage of that group that dies within one year. To illustrate the development of death rates in the Netherlands, in Figure 1 we displayed the observed death rates for males and females aged 65 for the period 1950-2006.

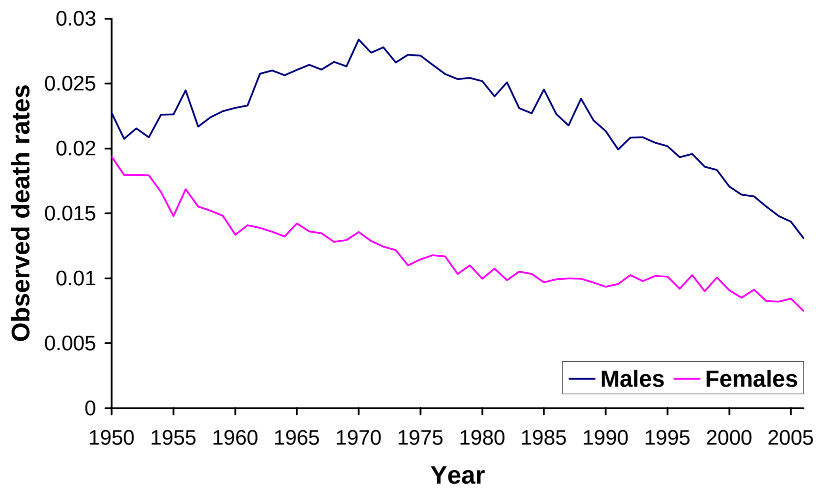


Figure 1: Observed death rates of people aged 65 between 1950 and 2006

In the Netherlands, mortality data is collected by the National Bureau of Statistics (in Dutch: Centraal Bureau voor de Statistiek (CBS)) since 1950. They provided the data that was used as input for our model to predict mortality. The applied model is designed by Lee and Carter in 1992. Nowadays, this is one of the most frequently used models worldwide, e.g. it is used by the American Census Bureau and the OECD to forecast future death probabilities.

The method of Lee-Carter is to capture age-specific trends from an observed period and extrapolate these trends into the future. With the model a forecast that is most likely to happen can be constructed, as well as a formulation for the uncertainty of this forecast. This enables us to make the financial consequences of longevity risk tangible. For a given level of certainty, we can predict the future expenses of a pension fund by a *range* which is linked to this level, given that the model is correct. For instance, when one wants to know the future expenses of a fund with a certainty of 90%, a corresponding range can be determined in such a way, that the probability that future expenses will lie within that range, is 90%, given that the model is correct.

Outline of the thesis

The goal of Chapter 1 is to provide background on mortality models, such as notation and concepts that will be frequently used, and the historic development of mortality modeling. Chapter 2 describes the method of Lee and Carter step by step. First the observed death rates are fitted using the Singular Value Decomposition (SVD), then historic trends are extrapolated with an appropriate ARIMA time series model. In Chapter 3 the Lee-Carter model is applied to Dutch death rates. We discuss the quality of the fit and explain our choice for a specific ARIMA model. Chapter 4 is dedicated to construct a forecast using the model defined in Chapter 3. Special attention is given to this topic because it directly influences the results of Chapter 5. An analytical way to forecast mortality is compared with simulations. In Chapter 5 two mortality forecasts are introduced that are frequently used in the Netherlands, the AG and the CBS prognosis. Two experiments are performed where these forecasts are compared with the implementation of Lee-Carter, to illustrate the effects of longevity on the financial situation of an average Dutch pension fund.

Acknowledgements

There are many people who have helped me to finish this thesis. First of all, I would like to thank all my supervisors. Rik and Cor, your feedback was indispensable to me and to the development of my work. Coenraad, thanks for all your interest and trust in my research and for the useful idea's you brought in. Arnold, I also appreciated your interest in my work and the fact that you were always there for me when I needed your help.
I owe a lot to Kiona and Andre, you often helped me out when I did not see the trees through the forest anymore. I like to thank Hewitt for supporting me so well and letting me participate in all the excursions.
Finally I like to thank my family, friends and especially Eva, for their understanding of my endless stories about this thesis.

Chapter 1

Mortality models

The objective of modeling is to draw conclusions from a given data set. A model attempts to describe a *pattern* in the data by a number of factors, which are believed to have a causal relation with this pattern. As a result, a low dimensional problem arises in which an optimum has to be found. It is important that long term patterns (trends) are separated from temporary events (noise) as a result of randomness, especially when the model is used for forecasting.

We are interested in mortality probabilities and the prediction of these probabilities in the future for doing actuarial calculations. The choice of a model to determine these probabilities is essential because it will influence the outcomes of the study. In this chapter we introduce notation and background in mortality models. In Section 1.1 we discuss the basic concepts from survival analysis, which will serve as a framework to analyze mortality.

The first reliable censuses of an entire population and the registration of the total number of deaths were done in Sweden around 1750 (Wilmoth, [29]). From this data a table with estimated mortality probabilities could be derived. Explanatory models of mortality are known from the early eighteenth century. Section 1.2 gives an overview of the most influential models from the past. Since the 1950's, data has become abundant and more reliable than in any other period before in history. The classic mortality models were replaced by more advanced ones that also described the *development* of mortality in time, to make them suitable for forecasting purposes. This resulted into a new type of mortality models qualified here as *modern mortality models*. This is the subject discussed in Section 1.3. In Section 1.4 we discuss how to order mortality data. We will explain the difference between *period* and *cohort* effects and show several ways to determine *death rates*.

1.1 Basic concepts

We denote the *age* of a person by x, $x \in \mathbb{R}^+$, where x is expressed in years. The random variable T is defined as the age at death, corresponding to a distribution function $F(x)$. The probability to survive until x is denoted by $S(x)$,

$$S(x) = 1 - F(x) = 1 - P(T \leq x), \quad x \geq 0.$$

The *hazard rate* $\mu(x)$ is the risk of instantaneous death at x,

$$\mu(x) = \lim_{h \downarrow 0} \frac{P(T \in (x, x+h] | T > x)}{h}. \tag{1.1}$$

When we assume $F(x)$ to be a differentiable function, (1.1) can be written as

$$\mu(x) = \frac{1}{1-F(x)} \frac{d}{dx} F(x) = -\frac{1}{S(x)} \frac{d}{dx} S(x) = -\frac{d}{dx} \log S(x). \tag{1.2}$$

Integrating $\mu(x)$ from t_1 to t_2, $0 \leq t_1 \leq t_2 < \infty$ leads to

$$\int_{t_1}^{t_2} \mu(s) ds = -\int_{t_1}^{t_2} \frac{d}{ds} \log S(s) ds = \log S(t_1) - \log S(t_2) = \log \frac{S(t_1)}{S(t_2)}.$$

This is equivalent to:

$$\frac{S(t_1)}{S(t_2)} = e^{\int_{t_1}^{t_2} \mu(s) ds} \quad \text{or} \quad \frac{S(t_2)}{S(t_1)} = e^{-\int_{t_1}^{t_2} \mu(s) ds}. \tag{1.3}$$

Another way to derive (1.3), is to write $\mu(x) = -\frac{1}{S(x)} \frac{d}{dx} S(x)$ as the boundary value problem:

$$\begin{aligned} S'(x) &= -\mu(x) S(x), \\ S(0) &= 1. \end{aligned}$$

A solution of this problem, assuming that $\mu(x)$ is integrable between 0 and x, is given by,

$$S(x) = e^{-\int_0^x \mu(s)ds}.$$

Taking t_1, t_2 as before we obtain the desired result,

$$\frac{S(t_1)}{S(t_2)} = \frac{e^{-\int_0^{t_1} \mu(s)ds}}{e^{-\int_0^{t_2} \mu(s)ds}} = e^{\int_0^{t_2} \mu(s)ds - \int_0^{t_1} \mu(s)ds} = e^{\int_{t_1}^{t_2} \mu(s)ds}. \tag{1.4}$$

It is often convenient to consider the remaining life span of a person who is alive at age x. We therefore define

$$R_x := T - x | T > x, \quad x \geq 0.$$

Note that $R_0 = T$. The distribution function of R_x is denoted by F_x and can be written in terms of $F(x)$:

$$F_x(a) = P(R_x \leq a) = \frac{F(x+a) - F(x)}{1 - F(x)}.$$

Let $p(x)$ be the probability that a person aged x is still alive at $x+1$, i.e.,

$$p(x) = P(R_x > 1) = P(T > x+1 | T > x) = \frac{P(T > x+1)}{P(T > x)} = \frac{S(x+1)}{S(x)}.$$

By choosing appropriate boundaries, it follows from (1.3) and (1.4) that,

$$p(x) = e^{-\int_x^{x+1} \mu(s)ds} = 1 - q(x), \tag{1.5}$$

where $q(x)$ is called the annual death probability of somebody alive at age x.

Life expectancy e_x is the expected remaining lifetime of a person aged x. This can be calculated by,

$$e_x = E[R_x] = \int_{y \geq x} y f_x(y) dy, \tag{1.6}$$

where $f_x(y)$ is the probability density function of R_x. Life expectancy from birth can be written as

$$e_0 = E[R_0] = E[T] = \int_{x \geq 0} x f(x) dx.$$

where $f(x)$ is the probability density function of T. For applications age is discretized and restricted to be integer valued. This discretized life expectancy $\tilde{e}_x$ is defined as the expected number of *complete* years that a person remains alive. Let us first consider the probability that someone aged x lives another h complete years (before dying):

$$P(h < R_x \leq h+1) = P(R_x \leq h+1 | R_x > h) P(R_x > h).$$

Note that

$$P(R_x \leq h+1 | R_x > h) = P(R_{x+h} \leq 1) = q(x+h)$$

and

$$P(R_x > h) = P(R_x > 1) P(R_{x+1} > 1) \cdots P(R_{x+h-1} > 1) = \prod_{j=0}^{h-1} p(x+j).$$

The expected number of complete years can be stated in terms of $q(x)$:

$$\begin{aligned} \tilde{e}_x &= \sum_{h=1}^{\infty} h P(h < R_x \leq h+1) = \sum_{h=1}^{\infty} h P(R_x \leq h+1 | R_x > h) P(R_x > h) \\ &= \sum_{h=1}^{\infty} h q(x+h) \prod_{j=0}^{h-1} p(x+j) = \sum_{h=1}^{\infty} h q(x+h) \prod_{j=0}^{h-1} (1 - q(x+j)). \end{aligned} \tag{1.7}$$

1.2 Classic mortality models

In the past there have been numerous attempts to describe mortality. Usually this was restricted to small groups, such as the German nobility in the twelfth century. After the Middle Ages, when trade expanded and welfare increased, there was a great demand for reliable mortality data. In London, where life insurances emerged on the market, John Graunt carried out a large study on mortality probabilities in 1662. He influenced the brothers Lodewijk and Christiaan Huygens, who calculated a life expectancy of 18.2 years for Dutch newborns. In Sweden reliable mortality data on a national scale appeared around 1750. Abraham de Moivre was the first person who tried to capture mortality with an analytical function, in 1729. He assumed a linear relation between survival and age:

$$S(x) = 1 - \frac{x}{\omega}, \tag{1.8}$$

where ω denotes the highest possible age (ω had to be determined by the modeler). The concept of a hazard rate was unknown at that time (Bernoulli introduced it in 1766). In order to compare de Moivre with the formula's of his successors, who usually modeled $\mu(x)$, we use (1.2) to rewrite (1.8) as,

$$\mu(x) = \frac{1}{\omega - x} \qquad \text{De Moivre (1729).}$$

Benjamin Gompertz (England, 1779-1865) introduced the first explanatory mortality model. He talked about man's resistance to death, which he believed decreased exponentially. His famous 'law of mortality' states:

$$\mu(x) = Be^{\Theta x} \qquad \text{Gompertz (1825),}$$

where $B, \Theta > 0$. Every year, one gets 'less resistance to death' leading to an increase of μ with a factor e^{Θ}. His model greatly improved the fit with observed mortality rates, especially for the ages over 35, the most important age group for (life) insurers. Makeham took the work of Gompertz a little further, resulting into a model which is later called the 'Gompertz-Makeham law of mortality',

$$\mu(x) = A + Be^{\Theta x} \qquad \text{Makeham (1867),}$$

where A is a positive constant. Makeham explained this new parameter as the risk of dying that is independent of someone's age. This could be certain diseases or (traffic) accidents. Because of this extension, the fit for higher ages is generally better. Despite the improvement, the real success of the model has always been the 'middle' ages, which roughly comes down to the ages 35 to 70. Until 1997, Gompertz-Makeham has been used by the Dutch actuarial society to estimate death probabilities. From the observed death rates in 2006, displayed in Figure 1.1, we can see that mortality between age 35 and 85 can be well approximated by an exponential function. This

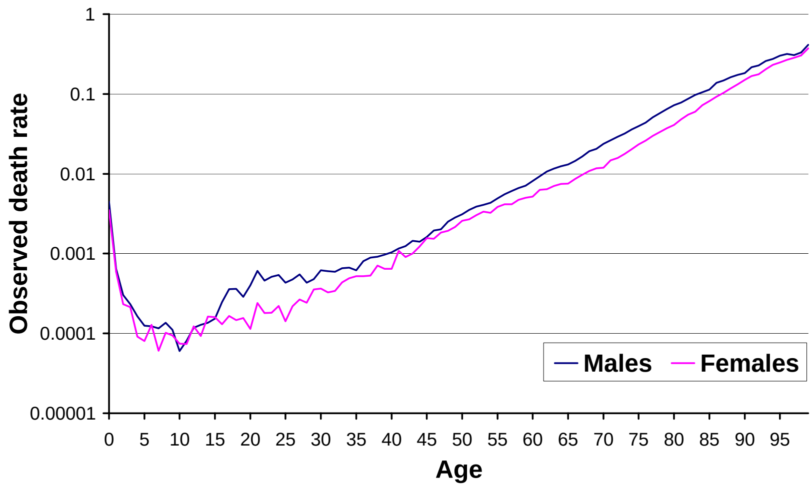


Figure 1.1: Observed death rates in 2006.

follows from the linearity in the chart between these ages and the logarithmic scaling on the vertical axis, which transforms an exponential course into a linear one.

The Gompertz-Makeham model had great influence on his successors. New models copied the exponential increase in μ, but improved the model, especially for lower ages. Two examples are the models of Thiele and Perks,

$$\mu(x) = A_1 e^{-B_1 x} + A_2 e^{-\frac{1}{2}B_2(x-c)^2} + A_3 e^{B_3 x} \qquad \text{Thiele (1872)},$$

$$\mu(x) = \frac{A + Be^{\Theta x}}{Ke^{-\Theta x} + 1 + Ce^{\Theta x}} \qquad \text{Perks (1932)}.$$

All parameters in these models are positive. Thiele, a Danish actuary, designed a model that involves three terms; besides the Gompertz term $A_3 e^{B_3 x}$, it has a separate term to describe high mortality rates for newborns: $A_1 e^{-B_1 x}$. The remaining term $A_2 e^{-\frac{1}{2}B_2(x-c)^2}$ describes the 'accident hump'. This term refers to the high risk that adolescents face to die in an accident. The hump can also be observed by the death rates for the ages between 16 and 25 (especially for male death rates) in Figure 1.1.

The British actuary Perks has been influenced by Gompertz' model too. His model is more simplistic than Thiele's; it only has 5 parameters, against 7 for Thiele. In the numerator we recognize Gompertz-Makeham. In the denominator, two terms are

incorporated to achieve a better fit for lower ($Ke^{-\Theta x}$) and higher ($Ce^{\Theta x}$) ages.
The latest adaptation to Gompertz is the model proposed by Heligman and Pollard in 1980:

$$\frac{q(x)}{1-q(x)} = A^{(x+B)^c} + De^{-E(\log x - \log F)^2} + GH^x \qquad \text{Heligman and Pollard (1980).}$$

Their model has 8 real-valued parameters which all have a demographic interpretation. It is heavily influenced by Thiele, as it has three terms to describe different mortality patterns, including a Gompertz term. They believed that a model for q, instead of μ, would result into a closer fit. This view was based on a research study, done by the institute of actuaries in Australia ([18], p. 50). The formulation of their model includes the term $\frac{q(x)}{1-q(x)}$, the value for $q(x)$ can be solved from this quotient.

The models discussed in this section can be qualified as *classic*. They have in common that they consist of only one variable: age. Wilmoth writes that: "Life expectancy has been increasing not just in industrialized societies but around the world. The rise in life expectancy at birth probably began before the industrial era, ..." [29], p. 1113. By using a classical model to calculate the life expectancy of a population, people will on average become older than expected. This will be discussed in more depth in the next section.

1.3 Modern mortality models

The models described in the previous section were not designed to reflect the *development* of mortality, which is a disadvantage. Actuaries attempt to make a projection of future expenses of a pension fund. This projection is partly based on the life-expectancy of the members and hence of *future* death probabilities. History shows that the average age of death in the Netherlands is increasing at a fast pace. In 1970, men died at an average age of 71 and women at age 75. In 2000, this gone up to 75 for men and 81 for women. However, until 2007 Dutch actuaries used current

death rates as an estimate for future death probabilities and they calculated future payments and premiums based on this information.
A consequence for a pension fund, when death rates turn out lower as expected, is that it gets confronted with higher expenses which affect the financial situation. This phenomenon is known as *longevity risk*. To protect a fund against longevity risk, actuaries used to conduct an *age-shift*. This means that a person is assumed to have the survival chances of a younger person (usually a couple of years) in the future. Since death rates show an upward trend with age, this age-shift obviously creates some security against a decrease of death rates in the near future. The problem with doing age-shifts is that it is arbitrary. An age shift can only be justified when the pattern of time improvement is approximated by the difference between death rates of consecutive ages. Another problem with doing age-shifts is that another static estimate is created; it might work well in the short run, but cannot be used in the long run because mortality is dynamic in time. The problem of longevity risk was recognized by the Dutch government, which imposed pension funds to take into account that mortality rates will continue to decrease in the future. The law was implemented on January 1st, 2007. Since then the demand for reliable forecasts has increased. We will call a model that incorporates time development a *modern mortality model*. An example is the *Lee-Carter* model, which will be discussed in the following chapters.

We conclude this section by extending the most important quantities of Section 1.1 to include a time variable. Let T_c be the time until death of an individual born at time c. The time-dependent hazard rate can be stated as $\mu_c(x)$,

$$\mu_c(x) = \lim_{h \downarrow 0} \frac{P(T_c \in (x, x+h] | T_c > x)}{h}.$$

The remaining life span $R_{x,t}$ gives the time until death of a (living) person aged x at time t:

$$R_{x,t} = T_{t-x} - x | T_{t-x} > x. \tag{1.9}$$

The probability that a person aged x at time t will survive until $t+1$ is given by

$$p(x,t) = P(R_{x,t} > 1) = P(T_{t-x} > x+1 | T_{t-x} > x).$$

The annual death probability reads

$$q(x,t) = 1 - p(x,t) = 1 - e^{-\int_x^{x+1} \mu_{t-x}(s)ds}. \tag{1.10}$$

Finally, the expected number of *complete* years that a person aged x at time t remains alive, is given by

$$\begin{aligned} \tilde{e}_{x,t} &= \sum_{h=0}^{\infty} hP(h < R_{x,t} \leq h+1) = \sum_{h=0}^{\infty} hP(R_{x,t} \leq h+1 | R_{x,t} < h) P(R_{x,t} < h) \\ &= \sum_{h=1}^{\infty} hq(x+h, t+h) \prod_{j=0}^{h-1} (1 - q(x+j, t+j)). \end{aligned} \tag{1.11}$$

1.4 Death rates

We have formulated a theoretical framework to analyze mortality in Section 1.1 and (at the end of) Section 1.3. In real life these quantities cannot be computed exactly, but estimates have be made by using data. In this research we are in particular interested in estimating $q(x,t)$, which was defined in (1.10), for the purpose of forecasting. Therefore time will be discretized into one year periods to determine observed *death rates*. Because of this discretization, from now we will restrict x and t to be integer valued.

The effect of time on mortality can be measured by period or by cohort. With a *period* approach we look at the change of mortality within a period, for instance a calender year. A *cohort* approach is used to observe the change of mortality on a cohort of people, like a group of people which all have the same year of birth. This is a very small difference, but can nicely be illustrated by a Lexis diagram. This diagram, founded by a German demographer, shows the relation between time and age on the horizontal and vertical axis respectively. It enables us to see the difference between the approaches. An example of a Lexis diagram is shown in Figure 1.2, where the solid line represents the (partially expected) course of life from the author, who is born on 17-02-1983. A cohort is taken to be the group of people who have the same year of birth and a period is chosen to be a calender year.

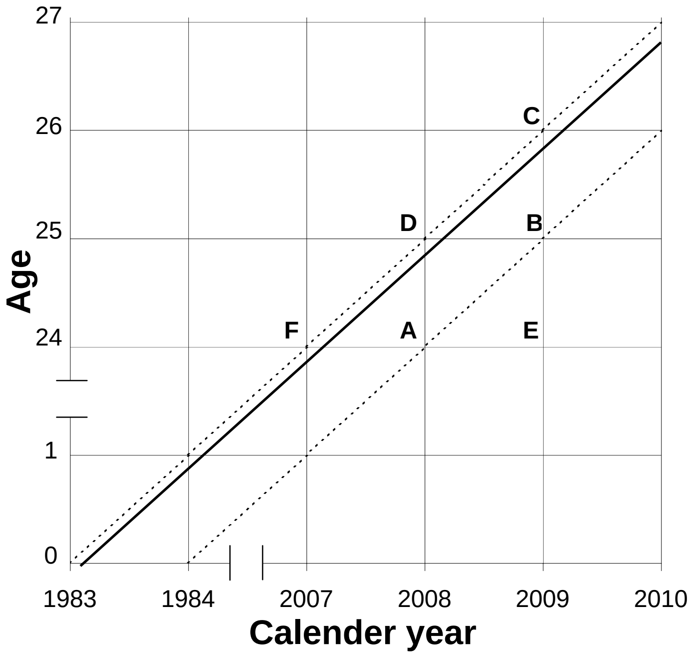


Figure 1.2: A Lexis diagram. The course of life from the author is denoted by the solid line.

For a given set of mortality data we can subdivide death cases according to one of the following approaches to create different types of death rates.

- **period-cohort**. The number of deaths during year t from people who are born in year c. In the Figure, deaths in $ABCD$ denote the group of people in 2008, who are born in 1983.

- **period-age**. The number of deaths in year t from people aged x. In the Figure, deaths in $AEBD$ denote the group of people aged 24 in 2008.

- **age-cohort**. This is the number of deaths from people, aged x, who are born in year c. In the Figure, deaths in $FABD$ denote the group of people aged 24, who are born in 1983.

Chapter 2

The Lee-Carter model

The Lee-Carter model [20] was designed in 1992 to predict future mortality probabilities of the US population. Nowadays the model is used by scientists around the globe and some major institutions, such as the United Sates Census Bureau and the OECD. Many modifications of the model have been proposed, among others by Bell [1], Booth [3], and Wilmoth [28].
In this chapter we give an outline of the Lee-Carter model. In the following chapters the model will be applied to Dutch mortality data to construct forecasts. The theory of this chapter will help to understand the results that will be presented in these chapters. Section 2.1 states the notation used and gives an outline of the method. The implementation of the model can be divided in two parts. The first part is to fit the model to the observed death rates, which will be described in Section 2.2. Secondly a time series model is used to predict future death probabilities, which is the topic of Section 2.3.

2.1 Outline of the method

Let us define the quantity that was originally modeled by Lee and Carter, the *central death rate* $M(x,t)$, which is a death rate based on the *period-age* approach (the different approaches were discussed in Section 1.4). It is obtained by

$$M(x,t) = \frac{D(x,t)}{N(x,t)}, \tag{2.1}$$

where $D(x,t)$ is the number of people aged x that die during calender year t and $N(x,t)$ is the number of x years olds at midyear.
According to the Lee-Carter model, $M(x,t)$ can be described as

$$M(x,t) \quad = \quad e^{a_x + b_x k_t + \varepsilon_{x,t}}. \tag{2.2}$$

The parameters a_x and b_x reflect age effects and k_t the development of mortality in time. The noise term $\varepsilon_{x,t}$ is assumed to have mean 0 and variance σ_x^2.
The implementation of the model can be divided in two parts,

- Fitting the model using the Singular Value Decomposition.
- Forecasting k_t with an ARIMA time series model.

We first look for parameters that fit the observed data in the least squares sense. For this purpose we define a set of age classes $\mathcal{X}$ and a set of observed calender years $\mathcal{T}$. Let p be the number of age classes and n be the number of observed years, leading to a total of $2p+n$ parameters to be estimated. We label the elements $x \in \mathcal{X}$ and $t \in \mathcal{T}$ in ascending order, so x_1 is the youngest age group and x_p denotes the highest age group. We use vector notation to write the parameters as $a := (a_1, \cdots, a_p)^T$, $b := (b_1, \cdots, b_p)^T$ and $k := (k_1, \cdots, k_n)^T$, where x and t are removed from the subscript for simplicity.
It is easy to see that (2.2) is an underdetermined model. Suppose that $\tilde{a}, \tilde{b}, \tilde{k}$ is a solution, then $\tilde{a} - \tilde{b}c, \tilde{b}, \tilde{k} + c$ is also a solution for every $c \in \mathbb{R}$, since

$$\tilde{a} - \tilde{b}c + \tilde{b}(\tilde{k} + c) = \tilde{a} + \tilde{b}\tilde{k}.$$

The same holds for $\tilde{a}, \tilde{b}c, \tilde{k}/c$. The constraints

$$\sum_{t \in \mathcal{T}} k_t = 0 \quad \text{and} \quad \sum_{x \in \mathcal{X}} b_x = 1 \tag{2.3}$$

are added to make sure that a solution is unique.
These constraints give the parameters an actual interpretation. Summing over t gives

$$\sum_{t\in\mathcal{T}} \log M(x,t) = \sum_{t\in\mathcal{T}} (a_x + b_x k_t + \varepsilon_{x,t}) = n a_x + \sum_{t\in\mathcal{T}} \varepsilon_{x,t}.$$

When we take expected value

$$E\left[\sum_{t\in\mathcal{T}} \log M(x,t)\right] = n a_x \quad \Longleftrightarrow \quad a_x = \frac{E\left[\sum_{t\in\mathcal{T}} \log M(x,t)\right]}{n}, \tag{2.4}$$

for every $x \in \mathcal{X}$ and conclude that a_x is the time-average, age specific, log death rate. The b_x are normalized and give the pace of mortality change compared to other ages. If b_{x_s} is high for some s, this means that *relative to other ages*, the mortality rate of people aged x_s changes rapidly. More specific, b_x satisfies:

$$\frac{dE[\log(M(x,t))]}{dt} = b_x \frac{dk_t}{dt}.$$

The k_t give the overall rate of change in mortality in time, which is typically a decreasing sequence, since death rates show a decreasing trend for almost all ages.
The second part of the implementation concerns the forecast. The values $k_1, \ldots, k_n$ are treated as a time series, which are extrapolated into the future to derive future (central) death rates using (2.2). This is done using the theory of ARIMA models. The construction of a forecast will be discussed in Section 2.3.

2.2 Estimating the parameters

In this section we describe how the parameters of (2.2) are estimated. We start with a set of *observed* central death rates $m(x,t)$, $x \in \mathcal{X}, t \in \mathcal{T}$. Equation (2.4) suggests that a_x can be estimated by

$$\hat{a}_x = \frac{\sum_{t\in\mathcal{T}} \log m(x,t)}{n},$$

for every $x \in \mathcal{X}$.

The other parameters b, k are estimated by the solution of the optimization problem

$$\begin{aligned} \min_{b,k} \quad & \sum_{x\in\mathcal{X}} \sum_{t\in\mathcal{T}} (\log m(x,t) - \hat{a}_x - b_x k_t)^2 \\ \text{s.t.} \quad & \sum_{x\in\mathcal{X}} b_x = 1, \quad \sum_{t\in\mathcal{T}} k_t = 0. \end{aligned} \tag{2.5}$$

The Singular Value Decomposition (SVD) can be used to find a solution. Let G be the $p \times n$ matrix with elements $g_{ij} = \log m(x_i, t_j) - \hat{a}_{x_i}$. The elements of G give the dispersion of the mean corrected logarithm of the observed central death rates.

The objective function of (2.5) can be rewritten as,

$$\min_{b,k} ||G - bk^T||_F^2, \tag{2.6}$$

where $||.||_F$ denotes the Frobenius matrix-norm defined as,

$$||G||_F = \left(\sum_i \sum_j g_{ij}^2 \right)^{1/2}.$$

The product bk^T is a matrix of rank one, so as a first step we look for the best rank one approximation of G. The SVD is a method to decompose a matrix of rank l into l matrices of rank 1. It can be represented as

$$G = \sigma_1 w_1 v_1^T + \sigma_2 w_2 v_2^T + \cdots + \sigma_l w_l v_l^T. \tag{2.7}$$

The scalars $\sigma_1 \geq \cdots \geq \sigma_l > 0$ are the *singular values* of G and $l = \text{Rank}(G)$. The vectors $w_1, \ldots, w_l$ and $v_1, \ldots, v_l$ are the left and right *singular vectors* of G respectively, corresponding to these singular values.

The $p \times p$ matrix GG^T and $n \times n$ matrix $G^T G$ are symmetric and nonnegative definite. Moreover, they both are of rank l and share the same set of l positive eigenvalues $\mu_1, \mu_2, \ldots, \mu_l$ (not necessarily distinct). The singular values of G are the square root of these eigenvalues. The left singular vectors are the orthogonal set of normalized eigenvectors of GG^T, corresponding to $\mu_1, \mu_2, \ldots, \mu_l$. The orthogonality property

follows from the fact that any real symmetric matrix can be diagonalized by an orthogonal matrix, for a proof see [24]. The right singular vectors, also orthogonal and of unit length, are obtained in a similar way from G^TG.
Let $G_i := \sigma_i w_i v_i^T$, $0 < i \leq l$, to write

$$G = G_1 + G_2 + \cdots + G_l.$$

The Frobenius norm of G_i is equal to σ_i as can be concluded from,

$$||G_i||_F^2 = \sigma_i^2 \sum_i \sum_j (w_i v_j^T)^2 = \sigma_i^2 \sum_i w_i^2 = \sigma_i^2. \tag{2.8}$$

The Frobenius *inner product* is defined as $G \cdot H = \sum_{i,j} g_{ij} h_{ij}$, which gives us the relation

$$G \cdot G = ||G||_F^2.$$

It is not hard to see that $G_i \cdot G_j = 0$ for $i \neq j$. Using this orthogonality property we get that

$$\begin{aligned} ||G||_F^2 &= ||G_1 + G_2 + \cdots + G_l||_F^2 = ||G_1||_F^2 + ||G_2||_F^2 + \cdots + ||G_l||_F^2, \\ &= \sigma_1^2 + \sigma_2^2 + \cdots + \sigma_l^2. \end{aligned}$$

Together $G_1, \ldots, G_l$ form an orthogonal *basis* of G. This suggests that the largest of these components is the best rank 1 approximation of G. Indeed, it is proved in [19] that G_1 is the best rank 1 approximation of G. The norm of the error reads

$$||G - G_1||_F^2 = \sigma_2^2 + \cdots + \sigma_l^2. \tag{2.9}$$

The best rank 1 approximation is unique if $\sigma_1 > \sigma_2$.
We now return to problem (2.6) and conclude that the solution must satisfy

$$\hat{b}\hat{k}^T = \sigma_1 w_1 v_1^T,$$

where we used the notation as in (2.7). This implies that $\hat{b}, \hat{k}$ are multiples of v_1, w_1. Write $\hat{b} = \alpha w_1$ and $\hat{k} = \beta v_1$, where $\alpha\beta = \sigma_1$ and $\alpha, \beta \in \mathbb{R}$. The question arises

whether we can always find α, β such that the constraints (2.3) are fulfilled.
Let $H := G^T G$. The columns of H sum up to zero, because the rows of G sum up to zero. This means that $\sum_i (Hu)_i = 0$ for every vector u. In particular this is true for every eigenvector corresponding to a nonzero eigenvalue, which we can write as $u = \frac{1}{\lambda} Hu$. Recall that the right singular vectors are equal to the normalized eigenvectors of H, thus the sum of the components of any right singular vector v_i is also zero. Thereby the second constraint of (2.3) is satisfied. The first constraint is satisfied by choosing α and β in the following way:

$$\alpha = \frac{1}{\sum_i (w_1)_i} \quad \text{and} \quad \beta = \sigma_1 \sum_i (w_1)_i.$$

Hence, the desired least squares solution is obtained by

$$\hat{b} = \frac{w_1}{\sum_i (w_1)_i}, \quad \hat{k} = \sigma_1 \sum_i (w_1)_i v_1. \tag{2.10}$$

So far the log of the observed death rates are minimized. Since

$$\sum_{x \in \mathcal{X}} \sum_{t \in \mathcal{T}} (\log m(x,t) - \log \hat{m}(x,t))^2 = \sum_{x \in \mathcal{X}} \sum_{t \in \mathcal{T}} \left(\log \left(\frac{m(x,t)}{\hat{m}(x,t)} \right) \right)^2,$$

where $\hat{m}(x,t)$ denotes the fitted central death rate, (the log of) the *quotient* $\frac{m(x,t)}{\hat{m}(x,t)}$ is minimized, regardless of the actual difference $|m(x,t) - \hat{m}(x,t)|$. In general this results into a better fit for ages where death rates are low. The $\hat{k}_t$ values are re-estimated in order to correct this distortion, while the other parameters are left the same. The new values of $\hat{k}_t$ are a solution of the following equation:

$$\sum_{x \in \mathcal{X}} d(x,t) = \sum_{x \in \mathcal{X}} n(x,t) \hat{m}(x,t) = \sum_{x \in \mathcal{X}} n(x,t) e^{\hat{a}_x + \hat{b}_x \hat{k}_t}, \quad \text{for all } t, \tag{2.11}$$

in which $d(x,t)$ and $n(x,t)$ are the realizations of $D(x,t)$ and $N(x,t)$ defined in (2.1). Solutions can be obtained using a numerical solver. The adjustment (2.11) has two advantages:

i) The adjustment avoids large differences between actual and fitted deaths. The re-estimation makes the total number of observed deaths in year t equal to the total number of fitted deaths. Ages with high death rates (for which $\hat{a}_x$ is high) receive more weight in (2.11), which should lead to a better fit for these ages.

ii) Age groups $n(x,t)$ which are large receive more weight in determining $\hat{k}_t$. This is favorable since a larger age group implies a larger sample.

2.3 The forecast

When the parameters have been determined, a forecast can be made. Since a_x and b_x are time-invariant, we will focus on k_t. Its values are interpreted as a time series which can be fit by an appropriate ARIMA model. We first consider a special class of ARIMA models: the ARMA models. These can be applied to any second order *stationary* time series $\{Y_t\}$, which means that there exist μ and γ such that

(i) $E[Y_t] = \mu$, independent of t.

(ii) $\text{Cov}(Y_t, Y_{t+h}) = \gamma(h)$, independent of t for each h.

Every ARMA model can be written in the following form:

$$Y_t - \phi_1 Y_{t-1} - \cdots - \phi_p Y_{t-p} = \mu + Z_t + \theta_1 Z_{t-1} + \cdots + \theta_q Z_{t-q},$$

where p, q give the order of the Auto-Regressive (AR) and the Moving-Average (MA) parts, respectively. Furthermore, the error terms Z_t are independent of each other and identically distributed. They will be called *white noise* innovations, with mean 0 and variance σ_Z^2, notated as $Z_t \sim \text{WN}(0, \sigma_Z^2)$.
When we observe a trend in a series of data, it might not be stationary. We can apply the differencing operator ∇ to attempt to make it stationary. This operator transforms the original sequence into a sequence of consecutive differences:

$$\nabla K_t = K_t - K_{t-1}. \tag{2.12}$$

When a time series $\{K_t\}$ has a linear trend, one time differencing will produce a stationary series $\{\nabla K_t\}$. The general idea of differencing is to apply the differencing operator repeatedly until the trend has disappeared. After that, one can try to fit an ARMA model to the differenced data. These two steps (differencing and ARMA fitting) together give us an ARIMA(p, d, q) model, where p and q are the same as before and d is the number of times the differencing operator is applied.

Lee and Carter mention that for U.S. data: "k declines roughly linearly from 1900-1989" and: "...short run fluctuations in k do not appear much greater in the first part of the period than they do in the second, with the exception of the influenza epidemic in 1918."([20], p. 662) The stable linear decline of k_t is the reason why their method is successful. The same pattern is observed by other countries of low mortality.
Lee and Carter fitted the estimated k_t by an ARIMA$(0, 1, 0)$ model, although it was found that an extra AR component made the model marginally superior. Hence, the model can be stated as

$$K_t = K_{t-1} + \mu + Z_t,$$

which is equivalent to a random walk with (linear) drift. Denote τ as the last year in which death rates are known. A forecast of h years in the future is calculated by

$$\begin{aligned} E[K_{\tau+h}] &= E[K_{\tau+h-1} + \mu + Z_{\tau+h}] = E[K_\tau + h\mu + Z_{\tau+1} + \cdots + Z_{\tau+h}] \\ &= k_\tau + h\mu. \end{aligned} \tag{2.13}$$

It was investigated by Lee and Carter how *parameter uncertainty*, i.e. the error that is made by fitting the paramaters, accounted for the total uncertainty of the forecast. Applying various methods, described in Appendix B of their article, they found that most of the uncertainty is captured by the model. However, "...taking account of parameter uncertainty increases the standard error of the forecast by less than 1% in the first year, by 6% after 10 years, by 25% after 50 years, and by 36% after 75 years."([20], p. 665)
The performance of the model was tested by using the data of the period 1900-1944

to forecast k_t for the *known* period 1945-1989, as well as the data from 1933-1962 to forecast 1963-1989. Both forecasts performed well, the actual data stayed well into a 95% prediction interval from the forecasts without including parameter uncertainty (the construction of prediction intervals will be discussed in Chapter 4). Another test is performed to test the stability of the model. By constructing forecasts for different base years, it was found that starting anywhere between 1930 and 1960 made little difference for the value of k_t in 2065. A lower value is found when starting at 1970 due to the rapid decline of death rates in the 1970's.

Chapter 3

Lee-Carter applied to the Netherlands

In this chapter we implement the model of Lee-Carter to mortality data from the Netherlands. In Section 3.1 we give a brief outline of the construction of Dutch death rates. The fitted rates and the estimated parameters $\hat{a}_x, \hat{b}_x$ and $\hat{k}_t$ are presented in Section 3.2. In Section 3.3 we derive which ARIMA time series model is most appropriate for the purpose of forecasting.

As far as we know the only one who studied future Dutch mortality probabilities using the Lee-Carter model is F. Gregorkiewicz [17], albeit his approach deviates from ours. In his research paper a number of steps is described to predict mortality probabilities. Sometimes these steps differ from the original Lee-Carter model. The most important one is that he applies the model to sex-specific death rates, while in the original paper a forecast is constructed for the sexes combined. In Section 3.4 we describe the difficulties when fitting Lee-Carter on sex specific rates and explain why we have chosen to follow the original approach of Lee-Carter.

3.1 Dutch mortality data

The CBS collects demographic information of the Dutch population from the municipalities. This data is freely available at StatLine, the data bank of the CBS; see [8]. A death rate can be constructed using the period-cohort approach (see Section 1.4 for the description of the different approaches). The number of registered deaths are categorized by age on December 31st given that this person would have been alive on that day. Population censuses are categorized by the age of a person on January 1st. We need to synchronize these data in order to compute a death rate. We therefore assume that people aged x on January 1st of year t, are born in year $t - x - 1$. At the time of writing, the following periods and cohorts were available:

$$t = 1950, \dots, 2006, \qquad x = 0, \dots, 98, 99+,$$

where t denotes calender year and x the age that a person will have at the end of the year given that he/she is alive. In this definition x is a *cohort age*, since all people have the same year of birth. The cohort age 99+ consists of all people that are aged 99 or higher (at the end of the year). We use the following definitions:

- $D(x, t)$: Number of deaths in year t of people that would have had age x at the end of t, $x \geq 0$.
- $N(x, t)$: Number of lives aged $x - 1$ at the beginning of year t, $x \geq 1$.
- $N(0, t)$: Number of newborns during year t.

The period-cohort death rate is obtained by

$$M(x, t) = \frac{D(x, t)}{N(x, t)}, \quad x \geq 0. \tag{3.1}$$

3.2 Quality of the fit

We are now ready to show some plots of the performance of the Lee-Carter model applied to Dutch mortality data. We use observed values $m(x, t)$, realizations of

$M(x,t)$ as was defined in (3.1), as input data and distinguish between ages $x \in \mathcal{X}$ (we shall use the word *age* to denote the *cohort age* x) and calender years $t \in \mathcal{T}$, where

$$\mathcal{X} = \{20, 21, \ldots, 98, 99+\}, \quad \mathcal{T} = \{1950, 1951, \ldots, 2006\}.$$

In Figure 3.1 the fit of the model for people aged 25, 45, 65 and 85 is displayed to give an idea of the accuracy. We see that most of the time the fit is good especially

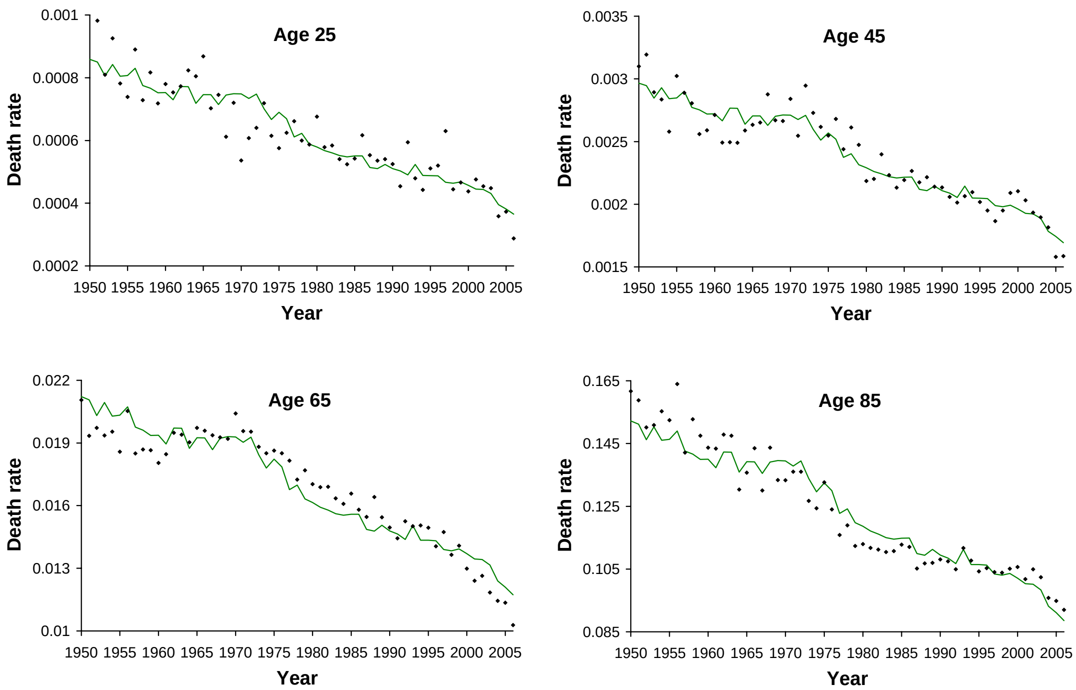


Figure 3.1: Death rates for selected ages. The observed rates (dots) and the fit by the Lee-Carter model (line).

for the last 30 years. The data is more volatile in the first half of the period than in the second half.

In Figure 3.2 the death rates are shown in a different perspective. Now we plotted the death rates of 1950 and 2006 for all ages and added the forecast for 2050. We see the best fit for old people except the very old ones, although this is partly due to the logarithmic scaling. For ages higher than 90 death rates converge and show a large variation because the number of people that reach these ages is limited.

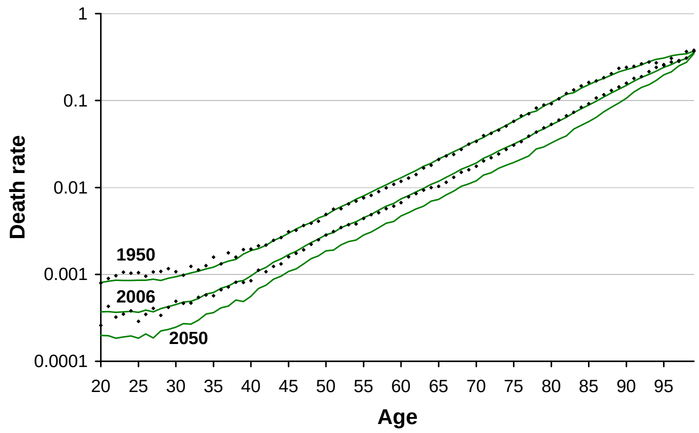


Figure 3.2: Death rates in 1950, 2006 and 2050 (forecast).

The parameters of the fitted model are shown in Figure 3.3. The first plot illustrates that average death rates per age show a monotonically increasing pattern from age 20. Looking at the trajectory of $\hat{b}_x$, we see that most improvement is realized for people

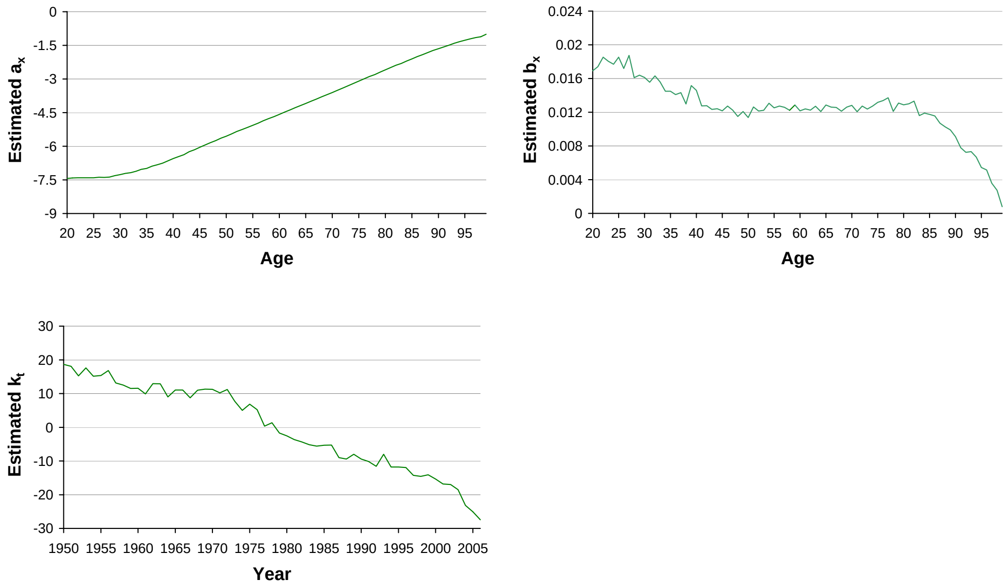


Figure 3.3: The estimated parameters $\hat{a}_x, \hat{b}_x$ (top) and $\hat{k}_t$ (bottom).

between 20 and 45. A sharp decrease is noticeable for ages beyond 85. This reinforces the earlier observation that death rates for the highest ages did hardly improve. This was also reflected in the prognosis for 2050, where we see that only a small improvement is expected. The time parameter $\hat{k}_t$ shows a decreasing trend which is close to linear just as in the American case. In the last three years an acceleration can be observed.

The measure R^2 is used to test how the model accounts for the variation of the data. Write the fitted rates as

$$\hat{m}(x,t) = e^{\hat{a}_x + \hat{b}_x \hat{k}_t},$$

then R^2 for $x \in \mathcal{X}$ can be defined as

$$R^2(x) = 1 - \frac{\sum_{t\in\mathcal{T}}(m(x,t) - \hat{m}(x,t))^2}{\sum_{t\in\mathcal{T}}(m(x,t) - \bar{m}(x))^2},$$

where $\bar{m}(x) = \frac{1}{n}\sum_{t\in\mathcal{T}} m(x,t)$.

The values of $R^2(x)$, plotted in Figure 3.4, give the proportion of the sample variance that is explained by the model sorted by age. For more than 88% of the age groups,

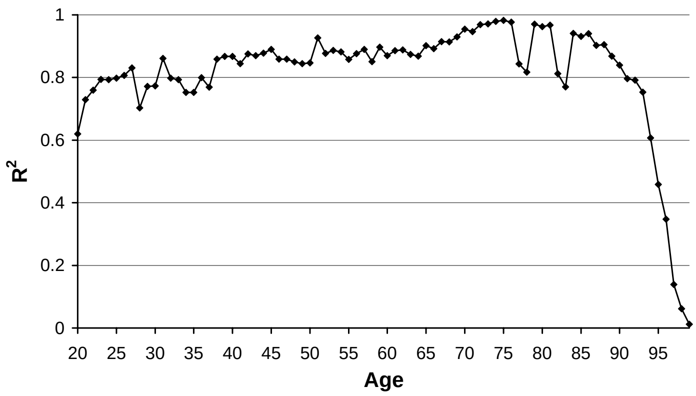


Figure 3.4: The fraction of explained variance per age group.

more than 75% of the variation is explained by the model. For 25% this is more than 90%. We see that the fit improves when x increases until age 75. An explanation is that young people have large variation in death rates because the total number of deaths is low. Another explanation is the re-estimation of k_t which gives more weight to ages with a high number of observed deaths. After age 75 the fit deteriorates. This can be explained by the fact that the number of people that reach these ages decreases quickly, which makes death rates more volatile (smaller sample). An overall measure for the goodness of fit method is obtained by calculating R^2 for all data

$$R^2 = 1 - \frac{\sum_{x\in\mathcal{X}}\sum_{t\in\mathcal{T}}(m(x,t)-\hat{m}(x,t))^2}{\sum_{x\in\mathcal{X}}\sum_{t\in\mathcal{T}}(m(x,t)-\bar{m}(x))^2}.$$

We find that 46.3% of the overall variation is explained by the model. When we leave out the ages higher than 88, 90.4% is explained.

We can obtain information about the errors by looking at the estimated errors $\hat{\varepsilon}_{x,t}$, which can be computed by

$$\hat{\varepsilon}_{x,t} = \log m(x,t) - \log \hat{m}(x,t).$$

We calculated the correlation matrix and observed a strong correlation across consecutive age groups. This is unfortunate since it violates the assumption that the errors are independent. Lee and Carter acknowledge this problem but argue that it is undesirable to add extra parameters into the model to describe this interdependence, because the effect would be marginal ([20], Appendix B).

3.3 Time series modeling

In this section we fit an ARIMA model to the estimated values $\hat{k}_t$ in order to forecast future values. Consider $\{\hat{k}_t\}$, $t = 1950, 1951, \ldots, 2006$, which were plotted in Figure 3.3. We can observe a decline that is approximately linear, so we apply the differencing operator ∇, defined in (2.12), to remove the trend. The resulting values are displayed in the top left chart of Figure 3.5. The horizontal line in the chart denotes the sample mean of the series.

We do not detect any trend nor a change in dispersion so we assume that the differenced values are stationary. The next step is to fit an ARMA(p,q) model to the sequence $\{\nabla\hat{k}_t\}$. Appropriate values for p and q can be derived from the Auto Correlation Function (ACF) $\rho(h)$ and the Partial Auto Correlation Function (PACF) $\alpha(h)$, $h \in \mathbb{N}$, which can be defined in terms of the Auto CoVariance Function (ACVF) $\gamma(h)$. Let $\{X_t\}$ be a time series, then $\gamma(h)$ is defined as

$$\gamma(h) = \mathrm{Cov}(X_t, X_{t+h}),$$

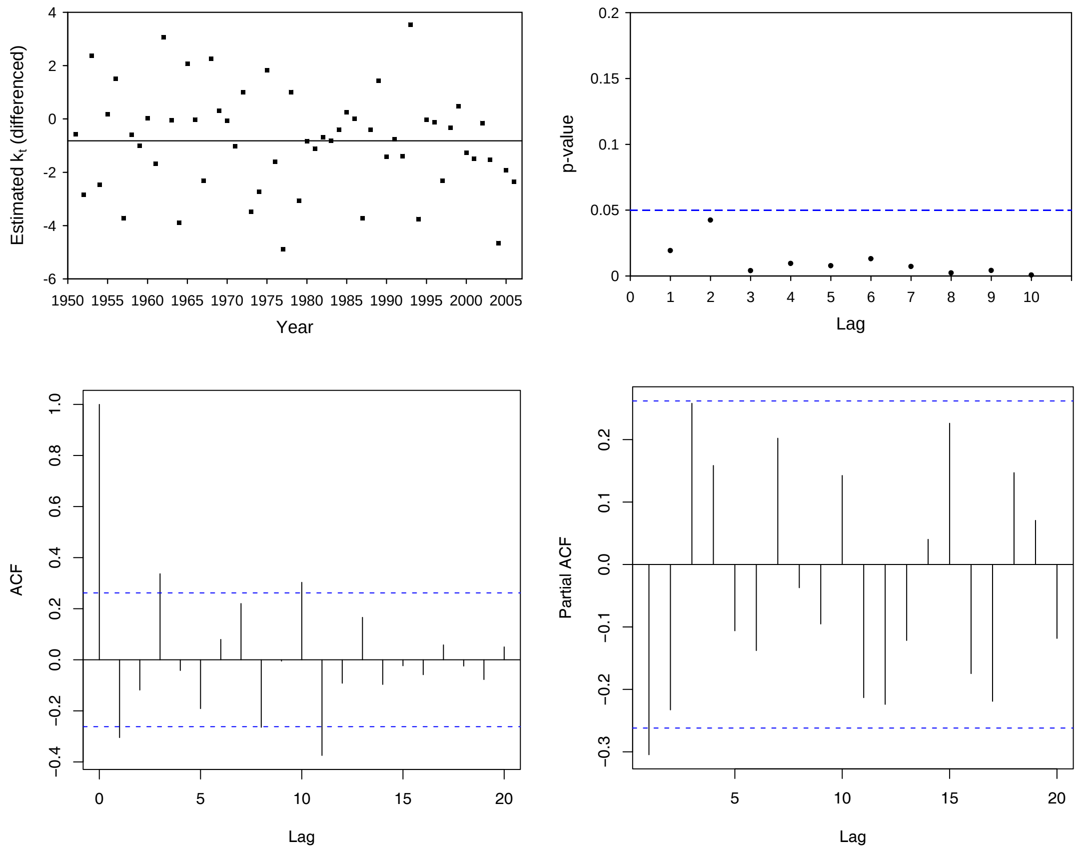


Figure 3.5: Examination of the sequence $\{\nabla \hat{k}_t\}$. **Top left**: the values (markers) around the sample mean. **Top right**: the Ljung-Box test statistic. **Bottom left**: the sample ACF. **Bottom right**: the sample PACF.

where h denotes the *lag*. The ACF and PACF are given by:

$$\rho(h) = \frac{\gamma(h)}{\gamma(0)} = \mathrm{Cor}(X_t, X_{t+h}), \quad h \geq 0, \tag{3.2}$$

$$\alpha(h) = (\Gamma_h^{-1}\gamma_h)_h, \quad h \geq 1, \tag{3.3}$$

where $(\Gamma_h^{-1}\gamma_h)_h$ is the the last element of the *vector* $\Gamma_h^{-1}\gamma_h$ which consists of

$$\Gamma_h = \begin{pmatrix} \gamma(0) & \gamma(1) & \dots & \gamma(h-1) \\ \gamma(1) & \gamma(0) & \dots & \gamma(h-2) \\ \vdots & \vdots & \ddots & \vdots \\ \gamma(h-1) & \gamma(h-2) & \dots & \gamma(0) \end{pmatrix} \quad \text{and} \quad \gamma_h = \begin{pmatrix} \gamma(1) \\ \gamma(2) \\ \vdots \\ \gamma(h) \end{pmatrix}.$$

We immediately conclude from (3.2) that $\rho(0) = 1$, the correlation of an observation with itself. For every $h > 0$, $\rho(h) = 0$ when the series X_t is uncorrelated. The ACF of a MA(q) process is 0 when $h > q$. The PACF is used to determine the number of AR components. For an AR(p) process, $\alpha(h) = 0$ for $h > p$. For applications, the ACF and PACF are usually unknown, as in our case, but they can be estimated from the observations.

The sample ACF $\hat{\rho}(h)$ and PACF $\hat{\alpha}(h)$ are determined in the same way as $\rho(h)$ and $\alpha(h)$ (in terms of the ACVF), but by replacing $\gamma(h)$ by $\hat{\gamma}(h)$, the sample ACVF, in (3.2) and (3.3) respectively. For a set of observations $x_1, x_2, \ldots, x_n$, $\hat{\gamma}(h)$ is defined as

$$\begin{aligned}\hat{\gamma}(h) &= \frac{1}{n}\sum_{t=1}^{n-h}(x_{t+h}-\bar{x})(x_t-\bar{x}),\ h<n, \quad \text{in which} \\ \bar{x} &= \frac{1}{n}\sum_{t=1}^{n} x_t.\end{aligned}$$

This estimator is biased even when we replace the factor $\frac{1}{n}$ by $\frac{1}{n-h}$. In order to draw conclusions from $\hat{\rho}(h)$ and $\hat{\alpha}(h)$ we need to make assumptions about its distribution. For a large sample size n, it can be shown that when $\rho(h) = 0$, $\hat{\rho}(h)$ is approximately normal distributed with $\mathrm{E}[\hat{\rho}(h)] = 0$ and $\mathrm{Var}(\hat{\rho}(h)) = 1/n$. Also, when $\alpha(h) = 0$, $\hat{\alpha}(h)$ is approximately normal distributed with $\mathrm{E}[\hat{\alpha}(h)] = 0$ and $\mathrm{Var}(\hat{\alpha}(h)) = 1/n$. A proof of these statements can be found in [4] p. 117. When we assume that our sample size is large enough (which is the case when we follow Brockwell & Davis, who suggest from $n = 50$ this assumption is reasonable, see [5] p. 60) we can construct confidence bounds. The probability that $\hat{\rho}(h), \hat{\alpha}(h)$ fall between the bounds $\pm\Phi_{1-\alpha/2}/\sqrt{n}$, where $\Phi_{1-\alpha/2}$ denotes the $1-\alpha/2$ quantile of the normal distribution, is $(1-\alpha)$.

We used the statistical program 'R' to determine $\hat{\rho}(h)$ and $\hat{\alpha}(h)$ for our differenced series $\{\nabla\hat{k}_t\}$ and plotted its values in the bottom charts of Figure 3.5, where the blue dotted lines denote the 95% confidence bounds. Hence, we expect approximately 95% of the values $\hat{\rho}(1), \hat{\rho}(2), \ldots \hat{\rho}(20)$ to fall between these bounds. Besides the first lag,

we see that five lags fall outside the bounds which suggests that there is correlation between the observations. In the chart of the sample PACF we see that only one value falls outside the bounds, which is the expected number for an uncorrelated series.
We can perform a statistical test whether $\{\nabla \hat{k}_t\}$ is generated by an independent and identically distributed (IID) sequence of random variables. If that is the case, an ARIMA$(0,1,0)$ is the most appropriate model. The null hypothesis is defined as:

$$H_0 : \ \{\nabla \hat{k}_t\} \text{ is generated by an IID sequence of random variables.} \tag{3.4}$$

Under the null hypothesis, $\rho(h) = 0$ for all h, so that the test statistic

$$\tilde{Q} = n \sum_{j=1}^{h} \hat{\rho}^2(j), \tag{3.5}$$

is approximately distributed as a chi-square distribution with h degrees of freedom $\chi^2(h)$. Here we test whether the first h autocorrelations are different from zero, so for large values of $\tilde{Q}$ this hypothesis will be rejected. Ljung and Box show in [22] that for finite sample size n the distribution of

$$Q = n(n+2) \sum_{h=1}^{j} \frac{\hat{\rho}^2(h)}{n-h},$$

is a closer approximation of $\chi^2(h)$ then $\tilde{Q}$ from (3.5). When a Ljung-Box is test performed, H_0 is rejected when $Q > \chi^2_{1-\alpha}(h)$, where $\chi^2_{1-\alpha}$ is the $1-\alpha$ quantile of the chi-square distribution with h degrees of freedom. We chose a significance level of $\alpha = 0.05$ and rejected the null hypothesis for lags $h = 1, \ldots, 10$. Hence, we conclude that the differenced data are correlated. The p-values are displayed in the top right chart of Figure 3.5.
We can use R to calculate the Akaike's information criterion (AIC) for several values of p and q, which are displayed in Table 3.1. The AIC balances the likelihood of the fit with the number of parameters. On the one hand we want a close fit with the observed data and adding parameters will in general improve this fit. On the other hand, we do not want an *overdetermined* model, because the series will be extrapolated to

Table 3.1: the AIC criterion for $p, q = 0, 1, \ldots, 4$

p/q	**0**	**1**	**2**	**3**	**4**
0	233.31	228.83	230.10	227.46	226.91
1	229.88	230.72	231.90	228.13	225.05
2	228.61	229.11	**222.2**	226.22	226.59
3	226.66	227.77	224.08	226.18	227.55
4	226.96	228.85	226.03	227.7	228.65

construct a forecast. Incorporating too many parameters will replicate noise from the past into the future, which can cause idiosyncratic features to appear in the forecast. The optimal AIC is given by the value closest to zero. The lowest AIC for our data is obtained when $(p, q) = (2, 2)$. This gives us an ARIMA$(2, 1, 2)$ model

$$K_t - K_{t-1} = \mu + Y_t, \tag{3.6}$$

where Y_t is a zero mean ARMA$(2, 2)$ model,

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \theta_1 Z_{t-1} + \theta_2 Z_{t-2} + Z_t, \quad Z_t \sim WN(0, \sigma_Z^2). \tag{3.7}$$

In Table 3.2 we show the estimates and the standard errors for the mean μ, the AR components ϕ_1, ϕ_2 and the MA components θ_1, θ_2 for our series.

Table 3.2: Estimates and standard erors for the coefficients of the ARIMA$(2, 1, 2)$ model

	μ	ϕ_1	ϕ_2	θ_1	θ_2
Estimate	-0.8291	-0.5239	-0.8836	0.2499	0.8754
se	0.1837	0.0868	0.0770	0.1007	0.1054

We plotted the fitted innovations $\hat{z}_t$, which result from plugging in the observed values $\hat{k}_t$ and the estimates from Table 3.2 into (3.6) and (3.7). We tested whether there is dependence between the innovations $\hat{z}_t$ in the same way as we did for $\nabla \hat{k}_t$ in Figure 3.5, and displayed them in Figure 3.6.

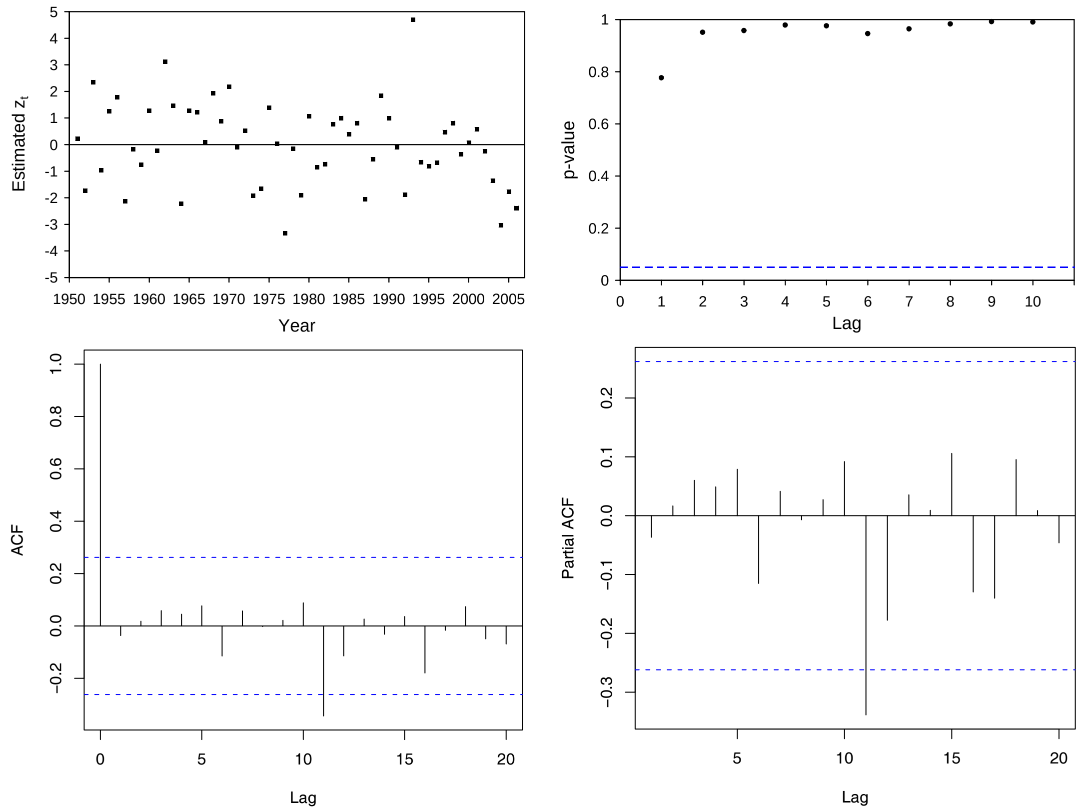
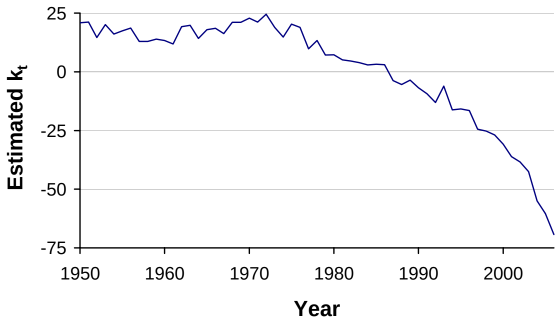


Figure 3.6: Examination of the residuals $\hat{z}_{1951}, \hat{z}_{1952}, \ldots, \hat{z}_{2006}$ after an ARIMA$(2, 1, 2)$ model is implemented. **Top left**: the values. **Top right**: the Ljung-Box test statistic. **Bottom left**: the sample ACF. **Bottom right**: the sample PACF.

In both the sample ACF and PACF we see that only one lag falls outside the dotted lines, which is the expected number for an uncorrelated series. The Ljung-box test gives us p-values which are all well over 0.05, so we cannot reject the hypothesis that $\hat{z}_t$ is the realization of an IID white noise sequence.

3.4 Fitting Lee-Carter on sex-specific death rates

One could imagine to apply the Lee-Carter model separately on the death rates of males and females to construct a sex-specific forecast, as Gregorkiewicz has done in 2006. However, Lee and Carter mention that the key to the success of their model is the steady, linear decline of k_t, and for the sex-specific approach this does not hold (in the Netherlands), as will be illustrated in this section. Fitting Lee-Carter on death rates from the period 1950-2006 and ages $0, \ldots, 98, 99+$ leads to $\hat{k}_t$ for males and females that are displayed in Figure 3.7.



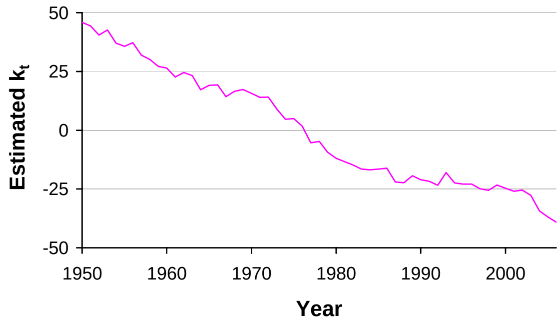


Figure 3.7: The fitted parameters $\hat{k}_t$ for males (left) and females (right).

We observe that the trend for males is different than for females. A linear trend for the female $\hat{k}_t$ seems reasonable but this is not the case for the males, which causes the difficulties. Gregorkiewicz also came to this conclusion, and decided to compensate for the non-linearity by leaving out a part of the observed period. For males he uses observed death rates from 1971 to fit the model, for females the years from 1980 are used. Because of this, estimated drift terms that equal $\hat{\mu} = -1.68$ for males and $\hat{\mu} = -0.98$ for females are found. The reason the drift for males is more than 1.5 times as large, lies in the development of mortality since 1950, as can be observed from Figure 3.7. A strong improvement for females can be seen until approximately 1980, followed by a slower improvement until 2006. For males, this development was the other way around. The question arises whether it is reasonable to expect that male death rates will continue to decrease at a (much) faster pace than the female death rates. We also tried several ways to extrapolate the male trajectory, but none of them gave a satisfactory result. Let us present our experiences.

For our purposes the ages on which people do not accrue pension are not important. The death rates of these ages influence the value of $\hat{k}_t$, so leaving them out might alter $\hat{k}_t$ in a favorable way. Unfortunately this was not the case, when we left out the ages 0-19 the trajectory became even worse as can be seen in Figure 3.8. We also fitted the model on a shorter period. When the model is fitted only on 1975-2006 the male trajectory slightly improved, albeit the female trajectory deteriorated, see the right chart of Figure 3.8. We experimented with different periods and ages, which

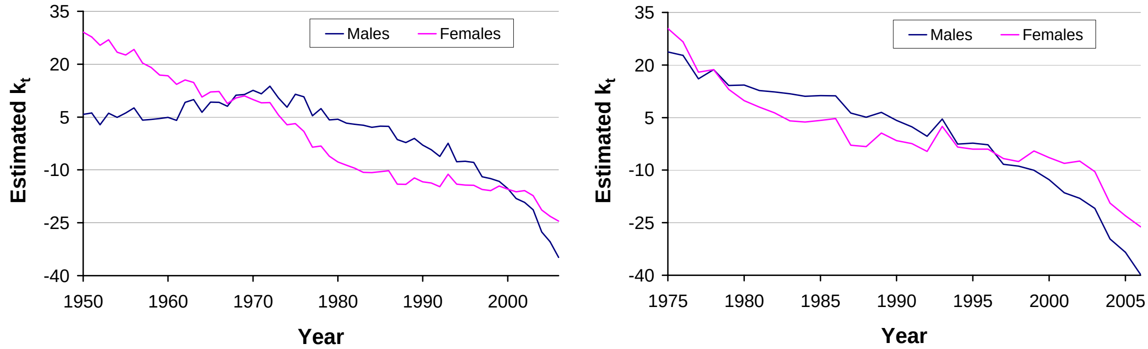


Figure 3.8: The fitted model where we left out ages 0-19 (left) or the period 1950-1974 (right).

leaded to the same conclusions.

We approximated the male trajectory by polynomials of low order to analyze its trend. In Figure 3.9 the first and second order approximation in the least squares sense are displayed. We see that the linear approximation is poor while the quadratic

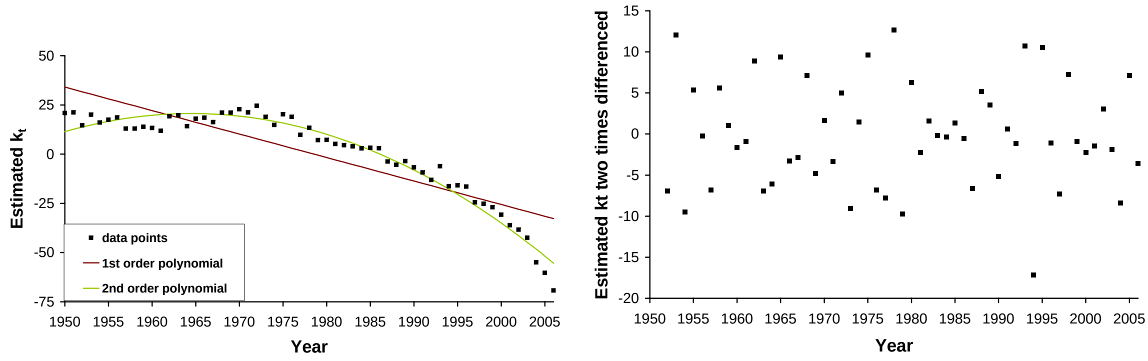


Figure 3.9: Trend approximation by a first and second order polynomial (left) and the values $\nabla^2 \hat{k}_t$ (right).

approximation is fairly good. Also displayed in Figure 3.9 are $\nabla^2 \hat{k}_t$, the values after differencing twice. We cannot observe a trend in this data, so it is reasonable to assume that they are stationary. This suggests that the $\hat{k}_t$ behave quadratically and hence we examined fitting an ARIMA$(p, 2, q)$ model. The lowest AIC value is attained when $(p, q) = (3, 0)$. In Figure 3.10 we have displayed forecasted $\hat{k}_t$ according to ARIMA$(3, 2, 0)$ together with a forecast for the female $\hat{k}_t$, for which we found that an ARIMA$(0, 1, 3)$ is the most appropriate model. The forecasts quickly

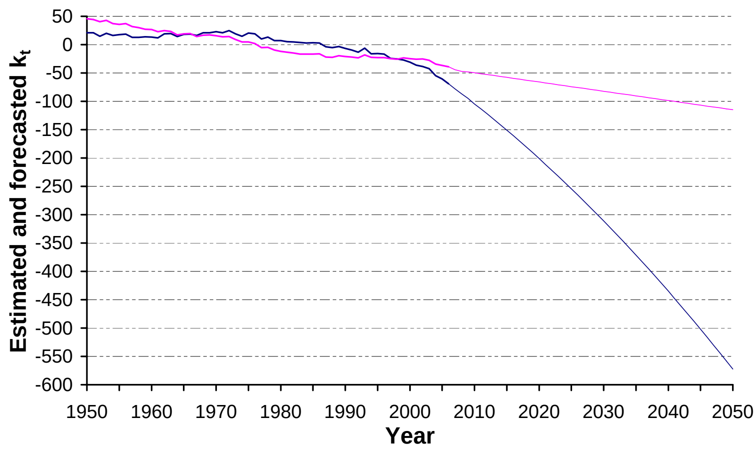


Figure 3.10: Sex specific forecasts for $\hat{k}_t$. For males an ARIMA$(3, 2, 0)$ model is used, for females an ARIMA$(0, 1, 3)$ model.

diverge. In 2050 most forecasted death rates for males are lower than for females. We compared these outcomes with the forecast of the unisex model, which will be presented in the next chapter, and concluded that the quadratic trend produces a forecast which is unlikely to happen.

Chapter 4

A forecast of Dutch mortality

In the previous chapter we implemented the Lee-Carter model on Dutch mortality rates. The purpose of this chapter is to investigate how we, given this model, should construct forecasts. Besides a forecast which is most likely to happen, we are also interested in the volatility of future death rates. In Section 4.1 we shall discuss several methods to determine a *prediction interval* for a given level α, $0 < \alpha < 1$, for which the probability that the realization will lie within the prediction interval is $(1 - \alpha)$. In Section 4.2 we test the model to see if it is suitable for forecasting. We fit the model for various periods in the past and compare predicted mortality with observed mortality.

4.1 Prediction intervals

In Section 3.3 we found that an ARIMA$(2, 1, 2)$ is most appropriate, i.e. the smallest AIC value, to model k_t. The fitted innovations $\hat{z}_t$ seem to be uncorrelated, based on the sample ACF, the sample PACF and the Ljung-Box test. From now we will adopt this model as the *true* model. Within the model, there are two aspects that have to be further explored:

i) The distribution of the innovations Z_t.

ii) The effect of parameter uncertainty.

In this section we will examine different ways to construct forecasted values $k_{\tau+h}$, where τ is the final year on which the model is fitted (in our case $\tau = 2006$). At first we ignore the effect of parameter uncertainty. The easiest way is to assume that Z_t is normally distributed with variance equal to the squared standard error $\hat{\sigma}_Z^2$. We fitted our model in R and found

$$\hat{\sigma}_Z \approx 1.571. \tag{4.1}$$

In Figure 3.6 the estimated innovations $\hat{z}_t$ were shown. A corresponding kernel density estimator with kernel equal to the standard normal density is displayed in Figure 4.1. Also plotted with dashed lines is the normal density with mean 0 and variance $\hat{\sigma}_Z^2$.

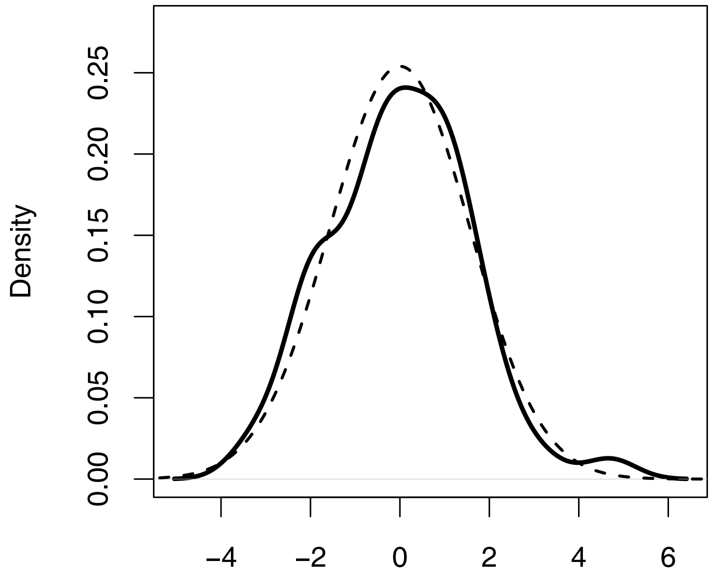


Figure 4.1: A kernel density estimator on $\hat{z}_t$ (solid line) compared with a normal density with mean 0 and variance $\hat{\sigma}_Z^2$ (dashed line).

The graphs are fairly close, so by the first generated forecast, in Subsection 4.1.1, we shall assume that the innovations are distributed as $Z_t \sim \mathcal{N}(0, \hat{\sigma}_Z^2)$. A forecast with minimum mean squared prediction error (MSPE) based on n observations, will be derived. Moreover, we will show how a prediction interval can be determined for a given level α. A prediction interval gives an indication of the volatility of the forecast, comparable to a confidence interval when an unknown parameter is estimated based on a sample. In Subsection 4.1.2 we will perform a simulation where instead of making an assumption on the distribution of Z_t, a bootstrap method is used. Finally, in Subsection 4.1.3 the effect of parameter uncertainty will be investigated by doing another simulation. Instead of generating future forecasts with a fixed drift term $\hat{\mu}$, at every iteration the drift will be drawn randomly as $\mu^* \sim \mathcal{N}(\hat{\mu}, \hat{\sigma}_\mu^2)$.

4.1.1 Analytical forecast

In this subsection we consider a special class of ARMA(p,q) models: the *causal* and *invertible* ARMA models with zero mean. At every ARMA process $\{X_t\}$ with nonzero mean μ we can apply a mean correction: $Y_t = X_t - \mu$ to obtain a nonzero process. Causality and invertibility can be checked from the parameters $\phi_1, \dots, \phi_q$, $\theta_1, \dots, \theta_p$. The ARMA$(2,2)$ process that was fitted on the $\nabla \hat{k}_t$ in Section 3.3 is a member of this class.

When $\{Y_t\}$ is a zero mean, causal and invertible ARMA(p,q) process with corresponding innovations $\{Z_t\}$, there exists sequences $\{\psi_j\}, \{\pi_j\}$, where $\sum_{j=0}^{\infty} |\psi_j| < \infty$ and $\sum_{j=0}^{\infty} |\pi_j| < \infty$, such that

$$Y_t = \sum_{j=0}^{\infty} \psi_j Z_{t-j} \quad \text{and} \tag{4.2}$$

$$Z_t = \sum_{j=0}^{\infty} \pi_j Y_{t-j}, \tag{4.3}$$

for all t. The sequence $\{\psi_j\}$ can be found recursively from

$$\psi_j - \sum_{k=1}^{p} \phi_k \psi_{j-k} = \theta_j, \quad \text{for } j = 0, 1, \dots, \tag{4.4}$$

where $\theta_0 = 1$, $\theta_j = 0$ for $j > q$ and $\psi_j = 0$ for $j < 0$. Similarly $\{\pi_j\}$ is found by

$$\pi_j + \sum_{k=1}^{q} \theta_k \pi_{j-k} = -\phi_j, \quad \text{for } j = 0, 1, \dots, \tag{4.5}$$

where $\phi_0 = -1$, $\phi_j = 0$ for $j > p$ and $\pi_j = 0$ for $j < 0$.

We write ${}_nP_\tau Y_{\tau+h}$ for a prediction of h years ahead based on a linear combination of $Y_{\tau-n+1}, Y_{\tau-n+2}, \ldots, Y_\tau$:

$$ {}_nP_\tau Y_{\tau+h} = a_0 Y_\tau + \cdots + a_{n-1} Y_{\tau-n+1}, $$

with minimum MSPE,

$$ E[(Y_{\tau+h} - {}_nP_\tau Y_{\tau+h})^2] = E[(Y_{\tau+h} - a_0 Y_\tau - \cdots - a_{n-1} Y_{\tau-n+1})^2]. $$

Note that from the way that ${}_nP_\tau Y_{\tau+h}$ is constructed, it follows that

$$ E[Y_{\tau+h} - {}_nP_\tau Y_{\tau+h}] = 0. $$

The MSPE can be interpreted as a quadratic function in $a_0, \ldots, a_{n-1}$ which is bounded below by zero. We can find the minimum value by solving

$$ \frac{\partial E[(Y_{\tau+h} - {}_nP_\tau Y_{\tau+h})^2]}{\partial a_j} = 0, \quad j = 0, \ldots, n-1. \tag{4.6} $$

By changing the order of (partial) differentiation and expected value, (4.6) can be reduced to:

$$ E\left[(Y_{\tau+h} - \sum_{i=0}^{n-1} a_i Y_{\tau-i}) Y_{\tau-j}\right] = 0, \quad j = 0, \ldots, n-1. \tag{4.7} $$

For some ARMA processes it is possible to derive the linear predictor with minimum MSPE directly from (4.7). For instance, the one-step predictor ${}_nP_\tau Y_{\tau+1}$, when $\{Y_t\}$ is an AR(p) process with $n > p$,

$$ Y_t = \phi_1 Y_{t-1} + \cdots + \phi_p Y_{t-p} + Z_t. \tag{4.8} $$

When we take $a_i = \phi_{i+1}$ for $i = 0, \ldots, p-1$ and $a_i = 0$ for $i = p, \ldots, n-1$, (4.7) becomes

$$ E\left[(Y_{\tau+1} - \sum_{i=0}^{p-1} \phi_{i+1} Y_{\tau-i}) Y_{\tau-j}\right] = E[Z_{\tau+1} Y_{\tau-j}] = 0, \quad j = 0, \ldots, n-1, $$

since $Z_{\tau+1}$ is uncorrelated with $Y_{\tau+1-n}, \dots, Y_\tau$. The corresponding MSPE reads

$$E[(Y_{\tau+1} - {}_nP_\tau Y_{\tau+1})^2] = E[Z_{\tau+1}^2] = \sigma_Z^2.$$

Now consider the case that $h = 2$. We can write

$$\begin{aligned} Y_{\tau+2} &= \phi_1 Y_{\tau+1} + \cdots + \phi_p Y_{\tau+2-p} + Z_{\tau+2} \qquad (4.9)\\ &= \phi_1(\phi_1 Y_\tau + \cdots + \phi_p Y_{\tau+1-p} + Z_{\tau+1}) + \phi_2 Y_\tau + \cdots + \phi_p Y_{\tau+2-p} + Z_{\tau+2} \\ &= (\phi_1^2 + \phi_2) Y_\tau + \cdots + (\phi_1\phi_{p-1} + \phi_p) Y_{\tau+2-p} + \phi_1\phi_p Y_{\tau+1-p} + \phi_1 Z_{\tau+1} + Z_{\tau+2}. \end{aligned}$$

When we put $a_0, \dots, a_{p-1}$ equal to the coefficients of $Y_\tau, \dots, Y_{\tau+1-p}$ from the last line of (4.9) and put $a_i = 0$ for $i = p, \dots, n-1$, we obtain

$$\begin{aligned} E\left[(Y_{\tau+2} - \sum_{i=0}^{p-1} a_i Y_{\tau-i}) Y_{\tau-j}\right] &= E[(Z_{\tau+2} + \phi_1 Z_{\tau+1}) Y_{\tau-j}] \\ &= 0, \quad j = 0, \dots, n-1. \end{aligned}$$

The corresponding MSPE reads

$$E[(Y_{\tau+2} - {}_nP_\tau Y_{\tau+2})^2] = E[(Z_{\tau+2} + \phi_1 Z_{\tau+1})^2] = (1 + \phi_1^2)\sigma_Z^2.$$

The coefficients for $Z_{\tau+1}, Z_{\tau+2}$ that are obtained in (4.9) can also be found from two iterations of (4.4) (for an ARMA(2, 0) process, $\psi_0 = 1$ and $\psi_1 = \phi_1$). Therefore, we can find ${}_nP_\tau Y_{\tau+h}$ for any $h \geq 1$ as follows. First determine the coefficients $a_0, \dots, a_{p-1}$, $\psi_0, \dots, \psi_{h-1}$ to write $Y_{\tau+h}$ as

$$Y_{\tau+h} = a_0 Y_\tau + \cdots + a_{p-1} Y_{\tau+1-p} + \psi_0 Z_{\tau+h} + \cdots + \psi_{h-1} Z_{\tau+1}.$$

It follows that

$$\begin{aligned} E\left[(Y_{\tau+h} - \sum_{i=0}^{p-1} a_i Y_{\tau-i}) Y_{\tau-j}\right] &= E\left[\left(\sum_{j=0}^{h-1} \psi_j Z_{\tau+h-j}\right) Y_{\tau-j}\right] \\ &= 0, \quad j = 0, \dots, n-1. \end{aligned}$$

Then the MSPE reads

$$E[(Y_{\tau+h} - {}_nP_\tau Y_{\tau+h})^2] = E\left[\left(\sum_{j=0}^{h-1} \psi_j Z_{\tau+h-j}\right)^2\right] = \sigma_Z^2 \sum_{j=0}^{h-1} \psi_j^2 := \sigma_Z^2(h). \qquad (4.10)$$

When we assume that $Z_t \sim \mathcal{N}(0, \sigma_Z^2)$,

$$Y_{\tau+h} - {}_nP_\tau Y_{\tau+h} \sim \mathcal{N}(0, \sigma_Z^2(h)). \qquad (4.11)$$

The probability that $Y_{\tau+h}$ will fall between the bounds ${}_nP_\tau Y_{\tau+h} \pm \Phi_{1-\alpha/2}\sigma_Z^2(h)$, where $\Phi_{1-\alpha/2}$ denotes the $1-\alpha/2$ quantile of the normal distribution, is $(1-\alpha)$. Therefore we will call

$$[{}_nP_\tau Y_{\tau+h} - \Phi_{1-\alpha/2}\sigma_Z(h), {}_nP_\tau Y_{\tau+h} + \Phi_{1-\alpha/2}\sigma_Z(h)], \qquad (4.12)$$

a $(1-\alpha)$ *prediction interval.*

When $\{Y_t\}$ is an ARMA(p,q) process with $q \neq 0$, constructing a prediction interval is more complicated. This can be illustrated by determining the one step predictor for a MA(1) process,

$$Y_t = \theta_1 Z_{t-1} + Z_t.$$

As usual, we need to find $a_0, \ldots, a_{n-1}$ such that

$$\begin{aligned} E\left[(Y_{\tau+1} - \sum_{i=0}^{n-1} a_i Y_{\tau-i})Y_{\tau-j}\right] &= E\left[(\theta_1 Z_\tau + Z_{\tau+1} - \sum_{i=0}^{n-1} a_i Y_{\tau-i})Y_{\tau-j}\right] \\ &= E\left[(\theta_1 Z_\tau - \sum_{i=0}^{n-1} a_i Y_{\tau-i})Y_{\tau-j}\right] \\ &= 0, \quad j = 0, \ldots, n-1. \end{aligned}$$

The problem is that we cannot construct Z_τ by a finite number of observations $Y_{\tau-n+1}, \ldots, Y_\tau$. We can interpret Y_τ as an infinite sum of past Z_t, $t \leq \tau$ with coefficients given by (4.4). Therefore we write

$$\theta_1 Z_\tau - \sum_{i=0}^{n-1} a_i Y_{\tau-i} = \sum_{j=0}^{\infty} \tilde{\psi}_j Z_{\tau-j} := R_n$$

where $\tilde{\psi}_0, \tilde{\psi}_1, \ldots$ depend on $a_0, \ldots, a_{n-1}$.

We state the prediction error as

$$Y_{\tau+1} - {}_nP_\tau Y_{\tau+1} = Z_{\tau+1} + (\theta_1 Z_\tau - a_1 Y_\tau - \cdots - a_n Y_{\tau+1-n}) = Z_{\tau+1} + R_n.$$

When we define $a_i = \theta_1 \pi_i$, $i = 1, 2, \ldots$, with π_i given by (4.5),

$$\lim_{n\to\infty} R_n = 0. \tag{4.13}$$

This suggests that the MSPE converges to $\sigma_Z^2(1)$:

$$\lim_{n\to\infty} E[(Y_{\tau+1} - {}_nP_\tau Y_{\tau+1})^2] = \sigma_Z^2(1). \tag{4.14}$$

For details about this limit we refer to the textbook written by Brockwell & Davis [5], in particular Section 2.5.

For the general case, let $\{Y_t\}$ be any zero mean, causal and invertible ARMA(p,q) process. Using (4.2) we can write the prediction error of h steps ahead as

$$\begin{aligned} Y_{\tau+h} - {}_nP_\tau Y_{\tau+h} &= \sum_{j=0}^{h-1} \psi_j Z_{\tau+h-j} + \left(\sum_{j=h}^{\infty} \psi_j Z_{\tau+h-j} - {}_nP_\tau Y_{\tau+h} \right) \\ &:= \sum_{j=0}^{h-1} \psi_j Z_{\tau+h-j} + \sum_{j=0}^{\infty} \tilde{\psi}_j^2 Z_{\tau-j}. \\ &:= \sum_{j=0}^{h-1} \psi_j Z_{\tau+h-j} + R_n. \end{aligned}$$

The corresponding MSPE can be written as

$$E[(Y_{\tau+h} - {}_nP_\tau Y_{\tau+h})^2] = \sigma_Z^2 \left(\sum_{j=0}^{h-1} \psi_j^2 + \sum_{j=0}^{\infty} \tilde{\psi}_j^2 \right) := {}_n\sigma_Z^2(h). \tag{4.15}$$

For large n, we expect $R_n \approx 0$ and hence ${}_n\sigma_Z^2(h) \approx \sigma_Z^2(h)$.

There exists algorithms that determine ${}_nP_\tau Y_{\tau+h}$ with corresponding MSPE from a set of observations and a given model, such as the *Innovations Algorithm* ([5], p. 71). It follows from (4.15) that when $Z_t \sim \mathcal{N}(0, \sigma_Z^2)$,

$$Y_{\tau+h} - {}_nP_\tau Y_{\tau+h} \sim \mathcal{N}(0, {}_n\sigma_Z^2(h)).$$

A prediction interval can be obtained via (4.12).

We used the function *predict* from R on the fitted values of our process (3.7) to construct ${}_nP_\tau y_{\tau+h}$, for $n = 56$, $\tau = 2006$ and $h = 1, \ldots, 44$. Using the recursion from (3.6) we computed the corresponding forecast $k_{\tau+h}$, which is displayed in Figure 4.2 with a 95% prediction interval, where $\hat{\sigma}_Z$ from (4.1) is used as an estimate for σ_Z.

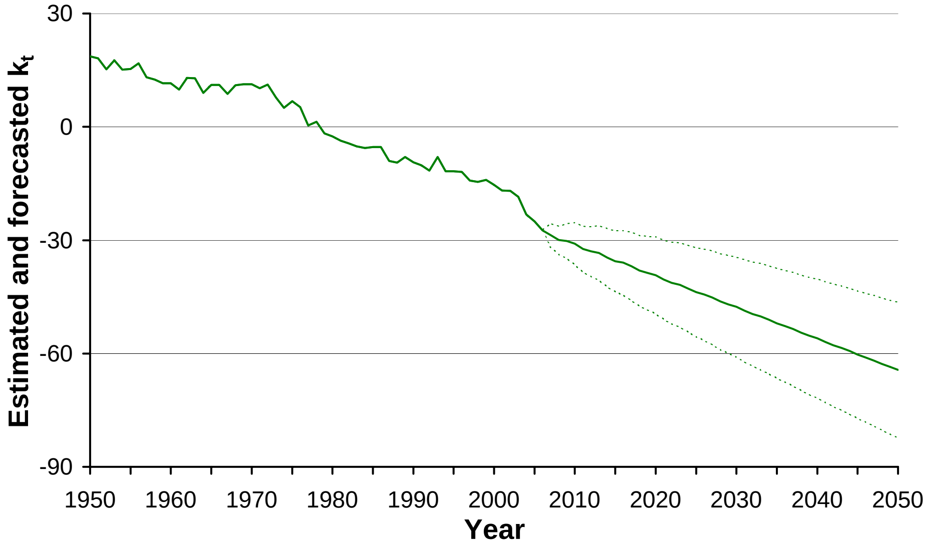


Figure 4.2: Forecasted k_t for the period 2007-2050 with a 95% prediction interval (dotted lines).

4.1.2 Simulation without parameter uncertainty

In this subsection we describe a simulation in which the error that was made by fitting the parameters is ignored and where innovations are generated by a *bootstrap* method. The objective is to examine how much the resulting forecast deviates from the forecast that was found in the previous subsection, where it was assumed that $Z_t \sim \mathcal{N}(0, \hat{\sigma}_Z^2)$. By the simulation, future innovations z_t^* are generated randomly from the observed innovations $\hat{z}_t$. In our case this means that we define the set

$$\mathcal{Z} = \{\hat{z}_{1951}, \hat{z}_{1952}, \ldots \hat{z}_{2006}\},$$

and draw z_t^* independently from $\mathcal{Z}$. Sequences $y_{\tau+1}^*, y_{\tau+2}^*, \ldots, y_{\tau+h}^*$ are constructed with the usual recursion formula of ARMA models. We need the fitted values $\hat{y}_{\tau-p+1}, \ldots, \hat{y}_\tau$ and $\hat{z}_{\tau-n+1}, \ldots, \hat{z}_\tau$ to start the process. In order for the simulation to make sense we require $n \geq \max(p, q)$. In Table 4.1 the construction of a single sequence is illustrated. From the N generated sequences an empirical distribution can be constructed.

Table 4.1: Generating one sequence $k^*_{\tau+1}, \ldots, k^*_{\tau+h}$.

Given estimates: $\hat{\phi}_1, \ldots, \hat{\phi}_p, \hat{\theta}_1, \ldots, \hat{\theta}_q, \hat{\mu}$
and fitted values $\hat{k}_\tau, \hat{y}_{\tau-p+1}, \ldots, \hat{y}_\tau, \hat{z}_{\tau+1-n}, \ldots, \hat{z}_\tau$, $n \geq \max(p, q)$.
Define $\mathcal{Z} = \{\hat{z}_{\tau+1-n}, \ldots \hat{z}_\tau\}$

for $i = 1 : h$ do
$y^*_{\tau+i} = \hat{\phi}_1 \tilde{y}_{\tau+i-1} + \cdots + \hat{\phi}_p \tilde{y}_{\tau+i-p} + \hat{\theta}_1 \tilde{z}_{\tau+i-1} + \cdots + \hat{\theta}_q \tilde{z}_{\tau+i-q} + z^*_{\tau+i}.$
$k^*_{\tau+i} = k^*_{\tau+i-1} + \hat{\mu} + y^*_{\tau+i}.$
end
where

$$\tilde{y}_t = \begin{cases} \hat{y}_t & \text{if } t \leq \tau; \\ y^*_t & \text{if } t > \tau; \end{cases} \qquad \tilde{z}_t = \begin{cases} \hat{z}_t & \text{if } t \leq \tau; \\ z^*_t & \text{if } t > \tau; \end{cases}$$

$k^*_\tau = \hat{k}_\tau$ and z^*_t independent drawings from $\mathcal{Z}$.

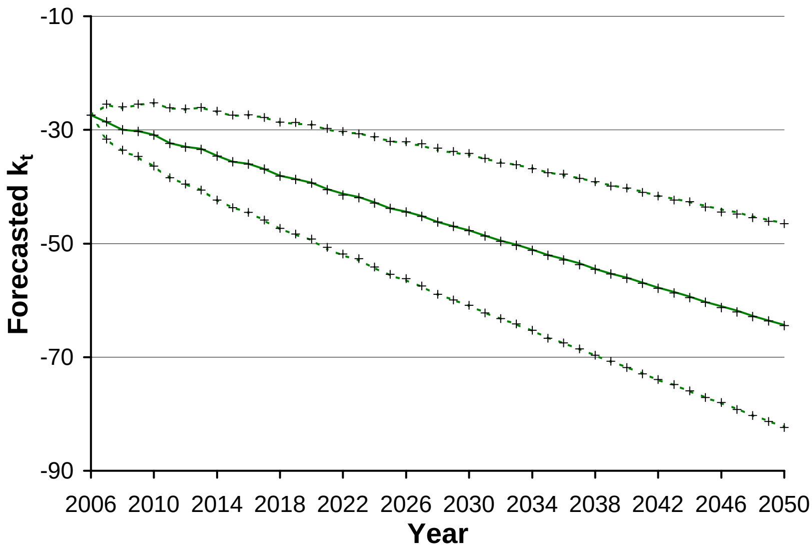


Figure 4.3: Forecasted k_t constructed by: a) the analytical forecast (green lines) and b) the 10000 simulations where parameter uncertainty is not included (black markers).

We performed the simulation for our model with $\tau = 2006$, $h = 44$, $N = 10000$ and computed the median, the 0.025 quantile and the 0.975 quantile of the values $k^*_{\tau+h}$. A comparison with the analytical forecast is displayed in Figure 4.3. We see that the difference between the forecasts is marginal and hence we conclude that the assumption that $Z_t \sim \mathcal{N}(0, \hat{\sigma}^2_Z)$, is reasonable.

4.1.3 Simulation with parameter uncertainty

In this subsection we describe a simulation to investigate how the forecast is affected when the parameters of the model are slightly perturbed. The reason to do this is that the fitted parameters are found by some optimization method, for instance maximum likelihood, and will deviate from the exact parameters. It is therefore important to know what the effects are on the forecast when the estimated parameters are a little altered.

At the start of every iteration, a random perturbation is added to the drift term μ. Instead of taking the drift equal to $\hat{\mu}$, forecasts are generated with $\mu^* \sim \mathcal{N}(\hat{\mu}, \hat{\sigma}^2_\mu)$. The standard error $\hat{\sigma}_\mu$ can be found via the matrix of second derivatives of the log-likelihood function of all parameters, its value for our model is displayed in Table 3.2. The ARMA parameters $\hat{\theta}^*_1, \ldots, \hat{\theta}^*_p, \hat{\phi}^*_1, \ldots, \hat{\phi}^*_q$ are obtained by fitting a zero mean ARMA(p, q) model on $\hat{k}_{\tau-n+2} - \hat{k}_{\tau-n+1} - \mu^*, \ldots, \hat{k}_\tau - \hat{k}_{\tau-1} - \mu^*$. Future innovations $z^*_{\tau+i}$ are generated as $z^*_{\tau+i} \sim \mathcal{N}(0, (\hat{\sigma}^2_Z)^*)$. The reason we do not use a random drawing of the set of fitted innovations to generate future innovations, is that the mean of this set can be significantly different from zero. One iteration of this simulation is illustrated in Table 4.2.

We performed this simulation for the same τ, h, N as before and displayed the median, the 0.025 and the 0.975 quantile in Figure 4.4, together with the analytical forecast from Subsection 4.1.1. The prediction bounds are more than 20% wider in 2030 and almost 33% wider in 2050.

Table 4.2: Generating one sequence $k^*_{\tau+1}, \ldots, k^*_{\tau+h}$ with parameter uncertainty.

Given drift estimates: $\hat{\mu}, \hat{\sigma}^2_\mu$ and fitted values $\hat{k}_{\tau-n+1}, \ldots, \hat{k}_\tau$.
Generate $\mu^* \sim \mathcal{N}(\hat{\mu}, \hat{\sigma}^2_\mu)$ and fit an ARMA(p,q) model on
$\hat{k}_{\tau-n+2} - \hat{k}_{\tau-n+1} - \mu^*, \ldots, \hat{k}_\tau - \hat{k}_{\tau-1} - \mu^*$.

Store the parameters $\hat{\phi}^*_1, \ldots, \hat{\phi}^*_p, \hat{\theta}^*_1, \ldots, \hat{\theta}^*_q, (\hat{\sigma}^2_Z)^*$,
and the fitted values: $\hat{y}^*_{\tau-p+1}, \ldots, \hat{y}^*_\tau, \hat{z}^*_{\tau-q+1}, \ldots, \hat{z}^*_\tau$

for $i = 1 : h$ do
$y^*_{\tau+i} = \hat{\phi}_1 \tilde{y}_{\tau+i-1} + \cdots + \hat{\phi}_p \tilde{y}_{\tau+i-p} + \hat{\theta}_1 \tilde{z}_{\tau+i-1} + \cdots + \hat{\theta}_q \tilde{z}_{\tau+i-q} + z^*_{\tau+i}$.
$k^*_{\tau+i} = k^*_{\tau+i-1} + \hat{\mu} + y^*_{\tau+i}$.
end
where

$$\tilde{y}_t = \begin{cases} \hat{y}^*_t & \text{if } t \leq \tau; \\ y^*_t & \text{if } t > \tau; \end{cases} \qquad \tilde{z}_t = \begin{cases} \hat{z}^*_t & \text{if } t \leq \tau; \\ z^*_t & \text{if } t > \tau; \end{cases}$$

$k^*_\tau = \hat{k}_\tau$ and $z^*_t \sim \mathcal{N}(0, (\sigma^2_Z)^*)$ for $t > \tau$ IID.

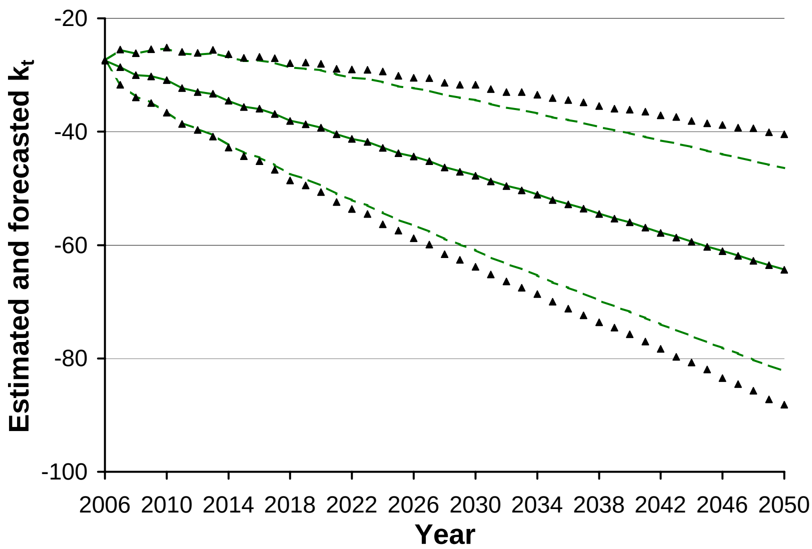


Figure 4.4: Forecasted k_t constructed by: a) the analytical forecast (green lines) and b) the 10000 simulations where parameter uncertainty is included (black markers).

4.2 Model evaluation

In this section we investigate whether the ARIMA$(2, 1, 2)$ model, with parameters from Table 3.2, will create reliable forecasts using the methods described in the previous section. Firstly, the volatility of the drift term $\hat{\mu}$ is examined between 1950 and 2006. Secondly, the model is fitted on several periods from the past and future values for k_t are forecasted until 2006. The values that are most likely to happen and 95% prediction bounds will be displayed and compared with the fitted $\hat{k}_t$ based on the entire period 1950-2006. This gives us an indication whether future mortality will be well predicted by our model. Thirdly, it is tested how these forecasts for k_t lead to forecasted values for the *flat life expectancy*, which is a measure of mortality in one calender year, defined in (4.16).

4.2.1 The drift

We calculated $\hat{\mu}$ for different periods in the past and displayed them in Table 4.3. Note that there is a large gap between the outcomes of the first two periods. The

Table 4.3: Comparison of the drift $\hat{\mu}$ for different periods.

period	$\hat{\mu}$
1950-1959	-1.0613
1950-1969	-0.5305
1950-1979	-0.7572
1950-1989	-0.7163
1950-1999	-0.6686
1950-2003	-0.6929
1950-2006	-0.8291

drift term corresponding to 1950-1959 is -1.0613 which is far below average. The drift from the period 1950-1969 reads $\hat{\mu} = -0.5305$, which is relatively high. When the period is 30 years or longer $\hat{\mu}$ is less volatile, which is fortunate. The estimate of $\hat{\mu}$ declines from -0.6929 based on 1950-2003, to -0.8291 when the last three years are incorporated. This is due to the sharp decrease in the years 2004, 2005 and 2006, as can be seen in the top left chart of Figure 4.5, where a scatter plot of $\hat{k}_t$ together with the best linear approximation in the least squares sense is displayed. We see that the values in these years are extremely low, which is unprecedented since the 1970's where a deviation from the approximation of multiple years can be observed, albeit in the opposite direction. The slope of the approximation equals -0.75 and when 1950-1976 is left out the slope is -0.78, which illustrates that overall the decline has been steady.

4.2.2 Forecasting k_t

Let us examine how the model predicts k_t based on several periods in the past, for years that have already been observed. For various τ, $1950 < \tau < 2006$, we fitted the model on the data from 1950 up to end including τ and predicted $k_{\tau+1}, k_{\tau+2}, \ldots, 2006$. We compared these values with the fitted values $\hat{k}_t$ from our complete set of data. For $\tau < 1982$ forecasted k_t turned out to be unreliable, because when fitting the parameters of the ARIMA$(2, 1, 2)$ model the estimates vary a lot between consecutive years. After 1982 these estimates were relatively stable over time, as was also concluded about $\hat{\mu}$ in the previous subsection. In Figure 4.5 the forecasts starting in 1982, 1989 and 1996 are displayed. Two pairs of prediction bounds are constructed, one using the techniques described in Subsection 4.1.1, and one using the empirical distribution from the simulation with parameter uncertainty included, see Table 4.2. The latter prediction bounds are wider than the former ones. For every year, $\hat{k}_t$ stays within all prediction bounds and is closely predicted until 2003. The value $\hat{k}_{2006}$ is underestimated by all forecasts. The shape of the observed values is often replicated by the forecasts, especially in the first ten years. This suggests that the AR and MA terms improve the quality of the forecast compared to the original ARIMA$(0, 1, 0)$ from Lee and Carter.

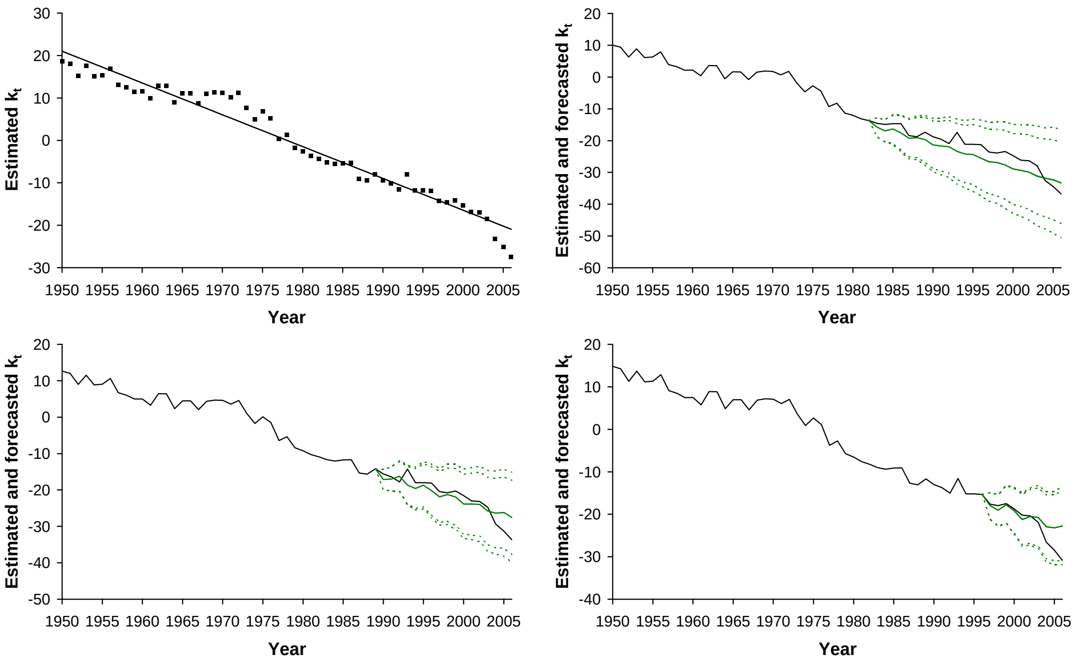


Figure 4.5: **Top left**: scatter plot of $\hat{k}_t$ and a first order approximation. **Top right**: $\hat{k}_t$ (black line) and forecasted values based on data from 1950-1982 (green line) with two pairs of 95% prediction bounds (green dotted). **Bottom left**: idem based on 1950-1989. **Bottom right**: idem based on 1950-1996.

4.2.3 Forecasting flat life expectancy

Another way to measure the performance of the model is by comparing the *flat life expectancy* FLE(x, τ), which is equal to life expectancy $\tilde{e}_{x,t}$, see (1.11), but by postulating that $q(x, t) = q(x, \tau)$ for all t:

$$\text{FLE}(x, \tau) = \sum_{i=1}^{\infty} iq(x + i, \tau) \prod_{j=0}^{i-1} (1 - q(x + j, \tau)). \tag{4.16}$$

When we replace $q(x, \tau)$ by $m(x, \tau)$, a realization of $M(x, \tau)$ which was defined in (3.1), we obtain the *observed* flat life expectancy from the Netherlands. Similarly the fitted FLE can be computed for a given mortality model. For the Lee-Carter model this means that $q(x, \tau)$ is replaced by $\hat{m}(x, \tau)$, where

$$\hat{m}(x, \tau) = e^{\hat{a}_x + \hat{b}_x \hat{k}_\tau}. \tag{4.17}$$

In Figure 4.6 we plotted FLE$(20, \tau)$, $1950 \leq \tau \leq 2006$, for the observed death rates, together with forecasts based on the same periods as in Subsection 4.2.2: 1950-1982, 1950-1989 and 1950-1996. The forecasts are obtained from plugging in the fitted rates (4.17) into (4.16).

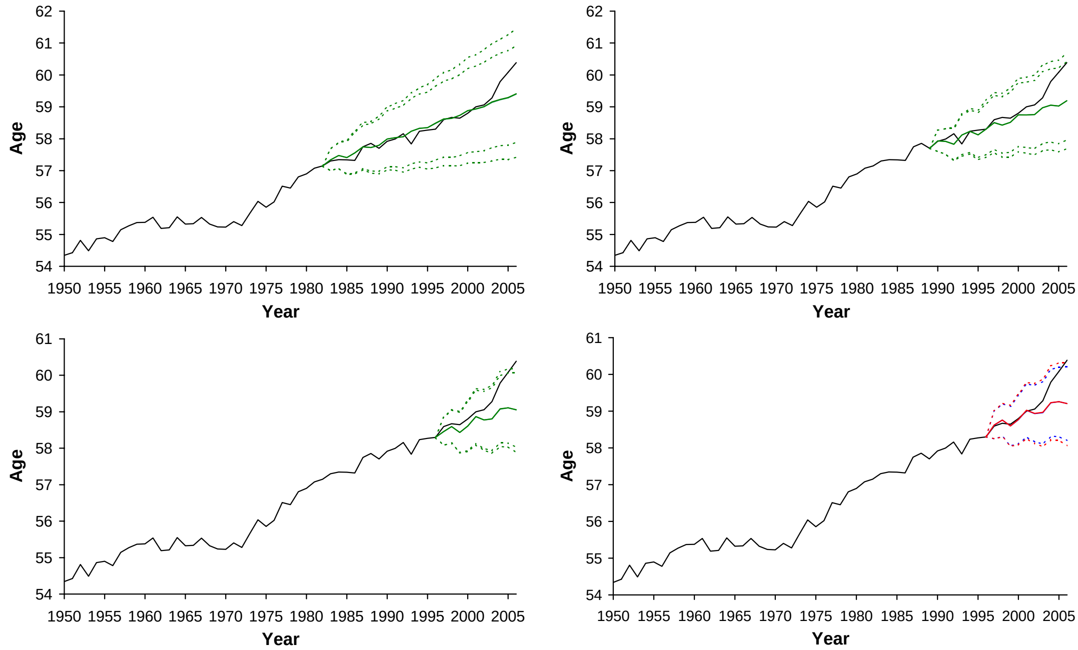


Figure 4.6: Flat life expectancy for people aged 20, observed (black) and forecasted (green and red), with 95% prediction intervals (green and red dotted), based various periods: **Top left**: 1950-1982. **Top right**: 1950-1989. **Bottom left**: 1950-1996. **Bottom right**: 1950-1996 (with a jump off correction).

The prediction bounds are constructed by plugging in the prediction bounds of k_t (from Figure 4.5) into (4.17), without including the fitting error $\varepsilon_{x,t}$. Lee and Carter found that when forecasting the FLE, k_t accounts for most of the uncertainty of the forecast ([20], Appendix B) which justifies this omission. They conclude that if a single death rate $m(x, t)$ is forecasted, $\varepsilon_{x,t}$ will add substantially to the uncertainty, but since the FLE is calculated as a product of $q(x, t)$ the fitting error is a product of $e^{\varepsilon_{x,t}}$ which is negligible since errors with different signs will cancel.

Again all forecasts are accurate until 2003. The observed FLE stays within all prediction bounds for the forecasts starting in 1982 and 1989. For the other forecast the observed FLE falls out of the analytical prediction bound for 2005 and 2006 and for the empirical bound in 2006. Together these three forecasts predicted 51 years, where only one or two values (depending on the bounds) have fallen outside the prediction bounds, which is less than 5%.

Lee and Miller published an evaluation of the Lee-Carter model in 2001 [21]. They propose an adaption for the estimate of a_x and k_t in the jump off year τ: "... the model would not fit the age-specific mortality data exactly in the jump off year, thus the initial conditions for the forecast would not be quite right. This situation inevitably would lead to error, which would be particularly important in the early years of the forecast"([21], p. 1113). Consider FLE$(20, \tau)$ with jump off year 1996. The fitted value in that year is 58.12, while the observed value is 58.30. When we apply the jump off correction we set $\hat{a}_x$ equal to the log of the age specific death rates of 1996 and put $\hat{k}_{1996} = 0$. The resulting forecast is slightly altered as can be seen in the bottom right chart of Figure 4.6. We see a small improvement in the first few years, and now only one value falls outside of the analytical prediction bounds. However, we expect that in the long run the forecast will not deviate a lot from the forecast without jump off correction. A risk of applying the correction is that the errors from the jump off year get extrapolated into the future.

Chapter 5

Financial consequences of using different mortality models

In this chapter we illustrate the consequences of longevity risk. We shall compare the forecast of our implementation of the Lee-Carter model applied to Dutch mortality rates, with the forecasts of the CBS (Dutch Bureau of Statistics) and the AG (Dutch Association of Actuaries). In Section 5.1 the method that is used by the CBS to construct their forecast, is discussed. Medical and sociological information is incorporated together with extrapolative techniques. Only one scenario of future mortality is provided. In Section 5.2 the forecast of the AG is described. For actuaries these are the most important rates since they are used to determine future expenses of a pension fund. Their method is, just like Lee-Carter, a purely extrapolative method. Although they provide only one scenario of future mortality, they describe a method to incorporate uncertainty in their forecast. The purpose of Section 5.3 is to give an introduction to actuarial calculations. We derive a formula for the *liabilities* at time τ of a pension fund, which can be defined as the expected future expenses of the fund to its participants to fulfill the commitments that were made up to and including τ. This formula will be used to compare the three mortality models in Section 5.4, in which the liabilities for an average Dutch pension fund are computed. The results of

Section 4.1 enable us to calculate the liabilities using the Lee-Carter model for different levels of uncertainty. In order to do this, the simulations defined in Table 4.1 and 4.2 will be extended to calculate the liabilities for every realization of mortality. Section 5.5 focuses on the difference between the AG forecast and the analytical forecast of Lee-Carter with minimum MSPE. We designed an experiment where an average fund is winding up and look at the financial consequences when mortality improvement is underestimated. In particular we shall focus on the disparity between young and old people.

5.1 The forecast of the CBS

Every two years the CBS publishes a sex-specific prognosis of future death probabilities. In 2000 and 2002 this forecast stretched until 2049, while in 2004 and 2006 the final year was 2050. The expected flat life-expectancy from birth $\mathrm{FLE}(0, 2049)$ has increased significantly between 2000 and 2006 as can be seen in Table 5.1. This is due to the sharp decrease of death rates in 2004 and 2005.

Table 5.1: Development of the forecasted $\mathrm{FLE}(0, 2049)$.

	2000	2002	2004	2006
Males	79.51	79.51	79.52	81.43
Females	82.6	82.51	82.61	84.13

The purpose of this section is to describe the method that is used by the CBS to calculate future death probabilities. It is a summary of the official document [15], which can be found on the CBS website.

Based on medical expertise, sociological arguments and historic development, a prognosis for 2018, 2034 and 2050 is made for age groups until 80. Interpolation is used to construct age and time specific death probabilities. For people older than 80, death probabilities are determined via extrapolation. A distinction is made between the following causes of death:

- Cancer, subdivided in lung-, breast-, prostate cancer and a group of other forms of cancer.
- Heart and vascular diseases.
- Diseases on the respiratory organs.
- Non-natural causes of death.
- Other causes of death.

A separate prognosis is made for age classes 0, 1-19, 20-49, 50-69, 70-79 and 80-99. People older than 99 are left out of the forecast. For people in group 1-19 natural death causes play a minor role. Non-natural causes like traffic accidents are the most common cause of death. For the ages 20-49 natural causes are also insignificant compared to the non-natural causes, in particular suicide. From age 50 the natural causes become the dominant factor.
The death of a young person is called a *premature death*. In many cases we can imagine that in the absence of a fatal cause, this person would have the same life-expectancy as anybody else with the same age. On the other hand the CBS speaks of a *geriatric death*, when the cause of death is less important for the remaining life span. Usually old people suffer from a general decline of health, so in the absence of the fatal cause of death a person would have died from something else within a short time. Therefore, no distinction between causes of death are incorporated in the forecast for the highest age group.
For illustration purposes, we now describe how the forecast for male deaths by lung cancer is obtained (see [15], p.66). About 85% of these deaths is due to smoking. In the fifties more than 90% of the male population smoked. This percentage has dropped to 40% in 1990 and stabilized until the new millennium. Since 2000 a new drop is visible. Anti-smoke advertisement and the right to work in a smoking-free

environment, a law that took effect in 2002, are possible explanations for the latest descent. The effect on mortality was not instantaneous. A downward trend has started only 20 years ago. Survival chances for patients of lung cancer did hardly improve since the seventies. Research has shown that only one out of every eight people is alive five years after the diagnosis.
The CBS expects the current improvements to continue until 2018. Between 2018 and 2034 a slower trend is expected, due to the stabilization of the percentage of smokers during the nineties. In the last interval 2034-2050 faster improvement is expected due to the most recent descent of the number of smokers, and the possibility that better treatment is available. The forecast for people aged 80 and older is obtained by linear regression to death rates of people from age classes 50-69 and 70-79.

The CBS method can be characterized as an *explanatory method*. They are convinced that the best way to predict mortality is by making intelligent considerations. "Examining different causes of death leads to more knowledge of the factors from the underlying process of mortality change"([15], p. 62). The risk of an explanatory method is the large amount of subjective judgement. Although our medical expertise nowadays is vast, there is no consensus about future developments. Throughout the past, explanatory models have tended to underestimate the improvement of the death probabilities.

5.2 The forecast of the AG

Most actuaries in the Netherlands are member of the Actuarial Society (in Dutch: 'Actuarieel Genootschap' (AG)), the professional association of Dutch actuaries. The AG was founded in 1888, and aims to encourage the sector and to maintain relations with international actuarial organizations. Every five years, the AG publishes (current) death probabilities based on mortality data of the CBS that can be used for financial purposes.

The mortality prognosis by the AG, which has been published only once so far (in 2007), differentiates between males and females. In next editions other factors will be included to discriminate between smokers/non smokers and level of income. The AG-prognosis and a description of their method can be found in [14]. It is interesting to note that the board of the AG have considered the Lee-Carter- and the CBS models, but in the end they have picked the *CRC model* to construct their forecast. CRC is an abbreviation of 'Commissie Referentietarief Collectief' which is a research team founded at the end of the 1980s consisting of people from the insurance business. AG motivates the choice of this model by the following arguments:

- CRC is a well known model to the Dutch market.
- CRC is transparent and relatively easy to understand.
- Outcomes of different models do not vary much.
- Trend uncertainty can easily be added.

The purpose of this section is to illustrate how the prognosis of the AG is constructed. As in the case of the CBS model, we only used the official source which lacks the details needed to exactly replicate the model and to reconstruct the rates that are published.

The CRC model assumes that for every age x the (sex-specific) annual death probability is reduced with a constant α_x:

$$q(x,t) = \alpha_x q(x,t-1), \qquad 0 \leq x \leq 120.$$

An error term is not included. Note that this is similar to Lee-Carter when k_t has a linear trend. The reduction factor α_x is computed in four steps:

1. The Van Broekhoven algorithm is applied to the observed death rates from the CBS from $\tau - n + 1$ up to and including τ, where τ is the final year of observation. This algorithm transforms death rates for people aged $x + \frac{1}{2}$ to rates for people

aged x. Van Broekhoven is a member of the CRC research team and also a co-author of [14]; see [6] for a description of his algorithm .

2. For the selected ages, a moving average filter of length five is applied to the adapted death rates.
3. The time-average reduction is computed.
4. The values resulting from 3. are smoothed by a moving average filter.

A future death rate of h years ahead is calculated from the last observed year τ,

$$m(x, \tau + h) = \alpha_x^h m(x, \tau). \tag{5.1}$$

When for some x, the future death rate for males becomes smaller than for females, the female reduction factor is adapted in a way that both death rates are equal in the final year of the forecast. The model was implemented based on observed death rates in $1988, \ldots, 2005$, the final year of their forecast is 2050. Data before 1988 are not used because a split in the trend of historic death rates was observed. For the ages 0-19 and 91-120, α_x is determined from death rates between 2000 and 2005.

Although the CRC model is deterministic, the AG describes a way to construct a prediction interval for $\text{FLE}(x, \tau)$ ([14], p. 44). First, the *most likely* estimate $m_{ml}(x, \tau + h)$ is defined as the future death rate that is most likely to happen, given that the model is correct. In this case

$$m_{ml}(x, 2005 + h) = \alpha_x^h m(x, 2005), \tag{5.2}$$

where α_x is calculated as described above. A corresponding most likely estimate for the flat life expectancy $\text{FLE}_{ml}(x, 2005 + h)$ is computed by plugging (5.2) into (4.16). Their 95% prediction interval is given by

$$\left[\text{FLE}_{ml}(x, 2005 + h) - 2.45S,\ \text{FLE}_{ml}(x, 2005 + h) + 2.45S\right], \tag{5.3}$$

where S is the trend uncertainty. This interval is constructed by generating predictions for $\text{FLE}(x, \tau + h)$, denoted by $\text{FLE}(x, \tau + h)_i$, where it is assumed that

$$\text{FLE}(x, \tau + h)_i \sim \mathcal{N}(\text{FLE}_{ml}(x, \tau + h), \sigma^2_{\text{FLE}(x,\tau+h)}), \tag{5.4}$$

and $\text{FLE}(x, \tau + h)_i$ IID realizations of $\text{FLE}(x, \tau + h)$ for every i. A prediction is obtained by computing reduction factors based on time intervals from the past. Denote ${}_{\tau_i,n}\alpha_x$ for the reduction factor that is based on mortality data from $\tau_i - n + 1, \ldots, \tau_i$. Future death rates are computed using the latest available data:

$$m(x, \tau + h)_i = {}_{\tau_i,n}\alpha_x^h m(x, \tau),$$

and $\text{FLE}(x, \tau + h)_i$ is computed in the usual way from $m(x, \tau+h)_i$. For k predictions the sample variance S_k^2, an estimate for $\sigma^2_{\text{FLE}(x,\tau+h)}$, is defined by:

$$\begin{aligned} S_k^2 &= \frac{1}{k-1} \sum_{i=1}^{k} (\text{FLE}(x, \tau + h)_i - \overline{\text{FLE}}(x, \tau + h)_k)^2 \quad \text{where} \\ \overline{\text{FLE}}(x, \tau + h)_k &= \frac{1}{k} \sum_{i=1}^{k} \text{FLE}(x, \tau + h)_i. \end{aligned}$$

Using (5.4) and the assumption that all predictions are independent, it follows that

$$\frac{\text{FLE}(x, \tau + h) - \text{FLE}_{ml}(x, \tau + h)}{S_k} \sim t(k - 1),$$

where $t(k - 1)$ is the Student's t-distribution with $k - 1$ degrees of freedom. A corresponding $(1 - \alpha)$ prediction interval can be obtained by

$$\left[\text{FLE}_{ml}(x, \tau + h) - t_{1-\alpha/2}(k - 1)S_k,\ \text{FLE}_{ml}(x, \tau + h) + t_{1-\alpha/2}(k - 1)S_k\right] \qquad (5.5)$$

where $t_{1-\alpha/2}(k - 1)$ denotes the $(1 - \alpha/2)$ quantile of the $t(k - 1)$ distribution. The AG has picked seven historic intervals, with $n = 20$, which explains the value $t_{0.975}(6) \approx 2.45$. Unfortunately only a few outcomes have been published, namely the forecasted $\text{FLE}(x, 2050)$ for $x = 0, 65$, which are shown in Table 5.2. We should remark that the AG uses another formula to calculate the FLE than was stated in (4.16). They assume that when somebody dies at age x, the life span has been (on average) $x + \frac{1}{2}$. So in order to compare the results of the AG with ours, one could subtract $\frac{1}{2}$ from the values of Table 5.2.

Table 5.2: Forecasted FLE$(x, 2050)$.

	x	$\text{FLE}_{ml}(x, 2050)$	0.025 quantile	0.975 quantile
Males	0	82.80	77.48	88.12
	65	19.81	15.89	23.73
Females	0	84.50	80.16	88.84
	65	21.42	17.40	25.44

Unlike the CBS, the AG has decided to use an extrapolative model to forecast mortality. They share with the CBS that they have chosen to use a model which is familiar to them. One of their arguments is that using another model will not lead to an outcome that is very different, and thereby justify the decision to refrain from more sophisticated models like Lee-Carter. Their prognosis stretches as far as 2050, which is 45 years ahead, based on only 18 years of data. Their prediction interval is very large compared to the prediction intervals we found, which will be discussed later.
In Figure 5.1 we displayed the FLE$(20, t)$, $t = 2006, \ldots, 2050$, for the forecasts of the CBS, the AG and our implementation of the Lee-Carter model with prediction bounds as in Figure 4.4. Since the AG and CBS have constructed a sex-specific forecast, we took the average rate of males and females to produce a sex independent rate. The Dutch population consists of more women than men (about 50.5% women in 2005) and because death rates are lower for women, the resulting death probabilities are slightly too low and hence the displayed FLE is a little too high. The most recent year that is used for the forecasts of CBS and AG is 2005, while for Lee-Carter we also included death rates that were observed in 2006. In order to make a fair comparison, we fitted Lee-Carter only on death rates between 1950 and 2005 to construct a forecast. Let us compare the prediction intervals from the AG with our prediction intervals. The length of their interval for FLE$(0, 2050)$ is 10.64 for men and 8.86 for women, while this is 5.01 for our interval where parameter uncertainty is included and 3.76 when only the uncertainty from the innovations Z_t is incorporated. It must be noted that by this comparison the difference is overrated, because our prediction

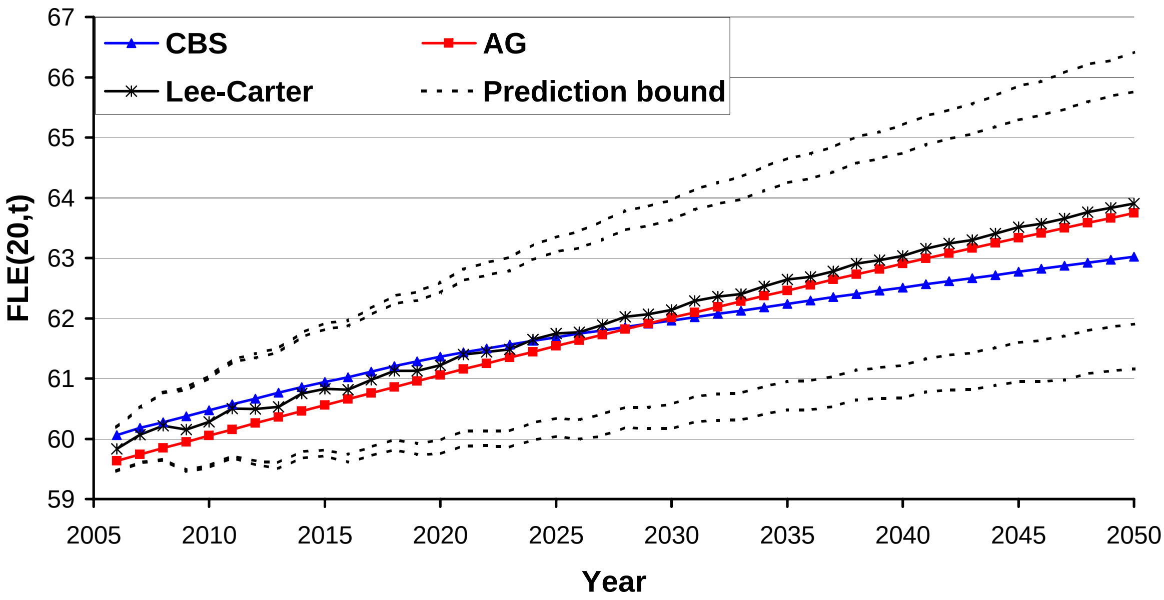


Figure 5.1: The FLE of a person aged 20.

interval does not include the ages 0-19. However this age group does not affect the FLE a lot since the death rates are low. As a matter of fact, the length of our prediction intervals for FLE(20, 2050) is still smaller than the AG prediction interval for FLE(65, 2050), which is 7.84 for men and 8.04 for women. This is a contradiction with the conviction of the AG, who say that forecasts of different models do not vary much.

5.3 Calculating the liabilities of a pension fund

Pension is a periodical payment that is obtained under certain circumstances. The most important are *Old-age Pension* (in Dutch: 'Ouderdoms Pensioen' (OP)), issued to people that have been retired and *Spouse Pension* (in Dutch: 'Nabestaanden Pensioen' (NP)), issued when a member passes away, to the surviving spouse. In the former case these payments stop when the member passes away and in the latter case when the widow(er) passes away. In this section we explain how the *liabilities* of a pension fund can be calculated. Liabilities can be defined as the expected amount of money that is needed to cover future expenses as a result from financial commitments that were made to the participants of the fund. We will focus on OP, the most important pension, liabilities of the NP can be determined in a similar way.

When somebody retires, a pension funds starts to pay out. Let us assume that this amount, which we call the *total accrued benefits* c, is paid at the beginning of each year t, given that this person is alive. The value of c depends on the number of years T that an employee has worked, his/her (annual) salary S and the pension plan rules. The goal of a pension plan is that, when a person reaches the retirement age (which is currently 65), he/she will have an income of $\alpha T \times S$, where α is called the *accrue rate*. Often α is around 0.02, so that an employee who has worked 40 years will receive 80% of its income as pension. Regardless of working history, everybody in the Netherlands who reaches the age of 65 receives statutory old age pension (in Dutch: 'Algemene Ouderdomswet' (AOW)) A. Because this income is guaranteed, for some part of S no benefits have to be accrued to realize the income goal. This part is called *offset* F, the value $S - F$ is called the *pension base* which is needed to determine c. For instance, when a pension plan has $\alpha = 0.02$ with an offset of $F = 10/8A$ the annual accrued pension is $0.02(S - 10/8A)$. Somebody who has worked 40 years then receives

$$40 \cdot 0.02(S - 10/8A) + A = 0.8S$$

as income. So far we stated salary, offset and AOW as if they are fixed in time, which is off course unrealistic. They typically increase once every year, so from now we will indicate the calender year when we use these quantities. We state two pension plans: the *final pay plan* which was most popular until the new millennium and the *career average plan* which is currently used by most pension funds. The aim of the career average plan is that a person receives a percentage of its career-average salary. For a person who has worked during the years $\tau - T + 1, \dots, \tau$,

$$c_\tau = \sum_{t=0}^{T-1} \alpha(S_{\tau-t} - F_{\tau-t}).$$

The final pay plan is equivalent with the career average plan, but now (a percentage of) the last salary that was earned is received:

$$c_\tau = T\alpha(S_\tau - F_\tau).$$

We denote $L_{\tau,\tau+h}$ as the liabilities, based on the information up to and including τ, of h years ahead. More specific, $L_{\tau,\tau+h}$ is determined at the beginning of year $\tau+1$ and denotes the payments that a pension fund expects to make at (the beginning of) $\tau+2, \tau+3, \ldots, \tau+h+1$, to cover the financial commitments that were made to the participants of the fund until $\tau+1$.

Consider a person aged $x \geq 65$ in year $\tau+1$, with total accrued benefits c_τ. A payment in the beginning of $t = \tau+2$ will be made to this person, only when he/she is alive at that time. Therefore the expected payment equals

$$c_\tau \cdot P(R_{x,\tau+1} \geq 1), \tag{5.6}$$

where $R_{x,\tau+1}$ denotes the remaining life span of someone alive at the beginning of year $\tau+1$ at an age of x. Since we are talking about a future payment, we can invest the money and receive some return rate. The expected payments that are due in $\tau+3, \ldots, \tau+h+1$ can also be invested. Denote by ${}_\tau r_{\tau+t}$ the expected annual return rate, which is based on the information up to and including τ, with a yield to maturity of t years. Every month the *Dutch National Bank* (DNB) publishes these return rates which are in conformity with market prices. Since January 1st 2007 pension funds are obliged to use these rates to determine the liabilities. The methods that are used to determine ${}_\tau r_{\tau+t}$ can be found in [12]. The liabilities h years ahead can now be obtained by

$$L_{\tau,\tau+h} = \sum_{t=1}^{h} \frac{c_\tau \cdot P(R_{x,\tau+1} \geq t)}{(1 + {}_\tau r_{\tau+t})^t}. \tag{5.7}$$

We now extend the analysis to a pension fund with ages $x \in \mathcal{X}$ in $\tau+1$. Let us denote $c_{x,\tau}$ as the total accrued benefits up to and including τ for *all* participants that were aged x in τ. The expected payment in $\tau+2$ for the group of people aged $x \geq 65$ in $\tau+1$ is given by $c_{x-1,\tau} \cdot P(R_{x,\tau+1} \geq 1)$. The liabilities of the entire fund can be calculated as

$$L_{\tau,\tau+h} = \sum_{x \in \mathcal{X}} \sum_{t=1}^{h} \frac{I_{[65,\infty)}(x+t) c_{x-1,\tau} \cdot P(R_{x,\tau+1} \geq t)}{(1 + {}_{\tau}r_{\tau+t})^t}, \tag{5.8}$$

where $I_{[65,\infty)}(x)$ is an indicator function to ensure that payments are only made to people that have reached the retirement age.

5.4 The price of longevity

In this section we examine the financial consequences of using the described models of the CBS, AG and Lee-Carter to calculate the liabilities $L_{\tau,\tau+h}$, where $\tau = 2005$ and $h = 1, \ldots, 45$. We create an imaginary fund which should resemble an average Dutch pension fund, to make sure the outcomes of the experiment are realistic.
For the participants of our fund we assume that:

- Everybody's working life lasts 40 years. A person starts working at the beginning of the year in which he/she will become 25 years old. In 2006 the fund consists of ages:
$$\mathcal{X} = \{26, \ldots, 99\}.$$
- The maximum age that a person can reach is 100.
- Everybody gets paid according to the career average plan with $\alpha = 0.02$ and offset $F = 10/7 \times A$.
- His/her salary has always been the modal salary.

To compute $L_{\tau,\tau+h}$, formula (5.8) is used. We have obtained ${}_{\tau}r_{\tau+h}$ and $c_{\tau,x}$ using the following information:

- Expected return rates ${}_{\tau}r_{\tau+h}$, where $\tau = 2005$ and $h = 0, 1, \ldots, 45$, are obtained from the DNB, see [12].

- For A we have taken the AOW in 2006 for a married person, $A = 8096.52$ euros.
- For the modal income we used the results of a survey that was held in 2005 by the CBS. The results are published in the StatLine data bank, see [8]. When we denote S_x to be the modal salary of a person aged x and F the offset, then the total accrued benefits for one person aged x is calculated by the sum

$$\alpha(S_{25} - F) + \alpha(S_{26} - F) + \cdots + \alpha(S_{\min(64,x)} - F).$$

- The average size of a Dutch pension fund is determined using the data from [13], the quarterly update of economic statistical information of the Netherlands provided by DNB. There we can find the number of people that accrue pension and the number of retired people sorted by age. The average fund is created by dividing these numbers by the total number of pension funds in 2006 which was 792. As a result our imaginary fund has 9860 participants.

The only thing that needs to be defined in order to compute $L_{\tau,\tau+h}$ is $P(R_{x,\tau+1} \geq t)$. Since

$$\begin{aligned} P(R_{x,\tau+1} \geq t) &= p(x,\tau+1) \cdot p(x+1,\tau+2) \cdots p(x+t-1,\tau+t) \qquad (5.9) \\ &= (1 - q(x,\tau+1)) \cdot (1 - q(x+1,\tau+2)) \cdots (1 - q(x+t-1,\tau+t)), \end{aligned}$$

model specific liabilities can be constructed, when $q(x,t)$ is replaced by $\hat{m}(x,t)$. In Figure 5.2 we displayed $L_{2005,t}$ for $t = 2006, \ldots, 2050$, where the forecasts of the AG, CBS and our implementation of Lee-Carter are used. The predictions by Lee-Carter are constructed by simulations: a forecast that is most likely to happen and two pairs of prediction bounds. Recall the simulations that were introduced in Section 4.1.1, which can be used to generate sequences $k^*_{\tau+1}, \ldots, k^*_{\tau+h}$. From every k^*_t, corresponding death rates $m^*(x,t)$ can be computed by (4.17). Subsequently, when these $m^*(x,t)$ are plugged into (5.8), a realization of $L^*_{\tau,\tau+h}$ is obtained. In the chart, we have displayed the median of 10000 iterations, where $L^*_{\tau,\tau+h}$ is generated at each iteration using the algorithm from Table 4.1. The first pair of prediction bounds is obtained

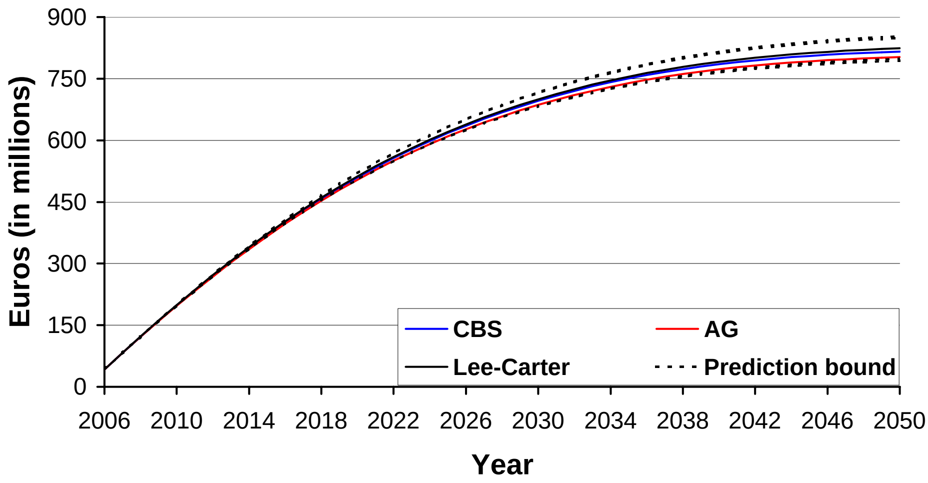


Figure 5.2: The liabilities $L_{2005,t}$.

by the 0.025 and 0.975 quantiles of this simulation. When the analytical forecast is used to compute $L_{\tau,\tau+h}$, the outcome is close to the median from the simulations. However the prediction interval that is obtained by plugging the analytical prediction bounds for k_t into (4.17) and then plug the resulting death rates into (5.8), is wider. We think the prediction interval using the simulation is more useful, because it incorporates the uncertainty of the entire process: the uncertainty in $m(x,t)$ together with the uncertainty in $L_{\tau,\tau+h}$, while the analytical prediction interval only incorporates the uncertainty in $m(x,t)$. The second pair of prediction bounds, with parameter uncertainty (wpu) included, are constructed by another simulation. Now the $L^*_{\tau,\tau+h}$ are generated using the algorithm described in Table 4.2.

In Figure 5.4 we displayed $L_{2005,2050}$ for the three models, and the levels of uncertainty that were described above. We observe that the value that is predicted by the AG is lower than for the CBS and the median of the Lee-Carter simulations. Also displayed in Figure 5.4 is a histogram with the relative difference between the AG and the other estimates for $L_{2005,2050}$. This shows us that a pension fund, which has used the prognosis of the AG to determine its liabilities, runs a significant risk to face higher expenses in the future, given that Lee-Carter is correct. The expected value of $L_{2005,2050}$ according to Lee-Carter is more than 2% higher than the value that is predicted by the AG. The difference between AG and the upper bounds of Lee-Carter is more than 5.5%. This means that when Lee-Carter is correct, the probability that future expenses until 2050 will be more than 5.5% higher than the prediction of AG, is 2.5%. In Figure 5.3 we have displayed kernel density estimators, which are based on the simulations from the liabilities until 2051. We see that according to Lee-Carter, the probability that future expenses will be higher than AG, is 94.3%, or 91.8% if parameter uncertainty is included. For the CBS, these rates are 72.4% and 69.1% respectively.

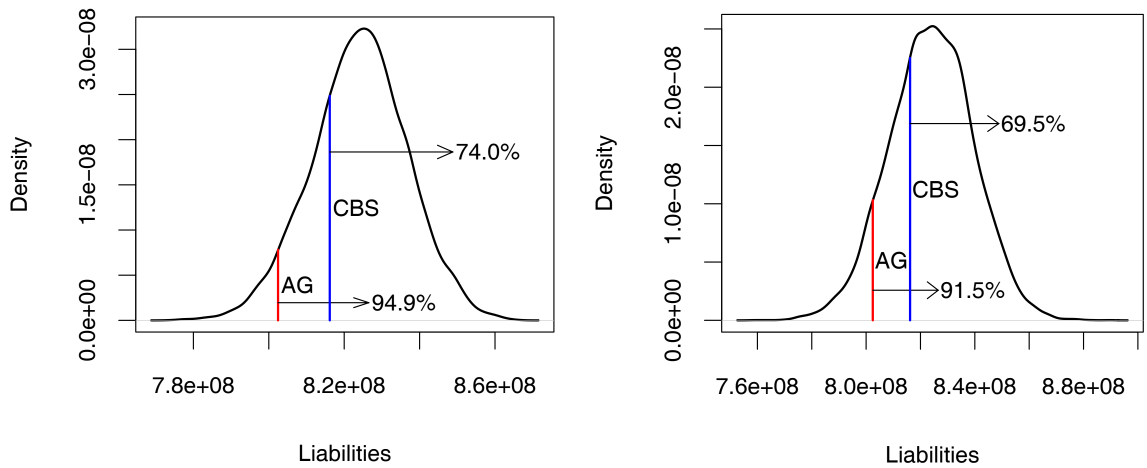


Figure 5.3: Kernel density estimators of the liabilities $L_{2005,2050}$, calculated by simulations of Lee-Carter, compared to the prognosis of CBS and AG. **Left**: without parameter uncertainty included. **Right**: with parameter uncertainty included

.

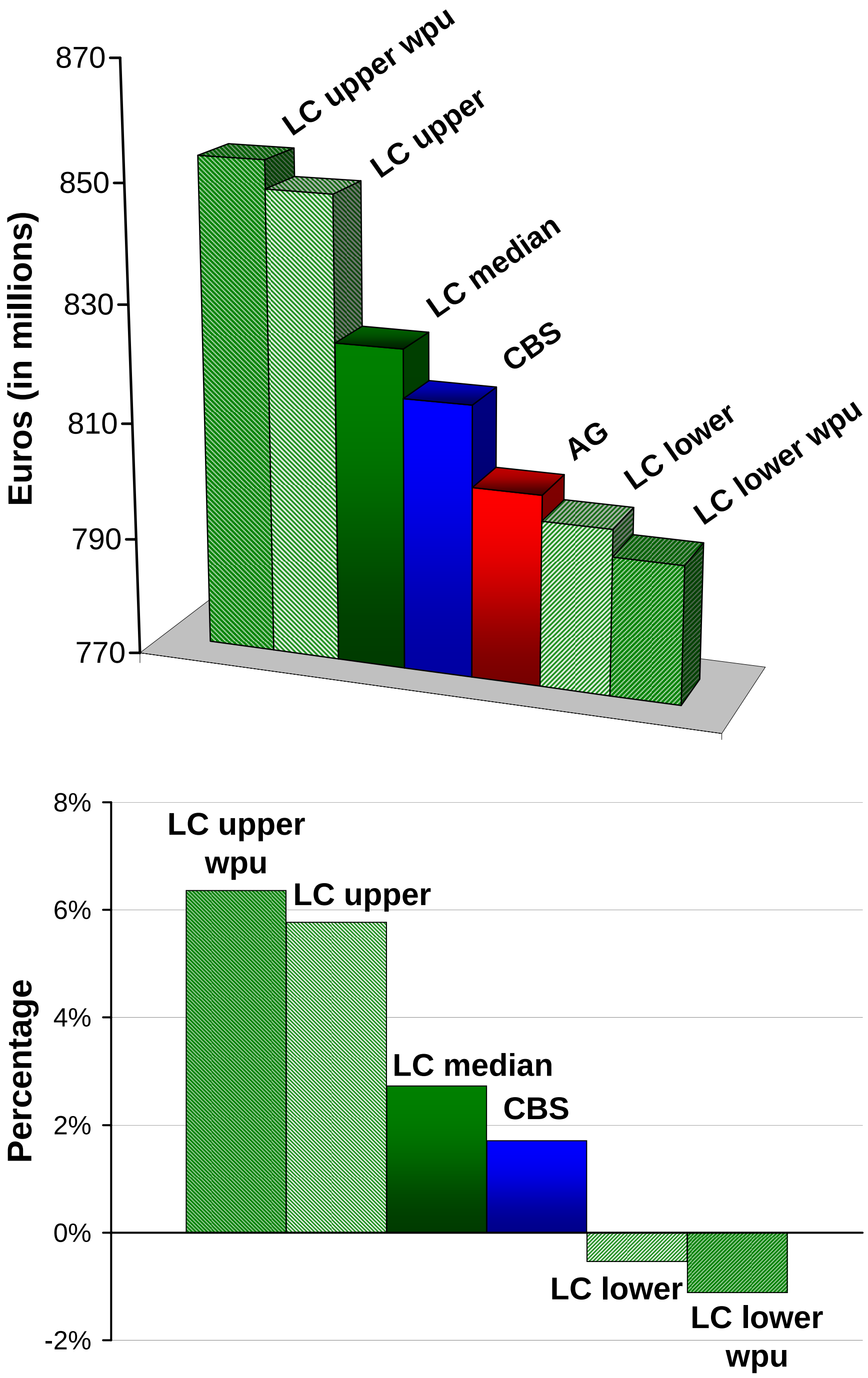


Figure 5.4: The liabilities up to and including 2050 $L_{2005,2050}$, calculated by the described methods. **Top:** The absolute amount of money needed to cover the liabilities. **Bottom:** The relative difference compared to the AG.

5.5 A generational disparity

In this section we perform an experiment about a pension fund that is winding up after year τ. In the future no new members nor premiums are accepted. However the board of the fund stays committed to pay the participants their pension on the beginning of the years $\tau+1,\ldots,\tau+h+1$. Moreover, they plan to pay an extra annual rate to compensate for future money devaluation (inflation), which is called the *indexation* rate i. The expected amount of money that is needed to cover these expenses can be calculated by incorporating indexation into (5.8),

$$V_{\tau,\tau+h} = \sum_{x\in\mathcal{X}}\sum_{t=1}^{h} \frac{I_{[65,\infty)}(x+t)c_{x-1,\tau}(1+i)^t \cdot P(R_{x,\tau+1}\geq t)}{(1+{}_\tau r_{\tau+t})^t}. \tag{5.10}$$

After year τ the board has to decide which model it wants to use to obtain predictions for $P(R_{x,\tau+1}\geq t)$, which determines the starting capital $C(\tau)$. After that, no more money will be added to the fund. As a result, when in $\tau+1$ death rates turn out differently as expected, it is possible that capital and expenses are no longer balanced. In that case a new indexation rate needs to be chosen for the remaining years in order to keep the financial situation healthy. When expenses are higher as expected (because less people die), i will be lowered and when costs turn out to be lower as expected, i becomes higher. Note that both situations are undesirable. When i is lowered, people that are not retired are in disadvantage, because they did not profit from the higher i in the previous year(s). On the other hand, retired people have reason to complain when i becomes higher, because they received a lower rate in the previous year(s). An interesting dilemma arises here. Young people, with many years to retirement, want a prudent mortality forecast to avoid the risk that when they retire, the capital has shrunk more than was anticipated, while older people rather have a risky forecast, because this leads to higher payments in the short run.
At every t, $t=1,\ldots,h-1$, the board has the opportunity to alter the indexation rate in a way that the current capital equals the forecasted expenses,

$$C(\tau+t) = V_{\tau+t,\tau+h}.$$

For the implementation of this experiment the following principles are used:

- The same pension fund is used as was described in Section 5.4, with the same accrued benefits $c_{x-1,\tau}$.
- The board has picked the forecasted mortality rates from the AG to predict $P(R_{x,\tau+1} \geq t)$, while in reality death rates will occur as predicted by the Lee-Carter model (the analytical forecast with minimum MSPE).
- At every t an indexation rate is chosen such that, when future death rates equal the expected death rates, $C(\tau + t + j) = V_{\tau+t+j,\tau+h}$ for $j = 0, \ldots, h - t$. The initial indexation rate is $i_\tau = 0.015$.
- The return rate ${}_\tau r_{\tau+t}$ is assumed to be constant during the entire experiment, i.e., ${}_\tau r_{\tau+t} = r$.

In the previous section we used return rates provided by the DNB. These rates are constructed for the situation that part of the capital will be invested for a long period. By this experiment the capital decreases and expenses are heavily influenced by the observed death rates so we need to have access to a large part of the capital at the beginning of every year. Therefore we decided to use a constant interest rate of $r = 0.04$. This rate is used by many actuaries as a rule of thumb for an *average* annual return rate.

We will denote $m_A(x,t)$ for the predicted death rates according to model A and $m_B(x,t)$ for death rates that are predicted by model B. For every t, the $m_A(x,t)$ are used to calculate *current expenses* $y(t)$ and the $m_B(x,t)$ are used to calculate expected future expenses, both *without indexation.* Hence, in our case 'A' stands for the AG prediction and 'B' for our implementation of Lee-Carter. When $y(t)$ is determined, we need to solve for which indexation rate $i_{\tau+t}$, expected future expenses and capital are equal. In the end we have a sequence $i_\tau, \ldots, i_{\tau+h}$, where i_τ is the indexation rate that was initially chosen. The algorithm that is used to run the experiment is described in Table 5.3. In Figure 5.5 a plot of the indexation rates for the years

Table 5.3: Generating the sequence $i_{\tau+1}, \ldots, i_{\tau+h}$.

Input: r, i_τ and $c_{x-1,\tau}, m_A(x,t), m_B(x,t)$ for $\tau < t \leq h$, $x \in \mathcal{X}$.
Determine $C(\tau)$ via (5.10), where $P(R_{x,\tau+1} \geq t)$ is found by plugging $m_A(x, \tau+t)$ into (5.9), $i = i_\tau$ and ${}_\tau r_{\tau+t} = r$.

for $t = 1 : h$ do
$c_{x-1,\tau} = (1 - m_B(x+t-1, \tau+t))c_{x-1,\tau}$ for $x \in \mathcal{X}$.
$y(\tau+t) = \sum_{x \in \mathcal{X}} I_{[65,\infty)}(x+t)c_{x-1,\tau}$.
Solve $i_{\tau+t}$ in:
$(1+r)C(\tau+t-1) = (1+i_{\tau+t})y(\tau+t) + V_{\tau+t,\tau+h}$
with $i = i_{\tau+t}$. Update:
$C(\tau+t) = (1+r)C(\tau+t-1) - (1+i_{\tau+t})y(\tau+t)$.
$c_{x-1,\tau} = (1+i_{\tau+t})c_{x-1,\tau}$ for $x \in \mathcal{X}$.
end

2006-2044 is displayed. We can see that indexation drops to a very low level. From 2040 the pension fund is even obliged to *lower* the pension payments. We did not show the indexation rates from 2045-2051 in the plot because these would disturb the scale. The rate of 2045 is -2.40% and this rate even drops to -13.32% in 2051.

This experiment is not meant to mimic a realistic situation. In reality there are many risks involved that influence future capital, most importantly investment risk, and they can not be neglected compared to the risk of longevity. However, by this experiment longevity risk is isolated to illustrate its effect when mortality improvement is underestimated. The experiment shows the importance of making accurate assessments when predicting mortality. We end this section by showing the relative benefit per age of the course $i_{\tau+t}$ from Figure 5.5. Let ι be the fairest indexation rate, which can always be determined afterwards, because the realized death rates are known. Fairest in this case means that i_t remains constant and $C(\tau+h) = 0$. At our experiment the fairest indexation would have been $\iota = 1.31\%$. Figure 5.6 shows per age the relative benefits from the course $i_{\tau+1}, \ldots, i_{\tau+h}$. It is a comparison of all payments that are received with the realized indexation rates, relative to the payments someone

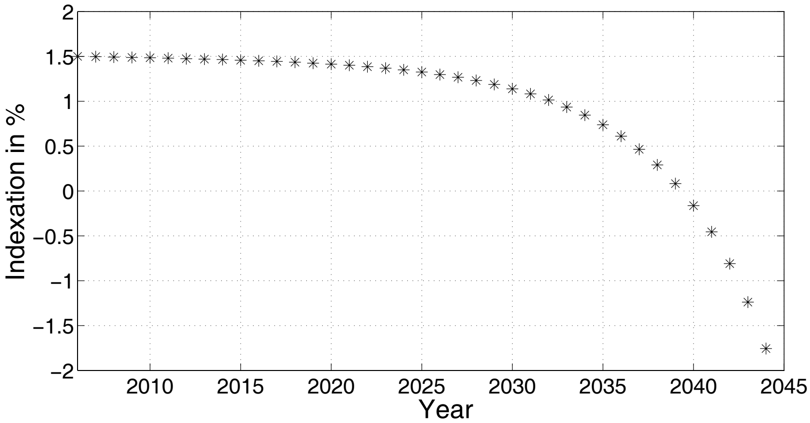


Figure 5.5: The development of the indexation rate between 2006 and 2051

would have received when a constant indexation rate of ι was used.

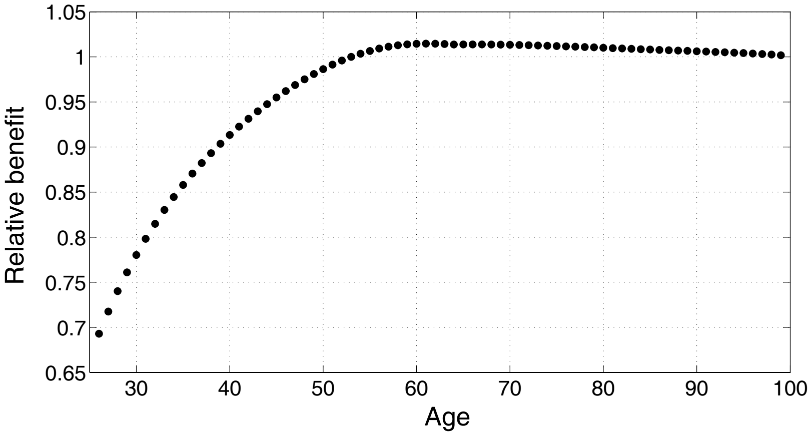


Figure 5.6: The relative benefit of the realized $i_{\tau+1}, \ldots, i_{\tau+h}$ compared to the fairest indexation rate ι for people aged x at τ.

Everybody older than 53 benefited from the way that i_t has developed. The people aged 61 the most, they have received 1.47% more than they would have done by a fixed indexation rate of ι. On the other hand, the people of age 26 only received 68.3% of what they would have received by fixed ι.

Conclusions and recommendations for further research

The goal of this thesis was to investigate the consequences of longevity risk for pension funds in the Netherlands. In order to do that we used the model of Lee-Carter to construct a forecast of Dutch mortality. We found that this model provides a good fit with the observed death rates. The fitted values $\hat{k}_t$ seem to have a linear trend, which is important for the performance of the model. In Chapter 4 we have examined several ways to construct a forecast. In Chapter 5 we used these forecasts to calculate the liabilities of an average Dutch pension fund, which were displayed in Figure 5.4. We observed that the expected amount of money that is needed to cover these liabilities is more than 2% higher than when the liabilities are calculated using the mortality prediction of the AG.

We have tested our model by applying it on death rates up to and including the years 1982, 1989 and 1996, and concluded that the forecasts from these years were close to the observed values until 2003, but that the decrease of death rates was underestimated in 2004, 2005 and 2006. This shows that the model has performed well over the past 25 years, and that mortality has not been overestimated. In order to obtain stable estimates for the parameters of the the model, we found that at least 30 years of observed death rates are needed. The AG has fitted its model on only 18 years of mortality data, to construct a forecast of 45 years ahead. This suggests that this forecast is susceptible to a substantial amount of parameter uncertainty. We have seen that the 95% prediction intervals that were published by the AG, are wider than

the ones we found.
For most ages, death rates are still decreasing rapidly and currently there are no signs that it is slowing down. Pension funds should therefore be aware that there is a considerable risk of longevity. One of the consequences of longevity risk is that young people will carry an unequal share of this burden in the future, as was illustrated in Section 5.5. Our implementation of Lee-Carter can be used to determine the *price* of longevity when future death rates are underestimated.

We can think of a couple of recommendations for further research. One of them is to apply the modifications of the Lee-Carter model, that have been proposed for instance by Wilmoth [28] or Renshaw and Haberman [26], to Dutch death rates. It would also be interesting to compare the Lee-Carter forecast with the forecast of other models that are popular at this moment, such as the mortality model constructed by Cairns, Blake and Dowd [10]. For people that have an economic background it can be interesting to investigate the risk of longevity on other financial products, such as a life insurance. In the United Kingdom it is investigated whether longevity risk can be traded, or *hedged*, with the risk of premature death. One could investigate whether these kind of trades are also applicable to the Dutch market.

Bibliography

[1] Bell, W.R. (1997) *Comparing and Assessing Time Series Methods for Forecasting Age-Specific Fertility and Mortality Rates*, Journal of Offical Statistics, Vol.13, No. 3, 1997,pp. 279-303.

[2] Bell, C.F., Miller, M.L. *Life tables for the United States Social Security Areas 1900-2100*, Actuarial Study No. 116
http://www.ssa.gov/OACT/NOTES/as116TOC.html

[3] Booth, H., Maindonald, J., Smith, L. (2002) *Age-time interactions in mortality projection: applying Lee-Carter to Australia*, Working papers in Demography No 85. Research school of social sciences, Australian National University.

[4] Brockwell, P.J., Davis, R.A., (1987) *Time Series: Theory and Methods*, Springer-Verlag New York, Inc.

[5] Brockwell, P.J., Davis, R.A., (2002) *Introduction to Time Series and Forecasting, Second Edition*, Springer-Verlag New York, Inc.

[6] van Broekhoven, H. (2002) *Market Value of Liabilities Mortality Risk: A Practical Model.* North American Actuarial Journal 6, no.2, 95-106.

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

[8] Centraal Bureau voor de Statistiek *StatLine databank*, see: http://statline.cbs.nl/statweb/

[9] Coughlan et al (2007), *LifeMetrics, Netherlands Longevity Index, Technical Document Supplement*, see: www.jpmorgan.com/pages/jpmorgan/investbk/solutions/lifemetrics/library

[10] Cairns, A.J.G., Blake, D., Dowd, K. (2005) *A Two Factor Model for Stochastic Mortality with Parameter Uncertainty*, Working paper.

[11] Coughlan et al (2007), *A toolkit for measuring and managing longevity and mortality risks*, see: www.jpmorgan.com/pages/jpmorgan/investbk/solutions/lifemetrics/library

[12] De Nederlandsche Bank (2005) *Vaststelling methode rentetermijnstructuur FTK*, see: http://www.dnb.nl/openboek/extern/id/nl/pf/40-157597.html

[13] De Nederlandsche Bank (2008) *Statistisch Bulletin September 2008*, see: http://www.dnb.nl/nieuws-en-publicaties/publicaties-dnb/statistisch-bulletin/index.jsp

[14] Donselaar, J., Attema, J.W., van Broekhoven, H.W.M., Roodenburg-Berkhout, L., Willemse, W.J., van Zijp, P.P.C. (2007) *AG; Over Sterfte en Overleven*, Actuarieel Genootschap, Woerden.

[15] van Duin, C., van der Meulen, A., Garssen, J. (2006) *Bevolkingsprognose 2006-2050: model en veronderstellingen betreffende de sterfte*, Bevolkingstrends. Statistisch kwartaalblad over de demografie van Nederland, Jaargang 54-4e kwartaal 2006.

[16] Gompertz, B., *On the nature of the function expressive of the Law of Human Mortality and on a new mode of determining the Value of Life Contingencies*, Philosophical Transactions Royal Society 36, pp. 513-585, W. Nicol, London, 1825.

[17] Gregorkiewicz, F. (2006), *Stochastic Modelling of Mortality Risks*, Master Thesis at the University of Amsterdam.

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

[19] Kalman, D., *A Singularly Valuable Decomposition: The SVD of a Matrix*, The College Mathematics Journal, Vol. 27, NO. 1, January 1996, www.american.edu/academic.depts/cas/mathstat/People/kalman/pdffiles/svd.pdf

[20] Lee, R.D., Carter, L.R. (1992) *Modeling and Forecasting U.S. Mortality*, Journal of the American Statistical Association, 87, 659-671.

[21] Lee, R.D., Miller, T. (2001) *Evaluating the Performance of the Lee-Carter Method for forecasting Mortality*, Demography, Vol. 38, No. 4., pp. 537-549.

[22] Ljung, G.M., Box, G.E.P. (1978) *On a measure of lack of fit in time series models*, Biometrika (1978), 65,2, 297-303.

[23] Makeham, W.M., *On the Law of Mortality*, Journal of the Institute of Actuaries, Vol. 13, pp. 335-340, 1867.

[24] Quandt, R.E., *Some Basic Matrix Theorems*, http://www.quandt.com/papers/basicmatrixtheorems.pdf

[25] Rawlings, J.O., Sastry, G.P., Dickey, D.A.,(1998) *Applied Regression Analysis: A Resarch Tool, 2nd edition*, Springer-Verlag New York, Inc.

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

[27] Willemse, W.J. (2001), *Computational Intelligence: Mortality models for the actuary*, PhD dissertation at the TU Delft.

[28] Wilmoth, J.R., *Computational Methods for Fitting and Extrapolating the Lee-Carter Model of Mortality Change*, Technical report. Department of Demography. University of California, Berkeley.

[29] Wilmoth, J.R., *Demography of longevity: past, present, and future trends*, Experimental Gerontology, Volume 35, Issues 9-10, December 2000, Pages 1111-1129