Quantifying Longevity Risk

Stochastically modeling uncertainty in future mortality and its effects on PFZW’s liabilities

Master Thesis

Evert Roobeek

Eindhoven University of Technology
Department of Mathematics and Computer Science

PGGM
Actuariele Verantwoording en Analyse
Actuarieel Advies en ALM

Supervisors TU/e: prof. dr.ir. Onno Boxma, prof.dr. Johan van Leeuwaarden
Supervisors PGGM: drs.ir. Marco van der Winden, drs. Job Stigter

July 2013
Acknowledgements

This thesis is the culmination of a year of research and a longer writing process. To arrive at the end of this process, episodes of writer’s block had to be overcome repeatedly and for succeeding in surmounting these hurdles, I owe a debt of gratitude to my supervisors, friends and family.

The research for this thesis was conducted during an internship at PGGM in Zeist in the departments of Actuariële Verantwoording en Analyse and Actuariaat en ALM. My colleagues provided me with a superb environment to work and learn about the actuarial issues confronting us today. Also, the atmosphere they created was one of great sympathy, conviviality and humor. In particular, I would like to thank my supervisors Marco van der Winden and Job Stigter for giving me the opportunity to conduct the final project for my master programme in their departments, for introducing me to the leading experts in the field and for patientely and constructively helping me to formulate an assessment of the results I obtained during our collaboration.

After my internship at PGGM, I spent a long time trying to perfect the research, to improve the conclusions and to bridge the gap between actuarial science and mathematical statistics. My supervisors at the department Eindhoven University of Technology, Onno Boxma and Johan van Leeuwaarden, were essential in demarcating the extent of my research and putting an end to my indefinite doubts. I thank them profoundly for their enduring, constructive support, mathematical suggestions and patience. I would also like to thank Jan van der Wal for casting his clear analysis over the preliminary results of my research and Koo Rijpkema for participating in my graduation committee.

Without the continued support, good-natured ribbing and well-organised weekends of diversion provided by my friends, the preceding years in Eindhoven would have been rather more tedious. Gentleman of “In Vino Veritas” and Batser, thank you for the unforgettable moments of laughter, evenings of joy and weekends of seemingly unbounded culinary delights and amusement.

Lastly, I would like to thank my family for their patience, care, love and support during the past difficult years. Sadly, in the time I spent contemplating the quality of my research and writing this thesis, our family lost my grandfather Jan Klaver and my uncle Michiel Klaver. Jan lived a rich and fulfilled life, had a famous sense of witty, Amsterdam-style humor and inspired me to make sure to enjoy life between pondering the great problems of the world. Michiel inherited his witty sense of humor and lived a successful and enterprising life, but unfortunately left us far too soon. He demonstrated to me the value of perseverance. I dedicate this thesis to their memory.

Evert Roobeek
July 2013
Contents

Acknowledgements 3

Contents 5

Introduction 7

Part I Stochastic mortality models 10
  1. Stochastic models to explain uncertainty in mortality 10
  2. Historic data of Dutch mortality 11
  3. Stochastic model based on Poisson distributed mortality 15
  4. Lee-Carter type models 16
     a. Lee-Carter model not sufficient 16
     b. Renshaw-Haberman add birth year to the equation 19
     c. Lee-Carter with cohorts 22
  5. Cairns-Blake-Dowd type models 25
     a. Original Cairns-Blake-Dowd model 25
     b. Extension of CBD with second-order effect for age 27
     c. Extension of CBD with cohorts and second-order effect for age 30
  6. Evaluation of results 34
  7. Prediction 36
     a. Life expectancy accelerates according to Renshaw-Haberman 36
     b. Cairns-Blake-Dowd models give consistent predictions 37
     c. Choice of basis of extrapolation can affect trend 39
  8. Conclusions 42

Part II Longevity Risk 44
  1. What is longevity risk? 44
  2. Sensitivity of a pension fund’s liabilities to longevity 44
  3. Mathematical definitions of risk 47
     a. Value at Risk 48
     b. Conditional Value at Risk 48
  4. Longevity risk calculated from stochastic models 49
     a. Risk for total population 49
     b. Distribution of risk over age and time 50
     c. Risk per birth cohort 52
     d. Pensioners carry a relatively small proportion of the longevity risk 54
  5. A non-stochastic measure of longevity risk 55
  6. Conclusions 56

Part III Alternative Scenarios 57
  1. New trends 57
  2. Simulations suggest linear relation 57
  3. Inverse model 59
  4. Alternative scenarios 60
     a. World champion scenario 60
     b. Americanization scenario 61
  5. Conclusions 62

Conclusions 64
Appendix A  Reflections and suggestions for further research

1. Mortality rate and the probability of dying
2. Likelihood distributions
   1. Binomial distribution
   2. Poisson distribution
3. Parsimonious regression models
4. Time series modeling
5. Cross-correlations between time series
6. Time series prediction
7. Model validation
   1. No validation period for time series
   2. Goodness-of-fit

Bibliography
Introduction

Ever since decent statistics on longevity are gathered, a generally increasing trend can be seen in life expectancy. Large deviations to that trend are seen in the early 1920’s as a result of the Spanish flu and in the 1940’s as result of the Second World War.

The evolvement of women’s life expectancy in various countries

With increasing longevity come larger pension liabilities. Since 2004 the mortality projections contain trends that emulate the increasing longevity, and pension funds use these to adjust the size of their liabilities.

PGGM administretes collective pension schemes. Its largest client is the pension fund for the healthcare and social work sector, Pensioenfonds Zorg en Welzijn (PFZW), which had assets worth approximately € 95 billion in 2010.

One of the most important factors in the estimation of a pension fund’s liabilities is the prediction of mortality. The changing views in the projections of life expectancy in the future have forced numerous Dutch pension funds to adjust the size of their liabilities and consequently their funding ratios. Funding ratios are central to the decision whether or not to adjust pension contributions and benefits.

Uncertainty in mortality projections directly translates to uncertainty about the size of liabilities, which in turn can have major effects on the pension fund’s policy. Therefore, it is necessary to quantify that uncertainty.

Longevity risk

The risk associated with uncertainty in mortality projections is called longevity risk. Based on the size of the corrections for mortality improvement that were made, it is possible to assume that the longevity risk is large. The possible size and the inherent long-term nature of the risk make it natural to be hedged by intergenerational solidarity, a corner stone of the Dutch pension system. Solidarity
between young and old forms the basis for the risk-sharing, cost-sharing and yield-sharing that underpins the Dutch pension system. However, before decisions can be made on how to spread the risk over all generations, we need an accurate estimate of longevity risk, not just a vague sense.

To quantify longevity risk, an extension is needed on the models that are put forward by the Dutch expert institutions Central Bureau of Statistics (CBS) and the Actuarial Society (AG). Their publications of mortality forecasts come without margins of error, while it is the uncertainty of these forecasts that induces longevity risk.

In order to allow for uncertainty, mortality is modeled as a stochastic variable dependent on age and time. The most well-known stochastic model for time-varying mortality was introduced by Lee & Carter (1992). More recently, an alternative model was published by Cairns, Blake and Dowd (2006). These two models and a number of variations on them were applied to Dutch historical mortality data in the research for this thesis. The models and the results of fitting them to the data are described in more detail in part I.

Although with probability one none of these models will correctly estimate future mortality, the stochastic uncertainty and the spread in observed trends form a basis on which we can calculate longevity risk. The approach followed is described in more detail in part II. There are various ways in which risk can be defined. Two variants, Value at Risk and Conditional Value at Risk, are defined in part II and calculations of longevity risk are made using these two definitions of risk.

The stochastic models applied in part I all come with the assumption that future changes in mortality will generally continue along the same lines as they have in the past. This assumption is arguably false. It is therefore proposed in part III of this thesis that it may be interesting to observe alternative scenarios in which mortality changes substantially. One scenario that was examined is that The Netherlands would experience a rise in life expectancy until 2050 to the projected level of the highest national life expectancy in the world, dubbed the ‘world champion scenario’. In the second scenario, the Dutch life expectancy converges to that of the United States.
Part I
Stochastic models for mortality

1 Stochastic models to explain uncertainty in mortality

The Central Bureau of Statistics (CBS) and the Actuarial Society (AG) both project mortality in a deterministic way. They estimate the expected mortality for a certain age in a certain year: a trend.

The perspective from which CBS explains and predicts mortality is by looking at causes of death. It has a vast amount of data on this subject. Trends for prediction are based on expert opinion. The AG applies an algorithm to smooth the CBS data and extrapolates for prediction. This way of modeling mortality doesn’t allow for the type of fluctuations that can be seen in historic data.

The fact that they only provide a trend means that not much is known about the uncertainty of their estimate, or their reliability. The perceived reliability of the CBS and AG estimates follows from the reputations of these institutions. But since they started estimating the longevity trend in 2004 CBS has had to readjust it both in 2006 and in 2008 and quite substantially. AG has only published one revision of its 2005 prediction with trend so far, also containing substantial readjustment.

As can be seen in the figure above, adjustments are quite large. In two years time, the projected life expectancy for 65 year olds, both male and female, have risen one year. Another observation is that in the AG projection, the trends for male and female life expectancy seem to be converging stronger than in both CBS estimates.

In financial terms, uncertainty is very strongly related to risk. Because there is no such thing as a free lunch, profit is usually only made when risk is involved. So not all risk needs to be mitigated. However, with longevity there is a risk of bad projections, but no strategic gain can be made by a pension fund if people die sooner than expected. In other words, there is no pay off. Therefore, longevity risk is a risk that one wants to mitigate. A stochastic model for mortality provides a structure for handling uncertainty and gives a basis for calculation of risk.
2 Historic data of Dutch mortality

The available data consist of tables of deaths and population size. Considering that the data is organized in time periods of one year and age brackets of one year as well, time and age will be treated like discrete variables. The time variable \( t \) ranges from 1950 to 2008 and by \( t=1950 \), the entire calendar year 1950 is addressed. Similarly, the age variable \( x \) only attains integer ages.

Now let \( D(t,x) \) denote the number of people that died in calendar year \( t \), who were aged \( x \) on January 1\(^{st}\), and let \( N(t,x) \) be the number of people aged \( x \) at the beginning of calendar year \( t \). The mortality rate \( m(t,x) \) can then be defined as the share of the people aged \( x \) at the beginning of the year \( t \), that die that year.

\[
m(t,x) = \frac{D(t,x)}{N(t,x)}
\]

To get an idea about how the mortality develops over time, the relative difference to the time-averaged mortality rate is shown in heat maps on the next page.

Three effects can clearly be observed. In both graphs, a red peak is located in the lower left corner. This indicates high child mortality in the early 1950s. These peaks can be explained by the fact that the children aged under 10 years old between 1950 and 1955 were born in or just after the Second World War and saw their already vulnerable health weakened further by the war circumstances.

Secondly, we see that both figures exhibit a darker shade of blue on the right hand side. This indicates decreasing mortality over time across all ages. Thirdly, diagonal lines of higher than average mortality can be observed. They correspond to birth years 1920 and 1946, leading to the conclusion that these effects are due to the Spanish flu and the Second World War, in particular the famish in the winter of 1945.

A fourth observation that can be made when comparing the two figures above is that in the 1950s, women aged 65 to 90 years old had a relatively high mortality rate with respect to later periods. This isn’t quite so much the case with men. This doesn’t mean that women of that age had lower survival probabilities than men. It is to be interpreted that age group 65 to 90 experienced a steeper decrease in mortality for women then it did for men.
Deviation as a percentage of the mortality rate to the average mortality rate over time for Dutch men aged 0-98 in the period 1950-2008.

Deviation as a percentage of the mortality rate to the average mortality rate over time for Dutch women aged 0-98 in the period 1950-2008.

The natural logarithm of the mortality rate is called the *force of mortality* (although no force in the physical sense of the word is meant to be implied). Depicting the force of mortality is usually preferred over depicting the mortality rate because it is more insightful.
Again, some striking differences between men and women can be seen. Firstly, the observation that mortality in the age group 65 to 90 years old decreased more for women than for men is corroborated by comparing the spread of the lines in the two graphs. In the figures above, a shift downwards relates to a decrease in mortality. At advanced ages, the spread is much larger for
women than for men, indicating that mortality of older women has decreased at a faster pace than mortality of older men.

Secondly, women under 40 are seen to be less likely to die than their male counterparts. The difference is most obvious around the age 20. There is a ‘bump’ in the mortality rate for men that age and not for women. This effect may be explained by a difference in the prevalence of risk-taking behaviour between young men and women.

Especially for men, but also for women, the most significant decrease in mortality is seen in the age groups younger than 30. However, the effect on overall life expectancy will be very small, because the number of deaths at those ages is very low. Therefore, a large relative change can come from a small absolute change.
3 Stochastic model based on Poisson distributed mortality

A straightforward way to set up a stochastic model for mortality is to define \( p(t, x) \) as the probability that a person aged \( x \) at beginning of year \( t \) dies that year. Then \( D(t, x) \), now a stochastic, is binomially distributed with parameters \( p = p(t, x) \) and \( n = N(t, x) \).

\[
D(t, x) \sim \text{Bin}(N(t, x), p(t, x)), \text{for all } t \text{ and } x
\]

The population size \( N \) is available in the CBS Statline database, but the probability \( p \) is unknown. Once \( p \) can be fixed, there is a way to simulate \( D(t, x) \), the number of deaths. The best, and most logical, estimate for \( p \) is the mortality rate \( m \). However, assigning a probability to each year and age bracket independently of one another is unrealistic. If a very big accident causes a relatively high number of teenagers to die, one would rather conclude that they had a severe stroke of bad luck than that they already had a higher probability of dying. Also, it leaves no room for prediction of future mortality.

So consider \( p(t, x) \) to be \( ED(t, x)/N(t, x) \), with \( ED \) the expected number of deaths. Once \( ED \) is fixed, the binomial distribution of \( D \) is entirely specified. However, before actually assigning some value to \( ED(t, x) \) in any way, a simplification can be made. A binomial distribution that has a fixed mean, as \( n \) tends to infinity can be approximated by a Poisson distribution with the same mean. The reason for this approximation is that it simplifies calculations.

\[
D(t, x) \sim \text{Poisson}(ED(t, x))
\]

The only thing left to do is to fix the parameter \( ED(t, x) \), the essential part of the model. In the following chapters, various regression models are fitted to the observed historical data \( D(t, x) \). All models under consideration were fitted by maximum likelihood estimation (MLE) methods. Because all models are based on the Poisson assumption, the likelihood that is to be optimized is deduced from the Poisson distribution. If \( \alpha \) denotes the ensemble of parameters of the stochastic model, then the log-likelihood of the estimation is given by the following equation.

\[
\ln L(\alpha; D(t, x), N(t, x)) = \sum_{t,x} \left( D(t, x) \ln \left( m(t, x; \alpha) N(t, x) \right) - m(t, x; \alpha) N(t, x) - \ln(D(t, x)) \right)
\]

After the maximizing likelihood estimation, the residuals will be examined to assess the goodness-of-fit. The numbers of deaths are assumed to be independently Poisson distributed. This also means that the standard deviation is assumed to be equal to the Poisson rate parameter \( \hat{m}(t, x)N(t, x) \), where \( \hat{m} \) is the modeled mortality rate. In order to compare the residuals, they are standardized. According to the central limit theorem the standardized residuals are asymptotically standard normally distributed.

\[
Z(t, x) = \frac{D(t, x) - \hat{m}(t, x)N(t, x)}{\sqrt{\hat{m}(t, x)N(t, x)}} \sim N(0,1)
\]

Note that calculating residuals using the Poisson distribution’s variance results in slightly smaller residuals, because the variance of the binomial distribution is slightly smaller than that of the Poisson approximation.
4 Lee-Carter type models

In 1992, Ronald Lee and Lawrence Carter introduced a stochastic model to forecast U.S. mortality (Lee & Carter, 1992). It has since then become a well-known stochastic model among actuaries. Since then, their model has been generalized to a family of Lee-Carter type models. The models are based on a regression on the logarithm of the mortality rate.

\[ \ln m(t, x) = \sum_{i=1}^{p} \beta_i(x)\kappa_i(t)\gamma_i(t - x) \]

The \( \beta_i \) are the parameter vectors dependent on age, the \( \kappa_i \) depend only on time and the \( \gamma_i \) are cohort parameters. Cohorts are introduced to model generations and are defined by the birth year \( t-x \) for all \( t \) and \( x \) under observation.

Although the general form presented above doesn’t imply a preference for one variable over the other, Lee-Carter type models are based on the perspective of age being the primary factor for mortality and time the secondary. This will be seen in more detail when the original Lee-Carter model will be introduced in detail.

a. Lee-Carter model not sufficient

Starting out with the Poisson process model and the data on the number of deaths \( D(t, x) \) and the size of the population \( N(t, x) \) we intend to estimate the mortality rate \( m(t, x) \).

A linear regression is applied on \( \ln(D(t, x)/N(t, x)) \). The model for the mortality rate consists of an age profile \( \beta_1(x) \) and a time dependent improvement of mortality \( \kappa(t) \), whose weight depends on age. Although this second age dependence seems a bit superfluous, it is needed to model the effect that for certain ages the improvement over time can be stronger than for others. This effect was seen to be especially significant for women.

\[ \ln m(t, x) = \beta_1(x) + \beta_2(x)\kappa(t) \]

To ensure that the parameters \( \beta_1 \), \( \beta_2 \) and \( \kappa \) are uniquely determined, two further constraints need to be introduced. By Lee and Carter, these are chosen to be

\[ \sum_{x} \beta_2(x) = 1 \]
\[ \sum_{t} \kappa(t) = 0 \]

These constraints imply that \( \beta_1(x) \) is the average force of mortality, which can be seen by summing the equation defining the model over \( t \) and applying the constraint on \( \kappa(t) \).

To predict future mortality rates, \( \kappa(t) \) is modeled as a time series. In most articles the used model is a random walk with drift. This means that the differences between subsequent values of \( \kappa(t) \) are assumed to be independent and normally distributed.

\[ k(t) - k(t - 1) \sim N(\mu, \sigma^2) \]
Here, $\mu$ is the drift parameter, that explains the trend in the time series. The parameter $\sigma$ captures the variability around the trend. This random walk model is used to predict future values of $\kappa(t)$, which are in turn inserted in the mortality rate to predict future mortality.

Modeling mortality in this way means that the assumption is made, that the baseline age-dependent mortality rate, $\beta_1(x)$, never changes in time and is as such a sort of biologically determined, inherent, age-dependent mortality rate. From this perspective, humans have always had a baseline probability to reach the age 100 equal to that of people living now, but the actual probability of reaching the age of 100 was diminished by the circumstances at the time. It also means that $\kappa(t)$ captures anything from medical progress to environmental changes.

*Results of the Lee-Carter model*

The results obtained from fitting the Lee-Carter model to the CBS data clearly mark the differences between men and women. To show this, parameters for men and women are presented in one combined graph.

The fitted parameter vectors $\beta_1$ (left) and $\beta_2$ (right) for men (blue) and women (green)

The upper two figures corroborate results that were obtained earlier when initial observations were made about the data. On the left we see that women have a lower average mortality than men. On the right we see again that the decrease in mortality is strongest for women between the ages of 65 and 90, as $\beta_2$ models the dependence of the decrease in mortality on age.
The lower figure adds to earlier observations that apparently, the decrease in mortality began later for men – only in the 1970s – than it did for women. Furthermore, the decrease in male mortality seems to be accelerated, while the decrease in female mortality has been going steadily at more or less the same pace since 1950.

The observations made from the parameter values corroborate much of the earlier observations that were made at first glance of the historic data. However, one would like to explain as much as possible of the mortality, even if it goes unnoticed at first glance. To see whether nothing has been missed, an assessment of the goodness-of-fit is made. If this model explains Dutch mortality well, we would see no apparent structure in the standardized residuals and these would be standard normally distributed.

![Standardized residuals of Lee-Carter model fitted on mortality data for Dutch men](image)

![Standardized residuals of Lee-Carter model fitted on mortality data for Dutch women](image)

Obviously, some structure is left in the residuals of the Lee-Carter model, both for men and women. If the fit would have been good, then the residuals would have been approximately standard normally distributed and values would hardly ever exceed 3. Clearly, this is not the case. The diagonals represent cohort effects. Extapolating, these diagonals can be linked to individuals that were born in 1920 and in 1946. As mentioned before, an explanation for these effects can be found
in the Spanish flu and the Second World War. Additionally, for men with birth year before 1890 cohort effects are seen as well, but these are spread across so many years that no particular event or disease can explain them.

b. Renshaw-Haberman add year of birth into the equation

As we have seen in the results of the Lee-Carter model, a clear generation effect persists for cohorts 1920 and 1946. There are several ways to add a parameter that helps explaining and modeling the cohort effect. Renshaw & Haberman (2005) chose to add a term to the regression that enables weighing the cohort effect with age. This is the variant that is considered here.

\[ \ln m(t, x) = \beta_1(x) + \beta_2(x) \kappa(t) + \beta_3(x) \gamma(t - x) \]

An additional normalization has to be introduced to ensure uniqueness of the model parameters, so \( \gamma \) is normalized to have mean 0 and \( \beta_3 \) sums to 1. When fitting this model, the cohort parameter for the first 5 and the last 5 birth years were put at zero before normalization. This was done because there is too little data available for an accurate estimation of these parameters.

Seeing that both the cohort effect and the general decrease in mortality are weighed with age parameter, this model may explain some of the decrease in mortality through the cohort effect, by means of low values for young cohorts.

\[ \beta_3(x) \text{ (left)} \] and \[ \gamma(t-x) \text{ (right)} \] for men

\[ \beta_3(x) \text{ (left)} \] and \[ \gamma(t-x) \text{ (right)} \] for women
A clear peak of the cohort parameter can be seen for 1946. The peak for 1920 seems small because the value for surrounding years is also high. This suggests that some effects of decreasing mortality are translated onto the cohort. In other words, according to these results some of the decrease in mortality comes from the existence of strong birth cohorts.

In order to check whether the two models really explain the decrease of mortality over time differently, we compare the parameters that the Renshaw-Haberman model has in common to Lee-Carter and see whether there is a significant difference.

The difference in the age profile parameter $\beta_1(x)$ between the two models is negligible. This points to the conclusion that at least this part of the model is stable and hence the major part of the dependence on age is modeled correctly.
The hypothesis that was made when first interpreting the cohort parameter values - that there is some interaction between the decrease in mortality and the cohort effect - seems to be supported by the values for the parameter $\kappa$ and $\beta_2$. Especially for men a remarkable difference can be observed between the two models. Ideally these differences should be very small, because that would imply stability of the model.

Disregarding the possible impurities of the Renshaw-Haberman model, it does eliminate the cohort effect almost completely as can be observed in the standardized residuals.
Clearly, the cohort effect is explained by the model, although the values are somewhat biased to the positive for men and to the negative for women. This is due to compensation for a number of outliers that are located from 1975 to 1980 for 98 year olds. These outliers are result of the high volatility of mortality at very high ages.

A consequence of the observed phenomenon of an exchange of weight between decreasing mortality and the cohort effect is that the Spanish flu effect is not completely eliminated. A better way of modeling the cohort effect could be to lose the dependence on age, in other words eliminate $\beta_3$ by setting it to 1.

c. Lee-Carter with cohorts

When setting $\beta_3$ equal to 1, the option of varying the impact of the cohort effect over age is cancelled and what remains is a cohort parameter that expresses more precisely the cohort effect due to specific birth years.

$$\ln m(t, x) = \beta_1(x) + \beta_2(x) \kappa(t) + \gamma(t - x)$$

As can be seen in the figure below, a clear correlation between adjoining birth years still exists in the simplification of the Renshaw-Haberman model. Because of the parallel estimation, all interdependence between the time-dependent decrease of mortality $\kappa(t)$ and the cohort parameter $\gamma$ cannot be removed. Especially with men, the cohort parameter for older cohorts adjusts for the change in $\kappa$. However, the peaks for 1920 and 1946 show much clearer.
Looking at the residuals, we see that the problematic fit for the highest ages in the period 1974-1984 has not improved with respect to the Renshaw-Haberman model. But if those values are disregarded as outliers, the residuals seem to be quite well balanced close to and around zero. There seems to be no remaining structure in the residuals.
Standardized residuals of the Lee-Carter+cohort model for Dutch women
5 Cairns-Blake-Dowd models

In 2006, Andrew Cairns, David Blake and Chris Dowd introduced an alternative to the Lee-Carter type of models. The dominant parameter vector models the decreasing mortality and the dependence on age is modeled in advance instead of estimated as part of the complete model (Cairns, Blake & Dowd, 2006). In these type of models a so-called logistic regression is applied on a transformation of the probability of dying within one year, \( q(t,x) \).

\[
q(t,x) = 1 - \exp(-m(t,x))
\]

a. Original Cairns-Blake-Dowd

The simplest Cairns-Blake-Dowd model consists of a general pattern of decreasing mortality, \( \kappa_1 \), and then a second time-dependent factor that depends linearly on age. If one remembers the force of mortality graphs, this relates only to mortality over 40 years of age.

\[
\text{logit } q(t,x) = k_1(t) + k_2(t)(x - \mu)
\]

Logit \( q(t,x) \) is the logarithm of the odds of dying, in which the odds mean the ratio between the probability of death and the probability of survival as defined above.

\[
\text{logit } q(t,x) = \ln \frac{q(t,x)}{1 - q(t,x)}
\]

The parameter \( \mu \) is the (non-weighted) average age under consideration in the data and is there to prevent adding age-effects to \( \kappa_1 \).

For the dominant parameter, \( \kappa_1 \), the same shape is seen as in the Lee-Carter model. The scale is different however, exactly because it is now the most important factor. Also, it is good to be reminded of the fact that we are dealing not with the logarithm of the probability of the dying \( m(t,x) \), but with a transformation of it. Therefore, the parameter value may not be compared one-to-one.
Scale is also an issue with the second parameter. If the parameter for the female population were to be drawn in the same graph as the one for the male population, a lot less variation would show because of the larger scale. When fitting trend for the $\kappa_2$ for women, a constant would be a good approximation, which leads to the conclusion that for women in this model the age dependence is approximately constant in time.

The obvious observation to make in the figure containing the standardized residuals of the CBD model for males is that rather serious cohort effects play a role. The well-known bad years of 1920 and 1946 show up again, but this time they are accompanied by positive cohort effects for men born in the 1880s and 1930s.
These larger cohort effects can also be interpreted in a different way. Observe that, looking in the direction of age, we have high, then low and then high values in 1950. This moves gradually to low-high-low in 1980 and then back again to high-low-high in 2008. An alternative to modeling this by cohorts only is to treat this as a second order effect of age that evolves through time. It is not an entirely out-of-the-box suggestion, because for the female data a strong case can be made for this effect.

This model has a number of issues that makes it ill-fitted to the Dutch mortality data. Not only is there a cohort effect, but the deviations are large everywhere due to a certain second-order effect for age that varies with time.

b. Extension of CBD with second order effect for age

Because it is the largest effect, the second order age-effect is the first one to be tackled. A second-order term is added, which looks rather involved because again there is a correction for the average, \( \sigma^2 \). It is applied so the parameter vector \( \kappa_4 \) regulates only the time-varying nature of this factor.

\[
\text{logit } q(t, x) = \kappa_1(t) + \kappa_2(t)(x - \mu) + \kappa_4(t)((x - \mu)^2 - \sigma^2)
\]

This extension of the model was introduced by Cairns, Blake and Dowd in a later article (Cairns, et al., 2009). To see whether the model is stable, we first compare parameters with the original CBD-model. The dominant parameter vector \( \kappa_1 \) is more or less stable, especially for men.
For men, even the second most important parameter hardly differs from the values obtained for the original model. However, there is an obvious divergence in the female graph, which should mean that second order effects are stronger for women in the last decades of the observed period.
It must be remarked that although the scale of these graphs might lead one to believe that this parameter were insignificant, these values are multiplied by the square of age minus some correction numbers, which are values up to 550. The values for men are markedly smaller than for women, but obey the model assumption of a gradually moving second order effect that returns to its original form after 1980. Furthermore, an increasing trend is observed for women throughout the time period, which means that the age dependence steadily becomes more non-linear.

In the standardized residuals we should only see a cohort effect in the ideal case. Alas, this case isn’t ideal. This addition only partly solves the problem. Some of the targeted deviations persist. This can be a consequence of centering the second order term around the average under observation, 69, as opposed to the age at which the deviations become most negative, around 75.
c. Extension of CBD with cohorts and second-order effect for age

After including the second order age effects in the CBD model, of course the cohort effects persist and still have to be addressed. Observing the figures of the standardized residuals, the addition of a cohort effect that can vary in age or time is an attractive idea, but one that is not pursued because of instability issues. The model already uses a large number of parameters and just adding a constant cohort effect will increase that significantly. Adding a time-varying cohort effect results in an instable model that has convergence issues as well as problems with the significance of parameters. Therefore only a constant cohort effect is added to the model in the form of $\gamma$.

$$\logit q(t, x) = \kappa_1(t) + \kappa_2(t)(x - \mu) + \gamma(t - x) + \kappa_4(t)((x - \mu)^2 - \sigma^2)$$

With the cohort effect being constant in time and age, it is assumed that the cohort effect is a fixed (dis-)advantage incurred at birth. On average over the generation of people born in the same year, the effect remains the same for one’s entire life.

Like in section 5.2, results are compared to the simpler versions of the model to be able to assess the stability of the parameters as an indication of the stability of the model.
The blue and green lines are hardly discernible, so the general decrease in mortality is estimated almost equally by the CBD models that have a second order factor for age. Adding the cohort effect has hardly had an impact on this parameter vector, which suggests that the general decrease of mortality as defined in these models, was already correctly estimated by the model without cohort effect.
Observing the graphs for the parameters that define the evolution of dependence on age through time together, the small differences of the cohort model with respect to the others are repeated, although maybe the level of $k_4$ for women has lowered a little. The fact that all $k$ are affected a little at most by adding the cohort parameter suggests that cohort factor is an independent factor to the rest. This is in line with the assumption that the cohort effect is incurred at birth.

As earlier results indicate that the cohort factor is independent of the others, the cohort parameter $\gamma$, can be interpreted as the deviation of a generation with respect to the general mortality. The peaks in 1920 and 1946 are genuine peaks. Also, elevation can be seen in the period 1940-1944. Good cohorts to be in are the youngest ones, who on top a general mortality decrease have a negative cohort affect. On either side of both graphs there are fixed flat lines, because the cohort parameter could not be calculated correctly, due to the small amount of data on those cohorts.
Looking at the standardized residuals, a structural deviation to model can still be seen. That was to be expected, because this remaining structural deviation, the yellow and blue bands, wasn’t addressed by adding a cohort parameter. The cohort effect is however completely eliminated in the female data and almost entirely for men. For men there seems to have occurred a shift in the cohort effects for cohorts in the 1870s. This is due to the abnormally high mortality some of these cohorts experienced at ages near 100, that are not in line with the relatively low mortality exhibited in their 70s, during the Second World War.
6 Evaluation of results

To compare the models in a somewhat less subjective manner than the interpretation of results, model selection criteria can be used. Selection criteria are used to compute a ‘score’ for each model. The model with the lowest score is the best according to that selection criterion. The criterion that will be used here is the Schwarz-Bayes information criterion (SBC) (Schwarz, 1978).

\[ SBC(\alpha) = -2 \ln L(\alpha; D(t, x), N(t, x)) + |\alpha| \ln nT \]

The value of the criterion depends on the log-likelihood of the estimated model, the number of parameters \(|\alpha|\), the number of age groups \(n\) and the length of the time period \(T\). The extension of a model with extra parameters increases the criterion value, but this increase can be compensated if the likelihood increases as well. In this way, the Schwarz-Bayes information criterion helps strike a balance between likelihood optimization and overfitting.

<table>
<thead>
<tr>
<th>Model</th>
<th>Men</th>
<th>Women</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lee-Carter</td>
<td>23256</td>
<td>19490</td>
</tr>
<tr>
<td>Renshaw-Haberman</td>
<td>18518</td>
<td><strong>18048</strong></td>
</tr>
<tr>
<td>Lee-Carter+cohort</td>
<td><strong>18356</strong></td>
<td>18066</td>
</tr>
<tr>
<td>Cairns-Blake-Dowd 1</td>
<td>23346</td>
<td>43363</td>
</tr>
<tr>
<td>Cairns-Blake-Dowd 2</td>
<td>21096</td>
<td>21504</td>
</tr>
<tr>
<td>Cairns-Blake-Dowd 3</td>
<td>18780</td>
<td>19721</td>
</tr>
</tbody>
</table>

Schwarz-Bayes selection criterion value

For men the lowest Schwarz-Bayes criterion value is attained for the Lee-Carter model with a simple cohort extension and for women the Renshaw-Haberman model scores best. According to this criterion, additions to a model, whether it be Lee-Carter or Cairns-Blake-Dowd, mostly lead to an improvement that was worth the extra parameters. The only exception is the better fit of the Lee-Carter+cohort model for men, which contains fewer parameters than Renshaw-Haberman.

However, even the twice extended CBD model was worse than the simple Lee-Carter model for women. This is a consequence of the problem with the Cairns-Blake-Dowd models for women, which is that the interaction between decreasing mortality with time and increasing mortality with age is not captured correctly by the basic model. A problem that the Lee-Carter model doesn’t have. This inability of the Cairns-Blake-Dowd model was imposed on it by assumptions that are made about the age effect. Lee-Carter hasn’t got the burden of needing the age dependence to be linear like in the standard CBD model, or quadratic like in the extensions of the CBD model. This freedom allows for a better fit, at the cost of more parameters.

<table>
<thead>
<tr>
<th>Model</th>
<th>Men</th>
<th>Women</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lee-Carter</td>
<td>19,7%</td>
<td>8,0%</td>
</tr>
<tr>
<td>Renshaw-Haberman</td>
<td>3,2%</td>
<td>2,9%</td>
</tr>
<tr>
<td>Lee-Carter+cohort</td>
<td>3,1%</td>
<td>2,8%</td>
</tr>
<tr>
<td>Cairns-Blake-Dowd 1</td>
<td>21,3%</td>
<td>57,0%</td>
</tr>
<tr>
<td>Cairns-Blake-Dowd 2</td>
<td>12,3%</td>
<td>15,5%</td>
</tr>
<tr>
<td>Cairns-Blake-Dowd 3</td>
<td>5,1%</td>
<td>9,5%</td>
</tr>
</tbody>
</table>

Percentage of standardized residuals significant at 1% level

The conclusion that can be drawn from the comparison of SBC values can be repeated when observing the percentage of significant differences between the model and the data. The assumption that underlies all models, that the residuals are Poisson distributed and the fact that the
standardized residuals should then be approximately standard normally distributed, is used to construct a measure of goodness-of-fit.

Given a certain confidence level, here chosen to be 99%, any residual that is located in either 0.5%-tail of the normal distribution is considered significant. If indeed the residuals are independent and standard normally distributed, then the number of significant residuals should be approximately 1% of all residuals. As can be seen in the table, none of the models attain this measure, although both cohort extension models of the Lee-Carter type come close.

While the numbers appoint Renshaw-Haberman to be very successful in modeling mortality, it is good to remember the transferring problem it had when it came to the cohort effect. Because of the dependence between the cohort factor and the other factors, the cohort parameter did not only reflect the birth year specific deviation from the trend, but carried some of the trend itself as well. This is a clear disadvantage to this model. A solution is to eliminate $\beta_3$ by setting it to 1 and thereby create a similar cohort parameter as in the third Cairns-Blake-Dowd model. The effect on the goodness-of-fit is small for women and slightly larger for men, but it leads to a more parsimonious model.
7 Prediction

The purpose of fitting models to historic data in this research project is to predict future mortality and give an idea of the uncertainty of this prediction. The various models that were fitted in earlier chapters will lead to a variety of predicted trends as well as varying sizes of statistical confidence intervals. However, the trends and confidence intervals will be shown to be very similar for most models.

In applying any time series model one usually assumes the time series is stationary, or at least can be made stationary by removing deterministic components or differencing. This implies that structural behavior of the time series in the future will be comparable to what has already happened. In other words, the assumption implied by applying a time series model is that mortality will evolve towards the future in a similar way to how it has evolved in the observed period of historic data.

The extrapolation that constitutes the prediction of future mortality is based on the fitting of a trend to the $\kappa$ parameter time series that model the development of mortality through time. In the first presentation of life expectancy projections by CBS and AG in Chapter 1, the modeling choice was made that the trend be based on the last 20 values of the $\kappa$ time series. In other words, a piecewise linear trend was assumed and the last piece was extrapolated for prediction. The effect this modeling choice has on the estimation of the trend and the confidence interval will be touched upon later in this chapter.

Another assumption made while extrapolating is that the cohort effect will tend to zero for cohorts that were not in the data. Because there is no information yet on these cohorts and $\gamma$ is 0 on average, setting new values to zero is the best estimate in theory. However, the question is whether this will fare well with the transferred decrease of mortality to the cohort parameter in the Renshaw-Haberman model.

The confidence intervals shown in the figures are also drawn from the time series prediction and are based on the 5% and 95% quantiles. The intervals are therefore 90% confidence intervals, which means that, according to the time series model and all assumptions that were made in applying it, the true value of life expectancy will lie in these intervals with a probability of 90%.

a. Life expectancy accelerates according to Renshaw-Haberman

The 90%-confidence intervals that were constructed from the time series extrapolation of the models expand rapidly. In 2050 the size of the confidence interval is about 17 years. The difference between the Lee-Carter model and the Renshaw-Haberman model lies in the estimation of the trend, but the statistical accuracy of this estimation in 2050 is almost equal for both models.

Although the cohort including Renshaw-Haberman model seemed to be a better fit than the Lee-Carter model, a problem with the interdependence of model parameters was observed. This also has some apparent consequences on the prediction. As the generations that live in the period of historic data die out, the contribution of their cohort effect to the mortality fades out and eventually life expectancy only varies due to the general decrease in mortality. The significantly steeper decrease in mortality in the Renshaw-Haberman model leads to acceleration in the increase of life expectancy.
Estimation of future life expectancy of Dutch 65 year old men (left) and women (right) according to Lee-Carter (blue) and Renshaw-Haberman (red) and Lee-Carter+cohort (green) with 95%-confidence intervals

The range between the three models’ 2050 predictions is larger for men, 7 years, than for women, 5 years. This shows again that the cohort effect is larger for men than for women. In fact, the rise in male life expectancy according to both cohort extension models of Lee-Carter is so large that it exceeds the confidence interval set by the original Lee-Carter model. This is an indication that the cohort effect must be included and hence that the Lee-Carter model disqualifies as a model for increase of male life expectancy.

b. Cairns-Blake-Dowd models give consistent predictions

The greater consistency of the parameter estimates in the three Cairns-Blake-Dowd variants translates to more consistency in the predictions obtained from these models. The cohort effect, though very obvious in the mortality data, is hardly noticeable in the life expectancy predictions. The small difference between the three trends can be explained by the $\kappa_1$ parameter dominating the trend estimation. Regarding its larger order of magnitude compared to the $\kappa_2$ and $\kappa_4$ parameters, this was to be expected.

Even so, the extension of the model with a second order effect for age improved the fit so much that it led to significantly smaller confidence intervals. In 2050 the difference between the size of the confidence interval of the original Cairns-Blake-Dowd model and its extensions amounts to 4 years that are mostly found in the lower tail. According to the extended Cairns-Blake-Dowd models the trend of increasing life expectancy is a lot less likely to revert.

The size of the confidence intervals of the extended models is approximately 17 years for women, comparable to the size of the Lee-Carter and Renshaw-Haberman confidence intervals. So even though the trend may differ significantly, the amount by which the true life expectancy may deviate from that trend is in the same order of magnitude. An exception must be made for the simple Cairns-Blake-Dowd model, due to the bad fit.
Prediction of future life expectancy of Dutch 65 year old men according to Cairns-Blake-Dowd (blue), with 2nd order age effect (red) and cohort effect (green)
c. **Choice of basis for extrapolation can affect trend**

Estimating the trend is done by extrapolating the linear trend present in the time-dependent parameters. Observing the graphs of the fitted \( \kappa \) parameter values, it can be seen that the choice of the basis of extrapolation has a large influence on the trend estimate. For example, if one were to base the estimate of the trend on the last 2 parameter values, a much steeper trend will be seen, than if one would base the trend on all observations. Taking only the last two values carries the assumption with it that only the last two years are indicative of future developments, while taking all values assumes that decrease in mortality as seen in the post-war period could repeat itself in the years to come.

For its 2008 prediction tables, the Dutch Actuarial Society (AG) fixed its predicted trend on the period 1988-2007, taking the last twenty years of observations as basis for the future development of life expectancy. Another alternative is looking for breakpoints in the \( \kappa \) parameter graphs. Observing the Lee-Carter \( \kappa \) parameter and the Cairns-Blake-Dowd \( \kappa_1 \) parameters for men, one might argue that a clear change of direction takes place between 1975 and 1977, so the period 1976-2008 might be a good basis. Note that this effect is eliminated by the Renshaw-Haberman model, which attributes it to the cohort effect.

To get an indication of the significance of the choice for either 1976 or 1988 as starting point for trend estimation, a comparison of results is made for each of the fitted models. In all figures the 1976-2008 based estimate is colored blue and the 1989-2008 based estimate is colored red.

A few general observations can be made at first glance. Firstly, with female life expectancy hardly any difference is notable, with the exception of the Renshaw-Haberman model and even there the difference is minor. The explanation for this lies in the steady decrease in mortality observed in almost the entire period of historic observation. The decrease was already approximately linear, so a linear extrapolation could be expected to be similar for different starting points.

Secondly, the elevated male life expectancy trend based on 1988-2008 with respect to that based on 1976-2008, expected from the observation of the 1976 breakpoint, shows in all graphs except the Renshaw-Haberman one.
Lee-Carter life expectancy predictions
for 65 year old Dutch men (left) and women (right)

Renshaw-Haberman life expectancy predictions
for 65 year old Dutch men (left) and women (right)

Cairns-Blake-Dowd life expectancy predictions
for 65 year old Dutch men (left) and women (right)
Although it is difficult to discern, in the cases where the trend is different, the variance changes a little as well. The confidence interval for the 1988 starting point based estimate is slightly smaller than the interval based on starting point 1976. The important difference is however the change in trend.
8 Conclusions

Some clear structure in mortality data
A first look at the Dutch mortality data reveals some clear structure. For one, there is an overall decrease in mortality for all ages, mostly for the youngest. Also two generations have clearly had a tougher life than others, people born in 1920 and 1946. These effects can be explained by the impact of the Spanish flu and of the 1945 winter famine. The Second World War also led to high child mortality in the early 1950s. These kids were born in or just after the war and may have experienced food shortages or other life threatening circumstances. Finally, male pensioners have seen a steeper decline in mortality than female elderly.

Stochastic model for calculating with uncertainty
As a way to introduce a type of uncertainty in mortality that can still be used for calculations, the number of deaths is modeled as a binomially distributed stochastic variable, dependent on the population size and a fixed, modeled probability of dying. The assumption of fixed probability leads to the approximation of the stochastic number of deaths by a Poisson distributed stochastic variable, only dependent on the modeled number of deaths. This allows for a simplification of calculations.

The basis of the stochastic model allows for maximum likelihood estimation, in order to obtain a best estimate on the probability of dying and for goodness-of-fit analysis based on standardized residuals.

Lee-Carter model lacks cohort parameter
The best-known actuarial model to explain decreasing mortality is the Lee-Carter model. It is based on an age factor and is extended with a factor for decreasing mortality in time, weighted by age. Although the Lee-Carter model captures the age-dependent decreasing mortality and the 1950s child mortality quite well, the 1920 and 1946 cohort effects are clearly not explained. Some smaller cohort effects in the 1880s show up as well. The model needs to be extended with a cohort parameter.

Renshaw-Haberman explains cohort effect but at a cost
The extension that is proposed by Renshaw and Haberman is age-dependent. This does a good job of eliminating the cohort effect from the residuals. However, it does so by explaining part of the decrease of mortality by low cohort parameter values for younger generations. This obfuscates the distinction between general improvement of living conditions, supposed to be modeled by $\kappa$, and the unlucky cohorts such as 1920 and 1946, supposed to be modeled by the cohort parameter $\gamma$. A solution can be to remove the multiplication of the cohort parameter with an age- or time-dependent factor.

Cairns-Blake-Dowd bad fit in two ways
In contrast to Lee-Carter, the Cairns-Blake-Dowd model is based on the decrease of mortality in time as the most important factor and a secondary factor that multiplies a linear dependence on age with a second time-dependent variable, allowing for varying decrease in mortality for different ages. It shows however that dependence of the used transformation of probability of dying is not at all linear as assumed. The deviations of the linear form vary in time. And just like the Lee-Carter model, the cohort effect is not accounted for.

Two extensions of Cairns-Blake-Dowd insufficient for good fit
In order to try and repair the shortcomings of the original Cairns-Blake-Dowd model, extensions with a time-dependent second-order effect for age and a cohort parameter independent of age are tried. The second-order effect does reduce the deviations to some extent, but a structural deviation in the
residuals is still left. The cohort parameter is successful in eliminating the cohort effect from the residuals. To come to a better fit, the rather tough assumptions on the dependence of age could be weakened. However, this would lead to an increase in parameters and thereby instability and convergence issues for the model.

Lee-Carter with simple cohort extension scores best
Comparing all models on the basis of the Schwarz-Bayes selection criterion, the simple cohort extension provides the best fit for both male and female Dutch mortality. It is slightly better than Renshaw-Haberman. However, by examining the residuals, it shows that none of the models attain a fit good enough to satisfy the condition of independently standard normally distributed residuals. The simple cohort extension of Lee-Carter again comes closest.

Prediction
Although the fit seems to be bad for most models, the predicted trends are similar with the exception of the Renshaw-Haberman model. With trend estimates based on the period 1989-2008, life expectancy of 65 year old men will rise from 82 to approximately 86 years in 2050, for women from 85 to approximately 88 years. This is slightly, less than one year, higher than the most recent CBS estimate.

As result of the Renshaw-Haberman model transferring general decrease in mortality to the cohort parameter, its estimation of the life expectancy increases much faster as older cohorts die out. Both 65 year old men and 65 year old women would on average live to the age of 92 by 2050. That is a significantly higher estimate.

The 95% confidence interval gives indication of the range in which the real value for the life expectancy will lie. For men, the increase projected by the Renshaw-Haberman model exceeds the Lee-Carter model confidence interval while that is not true the other way around, which indicates Lee-Carter is very likely false, according to the stochastic model. However, the fact that Renshaw-Haberman interchanges general decrease in mortality with the cohort effect, combined with the predictive assumption that the cohort effect will tend to zero for cohorts not in the data, is an obstacle to declaring Renshaw-Haberman the best predictive model.

The solution of simplifying the Renshaw-Haberman model gives a more credible alternative of a Lee-Carter type cohort model, but also has a significantly higher life expectancy.
Part II

Longevity risk

1 What is longevity risk?

After accepting that predictions of the longevity trend are inherently uncertain, the next step is to quantify that uncertainty. Traditionally, risk is an insurance related term. It was devised to put a number on how much one would need to cover almost all possible losses. Trivial worst case scenarios in which one loses everything are excluded. No one would take insurance on car if he would have to pay the entire value of the car twice.

Hence, intuitively, a definition of risk would be the amount of money one stands to lose in bad scenarios, or equivalently, the amount of money that could cover most losses. However, this definition of risk is hardly precise, so in this part of the thesis a number of different risk measures will be introduced. Most of them will be statistical risk measures, which can be applied on the stochastic models that were introduced in part I.

2 Sensitivity of a pension fund’s liabilities to longevity

The part of an individual’s pension, that is most important for the pension fund, is the pension amount that is built up as a consequence of participants paying pension premiums. The amount of euro’s that the participant can expect to add to his pension amount based on his yearly payment of premiums, is subject to yearly negotiations between employers and trade unions. Typically, pension funds try to adjust pension rights for inflation. The fact that these pension agreements are renegotiated each year means that pension rights are not set in stone, but in practice they are only adjusted in the very extreme of circumstances.

The grand total of all pension rights of a fund’s participants constitute the fund’s liabilities. The amount of euro’s that will be needed to be able to pay out all these pension rights is called the Voorziening Pensioenverplichtingen (VPV) or provision for pension liabilities. Important to know is that this number is calculated only with data of current participants of the pension fund: no attention is paid to possible future participants and their pension rights.

The calculation of the VPV is based for a large part on the mortality table, which estimates future mortality rates in order to correctly predict how long the pension fund is obligated to pay out pensions to its current participants. In the case of the PFZW, this is the most recent mortality trend published by AG, corrected for the population of the pension fund.

Most of the pension rights belong to women. The difference in pension rights built up is mostly made before retirement. In fact 90% of male pension rights are located with men between the ages of 45 and 75, whereas 90% of female pension rights are with women between the age of 35 and 70. This suggests that younger women build up more rights than younger men, possibly due to the characteristics of the pension fund’s population.
Sensitivity to mortality
However, it is not a given that as most of the pension rights lie with people in their fifties and sixties, also the longevity risk will be situated among them. To investigate this, we have to examine the impact of lower mortality on the VPV. In order to do this we observe what a relative decrease in mortality of 1% adds to estimated liabilities.

Depicted below is a heatmap of the impact of a decrease in mortality on the VPV. More precisely, for every age between 15 and 99 and for every year from 2009 to 2093, it was calculated how much pension rights (in euro) would have to be added to the VPV if for people of a given age in a given year, mortality would decrease with 1% relative to the mortality table on which the VPV is based. So the heatmap doesn’t show the effect of a general decrease with 1%, but rather the specific effects of a pinpointed decrease of mortality. In the picture, blue pixels mean very low impact and red pixels mean high impact.

The lower triangle is completely blue. A decrease in mortality there would have no impact on the VPV, because people who have not yet reached the age of 15 cannot be participants of the pension fund. Also, one can see that there is very little impact beyond 2050 and hardly any beyond the year 2060. This is because the only people who are now participants and will still be alive by then, are now young people, who haven’t yet built up a large amount of pension rights.

Furthermore, beyond the age of 90, no real effect seems to come of a change in mortality. Although they may have large pensions, their numbers are small, so if they live a couple of months longer on average – and the rest of the population doesn’t – no large adjustments will need to be made by the pension fund.
Very clearly, the red area is much larger for women than it is for men. The fact that more women are participants than men, explains this. Most of the impact seems to lie with men currently between the age of 60 and 80 and women between the age of 55 and 90. Although it is not entirely located along a diagonal line – indicating it would be allocated with specific birth cohorts – there seems to be a couple of birth cohorts that bear most red colored pixels in the sensitivity heat map.

The build-up of the impact per birth cohort can be seen when the impact is summated along diagonals. The result is an indication of how much the VPV would increase if for an entire birth cohort the mortality would structurally decrease with 1%. It is an indication and not an exact estimate, because the assumption is made that change of mortality is independent per year and per age. Although this is quite a strong assumption, calculations show that there is little difference (see appendix A.1).

One cohort that jumps out is the 1945 birth year cohort. This cohort has a generally higher mortality, as was seen in the first part of this thesis. An improvement of this mortality by 1% would not lead to ‘normal’ mortality rates, the generation would still suffer higher mortality than surrounding birth cohorts. Hence the impact would be limited.
Observing the cohort aggregated sensitivity of the VPV to a change in mortality, we see that men start to matter at a later age when it comes to mortality. Also, for men and women currently aged between 67 and 76, the sensitivity is almost equal, whereas it is larger for women of all ages outside that range.

Interesting may be to know how much of the sensitivity is located with pensioners. They will not build up any more pension rights, so the impact they have on the VPV could only be dependent on their mortality. Accumulating the sensitivity scores for the birth cohorts up to 1944, who turned 65 in 2009, the share of the sensitivity allocated with pensioners is roughly 25% for women and 35% for men.

Of course, the basis of this entire discussion about sensitivity has been a 1% relative decrease of mortality. In reality, the change of mortality will be different for various ages, across years. To get a more realistic view of what kind of impact on the VPV is possible, the models introduced in part I are used to simulate future mortality. The stochastic mortality models are used to get a sense of which particular age groups can be expected to experience a drop in mortality and hence where longevity risk really lies. But in order to draw conclusions about risk, we first have to know what it is and how to calculate it.

### 3 Mathematical definitions of risk

A unique mathematical definition of risk does not exist, but there are a number of properties that constitute a risk measure mathematically. The ones presented here were given by Elliot & Kopp (2005). Let $R$ be a stochastic variable denoting the net result of some transaction. If $R$ is negative a loss is incurred and if it is positive there is profit. A risk measure $\rho$ should firstly recognize that a safe bet is risk free.

\[
\text{If } R \geq 0 \text{ then } \rho(R) \leq 0
\]

Secondly, if a certain constant amount of money is combined with the stochastic result, i.e. a buffer, this decreases the risk by the amount of money in the buffer.

\[
\rho(R + b) = \rho(R) - b
\]

And a third property that the risk measure $\rho$ should also have is that the size of the risk is proportional to the size of the transaction.

\[
\rho(cR) = c \rho(R)
\]

Finally, there is a fourth property that ensures that diversification of risk pays off. Let $R_1$ and $R_2$ be the results of two transactions, then

\[
\rho(R_1 + R_2) \leq \rho(R_1) + \rho(R_2).
\]

A risk measure that has this property is called *sub-additive or convex* (Elliot & Kopp, 2005). If the risk measure has all of the four properties above, it is called a *coherent risk measure*.

A number of different measures of risk will be introduced and discussed. One well-known measure is Value at Risk, which is used throughout the world by financial institutions. Also an adaptation called Conditional Value at Risk will be introduced and its results discussed.
a. Value at Risk

Developed in the late 1980’s by analysts of the bank JP Morgan, the Value at Risk (VaR) measure is an intuitive application of the vague definition above. It is based on quantiles of the distribution of profit or loss.

\[ \text{VaR}_a(R) = m(R) - q_{1-a}(R) \]

Here, \( m \) is the median of the distribution and \( q_a \) the \( a \)th quantile. The quantile \( q_{1-a}(R) \) is defined as the value which is exceeded by stochastic \( R \) with probability \( a \). If the median of the distribution is 0, then VaR reduces to minus the \((1-a)\)th quantile. Most of the time the median is known. The statistic can then be corrected for the median and hence it can be left out of the definition.

![Graph showing VaR and Conditional Value at Risk](image.png)

Although VaR is easy to use, an important disadvantage of this risk measure is that it doesn’t give an indication of how bad the result may be, once it does beyond the VaR level. The VaR measure gives an indication of what a severe loss could be, but not how severe one may expect it to be in the case that a loss is incurred.

b. Conditional Value at Risk

The Conditional Value at Risk (CVaR) measure is a convex risk measure that bears a clear resemblance to VaR. It is defined as the mean of the 100a % worst results. In other words, it is an answer to the question how bad on average one could expect the result to be, given that it will be bad.

\[ \text{CVaR}_a(R) = m(R) - E(R \mid R < q_{1-a}) \]

Alternative names under which this risk measure is known are Expected Shortfall and Tail Conditional Expectation.
4 Longevity risk calculated from stochastic models

The models that were fitted in part I were used to make predictions based on time series analysis. This way of predicting came with a probability distribution which was used to construct confidence intervals. The two stochastic risk measures that are defined above can now be applied to those probability distributions.

When calculating the impact on the liabilities of stochastic mortality, 10000 simulations were made with each model. This facilitates the calculation of quantiles as well. The results for the impact on the liabilities estimate will be presented in a number of ways, starting with the aggregate risk for the total population. Results that will also be shown are aggregated risk per birth cohort.

Because the median of all simulations is not known in advance, results could not be corrected for it and the definition of the risk measures VaR and CVaR as used to calculate risk includes the a posteriori correction for the median.

a. Risk for total population

In the simulations the total impact, across all ages, of stochastic mortality according to the various models was calculated and then submitted to the VaR and CVaR risk measures. The results show that there is quite a large difference among models for the longevity risk of the male population.

<table>
<thead>
<tr>
<th>Model</th>
<th>Men</th>
<th>Women</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lee-Carter</td>
<td>490</td>
<td>1 333</td>
<td>1 823</td>
</tr>
<tr>
<td>Renshaw-Haberman</td>
<td>724</td>
<td>1 311</td>
<td>2 035</td>
</tr>
<tr>
<td>Lee-Carter+cohort</td>
<td>746</td>
<td>1 315</td>
<td>2 061</td>
</tr>
<tr>
<td>Cairns-Blake-Dowd</td>
<td>784</td>
<td>1 270</td>
<td>2 054</td>
</tr>
<tr>
<td>Cairns-Blake-Dowd 2</td>
<td>902</td>
<td>1 347</td>
<td>2 249</td>
</tr>
<tr>
<td>Cairns-Blake-Dowd 3</td>
<td>911</td>
<td>1 378</td>
<td>2 289</td>
</tr>
</tbody>
</table>

Longevity risk for PFZW by VaR_{0.975} measure, based on stochastic models (million euro)
<table>
<thead>
<tr>
<th>Model</th>
<th>Men</th>
<th>Women</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lee-Carter</td>
<td>500</td>
<td>1309</td>
<td>1809</td>
</tr>
<tr>
<td>Renshaw-Haberman</td>
<td>733</td>
<td>1287</td>
<td>2020</td>
</tr>
<tr>
<td>Lee-Carter+cohort</td>
<td>729</td>
<td>1325</td>
<td>2054</td>
</tr>
<tr>
<td>Cairns-Blake-Dowd</td>
<td>800</td>
<td>1268</td>
<td>2068</td>
</tr>
<tr>
<td>Cairns-Blake-Dowd 2</td>
<td>901</td>
<td>1373</td>
<td>2274</td>
</tr>
<tr>
<td>Cairns-Blake-Dowd 3</td>
<td>916</td>
<td>1357</td>
<td>2273</td>
</tr>
</tbody>
</table>

Longevity risk for PFZW by CVaR_{0.975} measure based on stochastic models (million euro)

\[ b. \quad \textit{Distribution of risk over age and time} \]

Every model that was fitted in part I, models the mortality in a different way, so the question can be asked what influence this may have on the estimation of longevity risk. Plotting the longevity risk attributed to certain age groups in a given year in heat maps, one can clearly see where the bulk of the risk is located.

According to the Lee-Carter model, the male part of the longevity risk lies mainly with the people of age 68 to 78 from 2015 to 2030. This corresponds to men born in the years 1938-1952. For women the hot spot of risk is located later, starting around 2025, and at higher ages, roughly corresponding to birth years 1950 – 1960, the baby boom generation.
Looking at all the different plots of the longevity risk, estimated by the various models, a certain similarity of form can be observed, but the differences are more interesting. For example, Renshaw-Haberman shifts the risk hot spot about 5 to 10 years into the future, but attributing it to the same generations. This can be explained by the higher estimate of life expectancy. This also accounts for the higher values.

The generation that bears the bulk of risk does not really differ. However the Lee-Carter model with simple cohort extension distributes the risk for men over a longer period.

This more stretched out distribution of risk is observed as well for all three of the Cairns-Blake-Dowd type models, with the only difference among them being the scale. Also for women, there seems little difference in distribution of risk. The heaviest peak lies around 2043 for women well into their eighties.
Heat map of longevity risk according to the Cairns-Blake-Dowd 2 model for men (left) and women (right)

Heat map of longevity risk according to the Cairns-Blake-Dowd 3 model for men (left) and women (right)

c. Risk per birth cohort

In order to have an indication of how the longevity risk is distributed over the population and because the heat maps suggest that risk strongly depends on birth year, the risk was also calculated per birth cohort. As can be seen, the bulk of the risk is accounted for by the people who are currently 40 to 70 years old, so who were born between 1940 and 1970. These are the people who have the highest paying jobs, have accumulated most pension benefits and will continue living for a long time.

Another observation, one that was already apparent in the figures of total longevity risk, is that where all models give a similar estimate for women, there is quite a lot of variation in the estimates for male longevity risk. This means that there is quite a bit of model risk involved in modeling male longevity risk.
Longevity risk measured by VaR \(0.975\) (left) and CVaR \(0.975\) (right), aggregated per cohort for men (red) and women (blue) (euro)

The shape of these graphs seems similar to those of the distribution of pension rights and of the sensitivity. To see whether this is a justified remark, the longevity risk as percentage of the liabilities was calculated as well.

Longevity risk as measured by VaR \(0.975\) as percentage of built up pension rights per cohort of five birth years for men (blue) and women (pink)

The longevity risk percentages are almost equal for men and women for those cohort that were not yet retired in 2009, gradually building up to about 5% for 50 year olds, followed by a slight descent towards 4% just before retirement.

However, after retirement there are decidedly differences. Some caution should be taken when viewing the results for high ages. There is quite a lot of volatility because small quantities are divided by other small quantities. Therefore, there is a degree of uncertainty in these figures. However, it is clear that there is a decidedly different trend for women and men. This might be due to the fact that women currently aged 90 are of a generation that as rule didn’t have many jobs and as a result didn’t build up much pension rights.
d. Pensioners carry a relatively small portion of the longevity risk

Calculating the cumulative distribution of risk over the birth cohort was done to estimate the share risk that is carried by the pensioners, to be able to compare this with the figure that we got when interpreting the sensitivity of the VPV. For men this share amounts to about 19%, for women about 13%. These figures are substantially lower than those for sensitivity. Thus, we see that the mortality models’ predictions of higher volatility of mortality for lower ages than for higher ages, account for limiting the share of the risk that is carried by pensioners.

<table>
<thead>
<tr>
<th>Model</th>
<th>Men</th>
<th>Women</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lee-Carter</td>
<td>15%</td>
<td>12%</td>
</tr>
<tr>
<td>Renshaw-Haberman</td>
<td>19%</td>
<td>15%</td>
</tr>
<tr>
<td>Lee-Carter+cohort</td>
<td>19%</td>
<td>12%</td>
</tr>
<tr>
<td>Cairns-Blake-Dowd</td>
<td>19%</td>
<td>13%</td>
</tr>
<tr>
<td>Cairns-Blake-Dowd 2</td>
<td>19%</td>
<td>13%</td>
</tr>
<tr>
<td>Cairns-Blake-Dowd 3</td>
<td>19%</td>
<td>13%</td>
</tr>
</tbody>
</table>

Share of longevity risk carried by passive population in the PFZW pension fund as measured by VaR<sub>0.975</sub>.

Beyond the age of 80, there is hardly any risk left. This contradicts that fact that more and more people will reach ages of 80 and higher. However, most of those who will reach very high ages are only in their forties now and therefore have built relatively little pension rights.

<table>
<thead>
<tr>
<th>Model</th>
<th>Men</th>
<th>Women</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lee-Carter</td>
<td>16%</td>
<td>12%</td>
</tr>
<tr>
<td>Renshaw-Haberman</td>
<td>19%</td>
<td>14%</td>
</tr>
<tr>
<td>Lee-Carter+cohort</td>
<td>19%</td>
<td>12%</td>
</tr>
<tr>
<td>Cairns-Blake-Dowd</td>
<td>20%</td>
<td>14%</td>
</tr>
<tr>
<td>Cairns-Blake-Dowd 2</td>
<td>19%</td>
<td>13%</td>
</tr>
<tr>
<td>Cairns-Blake-Dowd 3</td>
<td>19%</td>
<td>13%</td>
</tr>
</tbody>
</table>

Share of longevity risk carried by passive population in the PFZW pension fund as measured by CVaR<sub>0.975</sub>.

After correcting for the size of the estimated risk, we see that Lee-Carter structurally attributed more risk to younger aged people, compared to estimates with other models. This corroborates the observation made in the section (II 4a).
4 A non-stochastic measure of longevity risk

In the reports that pension funds write to inform the Dutch central bank (DNB) about the various categories of risk that the pension funds have acquired, there is also an accounting formula intended to represent the risk of uncertainty in mortality, called TSO. It is calculated by multiplying the value of the pension rights of a certain age group by a factor and summing the results. The factors are recommended by DNB.

<table>
<thead>
<tr>
<th>Age group</th>
<th>TSO-factor</th>
</tr>
</thead>
<tbody>
<tr>
<td>30-</td>
<td>0.1</td>
</tr>
<tr>
<td>30-35</td>
<td>0.095</td>
</tr>
<tr>
<td>35-40</td>
<td>0.085</td>
</tr>
<tr>
<td>40-45</td>
<td>0.075</td>
</tr>
<tr>
<td>45-50</td>
<td>0.06</td>
</tr>
<tr>
<td>50-55</td>
<td>0.045</td>
</tr>
<tr>
<td>55-60</td>
<td>0.035</td>
</tr>
<tr>
<td>60-65</td>
<td>0.025</td>
</tr>
<tr>
<td>65-70</td>
<td>0.02</td>
</tr>
<tr>
<td>70-75</td>
<td>0.02</td>
</tr>
<tr>
<td>75-80</td>
<td>0.02</td>
</tr>
<tr>
<td>80-85</td>
<td>0.015</td>
</tr>
<tr>
<td>85-90</td>
<td>0.01</td>
</tr>
<tr>
<td>90+</td>
<td>0.01</td>
</tr>
</tbody>
</table>

TSO-factors by age group as recommended by DNB

The TSO factors recommended by DNB do not differ for men and women. Also, the clear pattern is a descending TSO weight with increasing age. Using these factors to calculate longevity risk in a deterministic way, results in a value of € 630 million for men and € 1.56 billion for women, adding up to € 2.19 billion. The total amount fits in with those obtained with the stochastic models. However, because the factors are the same for men and women, the ratio of their TSO values is the same as the ratio of their pension amounts. This explains the difference in the results of male and female longevity risk between the stochastic models and this accounting formula.

The TSO weights can be taken to represent the share of the pension rights that should account for the longevity risk specific to that age group. However, comparing them to the values obtained when we calculated the risk per cohort in section 3, we see an entirely different picture.

The risk shares are not descending like the TSO weights. For low ages, the share of risk as calculated by stochastic models is much lower than the TSO weight and for high ages, the share stays larger than the TSO weights. This can be attributed to the fact that the moment young people will start drawing from their pensions is a long time away. That means that the interest rate discount will be large.
5 Conclusions

The total amount of risk is of the order of 2 billion euros
Results from the various models and the various risk measures all give a total longevity risk of 2 billion euro, with a maximum 10% difference for the most extreme estimates. The stability of these estimates is mostly due to the fact that the largest part of the risk lies with women and estimates for female longevity risk seem to be very stable across all models.

Male longevity risk is estimated very differently by different models
The variation in the estimates of the total longevity risk by different models can mostly be attributed to the variation in estimates of male longevity risk. Between the highest and the lowest estimate is a factor 2. The reference estimate for male longevity risk was chosen to be the estimate given by the Lee-Carter model extended with a cohort parameter, because that was seen to be the best fit in part I.

Same generations bear bulk of risk in all models
The bulk of the risk is bore by current 45-65 year old women and 50-70 year old men. These generations have built up most pension rights and are relatively close to retirement. However because different models allocate the largest improvements in mortality differently, the hot spot of risk is located at different ages for these generations.

Pensioners bear 19% of risk with men, 13% with women.
Although the total amount of longevity risk varies per model, the share that lies with pensioners is the same in almost all models: 19% for men and 13% for women.

Deterministic method leads to comparable estimate
Although the deterministic method supported by DNB is very different from stochastic models, the estimate of total longevity risk isn’t significantly different. Age specific factors differ a lot because DNB’s TSO weights do not seem to account for the interest rate discounting of pension rights for young people.
Part III
Alternative scenarios and inverse modeling

1. New trends

So far, the scope of the analysis has been limited to models being applied to Dutch historic data. However, as Oeppen & Vaupel (2005) shows, the evolution of life expectancy is positively correlated across nations in the developed world. This can be explained by the internationalization of medicine, the rapid adoption of new medical methods in other developed countries, as soon as resources are available. Also the cultural interweaving of these countries could lead to convergence of living habits. Americanization, the adoption of American culture by people on other parts of the world, is a good example of this phenomenon.

The models that have been used in parts I and II to predict future mortality cannot account for this correlation across countries. To make a start at computing the importance of this correlation effect, two alternative scenarios for the advance of Dutch life expectancy are defined. The first one, ‘Americanization’, is based on the convergence of Dutch life expectancy in 2050 to that of US citizens, using the same prediction models. The second one is a scenario in which the Dutch evolve to the longest living people in the world by 2050, the ‘world champion scenario’.

In the case of the world champion scenario, it is impossible to gather death and population size tables to serve as input for the mortality models defined in this report, as the champion country changes roughly every decade. However Oeppen & Vaupel showed there is quite a persistent law of linear advance in the world’s highest life expectancy with an increase of about 3 months per year. Using the predictions made in that article, this life expectancy would rise to 30 years beyond the age of 65.

In order to be able to make a reasonable estimate of the increase in pension liabilities, a direct link between the increase of liabilities and the increase of life expectancy is called for. This would mean a dramatic simplification with respect to mortality models that use hundreds of parameters, and comes at the cost of higher uncertainty.

The first step in establishing this direct relation is to model the effect of adjusting only the random walk drift parameter in dominant time-dependent parameter, kappa. This already narrows the number of parameters down to just one. The second step is to examine the effect of adjusting this drift parameter on the resulting life expectancy in 2050.

2. Simulations suggest linear relation between life expectancy and liabilities

The method described above tries to model the link between pension liabilities and life expectancy by using the fact that both are derived from the mortality table. Using the mortality models, a simulation of 10000 runs was made of these mortality tables. These simulated mortality tables were used to calculate remaining life expectancy at the age of 65 in 2050 and to compute the effect on the liabilities as well. These two quantities can then be combined. The results are presented in the form of scatterplots, showing the change in pension liabilities in euros as a function of change of remaining life expectancy at 65 in years.
The scatterplots exhibit a linear relation between a change in life expectancy and the increase in liabilities. By linear regression, we can quantify the linear dependence as well as the uncertainty of the linear dependence.

<table>
<thead>
<tr>
<th></th>
<th>Value (euro/year)</th>
<th>S.E. (euro/year)</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Men</td>
<td>3.72 $10^8$</td>
<td>3.49 $10^6$</td>
<td>&lt;0.0000001</td>
</tr>
<tr>
<td>Women</td>
<td>6.50 $10^8$</td>
<td>4.25 $10^6$</td>
<td>&lt;0.0000001</td>
</tr>
</tbody>
</table>

Just like before with the estimation of longevity risk, there are some differences between the various models. These differences result from the way mortality changes are distributed among ages and generations in different models, compounded by the specific distribution of sensitivity of the liabilities to mortality.
Scatterplot of change in pension liabilities (euro) as function of change in life expectancy (years), based on Cairns-Blake-Dowd 3 model simulations (women left, men right)

<table>
<thead>
<tr>
<th></th>
<th>Value (euro/year)</th>
<th>S.E. (euro/year)</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Men</td>
<td>$3.44 \times 10^8$</td>
<td>$2.92 \times 10^7$</td>
<td>&lt;0.0000001</td>
</tr>
<tr>
<td>Women</td>
<td>$6.29 \times 10^8$</td>
<td>$4.05 \times 10^6$</td>
<td>&lt;0.0000001</td>
</tr>
</tbody>
</table>

The standard errors reported are standard errors of the slope parameter. It gives a measure of how much the estimated value of the slope could deviate given the data. Given that a 95% confidence interval can be constructed by taking the value and 2 standard errors on each side, it is reasonable to assume that the given values are accurate up to one decimal point.

Performing residual analysis, it is noted that the residuals are uncorrelated to the remaining life expectancy at 65. The linear model seems to suffice to explain the dependence of the size of the pension fund’s OP liabilities. However, as certain as we are of the linearity, there still is quite a lot of deviation from the fitted line. The linear model explains 53-57% of the variance for men and ca. 70% for women as can be seen in the $R^2$ statistics.

<table>
<thead>
<tr>
<th></th>
<th>Lee-Carter</th>
<th>Lee-Carter-cohort</th>
<th>Cairns-Blake-Dowd 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Men</td>
<td>0.53</td>
<td>0.57</td>
<td>0.56</td>
</tr>
<tr>
<td>Women</td>
<td>0.69</td>
<td>0.70</td>
<td>0.69</td>
</tr>
</tbody>
</table>

$R^2$ statistic values for linear models of dependence life expectancy and OP liabilities

### 3. Inverse model

As in part II, the impact of a change in mortality for a certain age in a given year is calculated by multiplying the relative change in mortality as percentage by a sensitivity factor $s(t, x)$. This linear relation can easily be inverted to a relation linking mortality to impact on liabilities $I(t, x)$.

$$q_{\text{alt}}(t, x) = q_{\text{base}}(t, x) \left( 1 + \frac{I(t, x)}{s(t, x)} \right)$$

In order to use this, we would have to specify the impact on liabilities for each age and year to obtain a change in mortality, that we can then use to calculate life expectancy. However, this is not a user friendly way of calculating how mortality can change in order to achieve a certain effect on pension rights. Usually only the total impact is given. So for the purpose of applicability, we would like to reduce this inverse model to a direct link between life expectancy and impact on liabilities.
The approximate linear dependence that was found between a rise in life expectancy and a rise in pension rights, gives us a way to view longevity risk in terms of increasing life expectancy instead of increasing pension liabilities. This would be the inverse of calculating the increase in liabilities that results from increasing life expectancy. Assuming a certain combined value of assets, inverse modeling then makes it possible to calculate how much life expectancy would have to increase for the combined present value of liabilities to equal assets. More interestingly, we could calculate the increase in life expectancy that would result in a decrease of the cover ratio to 105%, the minimum acceptable according to the Dutch Pension Law.

As was shown in the scatterplots of the link between a change in life expectancy at retirement age and the effect on liabilities, different scenario’s with the same resulting life expectancy can result in the same effect on liabilities. And conversely, different effects on liabilities can result from the same increase in life expectancy, depending on the exact path the value of life expectancy takes between 2010 and 2050.

In the dataset of OP pension rights that was the basis for all calculations in this report, the total amount of liabilities is 53.3 billion euro, as a result of which 1 percentage point of cover ratio corresponds to 533 million euro. Using this we can invert the values found for the slopes of the linear models found earlier to find the change in life expectancy corresponding to a change of 1 percentage point of cover ratio.

<table>
<thead>
<tr>
<th>Model</th>
<th>Male</th>
<th>Female</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lee-Carter</td>
<td>1.4</td>
<td>0.8</td>
</tr>
<tr>
<td>Renshaw-Haberman</td>
<td>0.9</td>
<td>0.6</td>
</tr>
<tr>
<td>Lee-Carter+cohort</td>
<td>1.6</td>
<td>0.7</td>
</tr>
<tr>
<td>Cairns-Blake-Dowd 1</td>
<td>1.8</td>
<td>1.0</td>
</tr>
<tr>
<td>Cairns-Blake-Dowd 2</td>
<td>1.5</td>
<td>0.9</td>
</tr>
<tr>
<td>Cairns-Blake-Dowd 3</td>
<td>1.6</td>
<td>0.9</td>
</tr>
</tbody>
</table>

Years of additional life expectancy corresponding to a 1 percentage point decrease of cover ratio

A familiar divergence between results of the various models occurs. The markedly lower effect for Renshaw-Haberman can be partly explained by the fact that this model already projects a very high life expectancy. Deviations from the trend in this model, given the stochastic distribution, are therefore relatively small. Analogously, the CBD 1 model is more prudent in its projection of life expectancy, but comes with a relatively larger uncertainty around that prediction.

The other models all produce similar results of on average 1.5 years of life expectancy for men or on average 0.8 years of life expectancy for women corresponding to 1 percentage point of cover ratio.

4. Alternative scenarios

a. World champion scenario

According to Oeppen & Vaupel (2002) the increase of life expectancy at birth in the record holding country increases with 0.243 yr/yr for women and 0.222 yr/yr for men. Interestingly, if one continues to extend this diverging increase of life expectancy until 2050, remaining life expectancy at 65 years old for women will be 33.7 years, making record life expectancy for 65 year old women in 2050 as high as 98.7 years old, while remaining life expectancy for 65 year old men will amount to 25.5 years, bringing record male life expectancy at 65 to 90.5.
These values for life expectancy are well higher than the prediction made for Dutch men and women. According to Lee-Carter, the difference for men is 4.5 years, while for women it is as much as 10.1 years. In the table below, one can see the divergence in life expectancy and the corresponding increase in OP liabilities, which are calculated using the linear dependence modeled above.

<table>
<thead>
<tr>
<th></th>
<th>Years</th>
<th>Euro (billion)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lee-Carter</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Men</td>
<td>4.5</td>
<td>1.69</td>
</tr>
<tr>
<td>Women</td>
<td>10.1</td>
<td>6.59</td>
</tr>
<tr>
<td>Lee-Carter-cohort</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Men</td>
<td>2</td>
<td>0.675</td>
</tr>
<tr>
<td>Women</td>
<td>9.5</td>
<td>6.92</td>
</tr>
<tr>
<td>Cairns-Blake-Dowd 3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Men</td>
<td>4.4</td>
<td>1.48</td>
</tr>
<tr>
<td>Women</td>
<td>9.9</td>
<td>7.2</td>
</tr>
</tbody>
</table>

Predicted difference in years of life expectancy with world record holder in 2050 and corresponding extra pension liabilities if convergence to that life expectancy is attained

Depending on which model is used to predict Dutch life expectancy, the extra pension liabilities that would have to be accounted in the case of an adjustment of current predictions to the Netherlands becoming world record holders in 2050 for both men and women varies between 7.6 and 8.7 billion euro for the PFZW pension fund. This corresponds to an increase of 14%-15%. The sum is dominated by the female component because the gap between model predictions and world record holder prediction is large and the pension fund is more than twice as sensitive to an increase in life expectancy for women. This scenario is clearly beyond the longevity risk analysis carried out in part II.

b. Americanization scenario

The United States have never been world record holder of life expectancy. A plethora of possible explanations could be found. For instance, it could be argued that the absence of universal health care causes a larger disadvantage for lower classes than it does in European countries that have universal health care. Without going too far into why American life expectancy has been lower than Dutch, it is an historical fact and American projections continue to lie beneath Dutch.

Critics of American cultural influence in the Netherlands point out that the US is trend setter when it comes to developing lifestyle diseases, like obesity. Dutch numbers lag American by about 10 years in this example. It is estimated that a majority of Americans are overweight and 1 in 5 have developed Type 2 diabetes. If in ten years, the Netherlands attain similar numbers, this would have a dramatic effect on mortality. So the Americanization scenario, in which Dutch life expectancy converges to projected American life expectancy, may be interesting, because it is a scenario in which the increase of life expectancy is lower – as opposed to the world champion scenario – and because it is a scenario that can be interpreted as a realistic increase of lifestyle disease related mortality.

<table>
<thead>
<tr>
<th></th>
<th>Years</th>
<th>Euro (billion)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lee-Carter</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Men</td>
<td>0.3</td>
<td>0.111</td>
</tr>
<tr>
<td>Women</td>
<td>-0.86</td>
<td>-0.559</td>
</tr>
<tr>
<td>Lee-Carter-cohort</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Men</td>
<td>-2.85</td>
<td>-0.96</td>
</tr>
<tr>
<td>Women</td>
<td>-1.5</td>
<td>-1.09</td>
</tr>
<tr>
<td>Cairns-Blake-Dowd 3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Men</td>
<td>0.2</td>
<td>0.067</td>
</tr>
<tr>
<td>Women</td>
<td>-1.1</td>
<td>-0.800</td>
</tr>
</tbody>
</table>
Predicted difference in years of life expectancy with United States in 2050 and corresponding extra pension liabilities if convergence to that life expectancy is attained

Americanization would relieve PFZW’s OP liabilities by about 0.5 to 2 billion euro, a decrease of 1%-4%, depending on the model. It should be concluded that this effect is not really significant, because these changes in mortality fall in the 95%-confidence intervals for the Dutch life expectancy defined in part I. In other words, this scenario lies in the natural spread of possibilities predicted by the stochastic mortality models.

5. Conclusions

The models that we introduced in part I of this thesis were used to make predictions by extending historical trends into the future. However, the assumption that future developments will continue in line with historical ones is rather strong. Therefore, alternative scenarios should be considered to get idea of what might happen to PFZW’s liabilities if there will indeed be a significant deviation of the predicted trend.

Two scenarios are defined that are examined as alternatives to the Dutch trend, the world champion scenario and the Americanization scenario. The world champion scenario envisions that Dutch life expectancy will grow to be the highest in the world in 2050, where the assumption is made that world record life expectancy will continue to grow as postulated by Oeppen & Vaupel (2005). The Americanization scenario projects Dutch life expectancy to conform to that of Americans, building on the observation that Dutch society tends to follow a lot of trends set in the US.

Simulations suggest that the relation between deviations of life expectancy at 65 and corresponding changes in pension liabilities is approximately linear. However, depending on which model is used to model trend and deviations of life expectancy, different values of the parameter defining that relation are estimated. For women values as far apart as 630 and 730 million euro per year of extra life expectancy are recorded. For men this lies between 340 and 370 million euro. This relates to 1.2-1.4 percentage point lower cover ratio per additional year for women and 0.7 percentage point lower cover ratio per additional year for men.

The linear dependence can now also be used to define a simple inverse model, modeling a change of life expectancy in 2050 as function of the impact that would have on PFZW’s liabilities. Again different models give different results, Renshaw-Haberman markedly so, but generally a change of cover ratio by 1 percentage point corresponds to a deviation of 1.5 years with respect to the trend for men or 0.8 years for women.

The direct link between life expectancy at 65 and pension liabilities helps us find the impact of the world champion scenario. Because world record holders change through the years, there is no consistent historical mortality database for this scenario and we have to rely on the direct relation. It leads to the conclusion that especially female life expectancy would deviate significantly - by about 10 years - while male life expectancy increase by 2 to 4.5 years, depending on which model is used for the Dutch life expectancy trend. This combines to a 7.6-8.7 billion euro increase of OP liabilities, or 14-15 percentage point cover ratio.

Deviations are much smaller in the Americanization scenario. In fact, for men the deviations remain limited to several months, except for the Lee-Carter-cohort model, which projects significant higher life expectancy for men. For women, deviations are typically 1 to 1.5 years, also within the 95% confidence interval and hence not significant. Impact on OP liabilities would amount to 0.5-2 billion euro corresponding to 1-4 percentage point cover ratio.
Especially the world champion scenario shows that an alternative scenario can lead to much higher increases of pension rights than as estimated in the longevity risk analysis of part II. Relying only on the stochastic spread models with extension of historical trends overlooks scenarios that reach much further and impact a pension fund’s liabilities much harder. Although this report doesn’t incorporate a scenario in which life expectancy at 65 decreases, such scenarios are not unthinkable and hence should be examined as well.
Conclusions

Stochastic models for mortality
Mortality data show the development of various changes in mortality patterns in the period 1950-2008. These changes include a general decline of mortality over the years and an elevated mortality for birth years 1920 and 1945. The latter can be explained by particularly dire circumstances for infants at the time, namely the Spanish flu and the famine winter in the last year of World War II. These effects vary per age group and gender. However there are also deviations to these trends and effects, which leads to the consideration of stochastic models to model mortality.

Simplification of Renshaw-Haberman model is best fit
All of the models concerned incorporate a variable that models the general decline of mortality. The Lee-Carter model is extended with a cohort parameter that models the particular birth year effects and the Cairns-Blake-Dowd model is extended to allow for both second-order variation for age and birth year specific variation. The best fit on Dutch historical mortality data is attained by the Lee-Carter models with cohort parameter, Renshaw-Haberman and a simplification of it. However, the full Renshaw-Haberman exhibits signs of overfitting.

Predictions of future mortality
Using these models to predict future remaining life expectancy for 65 year olds, by modeling the time-dependent parameters as time series, leads to varying results, with Renshaw-Haberman projecting both women and men having life expectancy of 92 years at age 65 in 2050, from 82 for men and 85 for women in 2008. The other models lead to more moderate increases, to approximately 86 for men and 88 for women. However, these prediction all lie above the most recent CBS estimates (2008).

The 95%-confidence intervals that were constructed by modeling the time-dependent parameters as stochastic time series, are comparable in size among the Cairns-Blake-Dowd models, approximately 5 years to both sides. They are slightly smaller for the original Lee-Carter model and its simple cohort extension, approx. 4 years, and smallest for the Renshaw-Haberman model with approx. 3,5 years to both sides. Although the fact that the confidence interval is smallest for Renshaw-Haberman might suggest that it is the best predictive model, the problems with possible overfitting and resulting markedly higher prediction of life expectancy leads to the conclusion that its simplification is more credible and should be preferred.

Longevity risk consistent for women
The modeled stochastic distribution of mortality allows for the application of quantile-based risk measures like Value at Risk. Using the models found in part I, the value of overall longevity risk for the PFZW pension fund is calculated at approximately 2 billion euro, with a 10% deviation to the most extreme estimates. Most variation between models is encountered with men, differing by a factor 2 between the highest and lowest estimates. This is mainly due to differences in allocation of mortality decreases in time. The consistence of overall estimates comes from the fact that most pension rights and associate risk lie with women, whose mortality models are quite consistent as well. As a reference estimate for men the Lee-Carter model with simple cohort extension is chosen, because it was the best fit. A deterministic estimation method used by the Dutch central bank leads to an overall longevity risk of 2 billion euro as well.

Longevity risk concentrated in age groups
The bulk of longevity risk is located in the 45+ generations before retirement, who have built up most pension rights and are relatively close to retirement. It is observed that 13% of female longevity risk and 19% of male longevity risk is located with pensioners.
Alternative scenarios
Predicting future mortality with the stochastic models was based on the assumption that future developments will have similar effects as historical ones. This need not turn out to be the case. Therefore, realistic alternative scenarios were developed by comparing Dutch mortality to other countries.

The first scenario constitutes of The Netherlands growing to become world champions of life expectancy in 2050 according to the Oeppen & Vaupel law of world record life expectancy improvement. This would increase OP liabilities by approximately 8 billion euro, or 14-15% cover ratio. The second scenario is defined by The Netherlands converging to American life expectancy, which would result in a relatively small negative effect on OP liabilities of 0.5-2 billion euro, or about 1-4% cover ratio.

Focusing only on mortality projections that extend historical trends is common practice among actuaries but by limiting oneself to this model, one misses alternative realistic scenarios which could have large consequences for pension funds. It is therefore recommendable to include alternative scenarios in the analysis of pension liabilities.
Appendix A

Reflections and suggestions for further research

As was already mentioned in the acknowledgements, quite a long time was spent contemplating and trying to improve on the methods used and the results obtained in the research on quantifying longevity risk that was conducted during an internship at PGGM. This appendix contains reflections on the quality of the methods and results of the research presented in part I of the thesis. The aim of this appendix is to clarify and explain certain modeling decisions, evaluate the validity of modeling assumptions and provide suggestions for further research based on these reflections.

A.1 Mortality rate and the probability of dying

To reiterate the definition in chapter I.3, the mortality rate is defined as the number of deaths in proportion to the size of the population per time period. In this thesis the mortality rate is calculated over periods of one year.

\[ m_{t,x} = \frac{D_{t,x}}{N_{t,x}} \]  (A.1)

In actuarial science, individual deaths are modeled by a Poisson process with the mortality rate \( m_{t,x} \) as a time- and space-dependent parameter. This type of Poisson process is called non-homogenous. The time until (the first) death is then exponentially distributed. Because the mortality rate is taken to be piecewise constant as defined in equation A.1, the probability of death within one year can be formulated as

\[ q_{t,x} = 1 - e^{-m_{t,x}} \]  (A.2)

More information on non-homogenous Poisson processes can be found in section 2.4 of Ross (1996).

A.2 Likelihood distributions

It is assumed in chapter I.3 that the Poisson distribution function can be used as an approximation to the binomial distribution function with the intent to realize a faster convergence of the algorithms that fit mortality models on the data. In order to address concerns about the choice of distribution function for the maximum likelihood estimation of mortality model parameters, in this section a comparison is made between binomial and Poisson distributions, with the goal of finding any theoretical reasons why one should be preferred above the other.

A.2.1 Binomial distribution

A natural way to model mortality is to assume that deaths are binomially distributed with population size \( N \) and some probability of dying \( p \), which may depend on further model parameters \( \alpha \). Assuming that all spatial and temporal interdependence is modeled by \( p_{t,x}(\alpha) \), the likelihood function can be obtained from the binomial probability mass function:

\[ L(\alpha; D, N) = \prod_{t,x} \left( \frac{D_{t,x}}{N_{t,x}} \right) p_{t,x}(\alpha)^{D_{t,x}} \left( 1 - p_{t,x}(\alpha) \right)^{N_{t,x} - D_{t,x}} \]

The associated log-likelihood is of the form
\[
\ln L(\alpha; D, N) = \sum_{t,x} \left( \ln \left( \frac{D_{t,x}}{N_{t,x}} \right) + D_{t,x} \ln p_{t,x}(\alpha) + (N_{t,x} - D_{t,x}) \ln (1 - p_{t,x}(\alpha)) \right)
\]

The obvious choice for \( p \) is the one-year probability of dying \( q \), which may depend on further model parameters. Because Lee-Carter type models (chapter I.4) are based on a formula for \( \hat{m}_{t,x} \), the calculations below will also be in terms of \( \hat{m}_{t,x} \).

\[
\ln L(\alpha; D, N) = \sum_{t,x} \left( \ln \left( \frac{D_{t,x}}{N_{t,x}} \right) + D_{t,x} \ln (1 - e^{-\hat{m}_{t,x}(\alpha)}) - (N_{t,x} - D_{t,x})\hat{m}_{t,x}(\alpha) \right)
\]

In the optimization algorithm, the objective is to find a set of parameter values for which the gradient of the log-likelihood function with respect to the model parameters is 0. The gradient of the binomial log-likelihood function is

\[
\nabla_\alpha \ln L(\alpha; D, N) = \sum_{t,x} \left( D_{t,x} e^{-\hat{m}_{t,x}(\alpha)} - N_{t,x} + D_{t,x} \right) \nabla_\alpha \hat{m}_{t,x}(\alpha)
\]

\[
= \sum_{t,x} \left( \frac{D_{t,x}}{1 - e^{-\hat{m}_{t,x}(\alpha)}} - N_{t,x} \right) \nabla_\alpha \hat{m}_{t,x}(\alpha)
\]

The value of the gradient may explode due to the small values for the denominator in the first part of the summand for age groups where probability of dying is small.

It is possible to take \( m \) as an approximation for \( q \). In that case,

\[
\nabla_\alpha \ln L(\alpha; D, N) = \sum_{t,x} \left( \frac{D_{t,x}}{\hat{m}_{t,x}(\alpha)} - \frac{N_{t,x} - D_{t,x}}{1 - \hat{m}_{t,x}(\alpha)} \right) \nabla_\alpha \hat{m}_{t,x}(\alpha).
\]

This does not seem to be an improvement, as now the value of the gradient may diverge for age groups for which either probability of dying is very low or for which it is very close to 1. In theory, if \( m \) is estimated correctly, then the two parts of the summand cancel each other out, as both parts should yield values close to \( N_{t,x} \). However, in the first steps of the optimization algorithm, this may not be the case and divergence of the algorithm may be observed.

A.2.2 Poisson distribution

If \( N \) is large enough and \( p \ll 1 \), then the binomial distribution can be approximated by a Poisson distribution with parameter \( \lambda = N p \). The probability mass function for the Poisson distribution with parameter \( \lambda \) is

\[
P(X = k) = \frac{\lambda^k}{k!} e^{-\lambda}.
\]

Substituting \( \lambda = N_{t,x} p_{t,x}(\alpha) \) in the above and assuming all spatial and temporal interdependence is modeled by \( p_{t,x}(\alpha) \) gives the likelihood function

\[
L(\alpha; D, N) = \prod_{t,x} \left( \frac{N_{t,x} p_{t,x}(\alpha)}{D_{t,x}!} \right)^{D_{t,x}} e^{-N_{t,x} p_{t,x}(\alpha)}.
\]

The corresponding log-likelihood function is
\[
\ln L(\alpha; D, N) = \sum_{t,x} \left(D_{t,x} \ln(N_{t,x}p_{t,x}(\alpha)) - N_{t,x}p_{t,x}(\alpha) - \ln(D_{t,x}) \right).
\]

Again, there are two possible choices for \( p \). The choice \( q \) yields gradient

\[
\nabla_{\alpha} \ln L(\alpha; D, N) = \sum_{t,x} \left(D_{t,x} \frac{e^{-\bar{m}_{t,x}(\alpha)}}{1 - e^{-\bar{m}_{t,x}(\alpha)}} - N_{t,x}e^{-\bar{m}_{t,x}(\alpha)} \right) \nabla_{\alpha} \bar{m}_{t,x}(\alpha)
\]

\[
= \sum_{t,x} e^{-\bar{m}_{t,x}(\alpha)} \left( \frac{D_{t,x}}{1 - e^{-\bar{m}_{t,x}(\alpha)}} - N_{t,x} \right) \nabla_{\alpha} \bar{m}_{t,x}(\alpha)
\]

This log-likelihood function gradient continues to have the problem that its binomial counterpart had: if \( \bar{m}_{t,x} \) is close to 0, the value of the gradient may grow enough for the algorithm to diverge.

The approximation of \( q \) by \( m \) is natural, now that we have made the assumption \( p \ll 1 \) in order to allow the Poisson approximation. It results in log-likelihood gradient

\[
\nabla_{\alpha} \ln L(\alpha; D, N) = \sum_{t,x} \left( \frac{D_{t,x}}{\bar{m}_{t,x}(\alpha)} - N_{t,x} \right) \nabla_{\alpha} \bar{m}_{t,x}(\alpha).
\]

The possible convergence problem of this likelihood function is therefore the same as it was for the binomial distribution with probability parameter \( q \).

**Conclusion**

A comparison was made between binomial and Poisson distribution functions and choices for probability parameters \( q \) and \( m \). It was seen that divergence of the gradient may occur for age groups which have a small probability of dying, regardless of which distribution is chosen. In these cases \( m \) is a good approximation for \( q \) and therefore this choice does not impact convergence theoretically.

When one chooses the binomial distribution with parameter \( m \), then this divergence could also occur for age groups with high probability of dying. Furthermore, it is observed in practice that the variance of the number of deaths at high ages is relatively large, while, if mortality was modeled by a binomial distribution, this variance would decrease as the mortality rate increases. So even though the Poisson assumption may not hold for the data on higher age brackets, a model based on the binomial distribution might perform even worse.

**Discussion**

When fitting the mortality models with the Poisson distribution and parameter \( m \), no problems occurred as a result of the possibly unstable gradient. Rather, convergence was not achieved when the number of added parameters in \( \alpha \) grew too high or was internally highly correlated. So the observations made earlier may suggest that choosing the binomial distribution with probability parameter \( q \) should not pose any problems.

The algorithm used to fit the mortality models is a maximization algorithm that evaluates the log-likelihood function and its gradient in every iteration. If one should switch to the binomial distribution and \( q \), this would add a great amount of exponentiations that need to be computed in each iteration: twice for every \( t \) and \( x \). This slows down computation. This seems to be the only reason to prefer the choice of Poisson with parameter \( m \).
A.3 Parsimonious regression models

Parameter estimates are obtained in chapters I.4 and I.5 via regression with hundreds of parameters and then interpreted as fixed. When interpreting the results of the regression, it is observed that parameter vectors for age and time show an approximately linear increase or decrease. This suggests that the data could be described by a simpler model, which replaces an age dependent parameter of length 59 by an expression for a line containing 2 parameters.

Other parameter vectors have a more complicated shape. With the possible exception of the cohort parameter vector for the Cairns-Blake-Dowd extended model, all of the other parameter vectors show some structure that can be approximated by splines for example. This too would decrease the total number parameters to be estimated. A slightly worse fit is to be expected, but that can be offset by the drastic reduction in parameters. So in the sense of the Schwarz-Bayes information criterion introduced in chapter I.6, these reduced models can be preferable.

One type of parameters is effectively modeled by a simpler model and that is the time parameter. In each of the models, the primary effect for time is modeled as a random walk with drift time series. This is in effect a linear model. The deviations from this linear trend are used to calculate the variance of the time series process. This variance is the sole basis for the estimation of uncertainty in mortality projections.

A suggestion for further research is to investigate whether applying the approach used with the primary time parameter to all parameter vectors - finding a trend in the parameter vector and quantifying the uncertainty around this trend - can yield a broader basis for the estimation of variance of mortality predictions.

A.4 Time series modeling

A central step in the modeling process that leads to a calculation of longevity risk is the upgrading of $\kappa(t)$ from time dependent parameter to time series (chapter I.7). This modeling step determines the estimation of uncertainty in future mortality and hence longevity risk. Therefore, it deserves critical evaluation.

The assumption that a time series is adequately described by a random walk model with drift (RWD), implies that the random walk steps $\kappa(2) - \kappa(1)$, $\kappa(3) - \kappa(2)$, ..., $\kappa(t) - \kappa(t-1)$, ... are independent (chapter I.4, p. 16). This means that no significant autocorrelation should be observed.

However, there is autocorrelation in the time parameter vectors, so there is evidence that subsequent random walk steps may not be independent. Unfortunately, the historic data set is only 59 years long. The variance of estimates for autocorrelation is inversely proportional to the length of the data set (Chatfield 2004). This means that autocorrelation has to be larger than 0.25 to be considered significant. Also, such a short time series is arguably too short to be able to serve as a basis for advanced time series modeling, as will be argued in section A.7.

If one would be interested in extending the time series models to advanced models that can explain autocorrelation, then it is suggested one studies a well-known set of time series models: the set of autoregressive integrated moving average (ARIMA) models. These models are extensively described by Brockwell and Davis (1991) and Box and Jenkins (1976).

Denoting $\delta(t) = \kappa(t) - \kappa(t-1)$, the value for $\delta(t)$ is modeled to be a linear combination of former values of $\delta$ (the autoregressive part) and a linear combination of independent standard normally distributed innovations $Z_t$ (the moving average part):

$$\delta(t) - \mu = a_1(\delta(t-1) - \mu) + \cdots + a_p(\delta(t-p) - \mu) + \sigma(Z_t + b_1Z_{t-1} + \cdots + b_qZ_{t-q}).$$
The above holds for all $t$, with $\mu$, $\sigma$, $a_i$, and $b_j$ fixed. Modeling the time series this way may improve predictions of mortality, although mostly in the short term.

### A.5 Cross-correlation between time series

Just as time series can be autocorrelated, there may also be dependence between two time series, for example the time parameter vectors for men and women. Or there may be interdependence between the various time parameter vectors of the Cairns-Blake-Dowd models. Analogous to the ARIMA models for autocorrelated single time series there exist generalizations to multivariate ARIMA models, described in Lutkepohl (2005) and Reinsel (1993). There is an added layer of complexity, because multivariate models of this type are not necessarily uniquely identifiable. This problem is compounded by the already mentioned issue of a small data set.

### A.6 Time series prediction

The problematic brevity of the time period on which the models are calibrated, continues throughout the modeling process. In accordance with practice at the Actuarieel Genootschap (AG) the trend in the last 20-year period is used for prediction. The prediction period is 40 years. Philosophically, this is hard to defend. There is no obvious reason why the preceding 20-year period and not some other time period should be indicative of future mortality development. The only defense to be given is that it is so-called 'best practice' that is undertaken by experts (AG) and that is the defense offered in the thesis.

The difficulty I had with following this approach, is exactly what led me to the addition of part III to the thesis, in which it is argued that the uncertainty about the trend itself – let alone the variation around it - is such that several alternative scenarios should be investigated. Only two are presented in the thesis, but a clear suggestion for further research is to look into a multitude of alternative scenarios, including also scenarios defined in other ways than convergence to a certain trend.

### A.7 Model validation

As touched upon in the section on time series prediction (A.6), model validation may well be the most problematic issue of the fitting of mortality models to Dutch historic data.

#### A.7.1 No validation period for time series

The prediction variance that is displayed in chapter I.7 is a theoretical variance that follows from the assumption of the model. In time series research it is customary to compare the size of the variance to the observed deviations when the predictions are tested against withheld data (Chatfield 2004). To be able to assess the quality of a predictor that needs to act at a horizon of 42 years, a similar or ideally larger period should be withheld from the data for the purpose of validation. In a set of 59 years, not much is left if one withholds 40 years for validation. If one were to be orthodox when fitting the model, the trend and parameters would have to be calibrated on the time period 1950-1967, which is doomed to failure. Since no reasonable or useful period of validation could be selected, the choice was made to base prediction solely on the theoretical estimates based on the random walk time series model.

#### A.7.2 Goodness-of-fit

Chapter I.6 contains an intuitive approach to evaluate the standardized residuals. If indeed they were standard normally distributed as was hypothesized in chapter I.3, then only 1% of the standardized residuals should exceed 2.33. It is demonstrated that it is not even nearly true for most models. Not surprisingly, all conventional goodness-of-fit tests fail. It could therefore have been stated more decidedly that none of the models adequately model Dutch mortality.
An exception can be made for the Lee-Carter cohort model, which seems to leave no observable correlation in the residuals but merely suffers from excessive variance. This property is called overdispersion and may arise from the fact that the assumption of the Poisson model for residuals restricts the variance too much. It is assumed to be $\hat{m}_{t,x} N_{t,x}$, equal to the mean as a property of Poisson distributions.

A possible solution for this problem was found by Li, Hardy and Tan (2009). They model the total number of deaths of certain age in a given year as a sum of a number of stochastically weighted Poisson distributed deaths,

$$D_{t,x} = \sum_{i=1}^{P_x} D_{t,x}(i), \text{ with } D_{t,x}(i) | z_x(i) \sim \text{Poisson}(z_x(i) \hat{m}_{t,x} N_{t,x}).$$

Here, $P_x$ is the number of weights used, to be determined in advance, $z_x(i)$ are the weights. Denote the variance of the $z_x(i)$ by $\varphi_x$, then the resulting expression for the variance of $D_{t,x}$ becomes

$$\text{Var}(D_{t,x}) = \hat{m}_{t,x} N_{t,x} + \frac{P_x}{\varphi_x} (\hat{m}_{t,x} N_{t,x})^2$$

Now, the variance of the weights can act as an additional age-dependent parameter to be determined by maximum likelihood estimation, parallel to the other parameters. Of course, adding these parameters would improve goodness-of-fit at the cost of possible overdetermination, essentially the same dilemma as was described in section A.3. Information criteria would be needed to draw conclusions about the usefulness of additional parameters.
Bibliography


