Summary

This paper studies different stochastic mortality models with respect to its underlying assumptions. The Lee-Carter model, the rank-p SVD approximation model, the Weighted Least Squares model and the Poisson Bilinear model are discussed and applied to a Dutch data set. In an empirical analysis, it is illustrated how to obtain parameter estimates and how mortality rates can be forecasted. In a second step, the paper examines the model's underlying assumptions on a residual basis. Several tests are employed testing for normality, homogeneity of variance and autocorrelation. The test results seem to invalidate the models applicability due to failure in the underlying assumptions. In particular the crucial assumption of observational independence does not seem to hold, which may result in predicition intervals that are too narrow. Furthermore, a theoretical explanation for autocorrelation is given and an alternative model (multivariate ARIMA model) is proposed, that does not only rely on a weaker set of assumptions, but also requires less computational effort. Comparing the initial models' estimates with the estimates of the alternative model, it is concluded that all five stochastic mortality models tend to deliver similar estimates.

3.1 Introduction

Over the the last century, life expectancy increased significantly in developed countries. In particular, the Dutch experienced a great progress in longevity: whereas a Dutch new-born had a life expectancy of approximately 40 years in 1900, nowadays the life expectancy at birth is more than double. According to the [United Nations(2012)], the driving factors behind mortality reduction are medical accomplishments that progress through time. A profound knowledge of how longevity evolves over time is not only useful in the insurance business, but also for public retirement plans and medical care systems.

For the sake of prediction, it is tempting to extrapolate this historically stable trend and to project it to the future. Models that incorporate a stochastic component seem to be appealing instruments, because they result in prediction intervals stating how likely possible future values are. A mile-stone was set by Ronald Lee and Lawrence Carter in 1992: they proposed a stochastic model on the basis of a singular value decomposition and time series methods. Their model has become the most widely spread mortality prediction technique that is employed today by the U.S. Bureau of Census and by the United Nations (Girosi and King, 2007). Since Lee-Carter's publication, many researchers followed the stochastic approach and extended the original Lee-Carter model in various ways. Besides Lee and Carter, leading researchers in the field of stochastic mortality models are Wilmoth (weighted least squares), Brouhns et al. (Poisson bilinear) as well as Renshaw and Haberman (cohort effect). The mayority of the models rely on the method of maximum likelihood estimation assuming an underlying distribution and independence.

It is the purpose of this paper to investigate the validity of these underlying assumptions of established stochastic mortality models. In case of invalidity, the paper aims to explore the underlying causes leading to the assumptions' failure. Moreover, it attempts to find an alternative way of estimation based on a weaker set of assumptions.

The remaining part of the paper is organized as follows: Section 2 clearifies on the notation that is summarized in tabular form. In Section 3, several models are presented and their estimation and forecasting procedures are stated. In Section 4, three established models are fit to a Dutch data set and examined on a comparative basis. The subsequent Section 5 provides a goodness-of-fit analysis with focus on the models' residuals. In addition, a theoretical explanation for dependence is given. Section 6 discusses alternative approaches and thereupon presents the multivariate ARIMA model. Section 7 concludes the paper.

3.2 Notations

x	denotes the age group
t	denotes the calendar year
X	denotes the total number of age groups
T	denotes the total number of calendar years

$p_x(t)$	denotes the probability to survive for an individual at age x in year t
$q_x(t)$	denotes the probability to die for an individual at age x in year t
$e_x(t)$	denotes the expected remaining lifetime of an individual at age x in year t

a_x	denotes an age-specific constant
b_x	represents the age-specific patterns of change in mortality
k_t	denotes a time-varying index

$D_{x,t}$	number of deaths at age x in year t
$E_{x,t}$	exposure-to-risk at age x in year t
$m_{x,t}$	central death rate at age x in year t $(m_{x,t} \overset{\text{def}}{=} D_{x,t}/E_{x,t})$
$ln(m_{x,t})$	(natural-) log-central death rate at age x in year t
$\mu_{x,t}$	force of mortality at age x in year t

3.3 Stochastic Mortality Models

Mortality models are applied to data sets in order to estimate their model parameters. Of course, one cannot observe variables continuously, but only in discrete time. For instance the number of deaths can only be counted for intervals such as the number of deaths in the year interval $[t, t+1)$ and in the age interval $[x, x+1)$, where t and x are integers. Mortality data is available in tables, most frequently with the dimensions age and time. In lingo this is called a Lexis Plane. To overcome the discrete time problem, one needs to interpolate for variables that are continuous such as mortality rates or the force of mortality $\mu_{x,t}$. Therefore, mortality models usually assume that the force of mortality is constant within a cell. Having annual observations at each particular age, cell homogeneity is formally expressed by:

$$\mu_{x+\xi,t+\tau} = \mu_{x,t} \tag{3.1}$$

for all $\xi, \tau \in [0, 1)$ and x, t are integers. [Caselli, Vallin and Wunsch(2006)] show that the central death rate $m_{x,t}$ is a good raw estimator for the latent variable $\mu_{x,t}$. The central death rate $m_{x,t}$ plays an important role in most mortality models, because many variables of interest depend non-linearly on the force of mortality $\mu_{x,t}$.

3.3.1 Lee-Carter Model

An elegant approach that is simple to implement was proposed by [Lee and Carter(1992)]. In their original model, they set the central death rates in a logarithmic relation to its parameters. The model can be written as follows:

$$\ln(m_{x,t}) = a_x + b_x k_t + \varepsilon_{x,t}. \tag{3.2}$$

It is note-worthy that in this model a_x and b_x are age-specific constants that do not vary with time. In contrast, the parameter k_t is a time-varying index. The greatest advantage of the model is the easiness of parameter interpretation:

- a_x is an age-specific constant; $exp(a_x)$ is the general shape of the mortality schedule
- k_t is the time-varying overall mortality index
- b_x is the age-specific sensetivity to the time-varying index.

Another advantage of the model is that it prohibits negative central death rate values; Moreover, the model is quite popular as the computational effort is rather small compared to other models. Assuming that the error term $\varepsilon_{x,t}$ is a white noise process, i.e. $\varepsilon_{x,t} \overset{iid}{\sim} N(0, \sigma^2)$, one can estimate the parameters by maximum likelihood. The log-likelihood function is given by:

$$l(\mathbf{a}, \mathbf{b}, \mathbf{k}) = -\frac{TX}{2} ln(2\pi\sigma^2) - \frac{1}{2\sigma^2} \sum_{x,t} [\ln(m_{x,t}) - a_x - b_x k_t]^2. \tag{3.3}$$

Taking the first order derivatives with respect to a_x , b_x and k_t (for each x and each t) and setting these equations equal to zero leads to:

$$\sum_t [ln(m_{x,t}) - a_x - b_x k_t] = 0 \tag{3.4}$$

$$\sum_t k_t [ln(m_{x,t}) - a_x - b_x k_t] = 0$$

$$\sum_x b_x [ln(m_{x,t}) - a_x - b_x k_t] = 0.$$

It can be shown that finding maximum likelihood estimates for a_x, b_x and k_t is equivalent to minimizing the mean squared error. Therefore, consider (3.3): the constant part does not affect the optimization problem and can be dropped. Switching the sign of the second term while changing max to min, it is easy to see that the maximum likelihood estimates for a_x, b_x and k_t also minimize the mean squared error:

$$\min MSE = \sum_{x,t} \left[\ln\left(m_{x,t}\right) - a_x - b_x k_t\right]^2 . \tag{3.5}$$

As the alert reader might have noticed, there are infinitely many possibilities for a_x , b_x and k_t. Suppose one has found a_x^* , b_x^* and k_t^* that satisfy (3.5), it is easy to verify that $a_x' = a_x^*$, $b_x' = cb_x^*$ and $k_t^* = \frac{k_t^*}{c}$ for $c \in \mathbb{R}$ also solves the set of equations. To ensure identification of the parameters, Lee and Carter impose the following constraints:

$$\sum_t k_t = 0 \tag{3.6}$$

and

$$\sum_x b_x = 1. \tag{3.7}$$

Equation (3.6) combined with equation (3.4) implies that $\hat{a}_x = \frac{1}{T} \sum_t \ln\left(m_{x,t}\right)$. Estimates for b_x and k_t cannot be found by hand, but require a computational procedure called singular value decomposition. Lee-Carter estimated the model parameters in a three step approach:

1. estimate a_x by the mean log-rate for each x, i.e. $\hat{a}_x = \frac{1}{T} \sum_t \ln\left(m_{x,t}\right)$
2. estimate b_x and k_t by a first-rank SVD approximation
3. reestimate $\hat{k}_t$ to fit the observed number of deaths.

The first step is a simple spreadsheet calculation and needs no further explanation. The second step is a pure mathematical procedure. Estimates for b_x and k_t can easily be extracted after applying SVD to the de-meaned log central rate matrix. The detailed description is left to the appendix. In the third step, Lee and Carter adjust $\hat{k}_t$ in an ad hoc procedure to reproduce each year's actual number of deaths, i.e.

$$\sum_x D_{x,t} = \sum_x E_{x,t} exp(\hat{a}_x + \hat{b}_x \hat{k}_t).$$

They claim that the re-estimation is favorable as it avoids disparity between predicted and actual number of deaths. In the literature, this step is controversial, because by reestimating one gives up the maximum likelihood properties of $\hat{k}_t$.

3.3.2 Rank-p SVD Approximation

The rank-p SVD approximation model is a straight-forward extension of the original Lee-Carter model. [Renshaw and Haberman(2005)] as well as [Booth, Maindonald and Smith(2002)] altered Lee-Carter's procedure by carrying out a second-rank and fifth-rank SVD approximation respectively. The model can be expressed as follows:

$$\ln(m_{x,t}) = a_x + \sum_{i=1}^{p} b_x^{(i)} k^{(i)} + \varepsilon_{x,t}.$$

Note that the subscripts of the $k^{(i)}$'s are supressed as they differ for different values of i. It is common practice to use the SVD order for the $k^{(i)}$ such that $k^{(1)}$ carries the Lee-Carter time index t. For $i = 2$, Renshaw and Haberman set the index to $t - x$ to capture an observed cohort effect. Using a higher order SVD approximation one can, by construction, explain a higher amount of variance, however predicting higher order $k_t^{(i)}$'s is difficult as they may not follow linear patterns.

3.3.3 Weighted Least Squares

[Wilmoth(1993)] proposed another model using the same parameters suggested by Lee and Carter. Instead of minimizing (3.5), he advises to minimize the following equation:

$$\sum_{x,t} w_{x,t} \left[\ln(m_{x,t}) - a_x - b_x k_t\right]^2.$$

The model has two advantages: First, it can fit data sets that are constructed for specific death causes. In specific death cause data sets, some cells maybe equal to zero. To avoid taking the logarithm of zero, one can set the specific weight equal to zero. Second, it deals with heteroscedasticity observed in the Lee-Carter model. A common approach in WLS is to choose the weights equal to the inverse of the sample variance. Presuming that the number of deaths is exogeneous, Wilmoth proposes the weight $w_{x,t} = D_{x,t}$.

3.3.4 Poisson Bilinear Model

The Poisson bilinear model was proposed by [Wilmoth(1993)] based on [Brillinger's(1986)] approach to model the number of deaths, a counting variable, by a Poisson distribution. It is based on maximum likelihood and can be expressed as follows:

$$D_{x,t} \overset{ind}{\sim} Poisson\left(E_{x,t} m_{x,t}\right)$$

$$m_{x,t} = exp\left(a_x + b_x k_t\right)$$

Again, the model is overidentified such that the identification constraints (3.6) and (3.7) are needed. Note that this model allows for a heteroscedastistic error term that is crucial, according to Wilmoth, as the log central death rate fluctuates more at older ages due to small numbers of absolute deaths.

The parameters a_x, b_x and k_t can be estimated by maximum likelihood. The log-likelihood function is given by

$$l(\mathbf{a}, \mathbf{b}, \mathbf{k}) = \sum_{x,t} [D_{x,t}(a_x + b_x k_t) - E_{x,t} exp(a_x + b_x k_t)] + c, \tag{3.8}$$

where c is some constant. Finding parameters a_x, b_x and k_t that maximize (3.8) cannot be obtained using common statistical packages such as Stata or EViews, because a bilinear term is included. [Brouhns, Denuit and Vermunt(2002)] implemented Wilmoth's idea using a convenient program LEM, that is based on the Quasi-Newton method, to solve (3.8).

3.3.5 Forecast

Although [Lee and Carter(1992)] chose a different method of forecasting, today a conventional approch relies on bootstrapping. This method is analogue to [Brouhns, Denuit and Keilegoml(2005)]'s five step approach.

- *Step 1*
  generate N bootstrap sample tables

  Lee-Carter: $\ln\left(m_{x,t}^{n}\right) = \hat{a}_x + \hat{b}_x \hat{k}_t + \varepsilon_{x,t}^{n}$

  Rank-p SVD Approx.: $\ln\left(m_{x,t}^{n}\right) = \hat{a}_x + \sum_{i=1}^{p} \hat{b}_x^{(i)} \hat{k}^{(i)} + \varepsilon_{x,t}^{n}$

  WLS: $\ln\left(m_{x,t}^{n}\right) = \hat{a}_x + \hat{b}_x \hat{k}_t + w_{x,t}^{1/2} \varepsilon_{x,t}^{n}$

  Poisson: $D_{x,t}^{n} \sim Poisson\left(E_{x,t} e^{\hat{a}_x + \hat{b}_x \hat{k}_t}\right)$

- *Step 2*
  for each table n: reestimate the model parameters, i.e. $\hat{a}_x^n, \hat{b}_x^n$ and $\hat{k}_t^n$,

- *Step 3*
  for each n: estimate the ARIMA model and calculate potential values of k_{T+s}^{n} for a pre-specified period $[T, T+s]$

- *Step 4*
  for each n: calculate the point forecast for the variable of interest

- *Step 5*
  plot the N forecasts in a histogram and select an approximate $(1-\alpha)\%$ - interval

Bootstrapping is not a unique method, but there are various types of bootstrap schemes. For the Lee-Carter, the rank-p SVD approximation and the WLS model, one can either resample the models' residuals or use parametric bootstrap, i.e. $\varepsilon_{x,t}^{n} \overset{iid}{\sim} N\left(0, \hat{\sigma}^2\right)$. The parametric approach is also suitable for the Poisson bilinear model, finding $D_{x,t}^{n}$ by drawing independently from a Poisson distribution with mean $\left(E_{x,t} e^{\hat{a}_x + \hat{b}_x \hat{k}_t}\right)$. Regarding step three, note that one could allow for different ARIMA models. For instance, given n^*, one may get an ARIMA(1,1,0) whereas for a given n^{**} one may estimate an ARIMA(1,0,2). However, the common approach is that one only reestimates the parameters of the determined ARIMA model performed in the estimation stage. In step four, the *variable of interest* is not further specified. Usually, actuarials are interested in the expected remaining lifetime $e_x(t)$, the probability to survive $p_x(t)$ and/or the net single premium $a_x(t)$. Combining assumption (3.1) with the fact that the central death rate is an estimator for the force of mortality, [Brouhns, Denuit and Keilegoml(2005)] state that one can estimate the variables of interest by:

$$\hat{e}_x(t) = \frac{1 - exp\left(-\hat{m}_{x,t}\right)}{\hat{m}_{x,t}} + \sum_{k>0} \left(\prod_{j=1}^{k-1} exp\left(-\hat{m}_{x+j,t+j}\right) \right) \frac{1 - exp\left(-\hat{m}_{x+k,t+k}\right)}{\hat{m}_{x+k,t+k}}$$

$$\hat{p}_x(t) = exp\left(-\hat{m}_{x,t}\right)$$

and

$$\hat{a}_x(t) = \sum_{k>0} \left(\prod_{j=1}^{k-1} exp\left(-\hat{m}_{x+j,t+j}\right) \right) v^k$$

where v is the discount rate.

Generating N table samples is certainly computationally expensive. A less time-consuming approach was proposed by [Brouhns, Denuit and Vermunt(2002)] that avoids the generation of N tables and their model estimation. If the extraction method is based on maximum likelihood, one can draw N realizations from a multivariate normal distribution of the MLEs $\hat{a}_x, \hat{b}_x$ and $\hat{k}_t$. The question arises how to find variance estimates of such distributions. [Wilmoth(1993)] argued that variance estimates can be obtained by finding the second derivative of the log-likelihood function. Using the Delta method and exploiting the fact that MLEs are asymptotic efficient, the variance estimates can be extracted from the diagonal of the inverse of the Fisher information matrix. After drawing N realizations from the approximate multivariate normal distribution of the MLEs, one simply continues with step three mentioned above.

3.4 Empirical Analysis

In the following section, three of the four stochastic mortality models, discussed in the previous section, are implemented and set into comparison. Estimates of the traditional Lee-Carter (with and without the third step), the WLS and the Poisson bilinear model are calculated and analyzed.

3.4.1 Data

The fitting and forecasting is carried out on the basis of Dutch data. Data is provided by the [Human Mortality Database(2013)] and can easily be extracted from its website *www.mortality.org*. The human mortality database is a collection of all raw data from the national statistical offices. It provides Dutch data on central death rates, exposure-to-risk and number of deaths for different age and year intervals. The data is available for the full Dutch population as well as for gender subpopulations. To compare the three models, a set of Dutch data is selected that ranges from 1850 to 2009 with one-year intervals and covers the age groups from 0 to 98, i.e. 99 age groups. The sexes are combined and gender differences are disregarded. The selected data set is illustrated in a three-dimensional graph in Figure 3.1.

It is noteworthy, that the time range from 1850 to 2009 incorporates two world wars (1914-1918 and 1939-1945). Undoubtedly, the world wars have affected the mortality rates of the Dutch military and civil population. The question arises how to deal with the effect of the two world wars. This particular problem will be addressed in subsection 3.4.3.

3.4.2 Estimates

To obtain Lee-Carter parameter estimates for a_x, b_x and k_t, the two/three step procedure mentioned in subsection 3.3.1 was applied to the Dutch data set. The reestimation in the third step did not alter the k_t estimate significantly, thus no further distinction will be made with respect to the Lee-Carter k_t estimate. The WLS model was estimated by implementing [Wilmoth(1993)]'s normal equation algorithm. Different initial values for a_x, b_x and k_t were tried, all converging to the same result. The Poisson bilinear model was estimated by a computer program written by Vermunt.

Figure 3.2 shows the a_x estimates for all three models across age. It can be observed that the Lee-Carter, the WLS and the Poisson bilinear model produce nearly identical $\hat{a_x}$ estimates. Figure 3.3 shows the estimated relative speed of improvement, $\hat{b}_x$, for the three models across age. Again, the Lee-Cater, the WLS and the Poisson bilinear model seem to return smooth and alike estimates that reveal similar movements across age. Figure 3.4 illustrates the estimated value of time-varying index k_t from 1850 to 2009 for the Lee-Carter, the WLS and the Poisson bilinear model. It is remarkable that the WLS and the Poisson estimate nearly imitate each other. Furthermore, it is striking that all three models exibit peaks in the 1910s and 1940s. These peaks can be related to the first and second world war, that either directly (war causalities) or indirectly (e.g. starvation) increased the mortality rates. Although the Lee-Carter estimate follows the general movement of the other estimates, it deviates particularily in the most recent years that give rise to concern regarding forecasting: whereas the Lee-Carter $\hat{k}_t$ follows a moderate decrease since 1986, the Poisson and the WLS model estimate a sharp decrease in the time-varying index k_t. Nonetheless, all three k_t model estimates seem to follow a stochastic process with drift.

3.4.3 Forecasting k_t

In order to forecast the stochastic process of the time-varying index k_t an ARIMA time-series model is appealing. The question arises how to deal with the disruptions of the two world wars in the Dutch data set. Including the war years (1914-1918) and (1939-1945), all three k_t can be fit by an ARIMA(3,1,3) model. Sceptism is warranted as the model parameters seem not to be very stable. Excluding the war observations leads to an improved fit for all three k_t estimates. The estimated ARIMA(1,1,0) models are stable and are in line with other researchers' findings. Thus, it is explicitly assumed that a world war is a high unusual event that will not occur in the near future. An ARIMA(1,1,0) can be written as follows:

$$(1-L)(1-\phi L)k_t = \mu + \varepsilon_t, \tag{3.9}$$

where μ is the drift and $|\phi| < 1$. ε_t is assumed to follow a white noise process. Multiplying out and rearranging terms, equation (3.9) can be rewritten as:

$$k_t = (1+\phi)k_{t-1} - \phi k_{t-2} + \varepsilon_t. \tag{3.10}$$

Taking the expected value conditioned on the past and setting $t = T + s$ one receives

$$E(k_{T+s}|k_T) = (1+\phi)E(k_{T+s-1}|k_T) - \phi E(k_{T+s-2}|k_T),$$

which can be used to find point forecasts for k_{T+s} by an iterative procedure starting with $s = 1$.

Figure 3.5 shows the point forecasts of the time-varying indices for the year from 2010 until 2030 for the Lee-Carter, the WLS and the Poisson bilinear model. All four ARIMA(1,1,0) models have a similar trend, that are pairwise not statistically different from each other. For the Lee-Carter estimate the drift is -1.62, for the WLS estimate the drift is -1.64 and the drift for the Poisson bilinear model is -1.50. Therefore, the forecast values run almost parallel.

The next step in the routine of forecasting concerns prediction intervals. One distinguishes between marginal and simultaneous prediction intervals. Before explaining their differences, one needs another representation of k_t. Applying resubstitution to equation (3.10) one can see that k_t can be written as a linear function of past disturbances, i.e.

$$k_t = \left(\sum_{i=0}^{\infty} \lambda_i L^i\right) \varepsilon_t. \tag{3.11}$$

The λ_i's can be found by plugging (3.11) into the LHS and RHS of equation (3.10) and thereafter matching the coefficients. Following this procedure, one finds

- $\lambda_0 = 1$
- $\lambda_1 = 1 + \phi$
- $\lambda_i = (1 + \phi)\lambda_{i-1} - \phi\lambda_{i-2}$ for $i \geq 2$.

These λ_i's are components of both, marginal and simultaneous prediction intervals. [Box and Jenkins(1976)] have examined marginal $1 - \alpha$ intervals and constructed the s-step forecast interval the following way:

$$\left[E(k_{T+s}|k_T) \pm c_{\alpha/2} \left(\sum_{i=0}^{s-1} \lambda_i^2 \right)^{1/2} \sigma_\varepsilon \right],$$

where $c_{\alpha/2}$ is the critical value of the standard normal distribution. Combining the forecast intervals, one obtains an area. [Ravishanker, Hochberg and Melnick(1987)] reason that this marginal $1 - \alpha$ prediction interval is too optimistic as it does not imply that all future forecast values *jointly* will be in that defined area with a probability of $1 - \alpha$. Therefore, Ravishanker et al. introduced simultaneous ARIMA prediction intervals that are based on the joint probability distribution of the forecast errors. Its s-step forecast interval is constructed in the following way:

$$\left[E(k_{T+s}|k_T) \pm c \left(\sum_{i=0}^{s-1} \lambda_i^2 \right)^{1/2} \sigma_\varepsilon \right],$$

where c needs to be approximated. This can be achieved by Monte Carlo simulation. A more elegant approximation procedure was proposed by [Ravishanker, Hochberg and Melnick(1987)] using the Hunter bound and solving a simple maximum spanning tree problem with Kruskal's algorithm. Figure 3.6 presents the marginal and simultaneous 95% - prediction intervals of the Lee-Carter, the WLS and the Poisson k_t estimates. As expected, the simultaneous 95% - prediction intervals are wider than the marginal 95% - prediction intervals. Regarding the Poisson forecast intervals, the difference between marginal and simultaneous is relatively small. In contrast, the difference between the marginal and the simultaneous forecast interval of the Lee-Carter model seems to be much larger. This dissimilar difference in marginal and the simultaneous forecast intervals is attributed to the higher variance estimate in the ARIMA model.

3.5 Goodness of Fit

To test the models' underlying assumptions, the following section presents a goodness-of-fit analysis applied to the models estimated in the previous section. The following goodness-of-fit is in accord with Dowd et al. (2010)'s residual analysis and is based on four criteria:

1. the fraction of variance explained in each model
2. test for homogeneity of variance
3. test for normality
4. autocorrelation.

3.5.1 Residuals

Because the three models applied differ in their dependent variable, it is crucial to clearify how the residuals are calculated. The dependent variable in the Lee-Carter and in the WLS model is $\ln(m_{x,t})$. The fitted value in these models can be calculated by $\ln(\hat{m}_{x,t}) = \hat{a}_x + \hat{b}_x\hat{k}_t$. The Lee-Carter residuals $\hat{\varepsilon}_{x,t}$ can be calculated directly by substracting the fitted value from $\ln(m_{x,t})$. Regarding the WLS, one additionally needs to divide by $w_{x,t}^{1/2}$ after substracting. In reference to the assumptions stated in subsection three, the Lee-Carter and the WLS residuals follow a normal distribution. In contrast, the dependent variable in the Poisson bilinear model is the number of deaths $D_{x,t}$. The fitted value is denoted by $\hat{D}_{x,t}$ and is equal to $E_{x,t}exp(\hat{a}_x + \hat{b}_x\hat{k}_t)$. One possibility would be to standardize the Poisson residuals by using Pearson's approach, i.e.

$$\frac{D_{x,t} - \hat{D}_{x,t}}{\hat{D}_{x,t}^{1/2}}.$$

[Pierce and Schafer(1986)] showed that it is more appealing to standardize using Anscombe's approach in the case of a Poisson distribution:

$$\frac{D_{x,t}^{\frac{2}{3}} - \hat{D}_{x,t}^{\frac{2}{3}} + \frac{1}{9}\hat{D}_{x,t}^{-\frac{1}{3}}}{\frac{2}{3}\hat{D}_{x,t}^{\frac{1}{6}}}$$

leaving residuals that follow an approximately normal distribution. Figures **??** to **??** give scatter plots of the (standardized) residuals with respect to age and year in three-dimensional space. In all three models residual outliers can

Lee-Carter	**WLS**	**Poisson**
99.70%	99.56%	99.03%

Table 3.1: Fraction of Explained Variance

be detected for the world war years $1914-1918$ and $1939-1945$. Although the world war years are taken into account via the time-varying index k_t by increasing the general mortality for these time periods, it is clear that the world wars have affected distinguished age groups in a different way. Undoubtedly, during war the mortality of people that potentially be drafted into the army is higher than the mortality of people being out of the age-range to be called to be recruited. This effect cannot be captured by b_x. Thus, in the subsequent subsections the tests are applied to a subset of the residuals excluding the war years.

3.5.2 Explained Variance

A straight-forward indicator for the goodness of fit is the fraction of explained variance. This is calculated by the fitted sum of squares divided by the total sum of squares, i.e.

$$\frac{\sum_{x,t}(f_{x,t}-\hat{f}_{x,t})^2}{\sum_{x,t}(f_{x,t}-\bar{f})^2}.$$

Table 3.1 shows the fraction of explained variance. It can be infered that all three models seem to explain a high proportion of variance in their dependent variables. Keep in mind that the dependent variable of the Poisson bilinear model differs from the one of the other two models prohibiting a direct comparison.

3.5.3 Test for Homogeneity of Variance

All three models imply that the adjusted residual series across ages are homogeneous with respect to their variance. To test this implication one can employ [Hartley(1950)]'s variance ratio test for each of the three models. Given a specific model, the sample variance s_x^2 is calculated for each age group x. The Hartley's test statistic is

$$V_H = \frac{max(s_x^2)}{min(s_x^2)}.$$

Lee-Carter	**WLS**	**Poisson**
$V_H = 28.32\ p = 0.000$	$V_H = 27.53\ p = 0.000$	$V_H = 2.06\ p = 0.041$

Table 3.2: Hartley's Test

Test	**Lee-Carter**	**WLS**	**Poisson**
K.-Smirnov	$p = 0.000$	$p = 0.000$	$p = 0.000$
Lilliefors	$V_L = 0.052$ $p = 0.000$	$V_L = 0.058$ $p = 0.000$	$V_L = 0.051$ $p = 0.000$
Jarque-Bera	$V_{JB} = 50786$ $p = 0.000$	$V_{JB} = 22891$ $p = 0.000$	$V_{JB} = 85340$ $p = 0.000$

Table 3.3: Normality Tests

Under the null hypothesis of homogeneity, the critical values for 5% and 1% rejection rate, given a 99 x 148 residual matrix (12 war years excluded), are 2.04 and 2.19 respectively. Table 3.2 summarizes the test results. One can infer that the (standardized) residual series of the Lee-Carter and the WLS model do not have the same variance. In contrast, the Poisson bilinear model performs better in this test. Its null hypothesis of homogeneous variance can only be rejected at a 5% - significance level, but not at a 1% - significance level.

3.5.4 Normality Tests

The adjusted residuals of the three models can be tested for normality by a couple of tests. [Park(2008)] enlists a group of tests that are applicable: a common test is the Kolmogorov-Smirnov test that is based on the distance between the empirical distribution function and the (normal) cumulative distribution function. An adaption of the Kolmogorov-Smirnov test is the Lilliefors test, that does not specify the mean and variance of the normal distribution. Another common place test is the Jarque-Bera test. Although, it's validity is controversial, because it only checks the third (skewness) and forth central moment (kurtosis), it still serves for indication. Table 3.3 summarizes the test statistic and/or the respective p-value for each test/model combination. The test results unanimously reject the null hypothesis of normality for all three models at any conventional significance level. Combining these results with the results of the test for homogeneity in variance, one can conclude that the models' underlying distribution assumptions do not hold.

	Lee-Carter	**WLS**	**Poisson**
t - Autocorrelation	92.93%	95.96%	97.98%
x - Autocorrelation	91.22%	93.92%	96.62%
t-x - Autocorrelation	86.88%	91.88%	93.75%

Table 3.4: Autocorrelations

3.5.5 Autocorrelation Test

The Lee-Carter, the WLS and the Poisson bilinear model imply that the residuals are independent. To verify or to invalidate this implication one can test for serial autocorrelation in the residuals over the time span from 1850 to 2009 for each age group. Although autocorrelation can occur in higher lag orders, this section examines only the first order lag case by using [Durbin-Watson(1950)]. In the 99 x 148 residual matrix, one can test for autocorrelation with respect to three different axes. The first axis is the calendar year t (horizontal); the second axis is the age x (vertical); a third axis exists, called cohort, defined by $t - x$. It refers to the diagonal of the matrix and was already investigated by [Renshaw and Haberman(2005)]. Table 3.4 shows the rejection rate of the Durbin-Watson test evaluated at a 1% significance level for all nine model/axis combinations. The null hypothesis can clearly be rejected for all nine model/dimension combinations. Under the null-hypothesis of no serial correlation in the first order lag, one would have expected a value around 1%. It can be inferred that all three residual matrices produced by the different models have serial correlation in the first order lag in all three axes.

Having a closer look at the autocorrelation-function (ACF) and the partial autocorrelation function (PACF) across age, year and cohort, one finds that the ACFs decrease linearly, whereas the PACFs are caracterized by positive peaks in the first lag. This observation suggests that only first-order autocorrelation is persistent in the residuals, but not at higher lags.

The tests' results strongly indicate that the assumption of observational independence, presumed by the models, does not hold. In Figures **??** to **??**, the waving-behavior in the residuals also indicates that autocorrelation is persistent. The consecutive subsection provides a theoretical explanation for the observed autocorrelation.

3.5.6 A Theoretical Explanation for Autocorrelation

[Renshaw and Haberman(2005)] also observed a non-random pattern in the residuals, which they called cohort effect. They argue that the effect comes from following the same generation across a cohort. To capture this effect, they added a bilinear term to the Lee-Carter model. Although this modification of the Lee-Carter model possibly explains some autocorrelation across the cohort-axis, it cannot explain why autocorrelation is present across the age- and the time-axis.

There is a much more intuitive explanation for autocorrelation that is related to the way how the data is acquired. Figure 8 illustrates seven individuals drawn into the Lexis Plane with time on the x-axis and age on the y-axis.

Each individual is represented by a colored line that creates a $45\,^{\circ}$ angle with the x-axis. The $45\,^{\circ}$ angle comes from the fact that an increase in time by τ goes along with an increase in age by τ. The individuals' lines do not coincide, but are parallelly shifted due to different dates of birth. A small square at the end of an individual's line denotes the time/age - coordinate at which his or her death occured. The highlighted square in the center shows the Lexis Plane cell in which people are aged x in the calendar year t.

Observing one particular individual, say individual A, at each point on A's line, A has a certain force of mortality, $\mu_{x,t}^{(A)}$. It is rational to assume that for a sufficently small τ, the force of mortality at time $t+\tau$ (when A is aged $x+\tau$) is fairly close to A's force of mortality at time t, i.e.

$$\mu_{x+\tau,t+\tau}^{(A)} \approx \mu_{x,t}^{(A)}.$$

Generalizing this idea to a group G of individuals, one has $\mu_{x+\tau,t+\tau}^{(G)} \approx \mu_{x,t}^{(G)}$. Note that five out of the seven individuals, that were observed in the highlighted square, are again observed in square 2. Renshaw and Haberman observed exactly this process along the cohort. Moreover, approximately half of the seven individuals are also observed in square 1 (orange, green, yellow) and by analogue arguments approximately half are also observed in square 3 (red, purple, turquoise). If a group's average force of mortality at time $t+\tau$ is well-approximated by $\mu_{x,t}^{(G)}$ for small τ, then one actually expects first-order autocorrelation in the data. This effect of following the same group of people is far more general than Renshaw-Haberman's cohort effect. Take note that this effect cannot be captured by the current parameters a_x, b_x and k_t, since a_x only captures the general shape of the mortality schedule, k_t the general decrease of the mortality rates over time and b_x only the sensitivity to this time trend.

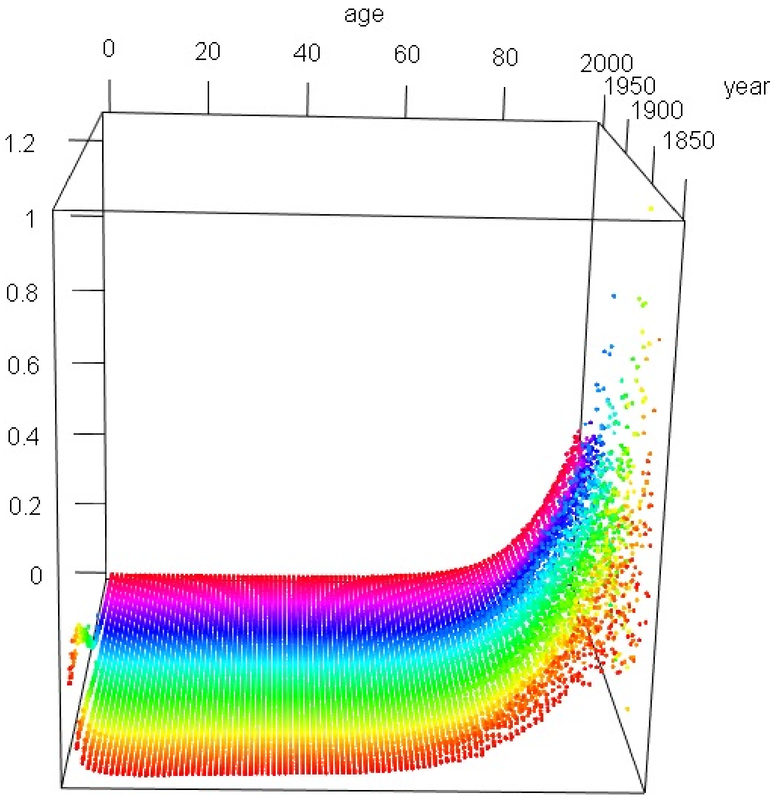


Figure 3.1: Dutch central death rates $m_{x,t}$

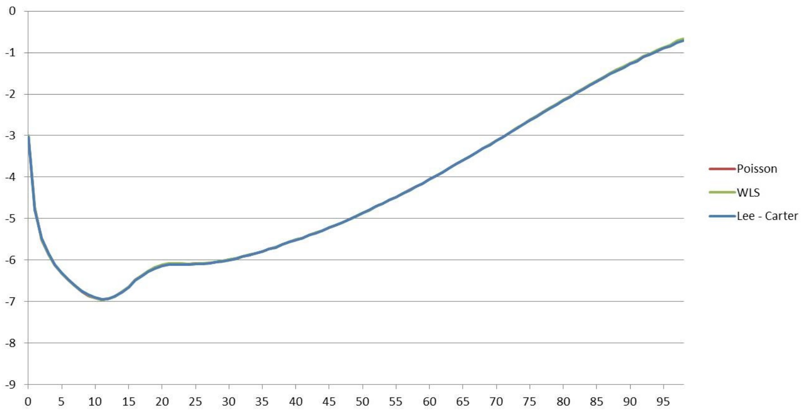


Figure 3.2: Estimates of parameter a_x

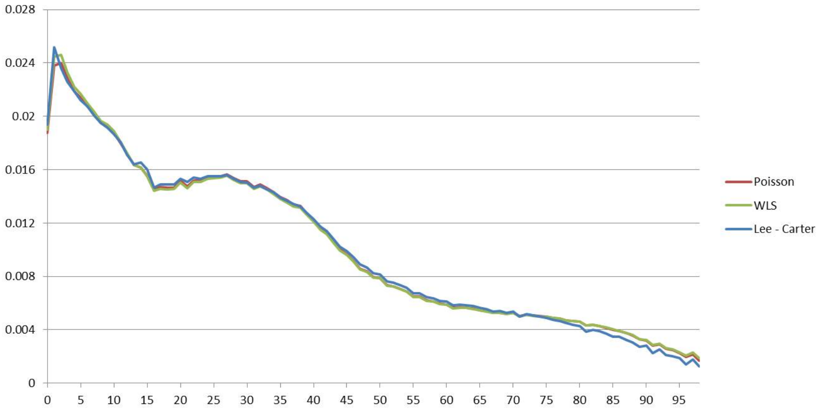


Figure 3.3: Estimates of parameter b_x

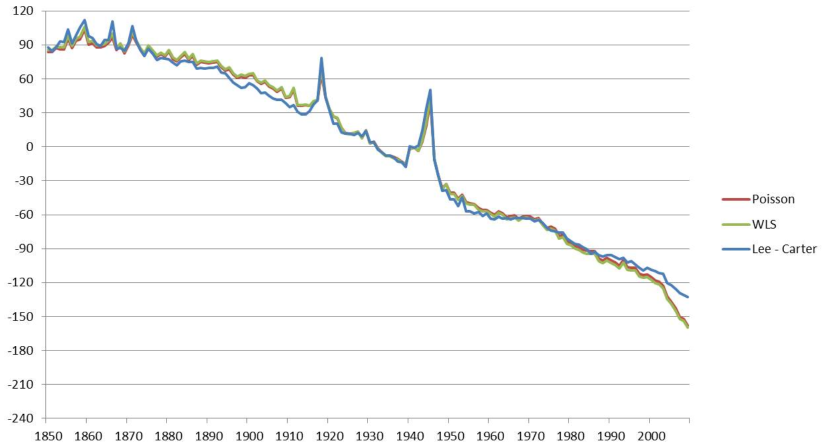


Figure 3.4: Estimates of parameter k_t

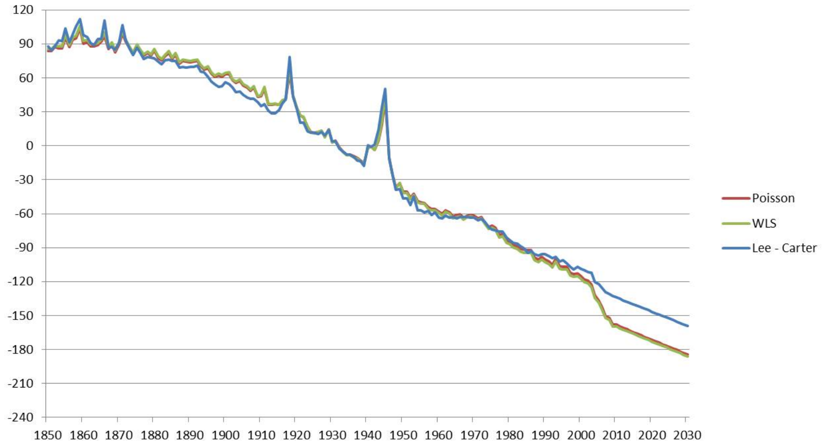
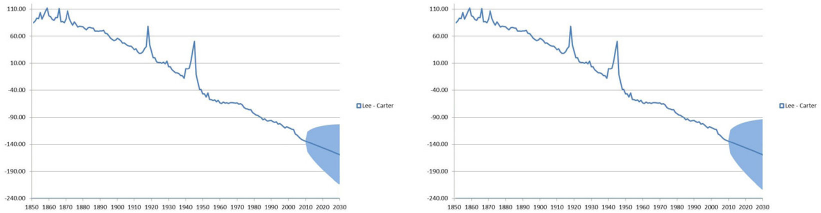
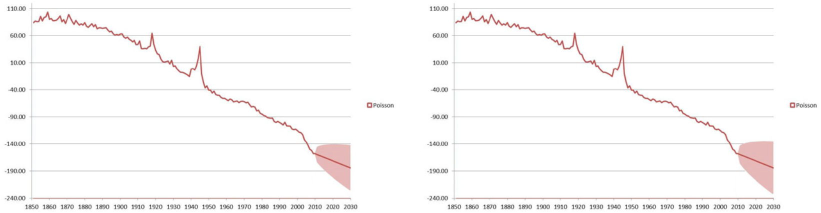


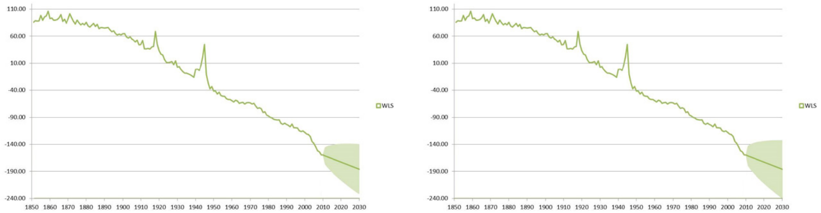
Figure 3.5: Point Forecasts of Parameter k_t



(a) Lee-Carter: Marginal 95% - P. I. (b) Lee-Carter: Simultaneous 95% - P. I.

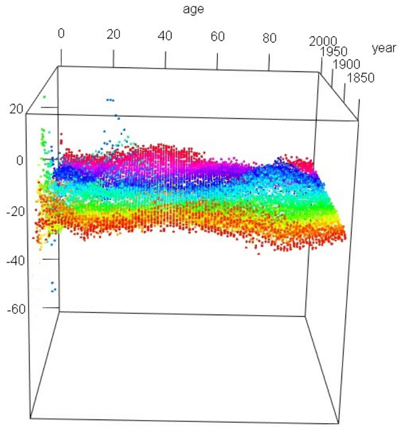


(c) Poisson: Marginal 95% - P. I. (d) Poisson: Simultaneous 95% - P. I.

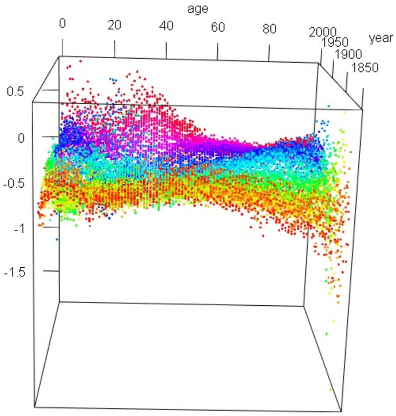


(e) WLS: Marginal 95% - P. I. (f) WLS: Simultaneous 95% - P. I.

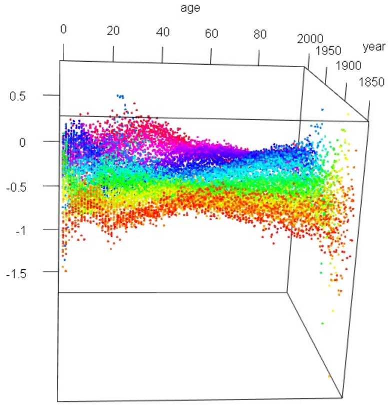
Figure 3.6: 95% - Prediction Intervals for k_{T+s}



(a) Poisson Anscombe residuals



(b) WLS residuals



(c) Lee-Carter residuals

Figure 3.7: (Standardized) Residuals

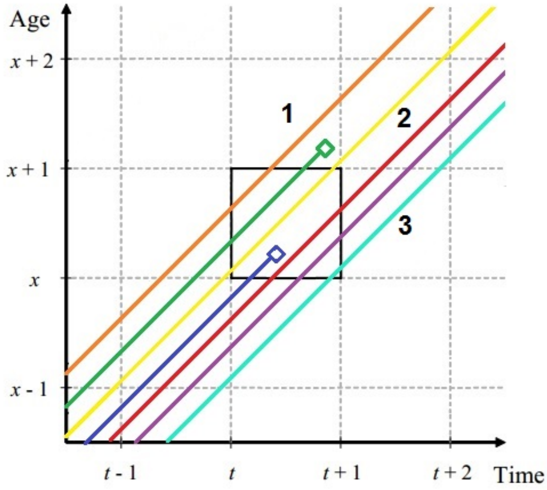


Figure 3.8: Lexis Plane

Bibliography

[Booth, Maindonald and Smith(2002)] Booth, H., Maindonald, J. and Smith, L. (2002). *Applying Lee-Carter under Condition of Variable Mortality Decline.* Population Studies, Vol. 56, No. 3, 325-336.

[Box and Jenkins(1976)] Box, G. E. P., and Jenkins, G. M. (1976). *Time Series Analysis Forecasting and Control.* San Francisco: Holden-Day.

[Bühlmann(1997)] Bühlmann, P. (1997). *Sieve Bootstrap for Time Series.* Bernoulli, 3, 123-148.

[Bühlmann(1998)] Bühlmann, P. (1998). *Sieve Bootstrap for Smoothing Nonstationary Time Series.* Annals of Statistics, 26, 48-83.

[Brillinger's(1986)] Brillinger (1986). *The natural variability of vital rates and associated statistics.* Bimetrics, Vol. 42, No. 4, 693-734.

[Brouhns, Denuit and Keilegoml(2005)] Brouhns, N., Denuit, M., Keilegom, I.V.(2005). *Bootstrapping the poisson Log-bilinear Regression Approach to the Construction of Projected Lifetables.* Scandinavian Actuarial Journal, 3, 212-224.

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

[Caselli, Vallin and Wunsch(2006)] Caselli, G.,Vallin, J., Wunsch, G. J.(2006). *Demography: Analysis and Synthesis.* Elsevier, Vol. 1, 117-119.

[Dowd et. al(2010)] Dowd et. al (2010). *Evaluating the goodness of fit of stochastic mortality models.* Insurance: Mathematics and Economics, 47, 255-265.

[Durbin-Watson(1950)] Durbin, J., and Watson, G. S. (1950). *Testing for Serial Correlation in Least Squares Regression, I.* Biometrika 37, 409-428.

[Eckart and Young(1936)] Eckart, C. and Young, G. (1936). *The approximation of one matrix by another of lower rank.* Psychometrika, Vol. 1, No. 3, 211-218.

[Girosi and King(2007)] Girosi F. and G. King(2007). *Understanding the Lee-Carter Mortality Forecasting Method.* Working paper, Harvard University, USA.

[Hartley(1950)] Hartley, H.O. (1950). *The Use of Range in Analysis of Variance.* Biometrica Trust, Vol. 37, No. 3/4, 271-280.

[Human Mortality Database(2013)] Human Mortality Database (2013). *Netherlands: Complete Data Series.* Retrieved from http://www.mortality.org on 10 April, 2013.

[Lee and Carter(1992)] Lee R.D. and L.R. Carter (1992). *Modeling and Forecasting U. S. Mortality.* Journal of the American Statistical Association, Vol. 87, No. 419, 659-671.

[Madsen, Hansen and Winther(2004)] Madsen, R.E., Hansen, L.K., Winther, O. (2004). *Singular Value Decomposition and Principal Component Analysis.*

[Park(2008)] Park, H. M. (2008). *Univariate Analysis and Normality Test Using SAS, Stata, and SPSS.* Indiana University, Indiana. USA.

[Pierce and Schafer(1986)] Pierce, D. A. and Schafer, D. W. (1986). *residuals in Generalized Linear Models.* Journal of the American Statistical Association, Vol. 81, No. 396, 977-986.

[Ravishanker, Hochberg and Melnick(1987)] Ravishanker, N., Hochberg, Y. , Melnick, E. L. (1987). *Approximate Simultaneous Prediction Intervals for Multiple Forecasts* Technometrics,Vol. 29, No. 3, 371-376.

[Renshaw and Haberman(2005)] Renshaw, A. and Haberman, S. (2005). *Mortality Reduction Factors Incorporating Cohort Effects.* Actuarial Research Paper, Cass Business School, London. UK.

[United Nations(2012)] United Nations, Department of Economic and Social Affairs, Population Division (2012). *World Mortality Report 2011.* United Nations publication.

[Wilmoth(1993)] Wilmoth J.R. (1993). *Computational Methods for fitting and Extrapolating the Lee-Carter Model of Mortality Change.* Technical Report, University of California, Berkeley. USA.