Modeling and forecasting duration-dependent mortality rates

Marcus C. Christiansen
Institut für Versicherungswissenschaften
Universität Ulm
D-89069 Ulm, Germany

Andreas Niemeyer
Institut für Versicherungswissenschaften
Universität Ulm
D-89069 Ulm, Germany

Lucia Teigiszerová
Institut für Versicherungswissenschaften
Universität Ulm
D-89069 Ulm, Germany

17th February 2014

Abstract: We study mortality data of disabled individuals and discuss parametric modeling approaches for the force of mortality. Empirical observations show that the duration since disablement has a strong effect on mortality rates. We incorporate duration effects by generalizing the Lee-Carter model, which is commonly used in mortality forecasting. For each proposed model, we discuss uniqueness and fitting techniques and calibrate them to mortality observations of the German Pension Insurance. Difficulties with coarse tabulation of the empirical data are solved by an age-period-duration Lexis diagram. Forecasting is demonstrated for an exemplary model, showing that duration dependence should not be neglected. While we see a clear longevity trend with respect to age, significant fluctuations but no systematic trend is observed for the duration effects.

Keywords: stochastic mortality; duration dependence; disability insurance; Lee-Carter model; Lexis diagram

1. Introduction

For the forecasting of future mortality rates, stochastic models have become more and more popular in demographic research, since they allow for describing not only a future best estimate development but also the uncertainty around a best estimate forecast. Applications range from governmental planning to private business, in particular insurance. In a seminal paper, Lee and Carter (1992) suggested a parametric model that allows stochastic forecasting of a populations mortality with respect to age and calendar year. Apart from being the first model of its kind, the popularity of this approach comes from (a) the reasonable fit for most Western countries, (b) the intuitive interpretation of the models parameters and (c) the simplicity and ease of use in practice. A number of authors further developed
the model of Lee and Carter (1992), for example Brouhns et al. (2002), Cairns et al. (2006), Hyndman and Ullah (2007), and Renshaw and Haberman (2006).

In the present paper we study the mortality of disabled people. The results are relevant for, e.g., the planning of health care infrastructure and the risk management of disability insurance portfolios. Based on empirical data, we suggest and discuss several stochastic forecasting models. All our models are extensions of the Lee-Carter approach. This way we try to keep the major advantages of the Lee-Carter approach, in particular the simple calibration, the fact that all parameters have an economic meaning, and the intuitive forecasting into the future. The most important covariates for the mortality rate of healthy people are age and calendar time. For disabled people, several studies have shown that the duration since disablement is a significant covariate (compare e.g. Segerer, 1993). Our data set confirms this observation. While the Lee-Carter model neglects duration dependence, all our extensions feature this covariate.

Our empirical study bases on a data set from the Research Data Center of the German Pension Insurance (Forschungsdatenzentrum der Rentenversicherung, FDR-RV, 2011), which contains mortality data for disabled policyholders in the German public disability insurance in the years 1994 to 2009. All our proposed parametric models are judged on the basis of this data set. Our results show that it is inevitable to consider duration dependence in order to achieve a reasonable fit to the empirical observations.

While the literature offers various Lee-Carter extensions, none of them features duration dependence for disabled people. Pitacco (2012) discusses duration dependence in the mortality of disabled people, looking from the perspective of private disability insurance, but he disregards calendar time dependence. Interestingly, Pitacco (2012) presents a model that links the mortality of disabled people with the mortality of healthy people, supposing some functional dependence. Some of our models discussed later on also have such a structure. Aro et al. (2013) model the termination probabilities of a disability insurance using logistic regression and fit there model to data of a Swedish insurance company. Renshaw and Haberman (2000) describe permanent health insurance with a Poisson regression. Both papers take into account duration dependence, but they do not show how their models can be forecast into the future. The classical Lee-Carter approach uses singular value decomposition in order to decompose the age and time matrix to single age and time vectors. Russolillo et al. (2011) extends the Lee-Carter model by a third factor that relates to the country and uses the Tucker3 method to decompose the age, time and country tensor to single age, time and country vectors. Unlike the singular value decomposition, the Tucker3 method is rather tricky and adds new problems. Christiansen et al. (2012) use a multivariate Lee-Carter generalization proposed in Hyndman and Ullah (2007) to model mortality, disability and reactivation in a disability insurance, but they also neglect duration dependence.

The paper is structured as follows. In Section 2 we explain our data set, we show how we calculate the empirical mortality rates, and we give some summary statistics. In Section 3 we fit the classical Lee-Carter model to our data. The main part of the paper is Section 4, where different extensions of the Lee-Carter model are proposed and discussed. The proofs of the propositions from Section 4 are presented in Section 5. In Section 6 we explain how to do forecasts, demonstrating the estimation of confidence intervals for one of our proposed models. Section 7 concludes.

2. Description of the data and estimation of the empirical mortality rates

In this section, we first describe the data set and then how the empirical mortality rates can be estimated from that data. Our data set comes from the Research Data Center of the German Pension Insurance (Forschungsdatenzentrum der Rentenversicherung, FDR-RV, 2011) and includes 1% of the total portfolio of pensioners at the end of the years 1993 to 2008 (sample size 282,415) and 10% of the discontinuations of pensioners due to mortality in the years 1994 to 2009 (sample size 85,131). The 1% subset and the 10% subset were randomly chosen from the total data. We only consider reduced earnings capacity pensions, which offer an income protection in case of partial or total disability. The data contains various covariates, but we consider only the age, the calendar time, and the duration
since disablement. In particular, we do not consider the gender. The German Pension Insurance payed the pensions monthly in advance, so that the last annuity within a year was paid at the beginning of December. An insured person was classified as being alive if he survived till the end of November. Analogously, if an insured person died in December, he was counted as dead in January. Therefore, we shift the calendar years by one month such that the year starts at the beginning of December and ends at the end of November.

For the duration \( d \) since disablement we consider 7 different classes, where the first class 0 means that the person is up to one year disabled, 1 between one and two years, and 6 between six and seven years. For durations of more than seven years the observed data is scarce, so we skip it. The data gives the duration only on a yearly time grid and, consequently, we can only study discrete durations. We study policyholders with age \( x \) between 40 and 59. For younger ages the data basis is small, and people with age 60 or older can alternatively receive old age pension benefits so that adverse selection effects interfere the data. In total, the covariates and their domains are

\[
x \in \{40, \ldots, 59\}, \ t \in \{1994, \ldots, 2009\}, \ d \in \{0, \ldots, 6\}.
\]

In the following theoretical discussion we use the general notation \( x \in \{x_1, \ldots, x_k\}, d \in \{d_1, \ldots, d_l\}, t \in \{t_1, \ldots, t_m\} \). We now discuss the estimation of the mortality rates \( m_{x dt} \). We generally assume that \( m_{x dt} \) is constant in between integer ages, integer durations and integer years, that is,

\[
m_{x dt} = m_{[x]} [d] [t], \quad x_1 \leq x < x_k + 1, \ d_1 \leq d < d_l + 1, \ t_1 \leq t < t_m + 1.
\]

Let \( \tilde{L}_{x dt} \) be the number of observed individuals at the beginning of year \( t \in \{t_1, \ldots, t_m\} \) that have age \( x \in \{x_1, \ldots, x_k\} \) and belong to duration class \( d \in \{d_1, \ldots, d_l\} \). Let \( \tilde{D}_{x dt} \) be the number of deaths observed in the time interval \([t, t+1)\) of individuals that have age \( x \) at death and that got disabled in year \( t - d \). Our data set offers the numbers \( \tilde{L}_{x dt} \) and \( \tilde{D}_{x dt} \) for \( x \in \{39, \ldots, 59\}, t \in \{1994, \ldots, 2009\}, \) and \( d \in \{0, \ldots, 6\} \). It is tempting to use the fraction \( \tilde{D}_{x dt} / \tilde{L}_{x dt} \) over \( \tilde{L}_{x dt} \) as an estimator for \( m_{x dt} \), but this would lead to a systematic error. We now explain how to avoid that systematic error by transforming \( \tilde{D}_{x dt} \) and \( \tilde{L}_{x dt} \) to an adequate death count \( D_{x dt} \) and exposure to risk \( L_{x dt} \).

Suppose that for each individual we observe the exact date of birth \( b \), the exact time of disablement \( e > i \). Then \( D_{x dt} \) corresponds in a three-dimensional Lexis diagram with dimensions \( b, i, d \) to the solid

\[
S(\tilde{D}_{x dt}) = \{(b, i, e) : b < i < e, t \leq e < t+1, x \leq e - b < x + 1, d - 1 \leq t - i < d\}.
\]

The number \( \tilde{L}_{x dt} \) can be interpreted as the exposure that corresponds to the solid

\[
S(\tilde{L}_{x dt}) = \{(b, i, e) : b < i < e, t \leq e < t+1, x \leq t - b < x + 1, d \leq t - i < d + 1\}.
\]

Since the three-dimensional Lexis diagram is difficult to illustrate, we show two two-dimensional projections in Figure 2.1 and Figure 2.2. The group of people \( \tilde{L}_{x dt} \) moves horizontally through the diagrams in Figure 2.1 and Figure 2.2, creating rectangular boxes. The deaths \( \tilde{D}_{x dt} \) correspond in Figure 2.2 to a regular box, but correspond in Figure 2.1 to the gray area. However, for estimating \( m_{x dt} \) we are rather interested in

\[
S(L_{x dt}) = \{(b, i, e) : b < i < e, t \leq e < t+1, x \leq e - b < x + 1, d \leq e - i < d + 1\},
\]

which describes the exposure of individuals in year \( t \) with age \( x \) and duration class \( d \), and

\[
S(D_{x dt}) = \{(b, i, e) : b < i < e, t \leq e < t+1, x \leq e - b < x + 1, d \leq e - i < d + 1\},
\]

which corresponds to the deaths at age \( x \) in duration class \( d \) in year \( t \). Both quantities, \( S(L_{x dt}) \) and \( S(D_{x dt}) \) have in Figures 2.1 and 2.2 to form of the gray areas. Once we have \( D_{x dt} \) and \( L_{x dt} \), we can estimate \( m_{x dt} \) by

\[
\hat{m}_{x dt} = \frac{D_{x dt}}{L_{x dt}}.
\]
Analogously to Pitacco et al. (2009, Section 3.3.3) we can show that is a maximum likelihood estimator for the mortality rate within a reasonable model.

In order to transform $\tilde{D}_{x,d,t}$ and $\tilde{L}_{x,d,t}$ to $D_{x,d,t}$ and $L_{x,d,t}$ we assume that within the solids $S(\tilde{D}_{x,d,t})$, $S(D_{x,d,t})$ and $S(\tilde{L}_{x,d,t})$, $S(L_{x,d,t})$ the deaths and the exposure to risk are approximately uniformly distributed. Then we can transform the number of deaths and the risk exposures by building the solids $S(D_{x,d,t})$ and $S(L_{x,d,t})$ as the union of the intersections with the solids $S(\tilde{D}_{x,d,t})$ and $S(\tilde{L}_{x,d,t})$,

$$L_{x,d,t} = \sum_{\tilde{x},\tilde{d},\tilde{t}} \frac{\text{Vol}(S(L_{x,d,t}) \cap S(\tilde{L}_{\tilde{x},\tilde{d},\tilde{t}}))}{\text{Vol}(S(\tilde{L}_{\tilde{x},\tilde{d},\tilde{t}}))} \tilde{L}_{\tilde{x},\tilde{d},\tilde{t}} = \frac{1}{3} \tilde{L}_{x,d,t} + \frac{1}{6} \tilde{L}_{x-1,d,t} + \frac{1}{6} \tilde{L}_{x,d-1,t} + \frac{1}{3} \tilde{L}_{x-1,d-1,t},$$

$$D_{x,d,t} = \sum_{\tilde{x},\tilde{d},\tilde{t}} \frac{\text{Vol}(S(D_{x,d,t}) \cap S(\tilde{D}_{\tilde{x},\tilde{d},\tilde{t}}))}{\text{Vol}(S(\tilde{D}_{\tilde{x},\tilde{d},\tilde{t}}))} \tilde{D}_{\tilde{x},\tilde{d},\tilde{t}} = \begin{cases} \frac{1}{2} \tilde{D}_{x,d,t} + \frac{1}{2} \tilde{D}_{x,d+1,t} & \text{for } d \geq 1, \\ \tilde{D}_{x,0,t} + \frac{1}{2} \tilde{D}_{x,1,t} & \text{for } d = 0. \end{cases}$$

Because of the uniform distribution assumption for solids, the number of deaths in the intersection $S(L_{x,d,t}) \cap S(\tilde{L}_{\tilde{x},\tilde{d},\tilde{t}})$ equals the number of deaths in $S(\tilde{L}_{\tilde{x},\tilde{d},\tilde{t}})$ times the percentage of the volume shared. Likewise we can justify the formula for the exposures. Note that $\text{Vol}(S(\tilde{D}_{\tilde{x},0,t})) = \frac{1}{2}$ while all other solids have a volume of one. An analogous transformation approach is used for age-period-cohort models, see e.g. Carstensen (2007). We still have the problem that $\tilde{L}_{x,-1,t}$ is needed for $L_{x,0,t}$, but we have no data for that. Therefore, in the following we approximate $L_{x,-1,t}$ by $\tilde{L}_{x,0,t}$. Consequently,

$$L_{x,0,t} = \frac{1}{2} \tilde{L}_{x,0,t} + \frac{1}{2} \tilde{L}_{x-1,0,t}. $$

In total,

$$\hat{m}_{x,d,t} = \begin{cases} \frac{1}{2} \tilde{D}_{x,d,t} + \frac{1}{2} \tilde{D}_{x,d+1,t} & d \geq 1, \\ \tilde{D}_{x,0,t} + \frac{1}{2} \tilde{D}_{x,1,t} & d = 0. \end{cases}$$

(2.1)

We apply (2.1) for the calculation of the empirical mortality rates from our data set and present some first impression of the data. First of all, we have 2240 mortality rates. Unfortunately, there are 4
of the $D_{xdt}$ equal to zero. We set the mortality rate to the lowest obtained value unequal to zero, because we want to consider the logarithm of the mortality rate. Figure 2.3 shows the logarithm of the empirical mortality rates. In Figure 2.3(a) the data are aggregated with respect to the duration and the age is aggregated into four intervals. We see that in the long term the mortality rate is slowly decreasing in time and that the morality improvement is better for the younger age groups. Interestingly, the curves are crossing, which means that mortality improvements are not consistent with respect to different age groups. We also find crossings in Figure 2.3(b), which shows the duration dependence of mortality and disregards the age-dependence. Mortality rates are strictly decreasing for increasing duration. This well-known empirical observation is usually explained as follows: When a policyholder gets disabled, he suffers from an injury or affliction that rapidly increases his mortality. The longer the duration of the injury, the better he learns to cope with the injury or affliction and the mortality rate moves back to the mortality rate of a non-disabled policyholder. Figures 2.3(c) and 2.3(d) show the time-dependence of the mortality in the different duration groups. The curves are rather flat, decreasing very slowly, that is, the duration dependence of mortality rates is not much changing with respect to time. Note that all four diagrams in Figure 2.3 use the same scale. Therefore we can already conclude that the mortality rate is much more sensitive to duration than to calendar time and age.

3. The Lee-Carter model and single effects

In this section we apply the Lee-Carter (LC) model to our data and analyze single effects, i.e. we fit models with only one parameter. So we have the Model MA $\log m_{xdt} = \alpha_x$, the Model MK $\log m_{xdt} = \kappa_t$, and the Model MG $\log m_{xdt} = \gamma_d$. The different models are summarized in Table 1. For all four models, we give the coefficient of determination ($R^2$), the Akaike information criterion (AIC), the Bayesian information criterion (BIC), and the Hannan–Quinn information criterion (HQC), expressing the goodness of fit. The last column shows the number of parameters of each model. The coefficient of determination is calculated as one minus the fraction of empirical variance of the residua divided by the empirical variance of the model. The three information criteria are usually defined with the Likelihood estimator. Since we estimated our models with a least squares approach, we use the representation via the empirical variance of the residua. The representations are equivalent when the residua are normally distributed (cf. Burnham and Anderson, 2002). However, this is not the case for
most of our models so that we use the representation with the variance of the residua as a definition, i.e.

\[ AIC = \log(\hat{\sigma}^2) + \frac{2p}{n}, \]

\[ BIC = \log(\hat{\sigma}^2) + \frac{p}{n}\log(n), \]

\[ HQC = \log(\hat{\sigma}^2) + \frac{2p}{n}\log(\log(n)), \]

where \( \hat{\sigma}^2 \) is the empirical variance of the residua, \( n \) the number of observations, and \( p \) the number of parameters. For our data we have \( n = 2240 \) so that we always have \( AIC \leq HQC \leq BIC \) for each model.

In the following the calibration method of the models is explained. We fit all models to the empirical mortality rates \( \hat{m}_{xdt} \) from Section 2. We generally use a least-squares estimation (cf. Pitacco et al., 2009, Section 5.2) that aims to minimize

\[ f(\alpha_x, \kappa_t, \ldots) := (\log(\hat{m}_{xdt}) - \log(m_{xdt}))^2 \rightarrow \min. \quad (3.1) \]

Our models use different sets of parameters, so the function \( f \) varies for different models. We search for minima by differentiating \( f \) and setting it equal to zero. Zero points are numerically calculated by
the multivariate Newton-Raphson method. For the simple models shown in this section this method converges very fast and always finds the minimum. This is not always the case for the more complex models in Section 4 for which we choose random starting points (normally distributed with mean 0 and standard deviation 10) and run this simulation several times until we do not find a new minimum anymore for several runs. We apply the Newton-Raphson method until the Newton-residuum is in the region of $10^{-10}$. To verify the results we check them by minimizing (3.1) by a standard line-search algorithm implemented in MATLAB. Both methods always gave that same results, so that we can rely on them.

Figure 3.1 gives the calibrated parameters of the models MA, MK, MG, and LC. The estimation is consistent with the summary statistics in Figure 2.3: mortality is largely increasing with age, we see a long-term longevity trend, mortality is decreasing with duration, and the longevity trend is stronger for younger ages. This observations is also conform with the literature. Surprisingly, the parameter $\alpha_x$ shows a decreasing mortality between age 55 and 60. As we see in Figure 2.3(a), this is not an effect of the model assumptions but is really observed in our data. A similar effect is observed in Kolster et al. (1998). The duration dependence is very strong for short durations but is nearly flat after 6 years. The duration dependence model MG clearly has the best fit and explains more then 70% of the empirical variance. This shows that taking respect of duration dependence is inevitable when modeling the mortality of disabled people. The calibrated parameters of the Lee-Carter model are consistent with the calibrated parameters of the models MA and MK. However, the general fit is very poor, which emphasizes the need to extend the Lee-Carter model by duration parameters.

### 4. Extensions of the Lee-Carter model

There are various ways to add duration dependence to the Lee-Carter model. Our first try is an additive component analogously to the basic age dependence $\alpha_x$ in the LC model.

**Definition 4.1 (Model M1).** We define the Model M1 as

$$\log m_{xdt} = \alpha_x + \gamma_d + \beta_x \kappa_t$$

with $x \in \{ x_1, \ldots, x_k \}, d \in \{ d_1, \ldots, d_l \}, t \in \{ t_1, \ldots, t_m \}$.

Without further restrictions, Model M1 is not uniquely identifiable. The following proposition shows how the make the model unique.

**Proposition 4.2.** If there is a $t$ with $\kappa_t \neq 0$, then the conditions

$$\sum_d \gamma_d = 0, \quad \sum_x \beta_x = 1 \quad \text{and} \quad \sum_t \kappa_t = 0 \quad (4.1)$$

make the model $\log m_{xdt} = \alpha_x + \gamma_d + \beta_x \kappa_t$ unique. For each model $\log m_{xdt} = \tilde{\alpha}_x + \tilde{\gamma}_d + \tilde{\beta}_x \tilde{\kappa}_t$ with $\sum_x \tilde{\beta}_x \neq 0$, there is an equivalent model $\log m_{xdt} = \alpha_x + \gamma_d + \beta_x \kappa_t$ that fulfills (4.1).

The proofs of this proposition and all following propositions are given in Section 5. Figure 4.1 shows the calibration of Model M1 to our data set, considering the conditions (4.1). In Table 2, see the Appendix, we give the corresponding goodness-of-fit parameters $R^2$, AIC, BIC and HQC. We see that $\alpha_x$, $\beta_x$ and $\kappa_t$ are almost identical to the same parameters in the Lee-Carter model, but in

<table>
<thead>
<tr>
<th>Model</th>
<th>$R^2$</th>
<th>AIC</th>
<th>BIC</th>
<th>HQC</th>
<th>par</th>
</tr>
</thead>
<tbody>
<tr>
<td>MA</td>
<td>0.6001</td>
<td>-0.5935</td>
<td>-0.5424</td>
<td>-0.5748</td>
<td>20</td>
</tr>
<tr>
<td>MK</td>
<td>0.6320</td>
<td>-0.5675</td>
<td>-0.5267</td>
<td>-0.5526</td>
<td>16</td>
</tr>
<tr>
<td>MG</td>
<td>0.7072</td>
<td>-1.7713</td>
<td>-1.7535</td>
<td>-1.7648</td>
<td>7</td>
</tr>
<tr>
<td>LC</td>
<td>0.1110</td>
<td>-0.6169</td>
<td>-0.4741</td>
<td>-0.5648</td>
<td>56</td>
</tr>
</tbody>
</table>

Table 1: The Lee-Carter model and single effects
contrast to the latter model the fit dramatically improved by introducing the additional parameter \( \gamma_d \). The estimation of \( \gamma_d \) is consistent with Model MG. A nice property of this model is that we can interpret \( \alpha_x + \beta_x \kappa_t \) as the mortality of the general population (largely healthy people) plus an additive disability effect \( \gamma_d \). The residua, illustrated in Figure A.1(a), still show a little bit of structure between the random noise, but seem to be acceptable. The same conclusion applies for the residua of all following models, see Figure A.1 (b) to (e). The low \( R^2 \) of Model MA motivates to simplify our model by skipping \( \alpha_x \), which we do in Model M1.1, see the appendix. The \( R^2 \) slightly reduces to \( R^2 = 0.8124 \), but with now 43 instead of 63 parameters, the BIC is better than before.

Instead of extending the Lee-Carter model by additive duration effects as before, we now try multiplicative duration parameters. We start with a factor on \( \alpha_x \).

**Definition 4.3** (Model M2). We define the Model M2 as

\[
\log m_{xdt} = \alpha_x \gamma_d + \beta_x \kappa_t
\]

with \( x \in \{x_1, ..., x_k\}, d \in \{d_1, ..., d_l\}, t \in \{t_1, ..., t_m\} \).

The following proposition shows how to make the model uniquely identifiable.
Proposition 4.4. If there are $d_1$, $d_2$ with $\gamma_{d_1} \neq \gamma_{d_2}$, $t_1$, $t_2$ with $\kappa_{t_1} \neq \kappa_{t_2}$, and $x_1$, $x_2$ with $\beta_{x_1} \neq 0$, $\beta_{x_2} \neq 0$, $\frac{\alpha_{x_1}}{\beta_{x_1}} \neq \frac{\alpha_{x_2}}{\beta_{x_2}}$, then the conditions

$$\sum_d \gamma_d = 1 \quad \text{and} \quad \sum_x \beta_x = 1$$

(4.2)

make the model $\log m_{xdt} = \alpha_x + \gamma_d + \beta_x \kappa_t$ unique. For each model $\log m_{xdt} = \tilde{\alpha}_x \tilde{\gamma}_d + \tilde{\beta}_x \tilde{\kappa}_t$ with $\sum_d \tilde{\gamma}_d \neq 0$ and $\sum_x \tilde{\beta}_x \neq 0$, there is an equivalent model that fulfills (4.2).

Figure 4.2 shows the calibrated parameters under the assumptions (4.2). The parameter $\alpha_x$ is very different now from the $\alpha_x$ in the Lee-Carter model. Instead of describing the basic mortality profile, it rather seems to balance the $\beta_x$, which has the same profile but flipped around. The $\gamma_d$ is also flipped around, but since it is multiplied with a negative $\alpha_x$, the fundamental duration effect is still the same. Looking at $\alpha_x$ and $\gamma_d$ together leads to the conclusion that the influence of the duration effect increases with increasing age. This seems to be confirm with Figure 2.3(b). The time trend $\kappa_t$ is similar to before. On the one hand, the model has a better fit than M1, even though it has the same number of parameters. The BIC is even the best of all models studied in this paper, see Table 2 in the appendix. On the other hand, the fact that $\alpha_x$ is nearly a mirror of $\beta_x$ raises doubts about the parametrization of the model. If $\alpha_x$ and $\beta_x$ are so similar, can we simplify the model by replacing $\beta_x$ by $\alpha_x$? We tried that in Model M2.1, see Figure A.2 in the appendix, but the result is not very convincing. The new $\alpha_x$ is very different from the old one.

In a next step we put a duration effect not on the first but on the second addend in the Lee-Carter model.

Definition 4.5 (Model M3). We define the Model M3 as

$$\log m_{xdt} = \alpha_x + \beta_x \delta_d \kappa_t$$

with $x \in \{x_1, \ldots, x_k\}, d \in \{d_1, \ldots, d_l\}, t \in \{t_1, \ldots, t_m\}$. 

Figure 4.1: Parameter Model M1, $\log m_{xdt} = \alpha_x + \gamma_d + \beta_x \kappa_t$
The following proposition gives some conditions for making the model unique.

**Proposition 4.6.** If there are $t_1, t_2$ with $\kappa_{t_1} \neq \kappa_{t_2}$ and $d_1, d_2$ with $\delta_{d_1} \neq 0, \delta_{d_2} \neq 0$, then the condition

$$\sum_x \beta_x = 1 \quad \text{and} \quad \sum_d \delta_d = 1$$

(4.3)

makes the model $\log m_{xdt} = \alpha_x + \beta_x \delta_d \kappa_t$ unique. For each model $\log m_{xdt} = \tilde{\alpha}_x + \tilde{\beta}_x \tilde{\delta}_d \tilde{\kappa}_t$ with $\sum_x \tilde{\beta}_x \neq 0$ and $\sum_d \tilde{\delta}_d \neq 0$, there is an equivalent model that fulfills (4.3).

On the basis of (4.3), Figure 4.3 shows the calibration of Model M3 to our data set. As in the model before, $\alpha_x$ is nearly a mirror of $\beta_x$. Again, this raises doubts about the parametrization. With now having a triple product $\beta_x \delta_d \kappa_t$, the parameters are more difficult to interpret because of the multiple interactions. The time trend $\kappa_t$ and the duration effect $\delta_d$ are similar to before. The fit (see the $R^2$) is not as good as for the Model M2, although they have the same number of parameters. In the previous section we learned that the duration effect is the most important effect. So in Model M3.1 and Model M3.2, see Table 2 in the appendix, we try what happens if we skip $\alpha_x$ and if we replace $\alpha_x$ by $\gamma_d$. Skipping $\alpha_x$ leads to clearly poorer results. Replacing $\alpha_x$ by $\gamma_d$ worsens the fit but increases the BIC. We also tried to skip $\beta_x$, see Model M3.3, which also significantly lowers the quality of the fit.

In a next step, we combine the two different duration parameters $\gamma_d$ and $\delta_d$ of Model M2 and Model M3 in a joint extended model.

**Definition 4.7 (Model M4).** We define the Model M4 as

$$\log m_{xdt} = \alpha_x \gamma_d + \beta_x \delta_d \kappa_t$$

with $x \in \{x_1, ..., x_k\}, d \in \{d_1, ..., d_l\}, t \in \{t_1, ..., t_m\}$.

Unique identifiability of the model can be achieved with the help of the following proposition.
Proposition 4.8. If there are $t_1$, $t_2$ with $\kappa_{t_1} \neq \kappa_{t_2}$, a parameter $d$ with $\gamma_d \neq \delta_d$, and $x_1$, $x_2$ with $\beta_{x_1} \neq 0, \beta_{x_2} \neq 0$, $\alpha_{x_1}/\beta_{x_1} \neq \alpha_{x_2}/\beta_{x_2}$, then the conditions
\[
\sum_d \gamma_d = 1, \sum_x \beta_x = 1, \quad \text{and} \quad \sum_d \delta_d = 1
\] (4.4)
make the model $\log m_{xdt} = \alpha_x + \beta_x \delta_d \kappa_t$ unique. For each model $\log m_{xdt} = \tilde{\alpha}_x \tilde{\gamma}_d + \tilde{\beta}_x \tilde{\delta}_d \tilde{\kappa}_t$ with $\sum_d \tilde{\gamma}_d \neq 0, \sum_x \tilde{\beta}_x \neq 0$ and $\sum_d \tilde{\delta}_x \neq 0$, there is an equivalent model that fulfills (4.4).

Figure 4.4 shows the calibrated parameters, using conditions (4.4).

The calibrated parameters of the present model are very similar to the calibrated parameters of Model M2, which is a submodel of the present model. Interestingly, the typical duration pattern that has been observed in the parameter $\delta_d$ before now moved to the parameter $\gamma_d$. The present model is also an extension of Model M3, but we find less similarities in the calibrated parameters. Figure 4.4(d) shows a new effect that has not been observed so far. However, as $\delta_d$ is one factor of a triple product, the interpretation is not obvious. It is not surprising that the Model M4 has a better fit than the Models M2 and M3, since we added extra parameters. But compared to the increase in the number of parameters, the increase in the $R^2$ is rather disappointing.

In a last and fifth step we allow for separate time trends with respect to age and duration. Including all possible pairs of the three dimensions age, time and duration yields the following model.

Definition 4.9 (Model M5). We define the Model M5 as
\[
\log m_{xdt} = \alpha_x \gamma_d + \beta_x \kappa_t^1 + \delta_d \kappa_t^2
\]
with $x \in \{x_1, ..., x_k\}, d \in \{d_1, ..., d_l\}, t \in \{t_1, ..., t_m\}$.

Making the model uniquely identifiable is more tricky here. The following proposition offers a solution.
Proposition 4.10. If there are parameters \( x, d \) with \( \beta_x \neq \frac{1}{x} \), \( \delta_d \neq \frac{1}{d} \), \( d_1, d_2 \) with \( \delta_d_1 \neq \delta_d_2 \), \( x_1, x_2 \) with \( \beta_x_1 \neq \beta_x_2 \), \( l > 1 \), and if the pairs \((\alpha_x)_x\) and \((\beta_x)_x\), \((\gamma_d)_d\) and \((\delta_d)_d\), \((\kappa)_t\) and \((\kappa^2)_t\) are assumed to be linearly independent, then the conditions

\[
\sum_d \gamma_d = 1, \quad \sum_x \beta_x = 1, \quad \sum_d \delta_d = 1
\]  

make the model \( \log m_{xdt} = \alpha_x \gamma_d + \beta_x \delta_d \kappa^1_t \) unique. For each model \( \log m_{xdt} = \tilde{\alpha}_x \tilde{\gamma}_d + \tilde{\beta}_x \tilde{\kappa}^1_t + \tilde{\delta}_d \tilde{\kappa}^2_t \) with \( \sum_d \tilde{\gamma}_d \neq 0 \), \( \sum_x \tilde{\beta}_x \neq 0 \) and \( \sum_d \tilde{\delta}_d \neq 0 \) there exists an equivalent model \( \log m_{xdt} = \alpha_x \gamma_d + \beta_x \kappa^1_t + \delta_d \kappa^2_t \) that satisfies the conditions (4.5).

Figure 4.5 shows the calibrated parameters on the basis of the assumptions (4.5). The duration effect \( \gamma_d \) has an inverse monotony compared to Model M4. The typical duration effect (decreasing mortality for increasing duration) is rather found in \( \delta_d \). (Note that \( \alpha_x \) and \( \kappa^2_t \) are negative.) It is also confusing that both age parameters \( \alpha_x \) and \( \beta_x \) are for the most part monotonically decreasing. Apart from the clear trend \( \kappa^1 \) with respect to age, the fundamental effects that we observed in the data in the previous section are somehow spread between the different parameters. This makes the interpretation
of the parameters less intuitive. Nonetheless this model offers new insight. First, $\delta_d$ shows that the variation of the duration effect with respect to time is nearly zero for short durations and is much more present for longer durations. Second, $\kappa_t^2$ reveals that there is no clear long-term time trend in the duration effects. Being the model with the most parameters, Model M5 has the best fit. Compared to the number of parameters, the advantage over the other models is rather disappointing. With no clear long-term time trend observed in the duration effect, there are doubts if adding a second time trend parameter is beneficial.

5. Proofs of the propositions

In this section we give the proofs of the propositions from the last section.

Proof of Proposition 4.2. To show the uniqueness we assume that $\alpha_x$, $\gamma_d$, $\beta_x$, $\kappa_t$ as well as $\tilde{\alpha}_x$, $\tilde{\gamma}_d$, $\tilde{\beta}_x$, $\tilde{\kappa}_t$ fulfill (4.1) and that

$$\alpha_x + \gamma_d + \beta_x \kappa_t = \tilde{\alpha}_x + \tilde{\gamma}_d + \tilde{\beta}_x \tilde{\kappa}_t.$$  

(5.1)
Summing (5.1) over \(d\) and \(t\) yields \(\alpha_x = \tilde{\alpha}_x\). Using this equality, summing over \(t\) implies \(\gamma_d = \tilde{\gamma}_d\), and summing over \(x\) shows that \(\kappa_t = \tilde{\kappa}_t\). Since there is a \(t\) with \(\kappa_t \neq 0\), we also get \(\beta_x = \tilde{\beta}_x\).

In order to prove the second statement of the proposition, we define \(\alpha_x = \tilde{\alpha}_x + \tilde{\beta}_x \frac{1}{m} \sum_t \tilde{\kappa}_t + \frac{1}{l} \sum_d \tilde{\gamma}_d\), \(\gamma_d = \tilde{\gamma}_d - \frac{1}{l} \sum_d \tilde{\gamma}_d\), \(\beta_x = \frac{\tilde{\beta}_x}{\sum \tilde{\beta}_x}\), and \(\kappa_t = (\tilde{\kappa}_t - \frac{1}{m} \sum_t \tilde{\kappa}_t) \sum \tilde{\beta}_x\). Straightforward calculations show that \(\alpha_x + \gamma_d + \beta_x \kappa_t = \tilde{\alpha}_x + \tilde{\gamma}_d + \tilde{\beta}_x \kappa_t\) and that \(\alpha_x, \gamma_d, \beta_x, \kappa_t\) fulfill conditions (4.1).

**Proof of Proposition 4.4.** Let \(\alpha_x, \gamma_d, \beta_x, \kappa_t\) and \(\tilde{\alpha}_x, \tilde{\gamma}_d, \tilde{\beta}_x, \tilde{\kappa}_t\) be equivalent models, i.e.

\[
\alpha_x \gamma_d + \beta_x \kappa_t = \tilde{\alpha}_x \tilde{\gamma}_d + \tilde{\beta}_x \tilde{\kappa}_t,
\]

that both satisfy the assumptions of the first statement in the proposition. Summing in equation (5.2) over \(x\) and \(d\), we get that \(\kappa_t = \tilde{\kappa}_t + c\) for constant \(c\) equal to

\[
c = \frac{1}{l} \sum_x (\tilde{\alpha}_x - \alpha_x) .
\]

Plugging that into (5.2), we obtain

\[
\alpha_x \gamma_d - \tilde{\alpha}_x \tilde{\gamma}_d + \tilde{\kappa}_t (\beta_x - \tilde{\beta}_x) + c \beta_x = 0.
\]

If there were an \(x\) with \(\beta_x \neq \tilde{\beta}_x\), then \(\kappa_t\) had to be constant in \(t\), which is a contradiction to the initial assumptions. Consequently, \(\beta_x = \tilde{\beta}_x\) for all \(x\).

In a next step, we sum (5.3) over \(d\) and get \(\alpha_x = \tilde{\alpha}_x - l \beta_x\). Plugging this result into (5.2), we get

\[
\tilde{\alpha}_x (\gamma_d - \tilde{\gamma}_d) + c \tilde{\beta}_x (1 - l \gamma_d) = 0.
\]

By assumption there is a \(d\) with \(1 - l \gamma_d \neq 0\), and so we obtain for \(x = x_1, x_2\) that

\[
c = \frac{\tilde{\alpha}_x (\gamma_d - \tilde{\gamma}_d)}{\tilde{\beta}_x (1 - l \gamma_d)} .
\]

For \(x = x_1, x_2\) the fraction \(\tilde{\alpha}_x / \tilde{\beta}_x\) differs, so for the specific \(d\) we need to have \(\gamma_d = \tilde{\gamma}_d\), which implies \(c = 0\). Thus we have \(\kappa_t = \tilde{\kappa}_t\) for all \(t\). With the side condition \(\sum \gamma_d = 1\) we also get \(\alpha_x = \tilde{\alpha}_x\) and \(\gamma_d = \tilde{\gamma}_d\).

The second statement of the proposition can be verified by defining \(\alpha_x = \tilde{\alpha}_x \sum \tilde{\gamma}_d, \gamma_d = \frac{\tilde{\gamma}_d}{\sum \tilde{\gamma}_d} \beta_x = \frac{\tilde{\beta}_x}{\sum \tilde{\beta}_x}, \) and \(\kappa_t = \tilde{\kappa}_t \sum \tilde{\beta}_x\). This way we obtain an equivalent model, i.e. \(\alpha_x \gamma_d + \beta_x \kappa_t = \tilde{\alpha}_x \tilde{\gamma}_d + \tilde{\beta}_x \tilde{\kappa}_t\) that satisfies conditions (4.2).

**Proof of Proposition 4.6.** Suppose that \(\alpha_x, \beta_x, \delta_d, \kappa_t\) and \(\tilde{\alpha}_x, \tilde{\beta}_x, \tilde{\delta}_d, \tilde{\kappa}_t\) are equivalent models, i.e.

\[
\alpha_x + \beta_x \delta_d \kappa_t = \tilde{\alpha}_x + \tilde{\beta}_x \tilde{\delta}_d \tilde{\kappa}_t,
\]

that satisfy the assumptions of the first statement of the proposition. Summing in equation (5.4) over \(x\) and \(d\), we get \(\kappa_t = \tilde{\kappa}_t + c\) for constant \(c\) equal to \(\frac{1}{l} \sum_x (\tilde{\alpha}_x - \alpha_x)\). Using this two equations and summing (5.4) over \(x\), we obtain

\[
-lc + \tilde{\kappa}_t (\delta_d - \tilde{\delta}_d) + \delta_d c = 0.
\]

If there were a \(d\) with \(\delta_d \neq \tilde{\delta}_d\), then we could solve (5.5) with respect to \(\tilde{\kappa}_t\) and would obtain an expression that does not depend on \(t\), which is a contradiction to the assumption that there are at least two different \(\tilde{\kappa}_t\). Moreover, we must have \(\delta_d = \tilde{\delta}_d\). So (5.5) is simplified to \((\delta_d - l)c = 0\). Since there are different \(\delta_d\), the constant \(c\) must be zero, which implies \(\kappa_t = \tilde{\kappa}_t\). Plugging these results into (5.4), we get

\[
\alpha_x - \tilde{\alpha}_x + \delta_d \kappa_t (\beta_x - \tilde{\beta}_x) = 0.
\]
By assumption there is a $d$ with $\delta_d \neq 0$. If there were an $x$ with $\beta_x \neq \beta_x$, we could solve the equation for $\kappa_t$ and would obtain an expression that does not depend on $t$, which is a contradiction to the initial assumptions. So we must have $\beta_x = \beta_x$ and also $\alpha_x = \tilde{\alpha}_x$.

In order to prove the second statement of the proposition, we define $\alpha_x = \tilde{\alpha}_x$, $\beta_x = \beta_x$, $\delta_d = \delta_d$, and $\kappa_t = \kappa_t(\sum \beta_x)(\sum \delta_d)$. This way we can always obtain an equivalent model, i.e. $\alpha_x + \beta_x \delta_d = \alpha_x + \beta_x \tilde{\delta}_d$, that satisfies the conditions (4.3).

**Proof of Proposition 4.8.** Let $\alpha_x, \gamma_d, \beta_x, \delta_d, \kappa_t$ and $\tilde{\alpha}_x, \tilde{\gamma}_d, \tilde{\beta}_x, \tilde{\delta}_d, \tilde{\kappa}_t$ be equivalent models, i.e.

$$\alpha_x \gamma_d + \beta_x \delta_d \kappa_t = \tilde{\alpha}_x \tilde{\gamma}_d + \tilde{\beta}_x \tilde{\delta}_d \tilde{\kappa}_t,$$

(5.6)

that fulfill the assumptions of the first statement of the proposition. Summing in equation (5.6) over $x$ and $d$, we necessarily get $\kappa_t = \tilde{\kappa}_t$ with $c = \sum_x (\tilde{\alpha}_x - \alpha_x)$. Plugging this equation into (5.6) and summing over $d$ leads to

$$(\alpha_x - \tilde{\alpha}_x) + \tilde{\kappa}_t (\beta_x - \tilde{\beta}_x) + c \beta_x = 0.$$

(5.7)

If we there were an $x$ with $\beta_x \neq \tilde{\beta}_x$, then $\tilde{\kappa}_t$ had to be constant, which is a contradiction to the assumptions of the proposition. Thus, $\beta_x = \beta_x$. Plugging the simplified condition (5.7) and our other results into (5.6), we get

$$\tilde{\alpha}_x (\gamma_d - \tilde{\gamma}_d) + \beta_x \tilde{\kappa}_t (\delta_d - \tilde{\delta}_d) + c \beta_x (\delta_d - \gamma_d) = 0.$$

(5.8)

By assumption there is an $x$ such that $\beta_x \neq 0$, and so we must have $\delta_d = \tilde{\delta}_d$, since otherwise $\kappa_t$ had to be constant. Under the assumptions of the proposition there is a $d$ and there are $x_1, x_2$ for which (5.8) implies that

$$c = \frac{\tilde{\alpha}_x (\gamma_1 - \tilde{\gamma}_d)}{\beta_x (\delta_d - \delta_1)}.$$

As $\tilde{\alpha}_x / \beta_x \neq \alpha_x / \beta_x$, this equation can only be true if for the specific $d$ we have $\gamma_d = \tilde{\gamma}_d$. Thus, we necessarily have $c = 0$. From this we can conclude that $\kappa_t = \tilde{\kappa}_t$, $\alpha_x = \tilde{\alpha}_x$, and $\gamma_d = \tilde{\gamma}_d$.

In order to prove the second statement of the theorem, we define the model $\alpha_x = \tilde{\alpha}_x \gamma_d, \gamma_d = \sum \beta_x, \delta_d = \sum \delta_d$, and $\kappa_t = \kappa_t(\sum \beta_x)(\sum \delta_d)$. This way we can always obtain an equivalent model, i.e. $\alpha_x \gamma_d + \beta_x \delta_d \kappa_t = \tilde{\alpha}_x \gamma_d + \beta_x \delta_d \tilde{\kappa}_t$, that satisfies conditions (4.4).

**Proof of Proposition 4.10.** Let $\alpha_x, \gamma_d, \beta_x, \kappa_t^1, \delta_d, \kappa_t^2$ and $\tilde{\alpha}_x, \tilde{\gamma}_d, \tilde{\beta}_x, \tilde{\kappa}_t^1, \tilde{\delta}_d, \tilde{\kappa}_t^2$ be equivalent models, i.e.

$$\alpha_x \gamma_d + \beta_x \kappa_t^1 + \delta_d \kappa_t^2 = \tilde{\alpha}_x \tilde{\gamma}_d + \tilde{\beta}_x \tilde{\kappa}_t^1 + \tilde{\delta}_d \tilde{\kappa}_t^2,$$

(5.9)

that both satisfy the assumptions of the first statement in the proposition. Summing the equation (5.9) over both $x, d$, over $d$ only, and over $x$ only leads to an equation system $A(\kappa_t^1, \kappa_t^2, \tilde{\kappa}_t^1, \tilde{\kappa}_t^2)^T = b$ with matrix $A$ equal to

\[
\begin{pmatrix}
 l & -l & k & -k \\
 l\beta_x & -l\beta_x & 1 & -1 \\
 1 & -1 & k\delta_d & -k\delta_d
\end{pmatrix}.
\]

If this matrix had a rank of three, then we obtain linear dependence between $(\kappa_t^1)_t$ and $(\kappa_t^2)_t$ (or $(\tilde{\kappa}_t^1)_t$ and $(\tilde{\kappa}_t^2)_t$). This is a contradiction to our initial assumptions, so the matrix $A$ may have at most a rank of two. By linear operations the matrix $A$ can be transformed to the following matrices with same rank as $A$,

\[
\begin{pmatrix}
 l & -l & k & -k \\
 0 & l(\beta_x - \beta_x) & 1 - k\beta_x & k\beta_x - 1 \\
 0 & 0 & k(\delta_d - \frac{1}{2}) & k(\frac{1}{2} - \delta_d)
\end{pmatrix},
\begin{pmatrix}
 l & -l & k & -k \\
 l(\beta_x - \frac{1}{2}) & l(\frac{1}{2} - \beta_x) & 0 & 0 \\
 1 - l\beta_x & l\beta_x - 1 & k(\delta_d - \tilde{\delta}_d) & 0
\end{pmatrix}.
\]
In order that the rank is always smaller than three, we must have \( \beta_x = \tilde{\beta}_x \) for all \( x \) and \( \delta_d = \tilde{\delta}_d \) for all \( d \).

We use these results and sum (5.9) over both \( t,d \), all \( \kappa \) Consequently, we conclude that the ARIMA(0,1,0) model is appropriate here. So we define

\[
\text{larger than 0}
\]

\( \kappa \) corresponds to duration effects and not to age effects. Because of the great similarities between the \( C \) of equations \( \kappa \)

In this section we forecast the mortality rate into the future. In Section 4 we have seen that almost all \( \kappa_t \) show a clear linear trend, except for \( \kappa^2_t \) in Model M5, which is the only time parameter that clearly corresponds to duration effects and not to age effects. Because of the great similarities between the \( \kappa_t \) in the different models, we exemplarily just show the forecasting for Model M2 here, which is the model with the best BIC.

We follow the concept in Section 5.7 of Pitacco et al. (2009) and model the time parameter \( \kappa_t \) by an ARIMA(0,1,0) model. Figure 4.2(d) shows that \( \kappa_t \) is not stationary and the Augmented Dickey-Fuller test applied to the first differences of \( \kappa_t \) has a \( p \)-value of less than 0.001, indicating that the differences of \( \kappa_t \) are stationary. The Shapiro-Wilk test has a \( p \)-value of 0.6974 and the Jarque-Bera test a \( p \)-value larger than 0.5, so that the null hypothesis that residuals are normally distributed is not rejected. Consequently, we conclude that the ARIMA(0,1,0) model is appropriate here. So we define \( \kappa_t \) by

\[
\kappa_t = \kappa_{t-1} + \mu_t + \epsilon_t , \quad \text{with} \quad \epsilon_t \sim \mathcal{N}(0, \sigma^2).
\]

We use standard maximum likelihood techniques to calibrate the ARIMA(0,1,0) model, using the functions \texttt{arima} and \texttt{estimate} in MATLAB. The estimated parameters are \( \mu = -0.5131 \) and \( \sigma = 1.1458 \).

We illustrate the forecast by showing the future mortality of an individual. We consider a \( x_0 = 40 \) year old individual in the year \( t = 2009 \) with a duration in the state disabled of \( d = 0 \) and forecast the mortality rate for the next 16 years. Since our empirical analysis showed that the duration effect vanishes with increasing duration and is nearly flat after 6 years, see Figure 4.2(b), we assume here that \( m_{x,d,t} = m_{x,6,t} \) for all \( d \geq 6 \). Figure 6.1 shows the logarithm of the future mortality rates. Since they are random, we give the median, the 75%-quantile, and the 95%-quantile. With increasing time, 

6. Forecasting of the mortality rates

In this section we forecast the mortality rate into the future. In Section 4 we have seen that almost all \( \kappa_t \) show a clear linear trend, except for \( \kappa^2_t \) in Model M5, which is the only time parameter that clearly corresponds to duration effects and not to age effects. Because of the great similarities between the \( \kappa_t \) in the different models, we exemplarily just show the forecasting for Model M2 here, which is the model with the best BIC.

We follow the concept in Section 5.7 of Pitacco et al. (2009) and model the time parameter \( \kappa_t \) by an ARIMA(0,1,0) model. Figure 4.2(d) shows that \( \kappa_t \) is not stationary and the Augmented Dickey-Fuller test applied to the first differences of \( \kappa_t \) has a \( p \)-value of less than 0.001, indicating that the differences of \( \kappa_t \) are stationary. The Shapiro-Wilk test has a \( p \)-value of 0.6974 and the Jarque-Bera test a \( p \)-value larger than 0.5, so that the null hypothesis that residuals are normally distributed is not rejected. Consequently, we conclude that the ARIMA(0,1,0) model is appropriate here. So we define \( \kappa_t \) by

\[
\kappa_t = \kappa_{t-1} + \mu_t + \epsilon_t , \quad \text{with} \quad \epsilon_t \sim \mathcal{N}(0, \sigma^2).
\]

We use standard maximum likelihood techniques to calibrate the ARIMA(0,1,0) model, using the functions \texttt{arima} and \texttt{estimate} in MATLAB. The estimated parameters are \( \mu = -0.5131 \) and \( \sigma = 1.1458 \).

We illustrate the forecast by showing the future mortality of an individual. We consider a \( x_0 = 40 \) year old individual in the year \( t = 2009 \) with a duration in the state disabled of \( d = 0 \) and forecast the mortality rate for the next 16 years. Since our empirical analysis showed that the duration effect vanishes with increasing duration and is nearly flat after 6 years, see Figure 4.2(b), we assume here that \( m_{x,d,t} = m_{x,6,t} \) for all \( d \geq 6 \). Figure 6.1 shows the logarithm of the future mortality rates. Since they are random, we give the median, the 75%-quantile, and the 95%-quantile. With increasing time,
also the age and the duration of the policyholder are increasing. In the first 3 years the mortality rate is considerably decreasing, which is due to the increasing duration. After that time the mortality rates are increasing, which comes from the age effect. It is in particular interesting that the rates are already increasing several years before 2016, where the duration is 7 and our extrapolation from above starts. Already from 2012 on the aging effect is stronger than the duration effect. The ageing effect is interfered by a longevity trend, but this is difficult to see here. As we would expect, the confidence intervals are widening in the first years, but in the last couple of years they are narrowing, which seems a little bit surprising. The reason is that although the confidence intervals for \( \kappa_t \) are throughout widening, the parameter \( \beta_x \) is moving towards zero around age 55, see Figure 4.2(c). As mortality improvements rather happen at younger ages in our data, there is less forecasting uncertainty about mortality improvements at higher ages. It also should be kept in mind that in our model the mortality improvement does not interact with the duration.

7. Conclusions

We proposed, calibrated and discussed several extensions of the Lee-Carter with the aim to take respect of duration effects in the mortality of disabled. By adding duration parameters, the model fit increases dramatically. All proposed models with duration parameters explain between 80% and 84% of the empirical variance. Taking into consideration that our data is grouped into 2240 classes and that many classes are sparse, this is a remarkable value. The residuals show more than just random noise but still seem to be acceptable.

With the differences in the fitting qualities being rather small for all our proposed models, we cannot offer a clear recommendation for one specific model. The models with triple factors have some appeal when adding a third dimension (the duration) to the two-dimensional Lee-Carter model. However, the interpretation of the parameters is not obvious, since the three factor products include several interactions. In contrast, Model M1, which introduces duration effects by a simple, additive parameter, has a very intuitive interpretation. The Model M2 has the best BIC.
While we observe significant mortality improvements with respect to age, no clear trend can be observed with respect to duration since disablement. However, more evidence is needed to confirm this observation. Because our data set contains only few individuals with long durations since disablement, we limit our data analysis to a maximal duration of six years. More sophisticated statistical techniques are needed for analyzing also longer durations.

A. Appendix

Figure A.1 shows the residua of the different models. On the x-axis are the years and on the y-axis the age. The 7 different durations are represented by the 7 figures for each model.

Figure A.1: Residua of the Models M1-M5

Table 2 summarizes the different models with the most important properties. Figure A.2 shows the parameter $\alpha_x$ of Model M2.1.
Table 2: Overview of the different models

<table>
<thead>
<tr>
<th>Model</th>
<th>( \log m_{xdt} = \alpha_x )</th>
<th>( R^2 )</th>
<th>AIC</th>
<th>BIC</th>
<th>HQC</th>
<th>no of par</th>
</tr>
</thead>
<tbody>
<tr>
<td>MA</td>
<td>( \log m_{xdt} = \alpha_x )</td>
<td>0.0601</td>
<td>-0.5935</td>
<td>-0.5424</td>
<td>-0.5748</td>
<td>20</td>
</tr>
<tr>
<td>MK</td>
<td>( \log m_{xdt} = \kappa_t )</td>
<td>0.0320</td>
<td>-0.5675</td>
<td>-0.5267</td>
<td>-0.5526</td>
<td>16</td>
</tr>
<tr>
<td>MG</td>
<td>( \log m_{xdt} = \gamma_d )</td>
<td>0.7072</td>
<td>-1.7713</td>
<td>-1.7535</td>
<td>-1.7648</td>
<td>7</td>
</tr>
<tr>
<td>LC</td>
<td>( \log m_{xdt} = \alpha_x + \beta_x \kappa_t )</td>
<td>0.1110</td>
<td>-0.6169</td>
<td>-0.4741</td>
<td>-0.5648</td>
<td>56</td>
</tr>
<tr>
<td>M1</td>
<td>( \log m_{xdt} = \alpha_x + \gamma_d + \beta_x \kappa_t )</td>
<td>0.8182</td>
<td>-2.1979</td>
<td>-2.0371</td>
<td>-2.1392</td>
<td>63</td>
</tr>
<tr>
<td>M2</td>
<td>( \log m_{xdt} = \alpha_x \gamma_d + \beta_x \kappa_t )</td>
<td>0.8290</td>
<td>-2.2593</td>
<td>-2.0986</td>
<td>-2.2060</td>
<td>63</td>
</tr>
<tr>
<td>M3</td>
<td>( \log m_{xdt} = \alpha_x + \beta_x \delta_d \kappa_t )</td>
<td>0.8239</td>
<td>-2.2298</td>
<td>-2.0691</td>
<td>-2.1711</td>
<td>63</td>
</tr>
<tr>
<td>M4</td>
<td>( \log m_{xdt} = \alpha_x \gamma_v + \beta_x \delta_d \kappa_t )</td>
<td>0.8301</td>
<td>-2.2596</td>
<td>-2.0810</td>
<td>-2.1944</td>
<td>70</td>
</tr>
<tr>
<td>M5</td>
<td>( \log m_{xdt} = \alpha_x \gamma_d + \beta_x \kappa_t + \delta_d \kappa_t^2 )</td>
<td>0.8404</td>
<td>-2.3077</td>
<td>-2.0883</td>
<td>-2.2276</td>
<td>86</td>
</tr>
</tbody>
</table>

\[ \text{M1.1 } \log m_{xdt} = \gamma_v + \beta_x \kappa_t \]
\[ \text{M2.1 } \log m_{xdt} = \alpha_x (\gamma_d + \kappa_t) \]
\[ \text{M3.1 } \log m_{xdt} = \beta_x \delta_d \kappa_t \]
\[ \text{M3.2 } \log m_{xdt} = \gamma_v + \beta_x \delta_d \kappa_t \]
\[ \text{M3.3 } \log m_{xdt} = \alpha_x + \delta_d \kappa_t \]

Figure A.2: Parameter \( \alpha_x \) of Model M2.1

References


