On the Modeling of Variable Mortality Trend Processes

Johannes Schupp
Institute for Finance and Actuarial Sciences (ifa), Ulm
& Institute of Insurance, Ulm University
j.schupp@ifa-ulm.de

October 2019

Abstract

Due to the observed and persistently strong improvements in life expectancy, the estimation of future mortality improvements and trends by actuaries and demographers gains more and more social and financial importance. Often, (stochastic) forecasting is performed by means of parametric mortality models, e.g. the models of Lee and Carter (1992), Cairns et al. (2006), or Plat (2009) that reduce the puzzle of mortality to a few interpretable parameters. Often a random walk with drift is used to project the time dependent parameters within such models. However, in some cases it seems to underestimate the uncertainty in future mortality rates due to its constant and fixed drift (see Börger et al. (2014) or Lee and Miller (2001)). In many countries, historical mortality trends seem to have changed once in a while and there exists nowadays a significant number of models with changing mortality trends with applications in different populations. Therefore, mortality forecasts should take this fact under consideration.

In this paper, we analyse different approaches that account for changing mortality trends. First, we analyse and compare different trend models qualitatively and deduce criteria to identify a suitable model. Then, we analyse and discuss existing methods to calibrate a model with changing trend. However, data about historical trend changes is sparse, and thus, the parameter estimation of these trend processes is vulnerably dependent on the concrete calibration method. Starting from existing approaches, we derive a stable calibration method by using historical observations to make consistent forecasts of future mortality. Within this calibration, we implicitly derive the distribution of the future trend process. Finally, we apply the trend model in a numerical example and show how the different components of uncertainty in the parameter estimates can be quantified.

1 Introduction

Increasing life expectancies can be observed in most countries and societies have to find ways to deal with the ageing population. Estimating future mortality rates is one of the most common tasks of actuaries or demographers, and in the literature, a considerable variety of models have been developed. Most popular are parametric mortality models that reduce information about mortality and longevity to only a few parameters in a lower dimensional setting. There exist models like Lee and Carter (1992) and Booth et al. (2002) with one time varying parameter, hereafter called period effect, or models with two time dependent period effects like Cairns et al. (2006) and Renshaw and Haberman (2006). Also models of higher dimension including cohort effects, curvatures, multiple populations (see Hunt and Blake (2014)) exist. For the prediction of the future mortality trends, it is necessary to forecast these period effects into the future. These two-stage stochastic mortality models are widely used in actuarial literature, see Cairns et al. (2008) or Haberman and Renshaw (2011). Here, trend processes like random walks, Autoregressive Moving Average (ARMA) models, trend stationary models, vector error correction models (VECM) can be used. To analyse, compare and further develop different trend processes will be the aim of this paper.

As Figure 1 indicates for the period effects of English and Welsh males calibrated with the Cairns et al. (2006) model (for specification and parameter estimation see Appendix A), the historical evolutions show significant changes in the trends over time. Similar findings can be made for basically any population worldwide. Therefore, a variety of models aims to identify and model trend changes in the historical mortality trend (see van Berkum et al. (2014), Coelho and Nunes (2011), Bai and Perron (2003), Chen (2009), Li et al. (2011)). These models can be subdivided into models that only aim to estimate optimal historical changing trends and then extrapolate the most recent trend into the future. However, assuming a constant trend for the entire prediction seems inconsistent with the historical observation of several significant trend changes. On the other hand, models like those of Hunt and Blake (2015), Sweeting (2011), Milidonis et al. (2011), Lemoine (2014), Hainaut (2012), and Börger and Schupp (2018) also include stochastic future trend changes in their respective predictions.

These models take the historically observed trend changes to estimate a distribution for future trend changes. However, data on historical trend changes is sparse, and thus, the estimation of the simulation model contains a considerable amount of uncertainty. If the magnitude of historical trend changes is rather inhomogeneous, the simulation of future trend changes will no longer be consistent with the observed historical trend changes as those will be hardly reproduced by

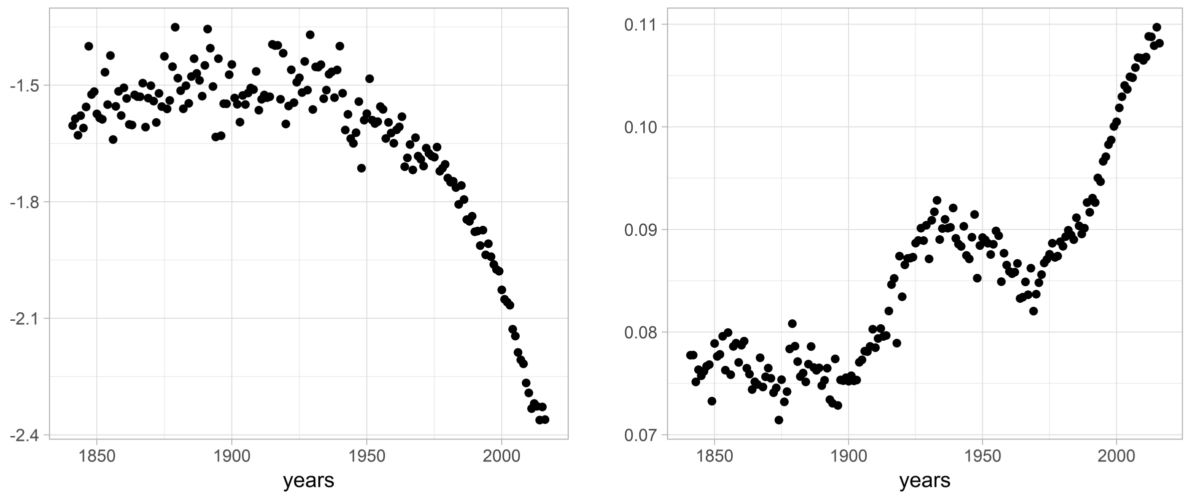

Figure 1: Period effects in the CBD model for English and Welsh males for ages 60 to 109

the distribution of future trend changes. In this paper, we avoid this inconsistency by including the distribution of future trend changes in the calibration of historical trends. We extend the calibration to a pure maximum likelihood estimate, including the distribution of future trend changes. However, it becomes apparent that this estimation is not feasible from a practical point of view. Therefore, we deduce an approach, based on the pure maximum likelihood estimation, that can be calibrated efficiently.

The remainder of this paper is structured as follows: In Section 2, we structurally compare different trend processes. Parameter estimation of these approaches is then discussed in Section 3. In particular, we present a calibration approach that incorporates the distribution of future trend changes to achieve more consistent forecasts. In Section 4, we focus on uncertainties in the parameter estimation and explain how the uncertainties can be accounted for in a practical application. Finally, Section 5 concludes.

2 Structural comparison of trend processes

As outlined in the Introduction, we compare now different approaches to model mortality trends in a population (summarized in Table 1 in Appendix B). In this section, we distinguish between two groups of trend processes: A larger group that focus on the changing historical trend processes and extrapolate the most recent trend and a smaller group of trend processes that also start with a changing historical trend process but continue to include trend changes also in future simulations. In both groups, difference stationary processes (random walk models) or trend stationary processes (trend models) can be used. Although these models are very similar, they also have a crucial difference that we can use to decide which model is suitable in a particular case.

2.1 Difference and trend stationary processes

Broadly known models based on random walks are well studied and often the first choice to model the period effect κ_t of a parametric mortality model. These models use cumulative, difference stationary error terms together with a common drift:

$$\kappa_t = \kappa_{t-1} + d + \epsilon_t.$$

One-year errors ϵ_t remain in the future evolution of κ_t and hereby influence estimates of κ_T where $T \geq t$. Hence, uncertainty in short-term and long-term forecasts of random walks with constant drifts are driven by the underlying volatility structure. In order to reflect the long-term uncertainty as observed in the period effects in Figure 1, a high degree of volatility would be necessary, which in turn would massively overestimate the short-term uncertainty. Therefore, a second source of uncertainty is required, e.g. a variable drift. More general models like the class of ARIMA(p,d,q) models include error structures for several years but are structurally also not suitable to reflect structural changes.[1]
Similar to the random walk models, trend models are another option to model period effects. These models use trend stationary errors around a linear trend. The easiest form of a trend model would be a simple linear regression, described by

$$\kappa_t = \hat{\kappa}_t + \epsilon_t,$$

where $\hat{\kappa}_t = \beta_0 + \beta_1 \cdot t$ with β_0 the starting value and β_1 the corresponding slope of the trend and $\hat{\kappa}_t$ the underlying period effect κ_t net of random fluctuations.

The formula of the random walk with drift and the formula of the trend model are very similar and basically differ in the error structure. With a difference stationary structure, the correlation coefficient of subsequent changes in the period effect ($Corr(\Delta\kappa_t, \Delta\kappa_{t-1}) = 0$ with $\Delta\kappa_t = \kappa_t - \kappa_{t-1}$) should be roughly zero. In a trend stationary model it can be shown that a negative correlation of subsequent changes in the period effect is expected ($Corr(\Delta\kappa_t, \Delta\kappa_{t-1}) = -1/2$ [2]). We calculated the empirical correlation coefficients for 56 period effects (standard CBD

[1] For applications of ARIMA (p,d,q) models in parametric mortality models, see Chan et al. (2014) and Richards et al. (2014).

[2] With the restriction of the trend component being treated as given. With changes in the trend process a closed

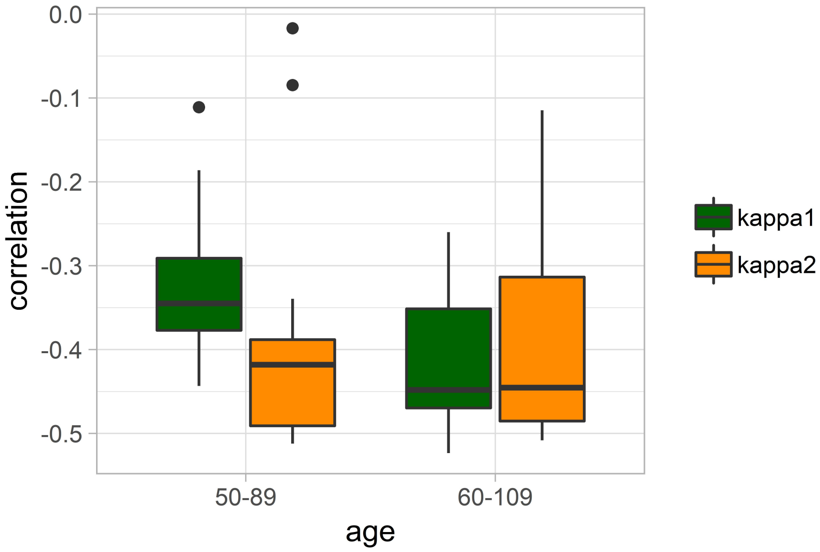

Figure 2: Boxplot of empirical correlation coefficients of subsequent changes in 60 period effects

model κ_t^1, κ_t^2 , males and females, different age spans ($50-89$, $60-109$) for England & Wales, Sweden , USA, Australia, France, Spain, Netherlands[3]) see Figure 2. Most period effects seem to be trend stationary. However, there are also exceptions (e.g. Swedish Males aged $50-89$ with a $Corr(\Delta\kappa_t, \Delta\kappa_{t-1}) = -0.11$), and therefore it is necessary to decide on a case by case basis which trend process is suitable.

2.2 Models without changes in the trend process in projections

Many authors discuss changes in the historical parameters of trend processes, e.g. a time dependent drift d_t in a random walk model or a time dependent slope β_t in a trend model. Booth et al. (2002) develop an approach for the specification of optimal periods for the estimation of the random walk with drift in the Lee and Carter (1992) model. This already implies the assumption of different mortality scenarios. An explicit change in the Lee and Carter (1992) model is tested in the approaches of Li et al. (2011) and Coelho and Nunes (2011), that find a change in almost every population under consideration. Applying the Zivot and Andrews (2002) procedure Li et al. (2011) test the null hypothesis that the period effects should be modelled with a random walk with constant drift against the alternative of a broken trend stationary mode. For England & Wales and the United States they suggest to model the underlying period effects $\hat{\kappa}_t$ with a trend

formula cannot be derived. However, since changes occur only rarely and are a priori unknown, the correlations should nevertheless provide a valuable indication for the model choice.

[3]We limited the analysis to data which is marked as reliable on the HMD website (see Human Mortality Database (2018)), e.g. years 1860 onwards for Sweden.

change and jump in in the mid 1970s. To our perspective, the trends appear rather continuously and jumps should be rather small. Moreover, jumps corresponding to pandemia should only influence the long term trend marginally. Also other authors apply continuous linear trends in related processes in mortality modeling, e.g. Gillings et al. (1981) use a piecewise linear trend in the modeling of perinatal death rates and Wilmoth (2000) model the maximum age at death with a piecewise linear trend. Also combinations of these models are possible, e.g. Njenga and Sherris (2011) use an autoregressive process around a common fixed trend.
O'Hare and Li (2015) test the significance of a trend change with a test for the residuals based on Ploberger and Krämer (1992) for different parametric mortality models[4] in four countries (UK, US, Netherlands and Australia) and found significant trend changes. Using the Bai and Perron (1998, 2003) method they also find an optimal trend change around the 1970s in any situation. The clarity of a trend change in the 1970s is very interesting, but the period effects in Figure 1 indicate further trend changes. Including the possibility of multiple structural changes van Berkum et al. (2014) allow for multiple trend changes. With the method of Bai and Perron (1998, 2003) the historical period effects are modelled as a random walk with multiple times changing drift, i.e. a time dependent drift $d = d_t$ with $(l-1)$ drift changes in $t_1, \dots, t_l$:

$$\kappa_t = \kappa_{t-1} + d_k + \epsilon_t, t \in (t_{k-1}, t_k].$$

Although the methods for the identification of historical trend changes vary, the idea for the projection of future mortality is based on the same assumption: after estimating the historical change(s), the projection model estimation period is limited to the data after the last trend change. Both for short-term forecasts and for the long-term best estimate, this trend based on a limited period seems reasonable. However, the existence of multiple historical trend changes in almost all applications suggests that trend changes may also be possible in the future. Especially for long-term forecasts, the assumption of a constant trend is therefore implausible and in particular inconsistent with the historical calibration.

2.3 Models with changes in the trend process in projections

There are models that address this inconsistency and include trend changes in simulations. Several authors propose models with a regime switch. Milidonis et al. (2011) and Lemoine (2014) propose switches between two fixed regimes in models with Brownian motion and autoregressive processes, respectively. Here the different regimes appear Markov with transitions in between. For instance, Hainaut (2012) uses two regimes $\eta_t = 1, 2$ with changes in the trend and volatility

[4]Models of Lee and Carter (1992), Cairns et al. (2006), Plat (2009) and O'Hare and Li (2015)

for French males and females (κ_t^i corresponds to the Lee-Carter model with two dimensions):

$$\begin{pmatrix} d\kappa_t^1(\eta_t) \\ d\kappa_t^2(\eta_t) \end{pmatrix} = \begin{pmatrix} a_1(\eta_t) \\ a_2(\eta_t) \end{pmatrix} dt + \begin{pmatrix} \sigma_1(\eta_t) dW_t^1(\eta_t) \\ \sigma_2(\eta_t) dW_t^2(\eta_t) \end{pmatrix} \qquad \eta_t = 1, 2$$

where W_t^1 and W_t^2 are correlated Brownian motions. These two regimes represent two scenarios: For males the regimes differ mainly in the volatility. For females the regimes differ in trend and volatility. However, the restriction to two rigid scenarios is not possible in many situations. Further refining van Berkum et al. (2014)'s approach, Hunt and Blake (2015) include random changes in the predicted drift. They reformulate the random walk model as:

$$\kappa_t = \kappa_{t-1} + d_t + \epsilon_t$$

where d_t is the possibly changing drift. If there is a change in the drift from year $t-1$ to t the magnitude of the trend change is modelled with ν_t, i.e.:

$$d_t = \begin{cases} d_{t-1} + \nu_t & \text{, if a trend change occurs} \\ d_{t-1} & \text{, if no trend change occurs} \end{cases}$$

On the basis of optimal historical drifts, Hunt and Blake (2015) assume that the frequency of changes remains constant over time, e.g. 2 changes in a 64-year period effect would indicate a trend change probability of $\frac{2}{64}$. In the event of a future drift change, the authors propose a decomposition of ν into the components sign and absolute magnitude. Both estimates are subject to significant uncertainty, as only a few historical changes are observable. For simplicity, they assume that positive drift changes are as likely as negative drift changes. The absolute magnitude of the drift change is modelled with a Pareto distribution. To ensure consistency with the historical calibration, the authors estimate the threshold parameter as the minimal size of a statistically significant drift change at a 99% confidence level. Hunt and Blake (2015) argue that using the ML-estimate (the minimum observed value) would truncate the simulated drift changes to the smallest observed. However, this seems to be a rather specific issue of the Pareto distribution and ML-estimators can be used when applying other distributions. The scale parameter of the Pareto distribution is chosen such that the distribution's mean coincides the sample mean, i.e. the historical mean of absolute trend changes. A disadvantage of the Pareto distribution is certainly that the monotony puts a lot of probability mass to values close to the minimum. With the constraint of identical means, this may lead to an overestimated variance.

There are two approaches that model a changing mortality trend as a continuous piecewise linear trend (see Sweeting (2011) and Börger and Schupp (2018)) and include the possibility of future trend changes in their approaches. The underlying trend process net of fluctuations $\hat{\kappa}_t$ uses the possibly changing trend to be recursively defined as

$$\hat{\kappa}_t = \hat{\kappa}_{t-1} + \beta_t$$

with β_t the possibly changing trend between year $t-1$ and t. In the case of a trend change between $t-1$ and t, the slope of the trend changes from β_{t-1} to $\beta_{t-1} + \lambda_t$. If there is no trend change, the current trend remains, i.e.:

$$\beta_t = \begin{cases} \beta_{t-1} + \lambda_t & \text{, if a trend change occurs} \\ \beta_{t-1} & \text{, if no trend change occurs.} \end{cases}$$

Sweeting (2011) and Börger and Schupp (2018) estimate the trend change frequency similar to the approach of Hunt and Blake (2015), i.e. assuming constant frequencies over time. Moreover, Sweeting (2011) observes a strong dependency between trend changes in the two period effects of the CBD-model. Börger and Schupp (2018) find no valid evidence for dependencies in the trend changes of the period effects. Whilst Börger and Schupp (2018) apply a similar decomposition as Hunt and Blake (2015), Sweeting (2011) uses a normal distribution to simulate the magnitude of future trend changes. Based on positive and negative historical trend changes this normal distribution would simulate numerous trend changes close to zero which is contradictory to the historical estimation, where trend changes close to zero would not have been detected. Börger and Schupp (2018) use a lognormal distribution for the absolute magnitude of trend changes. In contrast to the Pareto distribution, this distribution does not require a lower bound for trend changes, but will assign trend changes near zero a very low probability mass. If the historical trend changes are rather inhomogeneous, the simulation of future trend changes will no longer be consistent with the observed historical trend changes as those will be hardly reproduced by the distribution of future trend changes. We avoid this inconsistency by including the distribution of future trend changes in the calibration of historical trends in the next section.

3 Calibration of processes with variable drifts and variable trends

In the previous section, we compared different approaches to model the period effects of a parametric mortality model with trend processes. We proceed comparing different approaches of parameter estimation to make forecasts with a trend process. In this section, we first show how distributional assumptions are included in existing approaches and we illustrate why this is necessary. Then, we present a new calibration technique, that include the distribution of future trend changes and thus improve consistency of forecasts.

3.1 Statistical Analysis of historical trend processes

Some historical evolutions of mortality show a few one-year outliers due to extremal events. A popular example is the Spanish flue in 1918 with massive influence on annual death rates. Influences like wars, economic depressions or large migrations may affect mortality a few years. But they have no sustainable influence on a longer perspective trend. Thus, they should not be included in the trend model used for forecasts. Medical improvements reduced the effects of years with strong influenza or other diseases with immediate effects on annual death rates. The countries under consideration also showed an increase in population size. Combining these medical and statistical effects results in significantly decreasing volatility in annual mortality rates. The reasons put forward indicate that decreasing volatility seems to be sustainable, which is why we assume that volatility will not return to earlier levels. We can use statistical tools to account for outliers (see Grubbs (1951) and Appendix C) or structures in the errors like the general decrease in volatility, e.g. Börger and Schupp (2018) include a non-parametric CUSUM test (see Ploberger and Krämer (1992)).

Existing approaches, e.g. Sweeting (2011), Börger and Schupp (2018), and implicitly Hunt and Blake (2015) use this volatility assumption to subsequently calibrate trends that optimize the likelihood of the resulting errors. With iterative algorithms it is possible to update these estimates consecutively. For instance, if we have an estimate of a trend process with $k-1$ trend changes we can use the corresponding errors to estimate a variance for the calibration of k trend changes. For $k=0$ we can start with a constant variance and than update the estimate.

The method of Bai and Perron (1998, 2003) is a fast approach to identify trend changes based on the sum of squared residuals. With ordered data of length n, the dynamic algorithm performs estimates for the optimal trend for each possible partition. This can be done a-priori as there are only $\frac{n\cdot(n-1)}{2}$ partitions in an ordered sample. For any $1 \leq i_1 < i_2 \leq n$ the algorithm estimates a trend, such that the sum of squared residuals gets minimal. With a constant volatility assumption this is equivalent to a maximum likelihood estimation of normal errors:

$$l(\beta, c; \kappa_t, t_1+1, t_2) = \sum_{t=t_1+1}^{t_2} \left(\hat{\kappa}_t(\beta, c, t_1+1, t_2) - \kappa_t\right)^2$$

where $\hat{\kappa}_t(\beta, c, t_1+1, t_2) = c_{t_1+1,t_2} + \beta_{t_1+1,t_2} \cdot t$ in the case of a linear regression. Any combinations of multiple trend changes is a combination of these simple linear trends, therefore a calibration can be programmed very efficiently. For any trend process with m changes in

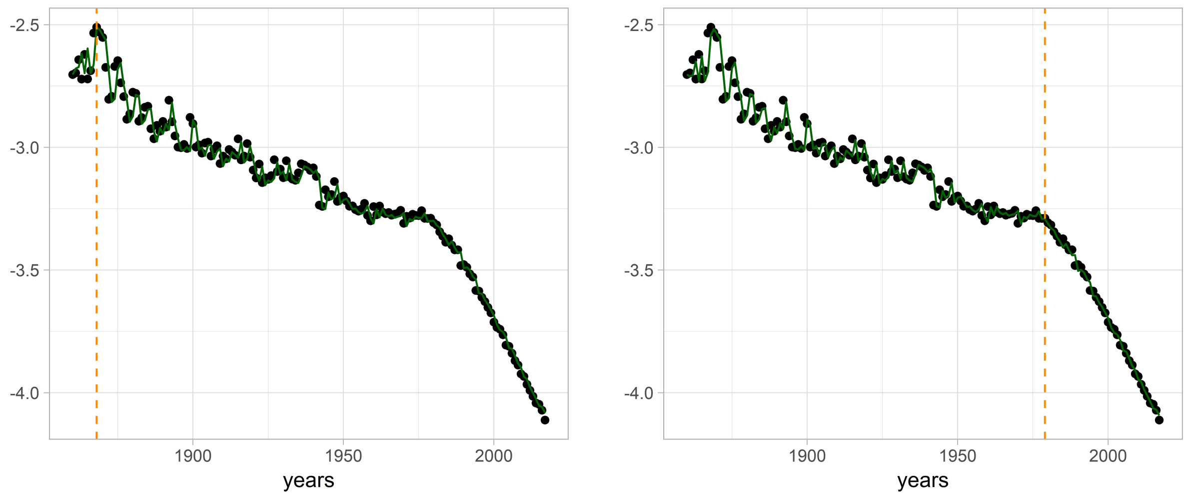

Figure 3: Period effect κ_t^1 for Swedish males for the ages $50-89$ (dotted) and a random walk with changing drift with (right) and without (left) heteroscedasticity (solid lines). The dashed lines indicate the year of the drift change.

$t_0 = 0, t_1, \cdots, t_m, t_{m+1} = n$ the likelihood optimal trend would be:

$$\sum_{i=0}^{m} l(\beta, c; \kappa_t, t_i + 1, t_{i+1}).$$

Hunt and Blake (2015) and van Berkum et al. (2014) differentiate period effects to estimate the optimal partition of a multiple times changing random walk with drift. This is necessary as the errors are difference stationary. They use BIC (see Burnham and Anderson (2002)) to identify a parsimony model.Longer period effects often show curious trends. This is mainly caused by the aforementioned heteroscedasticity in the period effects. Using the likelihood perspective, the method can be extended straightforwardly by using normed residuals, i.e.:

$$l(\beta, c; \kappa_t, t_1 + 1, t_2, \sigma_t^2) = \sum_{t=t_1+1}^{t_2} \frac{(\hat{\kappa}_t(\beta, c, t_1 + 1, t_2) - \kappa_t)^2}{\sigma_t^2} \tag{1}$$

A period effect, that should probably be modelled with a difference stationary trend process is the period effect of Swedish males aged $50 - 89$ ($Corr(\Delta\kappa_t, \Delta\kappa_{t-1}) = -0.11$ as outlined in Section 2). Figure 3 shows the estimated trend processes with and without consideration of heteroscedasticity. While the approach without heteroscedasticity shows curious effects, the approach with consideration of heteroscedasticity estimates a convincing trend change in 1979. In the case of a piecewise linear continuous trend process the method of Bai and Perron (1998, 2003) cannot be applied, as due to the continuity slopes are strongly dependent of each other.

Therefore, these trend processes require different types of calibration methods.

The method of Muggeo (2003) can be used to estimate an optimal trend process with respect to the number of trend changes and their positions simultaneously. Börger and Schupp (2018) transferred this approach to the field of mortality modeling and showed how this approach can be extended to include the mortality specific structures in volatility. Similar to the extended version of the Bai-Perron method (see Equation 3.1), the approach estimates optimal historical trends with a volatility assumption, i.e. $l(\beta, c, \tau; \kappa_t, \sigma_t^2)$, where the trend changes it's slope in the years τ. However, the continuity constraint of piecewise linear trends makes calibration considerably more difficult.

All presented calibration approaches take the historical trend process as the basis for estimating the parameters for a simulation. In most cases, there are only a few inhomogeneous trend changes. Therefore, the simulation of future trend changes will no longer be consistent with the sparse observed historical trend changes. To tackle this problem, there are basically two options: Adjusting the projection approach can be a possible solution, e.g. if a simulation only includes trend changes that would also have been detected within the historical trend calibration. However, the strong dependencies between trend changes due to the continuity constraint make this a challenging and not feasible task. The other possibility is to adjust the historical calibration such that only trends likely to be generated with the projection's distribution should be calibrated. In the next subsections, we show how to calibrate the latter and introduce a new calibration algorithm that leads to more consistent predictive models.

3.2 Combined Calibration based on likelihood

Starting a simulation in $t = 0$ based on an observed period effect $\kappa_{-N}, \ldots, \kappa_0$ the log-likelihood function l of a trend process from the set of trend processes $(\hat{\kappa}_{-N}, \ldots, \hat{\kappa}_0) \in \mathbb{R}^{N+1}$ is optimized. Assuming independent normally distributed errors we get:

$$\max_{\hat{\kappa}_{-N},\ldots,\hat{\kappa}_0} l(\hat{\kappa}_{-N}, \ldots, \hat{\kappa}_0) = \max_{\hat{\kappa}_{-N},\ldots,\hat{\kappa}_0} \sum_{i=-N}^{0} \log f_{K_i}(\kappa_i | \hat{\kappa}_i, \sigma_i^2),$$

where $f_{K_i} \sim \mathcal{N}(\hat{\kappa}_i | \sigma_i^2)$ with known σ_i^2. This is the main assumption of the approaches presented at the beginning of this section. However, forecasts are actually possible without explicit knowledge about the historical trend. Forecasts require an estimation of the starting value of the trend process, the starting slope of the trend and the parameters of the distributions used for forecasts, e.g. in the case of a lognormal distribution $\theta_T = (\hat{\kappa}_0, \beta_0, p, \sigma_{\mathcal{N}}, \sigma_{\mathcal{LN}}, \mu_{\mathcal{LN}})$ is required. The density function of a random trend process $K_{-N}, \ldots, K_0$ (with random changes in the slope at random points in time) can be formulated by:

$$\begin{aligned}
f(K|\theta_T) = {} & f_{\mathcal{N}}(\epsilon_0|\sigma^2_{\mathcal{N}}, \hat{\kappa}_0) \cdot f_{\mathcal{N}}(\epsilon_{-1}|\sigma^2_{\mathcal{N}}, \hat{\kappa}_0, \beta_0) \\
& \cdot \int_{\mathbb{R}^{N-1}} \prod_{s=2}^{N} g(\lambda_{-s-2}|p, \sigma^2_{\mathcal{LN}}, \mu_{\mathcal{LN}}) \\
& \cdot f_{\mathcal{N}}(K_{-s} - (\hat{\kappa}_0 - s\beta_0 + \sum_{l=1}^{s-1} l \cdot \lambda_{-(s-1-l)})|\sigma^2_{\mathcal{N}}, \hat{\kappa}_0, \beta_0) d\lambda_{-(N-2)}, \ldots, d\lambda_0,
\end{aligned}$$

where we can use a recursive formulation for the trend process:

$$\hat{\kappa}_{-s} = \hat{\kappa}_0 - s \cdot \beta_0 + \sum_{l=0}^{s-1} l \cdot \lambda_{-(s-1-l)} \qquad \forall s \in \mathbb{N}.$$

The density function g represents the random trend changes, i.e. $g(\lambda_i) = (1-p)$ in the case of no trend change ($\lambda_i = 0$) and $g(\lambda_i) = \frac{p}{2} \cdot f_{\mathcal{LN}}(|\lambda_i|)$ in the case of a trend change ($\lambda_i \neq 0$), where $p/2$ represents the probability of observing a positive or negative trend change. $f_{\mathcal{LN}}$ is the lognormal density function and represents the absolute magnitude of the trend change[5]. Consequently we can use this density to derive a likelihood formulation:

$$l(\theta_T|\kappa) = \log(f(\kappa|\theta_T)).$$

In the case of a random walk with changing drift, forecasts require a slidely different set of parameters, namely the starting drift d_0 instead of starting slope and starting value: $\theta_D = (d_0, p, \sigma_{\mathcal{N}}, \sigma_{\mathcal{LN}}, \mu_{\mathcal{LN}})$. Here, the density function of reformulates to:

$$\begin{aligned}
f(K|\theta_D) = {} & f_{\mathcal{N}}(\epsilon_{-1}|\sigma^2_{\mathcal{N}}, d_0) \\
& \cdot \int_{\mathbb{R}^{N-1}} \prod_{s=2}^{N} g(\lambda_{-s-2}|p, \sigma^2_{\mathcal{LN}}, \mu_{\mathcal{LN}}) \\
& \cdot f_{\mathcal{N}}(K_{-s} - (\kappa_{-s+1} - d_0 + \sum_{l=1}^{s-1} \lambda_{-(s-1-l)})|\sigma^2_{\mathcal{N}}, d_0) d\lambda_{-(N-2)}, \ldots, d\lambda_0,
\end{aligned}$$

where we can use $d_{-s} = d_0 - \sum_{l=1}^{s-1} \lambda_{-(s-1-l)}$.

By maximizing this function we could estimate the parameters for a joint distribution which has the highest likelihood of observing the period effect for one specific realization. However, this

[5]Note that the distributional assumptions can be modified, e.g. use a Pareto distribution for the trend changes magnitude.

high-dimensional integral can't be calculated analytically. Numerical techniques or Monte-Carlo techniques to calculate integrals failed due to the dominating pole of the symmetric function g around zero. The likelihood function of this variable trend model can be derived in an elegant mathematical way. However, with the available methods, convergence can't be achieved. Therefore, we propose to use a simplified Pseudo Maximum Likelihood approach, which has the essential advantages of this model and can be calibrated efficiently.

3.3 Combined Calibration based on Pseudo Likelihood

The structure of g, namely the dependencies between errors and trend changes on realized period effects made a calibration so difficult. If we look on historical period effects, there seems to be no evidence for a dependence between errors and trend changes. Therefore, we combine two likelihoods independently:

1. Likelihood of observing a specific process with k changes in the trend.
2. Likelihood of observing the error structure compared to the period effect under consideration given the specific process realization from 1.

Combining these two likelihoods, we get an alternative Pseudo Maximum Likelihood estimation to calibrate optimal historical trend processes with $k \in \mathbb{N}$ trend changes:

$$\tilde{l}(\beta, c, \tau; \kappa_t, \sigma_t^2, \mu_{LN}, \sigma_{LN}, p) = l_{\mathcal{N}}(\beta, c, \tau; \kappa_t, \sigma_t^2) + l_{\mathcal{LN}}(\beta; \mu_{LN}, \sigma_{LN}) + l_{\mathcal{B}}(p). \tag{2}$$

By independently combining the two likelihoods, we change the range of optimal trend processes to choose from with the information criteria from a pure error analysis as in the approaches of Hunt and Blake (2015) or Börger and Schupp (2018) to a combined error-inference analysis. This was also the purpose of the infeasible full likelihood approach. Nevertheless, this is no longer a pure likelihood optimization, nor a comparison of likelihood optimal trends for different numbers of trend changes. Instead, we compare and estimate 'optimal' realizations (errors and trend changes) from a combined but complex distribution.

This modification allows to estimate the parameters efficiently again. Again, we can apply the Muggeo (2003) algorithm to solve this optimization. For a fixed number of k trend changes, we can optimize:

$$\tilde{l}(\beta, \gamma, c; \kappa_t, \tau^{(0)}, \sigma_t^2, \mu_{LN}, \sigma_{LN}, p) = \sum_{j=1}^{k} \log\left(f_{\mathcal{LN}}(|\beta_j|; \mu_{LN}, \sigma_{LN})\right) + l(\beta, \gamma, c; \kappa_t, \tau^{(0)}, \sigma_t^2).$$

As the simulation parameters θ_T are an intrinsic part of the calibration it is necessary to include their estimation within the iterative calibration. Instead of updating only the error's volatility

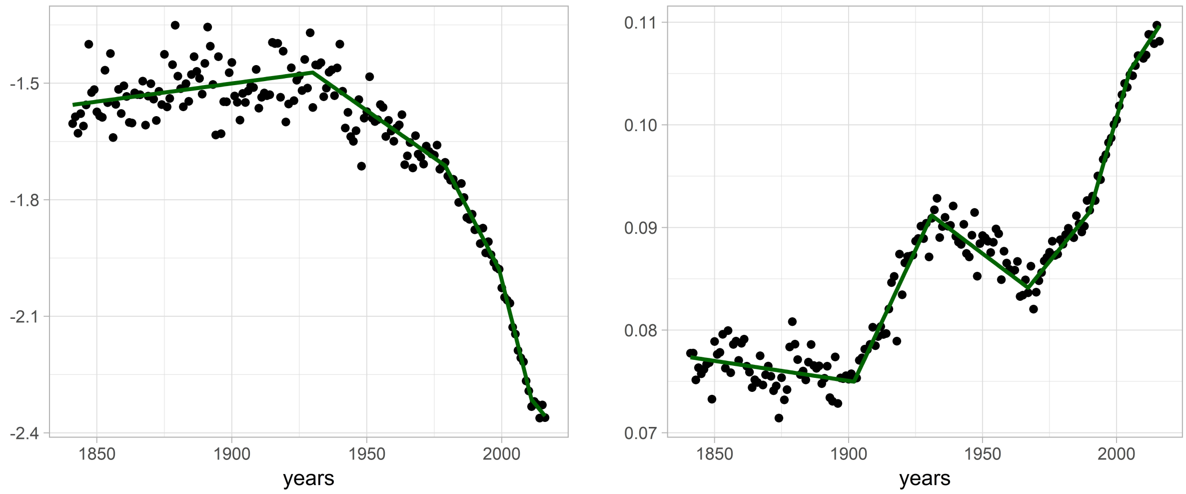

Figure 4: Historical period effects κ_t^1 (left) and κ_t^2 (right) for English and Welsh males (dotted). Pseudo Maximum Likelihood estimated best possible realization of underlying trend process (solid lines)

iteratively we include the parameters of the trend change distribution $(p, \mu_{\mathcal{LN}}, \sigma^2_{\mathcal{LN}})$ in the iterative scheme (see Appendix D).

The number of parameters to be estimated for a trend process realization with k trend changes is $(3k+3)$. Therefore, the BIC can be used to compare optimal realizations u_k for different k:

$$BIC = -2 \cdot (l_{\mathcal{N}}(\beta, c, \tau; \kappa_t, \sigma_t^2) + l_{\mathcal{LN}}(\beta; \mu_{LN}, \sigma_{LN}) + l_{\mathcal{B}}(p)) + (3k+3)\ln(n).$$

Applied to the period effects of English and Welsh males aged $60-109$ we observe four changes in the BIC-optimal trend process realization for κ_t^1 and five changes in the BIC-optimal trend process realization for κ_t^2 (see Figure 4). For fixed parameters $(\sigma_t^2, \mu_{\mathcal{LN}}, \sigma^2_{\mathcal{LN}}, p)$, it can be shown, that the Pseudo Likelihood part of the BIC is in expectation typically not constant. For instance,

$$\mathbb{E}\mathcal{L}_{\mathcal{LN}}(\beta) = k \cdot \frac{1}{2\sqrt{\pi\sigma^2_{\mathcal{LN}}}} \exp(-\mu_{\mathcal{LN}} + \frac{\sigma^2_{\mathcal{LN}}}{4})$$

is monotonically increasing in k. In this situation, the BIC is not only used to balance between under- and overfitting due to randomness. In addition, the BIC compensates systematic growth in the Pseudo Likelihood part with increasing numbers of trend changes. In most of the period effects studied, the BIC seemed capable of doing so. However, there are also cases where the BIC is unable to compensate for this systematic effect. With a normalized Pseudo Maximum Likelihood approach this systematic effect can be avoided. Therefore, we normalize the optimisation such

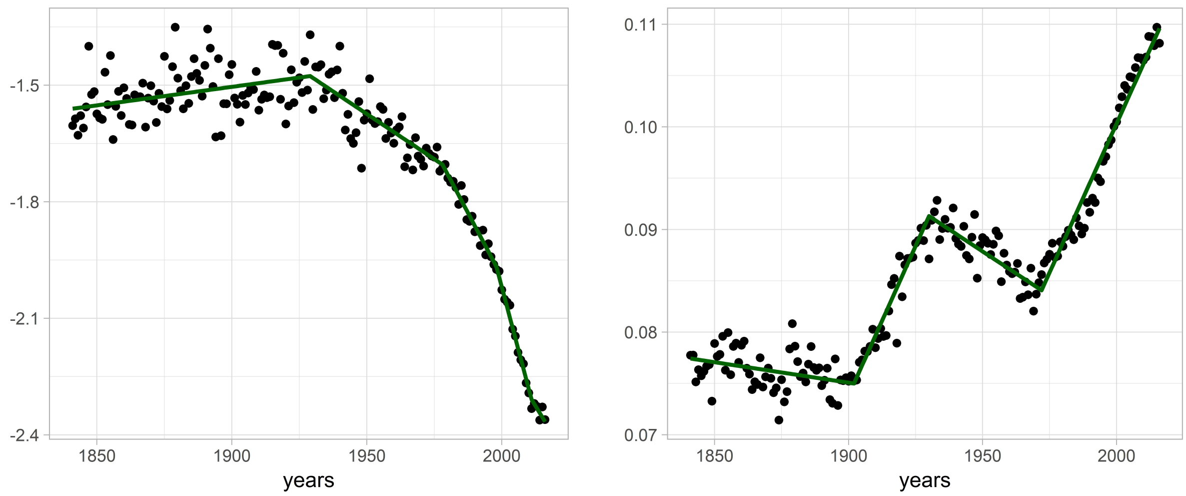

Figure 5: Historical period effects κ_t^1 (left) and κ_t^2 (right) for English and Welsh males (dotted). Normed Pseudo Maximum Likelihood estimated best possible realization of underlying trend process (solid lines)

that it is directly comparable for different numbers of trend changes, i.e. we replace Equation 2 by:

$$\tilde{l}(\beta, c, \tau; \kappa_t, \sigma_t^2, \mu_{LN}, \sigma_{LN}, p) = l_{\mathcal{N}}(\beta, c, \tau; \kappa_t, \sigma_t^2) + l_{\mathcal{LN}}(\beta; \mu_{LN}, \sigma_{LN}) + l_{\mathcal{B}}(p) - \log(\mathbb{E}\mathcal{L}_{\mathcal{N}} \cdot \mathbb{E}\mathcal{L}_{\mathcal{LN}} \cdot \mathbb{E}\mathcal{L}_{\mathcal{B}}). \tag{3}$$

Applied again to the period effects of English and Welsh males (see Figure 5), we observe basically identical optimal trend process realizations for κ_t^1. For κ_t^2 we observe an optimal realization with only three trend changes. However, also the realization with five trend changes still has a significant probability.

4 Parameter Uncertainty

The parameter estimation approaches presented in the previous section include a significant estimation risk. In this section, we show an efficient way to account for parameter uncertainty. The estimation risk of these models can be divided into two main categories (see Börger and Schupp (2018)): The uncertainty in the trend change parameters (p, μ, σ^2). Exemplary an overestimation (underestimation) of these parameters result in unrealistically wide (narrow) long-term prediction intervals. Secondly, the uncertainty in the starting values ($\hat{\kappa}_0, \beta_0$ (trend stationary) or d_0 (difference stationary)) affects especially short-term estimates directly. This uncertainty can vary extremely when comparing different countries and period effects as it is dominantly driven by the latest observations.

Some approaches to account for parameter uncertainty are computationally intensive and are therefore often disadvantageous in a practical use. With the assumption of deaths following a Poisson or Binomial distribution, Koissi et al. (2006) propose to recalibrate the period effects first and then re-estimate the parameters used for forecasts. In the situation of trend stationary period effects, the optimal historical realization needs to be recalibrated with the iterative scheme for each simulated period effect. Clearly, this is computationally expensive. Moreover, approaches based on recalibration can only comprehend the uncertainty in the starting values partially as the potential additional change would not have been detected simply due to the limited data afterwards and not due to random effects captured with the Poisson/Binomial distribution.
The approaches of the previous section estimate optimal realizations of trend processes by comparing estimates for different numbers of trend changes. The main uncertainty is about the actual number of trend changes. Due to random effects, it is possible that the optimal estimate overestimates or underestimates the number of trend changes. As a side-product of the calibration we already have estimates of the trend change parameters for different numbers of trend changes k denoted by $\zeta_k = (p_k, \mu_k, \sigma_k^2)$ and $\eta_k = (\hat{\kappa}_{0,k}, \beta_{0,k})$. The uncertainty in the starting values mainly arises due to limited data after the last trend change. For instance, the last trend in κ_t^1 only includes a few years and also a starting value and slope without that last trend change should be included in a simulation.

Börger and Schupp (2018) propose to adopt Bayesian weights to compare different parameter sets. We expand this approach, such that the uncertainty in the trend change parameters, in the starting values and in the volatility structure are covered jointly. The calibration provides parameter sets for different numbers of trend changes that can be used straightforwardly to assign probabilities to θ_k and η_k. Using the optimization in Equation 3 and let $\tilde{l}(\cdot, k) = \tilde{l}(\beta, c, \tau; \kappa_t, \sigma_t^2, \mu_{LN}, \sigma_{LN}, p)$ denote the Pseudo Maximum Likelihood for k trend changes:

$$\mathbb{P}(k \text{ trend changes}) = w_k = \frac{\exp(l(\cdot, k) - l(\cdot, \hat{k})}{\sum_{i=1}^{n} \exp(l(\cdot, i) - l(\cdot, \hat{k}))},$$

where $l(\cdot, \hat{k})$ is the overall optimal value of all possible k. This approach can also be applied to other optimizations based on Maximum Likelihood, i.e. also to approaches based on information criteria as in the approaches of Hunt and Blake (2015) or van Berkum et al. (2014).
We can use these weights to assign probabilities for the trend change parameters to account for the unclear number of actual trend changes, i.e.:

$$\mathbb{P}(\zeta = \zeta_k) = w_k \text{ and } \mathbb{P}(\eta = \eta_k) = w_k$$

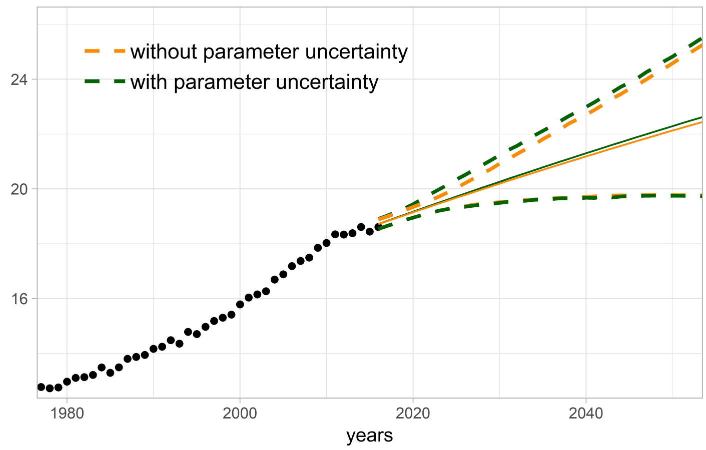

Figure 6: Historical and forecast 90% prediction intervals (dashed lines) and median (solid lines) for remaining period life expectancy of 65-year old males in England & Wales

Focussing on forecasts, the uncertainty in the error structure Σ_t^2 has a very small influence. The general decline in volatility seems to be permanent and therefore the covariance matrix of the errors is assumed to be constant in forecasts. Moreover, the trend stationary structure of a trend process involves only a year-by-year influence of errors in forecasts. Also the uncertainty around the central estimate is rather small as the estimate is based on significantly bigger sample as in the case of the trend changes. Therefore, it seems appropriate to forego this uncertainty. We estimate covariances from the variance estimate and the historical errors from each period effect combination κ_t^1 with k trend changes and κ_t^2 with m trend changes denoted by $\Sigma_{k,m}$. We combine different numbers of trend changes again, by assigning weights, i.e. with the associated weights $w_k^{(1)}$ for κ_t^1 and $w_m^{(2)}$ for κ_t^2 we deduce the following probabilities:

$$P(\Sigma = \Sigma_{k,m}) = w_k^{(1)} \cdot w_m^{(2)}.$$

With these specification we first draw in each simulation path $(\zeta, \eta, \Sigma) = \theta_T$ and second simulate the future evolutions of the period effects. Figure 6 shows the 90% prediction intervals of the period life expectancy for English and Welsh males for the ages $60 - 109$ forecasted with and without parameter uncertainty. If forecasts are based on one optimal historical realization only, the uncertainty would be underestimated. This is especially visible at the upper tail. Here, uncertainty is generated by the possibility that the latest slowdown in mortality improvements only was random and not the beginning of an ongoing different trend. Scenarios, where the recent slowdown in mortality improvements was not sustainable are taken in 7% of all simulation paths. Thus, both scenarios, a flattening and a persistence of the recent trends in mortality improvements, are well included in the prediction intervals.

5 Conclusion

Throughout the recent years a variety of stochastic mortality models have been developed and most of these models use one or more time dependent period effect. Forecasting these period effect(s) requires to analyse the historical shapes, where we can observe changes in the improvements of mortality rates in most countries.
In this work we have analysed approaches that account for a trend process with changing trend component. Existing methods for the calibration of a mortality model with trend changes use historical trends solely to derive the parameters for forecasts. If the historical trend changes were rather inhomogeneous, the simulation of future trend changes would be no longer consistent with the observed historical trend changes as those would hardly be reproducible by the distribution of future trend changes. We avoided this inconsistency as we included the distribution of future trend changes in the calibration of historical trends. We presented an alternative calibration approach that adjusted the historical calibration such that only trends likely to be generated with the projection's distribution are calibrated. With a full likelihood based approach, we have introduced a convoluted density for the process with variable trend. However, the numerical optimization failed due to the complexity of this function. Therefore, we have developed an estimation of the parameters that combines the the two likelihoods of observing a specific realization of a trend process with the likelihood of the resulting errors. By merging these two perspectives, we have outlined a procedure for calibrating consistent trend processes, that can be used straightforwardly in forecasts.

Finally, we have discussed parameter uncertainty in models with variable trend processes. The main uncertainty in the estimation of the parameters of a trend processes is certainly the possibility of underestimating or overestimating the actual number of historical trend changes due to noisy data. This uncertainty can be taken into account by considering optimal realizations for different numbers of trend changes in a simulation. Also realizations with (without) the last optimal trend change are included in forecasts, if there is a significant uncertainty about the starting values. The limited amount of data after a possible last trend change in the historical trend results in a considerable uncertainty about the last trend change. There could be an undetected additional change in the trend afterwards. Just as well it may be possible that the last detected trend change was only mistakenly accepted and only a result of random noise. Only with further observations of this trend, we can be sure about the actual development. For now, however, it is important to consider all possibilities in simulations. Existing approaches based on recalibration can not fully reflect this uncertainty.

We therefore conclude that the presented features allow to calibrate a mortality model with consistent variable trend processes. On the other hand, the presented methods with variable mortality trend represent significant improvements in several aspects over existing approaches and therefore constitute a valuable alternative.

Appendix

A Cairns-Blake-Dowd Model

The Cairns-Blake-Dowd (CBD) model (see Cairns et al. (2006)) is a broadly used parametric mortality model with two time dependent period effects describing the basic characteristics of mortality over age and time. Where the first period effect (κ_t^1) describes the general level of mortality over time, the second period effect (κ_t^2) describes the evolution for different ages. The formula of the standard CBD model transforms the logit of death probabilities into the two period effects:

$$\text{logit}(q_{x,t}) = \kappa_t^1 + (x - \bar{x}) \cdot \kappa_t^2,$$

where $\bar{x}$ is the mean of the considered age range, e.g. 84.5 when the age range under consideration is $60-109$. The period effects can be estimated with generalized (non-) linear models (see Villegas et al. (2015) and Currie (2016)) in an effective way. Let $d_{x,t}$ be the observed number of x-year old individuals died in year t which is considered as a realization of the random variable $D_{x,t}$, where

$$D_{x,t} \sim Poisson(m_{x,t} \cdot e_{x,t})$$

with given exposure $e_{x,t}$ and central mortality rate $m_{x,t}$. The corresponding Maximum Likelihood (ML) problem can be formulated as:

$$\mathcal{L}(d_{x,t}, \hat{d}_{x,t}) = \sum_x \sum_t d_{x,t} \cdot \log \hat{d}_{x,t} - \hat{d}_{x,t} - \log d_{x,t}!,$$

see, e.g. Villegas et al. (2015), where we can use the link function $g(\mathbb{E}(\frac{D_{x,t}}{E_{x,t}})) = \kappa_t^1 + (x - \bar{x}) \cdot \kappa_t^2$ to estimate the number of deaths for a given set (κ_t^1, κ_t^2) of period effects:

$$\hat{d}_{x,t} = e_{x,t} \cdot g^{-1}\left(\kappa_t^1 + (x - \bar{x}) \cdot \kappa_t^2\right).$$

Although we focus on the CBD model, the findings derived in this paper can be used for any parametric mortality model with period effects.

B Summaries - specification and calibration of trend processes

Trend	Stationarity	Forecasting	Model	Reference
Constant	d-Difference	historical drift	ARIMA(p,d,q)	Chan et al. (2014), Richards et al. (2014)
Variable	Difference	historical drift	RWD	van Berkum et al. (2014), Booth et al. (2002), Coelho and Nunes (2011)
		regime switch	Brownian motion	Milidonis et al. (2011), Lemoine (2014), Hainaut (2012)
		changing drift	RWD	Hunt and Blake (2015)
	Trend	historical trend	Trend Model	O'Hare and Li (2015), Gillings et al. (1981), Wilmoth (2000)
		historical trend	Trend Model with jumps	Li et al. (2011)
		changing trend	Trend Model	Sweeting (2011), Börger and Schupp (2018)

Table 1: Variants of models with trend processes

C Accounting for Outliers

One-year outliers in data with normally distributed errors can be identified with Grubbs (1951)' test statistic

$$G = \frac{\max_{i \in N} |Y_i - Y|}{S}.$$

The relevant null hypothesis is H_0 : no outliers in the normal distributed error data. Given H_0 the statistic G is t_{N-2} distributed. We propose to apply this approach iteratively on all sets of eleven adjacent data points. In the case of a rejection of H_0 (i.i.d errors) at the 1% significance level, we mark the outlier and assign a weight of zero. We analysed several period effects choosing seven to fifteen adjacent data points and only observed marginal differences. Hence, the choice of eleven seems to be very robust.

D Trend process Calibration

Muggeo (2003) proposes a method to fit a continuous and piecewise linear trend process to a time series with optimality criteria, with respect to the number of trend changes and their positions. First, a linear curve without any trend changes, i.e. a straight line, is fitted with a simple linear regression. The trend process with $k > 0$ trend changes is rewritten as:

$$\begin{aligned} \hat{\kappa}_t &= \hat{\kappa}_0 + \sum_{i=1}^{t} \beta_t \\ &= \hat{\kappa}_0 + \beta' \cdot (t \cdot \mathrm{I} - \tau)^+ \end{aligned}$$

with changes in the slope $\beta' = (\beta_0, ..., \beta_k)$ in the years $\tau = (t_0, \tau_1, ..., \tau_k)'$ and starting value $\hat{\kappa}_0$ of the trend process, i.e.:

$$\beta_t = \begin{pmatrix} \beta_0, \tau_0 \leq t < \tau_1 \\ \beta_0 + \beta_1, \tau_1 \leq t < \tau_2 \\ \vdots \\ \sum_{j=0}^{k-1} \beta_j, \tau_{k-1} \leq t < \tau_k \end{pmatrix}$$

The estimation of the parameters τ, β, and $\hat{\kappa}_0$, is a non-linear optimization problem. Muggeo (2003) suggests to use a Taylor expansion around $\tau^{(0)} = (t_0, \tau_1^{(0)}, ..., \tau_k^{(0)})'$ to simplify the optimization by iteratively solving a series of linear problems that converge to the solution of the non-linear problem. An expansion $\tau^{(0)}$, yields:

$$\begin{aligned}\hat{\kappa_t} &= \hat{\kappa}_0 + \beta' \cdot (t \cdot \mathrm{I} - \tau)^+ \\ &\approx \hat{\kappa}_0 + \beta' \cdot (t \cdot \mathrm{I} - \tau^{(0)})^+ - \beta' \cdot \Delta_\tau^{(0)} \cdot \mathbb{1}(t \cdot \mathrm{I} > \tau^{(0)}) \\ &= \hat{\kappa}_0 + \beta' \cdot (t \cdot \mathrm{I} - \tau^{(0)})^+ - \gamma' \cdot \mathbb{1}(t \cdot \mathrm{I} > \tau^{(0)}),\end{aligned}$$

where $\Delta_\tau^{(0)}$ is a $(k+1)\times(k+1)$-dimensional matrix with entries $(t_0 - t_0), (\tau_1 - \tau_1^{(0)}), ..., (\tau_k - \tau_k^{(0)})$ on its diagonal and $\gamma' = \beta' \cdot \Delta_\tau^{(0)}$. With this transformation, the resulting optimization problem is linear in β, γ, and $\hat{\kappa}_0$ and can be solved by maximizing the log-likelihood function

$$l(\beta, \gamma, c; \kappa_t, \tau^{(0)}) = -\frac{1}{2}\sum_{t=t_0}^{t_n}\left(\log(2\pi\sigma_t^2) + \frac{\kappa_t - \hat{\kappa}_0 - \beta' \cdot (t \cdot \mathrm{I} - \tau^{(0)})^+ + \gamma' \cdot \mathbb{1}(t \cdot \mathrm{I} > \tau^{(0)})}{\sigma_t}\right)^2. \tag{4}$$

Once the optimal parameter values $\beta^{(0)}$, $\gamma^{(0)}$, and $c^{(0)}$ are estimated, the years τ can be updated:

$$\tau_i^{(1)} = \tau_i^{(0)} + \frac{\gamma_i^{(0)}}{\beta_i^{(0)}}, \quad i = 1, ..., k.$$

Obviously, the $\beta_i^{(0)}$ need to be different from zero which is almost surely the case in practical applications (see Muggeo (2003)). The new estimate for τ can then be used for another Taylor expansion, and the estimation of β, γ, and c can be repeated. This iterative optimization stops as soon as the changes in the estimates for γ from one iteration to the next become insignificant. More precisely $|\gamma_i^{(j)} - \gamma_i^{(j-1)}| < 10^{-4}, \; i = 1, ..., k$ for a stop after j iterations. The final parameter estimates for β, γ, and $\hat{\kappa}_0$ can depend on the initial values $\tau^{(0)}$ for the Taylor expansion. Therefore, in order to minimize the risk of running into local optima, it is advantageous to use multiple starting values. Therefore, we generated 1000 sets of evenly distributed seeds. With this approach it is possible to estimate trend processes for different numbers of trend changes. Again, Different numbers of trend changes can be evaluated with information criteria. The parameters of the distribution can be updated after each estimate which led to an improvement in the likelihood. Starting with i.i.d. weights for the trend without a break, we can update the variance estimates according to a CUSUM test, i.e. the variance estimate of the optimal trend process with k trend changes can be used to estimate the variance of $k+1$ trend changes.

1. Determine initial estimates for the variances σ_t^2 of the normally distributed residuals by fitting a straight line and apply a CUSUM test to estimate constant levels of variance.

2. Determine an initial trend with one trend change assuming $N(0, \sigma_t^2)$-distributed residuals

and update the estimate of σ_t^2.

3. Determine an initial trend with two trend changes assuming $N(0, \sigma_t^2)$-distributed residuals.

4. Update the variance estimates σ_t^2 and the parameters of trend changes $\mu_{\mathcal{LN}}, \sigma_{\mathcal{LN}}^2, p$ based on the trends from (3).

5. Increase the number of trend changes by one, i.e. from N to $N+1$ for these steps.

6. Create an evenly distributed set S of starting values with $\dim S = 1000$. For each $s \in S$ proceed these steps:

   (a) Determine the optimal continuous and piecewise linear trend curve for $N+1$ trend changes based on the method proposed by Muggeo (2003) with starting values s using the optimization function under consideration (can be Equation 2, or Equation 3.

      i. Estimate an optimal trend process for $\tau^{(0)} = s$ based on $\tilde{l}(\beta, \gamma, c; \kappa_t, \tau^{(0)})$.

      ii. Update breakpoints according to

$$\tau_i^{(j)} = \tau_i^{(j-1)} + \frac{\gamma_i^{(j-1)}}{\beta_i^{(j-1)}}, \quad i = 1, ..., k$$

      iii. If $|\gamma_i^{(j)} - \gamma_i^{(j-1)}| < 0.0001, i = 1, \ldots, N+1$ assume the current trend change points are optimal; otherwise return to (i) with $\tau^{(j-1)}$ replaced by $\tau^{(j)}$.

   (b) Update the parameter estimates for $\sigma_t^2, \mu_{\mathcal{LN}}, \sigma_{\mathcal{LN}}^2, p$ based on the trend from (a) if the trend is better with respect to the optimization function than the previous trends.

7. Compare the optimal trends for N, $N-1$, $N-2$ and $N+1$ trend changes. If the value of the information criteria for $N+1$ trend changes is higher than $\max(N, N-1, N-2)$, return to (5); otherwise assume the trend process for $N-2$ trend changes to be the optimal trend process.

References

Bai, J. and Perron, P. (1998). Estimating and testing linear models with multiple structural changes. *Econometrica*, 66(1): 47–78.

Bai, J. and Perron, P. (2003). Computation and analysis of multiple structural change models. *Journal of Applied Econometrics*, 18(1): 1–22.

van Berkum, F., Antonio, K. and Vellekoop, M. (2016). The impact of multiple structural changes on mortality predictions. *Scandinavian Actuarial Journal*,2016(7): 581–603.

Booth, H., Maindonald, J. and Smith, L. (2002). Applying Lee-Carter under conditions of variable mortality decline. *Population Studies*, 56(3): 325–336.

Börger, M., Fleischer, D. and Kuksin, N. (2014). Modeling the mortality trend under modern solvency regimes. *ASTIN Bulletin*, 44(1): 1–38.

Börger, M. and Schupp, J. (2018). Modeling Trend Processes in Parametric Mortality Models. *Insurance: Mathematics and Economics*, 78, 369–380.

Burnham, K. P. and Anderson, D. R. (2002). *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach.* Springer Science & Business Media.

Burnham, K. P. and Anderson, D. R. (1999). *Understanding information criteria for seelection among capture-recapture of ring recovery models. Bird Study*, 46(1): 14–21.

Cairns, A. J., Blake, D. and Dowd, K. (2006). A Two-Factor Model for Stochastic Mortality with Parameter Uncertainty: Theory and Calibration. *Journal of Risk and Insurance*, 73(4): 687–718.

Cairns, A. J., Blake, D. and Dowd, K. (2008). Modelling and Management of Mortality Risk: A Review. *Scandinavian Actuarial Journal*, 2015(2–3): 79–113.

Chan, W.-S., Li, J. S.-H. and Li, J. (2014). The CBD mortality indexes: modeling and applications. *North American Actuarial Journal*, 18(1): 38–58.

Chen, H. and Cox, S. H. (2009). Modeling mortality with jumps: Application to mortality securitization. *Journal of Risk and Insurance*, 76(3): 727–751.

Coelho, E. and Nunes, L. C. (2011). Forecasting mortality in the event of a structural change. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 174(3): 713–736. *Statistical Modelling*, 279(4).

Currie, I. D. (2016). On fitting generalized linear and non-linear models of mortality. *Scandinavian Actuarial Journal*, 2016(4), 356–383.

Gillings, D., Makuc, D. and Siegel, E. (1981). Analysis of interrupted time series mortality trends: an example to evaluate regionalized perinatal care. *American journal of public health*, 71(1): 38–46.

Grubbs, F. E. (1950). Sample criteria for testing outlying observations. *Annals of Mathematical Statistics*, 21(1): 27–58.

Haberman, S. and Renshaw, A. (2011) A comparative study of parametric mortality projection models. *Insurance: Mathematics and Economics*, 48(1), 35–55.

Hainaut, D. (2012). Multidimensional Lee-Carter model with switching mortality processes. *Insurance: Mathematics and Economics*, 50(2): 236–246.

Human Mortality Database (2018). University of California, Berkeley (USA), and Max Planck Institute for Demographic Research (Germany). Available at www.mortality.org or www.humanmortality.de (data downloaded on November 01, 2018).

Hunt, A. and Blake, D. (2014). A general procedure for constructing mortality models. *North American Actuarial Journal*, 18(1): 116–138.

Hunt, A. and Blake, D. (2015). Modelling mortality and longevity risk management - chapter 4: Consistent mortality projections allowing for trend changes and cohort effects. *Doctoral thesis* Cass Business School.

Koissi, M.-C., Shapiro, A. F. and Högnäs, G. (2006). Evaluating and extending the Lee–Carter model for mortality forecasting: Bootstrap confidence interval. *Insurance: Mathematics and Economics*, 38(1): 1–20.

Lee, R. D. and Carter, L. R. (1992). Modeling and forecasting U.S. mortality. *Journal of the American Statistical Association*, 87(419): 659–671.

Lee, R. D. and Miller, T. (2001). Evaluating the performance of the Lee-Carter method for forecasting mortality. *Demography*, 38(4): 537–549.

Lemoine, K. (2014). Mortality regimes and longevity risk in a life annuity portfolio. *Scandinavian Actuarial Journal*, 2015(8): 689–724.

Li, J. S.-H., Chan, W.-S. and Cheung, S.-H. (2011). Structural changes in the Lee-Carter mortality indexes: detection and implications. *North American Actuarial Journal*, 15(1): 13–31.

Milidonis, A., Lin, Y. and Cox, S. H. (2011). Mortality regimes and pricing. *North American Actuarial Journal*, 15(2): 266–289.

Muggeo, V. M. (2003). Estimating regression models with unknown break-points. *Statistics in medicine*, 22(19): 3055–3071.

Njenga, C. N. and Sherris, M. (2011). Longevity risk and the econometric analysis of mortality trends and volatility. *Asia-Pacific Journal of Risk and Insurance* 5(2) 1–54.

O'Hare, C. and Li, Y. (2012). Explaining young mortality. *Insurance: Mathematics and Economics*, 50(1), 12–25.

O'Hare, C. and Li, Y. (2015). Identifying structural breaks in stochastic mortality models. *ASCE-ASME Journal of Risk and Uncertainty in Engineering Systems, Part B: Mechanical Engineering*, 1(2): 021001.

Plat, R. (2009). On stochastic mortality modeling. *Insurance: Mathematics and Economics*, 45(3): 393–404.

Ploberger, W. and Krämer, W. (1992). The CUSUM test with OLS residuals. *Econometrica*, 60(2): 271–285.

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

Richards, S., Currie, I. and Ritchie, G. (2014). A value-at-risk framework for longevity trend risk. *British Actuarial Journal*, 19(1): 116–139.

Sweeting, P. (2011). A trend-change extension of the Cairns-Blake-Dowd model. *Annals of Actuarial Science*, 5(2): 143–162.

Villegas, A. M., Kaishev, V. K. and Millossovich, P. (2015). 'StMoMo': An R package for stochastic mortality modelling.

Wilmoth, J. R. (2000). Demography of longevity: past, present, and future trends. *Experimental gerontology*, 35(9–10): 1111–1129.

Zivot, E. and Andrews, D. W. K. (1992). Further evidence on the Great Crash, the Oil-Price Shock, and the unit-root hypothesis. *Journal of Business & Economic Statistics*, 10(3): 251–270.