Modeling Trend Processes in Parametric Mortality Models

Matthias Börger

Institute for Finance and Actuarial Sciences (ifa), Ulm

Johannes Schupp (corresponding author)

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

December 2015

Keywords: longevity risk, mortality projection, parametric mortality models, mortality trend process, parameter uncertainty

Abstract

Parametric mortality models like those of Lee and Carter (1992), Cairns et al. (2006), or Plat (2009) typically include one or more time dependent parameters. Often, a random walk with drift is used to project these parameters into the future. However, longer time series of historical mortality data often show patterns which a random walk with drift is highly unlikely to generate. In fact, historical mortality trends often appear to be trend stationary around piecewise linear trends with changing slopes over time (see, e.g., Sweeting (2011) or Li et al. (2011)). Periods of lower (but rather constant) mortality improvements are followed by periods of higher improvements and vice versa.

In this paper, we propose an alternative trend process which builds on the patterns observed in the historical data. Future trend changes occur randomly over time, and also the trend change magnitude is stochastic. Furthermore, we show how the parameters of this trend process, in particular the probability of observing a trend change in a certain year and the distribution for the trend change magnitude, can be estimated from historical data. We also outline how uncertainty in the parameter estimates can be accounted for. Finally, we compare the trend process to other trend processes which have been proposed in the literature.

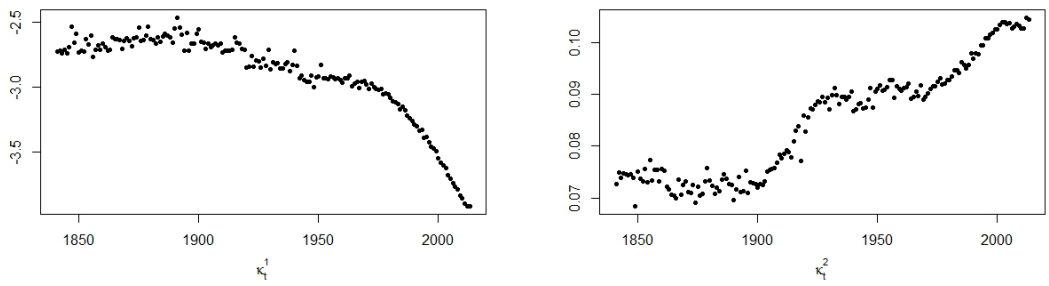

Figure 1: Time processes in the CBD model for English and Welsh males

1 Introduction

Longevity risk, i.e. the risk of insured or pensioners living longer than expected, has gained considerable attention over the last decades. The evolution of an increasingly active market for longevity risk transfers illustrates this. In order to measure longevity risk in annuity or pension portfolios, stochastic mortality models are required, and an enormous number of such models and model variants has been developed over the last decades. Most of them have a parametric structure which includes one or more time dependent parameters (time processes) to describe the evolution of mortality over time. In order to generate stochastic forecasts of future mortality, these parameters are projected into the future using stochastic processes. Obviously, it is crucial that these processes adequately project both the best estimate mortality evolution and the uncertainty of this evolution. Otherwise, risk management decisions will be based on deficient information, capital requirements will too high or too low, and longevity transactions will not be priced reasonably.

Figure 1 shows, exemplarily, the two time processes κ_t^1 and κ_t^2 in the model of Cairns et al. (2006, CBD model) for English and Welsh males.[1] As we can see from the evolution of κ_t^1, mortality has been generally decreasing over the last 173 years. The parameter κ_t^2 describes the increase of mortality with age in year t, and we can infer from its overall increase over time that mortality improvements have been stronger for younger ages than for older ages in general. For projecting $\kappa_t = (\kappa_t^1, \kappa_t^2)$ into the future, a two-dimensional random walk with drift is used in most cases, i.e.

$$\kappa_t = \kappa_{t-1} + d + CZ, \tag{1}$$

where d is a time constant drift vector, Z is a vector of standard normal innovations, and C is an upper triangular matrix with $V = CC'$ being the covariance matrix of the innovations.

The (one- or multi-dimensional) random walk has been a very popular choice for projecting the time processes in other stochastic mortality models as well. One reason for that certainly is its simplicity. In its two-dimensional form, still only five parameters need to be estimated from the historical data, i.e. the

[1] The CBD model is fitted to data from the Human Mortality Database (2015) for ages 50 to 89. For details on the model and the estimation of its parameters, we refer to Appendix A.

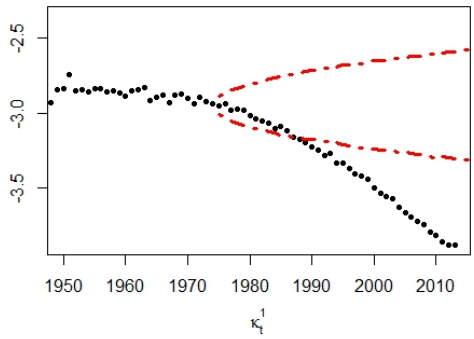

Figure 2: Back test for time process κ_t^1 in the CBD model for English and Welsh males; the time process is projected by a random walk with drift

two elements of the drift vector d and the three entries in the matrix C which determines the volatility of the innovation vector. However, the random walk's simplicity can also be problematic. Looking at data for the most recent decades only in Figure 1, the assumption of a time constant drift appears reasonable. However, looking farther into the past, the trends in the time processes seem to have changed several times. This observation can be made for basically any population.

Thus, a constant drift seems to be a reasonable assumption only for a limited period of time. When projecting the time processes into the far future, the possibility of further trend changes should be taken into account – which the random walk does not. Potential trend changes in the future imply that the confidence bounds generated by a random walk with drift might be too narrow in the long run (see, e.g., Lee and Miller (2001)). Figure 2 illustrates this by a back test in which a (one-dimensional) random walk is fitted to the κ_t^1 for English and Welsh males from 1956 to 1975 and then projected into the future.[2] In the long run, the realized κ_t^1 lie far outside even the 99% confidence interval. Börger et al. (2014) make an analogous observation for Dutch males.

Due to the random walk's structure, the width of a confidence interval it generates only depends on the volatility of the innovations. This volatility is fitted to annual random fluctuations in the historical data, and therefore, small (large) short term fluctuations automatically imply that long term trend uncertainty is also small (large). However, this is not reasonable in any case as the example of Liechtenstein and Switzerland shows. Due to Liechtenstein's population size, volatility is significantly larger than in Switzerland, and this also implies a larger parameter uncertainty in the projection. However, one would expect the long term trend uncertainty to be similar for both countries because of their very close political, economic, and social links. Thus, there is not necessarily a direct connection between short term volatility and long term trend uncertainty. Furthermore, Figure 1 suggests that annual random fluctuations are rather trend stationary around piecewise linear trends.

For the aforementioned reasons, we believe that the general adequacy of the random walk with drift for

[2]The length of the estimation period is rather arbitrary, but 20 years seems to be a usual choice.

projecting time processes in parametric mortality models is questionable, at least for long term projections. In fact, Figure 1 indicates that a trend process should have the following properties:

1. Random fluctuations are stationary around some underlying trend.
2. The underlying trend evolves continuously and is piecewise linear.
3. The slope of the underlying trend can change at random points in time and in both directions.

In this paper, we derive such a trend process. We first consider a one-dimensional version before we then discuss its generalization to a multi-dimensional version as required, e.g., for the CBD model. We also show how the parameters of the trend process can be estimated from historical data and how parameter uncertainty can be accounted for. To this end, we apply the method proposed by Muggeo (2003) to fit a continuous and piecewise linear curve to historical data. From the thus detected historical trend changes we can estimate the probability of observing a trend change in a certain year in the future as well as a distribution of its magnitude. Furthermore, we can assess the uncertainty in these estimates. Finally, we compare our trend process to other trend processes which have been proposed in the literature. Even though we mostly focus on the CBD model in the examples and applications in this paper, it is important to note that our trend process can be applied within basically any parametric mortality model. The CBD model is just a convenient choice as it is possibly the most simple of all multi-dimensional mortality models.

Changes in mortality trends and ways to account for them have already been discussed by other authors. In some sense, Booth et al. (2002) already detected trend changes when they proposed a procedure to determine the optimal estimation period for the random walk with drift in the Lee-Carter model. Their criterion searches for the longest data period for which the time process does not violate the assumption of linearity. Li et al. (2011) and Coelho and Nunes (2011) suggest procedures which are explicitly designed for the detection of trend changes within the Lee-Carter framework. They build on tests developed by Zivot and Andrews (2002) or Harvey et al. (2009) and Harris et al. (2009), respectively, which can however only detect one trend change in the time series under consideration. O'Hare and Li (2015) and van Berkum et al. (2014) address this shortcoming by applying the method of Bai and Perron (1998, 2003) which can detect and date multiple trend changes. Even though the methods for detecting historical trend changes vary, all the aforecited papers follow the same idea for projecting future mortality. The estimation periods for the parameters in the respective trend processes are restricted to the period ranging back to the most recent trend change, and the most recent (linear) trend is extrapolated into the future. Without any additional knowledge about future mortality, it seems reasonable to assume that the most recent historical trend is also the best estimate trend for the future. We also incorporate this idea in our trend process. However, the occurrence of trend changes in the past clearly indicates that trend changes may also occur in the future. Thus, assuming the same trend for the entire projection seems somewhat inconsistent with historical observations.

Therefore, several authors suggest incorporating regime switches in stochastic mortality projections. For instance, Milidonis et al. (2011) propose a Markov regime switching model with two states where the future mortality evolution is driven by a geometric Brownian motion, either with higher or lower volatility.

Lemoine (2014) extends this approach to multiple states using an autoregressive (AR) process instead of a Brownian motion. However, in both works, the trend parameters are the same in all regimes, and thus, the trends remain unchanged over time. In contrast, Hainaut (2012) allows for switches between regimes with different volatilities and trends in his multi-dimensional Lee-Carter model. Nevertheless, the range of possible trend changes is limited by the number of regimes and does not include unprecedented trend changes. Hunt and Blake (2014a) overcome this issue by drawing the magnitude of future trend changes from a Pareto distribution with the distribution's parameters fitted to historical trend changes. By drawing from this distribution, they can simulate random changes in the drifts of the random walks which they use to project the time processes in a "general procedure" model (see Hunt and Blake (2014b) for details on the "general procedure").

Richards et al. (2014) and Chan et al. (2014) propose time series processes which are more complex than the random walk. Richards et al. (2014) use an autoregressive integrated moving average (ARIMA) process of order (3,1,3) in the Lee-Carter model, while Chan et al. (2014) apply a two-dimensional vector ARIMA (VARIMA) process of order (5,1,0) in the CBD model. Both processes still include a constant drift term, but at the same time, the autoregressive terms can take up trends in the most recent data points. Börger et al. (2014) derive a time series process similar to an AR process which does not include a constant drift/trend anymore. Their process simply extrapolates the trend in the most recent data points, thus assuming that this extrapolation can serve as a best estimate for the future mortality evolution. They apply the process in an extended version of the model of Plat (2009).

Sweeting (2011) proposes a trend process which is structurally similar to our process. He also projects time processes (in the CBD model) by piecewise linear and continuous trends with stationary noise around the prevailing trends. We build on his findings and propose improvements where we see shortcomings. He derives the probability of observing a trend change in a certain year from significant historical trend changes, and thus, also the magnitude of simulated trend changes should be significant in general. However, Sweeting draws trend change magnitudes from a normal distribution with mean zero, and thus, the magnitudes will be close to zero in many cases. Furthermore, his detection of trend changes by the tests of Durbin and Watson (1950, 1951) and Chow (1960) is cumbersome and involves a certain degree of subjectivity. Therefore, we propose using the method of Muggeo (2003) in order to analyze the historical mortality evolution. Moreover, we discuss parameter uncertainty and explain how it can be accounted for.

The remainder of this paper is structured as follows: In Section 2, we introduce our trend process in a one-dimensional and a multi-dimensional setting. In particular, we explain how future trend changes are simulated. Parameter estimation is then discussed in Section 3 which includes the detection of historical trend changes in particular. In Section 4, we address uncertainties in the parameter estimation and explain how the uncertainties in the most relevant parameters can be accounted for. We then compare our trend process to alternative trend processes in Section 5. Finally, Section 6 concludes.

2 A Piecewise Linear Trend Process

In this section, we introduce a new trend process which complies with the three requirements set out in the Introduction. The underlying trend is continuous and piecewise linear with random changes in its

slope and stationary random fluctuations around the prevailing trend. We commence in a one-dimensional setting and then discuss how the trend process can be generalized to a multi-dimensional setting.

Let κ_t be an observable time process in a parametric mortality model (like the processes in Figure 1) and let $\tilde{\kappa}_t$ be the "true" underlying time process net of random fluctuations. Moreover, let d_t be the linear trend, i.e. the slope, of the process $\tilde{\kappa}_t$ at time t. Then the projection of κ_t is carried out in three steps:

1. Determine whether a trend change occurs between $t-1$ and t. If so, the trend d_{t-1} changes by λ_t; if not, the trend remains unchanged:

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

2. Derive the value of the underlying time process at time t:

$$\tilde{\kappa}_t = \tilde{\kappa}_{t-1} + d_t$$

3. Add some (normally distributed) random noise ϵ_t to obtain the value of the observable time process:

$$\kappa_t = \tilde{\kappa}_t + \epsilon_t \tag{2}$$

In Step 1, a trend change occurs with probability p which can be estimated from the frequency of trend changes in the historical data. Since we regard the current trend as the best information available on the future mortality trend, the trend change intensity λ_t should be specified such that the current trend is an unbiased estimator for the trend at any point in time in the future. This can be achieved most easily by requiring that positive and negative trend changes are equally likely and of the same magnitude. Therefore, we model the trend change intensity as the product of the absolute magnitude of the trend change M_t and its sign S_t, i.e. $\lambda_t = S_t \cdot M_t$, and we require that S_t assumes both values -1 and 1 with probability 0.5. For M_t, we propose a lognormal distribution as it generates positive values with probability 1 and has only little probability mass close to zero. Thus, the simulated trend change magnitudes are material in general. The parameters of the lognormal distribution, i.e. μ_M and σ_M^2, can be estimated from the absolute magnitudes of historical trend changes. The use of a lognormal distribution is also supported by Figure 3 which compares the pdf of the standard lognormal distribution with a histogram of all standardized historical trend change magnitudes in the CBD time processes for males and females in 15 industrialized countries with sufficiently long data history.[3] Standardization of the historical trend change magnitudes has been carried out for each population and each time process separately to ensure comparability. In the figure, we see that the lognormal pdf and the histogramm show similar patterns. In particular, the lognormal distribution with its heavy tail should be able to generate some rare but very large trend change magnitudes like they have occurred in the past. As an alternative to the lognormal distribution, Hunt and Blake (2014a) apply a Pareto distribution. However, this requires specifying a

[3]The countries under consideration are: Australia, Belgium, Canada, Denmark, England & Wales, Finland, France, Italy, the Netherlands, New Zealand, Norway, Spain, Sweden, Switzerland, and the United States. For more details on the data used, we refer to Section 3.

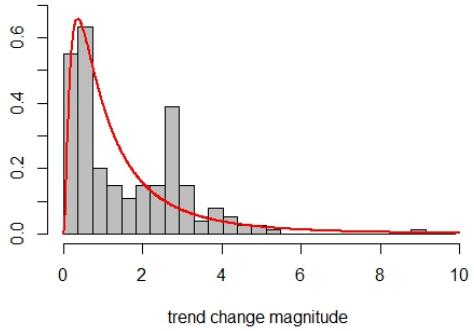

Figure 3: Standardized historical trend change magnitudes in the CBD time processes for 15 countries and pdf of the standard lognormal distribution

minimum magnitude for future trend changes which we find very difficult given the usually rather small number of observed historical trend changes. Sweeting (2011), on the other hand, does not separate the absolute trend change magnitude from its sign, but uses a Normal distribution with mean zero for the trend change intensity. As outlined in the Introduction, this approach is inconsistent with estimating the trend change probability p from significant historical trend changes only. Of course, p could be specified such that it also allows for trend changes with magnitudes very close to zero. However, this would imply modeling immaterial and unobservable trend changes which is counterintuitive.

The innovations ϵ_t in Equation (2) are assumed to be serially uncorrelated and normally distributed with mean zero and variance σ_ϵ^2. The variance can be estimated from historical residuals, and we assume that it remains constant over time. Of course, one could enhance the trend process further by also allowing the variance to change over time as it has done in the past (see Figure 1). However, that would add unnecessary complexity to our trend process since the innovations only have a limited effect on projection outcomes. In contrast to the random walk with drift, the innovations ϵ_t do not determine changes in the level of mortality but are simply random noise. Furthermore, Equation (2) could easily be extended by a transitory jump component if one wanted to take into account potential catastrophies or pandemics.

The actual projection of the trend process κ_t can obviously be carried out by iteratively applying the three step approach from above. Alternatively, one can first draw the number of years until the next trend change occurs from a geometric distribution with parameter p and then simulate the trend and the innovations up to that point in time.

In a multi-dimensional setting, dependencies between the different time processes, i.e. the components of the vector κ_t, need to be accounted for. For instance, Sweeting (2011) finds that trend changes in the two CBD time processes occur at once noticeably often, and he therefore derives trend change probabilities for simultaneous and individual trend changes. Of course, also dependencies between the trend change intensities should be accounted for in case of simultaneous trend changes. For the 30 populations we considered, we found six potentially simultaneous trend changes in κ_t^1 and κ_t^2, i.e. trend changes which

occurred in the same or subsequent years. Since one would already expect about nine such trend changes under the assumption of independence between κ_t^1 and κ_t^2, this is not a significant number. Therefore, we conclude that, at least in the CBD model, trend changes in the two time processes can be projected independently of each other. However, we typically found significant correlation between the historical innovations ϵ_t^1 and ϵ_t^2, and therefore, the innovations should be drawn from a multinormal distribution with corresponding correlation matrix Σ_ϵ^2.

3 Calibration of Trend Process Parameters

In the previous section, we introduced a trend process with piecewise linear trends and random fluctuations around the prevailing trend. In order to project future mortality by this process, the following set of parameters needs to be specified, where t_n denotes the last year of the available historical data:

- the value of $\tilde{\kappa}_{t_n}$ as the starting point of the projection,
- the current trend d_{t_n},
- the probability p of observing a trend change in a certain year in the future,
- the parameters μ_M and σ_M^2 of the lognormal distribution for the trend change magnitude,
- the covariance matrix Σ_ϵ^2 of the annual innovations ϵ_t.

All these parameters can be one- or multi-dimensional, depending on the number of time processes in the mortality model under consideration. In order to ease notation, we only discuss parameter estimation in detail for the one-dimensional case. In the multi-dimensional case, the same estimation approach can be applied to each time process individually since we assume that trend changes in the different time processes occur independently of each other. The only exception is the estimation of the covariance matrix Σ_ϵ^2 which needs to be carried out for all time processes simultaneously.

All parameters can be estimated from historical data. To this end, we propose fitting a piecewise linear and continuous trend curve to the historical time process κ_t. Once this is done, the parameters can be determined as follows:

- Set $\tilde{\kappa}_{t_n}$ to the value of the fitted trend curve in the final year of the historical data.
- Set d_{t_n} to the most recent trend observed in the historical data.
- Set p equal to the ratio of the number of observed trend changes and the length of the historical data period. This is the maximum likelihood estimate for p.
- Set μ_M and σ_M^2 to their maximum likelihood estimates based on the observed trend change magnitudes.
- Estimate Σ_ϵ^2 from the residuals $\kappa_t - \tilde{\kappa}_t$. In case the residuals are heteroscedastic, only a shortened set of most recent residuals should be used (more details below).

The crucial step in the calibration of our trend process clearly is the fitting of a trend curve to the historical κ_t process. This can be done in a robust and fast way by the method of Muggeo (2003). The concept behind this method is to first find, for any reasonable but fixed number of trend changes, the optimal trend curve. Thus, the trend changes are positioned and the trend change magnitudes determined such that the resulting trend curve fits the data as good as possible. This fitting is done by maximizing the likelihood function under the assumption of normally distributed residuals, i.e. $\epsilon_t \sim \mathcal{N}(0, \sigma_t^2)$. In a second step, the optimal trend curves for different numbers of trend changes are compared by some information criterion. Each additional trend change increases the number of parameters in the trend curve by two, i.e. the position of the additional trend change and the slope of the subsequent linear trend. The overall optimal trend curve is the one which minimizes the information criterion. For more details on the method of Muggeo (2003), we refer to Appendix B.

In the maximum likelihood estimation of the trend curves, heteroscedasticity can be allowed for. This is important in our case as we see in Figure 1 that annual fluctuations have decreased considerably over time. However, the variances σ_t^2 cannot be estimated as part of the method of Muggeo (2003), and we therefore propose using an iterative approach. Starting with some initial variance estimates, a preliminary optimal trend curve can be determined. Subsequently, the variance estimates can be updated based on the obtained residuals, and the trend curve fitting can be repeated. For the initial variance estimates, we follow the approach of Sweeting (2011), i.e. for each data point, we fit a regression line to this and the six closest data points and estimate raw variances from the seven residuals. A CUSUM test applied to these raw variance estimates then indicates when variances changed in the past and how large they were during which time period.[4] In particular, the CUSUM test eliminates random fluctuations in the raw variance estimates. In subsequent iterations, variance estimates can be derived in the same way from the residuals of the previous trend curve fitting. In summary, the optimal trend curve for a fixed number of trend changes can be determined as follows:

1. Estimate initial variances from regression lines and graduate them by applying the CUSUM test.
2. Determine the optimal continuous and piecewise linear trend curve by the method of Muggeo (2003) given the variance estimates from 1.
3. Update the variance estimates based on the residuals from 2. and another CUSUM test.
4. Update the trend curve based on the updated variances from 3.
5. Determine whether the trend curve from 4. differs significantly from that in the previous iteration in terms of number and positions of the trend changes as well as slopes of the linear trends. If so, return to 3.; otherwise the trend curve optimization is accomplished.

We applied this iterative approach to the CBD time processes for males and females aged 50 to 89 in the following 15 countries: Australia, Belgium, Canada, Denmark, England & Wales, Finland, France, Italy, the Netherlands, New Zealand, Norway, Spain, Sweden, Switzerland, and the United States. With

[4]For more details on the CUSUM test, we refer to Appendix C.

one exception, these are the countries for which reliable data is available in the Human Mortality Database (2015) for at least 75 years.[5] The only country with such a long data history which we did not consider is Iceland for which, due to the small population size, data is too noisy to determine trend changes. The requirement of 75 years of data is rather arbitrary. In fact, there is no minimum data history required, but parameter estimates are obviously more stable the more historical trend changes can be observed. We found that the iterative trend curve optimization converges very fast. For every time process we considered, 3 iterations were sufficient to obtain reliable parameter estimates. For the time processes in other mortality models, we would expect the trend curve optimization to converge similarly fast.

Exemplarily, Figure 4 shows the historic κ_t^1 and κ_t^2 processes in the CBD model for English and Welsh males as well as the optimal trend curves according to the Akaike Information Criterion (AIC), the Bayesian Information Criterion (BIC), and the Modified BIC (MBIC). These criteria are defined as

$$\begin{aligned} \text{AIC} &= -2\ln(\hat{L}) + 2 \cdot K, \\ \text{BIC} &= -2\ln(\hat{L}) + \ln(n) \cdot K, \\ \text{MBIC} &= -2\ln(\hat{L}) + \ln(n) \cdot \ln(\ln(n)) \cdot K, \end{aligned}$$

where $\hat{L}$ is the maximized likelihood, n the number of data points, and K the number of parameters in a fitted model. We consider AIC and BIC because they are well known and widely applied. The MBIC was proposed by Wang et al. (2009) (in a generalized form), and we use the variant which Muggeo and Adelfio (2011) found to perform best within their trend estimation method. It only differs from the BIC in that the penalty term is multiplied by $\ln(\ln(n))$. Of course, other information criteria like (generalized) cross validation or Takeuchi's Information Criterion could be used as well.[6] In both panels of Figure 4, we observe that BIC and MBIC yield plausible and very similar trend curves, and we made the same observation for almost every time process and every population we considered. The AIC, on the other hand, detects more trend changes for both time processes as additional parameters are penalized less strongly in the AIC compared to the BIC and MBIC. These additional trend changes even occur in subsequent years, i.e. in 1919 and 1920 for κ_t^1 and in 1861 and 1863 for κ_t^2. This clearly is an attempt to incorporate jumps, i.e. sudden and permanent changes in the level of mortality, and indicates that the AIC is not sufficiently robust with respect to short term effects in the historical data. Following the suggestion of Muggeo and Adelfio (2011), we therefore use the MBIC in all applications presented in the remainder of this paper.

The number of historical trend changes is rather small in general. For some time processes like the κ_t^1 process in the CBD model for English and Welsh males, we find only three trend changes. This implies considerable uncertainty in the parameter estimation. In the following section, we will discuss how this uncertainty can be assessed and accounted for based on the data which is available for the time process under consideration. Alternatively, trend changes from several similar populations could be combined to reduce parameter uncertainty. For instance, when projecting mortality for Sweden, also historical trend

[5] For each population, we omitted data which is marked as unreliable on the HMD website, e.g. data for years 1751 to 1860 for Sweden.

[6] A discussion of different information criteria and their application can be found, e.g., in Burnham and Anderson (2002).

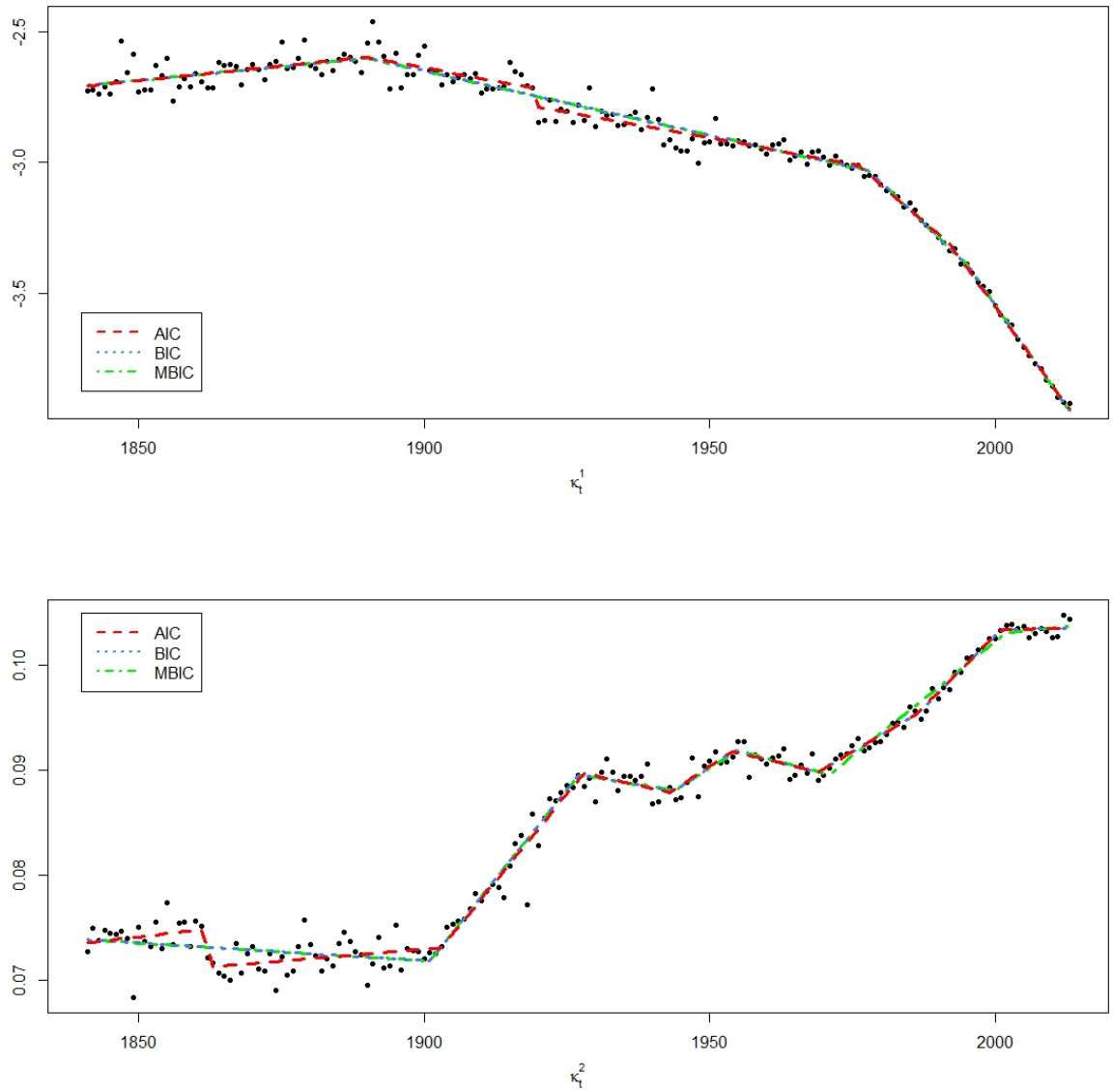

Figure 4: Time processes in the CBD model for English and Welsh males and optimal trend curves according to different information criteria

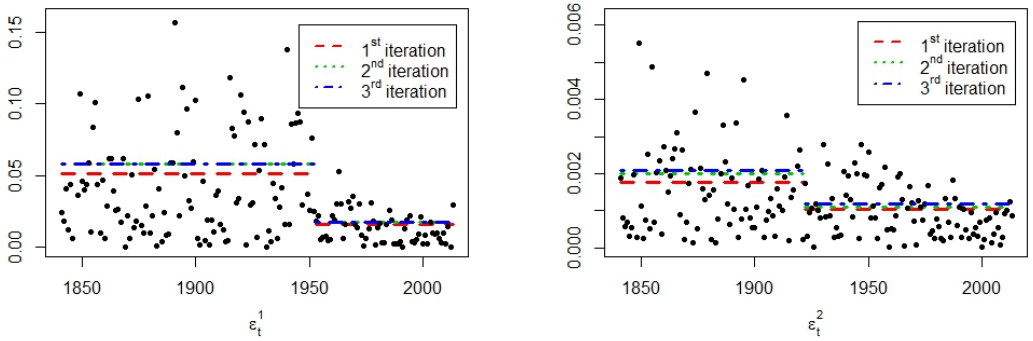

Figure 5: Absolute values of residuals from the trend curve optimization and their standard deviation estimates in subsequent iterations, all for the time processes in the CBD model for English and Welsh males

changes for, e.g., Finland and Norway could be considered in the parameter estimation. As usual in the context of multi-population mortality modeling, one would assume that data from closely related populations can provide additional information and thus enhance the modeling for the population one is particularly interested in (see, e.g., Jarner and Kryger (2011) or Li and Lee (2005)). This approach can be particularly helpful for small populations or populations with limited data history.

Finally, Figure 5 shows, for the time processes in the CBD model for English and Welsh males, the absolute values of the final residuals as well as the estimated standard deviations in the three iterations we carried out. While there is a visible change in the standard deviation (and thus also variance) estimates from the first to the second iteration, they change only very slightly from the second to the third iteration already. This underlines the fast convergence of the iterative trend curve optimization. Moreover, the CUSUM test indicates that, for both time processes, the variances changed only once during the observation period. In fact, these changes were already detected in the first iteration. The information provided by the CUSUM test can also be used to determine the optimal estimation period for the covariance matrix Σ_ϵ^2. In case κ_t is one-dimensional, Σ_ϵ^2 can simply be set equal to the most recent historical variance estimate. In case κ_t is multi-dimensional as in the CBD model, Σ_ϵ^2 can be estimated from the residuals starting in the year in which any of the variance estimates for the different time processes changed for the last time.

4 Parameter Uncertainty

The parameter estimation as proposed in the previous section usually involves a considerable amount of uncertainty. Therefore, in this section, we explain how this uncertainty can be assessed and taken into account when projecting mortality rates. We distinguish between two sets of parameters: the "starting values" for a projection, i.e. $\tilde{\kappa}_{t_n}$ and d_{t_n}, and the "trend change parameters", i.e. the probability p of observing a trend change in a certain year and the parameters μ_M and σ_M^2 of the lognormal distribution for the trend change magnitude. In comparison, the parameter uncertainty in the covariance matrix Σ_ϵ^2 is

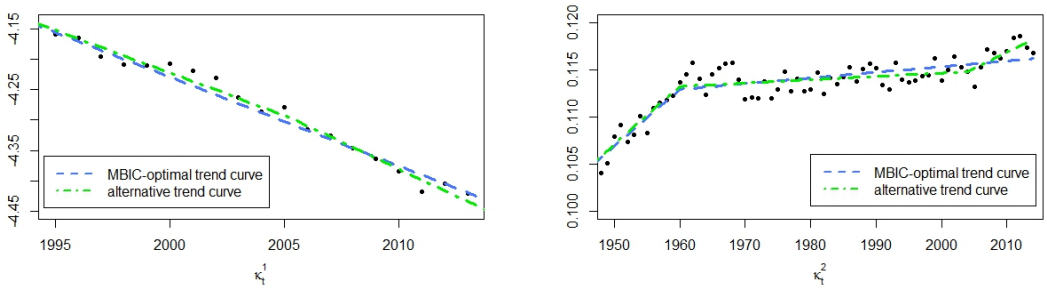

Figure 6: MBIC-optimal and alternative trend curves for the time processes in the CBD model for Swedish females

usually very small. Moreover, the impact of the additional uncertainty on projected mortality rates would only be marginal since Σ_ϵ^2 only drives the random fluctuations around the prevailing trend. For simplicity, we therefore neglect the uncertainty in the covariance matrix.

4.1 Parameter Uncertainty in the Starting Values

Figure 6 illustrates the uncertainty which we need to account for in the starting values $\tilde{\kappa}_{t_n}$ and d_{t_n}. The figure shows, for recent years, the time processes in the CBD model for Swedish females as well as the MBIC-optimal trend curves and alternative trend curves which also have high likelihoods.[7] We see that there is a certain chance that an additional trend change might have occurred in the final years of the data set which the parameter estimation did not detect due to the limited data available after this potential trend change. For κ_t^1, the potential impact of such an additional trend change seems rather small since the trend would have changed only slightly. For κ_t^2, however, such an additional trend change could have a massive impact on $\tilde{\kappa}_{t_n}^2$ and, in particular, $d_{t_n}^2$. Thus, the main uncertainty with respect to the starting values stems from potentially undetected trend changes in most recent years. In what follows, we explain how this uncertainty can be assessed and taken into account. Again, we only do this in detail for the case of one time process, i.e. if κ_t is one-dimensional. In the multi-dimensional case, we again propose assuming independence between the time processes. Thus, the same approach can be applied to each time process, i.e. each component of the vector κ_t, individually.

In order to quantify the uncertainty in the starting values, we consider alternative starting values which would be obtained if there was an additional trend change in most recent years. More precisely, we assume that the trend changes which have been detected in the parameter estimation in the previous section are fixed and that an additional trend change might have occurred in any year after the most recent of the originally detected trend changes. For each of those years for the additional trend change, we obtain a new optimal trend curve with different values for $\tilde{\kappa}_{t_n}$ and d_{t_n}. We denote these values by $\tilde{\kappa}_{t_n,s}$ and $d_{t_n,s}$, where $s = 1, ..., N-1$ refers to the different years in which the additional trend change can occur, and

[7] We use the example of Swedish females instead of English and Welsh males here as potential effects of allowing for parameter uncertainty become more obvious for this population. As we will see in Section 5, parameter uncertainty is rather small for English and Welsh males.

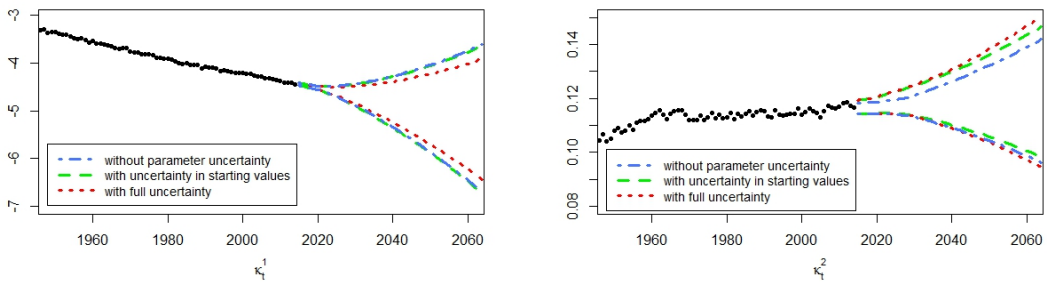

Figure 7: 90% confidence intervals with and without parameter uncertainty for projections of the time processes in the CBD model for Swedish females

N is the number of years of available data after the most recent of the originally detected trend changes. Moreover, we denote by $\tilde{\kappa}_{t_n,0}$ and $d_{t_n,0}$ the starting values from the original parameter estimation. All these potential starting values $\tilde{\kappa}_{t_n,s}$ and $d_{t_n,s}$, $s = 0, ..., N-1$ build up a combined empirical distribution for $\tilde{\kappa}_{t_n}$ and d_{t_n} from which starting values can then be drawn randomly for each path in a stochastic projection. In order to assign probabilities to the $\tilde{\kappa}_{t_n,s}$ and $d_{t_n,s}$, we propose adopting the concept of Bayesian weights as introduced by Burnham and Anderson (2002):

$$\mathbb{P}(\tilde{\kappa}_{t_n,s}, d_{t_n,s}) = \frac{\exp(-\frac{1}{2}(\mathrm{IC}_s - \mathrm{IC}_0))}{\sum_{i=0}^{N-1} \exp(-\frac{1}{2}(\mathrm{IC}_i - \mathrm{IC}_0))},$$

where IC_0 denotes the value of the information criterion under consideration for the originally fitted trend curve, and IC_s is the corresponding value for the case of an additional trend change in year s.

The effect of allowing for uncertainty in the starting values $\tilde{\kappa}_{t_n}$ and d_{t_n} is illustrated in Figure 7. It shows 90% confidence intervals for the time processes in the CBD model for Swedish females, with and without allowance for parameter uncertainty in the starting values. For κ_t^1, the confidence intervals are very similar in terms of both position and width which means that parameter uncertainty in $\tilde{\kappa}_{t_n}^1$ and $d_{t_n}^1$ is rather small. This coincides with our observations in Figure 6. For κ_t^2, on the other hand, the confidence intervals differ significantly. The position of the confidence interval with allowance for parameter uncertainty indicates that there is a decent chance that an additional upward trend change occurred after the most recent detected trend change. In fact, the probability for such an additional trend change is about 28%. Furthermore, the uncertainty with respect to undetected trend changes also affects the width of the confidence interval. As expected, long term uncertainty in κ_t^2 increases when we allow for the fact that we do not know the current mortality trend for sure. Clearly, allowing for uncertainty in the starting values $\tilde{\kappa}_{t_n}$ and d_{t_n} is important when modeling longevity risk.

4.2 Uncertainty in the Trend Change Parameters

Uncertainty in the parameters p, μ_M, and σ_M^2 can also be material since the number of historical trend changes is typically rather small. We propose accounting for this uncertainty in a similar fashion as for the starting values, i.e. we again build up a combined empirical distribution from which parameter values can be drawn randomly for each projection path. As in the previous subsection, we explain the approach in detail for one time process and propose applying the same approach to each time process individually in a multi-dimensional setting.

A combined empirical distribution for the trend change parameters p, μ_M, and σ_M^2 can be derived by varying the number of trend changes k which are to be detected in the historical data. For each (reasonable) number of trend changes, the method of Muggeo (2003) can provide an optimal trend curve in terms of likelihood from which potential trend change parameters p_k, $\mu_{M,k}$, and $\sigma_{M,k}^2$ can be estimated.[8] The empirical distribution build up by these trend change parameter sets for different values of k represents the uncertainty that the number of historical trend changes and thus also their magnitudes might be different from what was detected in the original parameter estimation. In order to assign probabilities to the different p_k, $\mu_{M,k}$, and $\sigma_{M,k}^2$, again the concept of Bayesian weights can be applied:

$$\mathbb{P}(p_k, \mu_{M,k}, \sigma_{M,k}^2) = \frac{\exp(-\frac{1}{2}(\mathrm{IC}_k - \mathrm{IC}_{\hat{k}}))}{\sum_{i=1}^{n} \exp(-\frac{1}{2}(\mathrm{IC}_i - \mathrm{IC}_{\hat{k}}))},$$

where $\hat{k}$ is the optimal number of trend changes. In practical applications, obviously, only those numbers of trend changes k need to be taken into account for which the Bayesian weights are at least slightly different from zero. In fact, for any time process we considered, we only had to compute weights for the optimal number of trend changes plus/minus three trend changes at most.

Uncertainty also arises from the fact that, due to the random fluctuations in the observable time process κ_t, the optimal trend curve can deviate from the "true" historical trend evolution even if the number of historical trend changes is detected correctly. However, in comparison, we regard this uncertainty as rather small particularly because it does not affect the estimated trend change probability p at all. Therefore, we neglect this source of uncertainty for simplicity.

Figure 7 also displays confidence intervals for the case of parameter uncertainty in both the starting values and the trend change parameters (full parameter uncertainty). For κ_t^2, the confidence interval slightly widens compared to the case of parameter uncertainty in the starting values only which is as one would expect. For κ_t^1, on the other hand, we observe that the confidence interval with full parameter uncertainty is more narrow than the confidence interval for the case of parameter uncertainty in the starting values only. This may be surprising at first sight, but simply indicates that $p_{\hat{k}}$, $\mu_{M,\hat{k}}$, and $\sigma_{M,\hat{k}}^2$ from the original parameter estimation in Section 3 might be too large. In fact, there is a 25% chance that the "true" number of trend changes is only three instead of the detected four. Thus, when allowing for parameter uncertainty, the trend change probability is only $p = \frac{3}{155}$ instead of $p = \frac{4}{155}$ in every fourth simulation path. Again, we find that parameter uncertainty can be material and should be accounting for.

[8] In fact, the method of Muggeo (2003) provides these trend curves automatically as part of the optimization process (see Section 3 and Appendix B).

5 Comparison with other Trend Processes

In this section, we compare our trend process to other trend processes which have been proposed in the literature. To this end, we project remaining period life expectancies for 60-year old males in England and Wales with the CBD model. The model has been fitted to data for ages 50 to 89 from 1841 to 2013, and we have found 3 and 6 trend changes in the historical time processes κ_t^1 and κ_t^2, respectively. The alternative trend processes which we consider are:

- the random walk with drift
- the VARIMA(5,1,0) process proposed by Chan et al. (2014)
- the random walk with variable drifts as introduced by Hunt and Blake (2014a)
- the piecewise linear trend process proposed by Sweeting (2011)

Thus, we have two time series processes with fixed drifts and two processes which allow for changes in their (linear) trends.

5.1 Comparison with Time Series Processes

For the (two-dimensional) random walk with drift (see Equation (1)), we consider fits to two different data sets: data for the most recent 20 years which seems to be a typical estimation period, and data after 2002 since 2002 is the year where the trend for any of the time processes changed for the last time. Thus, in the second fit, we use the trend change analysis to determine a kind of optimal estimation period for the random walk, i.e. the longest possible estimation period for which the assumption of a constant drift can be justified.[9]

The VARIMA(5,1,0) process has already been used by Chan et al. (2014) to project future mortality for English and Welsh males with the CBD model. It was found to be the best choice amongst all VARIMA(p,d,q) processes for that population. The increments in $\kappa_t = (\kappa_t^1, \kappa_t^2)$ are projected as

$$\Delta\kappa_t = \kappa_t - \kappa_{t-1} = C_0 + \sum_{i=1}^{5} \Phi_i \Delta\kappa_{t-i} + \epsilon_t,$$

where C_0 is a two-dimensional constant drift vector and the Φ_i are matrices of dimension 2×2. In order to estimate these parameters, we apply the approach proposed by Chan et al. (2014) to two different data sets: data after 1950 (as in Chan et al. (2014)) and the full data set.

Figure 8 shows 90% confidence intervals for the remaining period life expectancies based on projections with the different trend processes. First of all, we observe that the confidence intervals differ considerably in terms of both position and width. Thus, model risk with respect to the trend process is highly significant. We also find that parameter uncertainty in our trend process is rather small for the case of English and

[9] Note that the drift can still differ significantly from the most recent detected trend since, in standard time series estimation, the drift is only fitted to the first and the last data point in the estimation period and thus prone to random effects in these two data points.

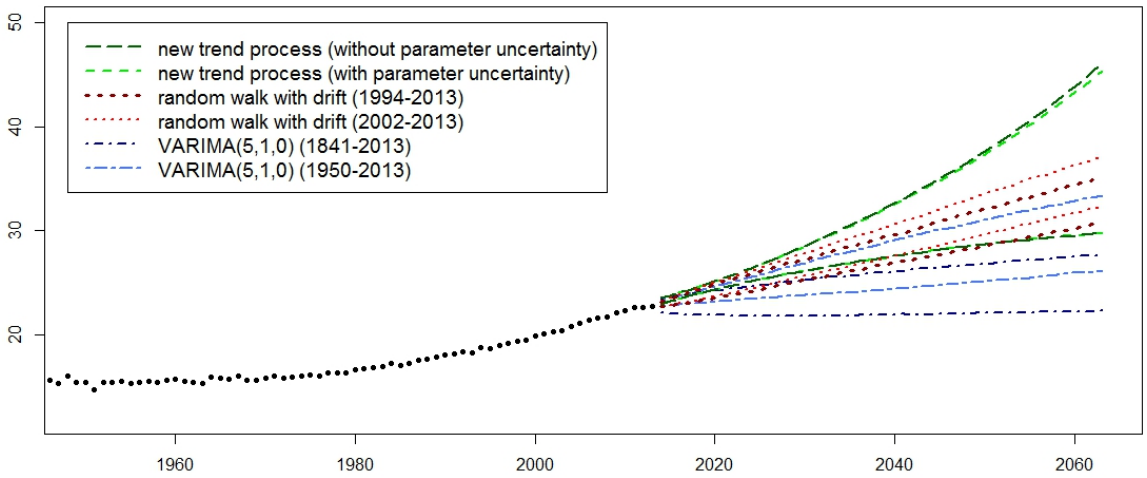

Figure 8: 90% confidence intervals for projected remaining period life expectancies of 60-year old English and Welsh males; comparison between new trend process and time series processes

Welsh males and that the projections based on the time series processes strongly depend on the estimation periods. The 90% confidence interval for one VARIMA process variant does not even include the other variant's central projection. Moreover, the VARIMA processes project life expectancies which appear rather small given the most recent increase in historical life expectancies. This observation particularly holds for the variant which is fitted to the full data set. The reason for this is the fixed drift which corresponds to the average change in the time processes over the estimation period. For both the VARIMA process and the random walk, the long term confidence intervals are rather narrow in the long run which has been criticized by, amongst others, Lee and Miller (2001) and is again mainly due to the fixed drift. Allowing for parameter uncertainty would widen the confidence intervals slightly, but the main issue of a fixed drift would remain. The newly introduced trend process overcomes this issue, and the projected uncertainty appears plausible.

5.2 Comparison with other Piecewise Linear Trend Processes

Next, we compare our trend process to the two alternative trend processes with variable linear drifts/trends. Hunt and Blake (2014a) use random walks with changing drifts in their "general procedure" model, and we transfer their approach to the time processes in the CBD model. More precisely, we consider a two-dimensional random walk where the drifts in both components can change independently of each other. Hunt and Blake (2014a) estimate the drift change probabilities in the same way as we estimate the trend change probabilities, i.e. as the ratios of the numbers of historical drift changes and the length of the data period. Drift changes can also be positive and negative with probability 0.5, but their magnitudes

are drawn from Pareto distributions instead of lognormal distributions. As in Hunt and Blake (2014a), we apply the method of Bai and Perron (1998, 2003) to data after 1950 in order to detect historical drift changes. Obviously, disregarding more than 100 years of available data is critical given the rare occurrence of drift/trend changes. In fact, Hunt and Blake (2014a) found two drift changes at most for any of the time processes in their model, and we also found only 1 drift change for κ_t^1 and no drift changes for κ_t^2. When we tried to extend the estimation period farther into the past, the method of Bai and Perron (1998, 2003) failed due to the large volatility in the time processes in the earlier years. The lack of historical drift changes implies that κ_t^2 is again projected as a standard random walk in the basic model variant. However, we also consider another variant which includes parameter uncertainty. Following Hunt and Blake (2014a), we apply the bootstrapping approach of Koissi et al. (2006) which means that the CBD model is refitted and the drift change analysis repeated in each simulation path. In some of these paths, a drift change can be detected for κ_t^2, and then, also κ_t^2 is projected as a random walk with variable drift. However, compared to our approach to parameter uncertainty, the approach of Hunt and Blake (2014a) is rather time consuming due to the full model re-estimation in each simulation path. Note also that the general issue of a random walk that annual random fluctuations drive long term uncertainty remains in the setting of Hunt and Blake (2014a), even though a second source of uncertainty is added in the form of drift changes.

The main difference between our trend process and that of Sweeting (2011) is the distribution for the trend change magnitude. Sweeting (2011) uses a normal distribution with mean zero, and he determines the standard deviation of this distribution as the root mean square of historical trend change magnitudes. As outlined in the Introduction, this implies that the magnitude of many projected trend changes will be close to zero which is inconsistent with the annual trend change probability being fitted to significant historical trend changes only. For simplicity and comparison, we fit Sweeting's trend process to the historical trend changes which we have detected by the method of Muggeo (2003).

Figure 9 again shows 90% confidence intervals for projected period life expectancies for 60-year old English and Welsh males. The confidence interval for Sweeting's trend process appears unrealistically wide which is particularly due to the distribution for the trend change magnitude. Setting the distribution's standard deviation equal to the root mean square of (significant) historical trend changes implies that, on average, projected trend changes are similar in size to the historical trend changes which is as desired. However, since the distribution has significant mass around zero, some projected trend changes have to be considerably stronger than the historical ones. These rather extreme trend changes lead to unrealistically wide confidence intervals. The confidence intervals for the random walks with variable drifts appear more plausible in the long run, both with and without allowance for parameter uncertainty. However, they are very wide in the short term which is also a typical feature of the random walk. In comparison, we again conclude that our trend process provides highly plausible mortality projections.

6 Conclusion

In order to assess longevity risk, stochastic mortality projections are required, and over the last decades, a variety of models has been proposed for this purpose. These models typically contain one or more time

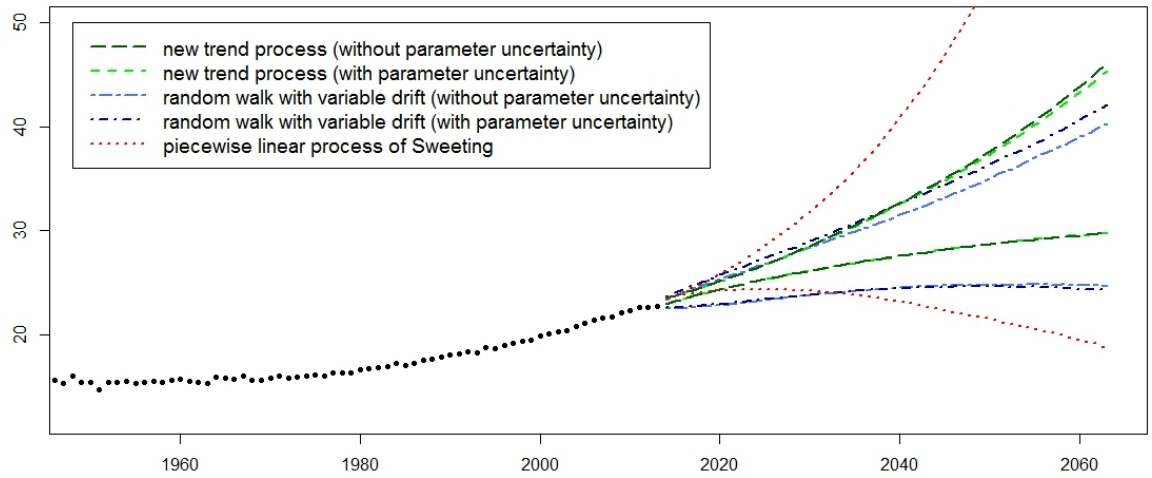

Figure 9: 90% confidence intervals for projected remaining period life expectancies of 60-year old English and Welsh males; comparison between new trend process and alternative processes with variable drifts/trends

dependent parameters which need to be projected into the future by stochastic processes. Often, a random walk with drift is used which has some shortcomings. Most prominently, the long term confidence intervals seem too narrow due to the fixed drift. In fact, historical time processes indicate that the mortality trend for basically any population changed occasionally. Therefore, it seems more reasonable to project mortality by a trend process which evolves piecewise linear with random changes in its slope.

We introduce such a trend process in which the projected linear trend can change every year with a certain probability. If a trend change occurs, its sign and its absolute magnitude are random. In fact, for the absolute trend change magnitude, we propose a lognormal distribution as it is continuous, has no probability mass at zero, and fat tails. Annual random fluctuations in mortality are modeled as gaussian noise around the prevailing trend. We first introduce a one-dimensional version of this trend process and then explain how it can be generalized to multiple dimensions.

The parameters of the trend process can be estimated from continuous and piecewise linear trend curves which are fitted to historical time processes. Such trend curves can be derived in a robust and fast way by the method of Muggeo (2003). They provide estimates for the current level and trend of the time process, i.e. the starting values for a stochastic projection, as well as historical trend changes which can be used to calibrate the trend change probability and the lognormal distribution for the trend change magnitude. In many cases, however, parameter uncertainty is material due to the limited amount of available historical data. This uncertainty can be assessed and accounted for efficiently by setting up empirical distributions for the trend process' most relevant parameters. From these distributions, parameter values can be drawn randomly for each simulation path. Alternatively, historical trend changes from several populations can

be combined to increase the data basis for parameter estimation. We find that allowing for parameter uncertainty can significantly affect both the width and the position of confidence intervals. Interestingly, taking parameter uncertainty into account can reduce the uncertainty in projected mortality rates.

Finally, we compare our trend process to alternative trend processes which have been proposed in the literature. We discuss advantages and disadvantages of the trend processes and find that projected life expectancies vary significantly in terms of both central projection and projected uncertainty. Thus, model risk with respect to the trend process is significant. Life expectancies projected by our trend process appear highly plausible in comparison to projections based on the other trend processes. We therefore conclude that our trend process constitutes a valuable alternative to existing trend processes.

Appendix

A The Cairns-Blake-Dowd Model

The model of Cairns et al. (2006) relies on the assumption that, for each year t, the logit of the probabilities of death follows a straight line with age, i.e.

$$\text{logit}\,(q_{x,t}) = \log\left(\frac{q_{x,t}}{1-q_{x,t}}\right) = \kappa_t^1 + (x-\bar{x})\cdot\kappa_t^2, \tag{3}$$

where $\bar{x}$ is the average age in the age range under consideration. Thus, the mortality rates in each year are determined by only two parameters. The time process κ_t^1 describes the evolution of the general level of mortality rates over time, while the time process κ_t^2 identifies changes in the slope of the mortality curve over time and thus differences in the mortality evolutions for different ages groups. The assumption of linearity in the logit of the probabilities of death implies that the model can only be applied to older ages in general as young age effects like the accident hump can not be accounted for.

The parameters in the CBD model are typically estimated using a maximum likelihood approach. Given exposures $E_{x,t}$ and (central) mortality rates $m_{x,t}$, the number of deaths $D_{x,t}$ is assumed to follow a Poisson distribution, i.e.

$$D_{x,t} \sim Poi(E_{x,t}\cdot m_{x,t}).$$

Thus, for each year t, one needs to find κ_t^1 and κ_t^2 which maximize the log-likelihood function

$$l_t(\kappa_t^1,\kappa_t^2;E_{x,t},d_{x,t}) = \sum_x \left(d_{x,t}\cdot\log\left(E_{x,t}\cdot m_{x,t}\left(\kappa_t^1,\kappa_t^2\right)\right) - \log(d_{x,t}!) - E_{x,t}\cdot m_{x,t}\left(\kappa_t^1,\kappa_t^2\right)\right),$$

where $d_{x,t}$ is the observed number of deaths. In order to express the mortality rates $m_{x,t}$ in terms of κ_t^1 and κ_t^2, we assume, as usual, a constant mortality intensity for each integer age. This implies the following relation between mortality rates and probabilities of death:

$$m_{x,t} = -\log(1-q_{x,t}).$$

Using Equation (3), we obtain:

$$m_{x,t}\left(\kappa_t^1, \kappa_t^2\right) = \log\left(1 + \exp\left(\kappa_t^1 + (x - \bar{x}) \cdot \kappa_t^2\right)\right).$$

B Trend Curve Fitting

Muggeo (2003) proposes a method to fit a continuous and piecewise linear trend curve to a series of ordered data. This curve fulfills optimality criteria with respect to the number of trend changes and their positions. In this section, we explain the method of Muggeo (2003) in detail for the case of time dependent data by considering the example of a time process κ_t, $t = t_0, ..., t_n$ in a parametric mortality model.

First, we fit a trend curve without any trend changes, i.e. a straight line, to the time process. This is done by maximizing the likelihood function for the residuals which are assumed to be normally distributed. Heteroscedasticity in the residuals is allowed for by iteratively updating the trend curve estimate and the variance estimates for the residuals (see Section 3). Next, we fit a trend curve with exactly one trend change. The trend change is positioned such that, again, the likelihood function for the residuals is maximized. More details on this (for any number of trend changes) follow below. We then compare the values of the MBIC (or any other information criterion under consideration) for zero and one trend changes. The number of parameters in a trend curve without trend changes is obviously two, and the inclusion of trend changes increases the number of parameters by two for each trend change, i.e. the time point of the trend change and the change in slope. If the newly included trend change reduces the MBIC, we assume that a trend change actually occurred. Subsequently, we increase the number of trend changes as long as additional trend changes reduce the MBIC. As soon as the MBIC does not decrease anymore, we make one final check whether the addition of two trend changes can reduce the MBIC further. If that is not the case, we are reasonably sure that any larger number of trend changes will lead to overfitting and that we have found the optimal number of trend changes. The fitted trend curve for this number of trend changes provides the optimal positions for these trend changes and the corresponding trend change intensities.

In what follows, we explain in detail how a trend curve with k trend changes can be fitted to the time process. Assuming a continuous and piecewise linear dependence of κ_t on t with k changes in the linear trend, we can write

$$\kappa_t = c + \beta' \cdot (t \cdot \mathrm{I} - \tau)^+ + \epsilon_t, \tag{4}$$

where

- $\tau = (t_0, \tau_1, ..., \tau_k)'$ denotes the time points at which the trend changes occur (including an artificial trend change at t_0 to ease notation),
- I is a $(k+1)$-dimensional vector of 1's,
- $\beta' = (\beta_0, ..., \beta_k)$ is the slope vector with β_0 the initial slope at $t = t_0$ and β_i the change in slope at τ_i, $i = 1, ..., k$,
- c is the intercept,

- $\epsilon_t \sim \mathcal{N}(0, \sigma_t^2)$ are residuals with potentially time dependent but known variance to allow for heteroscedasticity, and
- $(\cdot)^+ = max(\cdot, 0)$ element-wise.

In order to estimate the parameters τ, β, and c, a non-linear optimization problem needs to be solved. Muggeo (2003) suggests simplifying the optimization by iteratively solving a series of linear optimization problems whose solutions converge to the solution of the non-linear optimization problem. He simplifies the optimization problem by using a Taylor expansion for the vector $(t \cdot \mathrm{I} - \tau)^+$. An expansion around some initial values for the trend change time points, $\tau^{(0)} = (t_0, \tau_1^{(0)}, ..., \tau_k^{(0)})'$, yields

$$\begin{aligned}
(t \cdot \mathrm{I} - \tau)^+ &= \begin{pmatrix} (t - t_0)^+ \\ (t - \tau_1)^+ \\ \vdots \\ (t - \tau_k)^+ \end{pmatrix} \\
&\approx \begin{pmatrix} (t - t_0)^+ - \mathbb{1}(t > t_0)(t_0 - t_0) \\ (t - \tau_1^{(0)})^+ - \mathbb{1}(t > \tau_1^{(0)})(\tau_1 - \tau_1^{(0)}) \\ \vdots \\ (t - \tau_k^{(0)})^+ - \mathbb{1}(t > \tau_k^{(0)})(\tau_k - \tau_k^{(0)}) \end{pmatrix} \\
&= (t \cdot \mathrm{I} - \tau^{(0)})^+ - \Delta_\tau^{(0)} \cdot \mathbb{1}(t \cdot \mathrm{I} > \tau^{(0)}),
\end{aligned}$$

where $\mathbb{1}(\cdot)$ is the (element-wise) indicator function and $\Delta_\tau^{(0)}$ 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. This approximation can be inserted into Equation (4):

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

where $\gamma' = \beta' \cdot \Delta_\tau^{(0)}$. The resulting optimization problem is linear in β, γ, and c and can be solved by maximizing the log-likelihood function

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

Once the optimal parameter values $\beta^{(0)}$, $\gamma^{(0)}$, and $c^{(0)}$ have been determined, the estimates for the trend changes time points 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 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, we required $|\gamma_i^{(j)} - \gamma_i^{(j-1)}| < 10^{-4},\ i = 1, ..., k$ for a stop after j iterations.

The final parameter estimates for β, γ, and c can depend on the initial values $\tau^{(0)}$ for the Taylor expansion. Therefore, in order to minimize the risk of running into local maxima, we always carried out the iterative optimization for 1000 sets of randomly generated initial values. For any time process we considered, a significant number of optimization runs then led to the same and best parameter estimates so that we can be confident of having found the optimal solution.

C The CUSUM Test

A CUSUM test can be used to check whether a time series evolves around a constant mean or whether the mean changes over time. We use the CUSUM test to analyze, based on time series of raw variance estimates (see Section 3), whether the variance of the trend curve residuals is constant over time. Moreover, if the variance can not be assumed constant, the test indicates for which time periods variances should be estimated separately.

For a time series y_t, $t = t_0, ..., t_n$ (the raw variance estimates in our case), Ploberger and Krämer (1992) propose the following test statistic:

$$V_n(s) := \frac{1}{\hat{\sigma}_y \sqrt{n+1}} \sum_{t=t_0}^{\lfloor t_n \cdot s \rfloor} (y_t - \bar{y}),$$

where $\bar{y}$ and $\hat{\sigma}_y$ are the sample mean and the sample standard deviation of the time series. This test statistic is well defined on the interval $[0, 1]$. Under the null hypothesis of a constant mean in the y_t and for n to infinity, this statistic converges in distribution to a Brownian bridge, i.e.

$$V_n(s) \xrightarrow{d} B(s) = (W_s \mid W_1 = 0),\ s \in [0, 1],$$

where W_s is a Wiener process. A Brownian bridge is equal to zero at both endpoints and has an increasing uncertainty toward the middle of the bridge. Under the null hypothesis of a constant mean in the y_t, $|V_n(s)|$ exceeds a certain threshold with roughly the same probability as $|B(s)|$. If the mean in the y_t is not constant, the probability for $|V_n(s)|$ exceeding a certain threshold is (considerably) larger. Therefore, the null hypothesis is rejected if $|V_n(s)|$ exceeds the threshold which $|B(s)|$ only exceeds with a probability of 5%. The change in mean is then assumed to occur where $|V_n(s)|$ assumes its maximum value.

In order to detect multiple changes in the mean, the CUSUM test can be applied iteratively. Once a change in mean is detected, the data is split at that point, and the test is applied once again to each partition of the data set.

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.

Berkum, F. v., Antonio, K., and Vellekoop, M. (2014). The impact of multiple structural changes on mortality predictions. to appear in: *Scandinavian Actuarial Journal*.

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.

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

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.

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.

Chow, G. C. (1960). Tests of Equality Between Sets of Coefficients in Two Linear Regressions. *Econometrica*, 28(3): 591–605.

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.

Durbin, J. and Watson, G. S. (1950). Testing for serial correlation in least squares regression. I. *Biometrika*, 37(3-4): 409–428.

Durbin, J. and Watson, G. S. (1951). Testing for serial correlation in least squares regression. II. *Biometrika*, 38(1-2): 159–177.

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

Harris, D., Harvey, D. I., Leybourne, S. J., and Taylor, A. (2009). Testing for a unit root in the presence of a possible break in trend. *Econometric Theory*, 25(6): 1545–1588.

Harvey, D. I., Leybourne, S. J., and Taylor, A. (2009). Simple, robust, and powerful tests of the breaking trend hypothesis. *Econometric Theory*, 25(4): 995–1029.

Human Mortality Database (2015). 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 22 December 2015).

Hunt, A. and Blake, D. (2014a). Consistent mortality projections allowing for trend changes and cohort effects. Working Paper, Cass Business School.

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

Jarner, S. F. and Kryger E. M. (2011). Modelling Adult Mortality in Small Populations: The Saint Model *ASTIN Bulletin*, 41(2): 377–418.

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

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

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

Li, N. and Lee, R. (2005). Coherent Mortality Forecasts for a Group of Populations: An Extension of the Lee-Carter Method. *Demography*, 42(3): 575–594.

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.

Muggeo, V. M. and Adelfio, G. (2011). Efficient change point detection for genomic sequences of continuous measurements. *Bioinformatics*, 27(2): 161–166.

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).

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.

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.

Wang, H., Li, B., and Leng, C. (2009). Shrinkage tuning parameter selection with a diverging number of parameters. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 71(3): 671–683.

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.