It Takes Two:
Why Mortality Trend Modeling is more than modeling one Mortality Trend

Matthias Börger (corresponding author)
Institute of Insurance, Ulm University & Institute for Finance and Actuarial Sciences (ifa), Ulm
Helmholtzstraße 22, 89081 Ulm, Germany
Phone: +49 731 5031257. Fax: +49 731 5031239
Email: m.boerger@ifa-ulm.de

and

Jochen Russ
Institute of Insurance, Ulm University & Institute for Finance and Actuarial Sciences (ifa), Ulm
Helmholtzstraße 22, 89081 Ulm, Germany
Phone: +49 731 5031233. Fax: +49 731 5031239
Email: j.russ@ifa-ulm.de

Preliminary version – January 29, 2013

Abstract

Increasing life expectancy and thus decreasing mortality rates constitute a global trend that can be observed in almost all countries worldwide. Estimating the current rate at which mortality rates decrease and modeling the future rate of decrease is important for e.g. demographers and actuaries. This task is commonly referred to as mortality trend modeling.

Recent work, e.g. by (Sweeting, 2011) or (Li, et al., 2011) has established that in many countries the mortality trend appears to be a piecewise linear function. This can be used in stochastic mortality models by implementing trend components that generate a (stochastic) piecewise linear trend and some kind of random fluctuation around this trend.

We show that previously discussed versions of this approach have several shortcomings. In particular we show that one needs to distinguish between two different mortality trends: The actual mortality trend (AMT) prevailing at a certain point in time and the estimated mortality trend (EMT) that an observer would estimate given the realized mortality up to that point in time. The difference between these two results from the fact that the AMT is not observable and moreover an observer would not always be able to distinguish between a recent chance in the actual trend and a “normal” random fluctuation around the previous long term trend. Depending on the question at hand, the AMT or the EMT might be the relevant figure to use in analyses.

The paper provides a clear definition of and distinction between the actual mortality trend and the estimated mortality trend, discusses their connection, and explains which of the two is relevant for which kind of question. Moreover, a combined model for both trends including a stochastic start trend for the actual mortality trend is specified, and calibrated to mortality data.

1. Introduction and Motivation

Increasing life expectancy and thus decreasing mortality rates constitute a global trend that can be observed in almost all countries worldwide. Estimating the current rate at which mortality rates decrease and modeling the future rate of decrease is commonly referred to as mortality trend modeling. This task is important for e.g. demographers and actuaries and e.g. constitutes an important input for the pricing and reserving of annuity and pension products.

Many mortality models like the well-known Lee-Carter model (Lee & Carter, 1992) or the model by (Cairns, et al., 2006) incorporate one or more time dependent parameters that model the mortality trend. For many populations, the historical development of these time dependent parameters has been rather linear over the past few decades, as can be seen e.g. from Figure 1. Here, the time dependent parameters which describe the general level of mortality in the model of (Börger, et al., 2012) are plotted for a variety of countries. Therefore, in many applications, a linear extrapolation of the historic mortality trend parameters is used as a best estimate for the future: In stochastic projections, mortality scenarios are generated by a random walk with drift, where the drift is given by this best estimate.

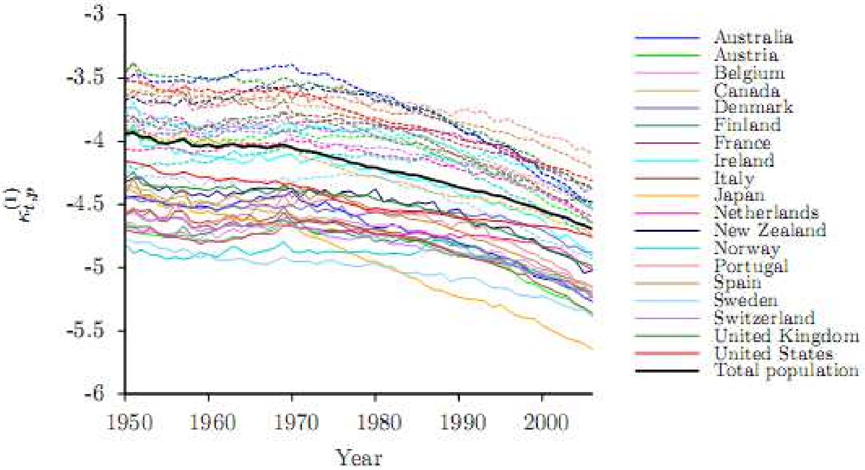

Figure 1: Exemplary time series for time dependent parameters which describe the general level of mortality (Source: (Börger, et al., 2012))

However, as Figure 1 already indicates and recent work, e.g. by (Sweeting, 2011) or (Li, et al., 2011) has established, there have in fact been significant changes of the 'speed of the reduction in mortality rates' and hence of the mortality trend parameters in many countries. They conclude that mortality trends appear to be piecewise linear functions and that in the past, random fluctuations around these piecewise linear trends occurred. This effect can be seen, e.g., in Figure 2 which shows the mortality trend parameters in the model of (Cairns, et al., 2006) for English and Welsh males aged 60 to 89.

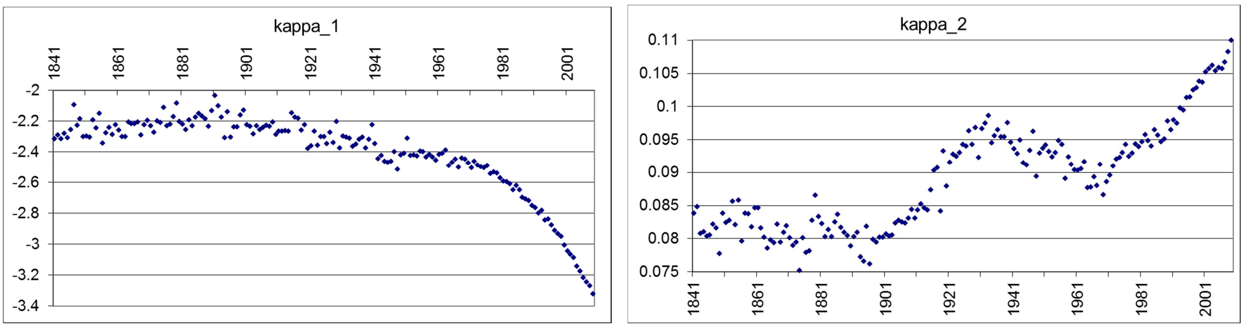

Figure 2: Mortality trend parameters in the model of (Cairns, et al., 2006) for English and Welsh males

This can be used in stochastic mortality models by implementing trend components that generate a (stochastic) piecewise linear trend and some kind of random fluctuation around this trend. Just extrapolating the most recent linear trend and modeling future mortality as a random walk around this drift, i.e. a difference stationary process appears too simple. In particular, this approach appears to underestimate long term uncertainty, as outlined, e.g., by (Sweeting, 2011) or (Börger, et al., 2012). Moreover, in case of a random walk with drift, the long term trend uncertainty only depends on the volatility of year by year random fluctuations. This is not a plausible assumption in general as closely related populations (in terms of political, economic, and social factors) with significantly different short term fluctuations must still be expected to have about the same long term uncertainty.

In this paper, we will discuss several shortcomings of previously discussed versions of this approach. Most importantly, we introduce the concept of two different mortality trends: The (unknown) actual mortality trend (AMT) prevailing at a certain point in time and the estimated mortality trend (EMT) that an observer would estimate given the realized mortality up to that point in time. Amongst other effects, the difference between these two trends results mainly from the fact that (in particular immediately after a trend change) an observer would not always be able to distinguish between a recent chance in the actual trend and a "normal" random fluctuation around the previous long term trend. We will show several aspects where not distinguishing between these two trends might have negative impact on the used models.

The first aspect is the start trend used for mortality projections. The focus of most existing mortality models is on modeling actual mortality and hence using the AMT. This particularly holds for the models by (Lee & Carter, 1992), (Cairns, et al., 2006), as well as (Sweeting, 2011). All these models assume that the future trend (or start trend, where trend changes are possible) for the projection of actual mortality is known. However, what we really know is just today's estimation for the current trend, i.e. the EMT. Hence, when modeling actual mortality and the AMT, the model should incorporate the fact, that the start value of the AMT is unknown. This could be achieved by using a stochastic start value for the AMT which should be somehow centered around the EMT and the distribution of which should reflect the confidence one has about the quality of the EMT. If it is assumed that today's EMT coincides with the AMT (whilst in reality it is just an estimate for the AMT), uncertainty about future mortality development is systematically underestimated, i.e. confidence bands are too narrow.[1]

A second aspect is that depending on the question at hand, the AMT or the EMT might be the relevant figure to use in analyses. Hence, for some applications a model for the future development of the AMT is needed, for other applications, a model for the future development of the EMT is needed. Often, AMT and EMT are even required simultaneously. A model that only captures one of these two trends might therefore not be suitable for certain analyses.

This can be seen from two simple examples (we will discuss some examples in more detail in Section 2): If one is interested in a confidence band for the cash-flow of annuities paid out from a book of annuity contracts, say, 10 years from now, survival rates derived from a model for the AMT would be relevant. If, however, one is interested in the question how reserves for the same book of annuity contracts can change from one year to the next, one needs to model actual mortality over one year as well as the evolution of the EMT over this year, since the EMT would be the basis for the calculation of policy reserves. Hence, a model with two components is required: (1) the development of actual mortality (based on some AMT) and (2) the development of the EMT which at any point in time is some function of the actual mortality up to that point in time.

Stochastic mortality trend modeling has already been discussed by several authors. First of all by (Sweeting, 2011) whose model also forms the basis for our AMT model. However, he does

[1] This uncertainty about the start trend is different from and in general more substantial than parameter uncertainty tackled e.g. by (Cairns, et al., 2006), cf. Section 3.

not consider a stochastic start trend which leads to a systematic underestimation of mortality/longevity risk. We also see a weakness in his approach for modeling the trend change intensity (see below). Moreover, Sweeting's model does not include a component for the EMT. (Zhu & Bauer, 2011), on the other hand, consider estimated mortality only. They derive different models, in particular forward mortality factor models, which describe changes and the uncertainty in mortality projections over time. In comparison to our approach, they do not only model the EMT as part of a parametric mortality model but entire generational mortality tables. Moreover, they show that commonly used mortality models like Lee-Carter do not adequately describe the uncertainty in mortality projections. We agree with this finding from a theoretically perspective since, in their standard specifications, those models are pure AMT models. As for Sweeting's model, the model of Zhu and Bauer can only be applied if risk from either realized or expected mortality is to be considered. In many situations, however, both components are necessary. For instance, when computing capital requirements over the 1-year time horizon of Solvency II, one is interested in realized mortality over one year and changes in expected mortality for the time beyond (see (Börger, 2010)).

Two models which aim to cover both realized mortality and changes in expected mortality (in particular in the Solvency II context) are the models of (Plat, 2011) and (Börger, et al., 2012). Although their modeling approaches differ somewhat – Plat considers mortality reduction factors while Börger et al. model log mortality rates – the common idea is to establish a connection between realized mortality and changes in expected mortality. More precisely, Plat models changes in mortality reduction factors and then determines realized mortality rates which are somewhat consistent with the change in expected mortality reductions: if expected mortality reductions increase, realized mortality rates must have been rather low in that year. Börger et al. take the opposite route and analyze how realized mortality influences the expectation of future mortality. However, both papers do not clearly distinguish between the AMT and the EMT. Finally, (Cairns, et al., 2011) and (Cairns, 2012) consider changes in mortality trends in the context of hedge effectiveness. Here, the authors find that changes in the EMT over the course of a longevity hedge can have a significant impact. However, they do not allow for changes in the AMT over time and do not account for the uncertainty of actually not knowing the AMT, in particular at the start of a simulation.

The remainder of this paper is organized as follows: In Section 2, we discuss some examples to clarify the relation and the difference between the AMT and the EMT. In Section 3, we introduce a new mortality model that simultaneously models the AMT (and hence actual mortality) as well as the EMT. We discuss the calibration of this model and how it can be applied in practice. Moreover, we present one approach how a stochastic start trend (as discussed above) can be integrated in our model. Section 4 discusses how the EMT should be derived from observed mortality, i.e. 'how fast' one should react to potential trend changes in the data. Finally, Section 5 concludes and gives an outlook for further research.

2. Differences and Relation between AMT and EMT

In this Section, we discuss some examples in order to clarify the relation and the difference between the AMT and the EMT and to motivate the need for a combined model for both, the AMT and the EMT.

As a first example, consider a portfolio of annuity contracts that is in run-off. The owner of the portfolio is interested in, say, a confidence band for future annuity payments or their present value. In such a run-off simulation, only actual future mortality and hence the AMT is of relevance. Nevertheless, even here, it is important to distinguish between the AMT and the EMT. One has to be aware, for instance, that today's AMT is not known and therefore the mortality trend at the start of the simulation is uncertain. We will get back to this issue in Section 3.2 by proposing a stochastic start trend to account for this uncertainty. We will see that taking this uncertainty into account can lead to significantly wider confidence bands for future mortality. If the owner of the portfolio is additionally interested in a confidence band for the reserves that need to be put up, the EMT has to be included in the simulation as well, since at any future time, the reserves depend on the then prevailing EMT and not on the then unknown AMT. We can conclude that even for a simple example, where at first glance only the AMT is of relevance, one needs to be aware of and consider the differences between the two different trends.

Now consider an insurer that has bought some kind of mortality derivative to reduce the risk resulting from deferred annuities with some kind of guaranteed annuitization options (GAO), where the insurer has guaranteed some minimum annuity amount or minimum annuitization rate at the end of the deferment period. The derivative will typically pay out at the end of the deferment period the present value of the annuity that was guaranteed to the insured. Of course, this present value will depend on the then prevailing EMT since the AMT is not known. The actual annuity payouts, however, will depend on the realized mortality after the deferment period and hence on the development of the AMT. Thus, when pricing the derivative or analyzing the derivative's payout, one is interested in the EMT at the end of the deferment period which is a function of realized mortality and hence the AMT during the deferment period. When analyzing the hedge efficiency from the insurer's perspective (cf. (Cairns, et al., 2011) and (Cairns, 2012)), in addition also the AMT after the deferment period is relevant. So, all these questions require a joint model for the AMT and the EMT.

As a final example, consider capital requirements under the new regulatory system Solvency II. Here, a percentile of the distribution of the change in reserves over one year is the key driver for the required capital. The reserves after one year will of course be calculated from an EMT, since the AMT is unknown. So the question is how the EMT can change over one year. Since the EMT in one year is a function of realized mortality within this year, the distribution of actual mortality has to be simulated first, using the AMT. In (Plat, 2011) and (Börger, et al., 2012), AMT and EMT are somehow implicitly jointly modeled in this context; however both approaches lack a clear formalization of and distinction between AMT and EMT.

3. A Combined Model for AMT and EMT

In this section, we fully specify and estimate a combined model for the AMT and the EMT. Exemplarily, we do this for the male population aged 60 to 89 in England and Wales. For this population, mortality data is available at the Human Mortality Database for years 1841 to 2009.[2] Of course, the model could be applied to any population where the data history is sufficiently long to estimate the probability and magnitude of changes in the AMT.

We commence with a description of the model structure and model estimation for the AMT in the following section. Subsequently, we discuss in detail the implications of not knowing the current AMT and how this uncertainty can be accounted for by a stochastic start trend. Finally, we extend our model by a component for the EMT.

3.1 The AMT Model Component

The basis for our AMT model is the stochastic mortality model proposed by (Sweeting, 2011) which itself is an extension of the model of (Cairns, et al., 2006). However, the AMT/EMT modeling concept which we introduce in this section is also applicable for many other parametric mortality models which include time dependent processes, e.g. the models of (Plat, 2009) or (Börger, et al., 2012). Note however, that we cannot recommend using our trend concept in the model of (Lee & Carter, 1992) since due to its age dependent but time constant parameters, this model should only be calibrated to rather short data periods. For longer data series, the fit of the Lee-Carter model typically becomes poor since the age dependent effects

[2] See www.mortality.org

are not as time constant in reality as presumed in the model structure. However, this is a general drawback of the Lee-Carter model (see, e.g., (Lee & Miller, 2001) for a thorough discussion).

In the model of (Cairns, et al., 2006), annual mortality rates are modeled as

$$logit(q_x) := \log\left(\frac{q_x}{1-q_x}\right) = \kappa_1 + \kappa_2 \cdot (x - \bar{x}),$$

where $\bar{x}$ is the average of all ages under consideration, i.e. $\bar{x} = 74.5$ in our case. For estimating this model, we use the freely available Lifemetrics Tool.[3] The resulting time series for κ_1 and κ_2 are shown in Figure 2.

As outlined in the Introduction, it seems plausible to project future mortality rates by extrapolating these time series linearly with random changes in the slope of the linear trends. Thus, for both time series, we need to specify the probability for a change in the trend as well as a distribution of the trend change intensity. For this, we analyze trend changes in the historical data.

The general idea behind this is to fit a series of linear trends, i.e. straight lines, to the data such that the residual sum of squares is minimized. The fitting is done using weights to account for the obvious heteroscedasticity in the data (see the Appendix for details on the weights applied). Moreover, we impose the following two constraints:

(i) Each line starts where the previous line ends, i.e. lines for adjacent parts of the data intersect in the year where the trend change occurs. Thus, looking at the time series as a whole, we obtain one continuous line with changing slope for each time series.

(ii) All trend changes must be significant with respect to the 1% significance level in a slightly modified Chow test (see the Appendix for details on this test).

By the first constraint, we rule out any jumps in the κ processes. As the data does not clearly indicate the need for such jumps this constraint seems reasonable also for the sake of simplicity. The rationale for the second constraint is that a larger number of trend changes would always reduce the residual sum of squares. In fact, by allowing for a trend change in every year, the residual sum of squares could be reduced to zero. Thus, some condition like the second constraint is necessary to ensure that only 'real' trend changes are detected. Otherwise, also random fluctuations would be identified as trend changes.

Note that we also considered the other tests which are described in (Sweeting, 2011) for determining the significance of a trend change, i.e. the Durbin Watson test or the t-test for a change in slope. However, we chose to use the Chow test since only this test accounts for the heteroscedasticity in the data. Nevertheless, the other tests helped us to decide on the significance level in the Chow test. Since the other tests seem to be more restrictive in general, we chose a rather small significance level of 1%.

Figure 3 shows the raw and fitted κ processes. For both κ_1 and κ_2, the optimal number of trend changes in the sense of minimizing the residual sum of squares is 7. The trend change years as well as the p-values of the Chow tests and the AMT for each subsequent time period are given in Table 1.

[3] See www.lifemetrics.org

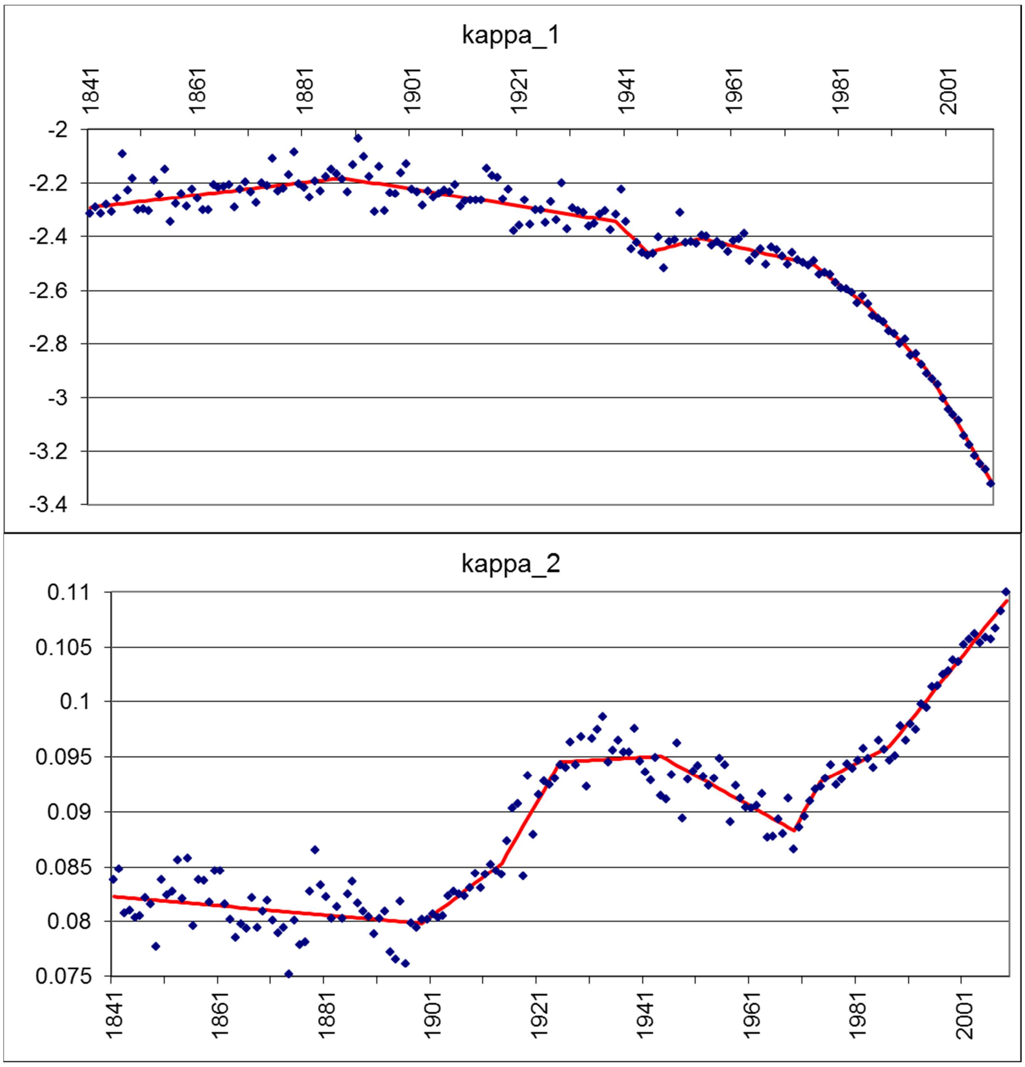

Figure 3: Raw and fitted κ processes for English and Welsh males

κ_1			κ_2		
Year	p-value	AMT	Year	p-value	AMT
1841		0.00236161	1841		-4.20E-05
1888	1.43E-12	-0.00320107	1899	7.69E-13	0.00036656
1939	0.00489344	-0.0192053	1914	3.76E-05	0.00083571
1945	0.00466214	0.00529551	1925	5.74E-08	2.77E-05
1955	0.00542198	-0.00469249	1944	0.00186355	-0.0002695
1976	1.23E-07	-0.0154744	1969	5.12E-06	0.00088583
1986	0.00848367	-0.0213147	1974	0.00989872	0.00024596
1997	7.44E-09	-0.0349844	1987	3.58E-05	0.00060327

Table 1: Detected AMT changes with p-values of Chow tests and AMT values

Based on these historical trend changes, we can now estimate the trend change probability p and the distributions of the trend change intensities $F_{\lambda_i}, i = 1,2$. Since we have observed 7 trend changes in 169 years of data for both κ_1 and κ_2, the intuitive estimate for the trend change probability is $p = 7/169$. Moreover, we assume that trend changes in κ_1 and κ_2 occur independently of each other as we have not found simultaneous changes in κ_1 and κ_2 in the data.

For the trend change intensities $\lambda_i, i = 1,2$, we assume a decomposition into two random variables,

$$\lambda_i = S_i \cdot M_i, i = 1,2,$$

where M_i describes the absolute magnitude of changes in the trend and S_i indicates whether the slope increases or decreases by the magnitude M_i. Thus, S_i follows a Bernoulli distribution with values -1 and 1 and probability 0.5. Since it is rather difficult to deduce a specific distribution for M_i from only 7 trend changes, we simply assume a Normal distribution and only estimate its parameters, μ_i and σ_i^2, from the available data, i.e. as the sample mean and sample variance of the absolute changes in the AMT in Table 1. The resulting parameters are provided in Table 2.

	$\boldsymbol{\kappa_1}$	$\boldsymbol{\kappa_2}$
$\boldsymbol{p}$	0.041420118	0.041420118
$\boldsymbol{\mu_i}$	0.012335376	0.00059077
$\boldsymbol{\sigma_i^2}$	4.32029E-05	9.28422E-08

Table 2: Trend change probabilities and parameters of the distribution of trend change intensities

The proposed decomposition of λ_i has several implications: First, the distribution F_{λ_i} is symmetric around zero which means that trend changes in both directions are equally likely and that their expected magnitude is the same for both directions. This coincides with common practice in mortality modeling where, in general, the most recent trend is extrapolated as a best estimate for the future mortality evolution. For our setting, this means that the current AMT is always the best estimate of future AMTs. Secondly, we have only little probability mass around zero but two “mass bumps” around the positive and the negative of the sample means. This is in line with the way we calibrated the trend change probability p. This probability is deduced from significant trend changes only and thus, in case a trend change is indicated in a simulation, the change in slope should be significant in general. Note that this is a difference to the approach of Sweeting. He also derives the trend change probability from significant historical trend changes but assumes a Normal distribution around zero for the trend change intensity. Thus, his approach tends to yield too many trend changes close to zero, i.e. trend changes which would not have been detected in the analysis of historical data. So in this sense the assumed distribution of future trend change intensities is not consistent to the calibration of trend change probabilities in Sweeting’s approach. In order to compensate for the numerous trend changes close to zero, the tails of Sweeting’s distribution for the trend change intensity are fairly wide as we will see later.

For any future year t, the processes κ_1 and κ_2 can then be projected as

$$\kappa_i(t) = \hat{\kappa}_i(t) + \varepsilon_i(t), i = 1,2,$$

where $\hat{\kappa}_i(t)$ denotes the actual value of the trend process and $\varepsilon_i(t)$ is Gaussian noise to allow for random fluctuations around this process. The process $\hat{\kappa}_i$ is simply projected as a straight line with $AMT_i(t)$ (the current AMT) as slope:

$$\hat{\kappa}_i(t) = \hat{\kappa}_i(t-1) + AMT_i(t), i = 1,2,$$

The AMT itself depends on whether a trend change occurred in $t-1$:

$$AMT_i(t) = \begin{cases} AMT_i(t-1) & , if\ no\ trend\ change\ occurred\ in\ t-1 \\ AMT_i(t-1) + \lambda_i(t-1) & , if\ a\ trend\ change\ by\ \lambda_i(t-1)\ occurred\ in\ t-1 \end{cases}$$

Thus, for any simulation of future mortality, we need to provide the value of the (2-dimensional) actual mortality processes and the (2-dimensional) AMT for the last year in the data set, i.e. $AMT(2009)$ and $\hat{\kappa}(2009)$ in our specific case. However, as indicated in the Introduction, we cannot observe the actual trend process and thus do not know their values in 2009 for certain. This issue is discussed in detail in the following section.

3.2 The Stochastic Start Trend

There are several reasons why we do not know $AMT(2009)$ and $\hat{\kappa}(2009)$ for certain. First, both parameters are estimated based on only a limited number of data points which implies parameter uncertainty. Secondly, we might have misestimated the exact year of the most recent trend change. The trend change might have occurred in the year(s) before or after the year we think it did and we have not been able to detect this due to random noise or lack of data. Thirdly, there might have been another trend change between the last detected one and today. We do not know whether deviations from the latest detected trend in most recent years are due to random noise or due to another trend change. Additional years of data are necessary to answer this question.[4]

Possible undetected or wrongly identified trend changes clearly imply the largest uncertainty in $AMT(2009)$ and $\hat{\kappa}(2009)$. In comparison, the parameter uncertainty about $AMT(2009)$ and $\hat{\kappa}(2009)$ seems comparably small. Therefore, in order to keep things as simple as possible, we primarily take into account the possibility of undetected trend changes in the following and assume that all detected trend changes and the AMTs between those trend changes have been determined correctly. However, if one wanted to also allow for parameter uncertainty this would be possible.[5]

The principle idea to account for undetected changes in the AMT is to specify joint distributions for $AMT_1(2009)$ and $\hat{\kappa}_1(2009)$ as well as $AMT_2(2009)$ and $\hat{\kappa}_2(2009)$ and to draw starting values randomly from these distributions for each simulation path.[6] We set up these joint distributions by assuming there was one additional trend change between the last detected one and 2009, i.e. between 1997 and 2009 for κ_1, and 1987 and 2009 for κ_2.[7] Such additional trend changes would lead to values for $AMT_i(2009)$ and $\hat{\kappa}_i(2009)$ as shown in Table 3. Here, the values for 1997 (for κ_1) and 1987 (for κ_2) denote the cases where there is no additional trend change. Regarding the probability for each of the $\left(AMT_i(2009), \hat{\kappa}_i(2009)\right)$ pairs, we need to assess, for each additional trend change, how likely it is that this trend change really occurred. To this end, we performed Chow tests and interpreted 1 minus the p-values as likelihood for each pair. The p-value of the last detected trend change is assumed zero – according to the assumption from above that all detected trend changes have been assessed correctly. In order to obtain the probabilities in Table 3, the likelihoods have been standardized such that they sum up to 1.

[4] It is possible that in the future, e.g. in the year 2015, the time series until 2015 will show that a significant trend change occurred, say, in the year 2007.

[5] For any given set of trend change years, the continuous line which describes the historical data is fitted by a linear model. Therefore, for each parameter, we obtain a distribution which describes the uncertainty in each parameter estimate. Parameter uncertainty in $AMT(2009)$ and $\hat{\kappa}(2009)$ can thus be accounted for by drawing from the distribution of the last detected AMT.

[6] This procedure is analogous to the way parameter uncertainty is typically accounted for in the random walk with drift approach (see, e.g., (Cairns, et al., 2006)).

[7] Note that, theoretically, there might have even been more than one undetected trend changes until 2009. However, we regard this as a rather unlikely case and therefore do not account for it.

Year	$\hat{\kappa}_1(2009)$	$AMT_1(2009)$	Probability	$\hat{\kappa}_2(2009)$	$AMT_2(2009)$	Probability
1987	n/a	n/a	n/a	0.10921	0.0006032	0.0776
1988	n/a	n/a	n/a	0.10915	0.0005959	0.0091
1989	n/a	n/a	n/a	0.10910	0.0005896	0.0222
1990	n/a	n/a	n/a	0.10909	0.0005886	0.0206
1991	n/a	n/a	n/a	0.10904	0.0005826	0.0302
1992	n/a	n/a	n/a	0.10899	0.0005763	0.0388
1993	n/a	n/a	n/a	0.10891	0.0005664	0.0517
1994	n/a	n/a	n/a	0.10884	0.0005570	0.0587
1995	n/a	n/a	n/a	0.10872	0.0005421	0.0673
1996	n/a	n/a	n/a	0.10863	0.0005297	0.0705
1997	-3.3140	-0.03498	0.3782	0.10851	0.0005109	0.0735
1998	-3.3152	-0.03529	0.1053	0.10841	0.0004944	0.0742
1999	-3.3158	-0.03549	0.1344	0.10829	0.0004726	0.0748
2000	-3.3151	-0.03530	0.0399	0.10825	0.0004571	0.0729
2001	-3.3152	-0.03533	0.0319	0.10810	0.0004166	0.0734
2002	-3.3162	-0.03570	0.0707	0.10802	0.0003817	0.0713
2003	-3.3170	-0.03606	0.0949	0.10811	0.0003748	0.0610
2004	-3.3158	-0.03569	0.0308	0.10857	0.0004516	0.0251
2005	-3.3151	-0.03545	0.0092	0.10888	0.0005058	0.0066
2006	-3.3141	-0.03497	0.0000	0.10926	0.0006193	0.0000
2007	-3.3152	-0.03579	0.0053	0.10967	0.0008537	0.0067
2008	-3.3213	-0.04454	0.0994	0.10995	0.0014015	0.0137
Mean	-3.3157	-0.03624		0.10859	0.0005164	

Table 3: Empirical distributions for the stochastic start trend

The last row in the table contains the means of $AMT_i(2009)$ and $\hat{\kappa}_i(2009)$. Interestingly, the means for $AMT_1(2009)$ and $\hat{\kappa}_1(2009)$ are significantly smaller than those values when assuming a last trend change in 1997. This indicates that the AMT for κ_1 is rather steeper downward sloping with time. Thus, simply assuming the last detected AMT as a start trend seems to overestimate κ_1 and future mortality rates. Hence, considering a stochastic start trend is not only important to capture the entire uncertainty in the future mortality evolution but also to properly assess its expected mean level at outset. For κ_2, we observe the same tendency but overall uncertainty regarding the $AMT_2(2009)$ seems larger than for $AMT_1(2009)$.

With the start trend and its distribution specified, we now have a full model for the AMT. As discussed in detail in Section 2, this model can be used to quantify mortality/longevity risk whenever one is only interested in the realized mortality evolution, e.g. when estimating the costs for a run-off of an annuity/pension portfolio.

For illustrative purposes and to compare our modeling approach to existing approaches, we now look at remaining period life expectancies at birth for a 60-year old in 2010 to 2050 in different models.[8] More precisely, we have considered the following modeling approaches:

(i) A 2-dimensional random walk with drift with the covariance matrix calibrated to the 20 most recent data points. After estimating the covariance matrix, we have set the drift to the most recent AMT to increase comparability with the other modeling approaches.

(ii) The approach of (Sweeting, 2011), recalibrated to the trend changes we have found in the historical data.

[8] Life expectancies are based on mortality rates for ages up to a maximum age of 140. For average scenarios, this very high maximum age does not have a significant impact; the life expectancies are very similar to those for a maximum age of 120. For extreme longevity scenarios, however, it seems reasonable and necessary to allow for such a high maximum age.

(iii) The new AMT model from the previous subsection without stochastic start trend.
(iv) The new AMT model from the previous subsection with stochastic start trend.

Figure 4 shows the medians (solid) as well as the 10^{th} and the 90^{th} percentile (dotted) of the life expectancies for all four approaches. Given the historical evolution, the median life expectancies look plausible in all four cases. For the first three approaches, they even coincide due to the same starting value for the AMT or drift, respectively. When considering a stochastic start trend, the median scenario is slightly higher. The reason for this becomes obvious from a comparison of the start trend distributions and the last detected AMTs (see Table 3). Most values of the start trend distributions are smaller than the last detected AMTs which implies that the start trend is typically steeper than the last detected AMT. Consequently, the median life expectancies are larger for the case with stochastic start trend.

In contrast to the medians, the percentiles for the AMT modeling approaches differ significantly. The random walk with drift yields the widest confidence intervals for the first years and by far the narrowest for later years. In particular, the width of the long term confidence intervals, with only 3.1 years in 2050, seems unrealistically small. This reconfirms the general issue that the random walk with drift is, in general, not able to properly reflect long term uncertainty.

Sweeting's approach implies an 80% confidence interval of 23.3 years in 2050 which seems extremely wide. At first sight, this is surprising as we have explained in the previous subsection that his model seems to yield too many changes in the AMT which are close to zero. However, calibrating the volatility of a mean-zero distribution for the trend change intensity to only highly significant historical trend changes implies a fairly large value for the volatility. Therefore, Sweeting's approach does not only seem to yield too many trend changes close to zero but also too many extreme trend changes.[9]

Based on our convoluted distribution for the trend change intensity, the average trend change may be about as strong as in Sweeting's approach but we get less extreme trend changes, as a kind of compensation for the less trend changes around zero. That is why the confidence bounds for our AMT approach are significantly narrower. Without stochastic start trend, we obtain a confidence interval of 7.7 years in 2050, with stochastic start trend, this confidence interval increases to 8.5 years. Both confidence intervals seem plausible from our point of view and we see how allowing for a stochastic start trend results in wider confidence intervals.

[9] Note that his effect also results from the fact that the assumed distribution of future trend change intensities is not consistent to the calibration of trend change probabilities in Sweeting's approach, cf. footnote **Fehler! Textmarke nicht definiert.**.

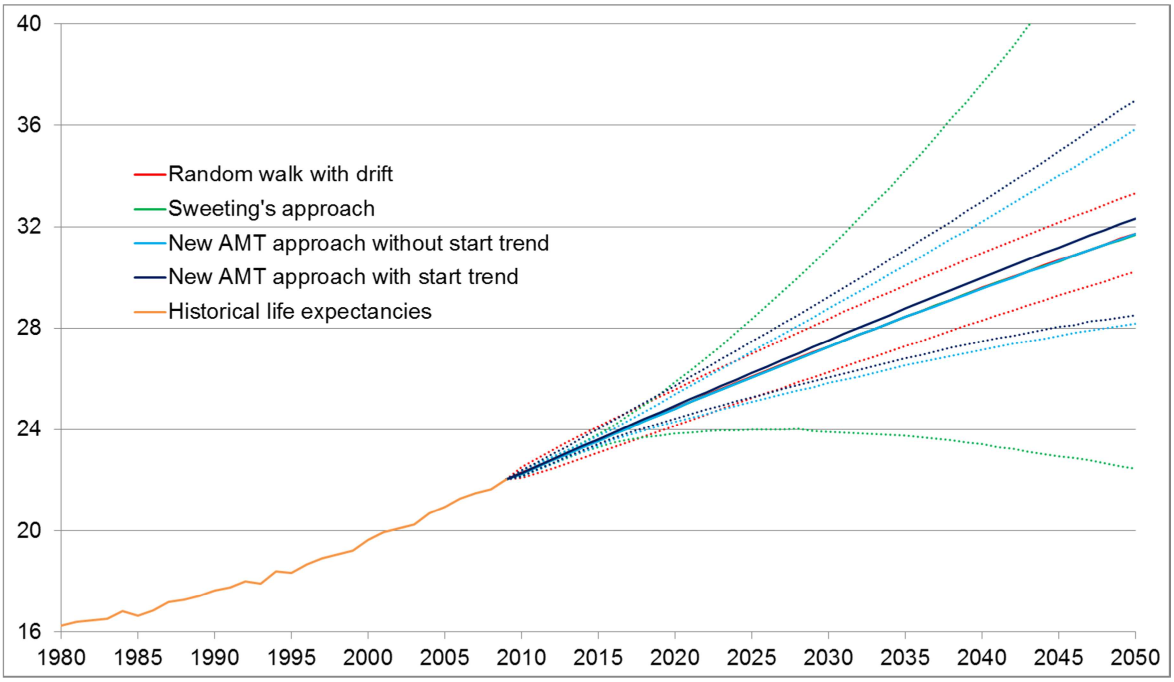

Figure 4: Remaining period life expectancies for a 60-year old based on different AMT modeling approaches

3.3 The EMT Model Component

As outlined the Introduction and in Section 2, often, one is not only interested in the realized future mortality evolution, but also in changes of best estimate mortality assumptions over a certain period of time (see Section 2 for examples). Thus, one needs to know the EMT, i.e. the best estimate for the future AMT given realized mortality up to a specific point in time.

Theoretically, this question is easy to answer as, the way we have specified our AMT model, the current AMT is always the best estimate for the future AMT. Unfortunately, the current AMT is not known to an observer who only observes realized mortality up to a specific point in time. This is the reason, why we have introduced the stochastic start trend in the previous subsection: The start trend distribution reflects our uncertainty about the current AMT and its mean therefore represents our best estimate for the AMT. Thus, the mean of the start trend distribution can also be interpreted as the expectation of the future AMT.

If one is interested in the EMT in the future, e.g. in one year for the computation of capital requirements under Solvency II, a similar start trend distribution would be required for those future points in time. The mean of this distribution would be the EMT. Thus, one would have to carry out analyses similar to the one we perform when determining the start trend distribution in the previous subsection. Such an approach is certainly possible but does not seem feasible from a practical point of view. We see two main reasons for this: First, in many applications the EMT at a future point in time has to be determined in each simulation path. Thus, we would require a start trend distribution in each path which implies a massive computational burden. Secondly, also the set of detected trend changes would have to be reassessed in each simulation path as the additional data from the simulation can indicate another significant trend change after the last one of the originally detected trend changes. This is particularly relevant when considering the EMT at the end of a longer simulation period. Therefore, alternative and more feasible methods need to be applied to determine the EMT at future points in time.

This could, in principle, be any existing approach for projecting the future mortality trend based on available historical data. However, since we assume the AMT to be piecewise linear an obvious approach for establishing the EMT is a linear regression to most recent data. Such an approach would also be easy to implement and efficient in terms of computing time. Since data points further in the past should generally have a smaller impact on the EMT, we introduce the following weights in the linear regression:

$$w_i(t-s) = \frac{1}{\left(1 + {}^{1}\!/_{h_i}\right)^{t-s}}, s \leq t \; and \; i = 1{,}2.$$

Here, t denotes the current year, $h_i \geq 0$ is a weighting parameter for κ_i, and the weights decrease going backwards in time, i.e. when the time lag $t-s$ increases. This approach is similar to the approach which (Börger, et al., 2012) use to project future mortality.

The crucial question is how h_i should be calibrated or, in other words, how many data points should be considered to determine the EMT. If we consider only very few we are able to follow changes in the AMT rather quickly. Thus, the EMT is likely to be close to the AMT if a trend has happened. On the other hand, we run the risk of the EMT falsely identifying random fluctuations as trend changes. Moreover, the EMT could be become rather instable since one additional data point would have a massive impact on the linear fit. On the other hand, in case of a fit to a rather long data history, the EMT would be more stable but it would take longer to identify a change in the AMT. We discuss this issue in detail in the following section and determine optimal values for h_i.

4. Estimation of the Future EMT

In this section, we first determine optimal values for the weighting parameters h_i. Thereafter, we compare the EMT modeling approach from the previous subsection with other approaches by analyzing how close to the current AMT the EMTs are.

We use separate weighting parameters for the two κ_i processes since the values of the weighting parameters particularly depend on two factors which can be rather different for both processes: the volatility of the random fluctuations around the AMT and the magnitude of trend changes. The larger the volatility, the larger the weighting parameter should be in general as more data significantly reduces the uncertainty in the weighted regression and thus the likelihood of misestimating the AMT. The impact of the magnitude of trend changes is contrary. The larger this magnitude, the more important it is for the EMT to reproduce the trend change quickly and hence, the smaller the weighting parameter should be.

We determine optimal values for the parameters h_i empirically by the following procedure: We simulate the AMT for the next 40 years 100,000 times and, for each path, derive the EMTs based on various weighting parameters.[10] For each weighting parameter, we then add up the 100,000 squared differences between the estimated EMTs and the AMTs. The optimal weighting parameter is the one which minimizes this mean squared error. For κ_1, this turns out to be $h_1 = 1.300$, for κ_2, it is $h_2 = 1.875$. These weighting parameters seem rather small at first sight as they assign almost 90% weight to only the last 4 data points for κ_1 and the last 5 data points for κ_2. However, given the very small volatility in the random fluctuations (see the last data points in each graph of Figure 3: Raw and fitted $\boldsymbol{\kappa}$ processes for English and Welsh males), this is not surprising. In fact, if we increase this volatility, the optimal weighting parameters become significantly larger. Thus, for other populations, the optimal weighting parameters may be different but we leave this question for future work. Note also that our optimal weighting parameters cannot be directly compared to those of (Börger, et al., 2012).

[10] The simulation horizon of 40 years is rather arbitrary. One should ensure though that the impact of the (fixed) historical data on the EMT is minimal to obtain parameter values which are valid rather generally. The random fluctuations around the AMT, $\varepsilon_i(t)$, are simulated as correlated Gaussian noise with mean zero and variances according to the variances in the residuals since the last detected trend changes, i.e. since 1997 for κ_1 and 1987 for κ_2. The correlation is forecast as the correlation between the residuals since 1987.

They increase the volatility of the random fluctuations by an add-on and their weighting parameters should therefore be larger.

Finally, we want to compare the EMT based on the optimal weighting parameters with alternative models for the EMT. More precisely, we consider 7 different EMT estimation approaches:

(i) Weighted regression with 'optimal' weighting parameters $h_1 = 1.3$ and $h_2 = 1.875$
(ii) Weighted regression with weighting parameters $h_1 = 0.8$ and $h_2 = 1.375$ (absolute reduction by 0.5)
(iii) Weighted regression with weighting parameters $h_1 = 1.8$ and $h_2 = 2.375$ (absolute increase by 0.5)
(iv) Weighted regression with weighting parameters $h_1 = h_2 = 1.5875$ (average of optimal h_i)
(v) Unweighted regression from last 5 data points for κ_1 and κ_2
(vi) Unweighted regression from last 10 data points for κ_1 and κ_2
(vii) Unweighted regression from last 20 data points for κ_1 and κ_2

The last three approaches are in the spirit of (Cairns, et al., 2011) who refer to the time frame for the regression as the recalibration window. The quantity we look at in this comparison is the remaining cohort life expectancy of a 60-year old 40 years from now. Thus, we again simulate the AMT for 40 years and, for each of the 1,000,000 paths, estimate the EMT according to the 7 outlined approaches. Based on these EMTs, we then forecast the cohort life expectancy and compute its deviations from the best estimate based on the AMT. We consider the cohort life expectancy here since it is highly dependent on the future mortality trend – in contrast to, e.g., the period life expectancy – and is similar in structure to annuity present values which actuaries are typically interested in.

Figure 5 shows the (middle part of the) empirical probability density functions for the differences in cohort life expectancies estimated based on the different EMT methods and the AMT. Interestingly, the optimal weighting parameters do not imply the largest probability mass at zero for the life expectancy differences. In that respect, the longer data series, i.e. 10 or 20 years, are favorable. This is not surprising as – in case no trend change occurs during the last years of the simulation – a longer data series for the EMT estimation leads to the EMT being closer to the AMT in general. However, if a trend change occurs in the last years of the simulation the longer data series for the EMT estimation leads to a larger error in the life expectancy estimation. This is why we observe rather fat tails for the approach with 20 years of data in particular. Here, we are back at the general trade-off in the EMT estimation between capturing trend changes as quickly as possible but also estimating linear trends as exactly as possible. The approaches with optimal weights, increased weights, and average weights seem to provide a more balanced solution to this trade-off problem than the approach with 20 years of data. They have slightly less probability mass at zero, but significantly more mass close to zero as well as thinner tails.

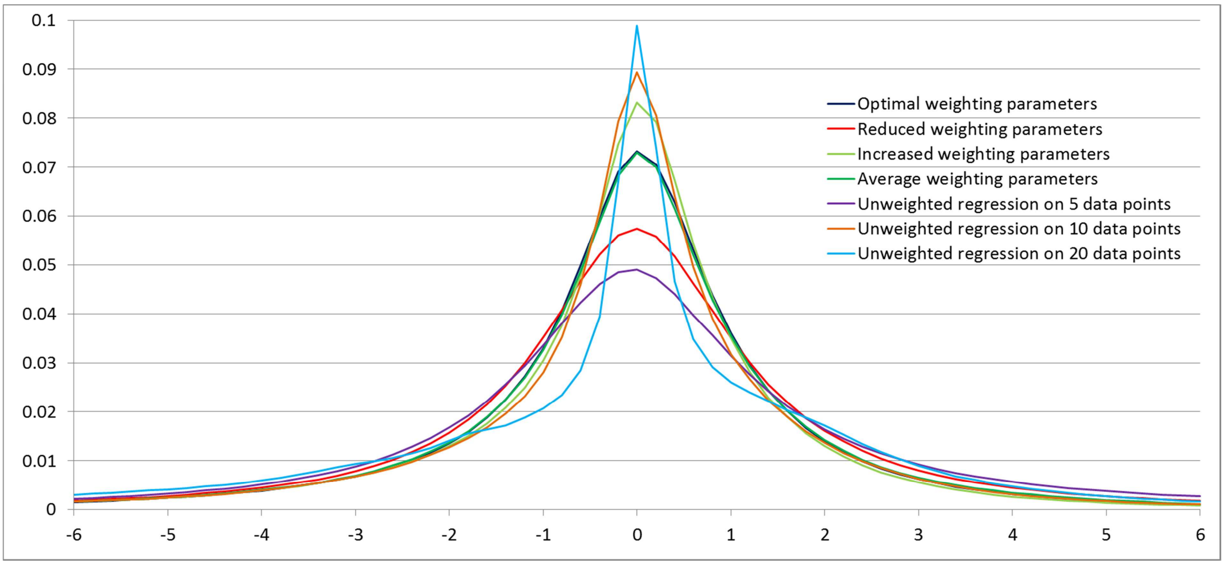

Figure 5: Empirical probability density functions of errors in estimation of expected cohort life expectancies

In order to be able to compare the EMT estimation approaches by a single number, the mean squared errors (mse) in estimating the cohort life expectancy are provided in Table 4: Errors in cohort life expectancy estimations. As expected, the mse is the largest for the unweighted regression with 20 data points. But also for 5 data points, the mse is significantly larger than for the remaining 5 approaches. Here, the data series is obviously too short and the EMT estimate is strongly exposed to the random fluctuations in the κ processes. If one wanted to use a fixed time interval for estimating the EMT, the optimal length of this interval would be 8 years. The mse would be 5.56738 in that case which is still (significantly) larger than for the weighted regression approaches with optimal, increased, or averaged weighting parameters. Since the mse for the case of increased weighting parameters is significantly smaller than for the case of reduced weighting parameters the EMT estimation seems to be less sensitive to an increase of weights than to a reduction.

EMT estimation method	Mean squared error	Root mean squared error
Optimal weighting parameters	5.136	2.266
Reduced weighting parameters	6.666	2.582
Increased weighting parameters	5.190	2.278
Average weighting parameters	5.312	2.305
Unweighted regression on 5 data points	10.188	3.192
Unweighted regression on 10 data points	5.801	2.409
Unweighted regression on 20 data points	11.682	3.418

Table 4: Errors in cohort life expectancy estimations based on different approaches for the EMT

In Table 4, we also show the root mean squared error as a measure of the “average” error in the estimation of the cohort life expectancy. This quantity nicely illustrates the impact of different choices of the EMT estimation approach for practical purposes. Let us consider a financial derivative which pays, at its maturity, a sum which is proportional to the remaining cohort life expectancy of a 60-year old as it would be estimated at maturity of the derivative.[11] The root mean squared error tells us how much the actual payoff of the derivative (based on the EMT), “on average”, deviates from the intended payoff, i.e. a payoff according to the AMT. When estimating the remaining life expectancy based on the EMT approach with optimal

[11] In practice, a derivative is more likely to pay according to the present value of an annuity. But since the annuity present value and the remaining cohort life expectancy are rather similar in structure our arguments should hold for both quantities.

weighting parameters, the derivate would pay for 2.266 years more or less than intended. In case of an EMT estimation based on the last 20 years of data, this value increases to 3.418, i.e. by more than 50%. Thus, if a pension fund bought such a derivative to hedge its longevity risk, in both cases, the derivative payoff might be smaller than required to cover future liabilities. But the gap between actual payoff and intended payoff significantly depends on the estimation approach for the EMT. To further illustrate this, the risk of underestimating the remaining life expectancy by 5 years or more is 7.7% in case of the unweighted regression on 20 years of data compared to only 3% in case of the weighted regression with optimal weighting parameter. The respective probabilities of misestimating the life expectancy by 10 years or more are 2.1% and 0.4%. This clearly shows how much the EMT estimation approach can impact the effectiveness of risk mitigation tools. Thus, the estimation approach for the EMT should always be chosen carefully and we recommend using a weighted regression approach.

5. Conclusions

In this paper, we have explained that for virtually all questions that require stochastic mortality modeling, a clear distinction between actual mortality and estimated future mortality is required. In particular, for models with a stochastic mortality trend, at any point in time the actual mortality trend that is being modeled and the estimated mortality trend that an observer would estimate based on the development up to that point in time are different.

We specified a concrete model where the AMT is modeled as a piecewise linear function. We identified and overcame weaknesses of existing trend chance models. In particular, we found that a simple random walk with drift systematically underestimates future uncertainty whereas, e.g., the model by (Sweeting, 2011) which is the basis for our approach, has some inconsistency between the assumed distribution of future trend change intensities and the calibration of trend change probabilities which leads to a significant overestimation of future uncertainty.

Since the start trend (i.e. the AMT at the start of the simulation) is unknown, just using today's estimate (i.e. today's EMT) as a start trend ignores one source of uncertainty. We therefore proposed to consider this uncertainty in the model by introducing a stochastic start trend.

Finally, we compared different approaches for deriving the EMT from observed actual mortality, concluded that a weighted regression appears suitable, and derived optimal weighting parameters.

Our approach opens potential for further research: First, it would be interesting to analyze, whether an application of our model to data from different countries leads to similar results, in particular whether trend changes in different countries are correlated. Also, our data indicates that including cohort effects and/or stochastic volatility in the model might be worthwhile. It would be interesting to analyze, how other parameters, e.g. optimal weighting parameters would chance in a model with stochastic volatility.

Appendix

A AMT Model Estimation and Testing for Trend Breaks

In this Appendix, we describe in detail the Chow test which we use to determine changes in the AMT. In general, the Chow test is a test to determine whether two separate regression lines describe some 2-dimensional data significantly better than only one regression (see (Chow, 1960)). The test statistic is

$$T = \frac{(RSS_C - RSS_1 - RSS_2)/k}{(RSS_1 + RSS_2)/(N_1 + N_2 - 2k)},$$

where RSS_C denotes the residual sum of squares in case only one regression line is fitted to the data, and RSS_1 and RSS_2 are the residual sums of squares for the two regression lines. The parameters N_1 and N_2 refer to the respective number of data points to which the two regression lines are fitted, and k is the number of parameters in the fitted model, i.e. $k = 2$ in case of a simple regression. This test statistic T follows an F-distribution with k and $(N_1 + N_2 - 2k)$ degrees of freedom.

For our situation, we need to slightly modify the Chow test. In the original test, the two lines are fitted independently of each other which is not feasible in our case. As we require a single continuous line (with changing slope) to describe the whole data set (see Figure 3), the two regression lines in the Chow test need to intersect. Moreover, the one or two (intersecting) regression lines cannot be fitted independently of the rest of the data set. They need to have common points with the previous straight line and the subsequent straight line in the years of the previous and the next trend change, respectively. To allow for this, we always fit a continuous line to the whole data set and allow for changes in the slope in those years where we assume trend changes. Thus, for a Chow test for a trend change in a certain year, two continuous lines have to be fitted: one which includes the trend change to be tested and one which does not. The first fit corresponds to the case of two lines describing the data between the previous and the next trend break (in relation to the trend change which is tested) and the second fit corresponds to the case of only a single line describing the data between the previous and the next trend change. The residual sums of squares RSS_C, RSS_1, and RSS_2 in the Chow test are then the respective components of the residual sums of squares of the overall fits. Note that due to the constraints imposed on the regression lines, the test statistic T does not have to follow an F-distribution exactly anymore. However, this should still be the case approximately.

Since we always fit a continuous line to the whole data set, we are highly exposed to heteroscedasticity in the data. For both time series, we observe significantly larger volatility in the earlier data points (see Figure 2) which implies the need for weighting in the least squares fitting of the continuous lines. Therefore, we fit a regression line to each set of 7 adjacent data points and use the inverse of the sample variance as weight for the central of the 7 data points. For the first 3 data points we assume the weight from the 4th data point and for the last 3 data points we use the weight of the 4th from last data point. This is the same approach as in (Sweeting, 2011).

Even though we use the same model as (Sweeting, 2011) in order to describe the historical mortality trend evolution, and find about as many trend changes as he does – Sweeting finds 7 trend changes for κ_1 and 8 trend changes for κ_2 – the estimation approaches differ significantly. Sweeting uses the Durbin Watson statistic to determine the location of potential trend changes and then drops those candidate years again for which tests do not clearly indicate a significant trend change. For these tests, he uses data which, in relation to the candidate year he is testing, starts at the candidate year for the previous trend change and ends at the candidate year for the next trend change. In case the exact location of a trend change is unclear, i.e. if there are several adjacent years for which the tests have indicated significant trend changes, he picks the year for which the p-value of the Chow test is minimized.

This approach is suboptimal from our point of view for two reasons. First, the choice of the data periods for the tests is problematic as the significance of a trend change strongly depends on the (length of the) time periods before and after the trend change. For instance, if one considered the whole time series for κ_1 (see Figure 2), a significant trend change could be observed for basically any year, so also for, say, 1880. However, if one uses a shorter time interval, e.g. from 1841 to 1890, there is certainly not a significant trend change in 1880. Therefore, the significance of a trend change should not be assessed between any two candidate years which may be omitted or shifted later on. Instead, significance tests should be carried out for a complete and fully fixed set of trend changes years. In case a trend change year is altered, the significance tests for all years should be repeated. Secondly, we regard the exact positioning of trend changes by minimization of the Chow test p-value as suboptimal. We think this task should be embedded in the model fitting, i.e. the trend change years should be chosen such that the residual sum of squares is minimized.

References

Börger, M., 2010. Deterministic Shock vs. Stochastic Value-at-Risk - An Analysis of the Solvency II Standard Model Approach to Longevity Risk. *Blätter der DGVFM*, Volume 31, pp. 225-259.

Börger, M., Fleischer, D. & Kuksin, N., 2012. Modeling Mortality Trend under Modern Solvency Regimes. *Working Paper, Ulm University, Ulm*.

Cairns, A., 2012. Robust Hedging of Longevity Risk. *Working Paper, Heriot-Watt University Edinburgh*.

Cairns, A., Blake, D. & Dowd, K., 2006. A Two-Factor Model for Stochastic Mortality with Parameter Uncertainty: Theory and Calibration. *Journal of Risk and Insurance*, Volume 73, pp. 687-718.

Cairns, A., Dowd, K., Blake, D. & Coughlan, G., 2011. Longevity Hedge Effectiveness: A Decomposition. *Working Paper, Heriot-Watt University, Edingburgh*.

Chow, G., 1960. Tests of Equality Between Sets of Coefficients in Two Linear Regressions. *Econometrica*, Volume 28, pp. 591-605.

Lee, R. & Carter, L., 1992. Modeling and Forecasting U.S. Mortality. *Journal of the American Statistical Association*, Volume 87, pp. 659-671.

Lee, R. & Miller, T., 2001. Evaluating the Performance of the Lee-Carter Method for Forecasting Mortality. *Demography*, Volume 38, pp. 537-549.

Li, J., Chan, W. S. & Cheung, S. H., 2011. Structural Changes in the Lee-Carter Mortality Indexes: Detection and Implications. *North American Actuarial Journal*, Volume 15, pp. 13-31.

Plat, R., 2009. On Stochastic Mortality Modeling. *Insurance: Mathematics and Economics*, Volume 45, pp. 393-404.

Plat, R., 2011. One-year Value-at-Risk for Longevity and Mortality. *Insurance: Mathematics and Economics*, Volume 49, pp. 462-470.

Sweeting, P., 2011. A Trend-Change Extension of the Cairns-Blake-Dowd Model. *Annals of Actuarial Science*, Volume 5, pp. 143-162.

Zhu, N. & Bauer, D., 2011. Coherent Modeling of the Risk in Mortality Projections: A Semi-Parametric Approach. *Working Paper, Georgia State University, Atlanta*.