Trend Processes in Mortality Models and Management of the Longevity Risk

Dissertation zur Erlangung des akademischen Grades eines Doktors der Wirtschaftswissenschaften (Dr. rer. pol.) an der Fakultät für Mathematik und Wirtschaftswissenschaften der Universität Ulm

Vorgelegt von:
Johannes Schupp

Amtierender Dekan:
Professor Dr. Martin Müller

Gutachter:
Professor Dr. Hans-Joachim Zwiesler
Professor Dr. Jochen Ruß

Tag der Promotion:
03.03.2020

Acknowledgments

Many people played essential roles in the development of this dissertation. First of all, I would like to call my partner Lisa Hirschauer. Thank you for being a reliable source of support, confidence, and love. This work would not have been possible without you, and every day I'm glad to have met you on the train once upon a time.

I owe special gratitude to my supervisors Hans-Joachim Zwiesler and Jochen Ruß. They always had the time and joy to guide and assist me in completing this work and contributed essentially to the quality of this result and to my employer ifa for providing an excellent framework for this dissertation.

I owe special gratitute to Matthias Börger for supervising this dissertation. The countless hours of creative and fruitful thinking and discussion that finally led to this work will always be remembered in gratitude. I am very thankful to Martin Genz, Arne Freimann and Justin Schoenfeld for discussions on topics related to the modelling of trends in parametric mortality models.

Ulm, May 2020 Johannes Schupp

Contents

Overview of Research Papers

Research papers included in this dissertation

1. Börger, M. and Schupp, J. (2018). Modeling Trend Processes in Parametric Mortality Models *Insurance: Mathematics and Economics*, 78:369–380.
2. Schupp, J. (2019). On the Modeling of Variable Mortality Trend Processes. Working Paper, Ulm University.
3. Börger, M., Ruß, J., and Schupp, J. (2019). It Takes Two: Why Mortality Trend Modeling is more than Modeling one Mortality Trend. Working Paper, Ulm University.
4. Börger, M., Schoenfeld, J., and Schupp, J. (2019). Calibrating Mortality Processes with Trend Changes to Multi-Population Data. Working Paper, Ulm University.

Co-Authorship

Professor Dr. Jochen Ruß

Jochen Ruß is managing partner of the Institut für Finanz- und Aktuarwissenschaften (ifa) in Ulm. The focus of his consulting work is on the development of innovative life insurance products in particular with new forms of guarantees, as well as actuarial topics in the area of traded endowments and life settlements.
Jochen is also adjunct professor for actuarial science at Ulm University and a lecturer at Ludwig-Maximilians-Universität Munich. Furthermore, he is a member of the German Society for Actuarial and Financial Mathematics, the American Risk and Insurance Association and the Asia-Pacific Risk and Insurance Association and associated member of the Munich Risk and Insurance Center.

Professor Dr. Matthias Börger

Matthias Börger is a professor at the Technical University Nuremberg. His main area of expertise is insurance risk management, with a focus on the modeling and management of biometric risks, in particular longevity risk. Matthias holds a PhD in actuarial mathematics from Ulm University. He is also a CERA and a member of the German Society of Actuaries.

Justin Schoenfeld

Justin Schoenfeld is a master student at Ulm University. He holds a Master of Science degree in mathematics from Ulm University.

Research Context and Summary of Research Papers

1. Field of Research

Reliable predictions of future mortality are important for private insurance companies, social security systems and pension funds as otherwise reserves or safety margins for annuities or pensions can be insufficient. This risk, i.e. the risk of insured living longer than expected, is called longevity risk. This cumulative thesis contributes to the literature on this risk and how it can be modeled and quantified.

For decades, life expectancies have been improving significantly in most developed countries. Oeppen and Vaupel (2002) analysed the worldwide record life expectancy and found an extraordinarily constant improvement of approximately 2.5 years per decade for the last 160 years. However, the underlying drivers have varied over time. In the late 19th and early 20th century, dominant improvements in infant mortality can be observed. Throughout the 20th century on the other hand, medical improvements led to a decline of mortality corresponding to several illnesses, e.g. cancer or cardiovascular diseases. Also changes in the socio-economic behaviour as the decline of smoking improved the death probabilities of adults. In recent years, however, mortality rates in some countries, e.g. United States (US), England & Wales, Taiwan, and Germany, have improved at an unexpectedly slower pace. In the US and Canada, mortality for some age segments, for example, is even on the rise due to the increase of deaths caused by opioid drug overdoses, see Holman et al. (2019) and Scholl et al. (2019).

These positive (and negative) developments and the corresponding risk drivers and also their impact became apparent rather slowly. Therefore, the aggregation of this potpourri of causes led to a – in a mathematical sense – continuous mortality evolution over time, which is most commonly called mortality trend. Remarkably, there have been several changes in the mortality trend over time. Periods with a rather steep mortality trend have been followed by periods with a flatter mortality trend and vice versa. The question of whether further improvements in longevity are possible and to which amount is highly controversial. Also in the scientific community the future evolution of life expectancies is an controversially discussed question. One group of scientist – 'the pessimists' – say that the required medical advances would be to extensive to achieve significant further improvements in longevity, see Olshansky et al. (2005). The opposite group of researches around the gerontologist Aubrey de Grey – 'the optimists' – say that there is virtually no limit to human life span as soon as research focusses more on ageing and how to decelerate it than on disease control.

Expected future changes in the life expectancy can be incorporated straightforwardly in the pricing of annuities or pensions. The inclusion of expected changes in the life expectancy does not pose any risk itself, as the increased life expectancy simply leads to correspondingly adjusted annuity conversion rates in an otherwise constant environment. However, pensions from private insurance or from social security systems traditionally include a guaranteed annuity level or guaranteed annuity conversion rates. If actual life expectancy for all retirees on average is significantly higher than the original expectation - on the basis of which the guarantee was issued - then the insurers need to hold additional reserves of a significant amount.[1] The effect on required reserves is huge: every additional year of unexpected life expectancy improvement increases the reserves by 3-5%, see OECD (2014). If the uncertainty is not taken into account appropriately and a case of unexpectedly high longevity occurs, this may result in critical implications, e.g. pension cuts.

Longevity risk has also important implications for the diversification of risks of an insurance company. The principle of diversification via pooling of risks across a large number of insured fails as improvements in life expectancies have a comparable impact on all insured persons - as simply everybody will live on average a little longer. As explained before, the drivers of increasing longevity lead to long term trends, and hence, also the diversification over time is not possible. Therefore, longevity risk is a systematic risk that needs to be quantified and thoroughly managed. The extensive effects of underestimating life expectancy make longevity risk a major insurance risk.

An adequate quantification and management of the risk of underestimating the life expectancy is of general public interest as the implications for individuals can be ruinous. If a scenario of unexpectedly increased longevity appears, pensions from different sources can be affected in a similar way. This cumulative influence emphasize the importance of longevity risk and the modeling of future mortality trends.

2. Motivation and Objectives

Demographers or actuaries can use models to estimate the future evolution of mortality. This central estimate is called best estimate scenario. From a risk management perspective, it is not sufficient to focus only on such a best estimate scenario. Instead, the uncertainty surrounding

[1] The low interest rate environment intensifies this effect substantially.

this best estimate scenario must also be considered in order to get a complete picture of the longevity risk involved. A sophisticated analysis of the longevity risk of social security systems, insurance companies or pension funds therefore particularly requires stochastic scenarios of future mortality, where all possible scenarios should be considered. In scenarios of extraordinarily decreasing mortality rates, the required reserves for a pension or annuity portfolio increase significantly. A risk adequate representation of the uncertainty about future mortality in risk management systems is therefore of fundamental interest of all involved stakeholders. As part of the Solvency II directive, this was also recognized and so there is also a panel for longevity risk.

The historical evolution of mortality has been analysed by several authors. Trend changes in the underlying mortality trend were repeatedly identified among others by Li et al. (2011), Chen and Cox (2009) and Coelho and Nunes (2011). Since historical trend changes can be observed, it seems necessary that stochastic trend changes are possible in the future as well. This necessity was incorporated in only a few approaches. For instance, Milidonis et al. (2011) and Hainaut (2012) included stochastic regime switches but limited their simulations to only two possible scenarios of future mortality. This limitation seems too restrictive as also the historical evolution of mortality was multifaceted. Sweeting (2011) and Hunt and Blake (2015b) included random trend changes in their simulation models of future mortality, but had several shortcomings in the estimation. This leads to the research question:

Research Question 1: How can we specify a stochastic trend model that is consistent with historical observations of trend changes? How can we calibrate this model based on data on historical trend changes and does this model generate reasonable forecasts of future mortality?

Existing trend models estimate historical trend changes based on observed data and parametrize the simulation models to generate a distribution for future mortality. However, the estimation of the historical trends contains a considerable amount of uncertainty. Moreover, in many cases, the magnitude of historical trend changes is rather inhomogeneous. All these aspects make the calibration of historical trend changes very difficult. Due to the sparsity and inhomogeneity, the calibration algorithms may have difficulties to estimate robust trends. Hence, this issue leads to the second research question:

Research Question 2: How can we specify a stochastic trend model that can handle the sparse data about historical trend changes? How can we modify and further develop the calibration of models with changing mortality trends to include the distribution of future trend changes in the calibration of the historical trend changes?

The majority of existing mortality models have been developed to project the actual mortality of an underlying population. However, in reality the actual mortality trend is not directly observable as we can only estimate the mortality trend based on the observable mortality data. The Human Mortality Database or national statistical offices like destatis offer publicly available information about national populations. Here, annual noise as a result of e.g. heat waves or strong influenza seasons influence the annual number of deaths significantly. A mortality trend can be estimated from this mortality data. But it is important to emphasize that this trend can only be an estimate and not the actual, unobservable trend. The estimate can be blurred by the volatile mortality data. These two trends might differ even more extremely in the presence of trend changes. Depending on the question at hand one or the other (or even both trends simultaneously) need to be used. If we do not distinguish, conclusions based on such trend assumptions may be inaccurate or misleading.

In the academic literature, this distinction has not been described so far. For instance, Richards et al. (2014) do not distinguish and use the Lee and Carter (1992) model to quantify the solvency capital requirement for longevity risk under Solvency II, where both the actual mortality trend and the estimated mortality trend are required. The best estimate assumptions of Solvency II always require an estimated mortality trend. Additionally, the potential change over one year is required, i.e. the potential development of the estimated mortality trend over one year needs to be analysed. With one year of additional information about deaths the estimated mortality trend can change. In a simulation, this additional year of information is generated by the actual mortality trend. Hence, a model with two somehow linked trend components is required. Also in many other situations, e.g. for the analysis of longevity hedges, a model that incorporates and connects the actual mortality trend with the estimated mortality trend is necessary. These requirements lead to the third research question:

Research Question 3: Why is it necessary to differentiate between two mortality trends? In the presence of trend changes, how can we link a model for the actual mortality trend with a model for the estimated mortality trend? What is the consequence from a risk management perspective if we clearly differentiate between the two mortality trends?

Unfortunately, data on trend changes is sparse and not sufficient in some cases to make reliable forecasts. These arguments also apply to insurance portfolios. Here, in addition, the smaller sub-population of insured in a portfolio compared to the general population makes it more difficult or sometimes impossible to estimate trends reliably. Even for countries with very long observation histories typically only a few clear trend changes can be identified. In most countries, there are only two to four historical trend changes. The intuitive approach to handle sparse data is to increase the sample size somehow. Closely related populations are exposed to the same fundamental drivers and changes in mortality, and therefore, it seems promising to aggregate data on mortality trend changes from different populations in any case. Moreover, a thorough analysis of the parameter uncertainty from a single population perspective and after the aggregation of data from different populations is necessary. These issues lead us to the forth research question:

Research Question 4: How can we quantify the parameter uncertainty in the presence of sparse trend changes? How can we use data on trend changes from several populations to increase the confidence about forecasts for a single population? What is the influence on parameter uncertainty when data about trend changes is aggregated?

As longevity risk constitutes a material risk for insurers, pension funds, but also for society, the aforementioned research questions are clearly relevant from a practical perspective. The four papers included in this dissertation contribute to the literature about longevity risk and significantly improve the modeling and quantification of this risk.

3. Summary of Research Papers

Research Paper 1: Modeling Trend Processes in Parametric Mortality Models

This paper is joint work with Matthias Börger and it has been published in volume 78 of *Insurance: Mathematics and Economics* in 2018. It discusses trend processes for the forecasting of time dependent parameters in parametric mortality models like the Lee and Carter (1992) or Cairns et al. (2006) model. These stochastic forecasts can then be used to derive stochastic scenarios of future mortality. Reliable projections are essential for a reliable quantification and a thorough management of longevity risk. Stochastic forecasts are typically derived by a random walk with drift. However, the historical patterns of these time dependent parameters show clear signs for several trend changes which cannot be incorporated in a random walk with drift model. And hence, a back test shows, that in a long run, the long-term uncertainty is insufficiently incorporated. Therefore, also other authors analysed the historical evolution closely and proposed models that include historical trend changes, see among others Sweeting (2011), Lemoine (2014), and Coelho and Nunes (2011).

Many of these models show a structural shortcoming: they project the future mortality trend by extrapolating the most recent trend assumption. This seems only reasonable for a central estimate. With the explicit goal to incorporate also the uncertainty around this central estimate, however, this seems insufficient. In the presence of several clear historical trend changes, it seems necessary to include potential further trend changes also in forecasts. This shortcoming was addressed in the model of Sweeting (2011) by including also stochastic trend changes in the future. We identify several inconsistencies in the trend model of Sweeting (2011) and therefore propose an alternative trend model. Stochastic forecasts require assumptions about the distribution of future trend changes. The paper legitimizes the assumptions made on the basis of historically observed trends.

The paper continues by presenting an approach to account for the significant estimation uncertainty by adopting the concept of Bayesian weights, see Burnham and Anderson (2002). The estimation of trends typically involves different sources of uncertainty. First, the trend estimation is disturbed by annual fluctuations around the trend due to heat waves, strong winters, influenza, etc. Therefore, it is possible to misestimate the actual number and the magnitude of historical trend changes. The second source analysed is the uncertainty about the latest trend, i.e. the trend a simulation starts with. If the starting values of a simulation indicate a too flat (steep) mortality trend, the central estimate of predictions will be too small (high). Simply due to the limited data after a potential last trend change, it is possible that this trend change is not identified. This uncertainty cannot be incorporated by existing approaches based on resampling (see Koissi et al. (2006)). The presented approach allows to account for both sources of parameter uncertainty efficiently.

The last section of the paper compares the new trend model with different approaches with and without trend changes that have been proposed in the literature. The forecasts of the remaining period life expectancies of 60-year old English and Welsh males vary significantly in terms of both central projection and uncertainty around the central projection. Compared to other approaches, the proposed trend model generates highly plausible forecasts for the remaining period life expectancy. This shows that model risk with respect to the trend process is significant. This should be of particular interest for everybody who quantifies longevity risk of a pension or annuity portfolio. For instance, reserves of a portfolio of (deferred) annuities differ significantly depending on the model used to project scenarios of future mortality.

Research Paper 2: On the modeling of Variable Mortality Trend Processes

This paper addresses the modeling of trend processes with trend changes with a focus on a consistent historical calibration. For this purpose, the paper analyses and further develops existing calibration approaches.

A significant amount of models aims to identify the historical trend changes with data of several populations (see Booth et al. (2002), Coelho and Nunes (2011), Berkum et al. (2014), Li et al. (2011), and Chen and Cox (2009)). Most of these models then extrapolate the most recent trend into the future. Also, a smaller amount of models (see Hunt and Blake (2015a), Sweeting (2011), Milidonis et al. (2011), Lemoine (2014), and Börger and Schupp (2018)) include potential further trend changes in simulations, accordingly. The aforementioned approaches initially estimate optimal historical trends, e.g. continuous piecewise linear trends with trend changes once in a while, and then estimate a distribution for future trend changes based on the few historical trend changes in a separate second step.

Under the fundamental assumption that the trends of the past are to some degree meaningful for the future trends, it is important that the future trends have patterns similar to those of the past. In many cases, however, the magnitude of historical trend changes is rather inhomogeneous. When deriving a distribution from only a few and inhomogeneous observations, it may happen that the simulated future trend changes show patterns significantly different to those of the historical trend changes.

Therefore, it seems reasonable to consider the distribution of future trend changes already in the calibration of historical trends. The paper closes this gap and links past and future trend changes more closely. Hereby, simulated mortality trend changes can be closer to estimated historical trend changes.

First, existing models with trend changes are analysed with respect to the already included distributional assumptions. Then an extension is presented that links the estimation of historical trend changes with the distribution for future trend changes. The presented approach is very efficient as the distribution parameters for future simulations of trends are already estimated as part of the historical trend calibration. Using examples, the paper shows how the inclusion of distributional assumptions affects the estimation of historical trends. In particular, it avoids the estimation of rather implausible historical trend changes. Under given distributional assumptions, the historical trend changes should be represented by that assumptions. Hereby, the estimated trends are smoother and the calibration itself gets more robust. Finally, parameter uncertainty is analysed.

If the connection between historical trend changes and simulated future trend changes fails, longevity risk may be significantly missestimated. This risk should be of particular interest for everybody who quantifies longevity risk with a mortality model with trend changes.

Research Paper 3: It Takes Two: Why Mortality Trend Modeling is more than Modeling One Mortality Trend

This paper is joint work with Matthias Börger and Jochen Ruß and it has been submitted to *Insurance: Mathematics and Economics*. It adresses the clear distinction of two mortality trends, the (unobservable) actual mortality trend (AMT) and the estimated mortality trend (EMT). An observer like an actuary in an insurance company will not be able to perfectly estimate the unobservable actual mortality trend from observed deaths and exposures. In addition, trends that have not been persistent for a longer time period are difficult to estimate, e.g. it can happen that a trend change in one of the most recent years has not yet been recognized.

In a simulation, the AMT is necessary for all quantities that are based on the actual mortality evolution, e.g. payouts of an annuity portfolio that are dependent on the actual number of deaths in the portfolio in a specific future year. If the simulated actual number of deaths (based on the AMT) is higher, the annuity provider has smaller payouts, and vice versa. Although the AMT is simulated it would not be observable by the actuary, and thus, it is still necessary to build a model for the EMT. The amount of reserves for a book of annuities or the annuity conversion rates of deferred annuities are dependent on the expectation about the mortality trend, i.e. the EMT is relevant. If the risk manager estimates a steep (flat) mortality trend with large (small) future improvements, the reserves will be higher (lower). In a simulation, this estimate changes year-by-year based on the prevailing (simulated) information about deaths and exposures. A sophisticated analyses of the longevity risk of an insurance company or a pension fund requires the simulation of both, the actual mortality trend, e.g. for the required payouts, and the estimated mortality trend, e.g. for the evolution of the reserves.

Depending on the applications at hand a model for the future development of the AMT or a model for the future development of the EMT is required, and frequently AMT and EMT need to be considered jointly. Therefore, it is necessary to distinguish between AMT and EMT carefully and a model is required that captures both mortality trends simultaneously and consistently. Otherwise, insurance risks and in particular longevity risk might be misinterpreted.

In the literature, this distinction of two trends has been mostly ignored. Most existing mortality models are models for the AMT.[2] However, these models are often also used when technically an EMT is necessary. In many examples, see, for example Richards et al. (2014), Cairns (2013), and Cairns et al. (2014), this EMT estimate is then treated merely as a recalibration of the AMT model based on the then available data. Hereby, they ignore the fact that there are numerous situations where both trends are needed at the same time and that it therefore can not be just a recalibration risk.

The paper proceeds by presenting the AMT model proposed by Börger and Schupp (2018) and further refined by Schupp (2019). The specification of the AMT implies that the prevailing AMT is always the (unobservable) best estimate for the future mortality trend. It therefore seems reasonable to estimate the EMT as a linear trend calibrated to the most recent available data. In the presence of trend changes however, it is necessary to discuss the length of the included data. Observations from earlier trends disturb the estimation of the EMT. If, on the other hand, an insufficient amount of observations is taken into account, the EMT can change

[2]There are also approaches that model the EMT, see among others Plat (2009), Bauer et al. (2010), Bauer et al. (2008), or Hunt and Blake (2015a)

significantly from one year to the other. Therefore, it is necessary to balance between these two extremes. In the paper, several efficient approaches to estimate the weights of a linear regression are derived, calibrated and tested. Compared to existing approaches, the results appear highly plausible.

The last section of the paper quantifies the importance of a clear distinction for the longevity risk. Three practical examples where mortality assumption are required show how important it is to distinguish between AMT and EMT. If the distinction fails, longevity risk is significantly missestimated in all three examples – sometimes underestimated, sometimes overestimated.

Research Paper 4: Calibrating Mortality Processes with Trend Changes to Multi-Population Data

This paper is joint work with Matthias Börger and Justin Schoenfeld and it has been accepted for presentation at the 2020 Living to 100 Society of Actuaries International Symposium, Orlando, FL (USA), where a conference monograph is published subsequently. It proposes possibilities to use data on trend changes from several populations to improve the reliability of the calibration for a single population.

The paper uses and further refines the trend model of Börger and Schupp (2018) and Schupp (2019). Especially, the modeling of the two main sources of parameter uncertainty are refined. The first category is the trend change parameters. These are the parameters of the distribution that can be used to simulate random future trend changes. The paper shows how calibrated trends with different numbers of historical trend changes can be combined to derive central estimates for the trend change parameters. In addition to the central estimate, the uncertainty around this estimate is analysed and decomposed in a sophisticated analysis. The second category is the starting values of a simulation. This uncertainty particularly arises from a potentially not detected additional trend change in one of the recent years. Certainty whether or not such a trend change occurred is gained once additional years of observations become available.

Subsequently, the parameters of the trend model are compared across calibrations for males and females worldwide with data from the Human Mortality Database. The trend change parameters show similar patterns across the 32 populations under consideration.[3] This substantial expansion of the information about trend changes means that these similarities between populations can be used to obtain much more reliable estimates also for individual populations.

[3]The uncertainty about the starting values shows highly population specific patterns. Here, an aggregation of parameter estimates does not seem plausible.

Consequently, in the paper six approaches to aggregate data about trend changes from several populations are proposed. Considering the significant uncertainty, it is justifiable to assume that trend change parameters are actually equal across populations. The first approaches aim to aggregate data to estimate this common unknown distribution. Furthermore, it is possible to relax the assumption such that trend change parameters are assumed to only come from the same distribution. In many cases one is confident about the population specific parameter estimate from the population one is particularly interested in, e.g. when the length of the data history is large. Then, it seems reasonable to increase the influence of standalone trend change parameters. For this case approaches based on credibility and based on Bayesian probability theory are presented.

The final section analyses the influence of different aggregation approaches on the forecasts for a single population (US males) exemplary. In this specific case, the central parameter estimates are significantly smaller and also the parameter uncertainty is reduced. Interestingly, also the contrary case can be possible when aggregation compensates for potentially underestimated parameters. The results of this paper should be interesting for everyone who aims to better understand the parameter uncertainty in a stochastic mortality model with trend changes.

References

Bauer, D., Börger, M., and Russ, J. (2010). On the Pricing of Longevity-Linked Securities. *Insurance: Mathematics and Economics* 46(1): 139–149.

Bauer, D., Börger, M., Russ, J., and Zwiesler, H. (2008). The Volatility of Mortality. *Asia-Pacific Journal of Risk and Insurance* 3: 184–211.

Berkum, F. v., Antonio, K., and Vellekoop, M. (2014). The impact of multiple structural changes on mortality predictions. *Scandinavian Actuarial Journal* 2016(7): 581–603.

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

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

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

Cairns, A. J. (2013). Robust Hedging of Longevity Risk. *Journal of Risk and Insurance* 80(3): 621–648.

Cairns, A. J., Blake, D., Dowd, K., and Coughlan, G. (2014). Longevity Hedge Effectiveness: A Decompositionn. *Quantitative Finance* 14(2): 217–235.

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.

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

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

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

Holman, R., MacDonald, C., and Miller, P. (2019). U.S. Population Mortality Observations - Updated with 2017 Experience. *Society of Actuaries.*

Hunt, A. and Blake, D. (2015a). Forward Mortality Rates in Discrete Time I: Calibration and Securities Pricing. *Working Paper, Cass Business Schooll.*

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

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

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

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

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

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

OECD (2014). Mortality Assumptions and Longevity Risk. Implications for Pension Funds and Annuity Providers. *OECD Publishing*. www.oecd.org/publications/mortality-assumptions-and-longevity-risk-9789264222748-en.htm.

Oeppen, J. and Vaupel, J. (2002). Broken Limits to Life Expectancy. *Science* 296: 1029–1031.

Olshansky, S. J., Passaro, D. J., Hershow, R. C., Layden, J., Carnes, B. A., Brody, J., Hayflick, L., Butler, R. N., Allison, D. B., and Ludwig, D. S. (2005). A potential decline in life expectancy in the United States in the 21st century. *New England Journal of Medicine* 352(11): 1138–1145.

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

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

Scholl, L., Seth, P., Kariisa, M., Wilson, N., and Baldwin, G. (2019). Drug and Opioid-Involved Overdose Deaths - United States, 2013–2017. *MMWR Morb Mortal Wkly Rep* 67: 1419–1427.

Schupp, J. (2019). On the Modeling of Variable Mortality Trend Processes. *Working Paper, Ulm University*.

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

Research Papers

1. Modeling Trend Processes in Parametric Mortality Models

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

URL: https://www.sciencedirect.com/science/article/pii/S0167668717304821
DOI: 10.1016/j.insmatheco.2017.09.024

Modeling Trend Processes in Parametric Mortality Models

Matthias Börger
Institute for Finance and Actuarial Sciences (ifa)
Lise-Meitner-Straße 14, 89081 Ulm, Germany
Phone: +49 731 20644257
Email: m.boerger@ifa-ulm.de

Johannes Schupp
Institute for Finance and Actuarial Sciences (ifa), Ulm & Institute of Insurance, Ulm University
Lise-Meitner-Straße 14, 89081 Ulm, Germany
Phone: +49 731 20644241
Email: j.schupp@ifa-ulm.de

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.

Keywords

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

Acknowledgments

The authors are very grateful to David Blake, Jochen Ruß and anonymous referees for comments and discussions which significantly improved the paper.

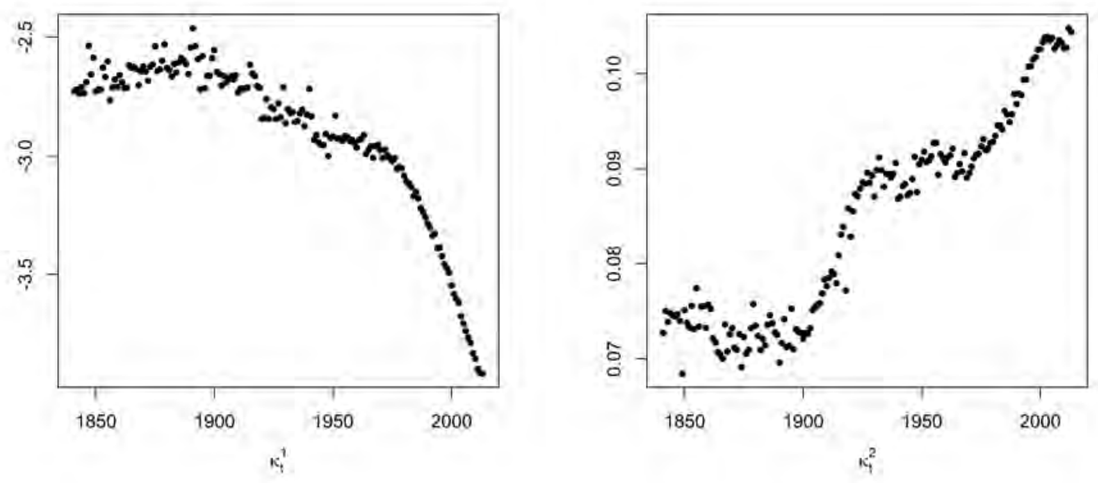


Figure 1.: Period effects in the CBD model for English and Welsh males

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 have been developed over the last decades. Most of them have a parametric structure which includes one or more time dependent parameters (period effects) 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 period effects 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.

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

[1] The CBD model is fitted to data from the Human Mortality Database (2018) for ages 50 to 89. For details on the model and the estimation of its parameters, we refer to Appendix A.

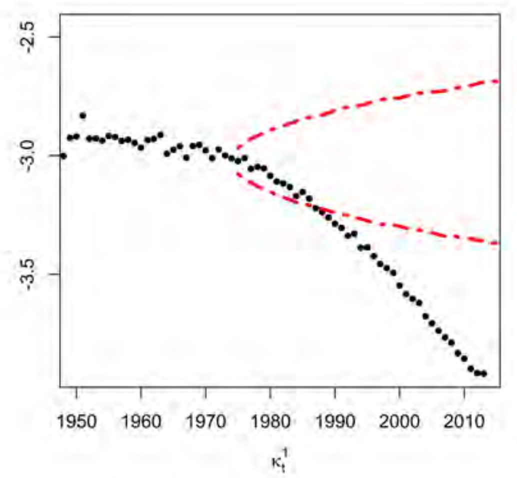


Figure 2.: Back test for period effect κ_t^1 in the CBD model for English and Welsh males; the period effect is projected by a random walk with drift and the dashed lines show the 99% confidence interval

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 period effects in other stochastic mortality models as well. One reason for that certainly is its simplicity. In its two-dimensional form, only five parameters need to be estimated from the historical data, i.e. the 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 period effects 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 period effects 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.

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

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 instead trend stationary around piecewise linear trends.

For the aforementioned reasons, we believe that the general adequacy of the random walk with drift for projecting period effects 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. Moreover, it does not include any time constant parameters as, e.g., in the mixed time and age term $\beta_x \cdot \kappa_t$ in the Lee and Carter (1992) model where the assumption of constant β_x over longer time horizons is questionable. Moreover, it does not include any time constant parameters as, e.g., in the mixed time and age term $\beta_x \cdot \kappa_t$ in the

Lee and Carter (1992) model where the assumption of constant β_x over longer time horizons is questionable.

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 period effect 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 (1992) 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 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 previously mentioned 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 (2015) 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 period effects in a "general procedure" model (see Hunt and Blake (2014) 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 period effects (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 period effect in a parametric mortality model (like the period effects in Figure 1) and let $\tilde{\kappa}_t$ be the "true" underlying period effect net of random fluctuations. Moreover, let d_t be the linear trend, i.e. the slope, of the period effect $\tilde{\kappa}_t$ 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 period effect 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 period effect:

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

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 period effects 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 period effect separately to ensure comparability. In the

[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.

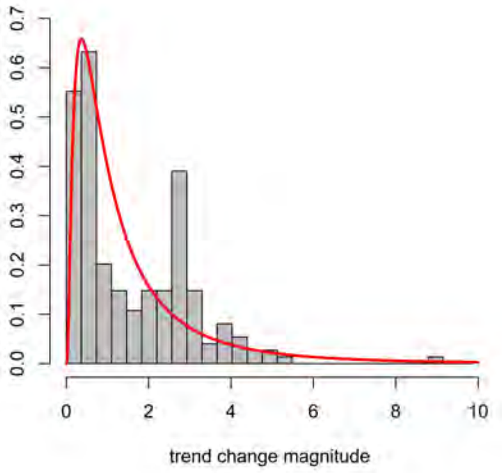


Figure 3.: Standardized historical trend change magnitudes in the CBD period effects for 15 countries and pdf of the standard lognormal distribution

figure, we see that the lognormal pdf and the histogram 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 (2015) apply a Pareto distribution. However, this requires specifying a minimum magnitude for future trend changes which we find impractical 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 period effect 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 period effects, 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 period effects occur quite frequently, 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 period effects 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 period effects 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 period effect individually since we assume that trend changes in the different period effects 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 period effects 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 period effect 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 period effect. 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 period effects 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 one exception, these are the countries for which reliable data is available in the Human Mortality Database (2018) 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 period effect we considered, 3 iterations were sufficient to obtain reliable parameter estimates. For the period effects 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 period effects in the CBD model for English and Welsh males as well as the optimal trend curves according to the AIC, the BIC, and the

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

[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.

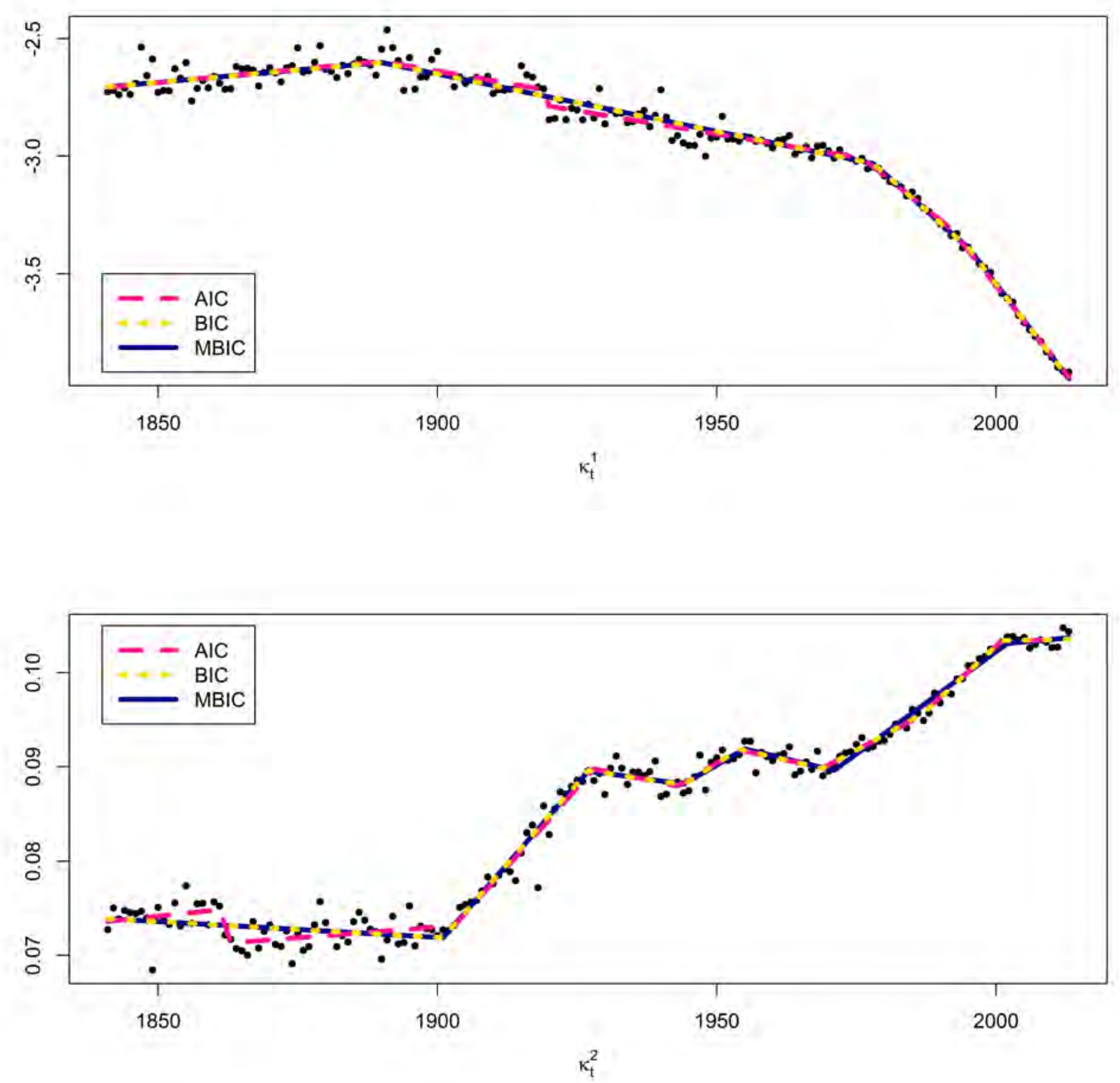


Figure 4.: Period effects in the CBD model for English and Welsh males and optimal trend curves according to different information criteria

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 period effect and every population we considered. The AIC, on the other hand, detects more trend changes for both period effects 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 spurious short term effects/trends which the method of Muggeo (2003) might detect 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 period effects like the κ_t^1 period effect 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 period effects 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 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 M. (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 period effects in the CBD model for English and Welsh males,

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

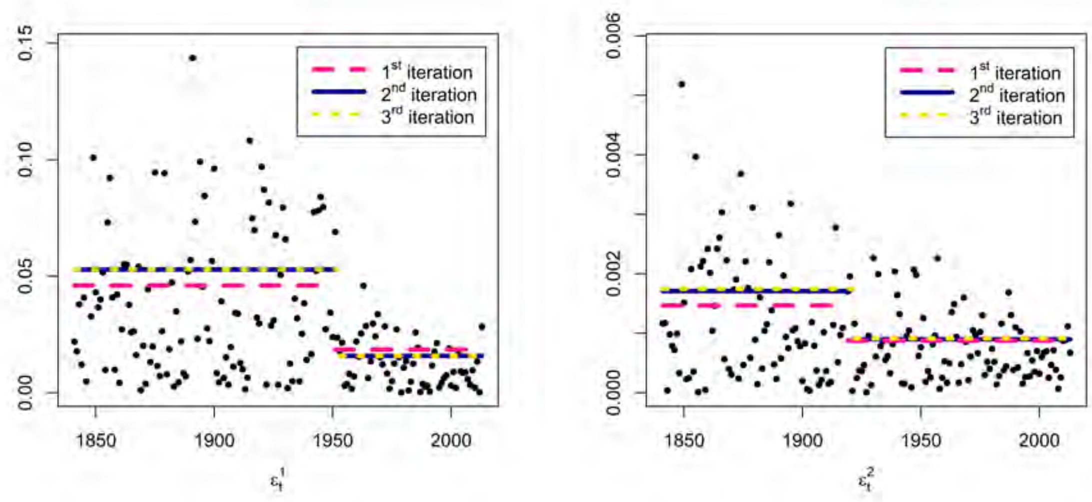


Figure 5.: Absolute values of residuals from the trend curve optimization and their standard deviation estimates in subsequent iterations, all for the period effects in the CBD model for English and Welsh males

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 period effects, 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 period effects changed for the last time. The historical decrease in variance is mainly driven by the reduction of deaths due to infectious diseases as well as increases in the population size under consideration. Therefore, the observed decrease in variance can be assumed to be permanent. Moreover, as outlined in Section 2, Σ_ϵ^2 only has a minor impact in our projection approach which justifies a rather simple calibration.

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

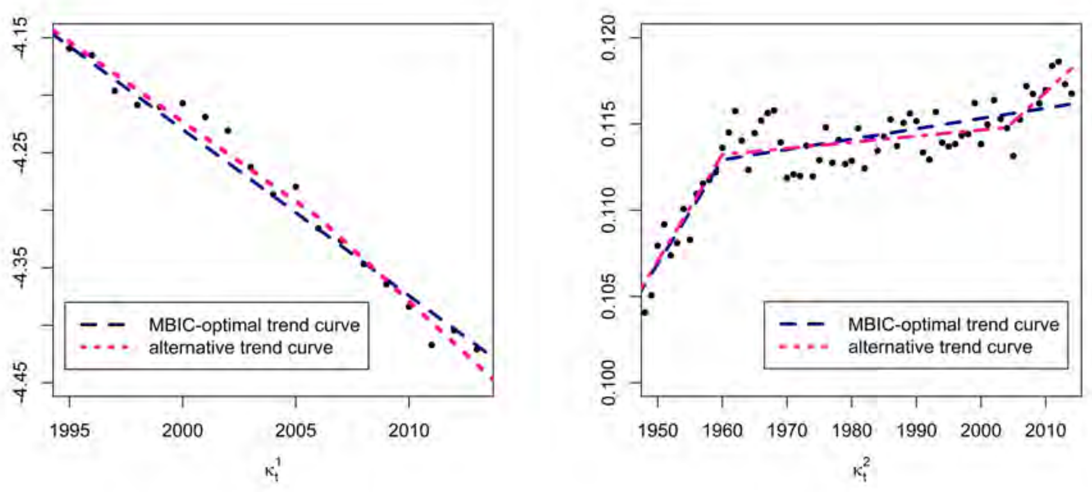


Figure 6.: MBIC-optimal and alternative trend curves for the period effects in the CBD model for Swedish females

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 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 period effects 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

[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.

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 period effect, i.e. if κ_t is one-dimensional. In the multi-dimensional case, we again propose assuming independence between the period effects. Thus, the same approach can be applied to each period effect, 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 [8]. 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 N is the number of years of available data after the most recent of the originally detected trend changes. For example, for κ_t^2, data is available up to 2013 with the latest trend change occurring in 2002. Thus, $N = 11$, $\tilde{\kappa}_{t_n,1}$ and $d_{t_n,1}$ denote the starting values for the case of an additional trend change in 2003, $\tilde{\kappa}_{t_n,2}$ and $d_{t_n,2}$ for the case of an additional trend change in 2004, etc. 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 period effects 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

[8]For simplicity, we only consider the impact of one additional trend change here. Scenarios of two or more additional trend changes are obviously possible but very unlikely in general.

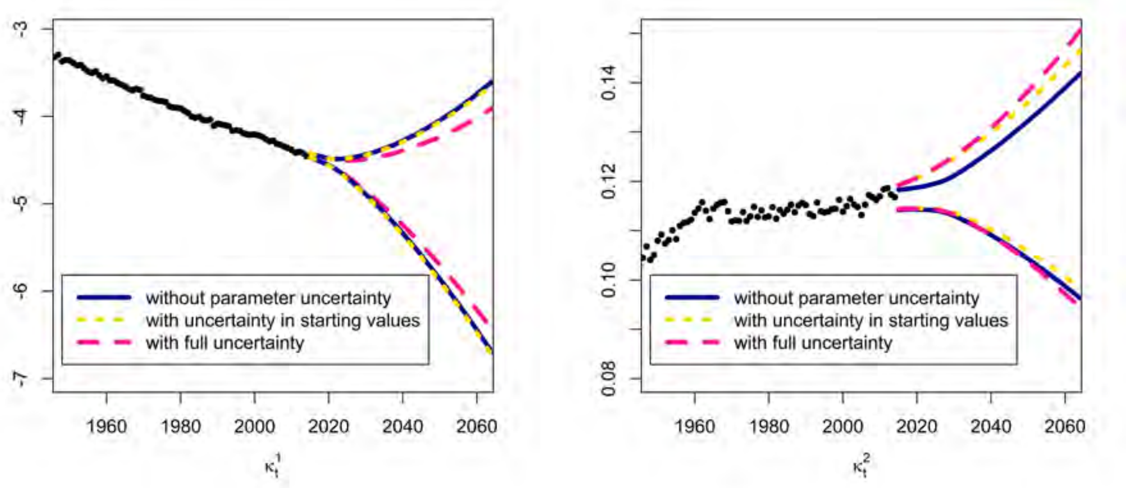


Figure 7.: 90% confidence intervals with and without parameter uncertainty for projections of the period effects in the CBD model for Swedish females

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 period effect and propose applying the same approach to each period effect 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.[9] 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

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

original parameter estimation. In order to assign probabilities to the different p_k, $\mu_{M,k}$, and $\sigma^2_{M,k}$, again the concept of Bayesian weights can be applied:

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

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 period effect 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 period effect 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 $\sigma^2_{M,\hat{k}}$ 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 accounted for.

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 period effects 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 (2015)
- 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 period effects 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.[10]

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 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

[10]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.

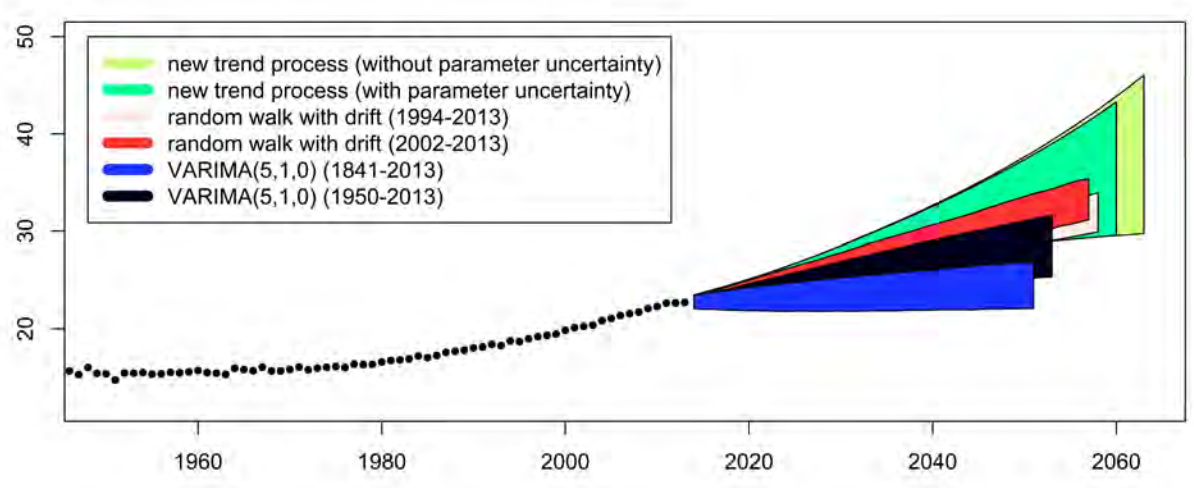


Figure 8.: 90% confidence intervals for projected remaining period life expectancies of 60-year old English and Welsh males; comparison between new trend process and time series processes

corresponds to the average change in the period effects 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 (2015) use random walks with changing drifts in their "general procedure" model, and we transfer their approach to the period effects 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 (2015) 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 (2015), 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 (2015) found two drift changes at most for any of the period effects 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 period effects 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 (2015), 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 (2015) 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 (2015), 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.

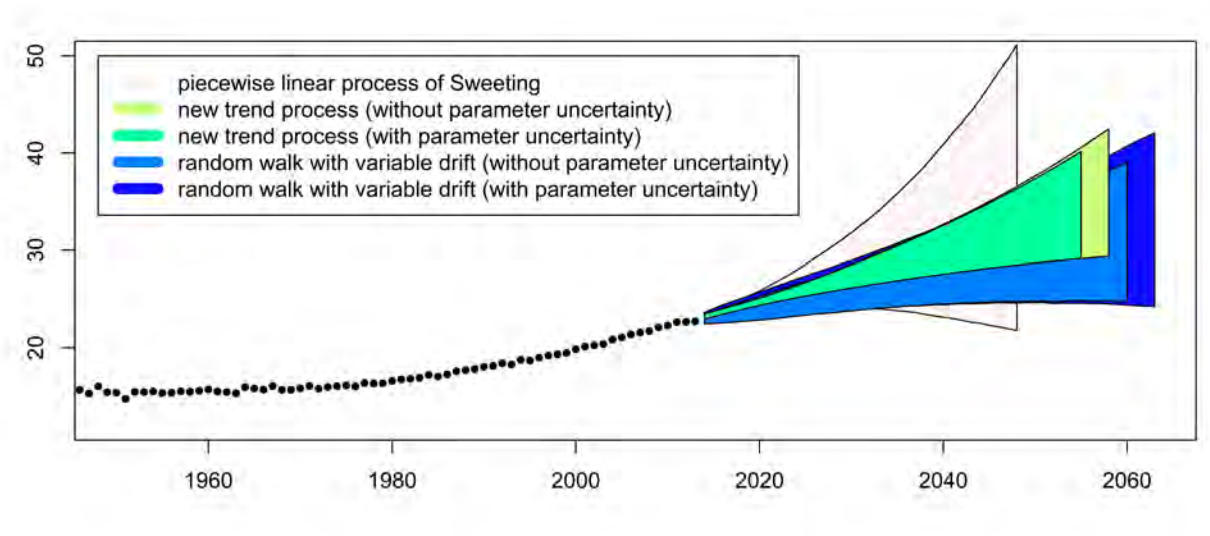


Figure 9.: 90% confidence intervals for projected remaining period life expectancies of 60-year old English and Welsh males; comparison between new trend process and alternative processes with variable drifts/trends

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 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 period effects 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 period effects. 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 period effect, 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.

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

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 period effect κ_t^1 describes the evolution of the general level of mortality rates over time, while the period effect κ_t^2 identifies changes in the slope of the mortality curve over time and thus differences in the mortality evolutions for different age 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 cannot 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 period effect 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 period effect. 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 period effect. 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 period effect 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 cannot 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. *Scandinavian Actuarial Journal* 2016(7): 581–603.

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

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

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

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

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

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

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

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

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

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.

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.

2. On the Modeling of Variable Mortality Trend Processes

Source:
Schupp, J. (2019). On the Modeling of Variable Mortality Trend Processes, Working Paper, Ulm University.

On the modeling of variable mortality trend processes

Johannes Schupp
Institute for Finance and Actuarial Sciences (ifa), Ulm & Institute of Insurance, Ulm University
Lise-Meitner-Straße 14, 89081 Ulm, Germany
Phone: +49 731 20644241
Email: j.schupp@ifa-ulm.de

Abstract

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

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

1. Introduction

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

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

These models take the historically observed trend changes to estimate a distribution for future trend changes. However, data on historical trend changes is sparse, and thus, the estimation of the simulation model contains a considerable amount of uncertainty. If the magnitude of historical trend changes is rather inhomogeneous, the simulation of future trend changes will no longer be consistent with the observed historical trend changes as those will be hardly reproduced by the distribution of future trend changes. In this paper, we avoid this inconsistency by including the distribution of future trend changes in the calibration of historical trends. We extend the calibration to a pure maximum likelihood estimate, including the distribution of

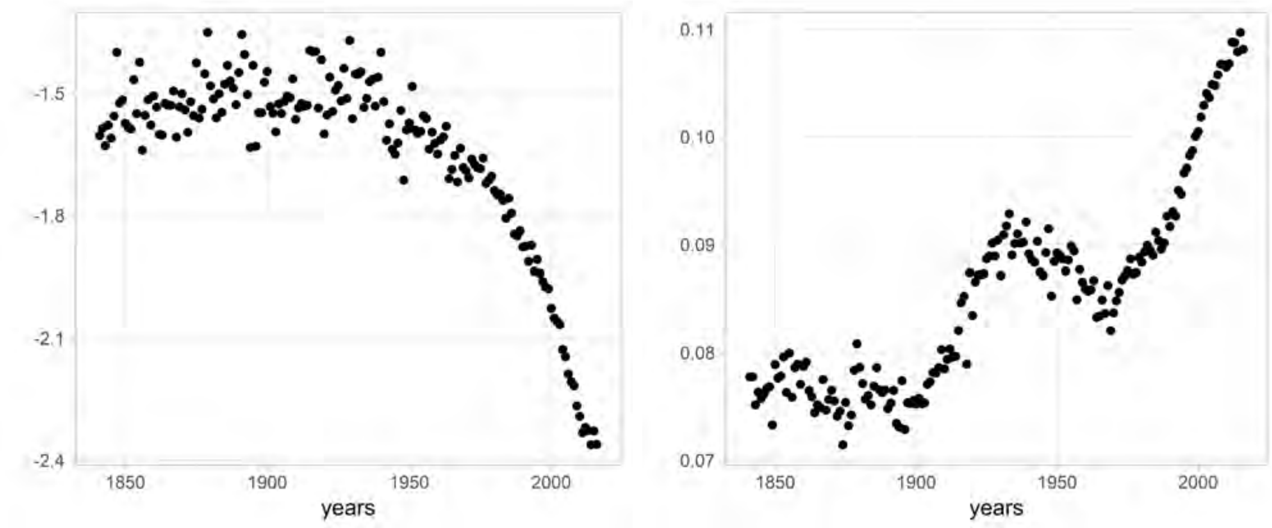


Figure 1.: Period effects in the CBD model for English and Welsh males for ages 60 to 109

future trend changes. However, it becomes apparent that this estimation is not feasible from a practical point of view. Therefore, we deduce an approach, based on the pure maximum likelihood estimation, that can be calibrated efficiently.

The remainder of this paper is structured as follows: In Section 2, we structurally compare different trend processes. Parameter estimation of these approaches is then discussed in Section 3. In particular, we present a calibration approach that incorporates the distribution of future trend changes to achieve more consistent forecasts. In Section 4, we focus on uncertainties in the parameter estimation and explain how the uncertainties can be accounted for in a practical application. Finally, Section 6 concludes.

2. Structural comparison of trend processes

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

2.1. Difference and trend stationary processes

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

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

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

Similar to the random walk models, trend models are another option to model period effects. These models use trend stationary errors around a linear trend. The easiest form of a trend model would be a simple linear regression, described by

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

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

The formula of the random walk with drift and the formula of the trend model are very similar and basically differ in the error structure. With a difference stationary structure, the correlation coefficient of subsequent changes in the period effect ($Corr(\Delta\kappa_t, \Delta\kappa_{t-1}) = 0$ with $\Delta\kappa_t = \kappa_t - \kappa_{t-1}$) should be roughly zero. In a trend stationary model it can be shown that a negative correlation of subsequent changes in the period effect is expected ($Corr(\Delta\kappa_t, \Delta\kappa_{t-1}) = -1/2$ [2]). We calculated the empirical correlation coefficients for 56 period effects (standard CBD model κ_t^1, κ_t^2 , males and females, different age spans ($50-89$, $60-109$) for England & Wales, Sweden , USA, Australia, France, Spain, Netherlands[3]) see Figure 1. Most period effects seem to be trend stationary. However, there are also exceptions (e.g. Swedish Males aged $50-89$

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

[2] With the restriction of the trend component being treated as given. With changes in the trend process a closed formula cannot be derived. However, since changes occur only rarely and are a priori unknown, the correlations should nevertheless provide a valuable indication for the model choice.

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

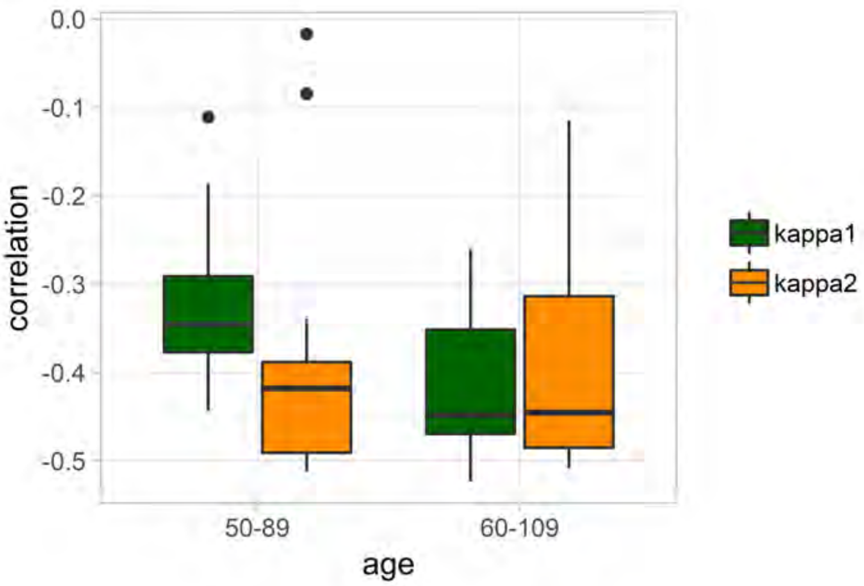


Figure 2.: Boxplot of empirical correlation coefficients of subsequent changes in 60 period effects

with a $Corr(\Delta\kappa_t, \Delta\kappa_{t-1}) = -0.11$), and therefore it is necessary to decide on a case by case basis which trend process is suitable.

2.2. Models without changes in the trend process in projections

Many authors discuss changes in the historical parameters of trend processes, e.g. a time dependent drift d_t in a random walk model or a time dependent slope β_t in a trend model. Booth et al. (2002) develop an approach for the specification of optimal periods for the estimation of the random walk with drift in the Lee and Carter (1992) model. This already implies the assumption of different mortality scenarios. An explicit change in the Lee and Carter (1992) model is tested in the approaches of Li et al. (2011) and Coelho and Nunes (2011), that find a change in almost every population under consideration. Applying the Zivot and Andrews (1992) procedure Li et al. (2011) test the null hypothesis that the period effects should be modelled with a random walk with constant drift against the alternative of a broken trend stationary mode. For England & Wales and the United States they suggest to model the underlying period effects $\hat{\kappa_t}$ with a trend change and jump in in the mid 1970s. To our perspective, the trends appear rather continuously and jumps should be rather small. Moreover, jumps corresponding to pandemia should only influence the long term trend marginally. Also other authors apply continuous linear trends in related processes in mortality modeling, e.g. Gillings et al. (1981) use a piecewise linear trend in the modeling of perinatal death rates and Wilmoth (2000) model the maximum age at death with a piecewise linear trend. Also combinations of these models are possible, e.g. Njenga and Sherris (2011) use an autoregressive process around a common fixed trend.

O'Hare and Li (2015) test the significance of a trend change with a test for the residuals based on Ploberger and Krämer (1992) for different parametric mortality models[4] in four countries (UK, US, Netherlands and Australia) and found significant trend changes. Using the Bai and Perron (1998, 2003) method they also find an optimal trend change around the 1970s in any situation. The clarity of a trend change in the 1970s is very interesting, but the period effects in Figure 1 indicate further trend changes. Including the possibility of multiple structural changes Berkum et al. (2014) allow for multiple trend changes. With the method of Bai and Perron (1998, 2003) the historical period effects are modelled as a random walk with multiple times changing drift, i.e. a time dependent drift $d = d_t$ with $(l-1)$ drift changes in $t_1, \dots, t_l$:

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

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

2.3. Models with changes in the trend process in projections

There are models that address this inconsistency and include trend changes in simulations. Several authors propose models with a regime switch. Milidonis et al. (2011) and Lemoine (2014) propose switches between two fixed regimes in models with Brownian motion and autoregressive processes, respectively. Here the different regimes appear Markov with transitions in between. For instance, Hainaut (2012) uses two regimes $\eta_t = 1, 2$ with changes in the trend and volatility for French males and females (κ_t^i corresponds to the Lee-Carter model with two dimensions):

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

where W_t^1 and W_t^2 are correlated Brownian motions. These two regimes represent two scenarios: For males the regimes differ mainly in the volatility. For females the regimes differ in trend and volatility. However, the restriction to two rigid scenarios is not possible in many situations.

[4]Models of Lee and Carter (1992), Cairns et al. (2006), Plat (2009) and O'Hare and Li (2012)

Further refining Berkum et al. (2014)'s approach, Hunt and Blake (2015) include random changes in the predicted drift. They reformulate the random walk model as:

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

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

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

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

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

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

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

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

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

3. Calibration of processes with variable drifts and variable trends

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

3.1. Statistical Analysis of historical trend processes

Some historical evolutions of mortality show a few one-year outliers due to extremal events. A popular example is the Spanish flue in 1918 with massive influence on annual death rates. Influences like wars, economic depressions or large migrations may affect mortality a few years.

But they have no sustainable influence on a longer perspective trend. Thus, they should not be included in the trend model used for forecasts. Medical improvements reduced the effects of years with strong influenza or other diseases with immediate effects on annual death rates. The countries under consideration also showed an increase in population size. Combining these medical and statistical effects results in significantly decreasing volatility in annual mortality rates. The reasons put forward indicate that decreasing volatility seems to be sustainable, which is why we assume that volatility will not return to earlier levels. We can use statistical tools to account for outliers (see Grubbs (1950) and Appendix C) or structures in the errors like the general decrease in volatility, e.g. Börger and Schupp (2018) include a non-parametric CUSUM test (see Ploberger and Krämer (1992)).

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

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

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

where $\hat{\kappa}_t(\beta, c, t_1+1, t_2) = c_{t_1+1,t_2} + \beta_{t_1+1,t_2} \cdot t$ in the case of a linear regression. Any combinations of multiple trend changes is a combination of these simple linear trends, therefore a calibration can be programmed very efficiently. For any trend process with m changes in $t_0 = 0, t_1, \cdots, t_m, t_{m+1} = n$ the likelihood optimal trend would be:

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

Hunt and Blake (2015) and Berkum et al. (2014) differentiate period effects to estimate the optimal partition of a multiple times changing random walk with drift. This is necessary as the errors are difference stationary. They use BIC (see Burnham and Anderson (2002)) to identify a parsimony model.Longer period effects often show curious trends. This is mainly caused by

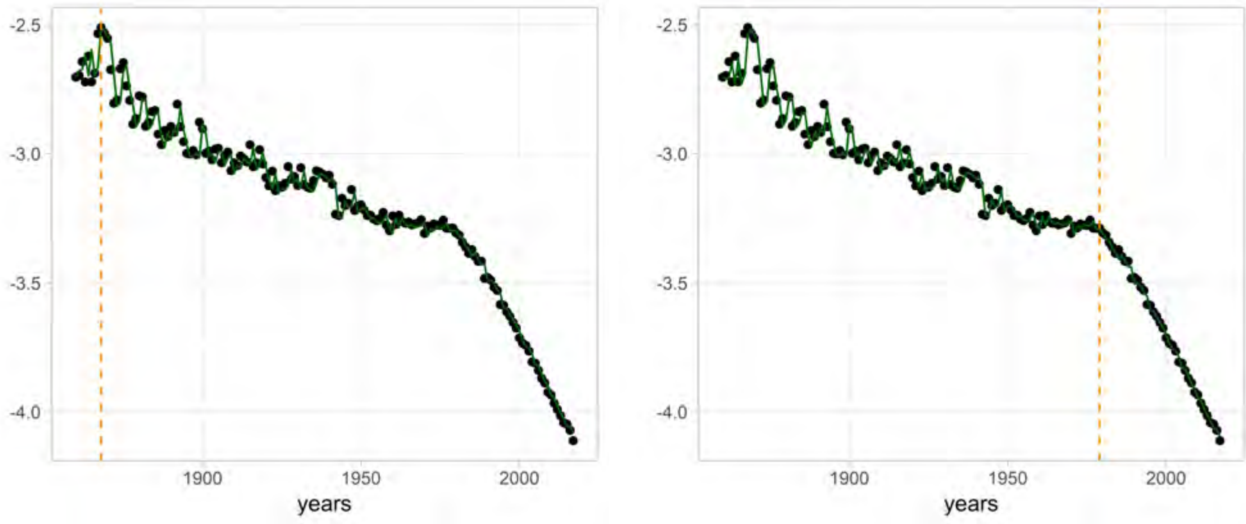


Figure 3.: Period effect κ_t^1 for Swedish males for the ages $50-89$ (dotted) and a random walk with changing drift with (right) and without (left) heteroscedasticity (solid lines). The dashed lines indicate the year of the drift change.

the aforementioned heteroscedasticity in the period effects. Using the likelihood perspective, the method can be extended straightforwardly by using normed residuals, i.e.:

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

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

The method of Muggeo (2003) can be used to estimate an optimal trend process with respect to the number of trend changes and their positions simultaneously. Börger and Schupp (2018) transferred this approach to the field of mortality modeling and showed how this approach can be extended to include the mortality specific structures in volatility. Similar to the extended version of the Bai-Perron method (see Equation 3.1), the approach estimates optimal historical trends with a volatility assumption, i.e.

$$l(\beta, c, \tau; \kappa_t, \sigma_t^2),$$

where the trend changes it's slope in the years τ. However, the continuity constraint of piecewise linear trends makes calibration considerably more difficult.

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

3.2. Combined Calibration based on likelihood

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

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

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

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

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

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

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

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

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

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

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

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

3.3. Combined Calibration based on Pseudo Likelihood

The structure of g, namely the dependencies between errors and trend changes on realized period effects made a calibration so difficult. If we look on historical period effects, there seems

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

to be no evidence for a dependence between errors and trend changes. Therefore, we combine two likelihoods independently:

1. Likelihood of observing a specific process with k changes in the trend.

2. Likelihood of observing the error structure compared to the period effect under consideration given the specific process realization from 1.

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

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

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

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

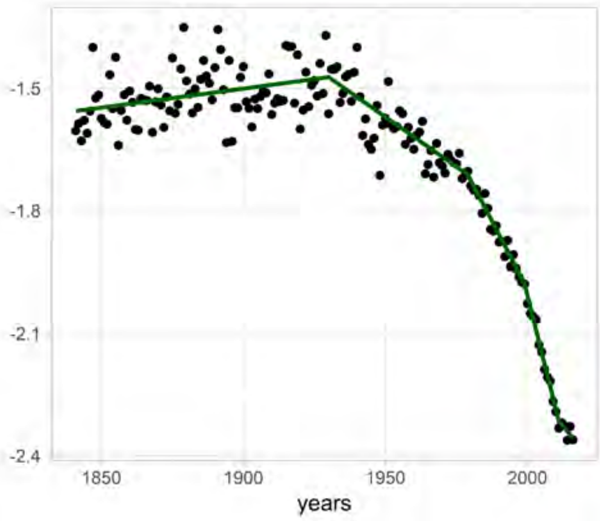
$$\tilde{l}(\beta, \gamma, c; \kappa_t, \tau^{(0)}, \sigma_t^2, \mu_{LN}, \sigma_{LN}, p) = \sum_{j=1}^{k} \log\left(f_{\mathcal{LN}}(|\beta_j|; \mu_{LN}, \sigma_{LN})\right) + l(\beta, \gamma, c; \kappa_t, \tau^{(0)}, \sigma_t^2).$$

As the simulation parameters θ_T are an intrinsic part of the calibration it is necessary to include their estimation within the iterative calibration. Instead of updating only the error's volatility iteratively we include the parameters of the trend change distribution $(p, \mu_{\mathcal{LN}}, \sigma_{\mathcal{LN}}^2)$ in the iterative scheme (see Appendix D).

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

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

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



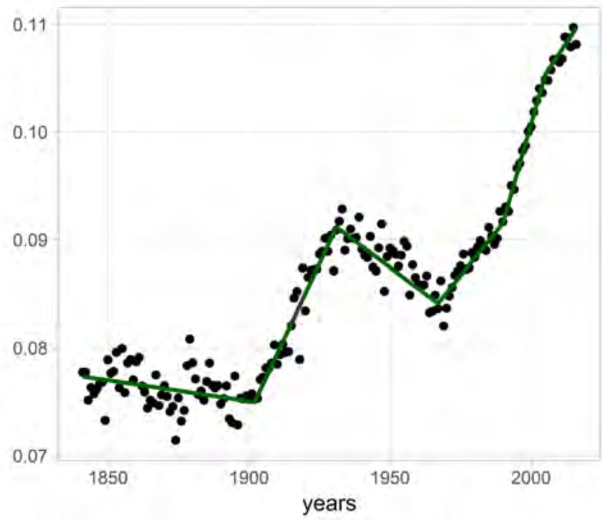


Figure 4.: Historical period effects κ_t^1 (left) and κ_t^2 (right) for English and Welsh males (dotted). Pseudo Maximum Likelihood estimated best possible realization of underlying trend process (solid lines)

For instance,

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

is monotonically increasing in k. In this situation, the BIC is not only used to balance between under- and overfitting due to randomness. In addition, the BIC compensates systematic growth in the Pseudo Likelihood part with increasing numbers of trend changes. In most of the period effects studied, the BIC seemed capable of doing so. However, there are also cases where the BIC is unable to compensate for this systematic effect. With a normalized Pseudo Maximum Likelihood approach this systematic effect can be avoided. Therefore, we normalize the optimisation such that it is directly comparable for different numbers of trend changes, i.e. we replace formula 3 by:

$$\tilde{l}(\beta, c, \tau; \kappa_t, \sigma_t^2, \mu_{LN}, \sigma_{LN}, p) = l_{\mathcal{N}}(\beta, c, \tau; \kappa_t, \sigma_t^2) + l_{\mathcal{LN}}(\beta) + l_{\mathcal{B}}(p) - \log(\mathbb{E}\mathcal{L}_{\mathcal{N}} \cdot \mathbb{E}\mathcal{L}_{\mathcal{LN}} \cdot \mathbb{E}\mathcal{L}_{\mathcal{B}}). \quad (4)$$

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

4. Parameter Uncertainty

The parameter estimation approaches presented in the previous section include a significant estimation risk. In this section, we show an efficient way to account for parameter uncertainty.

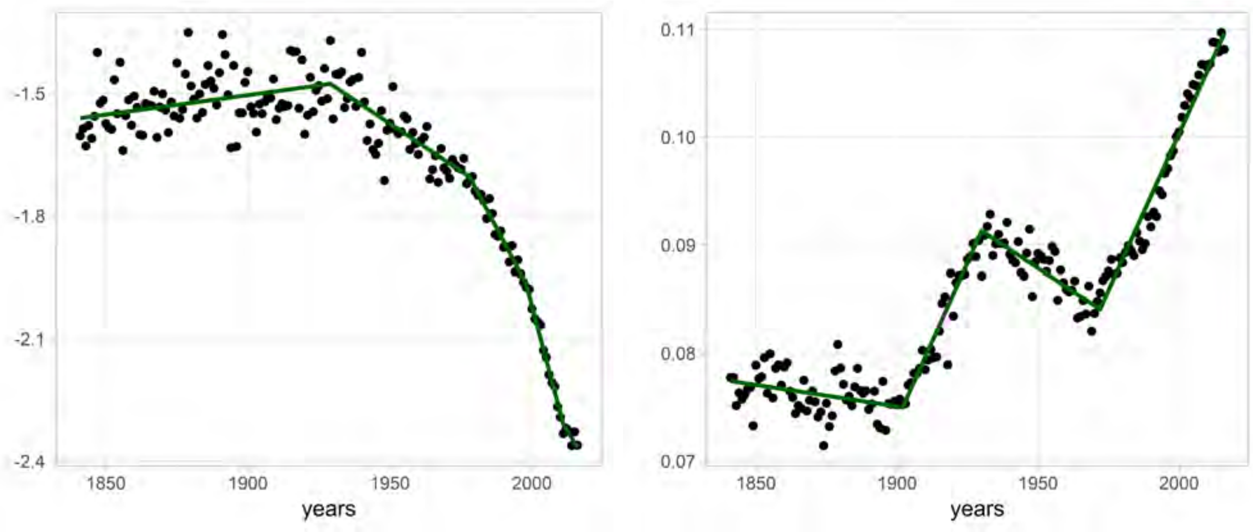


Figure 5.: Historical period effects κ_t^1 (left) and κ_t^2 (right) for English and Welsh males (dotted). Normed Pseudo Maximum Likelihood estimated best possible realization of underlying trend process (solid lines)

The estimation risk of these models can be divided into two main categories (see Börger and Schupp (2018)): The uncertainty in the trend change parameters (p, μ, σ^2). Exemplary an overestimation (underestimation) of these parameters result in unrealistically wide (narrow) long-term prediction intervals. Secondly, the uncertainty in the starting values ($\hat{\kappa}_0, \beta_0$ (trend stationary) or d_0 (difference stationary)) affects especially short-term estimates directly. This uncertainty can vary extremely when comparing different countries and period effects as it is dominantly driven by the latest observations. Some approaches to account for parameter uncertainty are computationally intensive and are therefore often disadvantageous in a practical use. With the assumption of deaths following a Poisson or Binomial distribution, Koissi et al. (2006) propose to recalibrate the period effects first and then re-estimate the parameters used for forecasts. In the situation of trend stationary period effects, the optimal historical realization needs to be recalibrated with the iterative scheme for each simulated period effect. Clearly, this is computationally expensive. Moreover, approaches based on recalibration can only comprehend the uncertainty in the starting values partially as the potential additional change would not have been detected simply due to the limited data afterwards and not due to random effects captured with the Poisson/Binomial distribution. The approaches of the previous section estimate optimal realizations of trend processes by comparing estimates for different numbers of trend changes. The main uncertainty is about the actual number of trend changes. Due to random effects, it is possible that the optimal estimate overestimates or underestimates the number of trend changes. As a side-product of the calibration we already have estimates of the trend change parameters for different numbers of trend changes k denoted by $\zeta_k = (p_k, \mu_k, \sigma_k^2)$ and $\eta_k = (\hat{\kappa}_{0,k}, \beta_{0,k})$. The uncertainty in the starting values mainly arises due to limited data after the last trend change. For instance, the last trend in κ_t^1 only includes a few years and also a starting value and slope without that last trend change should be included in a simulation. Börger and Schupp (2018) propose to adopt Bayesian weights to compare different parameter sets. We expand this approach, such that the uncertainty in the trend change parameters, in the starting values and in the volatility structure are covered jointly. The calibration provides parameter sets for different numbers of trend changes that can be used straightforwardly to assign probabilities to θ_k and η_k. Using the optimization in Equation 4 and let $\tilde{l}(\cdot, k) = \tilde{l}(\beta, c, \tau; \kappa_t, \sigma_t^2, \mu_{LN}, \sigma_{LN}, p)$ denote the Pseudo Maximum Likelihood for k trend changes:

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

where $l(\cdot, \hat{k})$ is the overall optimal value of all possible k. This approach can also be applied to other optimizations based on Maximum Likelihood, i.e. also to approaches based on information criteria as in the approaches of Hunt and Blake (2015) or Berkum et al. (2014).

We can use these weights to assign probabilities for the trend change parameters to account for the unclear number of actual trend changes, i.e.:

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

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

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

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

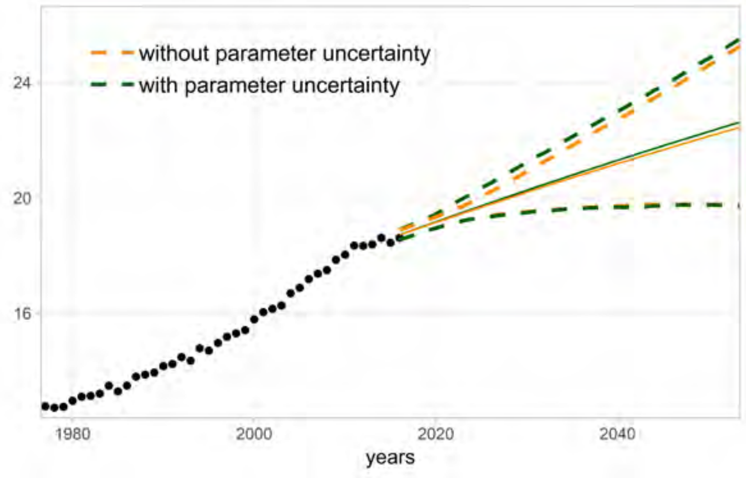


Figure 6.: Historical and forecast 90% prediction intervals (dashed lines) and median (solid lines) for remaining period life expectancy of 65-year old males in England & Wales

paths. Thus, both scenarios, namely a flattening and a persistence of the recent trends in mortality improvements, are well included in the prediction intervals.

5. Conclusion

Throughout the recent years a variety of stochastic mortality models have been developed and most of these models use one or more time dependent period effect. Forecasting these period effect(s) requires to analyse the historical shapes, where we can observe changes in the improvements of mortality rates in most countries.

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

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

Appendix

A. Cairns-Blake-Dowd Model

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

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

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

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

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

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

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

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

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

B. Summaries - specification and calibration of trend processes

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

Table 1.: Variants of models with trend processes

C. Accounting for Outliers

One-year outliers in data with normally distributed errors can be identified with Grubbs (1950)' test statistic

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

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

D. Trend process Calibration

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

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

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

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

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

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

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

$$l(\beta, \gamma, c; \kappa_t, \tau^{(0)}) = -\frac{1}{2} \sum_{t=t_0}^{t_n} \left(\log(2\pi\sigma_t^2) + \frac{\kappa_t - \hat{\kappa}_0 - \beta' \cdot (t \cdot \mathrm{I} - \tau^{(0)})^+ + \gamma' \cdot \mathbb{1}(t \cdot \mathrm{I} > \tau^{(0)})}{\sigma_t} \right)^2 . \tag{5}$$

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

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

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

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

2. Determine an initial trend with one trend change assuming $N(0, \sigma_t^2)$-distributed residuals and update the estimate of σ_t^2.

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

als.

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

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

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

   a) Determine the optimal continuous and piecewise linear trend curve for $N+1$ trend changes based on the method proposed by Muggeo (2003) with starting values s using the optimization function under consideration (can be Equation 3, or Equation 4.

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

      ii. Update breakpoints according to

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

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

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

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

References

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

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

Berkum, F. v., Antonio, K., and Vellekoop, M. (2014). The impact of multiple structural changes on mortality predictions. *Scandinavian Actuarial Journal* 2016(7): 581–603.

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

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

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

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

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

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.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Source:
Börger, M., Ruß, J., and Schupp, J. (2019). It Takes Two: Why Mortality Trend Modeling is more than Modeling one Mortality Trend, Working Paper, Ulm University.

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

Matthias Börger
Technische Hochschule Nürnberg Georg Simon Ohm, Nuremberg
Keßlerplatz 12, 90489 Nürnberg, Germany

Jochen Russ
Institute for Finance and Actuarial Sciences (ifa), Ulm & Institute of Insurance, Ulm University
Lise-Meitner-Straße 14, 89081 Ulm, Germany
Phone: +49 731 20644233
Email: j.russ@ifa-ulm.de

Johannes Schupp
Institute for Finance and Actuarial Sciences (ifa), Ulm & Institute of Insurance, Ulm University
Lise-Meitner-Straße 14, 89081 Ulm, Germany
Phone: +49 731 20644241
Email: j.schupp@ifa-ulm.de

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. In many applications however one needs to carefully distinguish between two different mortality trends: The actual (but unobservable) mortality trend (AMT) prevailing at a certain point in time and the estimated mortality trend (EMT) that an observer would estimate given the (observable) realized mortality up to that point in time. Since the AMT is not observable, an actuary or demographer might misestimate the AMT at any point in time. In particular, he would typically not 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 future AMT or the future EMT or both need to be considered and modeled 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 numerically efficient combined model for both trends is specified and calibrated to mortality data. The model component for the actual mortality trend builds on recent findings that mortality appears to evolve log-linear over time with random changes in slope. The model component for the estimated mortality trend is specified such that, given the assumed dynamics for the actual mortality trend, the estimated mortality trend matches the actual trend as close as possible. This provides valuable information on how best estimate mortality assumptions should be derived from the available data in general. Finally, we apply the combined model in practical examples and illustrate the importance of distinguishing between AMT and EMT. We show that, if the AMT is wrongfully assumed observable, the hedge effectiveness of a longevity hedge or the SCR for longevity risk are typically misestimated significantly.

1. Introduction

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 demographers and actuaries and constitutes an important input for the pricing, reserving, and risk management of annuity and pension products in particular.

In this paper, we deal with an aspect that is often ignored in the existing literature, i.e. the fact that one needs to distinguish between two different mortality trends: The (unobservable) 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. These trends differ, since obviously, an observer will not be able to perfectly estimate the (unobservable) actual mortality trend. Moreover, an observer might not always be able to distinguish between a recent chance in the actual trend and a simple random fluctuation around the previous long term trend. This is particularly relevant immediately after a trend change or immediately after a rather strong random fluctuation. We will show that not distinguishing between these two trends might have significant undesired consequences.

Even very simple examples can motivate why one needs to distinguish between the two trends (we will discuss some examples in much more detail in Section 4 where we will also quantify the error that would result from not distinguishing between the two trends). As a first example, assume that one is interested in a confidence band for the cash-flow of a portfolio of annuity contracts over the next, say, 10 years or its present value. In such a run-off simulation, only actual future mortality and hence the survival rates derived from a model for the AMT need to be used [1].

As a second example, assume that one is interested in the question how reserves for the same book of annuity contracts can change from one year to the next. Now, one needs to model actual mortality over one year as well as the evolution of the EMT over this year, since the EMT (i.e. an observes estimate for the then current trend) would be the basis for the calculation of policy reserves. Hence, a model with two components is required: the development of actual mortality (based on the AMT) and the development of the EMT which at any point in time is some function of the realized mortality up to that point in time.

These simple examples already show that for some applications a model for the future development of the AMT is needed, for other applications, a model for the future development of

[1]Nevertheless, even here, it is important to distinguish between the AMT and the EMT since one has to be aware, 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 below.

the EMT is needed, and frequently AMT and EMT need to be modelled simultaneously. A model that only captures one of these two trends might therefore not be suitable for certain analyses.

Therefore, in this paper, we specify a mortality model that simultaneously and consistently projects AMT and EMT. We choose a model structure that makes the combined AMT/EMT model highly efficient also in Monte Carlo simulations where the EMT needs to be estimated in each simulation path: For the AMT component, we use the model proposed by Börger and Schupp (2018) and further refined in Schupp (2019). This model builds on the model structure of Cairns et al. (2006), and the trend processes for the time dependent parameters take into account recent findings that mortality appears to evolve log-linear over time with random changes in slope. We show how the model parameters can be estimated from historical data and discuss how parameter uncertainty, in particular in the unobservable current AMT, can be accounted for. Finally, we illustrate the adequacy of the AMT model component and its parameter estimation by a concrete example calibration for English and Welsh males.

For the EMT component, we discuss and compare different model specifications. The EMT should match the AMT as closely as possible, and therefore, the assumed dynamics for the AMT provide valuable information on how the EMT should be specified. Since we assume the AMT to evolve piecewise linearly, it appears reasonable to also model a linear EMT. We therefore consider a linear regression approach and analyze the effect of different weightings within the regression. Obviously, more recent data points are more informative with respect to the current AMT than data points further in the past. The optimized weightings which are derived in this paper also provide a general indication how mortality trends should be estimated from available data.

Finally, we apply the combined model in practical examples and illustrate the importance of distinguishing between AMT and EMT. If the AMT is wrongfully assumed observable, the hedge effectiveness of a longevity hedge or the SCR for longevity risk are typically misestimated significantly.

The majority of the mortality models proposed in the literature have been developed in order to project actual mortality. Often, these models are also used to derive the EMT, typically in form of the (deterministic) central projection. For instance, Richards et al. (2014) use the Lee-Carter model (see Lee and Carter (1992)) both as an AMT and EMT model in order to compute the Solvency Capital Requirement (SCR) for longevity risk under Solvency II. They project actual mortality for one year, then recalibrate the model based on the extended data set, and finally interpret the central projection of the recalibrated model as the EMT after one year. Cairns (2013), Cairns et al. (2014), and also Cairns and El Boukfaoui (2019) essentially apply the same concept in the field of longevity hedging, but explicitly distinguish between a simulation model for the AMT and a valuation model for the EMT. These models can be chosen independently of each other. However, the repeated full (re)calibration of a valuation model as part of a Monte Carlo simulation framework is typically very expensive from a computational point of view. For this reason, Börger et al. (2014) propose a model in which only the mortality trend (as opposed to the full model) is reestimated based on the available additional data. Thus, they combine the projection of actual and estimated mortality in a single model, but create some inconsistency by interpreting the EMT also as the AMT at any point in time.

All the aforecited papers have in common that first actual mortality is simulated and then the EMT is derived from an extended data series. Plat (2009) turns this intuitive approach around when he proposes a stochastic EMT model in which, in a second step only, actual mortality rates are derived such that they are consistent with the simulated EMT change; for instance, if expected mortality increases from one year to the next, the actual mortality rates must have been larger than anticipated. Since the simulation of actual mortality is rather cumbersome without an explicit AMT, the model is particularly designed for applications in which the EMT is required in each year, e.g. for SCR computations. A similar argument holds for the class of forward mortality models. These models primarily project changes in expected mortality rates, but also provide actual mortality rates when expected mortality turns into realized mortality. Forward mortality models have been proposed by Bauer et al. (2008), Bauer et al. (2010), Zhu and Bauer (2011), or Hunt and Blake (2015, 2016). However, the specification and calibration of these models is fairly complex.

The remainder of this paper is organized as follows: In Section 2, we introduce the AMT component of our combined mortality model. We discuss how its parameters can be estimated and how parameter uncertainty can be accounted for. Moreover, we derive a full calibration for the case of English and Welsh males. Section 3 discusses how the EMT should be derived from observed mortality. In particular, we compare different approaches for weighting the available data such that the EMT adequately reacts to trend changes in the AMT. In Section 4, we provide three practical examples in which both, the AMT and the EMT are required, and we show how important it is to properly distinguish between these two trends. Finally, Section 5 concludes.

2. AMT Model Component

As outlined in the Introduction, we will now derive a full specification of a mortality model that consists of two components: the model for the unobservable actual mortality trend (AMT) and the model for the estimated mortality trend (EMT) an actuary or demographer would derive from the observable data available at any point in time. In this section, we introduce the AMT model component. The EMT model component will be discussed in Section 3.

2.1. AMT Model Specification

For the AMT, we use the model proposed by Börger and Schupp (2018) and further refined by Schupp (2019). This model builds on the well-known CBD model of Cairns et al. (2006), i.e. annual probabilities of death are modeled as

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

where $\bar{x}$is the average age of the age range under consideration. The time dependent parameter $\kappa_t^{(1)}$ determines the general level of mortality, whereas the slope parameter $\kappa_t^{(2)}$ describes the increase of mortality with age.

In many cases, these time dependent parameters are projected by a two-dimensional random walk with drift as originally proposed by Cairns et al. (2006). However, Figure 1 indicates that assuming a constant and fixed drift might not be a reasonable assumption particularly for long-term mortality projections. The figure shows the logarithm of probabilities of death for 65-year old males in different countries all over the world; the data has been obtained from the Human Mortality Database (2018). We observe for all populations that mortality trends appear linear, but change their slope once in a while. Thus, in order to project mortality consistently with historical observations, it appears reasonable to allow for trend changes also in the future.

Börger and Schupp (2018) show that a random walk with constant drift might significantly underestimate the long-term uncertainty in future mortality and therefore propose a trend process which projects piecewise linear trends with random changes in slope. [2] For any future year t, the 'observable' processes $\kappa_t^{(1)}$ and $\kappa_t^{(2)}$ are modeled as the sum of the actual but unobservable true mortality processes and some random noise:

$$\kappa_t^{(i)} = \hat{\kappa}_t^{(i)} + \epsilon_t^{(i)}.$$

The noise term $\epsilon_t^{(i)}$ accounts for annual fluctuations which are, e.g., due to flu waves, very hot summers, or catastrophes. The vector $\epsilon_t = \left(\epsilon_t^{(1)}, \epsilon_t^{(2)}\right)$ is assumed to follow a two-dimensional Normal distribution with mean zero and covariance matrix Σ.[3]

The actual mortality processes $\kappa_t^{(i)}, i = 1, 2$ are projected linearly with the current unobservable AMTs $d_t^{(i)}, i = 1, 2$ as slopes:

$$\hat{\kappa}_t^{(i)} = \hat{\kappa}_{t-1}^{(i)} + d_t^{(i)}, i = 1, 2.$$

[2]For a literature overview on alternative models with time dependent drift see e.g. Börger and Schupp (2018)

[3]Note that the covariance matrix is assumed constant over time even though heteroscedasticity can usually be observed in the historical data (see Figure 1). However, since the noise does not impact long-term mortality evolutions (in contrast to the innovations in the random walk), this simplification appears acceptable.

The AMTs remain unchanged until the next trend change occurs:

$$d_t^{(i)} = \begin{cases} d_{t-1}^{(i)} & \text{, if no trend change occurred in } t-1 \\ d_{t-1}^{(i)} + \lambda_{t-1}^{(i)} & \text{, if a trend change by } \lambda_{t-1}^{(i)} \text{ occured in } t-1 \end{cases}$$

The trend change intensities $\lambda_t^{(i)}$, are derived as the product of their absolute magnitudes $M_t^{(i)}$ and their signs $S_t^{(i)}$:

$$\lambda_t^{(i)} = S_t^{(i)} \cdot M_t^{(i)}, i = 1, 2.$$

Based on analyses of historical trend changes, Börger and Schupp (2018) model the magnitudes $M_t^{(i)}$ by lognormal distributions $\mathcal{LN}(\mu^{(i)}, \sigma^{(i)})$. For the signs $S_t^{(i)}$, they use a Bernoulli distribution with attainable values -1 and 1 and probability 0.5 each. The probability of observing a trend change in $d_t^{(i)}$ in any particular year is denoted by $p^{(i)}, i = 1, 2$. Moreover, trend changes in $d_t^{(1)}$ and $d_t^{(2)}$ are assumed to occur independently as indicated by the occurrences of trend changes in the historical data for a large set of populations.

This decomposition of the trend change intensities has several convenient implications. First, the distributions of future trend changes are symmetric, i.e. the slope in the AMT increases and decreases with equal probability and magnitude. Thus, the prevailing AMT is always the best estimate for the trend at any future point in time. Furthermore, the distribution of $\lambda_t^{(i)}$ has no probability mass at zero and only very little mass around zero. Thus, simulated trend changes can be considered as rather 'significant'. At the same time, the heavy tail of the Lognormal distribution implies that strong trend changes can occur which is in line with some of the trend changes we observe in Figure 1 (e.g. in Sweden around 1980).

2.2. Parameter Estimation and Uncertainty

The estimation of the model parameters consists of two main steps. First, the CBD model is fitted to historical data. Here, the historical time series should be as long as possible such that it contains as many trend changes as possible. In the second step, the parameters of the trend processes $\kappa_t^{(i)}$ are estimated from the historical realizations $\kappa_t^{(i)}, t \leq t_0$, where t_0 denotes the final year of the historical data, i.e. the starting point of a simulation. The parameters to be estimated are:

- the starting values for the actual trend processes $\kappa_{t_0}^{(i)}$,
- the prevailing AMTs $d_{t_0}^{(i)}$,
- the probabilities of observing a trend change in a certain year, $p^{(i)}$

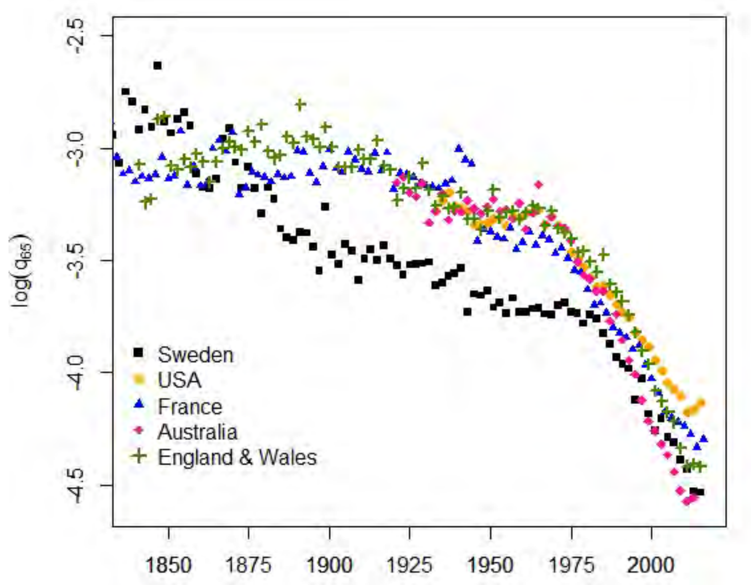


Figure 1.: Logarithm of probabilities of deaths for 65-year old males in selected countries

- the parameters of the Lognormal distributions for the trend change magnitudes, $\mu^{(i)}$ and $\sigma^{(i)}$,
- and the covariance matrix Σ of the two-dimensional noise vector ϵ_t.

The parameter estimation is carried out separately for each $\kappa_t^{(i)}$ process and the correlation between $\epsilon_t^{(1)}$ and $\epsilon_t^{(2)}$ is estimated once all other parameters have been determined. As Schupp (2019) explains, parameter estimation is complex due to the dependence of realized $\kappa_{t,t\leq t_0}^{(i)}$ on potential but unknown trend changes in previous years. In particular, a full maximum likelihood estimation of all model parameters seems unfeasible. Therefore, we estimate parameters from specific realizations for the actual trend process $\kappa_t^{(i)}, t \leq t_0$.

For a trend process realization with k past trend changes the obvious estimate for the future trend change probability is $p^{(i)} = \frac{k}{N}$, where N is the number of data points in the time series $\kappa_t^{(i)}, t \leq t_0$. For each of the (rather low number of) potential values for $p^{(i)}$ or k, respectively, the other parameters are then optimized using a pseudo maximum likelihood algorithm. For details on this rather complex algorithm, we refer to Schupp (2019).

We then have a full set of parameter estimates for each number of possible past trend changes k. From these parameter sets, some 'optimal' parameter estimates could be chosen given some optimality criterion. However, since the parameter estimation involves a certain degree of uncertainty, we rather interpret these parameter sets as possible outcomes of an empirical distribution from which parameter values can be drawn randomly for each simulation path. The probabilities for the different parameter sets S_k are derived as Bayesian weights (see Burnham and Anderson (2002)):

$$P(S_k) = \frac{\exp\left(-\frac{1}{2}\left(BIC(S_k) - min_i(BIC(S_i))\right)\right)}{\sum_{j=0}^{N-1} \exp\left(-\frac{1}{2}\left(BIC(S_j) - min_i(BIC(S_i))\right)\right)},$$

where $BIC(S_k)$ denotes the Bayesian Information Criterion for the parameter set S_k. It can be determined from the likelihood of a specific trend process realization which is a byproduct of the pseudo maximum likelihood algorithm (see Schupp (2019) for details); the number of parameters in the penalty term of the BIC corresponds to the degrees of freedom of a trend process realization with k trend changes, i.e. $3 \cdot k + 3$.

2.3. Example Calibration

In this subsection, we present a model calibration for English and Welsh males. We use the entire data set which is available in the Human Mortality Database (2018) for ages 60 to 109, i.e. from 1841 to 2016. Figure 2 shows the historic trend processes $\kappa_t^{(1)}$ and $\kappa_t^{(2)}$ and the best possible realizations for the actual trend processes $\hat{\kappa}_t^{(1)}$ and $\hat{\kappa}_t^{(2)}$ given different numbers of past trend changes k; the depicted k are those with Bayesian weights significantly different from zero. We observe that there might have been three, four, or five trend changes for $\kappa_t^{(1)}$ and five, six, or eight trend changes for $\kappa_t^{(2)}$.

Table 1 and Table 2 provide the estimates for $p^{(i)}, \mu^{(i)}, \sigma^{(i)}, \hat{\kappa}_{t_0}^{(i)}, \hat{d}_{t_0}^{(i)}$, and $\Sigma_{(i,i)}$ as well as the probabilities for all parameter sets. The trend change probability $p^{(i)}$ obviously increases with k. The mean of the trend change magnitude $\mu^{(i)}$, on the other hand, is rather constant. Combining these observations implies that the different parameter sets yield long-term mortality projections which differ significantly in terms of uncertainty. For $\kappa_t^{(2)}$, however, almost all probability is concentrated at $k = 5$, i.e. parameter uncertainty is rather small. For $\kappa_t^{(1)}$ most probability is assigned to the case $k = 4$, but there is also a significant probability for the case $k = 3$. Interestingly, the start values for the actual trend process, $\hat{d}_{t_0}^{(1)}$, differ significantly. While a trend change in 2011 is very likely, there is not yet enough data to be fairly sure (see also the difference between the orange and blue curves in the top panel of Figure 2). Thus, parameter uncertainty is significant here.[4]

Finally, Table 3 contains the covariance estimates between $\epsilon_t^{(1)}$ and $\epsilon_t^{(2)}$ for all combinations of parameter sets for $\kappa_t^{(1)}$ and $\kappa_t^{(2)}$. They are derived from the correlations in the noise terms of the best possible trend process realizations and the variance estimates $\Sigma_{(1,1)}$ and $\Sigma_{(2,2)}$.

[4]We also tested the inclusion of additional trend changes after the most recent detected one for each k as done in Börger and Schupp (2018). However, the likelihood for such additional trend changes was negligible in any case.

k	$p^{(1)}$	$\mu^{(1)}$	$\sigma^{(1)}$	$\hat{\kappa}_{t_0}^{(1)}$	$\hat{d}_{t_0}^{(1)}$	$\Sigma_{(1,1)}$	$P(S_k)$
3	0.0170	-4.836	0.2398	-2.403	-0.0235	$6.62 \cdot 10^{-4}$	5.63%
4	0.0227	-4.520	0.4698	-2.356	-0.0084	$2.26 \cdot 10^{-4}$	94.00%
5	0.0284	-4.543	0.4387	-2.360	-0.0086	$2.63 \cdot 10^{-4}$	0.37%

Table 1.: Empirical distribution for the AMT model parameters in $\hat{\kappa}_t^{(1)}$

k	$p^{(2)}$	$\mu^{(2)}$	$\sigma^{(2)}$	$\hat{\kappa}_{t_0}^{(2)}$	$\hat{d}_{t_0}^{(2)}$	$\Sigma_{(2,2)}$	$P(S_k)$
5	0.0284	-7.431	0.1355	0.1096	$3.95 \cdot 10^{-4}$	$4.64 \cdot 10^{-7}$	98.90%
6	0.0341	-7.526	0.3503	0.1096	$3.91 \cdot 10^{-4}$	$4.96 \cdot 10^{-7}$	1.05%
7	0.0398	-7.553	0.5053	0.1087	$3.85 \cdot 10^{-4}$	$4.46 \cdot 10^{-7}$	0.00%
8	0.0455	-7.659	0.4697	0.1095	$3.92 \cdot 10^{-4}$	$3.94 \cdot 10^{-7}$	0.05%

Table 2.: Empirical distribution for the AMT model parameters in $\hat{\kappa}_t^{(2)}$

This fully calibrated AMT model can be used to answer questions for which only the realized future mortality evolution is required, e.g. to simulate a portfolio run-off. In order to illustrate the plausibility of the mortality scenarios which the model generates, Figure 3 shows the median and mean projections as well as the 95 % prediction intervals for the remaining period life expectancy of 65-year old English and Welsh males. We see that, in terms of AMT, the mean projection lies somewhere between the historical trends before and after 2011. This is in line with the empirical distribution for $\hat{d}_{t_0}^{(1)}$ which assigns substantial weight to both values for the AMT. The median projection yields slightly smaller life expectancies than the mean projection which illustrates the 'upward' skewness implied by the empirical distribution for $\hat{d}_{t_0}^{(1)}$. The prediction intervals look highly plausible in the sense that the model output permits both, an ongoing steep increase in life expectancy as well as a leveling off in the near future followed by a long term decrease.[5]

3. EMT Model Component

In this section, we introduce the EMT component of our mortality model. As explained in the Introduction, the EMT is required in addition to the AMT in many practical applications, e.g. for the computation of risk capital over limited time horizons or for computing terminal payoffs of longevity hedging instruments with a fixed term. Therefore, we analyze and compare possible

[5]Comparisons of the presented trend process with alternative processes within the CBD model can be found in Börger and Schupp (2018)

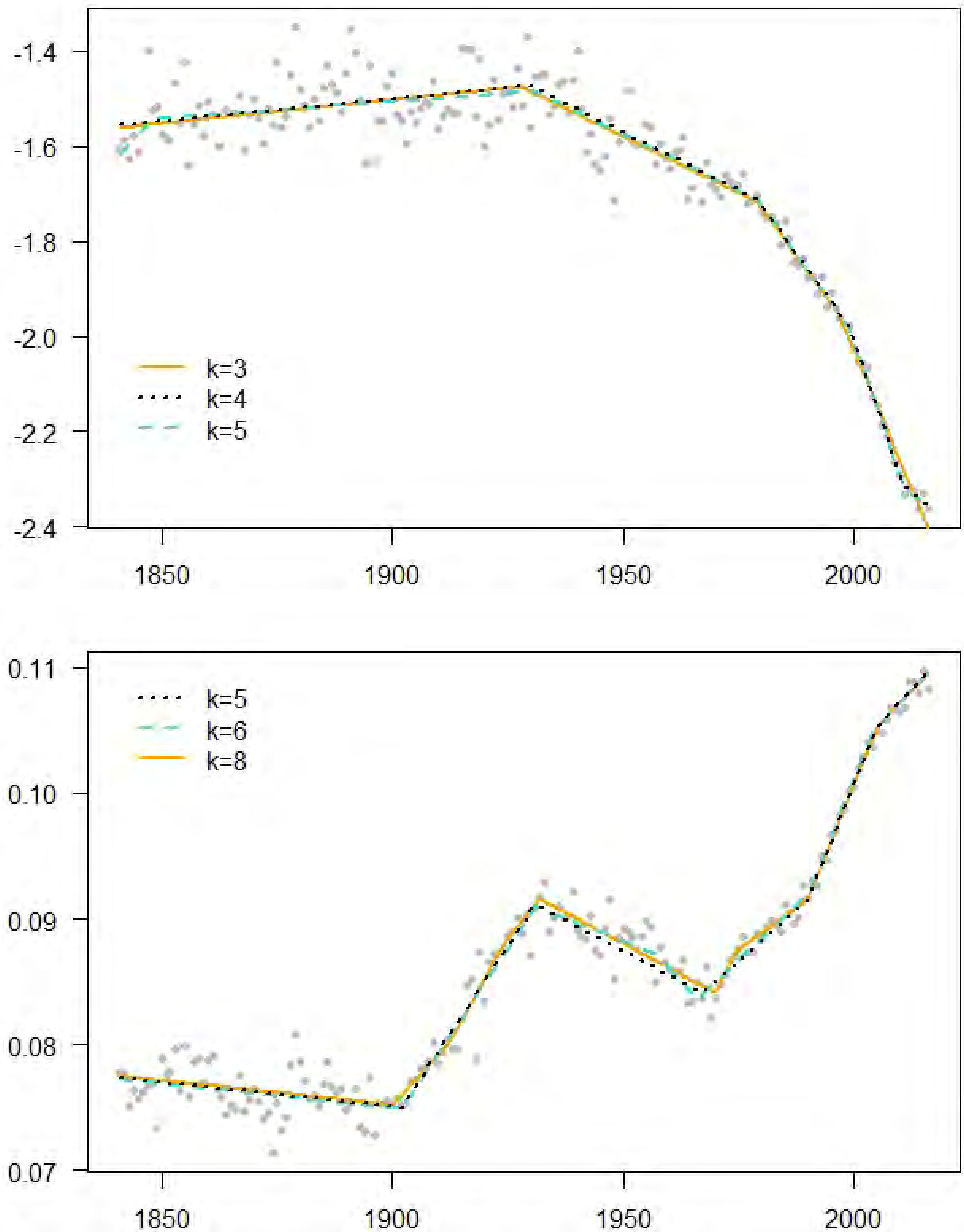


Figure 2.: Historical trend processes $\kappa_t^{(1)}$ (top) and $\kappa_t^{(2)}$ (bottom) for English and Welsh males (dots) and best possible realizations for the actual trend processes $\hat{\kappa}_t^{(1)}$ and $\hat{\kappa}_t^{(2)}$ given different numbers of trend changes k (lines)

	$\Sigma_{(1,2)}$	$\kappa_t^{(2)}$ $k=5$	$k=6$	$k=7$
$\kappa_t^{(1)}$	$k=3$	$9.25 \cdot 10^{-6}$	$9.43 \cdot 10^{-6}$	$8.80 \cdot 10^{-6}$
	$k=4$	$7.06 \cdot 10^{-6}$	$8.00 \cdot 10^{-6}$	$7.16 \cdot 10^{-6}$
	$k=5$	$6.86 \cdot 10^{-6}$	$7.76 \cdot 10^{-6}$	$7.09 \cdot 10^{-6}$

Table 3.: Covariance estimates between $\epsilon_t^{(1)}$ and $\epsilon_t^{(2)}$ for all combinations of parameter sets for $\kappa_t^{(1)}$ and $\kappa_t^{(2)}$

approaches for deriving the EMT, outline shortcomings of commonly used EMT approaches, and discuss the availability of information in order to derive the EMT at any point in time.

3.1. EMT Modeling Setup

Since the AMT is unobservable, any prediction of future mortality must build on some estimate for the AMT. For time t_0, one example for a (probabilistic) estimate has already been presented in Section 2.2. In many practical applications, however, within simulations an estimate for the AMT is required also at later points in time. Here, typically a point estimate is sufficient and we denote such a deterministic EMT at some point in time T by $\tilde{d}_T^{(i)}$.[6] An obvious choice for a deterministic EMT at time t_0 would be the mean of the empirical distribution for $\hat{d}_T^{(i)}$. For English and Welsh males at $t_0 = 2016$, this would be $\tilde{d}_{2016}^{(1)} = -0.00925$ and $\tilde{d}_{2016}^{(2)} = 0.000395$. In principle, it would be possible to determine the EMT in this way also within a simulation model at any future point in time. However, this would be fairly complex and time consuming since empirical distributions would have to be derived in every simulation path for every relevant point in time. Hence, more efficient alternatives need to be considered in practical applications.

As explained in Section 2.1, under our assumptions the prevailing AMT is always the (unobservable) best estimate for the future mortality trend. If we believe that the piecewise linear AMT describes reality in an adequate way, it is obvious to derive the EMT as the linear trend in the most recent available data.[7] We note that in practice, the $\kappa_t^{(i)}$ processes are not directly observable. However, by fitting the CBD model structure to raw mortality rates up to time T, an observer would obtain the same $\kappa_t^{(i)}, t \leq T$ as projected in the AMT simulation. Thus, the $\kappa_t^{(i)}$ can be derived from observable data. Then the EMT can be derived from the $\kappa_t^{(i)}$ which are directly available from the AMT simulation up to T.[8] This makes our modeling framework

[6]EMT-related quantities are denoted by $\tilde{\cdot}$ and AMT-related quantities by $\hat{\cdot}$ throughout the paper.
[7]Note that, by this approach, the (unknown) level of the mortality process, $\kappa_T^{(i)}$, is estimated as a byproduct.
[8]When deriving the EMT in practice, the $\kappa_t^{(i)}$ would typically be available only up to a few years before T as data collection and preparation takes its time. However, this does not affect the general concept presented in this paper, and for simplicity we therefore assume that mortality data is available directly.

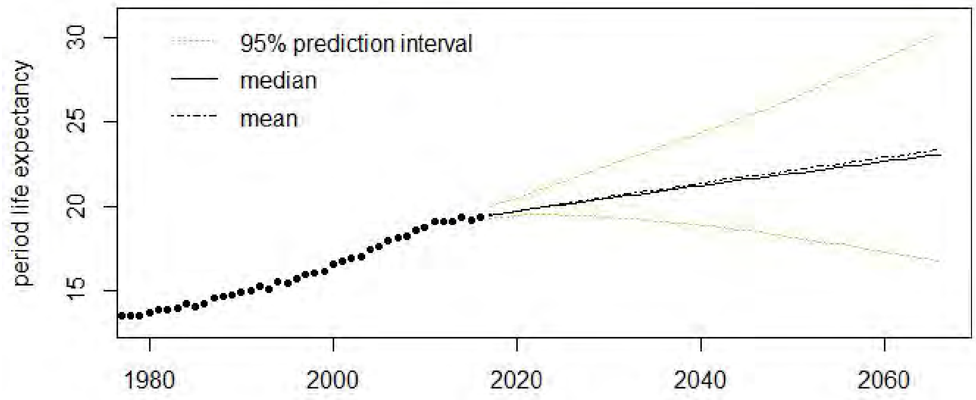


Figure 3.: Historical and forecast period life expectancies for 65-year old males in England and Wales

very efficient. In other mortality models, this might not be the case. For instance, in the model of Lee and Carter (1992), the observer would estimate a process $\kappa_t, t \leq T$ which differs from the simulated one due to interactions between the trend process and the age dependent parameters. Here, additional information in form of new data impacts κ_t estimates for previous years. Thus, the κ_t process could not be assumed observable, and the EMT would need to be determined from the observed raw mortality rates.

The common approach for estimating a mortality trend is to determine the linear trend within a time series of data whose length is usually fixed based on data availability constraints, expert judgement, or irregularities in the data. In such a setting, all data points usually have the same impact on the trend estimation. However, if mortality is assumed to follow a piecewise linear trend with changing slope, the most recent data points are obviously more informative with respect to the prevailing AMT than data points further in the past. In fact, if a trend change occurred in the recent past, all data points before this trend change even blur the EMT derivation instead of supporting it. Consequently, the most recent data points should have more weight in the trend estimation. If, however, too much weight was given to just a few data points, the EMT could pick up trend changes in the AMT rather quickly, but at the expense of increasing the risk of wrongly identifying noise as a trend change.

3.2. Comparison of EMT Weight Specifications

We will now analyze and compüare the impact of different weight specifications for our example population of English and Welsh males. Obviously, the EMT should be a good estimate for the

AMT under the assumed dynamics for the AMT. Hence we chose weightings such that the EMT is as close as possible to the AMT 'on average', i.e. the AMT and EMT model components are consistent. The base case is 'constant weights' for an estimation period whose length is optimized. When deriving an EMT at time T, the weight for the data point in year $t \leq T$ is

$$w_{const}(t,T) = \begin{cases} 1 & , \text{if } T - h_{const} < t \leq T \\ 0 & , \text{if } t \leq T - h_{const} \end{cases}$$

where h_{const} is the length of the estimation period, which is the parameter needed in the base case. As alternatives we consider weights which decay linearly or exponentially going backward in time:

$$w_{lin}(t,T) = \max\left(0; 1 - \frac{1}{h_{lin}} \cdot (T-t)\right), t \leq T,$$

$$w_{exp}(t,T) = \frac{1}{\left(1 + \frac{1}{h_{exp}}\right)}, t \leq T.$$

Here, the parameters h_{lin} and h_{exp} determine the speed of the linear or exponential decay, respectively.[9] Note that the weights $w_{exp}(t,T)$ have already been proposed and used in a similar context by Börger et al. (2014).

In order to optimize the parameters h_{const}, h_{lin} and h_{exp} (which we will refer to as weighting parameters in what follows) we consider two different criteria and compare the results. In the first approach, we determine the weighting parameters such that the EMTs $\tilde{d}_T^{(i)}$, estimated from $\kappa_{1841}^{(i)}, \ldots, \kappa_T^{(i)}$, are as close as possible in terms of mean squared error to the AMTs $\hat{d}_T^{(i)}, i = 1, 2$. In the second approach, we minimize the mean squared error between the remaining cohort life expectancy of a 65-year old based on the EMT at time T, $\tilde{e}_{(65,T)}$, and the actual cohort life expectancy. We denote the latter by $\hat{e}_{(65,t_\omega)}$ and derive it from an AMT simulation from T to t_ω, where t_ω is the point in time when the cohort has died out. Here the weighting parameters for $\tilde{d}_T^{(1)}$ and $\tilde{d}_T^{(2)}$ are determined simultaneously and also the estimates for $\hat{\kappa}_T^{(i)}$ impact the weighting parameter calibrations. We consider the cohort life expectancy since it is highly dependent on the future mortality trend - in contrast to, e.g., the period life expectancy - and since it is similar in structure to present values of annuities which are of high relevance to actuaries. Thus, in total we compare six different sets of parameter estimates (two optimization criteria x three weight specifications).

In order to ensure that our parameter estimates are not biased by the fixed historical $\kappa_t^{(i)}, t \leq t_0$, we first simulate the AMT model 40 years into the future and optimize the weighting parameters

[9] In the case of exponential weighting, the weights are different from zero for all t. Nevertheless, a limited estimation period of sufficient length can be considered for practical purposes as the impact of data points far in the past is negligible due to the fast decay.

	weights	constant	linear	exponential
$\tilde{d}_T^{(1)}$	$h^{(1)}$	9.0	10.5	2.1
	mse	$1.45 \cdot 10^{-5}$	$1.31 \cdot 10^{-5}$	$1.35 \cdot 10^{-5}$
$\tilde{d}_T^{(2)}$	$h^{(2)}$	8.0	9.5	1.8
	mse	$3.19 \cdot 10^{-8}$	$2.88 \cdot 10^{-8}$	$2.95 \cdot 10^{-8}$

Table 4.: Optimal weighting parameters and mse's for both EMTs $\tilde{d}_T^{(1)}$ and $\tilde{d}_T^{(2)}$ based on the optimization criterion 'EMT close to AMT'

based on the EMTs at $T = t_0 + 40$. The mean squared errors (mse's) for each set of weighting parameter estimates are computed from 100,000 simulation paths.

Table 4 shows the optimal weighting parameters and the mse's for both EMTs $\tilde{d}_T^{(1)}$ and $\tilde{d}_T^{(2)}$ based on the first optimization criterion (EMT close to AMT). Figure 4 illustrates the resulting weights where the solid lines correspond to $\tilde{d}_T^{(1)}$ and the dashed lines to $\tilde{d}_T^{(2)}$. We observe that the weighting parameters for $\tilde{d}_T^{(2)}$ are generally smaller than for $\tilde{d}_T^{(1)}$. The main reason is that the trend change probability for $\hat{\kappa}_T^{(2)}$ is on average larger than for $\hat{\kappa}_T^{(1)}$, and therefore, $\tilde{d}_T^{(2)}$ needs to react to trend changes more often. Furthermore, we see that the linear and exponential weights outperform the constant weights in terms of mse. This is as expected since in the case of constant weights (i.e. the case of no weighting within an estimation period of optimal length) older data points which possibly lie before the most recent trend change materially bias the EMT derivation. This observation further supports the idea that weighting should be considered in general when deriving mortality trends from historical data. The mse differences between linear and exponential weights are rather small in this example of English and Welsh males, with linear weights having the slight edge. However, we have found that for other populations the exponential weights outperform the linear weights. This particularly depends on the likelihood of trend changes, their magnitudes, and the volatilities of the noise terms in the $\kappa_t^{(i)}$ processes. More precisely, the smaller the noise's volatility and the larger the trend change probability and magnitude, the better the exponential weights perform with their more pronounced focus on the most recent data points.

Table 5 displays the optimal weighting parameters with their corresponding mse's for the second optimization criterion ($\tilde{e}_{(65,T)}$ based on EMT close to actual life expectancy $\hat{e}_{(65,t_\omega)}$. It also gives probabilities that the life expectancy estimates $\tilde{e}_{(65,T)}$ are smaller (larger) than 95% (105%) of the actual life expectancies $\hat{e}_{(65,t_\omega)}$. Moreover, for comparison we have added results for a fixed estimation period of 30 years (i.e. constant weights within a predetermined 30-year estimation period) as an example for an estimation period which is typically used in practice.

First of all we observe that the optimal weighting parameters in Table 5 are very similar to those in Table 4. Thus, the concrete optimization criterion has only a small impact on the weights.

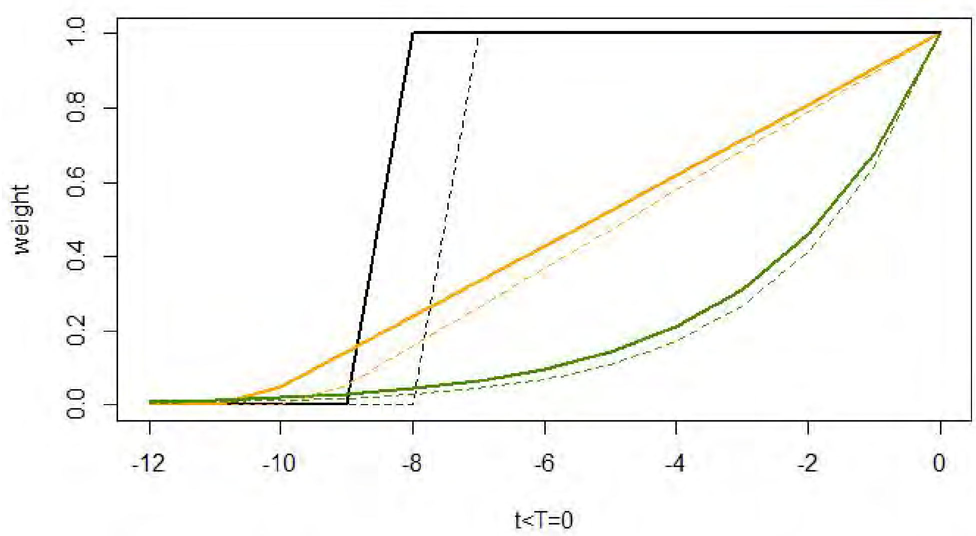


Figure 4.: Optimal weights for both EMTs $\tilde{d}_T^{(1)}$ (solid lines) and $\tilde{d}_T^{(2)}$ (dashed lines) based on the optimization criterion 'EMT close to AMT'

weights	constant	linear	exponential	fixed period
$h^{(1)}$	10.0	11.0	2.2	30
$h^{(2)}$	8.0	10.5	2.1	30
mse	1.59	1.56	1.58	3.04
$P(\tilde{e}_{65,T} < 95\% \cdot \hat{e}_{65,t_\omega})$	8.7%	8.5%	8.7%	15.0%
$P(\tilde{e}_{65,T} > 105\% \cdot \hat{e}_{65,t_\omega})$	7.3%	7.1%	7.0%	11.1%

Table 5.: Optimal weighting parameters and mse's for both EMTs $\tilde{d}_T^{(1)}$ and $\tilde{d}_T^{(2)}$ based on cohort life expectancy optimization criterion

Furthermore, the linear weights again outperform the alternatives in terms of mse. However, as also indicated by the misestimation probabilities, the differences between the three weighting approaches are fairly small. Results for the fixed estimation period of 30 years, on the other hand, are considerably worse. Compared to the case of an estimation period with optimal length (constant weights approach), the mse is about 91% and the combined misestimation probability 10.1 percentage points (i.e. 26.1% - 16.0%) larger. Clearly, available information on the dynamics of actual mortality should be taken into account when deriving best estimate mortality projections. The estimation period should be optimized based on this information, independent of whether weighting is applied or not.

We would like to stress that the optimal weighting parameters were derived assuming a concrete specification for the dynamics of the true (unknown) AMT. If the real dynamics of future mortality deviate from this assumption, other parameters might be optimal. We consider this as an aspect of model risk and will not further deal with this issue in this paper.

Coming back to our example of English and Welsh males aged 65 and older, the EMTs $\tilde{d}_T^{(1)}$ and $\tilde{d}_T^{(2)}$ at future points in time T should be derived using linear weights; the weighting parameters should be in the ranges as given in Table 4 and Table 5.

4. Distinction between AMT and EMT

In this section, we consider several practical examples to illustrate how important it is to differentiate between AMT and EMT. At the same time, we use numerical examples to explain how the combined AMT/EMT model from the previous sections can be applied in practice and how risk is misestimated if the difference between AMT and EMT is ignored. For the EMTs, we always use linear weights with weighting parameters as given in Table 5. We assume that annuitants'/pensioners' mortality rates are exactly as for males in England and Wales and that the number of annuitants/pensioners is large enough such that there is no unsystematic mortality risk. Furthermore, in oder to focus on effects resulting from stochastic mortality, we assume throughout this section that interest rates are deterministic and constant at 2%.

4.1. Example 1: Hedge Effectiveness of a Value Hedge

In this first example, we consider a pension fund whose members are all aged 45 at t_0. Some hedge provider has offered a value hedge agreement in order to mitigate the risk of the pension fund being underfunded at the members' retirement age of 65. Thus, the hedge agreement has a term of 20 years and pays the pension fund the best estimate liabilities according to up-to-date mortality assumptions at $T = t_0 + 20$. We assume that the hedge provider and the pension fund's trustees have agreed on deriving mortality assumptions at time T based on the EMT with optimal weighting (see Section 3.2). The premium for this hedge, payable at expiry, is the pension fund's expected liabilities as estimated at t_0 plus some hedging fee. Thus effectively, the hedge provider will fill up the pension fund's liabilities at time T if required and will receive money if the pension fund's liabilities are smaller than expected. While the hedge agreement transfers the longevity risk up to time T (including the reserving risk at time T), the pension fund will still be exposed to the risk that mortality after time T might not evolve as expected. Thus the remaining risk corresponds to the difference between the value of the actual pension payouts and its estimate at time T, i.e. the payoff of the value hedge. This is the risk we consider in more detail. In fact, it consists of two components: the risk that the mortality assumptions at time T are inadequate, i.e. the risk that the EMT differs from the AMT, and the risk that the AMT might change during the pension payout phase.

Assume for a moment that, when assessing this risk at t_0, the pension fund's trustees do not distinguish between AMT and EMT. Instead, they (wrongfully) assume that the AMT which they simulate up to time T is observable and compute the hedge payoff according to this AMT. Then for the trustees, the risk of the pension fund appears to be in the following difference of present values (PV) at time T:

$$PV_T(\text{pension payouts}|AMT_{t_\omega}) - PV_T(\text{pension payouts}|AMT_T), \tag{1}$$

where t_ω denotes the year when every pensioner has died and AMT_{t_ω} the realized AMT up to that year. However, in this case the trustees would neglect the first component of the risk the pension fund is exposed to, i.e. the risk of inadequate mortality assumptions at time T based on which the hedge payoff is derived. The actual risk the trustees are facing is rather given by

$$PV_T(\text{pension payouts}|AMT_{t_\omega}) - PV_T(\text{pension payouts}|EMT_T), \tag{2}$$

Figure 5 shows histograms for both differences based on 100,000 simulation paths and for 1£of annual pension payout. Obviously, the distribution for the risk described by equation (2) is significantly wider. In fact, its variance is 0.24 compared to only 0.15 in the case where the AMT is assumed to be observable. Thus, given the risk without a hedge, i.e. $Var(PV_T(\text{pension payouts}|AMT_{t_\omega}) = 1.90$, the trustees would assume a hedge effectiveness of $1 - \frac{0.15}{1.90} = 92.1\%$. The true hedge effectiveness however is only 87.2%.[10] Thus, ignoring the fact that the AMT is not observable makes the hedge appear better than it actually is.

4.2. Example 2: Safety Margins in Annuity Conversion Rates

In a second example, we now consider an insurer with a portfolio of unit-linked deferred annuities. We again assume that all policy holders are aged 45 at t_0. At the end of the accumulation phase (at $T = t_0 + 20$), the fund value is converted into a life-long annuity for which we assume a technical interest rate of 2%. However, the insurer does not want to convert at the actuarially fair rate, but intends to fix the rate such that the probability for losses from increasing longevity during the payout phase amounts to 1%. The surplus which arises in the 99% other cases may be (partially) credited to the policy holders as discretionary benefits. For the sake of simplicity,

[10]Note that we implicitly assume here that actual mortality evolves according to the dynamics in the AMT model component. Thus, we do not consider any model risk in the AMT model. Taking this model risk into account would reduce the hedge effectiveness in both cases, but the essence of this example would be unaffected.

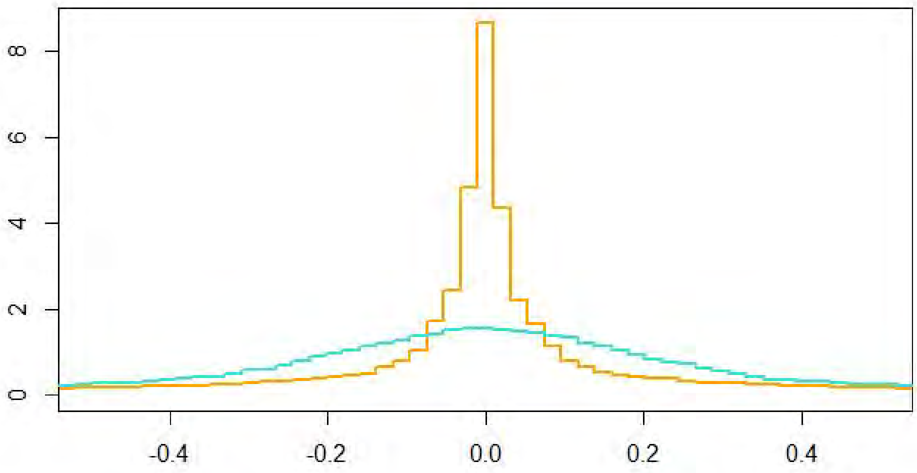


Figure 5.: Histograms for differences in Equation (1) (yellow line) and Equation (2) (blue line) for 1£of annual pension payout

we assume that the insurer is trying to achieve this by implementing a safety margin in form of a fixed percentage reduction of the actuarially fair conversion rate.

The risk of the insurer consists of the same two components as in the first example: the EMT might be an imprecise estimate of the AMT and actual longevity might increase after time T. Obviously, both components need to be taken into account when deriving the conversion rate reduction. Thus, the insurer is interested in the 99^{th} percentile of the distribution in Equation (2) (with annuity payouts instead of pension payouts) and printed in blue in Figure 5. He can then compute the percentage reduction of the actuarially fair annuity conversion rate as the ratio of this percentile and the average of the fair annuity conversion rates in the 100,000 simulation paths, i.e. $PV_T(\text{annuity payouts}|EMT_T) = 17.58$. With the 99^{th} percentile being equal to 1.68, the percentage reduction is then 9.6%.

However, if the insurer wrongfully assumes the AMT to be observable, he will determine the 99^{th} percentile of the distribution in Equation (1) (printed in yellow in Figure 5). In this case the percentage reduction of $PV_T(\text{annuity payouts}|AMT_T)$ (which the insurer would wrongfully assume to be the actuarially fair annuity conversion rate) would be only 7.8%. Thus, the insurer would underestimate the risk significantly. In fact, the probability of suffering losses during the annuity payout phase would be 1.7% instead of the intended 1%.

4.3. Example 3: SCR for Longevity Risk

As a final example, we consider an insurer who wants to develop an internal model for Solvency II. Thus the insurer needs to determine the SCR for longevity risk (as any SCR) as the 99.5% Value-at-Risk of the basic own funds (which is essentially the difference between assets and liabilities) over a 1-year time horizon. In this setting, longevity risk consists of two components: the risk of less than expected annuitants dying during the next year and the risk of an unfavorable change during the next year in longevity assumptions for the time beyond. Typically, the second component is more relevant.

For simplicity, we assume that only the liabilities are exposed to longevity risk (i.e. there are no hedge instruments on the asset side) and that there is no loss-absorbing capacity of technical provisions. Following Börger (2010), the insurer can then compute the SCR at time t as the 99.5^{th} percentile of the change in liabilities ΔBEL_{t+1} from time t to $t+1$ due to longevity:

$$\Delta BEL_{t+1} = (BEL_{t+1} + CF_{t+1}) \cdot \frac{1}{(1+r)} - BEL_t,$$

where r is the risk-free interest rate (2% in our case), CF_{t+1} denotes the cashflows (benefits plus expenses minus premiums) of the longevity prone business between t and $t+1$, and BEL_t is the best estimate liabilities at time t. In order to evaluate the distribution of ΔBEL_{t+1}, the insurer needs a combined AMT/EMT model. More precisely, the EMT component is needed to compute the best estimate liabilities and the AMT component to simulate the realized mortality evolution over the 1-year time horizon.

As mentioned above, the change in best estimate liabilities and thus the change in the EMT over one year essentially determines the SCR. Therefore, it is important to ensure that the assumed variability in the EMT over time is realistic. If the impact of the additional data point (for the one year) on the EMT update is too small (e.g. in case of a long estimation period), annual changes of the EMT will likely be underestimated. If the impact of the additional data point is too large, annual EMT changes will be overestimated. Here, the findings in Section 3.2 can be very helpful. The optimal weights give a clear recommendation how large the impact of new data points should be and thus by how much the EMT might change over one year.

Now assume that - as part of the internal model validation process - the insurer wants to analyze how large the SCR at some point in time T can be, depending on the mortality evolution up to that point in time. Besides the new data point in year $T+1$ and the weighting in the EMT derivation, the structure (in particular potential trend changes) in the mortality data up to time T is the main determinant for EMT changes over one year. As in the second example, the insurer sets $T = t_0 + 20$. He simulates 10,000 outer paths up to time T in order to average over virtually all potential mortality evolutions. Then for each of these outer paths, 10,000 inner

1-year paths are simulated in order to empirically derive the 99.5^{th} percentile of the change in liabilities ΔBEL_{T+1}. Exemplarily for a portfolio of annuitants, he considers the cohort of 75-year olds who receive combined annual payments of 1£at time T. For simplicity, there are no expenses and premiums.

The blue curve in Figure 6 is the density of the distribution for the SCR at time T depending on the mortality evolution up to T. On average, the SCR is 0.40 with a (sample) variance of 0.0064 and (sample) standard deviation of 0.08. The considerable coefficient of variation of about 20% shows that the SCR can vary significantly due to the variability in the EMT changes over one year.

Once again, we want to illustrate how important it is to distinguish between AMT and EMT. If the insurer wrongfully assumes the AMT to be observable and computes the best estimate liabilities based on the AMT, he will obtain the distribution given in yellow in Figure 6. On average, the SCR would be 0.73 and thus significantly larger than in the EMT case. Clearly, longevity risk would be overstated substantially. The reason is that the AMT is very likely to exhibit massive trend changes in the extreme 1-year scenarios which are relevant for the SCR. Annual changes in the EMT, on the other hand, are not that strong as the EMT does not pick up trend changes instantly. Due to the weighted linear regression, trend changes are only taken into account over several years.

While assuming the AMT to be observable implies an overestimation of risk in this example, it was the opposite in the other examples. The fact that misestimations can occur in both directions once again underlines the necessity to clearly distinguish between AMT and EMT.

5. Conclusion

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 have specified a concrete model that simultaneously and consistently projects AMT and EMT. In line with historic observations, the AMT is modeled as a piecewise linear trend with random changes in slope. Since the AMT at the start of a simulation is unknown, parameter uncertainty needs to be taken into account, and we show how this can be done. Given the assumed dynamics for the AMT, the EMT is best derived by a weighted linear regression on

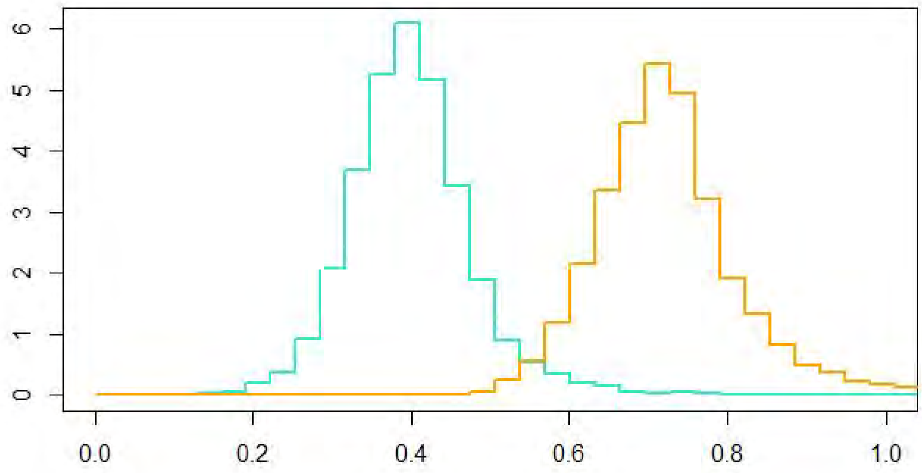


Figure 6.: Histograms for the SCR at time T depending on the mortality evolution up to T under the assumption that the AMT is observable (yellow line) or unobservable (blue line)

the most recent observable data. In our modeling framework, this can be done on the output of the AMT model component directly, which makes the framework very efficient in Monte Carlo simulations.

Finally, we have considered several practical examples to show how our model can be applied and to quantify the error that results if the difference between AMT and EMT is ignored: If the AMT is wrongfully assumed observable, risk is significantly misestimated in all our examples - sometimes underestimated, sometimes overestimated. Therefore, this research should be of interest to anybody who is concerned with mortality projections or longevity risk. This particularly holds for public pension schemes, pension funds, and insurers, but also for regulators and auditors.

References

Bauer, D., Börger, M., and Russ, J. (2010). On the Pricing of Longevity-Linked Securities. *Insurance: Mathematics and Economics* 46(1): 139–149.

Bauer, D., Börger, M., Russ, J., and Zwiesler, H. (2008). The Volatility of Mortality. *Asia-Pacific Journal of Risk and Insurance* 3: 184–211.

Börger, M. (2010). Deterministic Shock vs. Stochastic Value-at-Risk - An Analysis of the Solvency II Standard Model Approach to Longevity Risk. 31(2): 225–259.

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

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

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

Cairns, A. J. (2013). Robust Hedging of Longevity Risk. *Journal of Risk and Insurance* 80(3): 621–648.

Cairns, A. J., Blake, D., Dowd, K., and Coughlan, G. (2014). Longevity Hedge Effectiveness: A Decompositionn. *Quantitative Finance* 14(2): 217–235.

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

Cairns, A. J. and El Boukfaoui, G. (2019). Basis risk in Index-Based Longevity Hedges: A Guide for Longevity Hedgers. *North American Actuarial Journal*: 1–22.

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

Hunt, A. and Blake, D. (2015). Forward Mortality Rates in Discrete Time I: Calibration and Securities Pricing. *Working Paper, Cass Business Schooll*.

Hunt, A. and Blake, D. (2016). Forward mortality Rates in Discrete Time II: Longevity Risk and Hedging Strategies. *Working Paper, Cass Business School*.

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

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

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

Schupp, J. (2019). On the Modeling of Variable Mortality Trend Processes. *Working Paper, Ulm University.*

Zhu, N. and Bauer, D. (2011). Coherent Modeling of the Risk in Mortality Projections: A Semi-Parametric Approach. *Tech. Rep. 678, Georgia State University, Atlanta.*

4. Calibrating Mortality Processes with Trend Changes to Multi-Population Data

Source:
Börger, M., Schoenfeld, J., and Schupp, J. (2019). Calibrating Mortality Processes with Trend Changes to Multi-Population Data, Working Paper, Ulm University.

The Society of Acutaries will publish the article based on this working paper version, i.e. the status of the paper is 'accepted'. With publication, the Society of Actuaries owns the copyright. However, a publication as part of a dissertation is kindly permitted by the Society of Actuaries.

Calibrating Mortality Processes with Trend Changes to Multi-Population Data

Matthias Börger
Technische Hochschule Nürnberg Georg Simon Ohm, Nuremberg
Keßlerplatz 12, 90489 Nürnberg, Germany

Justin Schoenfeld
Institute of Insurance, Ulm University
Helmholtzstraße 20, 89081 Ulm, Germany

Johannes Schupp
Institute for Finance and Actuarial Sciences (ifa), Ulm & Institute of Insurance, Ulm University
Lise-Meitner-Straße 14, 89081 Ulm, Germany
Phone: +49 731 20644241
Email: j.schupp@ifa-ulm.de

Abstract

The uncertainty in future mortality rates is typically quantified by stochastic mortality models. To this end, the time dependent parameters in these models are projected by stochastic processes. Thus, the choice of these processes and their calibration have a crucial impact on estimates of future uncertainty. Since the commonly applied random walk with drift process has some structural shortcomings (see e.g. Börger et al. (2014)), alternative processes with random changes in the long-term mortality trend have been proposed by several authors. Such trend changes can be observed in the historical data for almost every population. However, data on such trend changes is sparse, and thus, the parameter estimation of these trend processes involves a significant degree of uncertainty. In this paper, we explain how data on trend changes from several populations can be combined in order to improve the reliability of trend process calibrations for individual populations. We discuss different assumptions on the "similarity" of parameters for different populations and implement those assumptions for the case of the trend change process proposed by Börger and Schupp (2018). In a numerical example we find

that the impact on parameter estimates can be substantial. Thus, relying on the sparse data for individual populations only can lead to significant misestimation of future mortality and its uncertainty.

1. Introduction

Annuity providers, pension funds, life insurers, and social security systems heavily rely on forecasts of future mortality. For risk management purposes in particular, stochastic mortality models are required in order to quantify the uncertainty in future mortality rates. Therefore, a large number of such models have been developed over the last decades, e.g. the models of Lee and Carter (1992) and Cairns et al. (2006). Most of these models contain one or more time dependent parameters, often referred to as period effects. These parameters are typically projected into the future by stochastic trend processes in order to generate scenarios of future mortality. Obviously, the choice of a specific stochastic trend process has a crucial impact on the forecasts, both in terms of the central, median, or best estimate projection scenario as well as in terms of the uncertainty around this scenario.

Often, a random walk with drift is applied. It is a simple process with a clear parameter interpretation, and it nicely extrapolates the rather linear trends which have been observed in the period effects for many populations over the last decades. However, the random walk with drift also has some structural shortcomings. Most prominently, long-term uncertainty is often underestimated since the drift is fixed and stochasticity is only contained in the annual innovations (see, e.g., Börger et al. (2014) or Börger and Schupp (2018)). This issue is illustrated by Figure 1 which shows the logarithm of probabilities of death for 65-year old males in different countries from all over the world; the data has been obtained from the Human Mortality Database (2019). We observe the aforementioned rather linear trends in most recent decades, but we also see that the drifts/trends in the log probabilities of death have also changed significantly once in a while in the past. A random walk with fixed drift is not able to generate such patterns, and in particular, it does not adequately allow for the uncertainty which arises from potential trend changes.

For this reason several authers have proposed trend processes which explicitly allow for trend changes. Hainaut (2012) uses a random walk with drift where the parameters of the random walk (and its drift in particular) can switch between different regimes. Hunt and Blake (2015) allow for a continuous range of future mortality trends by simulating drift changes by a Pareto distribution. Other authors like Sweeting (2011) and Börger and Schupp (2018) have skipped the random walk concept entirely and have instead proposed trend stationary processes with piecewise linear trends where the slope of the linear trend changes randomly over time. However, all these different approaches to stochastic modeling of trend changes have one thing in common: Data on trend changes and their magnitudes is sparse, and therefore, uncertainty in the estimation of the processes' parameters is substantial in general. Even for populations with rather long data histories, typically only a few historical trend changes can be observed. For populations with shorter data histories, e.g. starting after World War II, reliable param-

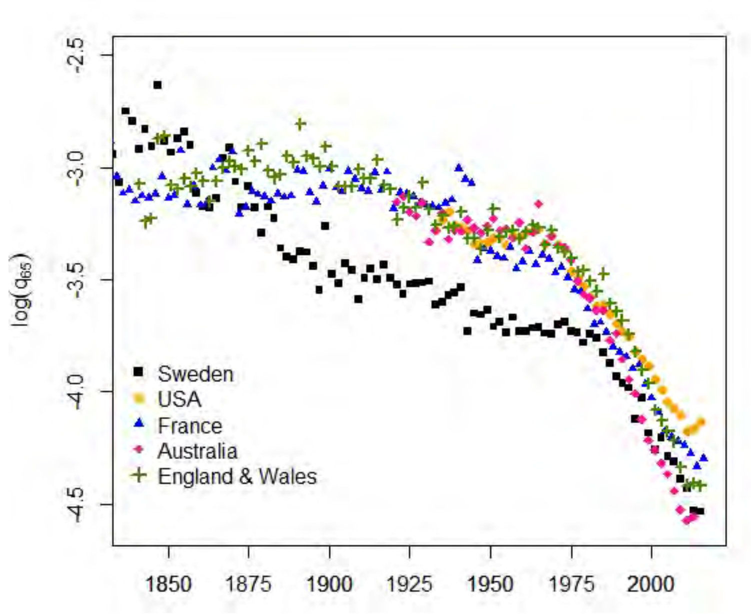


Figure 1.: Logarithm of probabilities of deaths for 65-year old males in selected countries

eter estimations are often impossible. This clearly limits the applicability of trend process with random trend changes so far. This paper addresses the issue of parameter estimation for these processes and makes propositions how reasonable calibrations can be obtained also for populations with short data histories.

A common concept for reducing parameter uncertainty is to enlarge the data base for parameter estimation by aggregating data from several populations. In this paper, we analyze how this idea can be applied in order to improve the parameter estimation of trend change processes. Exemplarily, we do this for the trend process proposed by Börger and Schupp (2018) and further refined in Schupp (2019) This allows us to illustrate different data aggregation approaches which could easily be applied to other trend change processes, too.

The remainder of this paper is structured as follows: In the following section, we introduce the trend process of Börger and Schupp (2018) and its application within the CBD mortality model of Cairns et al. (2006). We briefly summarize the parameter estimation in a single population setting and discuss the issue of parameter uncertainty. Finally, we provide a concrete example of the trend process for the population of US males. In Section 3, we compare parameter estimates as well as the respective uncertainties for different populations. Possible approaches to improving parameter estimates and reducing the associated uncertainty are then presented in Section 4. The theoretical discussion of these approaches is complemented by a numerical example in Section 5. Finally, Section 6 concludes.

2. Trend Change Mortality Process

2.1. Specification of Trend Change Process

Börger and Schupp (2018) and Schupp (2019) apply their trend change process to the time dependent parameters in the well-known CBD mortality model of Cairns et al. (2006). In the CBD model, annual probabilities of death are described as

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

where $\bar{x}$ is the average age of the age range under consideration. The time dependent parameter $\kappa_t^{(1)}$ determines the general level of mortality, whereas the slope parameter $\kappa_t^{(2)}$ describes the increase of mortality with age.

Building on the observations from Figure 1, Börger and Schupp (2018) and Schupp (2019) propose a trend process which projects piecewise linear trends with random changes in slope. For any future year t, the 'observable' processes $\kappa_t^{(1)}$ and $\kappa_t^{(2)}$ are modeled as the sums of underlying but unobservable true mortality processes and random noise terms:

$$\kappa_t^{(i)} = \hat{\kappa}_t^{(i)} + \epsilon_t^{(i)}, i = 1, 2.$$

The noise term $\epsilon_t^{(i)}$ accounts for annual fluctuations which are, e.g., due to flu waves, very hot summers, or catastrophes. The vector $\epsilon_t = \left(\epsilon_t^{(1)}, \epsilon_t^{(2)}\right)$ is assumed to follow a two-dimensional Normal distribution with mean zero and covariance matrix Σ.[1]

The underlying mortality processes $\hat{\kappa}_t^{(i)}, i = 1, 2$ are projected linearly with current slopes $\hat{d}_t^{(i)}, i = 1, 2$ as slopes:

$$\hat{\kappa}_t^{(i)} = \hat{\kappa}_{t-1}^{(i)} + d_t^{(i)}, i = 1, 2.$$

The slopes remain unchanged until the next trend change occurs:

$$\hat{d}_t^{(i)} = \begin{cases} \hat{d}_{t-1}^{(i)} & \text{, if no trend change occurred in } t-1 \\ \hat{d}_{t-1}^{(i)} + \lambda_{t-1}^{(i)} & \text{, if a trend change by } \lambda_{t-1}^{(i)} \text{ occured in } t-1 \end{cases}$$

The trend change intensities $\lambda_t^{(i)}$, are derived as the product of their absolute magnitudes $M_t^{(i)}$ and their signs $S_t^{(i)}$:

$$\lambda_t^{(i)} = S_t^{(i)} \cdot M_t^{(i)}, i = 1, 2.$$

[1]Note that the covariance matrix is assumed constant over time even though heteroscedasticity can usually be observed in the historical data (see Figure 1). However, since the noise does not impact long-term mortality evolutions (in contrast to the innovations in the random walk), this simplification appears appropriate. In the parameter estimation, heteroscedasticity is accounted for though.

Based on analyses of historical trend changes, Börger and Schupp (2018) propose modeling the magnitudes $M_t^{(i)}$ by Lognormal distributions $\mathcal{LN}(\mu^{(i)}, \sigma^{(i)})$. For the signs $S_t^{(i)}$, they use a Bernoulli distribution with attainable values -1 and 1 and probability 0.5 each. The probabilities of observing trend changes in $\hat{\kappa}_t^{(i)}$ in any particular year are denoted by $p^{(i)}, i = 1, 2$. Moreover, trend changes in $\hat{d}_t^{(1)}$ and $\hat{d}_t^{(2)}$ are assumed to occur independently as indicated by the occurrences of trend changes in the historical data for a large set of populations (see Börger and Schupp (2018)).

The decomposition of the trend change intensity into absolute magnitude and sign has several convenient implications. First, the distribution of future trend changes is symmetric, i.e. the slope increases and decreases with equal probability and magnitude. Thus, the prevailing trend (even though unobservable)[2] is always the best estimate for the trend at any future point in time. Furthermore, the distribution of $\lambda_t^{(i)}$ has no probability mass at zero and only very little mass around zero. Thus, simulated trend changes can be considered as rather 'significant'. At the same time, the heavy tail of the Lognormal distribution implies that strong trend changes can occur which is in line with some of the trend changes we observe in Figure 1 (e.g. for Sweden around 1980).

2.2. Parameter Estimation and Uncertainty

In this subsection we explain how the parameters of the trend processes $\kappa_t^{(i)}$ can be estimated from data for an individual population, and we particularly discuss the uncertainty involved. We assume that historical realizations $\kappa_t^{(i)}, t \leq t_0$ are given, where t_0 denotes the final year of the historical data, i.e. the starting point of a simulation. The parameters to be estimated are:

- the probabilities of observing a trend change in a certain year, $p^{(i)}$
- the parameters of the Lognormal distributions for the trend change magnitudes, $\mu^{(i)}$ and $\sigma^{(i)}$,
- the starting values for the underlying but unobservable trend processes $\hat{\kappa}_{t_0}^{(i)}$,
- the prevailing slopes of these trend processes $\hat{d}_{t_0}^{(i)}$,
- and the covariance matrix Σ of the two-dimensional noise vector ϵ_t.

[2]See Börger et al. (2019) for a thorough discussion on the observability of mortality data and trends as well as implications for mortality modeling in practical applications.

The parameter estimation is carried out separately for each $\kappa_t^{(i)}$ process, and the covariance matrix Σ is estimated in the very last step. As Schupp (2019) explains, parameter estimation is complex due to the common dependence of realized $\kappa_t^{(i)}, t \leq t_0$ on potential but unknown trend changes in previous years. In particular, a full maximum likelihood estimation of all model parameters seems impossible from a practical point of view. Therefore, we rely on the pseudo maximum likelihood approach proposed by Schupp (2019). An iterative algorithm determines, for any fixed number of trend changes k, (a) the specific realization of the underlying trend process with k trend changes, $\hat{\kappa}_{t,k}^{(i)}, t \leq t_0$, which is closest to the actual data $\kappa_t^{(i)}, t \leq t_0$ in terms of likelihood, and (b) parameter values as (pseudo) maximum likelihood estimates which are consistent with this trend process realization. More precisely, starting with some initial parameter values, the best possible trend process realization $\hat{\kappa}_{t,k}^{(i)}, t \leq t_0$ is determined. Then the parameter values are updated based on this realization and the contained trend changes in particular. Next, these updated parameter values are applied in an update of $\hat{\kappa}_{t,k}^{(i)}, t \leq t_0$. This iterative algorithm typically converges after only very few steps, and we refer to Schupp (2019) for more details.

From the sets of (pseudo) maximum likelihood estimates $p_k^{(i)}, \mu_k^{(i)}, \sigma_k^{(i)}, \hat{\kappa}_{t_0,k}^{(i)}, \hat{d}_{t_0,k}^{(i)}$, and $\Sigma_{(i,i),k}$ for different numbers of trend changes k, the final parameter estimates and their respective uncertainties can be determined as follows (a numerical example is provided in the next subsection). The trend change parameters $\theta^{(i)} = \left(p^{(i)}, \mu^{(i)}, \sigma^{(i)}\right)$ are computed as weighted averages of the estimates $\theta_k^{(i)} = \left(p_k^{(i)}, \mu_k^{(i)}, \sigma_k^{(i)}\right)$:

$$\theta^{(i)} = \sum_k w_k^{(i)} \cdot \theta_k^{(i)}, \tag{1}$$

where the weights $w_k^{(i)}$ sum up to one and are based on a relative likelihood measure similar to Bayesian weights. Thus, most weight is applied to the $\theta_k^{(i)}$ for which the respective best possible trend process realization fits the actual data best; for more details on these weights, we refer to Schupp, 2019. Even though k can range from zero to the number of data points in theory, in practice only a few values for k need to be considered since most weights $w_k^{(i)}$ are effectively zero.

In estimating the trend change parameters, uncertainty mainly arises from two sources: First, the actual number of historical trend changes cannot be determined exactly as the random noise affects the search for the best possible trend process realizations $\hat{\kappa}_{t,k}^{(i)}, t \leq t_0$. In other words, the trend process realizations which are found for different numbers of trend changes k may fit the actual data similarly well in terms of likelihood. This issue is already taken into account above when deriving the central parameter estimates as weighted averages. Nevertheless, the uncertainty around these central parameter estimates must not be neglected. The second source of uncertainty lies in the estimation of the trend change parameters from a limited (and typically small) number of trend changes k, even if we assume to know the exact number. In order to quantify the overall parameter uncertainty, we search for a combined (approximate) standard error for both sources of uncertainty. With respect to the second source of uncertainty, the (pseudo) maximum likelihood estimation provides covariance matrices of (approximate) standard errors for each value of k which we denote by $SE_k^{(i)}$. Moreover, we assume that this component of parameter uncertainty in $\theta_k^{(i)}$ can be expressed by some distribution $F_k^{(i)}$. Concerning the unclear number of actual trend changes, the weights $w_k^{(i)}$ provide probabilities for each possible number of trend changes k. Therefore, we assume that the overall parameter uncertainty in $\theta^{(i)}$ is described by the distribution $F^{(i)} = \sum_k w_k^{(i)} \cdot F_k^{(i)}$. Then it can be shown that the covariance matrix of overall standard errors is given by

$$SE^{(i)} = \sum_k w_k^{(i)} \cdot \left(SE_k + \left(\theta_k^{(i)} - \theta^{(i)}\right) \cdot \left(\theta_k^{(i)} - \theta^{(i)}\right)' \right). \tag{2}$$

To summarize the estimation of the trend change parameters, we have central estimates according to Equation (1) and a covariance matrix of standard errors according to Equation (2). Thus, when projecting the future mortality evolution, parameter uncertainty can be taken into account by randomly drawing, for each simulation path, parameter values from a suitable distribution with according mean vector and covariance matrix.

For the estimation of the starting values of a simulation, $\hat{\kappa}_{t_0}^{(i)}$ and $\hat{d}_{t_0}^{(i)}$, and the associated uncertainty, we follow a different approach. Parameter uncertainty here mainly arises from the uncertainty when the most recent trend change has occurred. When determining the best possible trend process realizations $\hat{\kappa}_{t,k}^{(i)}, t \leq t_0$, a trend change in recent years may be detected for some values of k but not for others. In order to confirm the (non-)occurrence of such a trend change, a couple of additional years of data would be required. As long as this data is not available, however, uncertainty with respect to the starting values may be substantial, depending on the magnitude of the potential trend change.[3] In a simulation of future mortality, obviously both cases, i.e. with and without the potential trend change, should be taken into account. However, since the potential trend change either has occurred or not, we cannot specify central parameter estimates. Instead, we have different estimates for $\hat{\kappa}_{t_0,k}^{(i)}$ and $\hat{d}_{t_0,k}^{(i)}$ with different probabilities/weights $w_k^{(i)}$, and for each simulation path, starting values should be drawn randomly from this empirical distribution. A numerical example is provided in the next subsection.

A central estimate for the covariance matrix Σ of the noise vector ϵ_t can be derived analogously to Equation (1). If we denote by $\Sigma_{k,m}$ the sample covariance matrix for the case of k trend changes in $\hat{\kappa}_{t,t\leq t_0}^{(1)}$ and m trend changes in $\hat{\kappa}_{t,t\leq t_0}^{(2)}$, the final estimate is given as

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

[3] In comparison, when the most recent trend change is assumed to be known, i.e. for a fixed k, the uncertainty in regressing the starting values from the available data appears negligible.

Compared to the uncertainty in the trend change parameters, the uncertainty in Σ is negligible as it is estimated from rather large samples of residuals. Furthermore, the impact of the noise vector ϵ_t in projections of future mortality is very limited. Therefore, parameter uncertainty can be ignored here.

2.3. Example Calibration

In this subsection we present a full model calibration for US males including specifications of the parameter uncertainties involved. We use the entire data set which is available in the Human Mortality Database (2019) for ages 60 to 109, i.e. from 1933 to 2016. Figure 2 shows the historic trend processes $\kappa_t^{(1)}$ and $\kappa_t^{(2)}$ and the best possible realizations for the underlying trend processes, $\hat{\kappa}_{t,k}^{(1)}$ and $\hat{\kappa}_{t,k}^{(1)}$, for the relevant numbers of trend changes k, i.e. for those k with $w_k^{(i)} > 0$. Table 1 and Table 2 provide the corresponding parameter estimates and weights/probabilities. We find that the number of actual trend changes is very likely to lie between 2 and 5 or 6, respectively. While some parameter values are very similar for different k, e.g. the $\mu_k^{(i)}$, others vary substantially. Unsurprisingly, this particularly holds for the trend change probabilities $p_k^{(i)}$. This observation clearly underlines why it is important to account for the fact that one cannot clearly observe the exact number of trend changes.

By applying Equation (1), the central parameter estimates can then be derived:

$$\theta^{(1)} = \left(p^{(1)}, \mu^{(1)}, \sigma^{(1)}\right) = (0.0451, -4.4867, 0.2290)$$

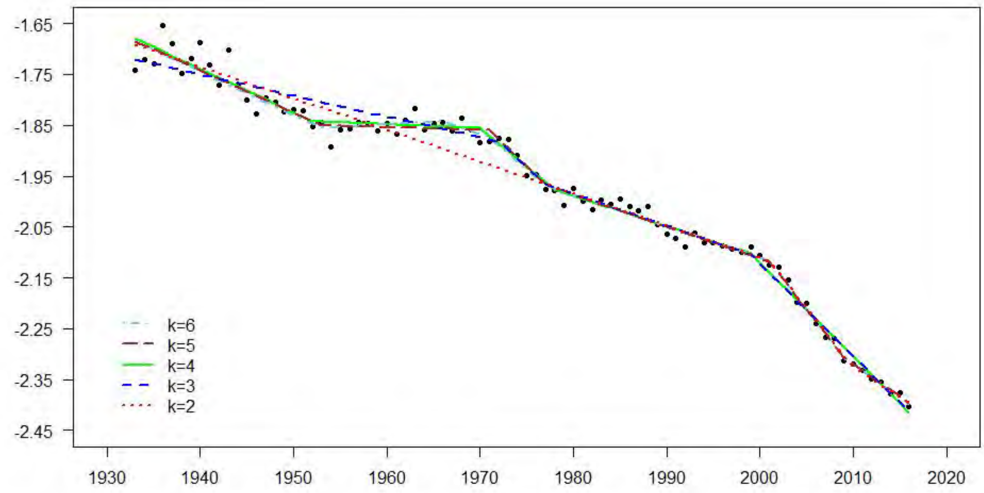
$$\theta^{(2)} = \left(p^{(2)}, \mu^{(2)}, \sigma^{(2)}\right) = (0.0480, -7.3466, 0.3932)$$

Furthermore, from Equation (2) we obtain the following covariance matrices of standard errors:

$$SE^{(1)} = \begin{pmatrix} 6.906 \cdot 10^{-4} & -1.245 \cdot 10^{-3} & 2.812 \cdot 10^{-4} \\ -1.245 \cdot 10^{-3} & 2.695 \cdot 10^{-2} & 2.047 \cdot 10^{-3} \\ 2.812 \cdot 10^{-4} & 2.047 \cdot 10^{-3} & 8.691 \cdot 10^{-3} \end{pmatrix}$$

$$SE^{(2)} = \begin{pmatrix} 5.929 \cdot 10^{-4} & -2.127 \cdot 10^{-4} & -1.108 \cdot 10^{-4} \\ -2.127 \cdot 10^{-4} & 4.469 \cdot 10^{-3} & -1.534 \cdot 10^{-3} \\ -1.108 \cdot 10^{-4} & -1.534 \cdot 10^{-3} & 2.396 \cdot 10^{-2} \end{pmatrix}$$

Comparing the (one-dimensional) standard errors, i.e. the roots of the diagonal entries of the $SE^{(i)}$, with the central parameter estimates, we find coefficients of variation between 40% and 60% for the $p^{(i)}$ and $\sigma^{(i)}$. Thus, parameter uncertainty is huge for these parameters. For the $\mu^{(i)}$, on the other hand, the coefficients of variation are only around -3%.



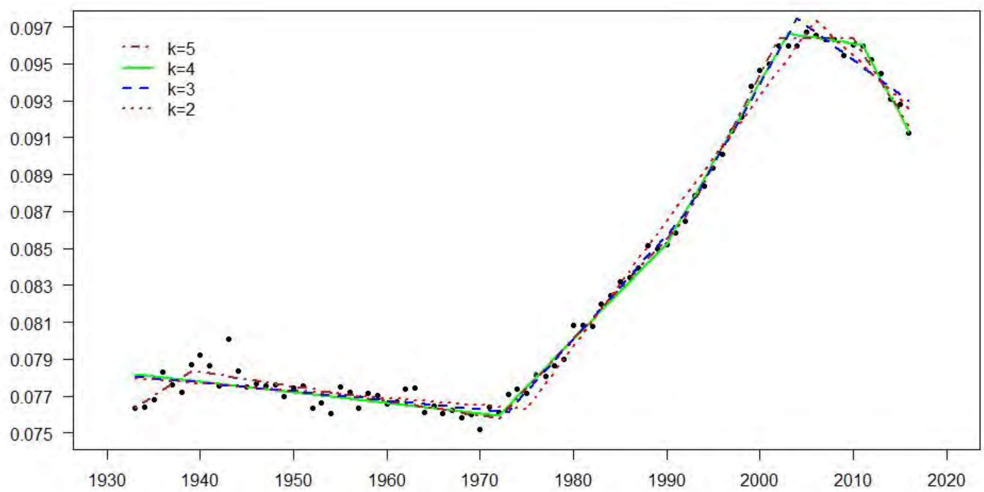


Figure 2.: Historical trend processes $\kappa_t^{(1)}$ (top) and $\kappa_t^{(2)}$ (bottom) for US males and best possible realizations for the underlying trend processes, $\hat{\kappa}_{t,k}^{(1)}$ and $\hat{\kappa}_{t,k}^{(2)}$

k	$p_k^{(1)}$	$\mu_k^{(1)}$	$\sigma_k^{(1)}$	$\hat{\kappa}_{t_0,k}^{(1)}$	$\hat{d}_{t_0,k}^{(1)}$	$w_k^{(1)}$
2	0.0244	−4.2823	0.1993	−2.3957	−0.0118	0.214
3	0.0366	−4.4642	0.0814	−2.4133	−0.0182	0.034
4	0.0488	−4.5431	0.2414	−2.4148	−0.0128	0.668
5	0.0610	−4.397	0.3033	−2.3959	−0.0128	0.008
6	0.0732	−4.5850	0.2607	−2.3981	−0.0132	0.076

Table 1.: Estimates for trend change parameters $p_k^{(1)}$, $\mu_k^{(1)}$, $\sigma_k^{(1)}$ and starting values $\hat{\kappa}_{t_0,k}^{(1)}$,$\hat{d}_{t_0,k}^{(1)}$ for different numbers of trend changes k and corresponding best possible realizations $\hat{\kappa}_{t,k}^{(1)}$ for the underlying trend process

k	$p_k^{(2)}$	$\mu_k^{(2)}$	$\sigma_k^{(2)}$	$\hat{\kappa}_{t_0,k}^{(2)}$	$\hat{d}_{t_0,k}^{(2)}$	$w_k^{(2)}$
2	0.0244	−7.0009	0.2368	$9.259 \cdot 10^{-2}$	$-4.753 \cdot 10^{-4}$	0.024
3	0.0366	−7.7383	0.5627	$9.259 \cdot 10^{-2}$	$-3.739 \cdot 10^{-4}$	0.079
4	0.0488	−7.3497	0.3858	$9.132 \cdot 10^{-2}$	$-9.440 \cdot 10^{-4}$	0.839
5	0.0610	−7.3972	0.3294	$9.166 \cdot 10^{-2}$	$-7.885 \cdot 10^{-4}$	0.057

Table 2.: Estimates for trend change parameters $p_k^{(2)}$, $\mu_k^{(2)}$, $\sigma_k^{(2)}$ and starting values $\hat{\kappa}_{t_0,k}^{(2)}$,$\hat{d}_{t_0,k}^{(2)}$ for different numbers of trend changes k and corresponding best possible realizations $\hat{\kappa}_{t,k}^{(2)}$ for the underlying trend process

The figure and tables also show substantial parameter uncertainty for the starting values $\hat{\kappa}_{t_0,k}^{(i)}$ and $\hat{d}_{t_0,k}^{(i)}$. For $\kappa_t^{(1)}$ and $k = 3, 4$, the most recent trend change is detected in 1999; the probability for this being the most recent actual trend change is about 70%. However, there is also a 30% chance that the most recent trend change occurred in fact in 2009 or 2010 as detected for $k = 2, 5, 6$.[4] Similarly, we find the most recent trend change in $\kappa_t^{(2)}$ in 2004 or 2006 (for $k = 2, 3$ and with probability of about 10%) or in 2010 or 2011 (for $k = 4, 5$ and with probability of about 90%). In both cases, the $\hat{d}_{t_0,k}^{(i)}$ differ with k in particular, and this uncertainty should be taken into account in projections of future mortality. In contrast to the example at hand, the most recent trend change may be very clear for other populations and/or at other points in time. Thus, parameter uncertainty in the starting values is highly case specific and may even be negligible in some cases.

Also the question whether parameter uncertainty in the starting values can be reduced by combining insights from several populations can only be answered individually for each specific case. Possibly, a potential trend change for one population can be confirmed by detecting similar trend changes for other populations with likely the same reason of occurrence. However, trend changes may also be specific to a single population such that insights from other populations can even be misleading. Due to this need for a case specific analysis, we will not discuss this question further in this paper.

For completeness, the estimate for the covariance matrix of the noise vector is

$$\Sigma = \begin{pmatrix} 1.773 \cdot 10^{-4} & 3.401 \cdot 10^{-6} \\ 3.401 \cdot 10^{-6} & 2.092 \cdot 10^{-7} \end{pmatrix}$$

In order to illustrate the issue of parameter uncertainty, Figure 3 shows projections of remaining period life expectancies of 65-year old US males with and without parameter uncertainty. Parameter uncertainty in the starting values is taken into account by randomly drawing from

[4] Due to the noise it is obviously difficult to exactly date a trend change. Therefore, we assume that the same trend change may be detected in subsequent years for different k.

the empirical distribution of starting values as given in Table 1 and Table 2 for each simulation path. For the case without parameter uncertainty, the starting values with the largest probability $w_k^{(i)}$ are considered. The path dependent trend change parameters $p^{(i)}$, $\mu^{(i)}$, and $\sigma^{(i)}$ are drawn from correlated (one-dimensional) Beta, Normal, and Gamma distributions, respectively.[5] The Beta and Gamma distributions are chosen in order to ensure that the trend change probabilities always lies between zero and one and that the standard deviations of the trend change magnitudes are always positive. Slight inconsistencies with the standard errors being derived under the assumption of Normality are accepted.

We find that the mean projection changes significantly in the case where parameter uncertainty is taken into account. The reason is that, for both $\kappa_t^{(1)}$ and $\kappa_t^{(2)}$, the most recent trend change could not be clearly determined and that different starting slopes are assigned significant probabilities in particular. The prediction intervals widen only slightly when allowing for parameter uncertainty. This is somewhat surprising at first sight given the substantial standard errors above. However, the parameters $p^{(1)}$ and $\mu^{(1)}$ are negatively correlated for US males which means that in case the likelihood of trend changes is higher (larger $p^{(1)}$, their magnitude is likely to be smaller (smaller $\mu^{(1)}$). For other populations, however, we have found positive correlation between these parameters leading to significantly wider prediction intervals when allowing for parameter uncertainty. This illustrates once again that it is reasonable to combine observations from different populations. The case with (population specific) parameter uncertainty will also be the benchmark in the numerical example later on.

3. Comparison of Population Specific Trend Process Calibrations

We commence our multi-population analysis with a comparison of population specific estimates of the trend change parameters $\theta^{(i)} = \left(p^{(i)}, \mu^{(i)}, \sigma^{(i)}\right), i = 1, 2$ and their associated uncertainty. To this end, we calibrate the CBD model and subsequently the trend process from Section 2 to the male and female populations from the following countries: Australia, Austria, Canada, Denmark, England & Wales, Finland, France, Italy, Japan, the Netherlands, New Zealand (non-Maori), Norway, Sweden, Switzerland, the United States, and West Germany.[6] For each

[5] More precisely, for each simulation path, a three-dimensional Normal vector with mean equal to the central parameter estimates and covariance matrix $SE^{(i)}$ is generated, and the first and third component of the Normal vector are then transformed to Beta and Gamma distributed values with unchanged mean and variance.

[6] Other countries for which data is available in the HMD have been omitted for different reasons: The populations are so small that the noise is too strong to detect trend changes (e.g. Iceland), reliable data is only available for a few decades (e.g. Portugal), or data is missing for some years (e.g. Belgium).

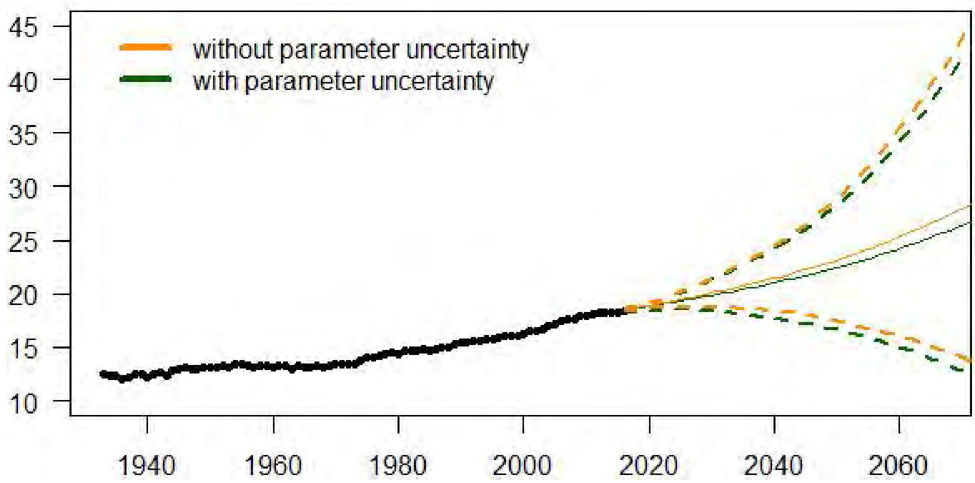


Figure 3.: Mean projections (solid lines) and 90% prediction intervals (dashed lines) of remaining period life expectancies for 65-year old US males with and without parameter uncertainty

population we use the entire HMD data history for ages 60 to 109, as long as it is not explicitly marked as unreliable in the HMD (as e.g. for Sweden before 1860).

Figure 3 shows, for each of the 32 populations, the central estimates and 95% confidence intervals for the trend change parameters. The confidence intervals are derived from Beta, Normal, and Gamma distributions, respectively, which have been calibrated to the population specific central parameter estimates and covariance matrices of standard errors (as explained for US males in Subsection 2.3). We find that uncertainty is substantial for most parameters and populations. This particularly holds for the trend change probabilities. In fact, the central parameter estimates lie within the confidence intervals for most other populations. In comparison, uncertainty in $\mu^{(1)}$ and $\mu^{(2)}$ is smaller which is in line with our findings for US males in Subsection 2.3. Nevertheless, the central parameter estimates are rather similar between the different populations, in particular for $\mu^{(1)}$. The parameter uncertainty in $\sigma^{(1)}$ and $\sigma^{(2)}$ is exceptionally large for a few populations, but again the central parameter estimates lie within the confidence intervals for most other populations.

Given the similarities between parameter estimates for many populations and given the substantial parameter uncertainties, there is reason to believe that parameter calibrations can be improved by aggregating data from different populations. Thus, the question arises which populations should be considered, i.e. which populations can provide insights for the population one is particularly interested in. When modeling multi-population mortality, typically populations with close economic and political links are taken into account, which have thus experienced similar historical mortality evolutions. However, this may not always be a suitable approach when calibrating trend change processes. Populations which are very closely linked may have essentially experienced the same trend changes. Thus, in order to substantially enrich the data base on trend changes, also populations with weaker links and different historical trend process patterns should be considered. In that case, some assumption needs to be made with respect to the similarity (or even equality) of the underlying trend change parameters. We will address this issue in the following section.

4. Trend Process Calibrations to Multi-Population Data

In this section, we will explain different approaches how data from multiple populations can be combined to obtain calibrations for the trend change parameters $\theta^{(i)} = \left(p^{(i)}, \mu^{(i)}, \sigma^{(i)}\right)$. This can be particularly helpful for populations with insufficient data for an individual trend process estimation and also to reduce parameter uncertainty for populations with longer data histories. These data aggregation approaches can be categorized according to whether the trend change parameters for different populations are assumed to be equal or only to come from the same distribution. The assumption of equal parameters for all populations can be motivated by the observations from Figure 3. Parameter uncertainty is huge for most populations and parameters, and it cannot be ruled out that the trend change parameters are equal for all populations. Nevertheless, one may want to relax this assumption, in particular if substantial data is available for the population one is particularly interested in and this data indicates that the population's parameters may be different from those of other populations.

For the remainder of this paper, we assume that data is available for a set of populations P, and the index $\cdot_p$ denotes specific parameter estimates etc. for population $p \in P$. Moreover, let N_p denote the number of data points for population p. Finally, let $p^* \in P$ be the population whose future mortality evolution one wants to project. A numerical example for the proposed data aggregation approaches is provided in the next section.

Parameter estimation from entire data set

Assuming equal parameters for all populations, the most consistent approach certainly is to estimate the parameters from data for all populations simultaneously. This would stabilize the parameter estimation and should reduce parameter uncertainty substantially. However, even though this is a desirable approach from a theoretical perspective, it can be difficult to implement for a large number of populations in practice. For the estimation algorithm proposed in Subsection 2.2, we have found this to be hardly feasible.

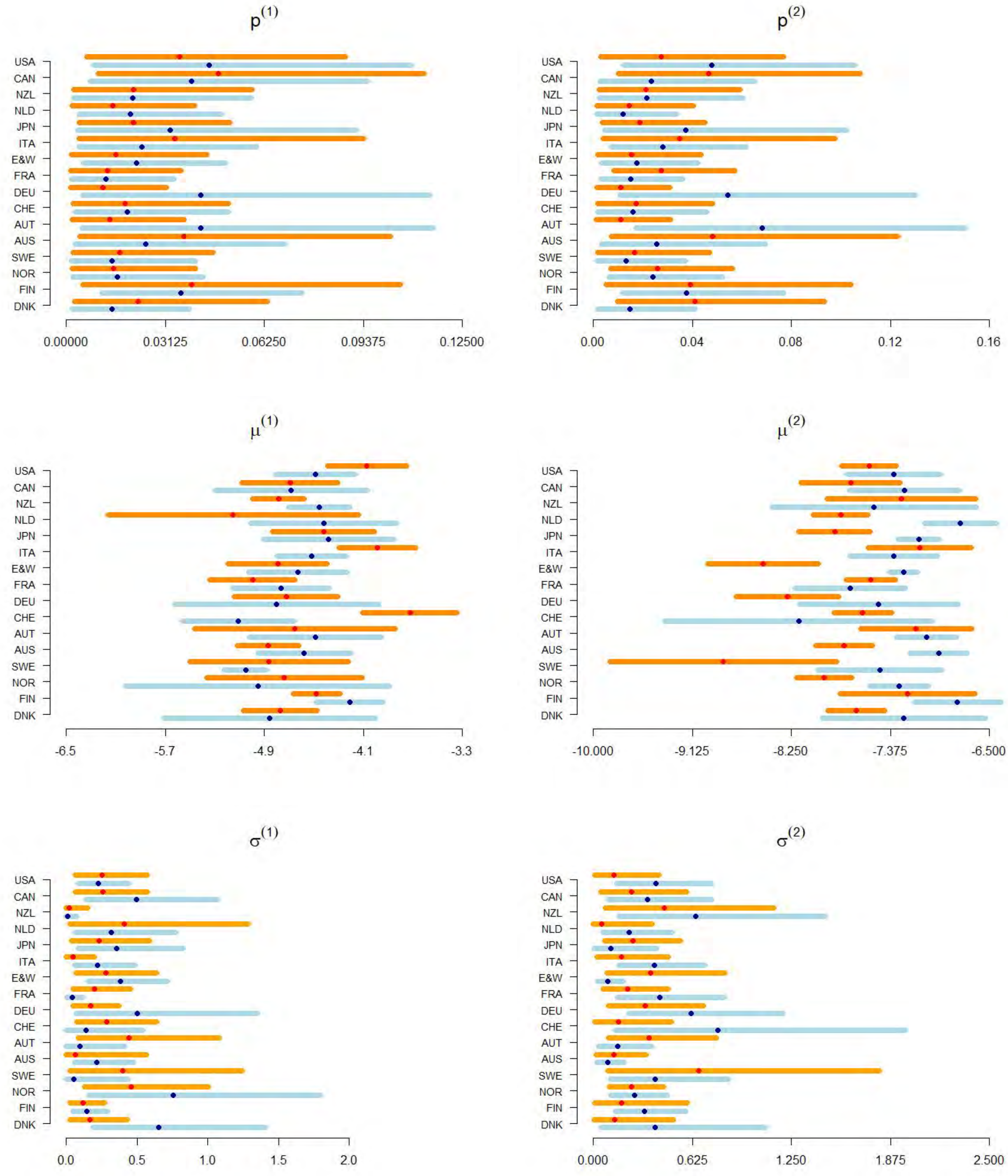


Figure 4.: Central estimates and 95% confidence intervals for the parameters $p^{(1)}$, $\mu^{(1)}$, $\sigma^{(1)}$(left) and $p^{(2)}$, $\mu^{(2)}$, $\sigma^{(2)}$ (right) for males (blue) and females (orange).

Parameter estimation from observed trend changes

Alternatively, the trend change parameters can be estimated from the historical trend changes for all populations. These trend changes, more precisely their occurrences and magnitudes, are determined for each population individually. This obviously implies a slight distributional inconsistency as the historical trend changes are not assumed to be generated by the same set of trend change parameters, but it makes the approach practically feasible. Given the historical trend changes, the trend change parameters are then estimated e.g. via a maximum likelihood approach. Denoting by $|\lambda^{(i)}|_{p,k}$ the vector of absolute trend change magnitudes for the case of k trend changes for population p, the following likelihood function should be maximized:

$$\begin{aligned} L(\theta^{(i)}) &= \prod_{p\in P}\prod_{k=0}^{N_p}\left(L_{lognormal}\left(|\lambda^{(i)}|_{p,k};\mu^{(i)},\sigma^{(i)}\right)\cdot L_{bernoulli}(k;p^{(i)})\right)^{w^{(i)}_{p,k}} \\ &= \prod_{p\in P}\prod_{k=0}^{N_p}\left(\prod_{j=1}^{k} f_{lognormal}\left(\left(|\lambda^{(i)}|_{p,k}\right)_j;\mu^{(i)},\sigma^{(i)}\right)\cdot p^{(i)k}\cdot\left(1-p^{(i)}\right)^{N_p-k}\right)^{w^{(i)}_{p,k}} \end{aligned}$$

where $f_{lognormal}$ denotes the probability density function of the lognormal distribution, and the weights $w^{(i)}_{p,k}$ account for the "relevance" of the different trend change realizations. The maximum likelihood estimation also provides a covariance matrix of (approximate) standard errors as a representation of the remaining parameter uncertainty.

Weighted average of population specific parameter estimates

Instead of equal parameters for all populations, we now assume that the parameter values for each population only come from the same distribution of possible parameter values. In this case, the simplest approach to obtain a set of parameter values for population p^* is to take the average of all population specific parameter estimates $\theta^{(i)}_p = \left(p^{(i)}_p, \mu^{(i)}_p, \sigma^{(i)}_p\right)$. This approach is particularly applicable if hardly any information is available on the true trend change parameters for population p^*. A weighted average can be applied in order to account for the relevancy of each population or the credibility of its parameter estimates. A larger weight would then be assigned to a population if, e.g., it is expected to be very informative for population p^* or it has a comparably long data history. Denoting by v_p the weight for each population, the common parameter estimates would be

$$\theta^{(i)} = \sum_{p\in P} v_p\cdot\theta^{(i)}_p,$$

with weights e.g. according to the data history, $v_p = \frac{N_p}{\sum_{q\in P} N_q}$. The uncertainty associated with these parameter estimates can be determined as the (weighted) sample covariance matrix of the population specific parameter sets $\theta_p^{(i)}$. For a simulation of future mortality, the same approach as in Subsection 2.3 can then be applied.

Parameter sampling from empirical distribution

Alternatively, a (three-dimensional) empirical distribution can be derived from the population specific parameter estimates $\theta_p^{(i)}$, and parameter values can be drawn randomly from this distribution for each simulation path. In analogy to the weighted average above, the different parameter sets in the empirical distribution can be assigned different probabilities. We denote this empirical distribution by $F_{\theta^{(i)}}$ with $P(\theta_p^{(i)}) = v_p$ for $p \in P$ and zero otherwise. The outcomes of this approach should be very similar to those for the weighted average approach. The means of the randomly drawn parameter values are the same by construction, and also the simulated uncertainty in the parameter values should be comparable if the set of populations is large. In that case, the empirical distribution should be reasonably similar to the (theoretical) distribution parameters are drawn from in the weighted average approach.

Credibility approach

In the previous two approaches we assumed that hardly any information is available on the true trend change parameters for population p^*. However, typically at least some information is available, e.g. in form of the $\theta_p^{(i)}$ for all populations $p \in P$. Hence, when simulating mortality for population p^*, the parameter estimates $\theta_{p^*}^{(i)}$ should be assigned a larger probability than parameter estimates which are significantly different. This can be achieved rather easily in a credibility approach where the probability v_{p^*} is increased to emphasize the population specific information. Thus, this approach represents a compromise between the uncertain estimates for the true population specific parameters and a larger reference group of parameter sets which may differ from the true population specific parameters.

Bayesian approach

Alternatively to the credibility approach, a Bayesian approach can be applied. We again assume that the parameter values for all populations come from the same, but unknown distribution. This is the prior distribution in the Bayesian setting, and we approximate it by the empirical distribution $F_{\theta^{(i)}}$. Without any further knowledge about population p^*, parameter values would be drawn from this prior distribution as in the sampling approach above. The realized $\kappa_{t,p^*}^{(i)}$ processes however provide additional information on likely parameter values $\theta_{p^*}^{(i)}$. Unfortunately, we cannot specify the likelihood $L(\kappa_{t,p^*}^{(i)}, t \leq t_0 | \theta)$ for the realized $\kappa_{t,p^*}^{(i)}$ process being generated by some parameter set θ. Therefore, in line with the parameter estimation in Subsection 2.2, we instead consider likelihoods for best possible trend process realizations for different numbers of trend changes k which should be approximately proportional to $L(\kappa^{(i)}_{t,p^*}, t \leq t_0|\theta)$:

$$L(\kappa^{(i)}_{t,p^*}, t \leq t_0|\theta) = \sum_{k=0}^{N_p} L(\kappa^{(i)}_{t,p^*}, t \leq t_0|\theta, k) \propto \sum_{k=0}^{N_p} \hat{L}(\hat{\kappa}^{(i)}_{t,p^*}(\theta), t \leq t_0|\theta, k).$$

Here $L(\kappa^{(i)}_{t,p^*}, t \leq t_0|\theta, k)$ denotes the likelihood under the condition that the data has been generated with k trend changes and $\hat{L}(\hat{\kappa}^{(i)}_{t,p^*}(\theta), t \leq t_0|\theta, k)$ is the likelihood as defined in Schupp (2019). In order to avoid the computationally expensive iterative algorithm to obtain the latter likelihood for any parameter set θ, we approximate the best possible trend process realizations $\hat{\kappa}^{(i)}_{t,p^*}(\theta), t \leq t_0$ by those from the individual parameter estimation, $\hat{\kappa}^{(i)}_{t,p^*}(\theta^{(i)}_{p^*}), t \leq t_0$. Even though the latter have been determined under different parameter estimates, we can assume them to be reasonably similar to those for the parameter set θ since we have observed that the optimal positions of the k trend changes are the same for almost all reasonable parameter sets in general. Thus, we have

$$L(\kappa^{(i)}_{t,p^*}, t \leq t_0|\theta) \propto \sum_{k=0}^{N_p} \hat{L}(\hat{\kappa}^{(i)}_{t,p^*}(\theta), t \leq t_0|\theta, k) = \sum_{k=0}^{N_p} \hat{L}(\hat{\kappa}^{(i)}_{t,p^*}(\theta^{(i)}_{p^*}), t \leq t_0|\theta, k) =: \hat{L}^{(i)}_{p^*}(\theta).$$

Finally, we can derive the posterior distribution $F_{\theta^{(i)}|\kappa^{(i)}_{t,p^*}}$, i.e. the probability for parameter set $\theta^{(i)}_p$ is

$$P(\theta^{(i)}_p|\kappa^{(i)}_{t,p^*}, t \leq t_0) = \frac{1}{c} \cdot v_p \cdot \hat{L}^{(i)}_{p^*}(\theta^{(i)}_p),$$

where c is a constant such that the probabilities sum up to one. This posterior distribution describes the remaining population specific parameter uncertainty, and in a simulation, path dependent parameter values should be drawn from this distribution.

5. Numerical Example

In this section we apply the different data aggregation approaches from Section 4 to the case of US males. Our set of reference populations P consists of the 32 populations which we already considered in Section 3 and the weights/probabilities v_p are derived according to the number of available data points. In the credibility approach we set $v_{p^*} = 0.5$ and reduce all other probabilities proportionally.

Figure 4 shows the central estimates of the trend change parameters and their 95% confidence intervals for the individual calibration for US males and the different data aggregation approaches. We see that the central parameter estimates can vary between the approaches and that parameter uncertainty is not necessarily reduced compared to the individual calibration. We will now explore why this is the case.

Starting with the maximum likelihood estimation based on all historical trend changes, we find that the central estimates of the trend change probabilities $p^{(i)}$ are considerably smaller when data is aggregated. A comparably large number of trend changes has been observed for US males and this is now compensated for. Parameter uncertainty reduces substantially for all three trend change parameters as expected. This means that the same trend change probability and Lognormal distribution may be assumed in order to generate trend changes for different populations. The comparably large central estimates for the $\sigma^{(i)}$ compensate for the fact that the mean of the trend change magnitudes is now derived from all trend changes and not only those of a particular population.

The results for the weighted averages and the sampling from the empirical parameter distribution are very similar as expected. The trend change probabilities $p^{(i)}$ are reduced again compared to the individual calibration, and also parameter uncertainty is smaller. In contrast, parameter uncertainty in the magnitude parameters $\mu^{(i)}$) and $\sigma^{(i)}$has increased. However, this is primarily uncertainty which arises from the assumption of a distribution for population specific parameter values instead of the assumption of one fixed parameter set for all populations. Thus, it is systematic uncertainty as opposed to rather unsystematic uncertainty arising from parameter estimation from limited data. Only a small portion of the depicted parameter uncertainty can be credited to the randomness in the 32 population specific parameter estimates which are used to approximate the true but unknown distribution of parameter values.

Also the results from the credibility and the Bayesian approaches are rather similar. Both approaches build on the assumption of an unknown distribution for population specific parameters and determine some trade-off between the population specific parameter estimates and those from other populations. In the credibility approach, 50% probability is assigned to the parameter estimates for US males, while it is about 40% for $\theta_{p^*}^{(1)}$ and 11% for $\theta_{p^*}^{(2)}$ in the Bayesian approach. The remaining probability is assigned to other parameter sets, though based on different principles. Again we find reduced trend change probabilities, but less reduced than for the other aggregation approaches due to the substantial weight for the US male parameter set. The same applies to the magnitude parameters which also lie between the individual estimates and the parameter values from the other aggregation approaches. We also observe that, for all parameters, uncertainty remains substantial or even increases compared to the individual calibration. Again, this is primarily systematic uncertainty arising from the assumption of a distribution for population specific parameter values. Given the limited data on trend changes for US males, substantial uncertainty still remains with respect to the credibility of the population specific parameter estimates.

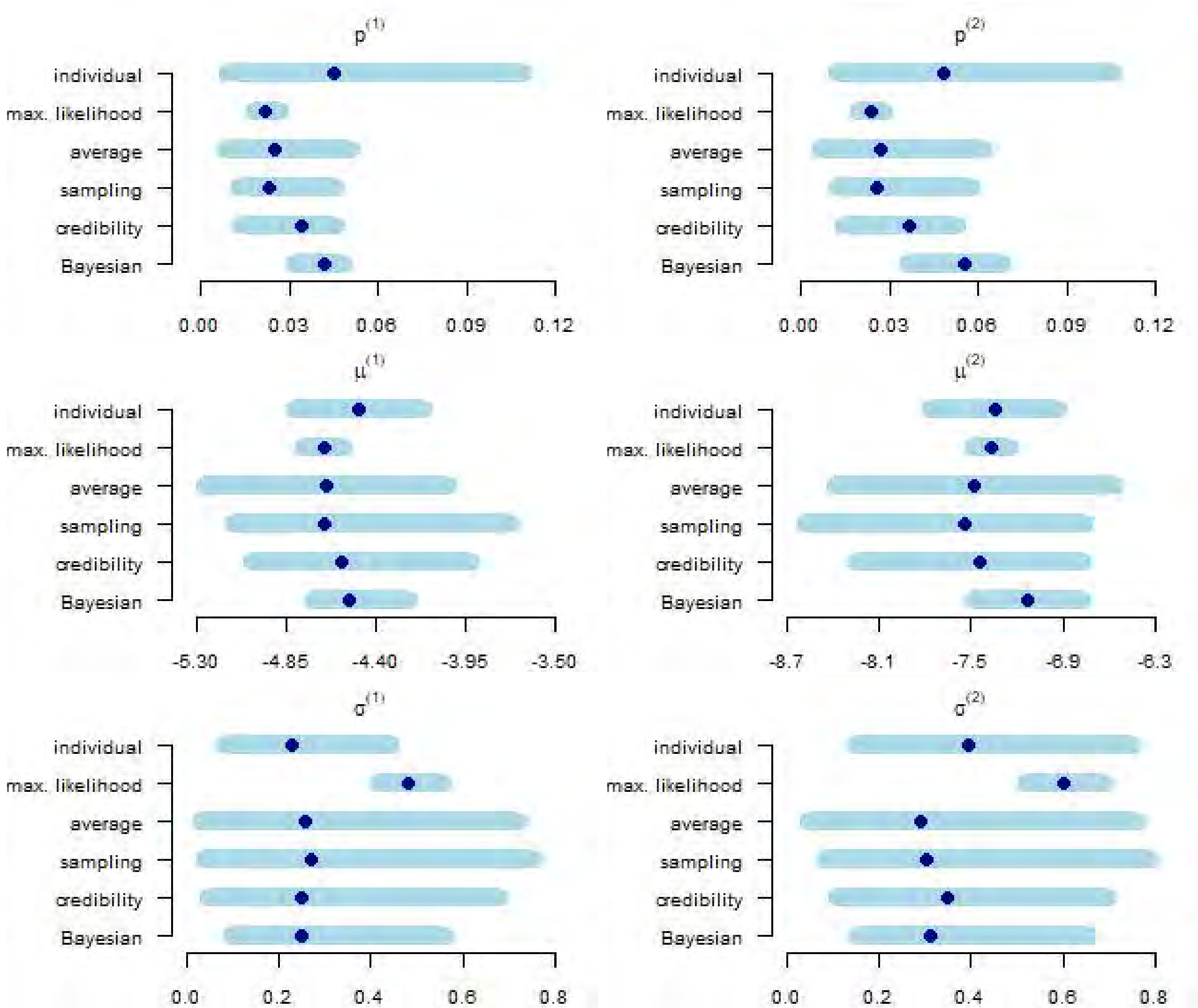


Figure 5.: Central estimates and 95% confidence intervals for the parameters $p^{(1)}$, $\mu^{(1)}$, $\sigma^{(1)}$ (left) and $p^{(2)}$, $\mu^{(2)}$, $\sigma^{(2)}$ (right) for US males based on multi-population data.

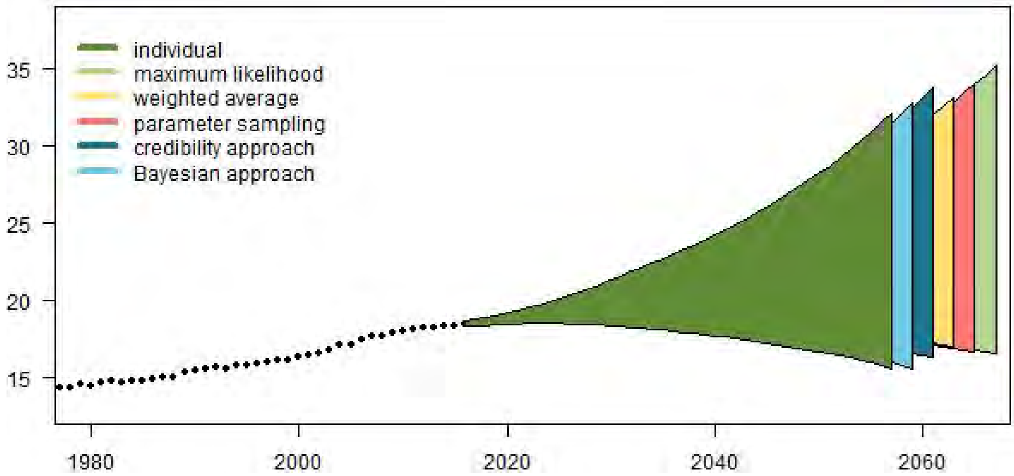


Figure 6.: 90% prediction intervals for remaining period life expectancies of 65-year old US males based on different trend process calibrations

Finally, we compare the different data aggregation approaches by projecting remaining period life expectancies for 65-year old US males. In each case, parameter uncertainty is accounted for by drawing from Beta, Normal, and Gamma distributions with case specific parametrizations. The starting values are modeled as explained in Subsection 2.3 in any case. Figure 5 shows the 90% prediction intervals for all approaches considered. The widest prediction intervals can be observed for the population specific calibration. This is in line with the comparably large central parameter estimates for $p^{(i)}$ and $\mu^{(i)}$ which we observe in Figure 4. The maximum likelihood approach yields the most narrow prediction intervals, mainly due to comparably small central parameter estimates for $p^{(i)}$ and $\mu^{(i)}$ and the small parameter uncertainty for all parameters. Prediction intervals for the weighted parameter averaging and the parameter sampling are very similar which is again in line with observations from Figure 4. The widths of the prediction intervals are between those for the aforementioned approaches. Thus, uncertainty is particularly reduced compared to the individual calibration where the trend change probabilities may have been overestimated simply by chance. The credibility and the Bayesian approach yield prediction intervals between the individual calibration and the sampling approach as they define a kind of 'mixture' of parameter estimates from these approaches.

6. Conclusion

In this paper, we have discussed the issue of parameter uncertainty in mortality processes with trend changes. Due to the limited number of observed historical trend changes, parameter uncertainty is substantial in general. This particularly holds for those parameters which determine future trend changes as part of stochastic projections. We have identified the main sources of this uncertainty and have explained how it can be quantified for the trend process of Börger and Schupp (2018). A comparison of 32 populations shows that central parameter estimates can vary considerably when trend processes are calibrated for each population individually. However, due to the substantial uncertainty associated with these estimates, it is not clear whether this is mainly due to random effects in the few trend changes which have been observed for each population.

In order to improve the reliability of trend process calibrations, we have then discussed different approaches for aggregating data on trend changes from several populations. This includes approaches which assume the true parameter values to be equal for all populations under consideration as well as approaches where the parameter values for different populations are only assumed to come from the same, but unknown distribution. A maximum likelihood parameter estimation based on the historical trend changes for all populations shows that the assumption of equal trend change probabilities and magnitudes for all populations may be reasonable. Parameter uncertainty can be reduced substantially here.

When allowing for different parameter values from the same underlying distribution for each population, parameter uncertainty reduces only slightly or even increases compared to the population specific calibrations. Depending on how different the population specific parameter estimates are, the uncertainty in the empirical distribution build from these estimates can be larger than the uncertainty in estimating the population specific parameters in the first place. Nevertheless, central parameter estimates can change substantially when aggregating parameter estimates in a common distribution. As we have seen in a numerical example, this may prevent overestimation (underestimation) of the uncertainty in the future mortality evolution in case the population specific parameter estimates may have been rather large (small) simply by chance.

In conclusion, we have found that parameter uncertainty can be much better understood when data from different populations is aggregated. Furthermore, the reliability of trend process calibrations can be improved by reducing random effects in population specific parameter estimates. Whether this leads to more narrow or wider prediction intervals for the quantities of interest like future life expectancies, depends on several factors: most importantly, the population specific parameter estimates, the set of reference populations, and the assumption on how the parameter values for different populations relate to each other.

References

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

Börger, M., Russ, J., and Schupp, J. (2019). It Takes Two: Why Mortality Trend Modeling is more than modeling one Mortality Trend. *Working Paper, Ulm University.*

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

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.

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

Human Mortality Database (2019). University of California, Berkeley (USA) and Max Planck Institute for Demographic Research (Germany). Available at www.mortality.org (data downloaded on February 01, 2019).

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

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

Schupp, J. (2019). On the Modeling of Variable Mortality Trend Processes. *Working Paper, Ulm University.*

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

Curriculum Vitae

Die Inhalte dieser Seite wurden aus Datenschutzgründen entfernt.

Die Inhalte dieser Seite wurden aus Datenschutzgründen entfernt.