Coherent Projections of Age, Period, and Cohort Dependent Mortality Improvements*

Matthias Börger und Marie-Christine Aleksic

Preprint Series: 2011-05

Fakultät für Mathematik und Wirtschaftswissenschaften
UNIVERSITÄT ULM

Coherent Projections of Age, Period, and Cohort Dependent Mortality Improvements*

Matthias Börger†

Institute of Insurance, Ulm University & Institute for Finance and Actuarial Sciences (ifa), Ulm

Helmholtzstraße 22, 89081 Ulm, Germany

Phone: +49 731 50 31257. Fax: +49 731 50 31239

Email: m.boerger@ifa-ulm.de

Marie-Christine Aleksic

Institute of Insurance, Ulm University

July 2011

Abstract

The projection of future mortality experience constitutes a challenge for both actuaries and demographers. Some of the currently used projections have several shortcomings which may pose a serious threat to insurers and social security systems.

In this paper, we propose a new projection methodology which overcomes these shortcomings. Our model allows mortality improvements to depend on age, period, and cohort and provides highly plausible forecasts. Moreover, it is very flexible with respect to the level of future mortality improvements. This allows us to derive coherent projections for several populations simultaneously, e.g. males and females of the same country or populations from closely related countries. We observe that the incorporation of information about the mortality experience of other populations can have a significant impact on the projection for a given population. In order to illustrate our methodology, we derive fully specified projections for German males and females as members of a large reference set of European populations.

*The authors are very grateful to Jochen Ruß and Hans-Joachim Zwiesler for their valuable comments and support.

†Corresponding Author

1 Introduction

Longevity risk, i.e. the risk of insured/pensioners living longer than expected, is one of the most serious insurance risks. It is most relevant for pension funds, annuity providers, and social security systems. In the past, gains in life expectancy or equivalently, improvements in mortality have been underestimated consistently in most industrialized countries. The revision of mortality projections, i.e. the realization of longevity risk, has then led to the requirement of additional funds to support increasing liabilities. This poses a serious threat to any (financial) institution concerned with the provision of survival benefits.

Longevity risk has always been present but its significance has gained considerably in recent decades. Riskless yields in the financial markets have fallen considerably in many countries leaving less funds for the provision of additional reserves. At the same time, the size of longevity risk in the private sector has increased. Benefits from social security systems have been reduced in many countries which in turn has increased the demand for private annuities and occupational pensions. This demand is often supported by tax incentives, either for products with mandatory annuitization or to make annuitization more attractive to the policyholder than taking the lump sum payment.

In order to minimize longevity risk, actuarial and social security institutions have steadily looked to improve their mortality projections. However, graphical analyses reveal that some currently used projections still seem questionable. As an example, in Figure 1 we plot the annual mortality improvements embedded in the standard mortality table for reserving for private annuity business in Germany, i.e. the table DAV 2004 R. In the left panel, we see historical mortality improvements for West German males up to 2008 and projected best estimate improvements thereafter.[1] The plot reveals several issues which can be identified for many existing projections and it indicates very nicely what the focus in the derivation of new projections should be on:

- We observe a structural break between historical and forecast mortality improvements. In reality, the transition will almost certainly be smooth.
- The projection assumes a rapid slowdown in mortality improvements over the next years which cannot be motivated from the historical data.
- The historical data contains significant diagonal structures, i.e. cohort effects, which are not extrapolated into the future.

The right panel of Figure 1 shows the projection of the DAV 2004 R table including margins. The structure of the projected mortality improvements still looks critical and even with margins, mortality improvements seem to be underestimated for some ages at least for the next years. One may argue that the projection should be sufficient for a portfolio of contracts with a widely spread age distribution.

[1] The historical data is obtained from the Human Mortality Database (2011) and we apply P-splines to smooth the mortality rates before computing the improvements. To support interpretability and comparability of different heat charts, here and throughout this paper, we sometimes cap rather extreme values. We use data for West Germany only as the projection was derived from the same data set. The trend parameters in the projection are set to $T_1 = 10$ and $T_2 = 15$. For more details on this projection and its parameters, we refer to DAV (2004).

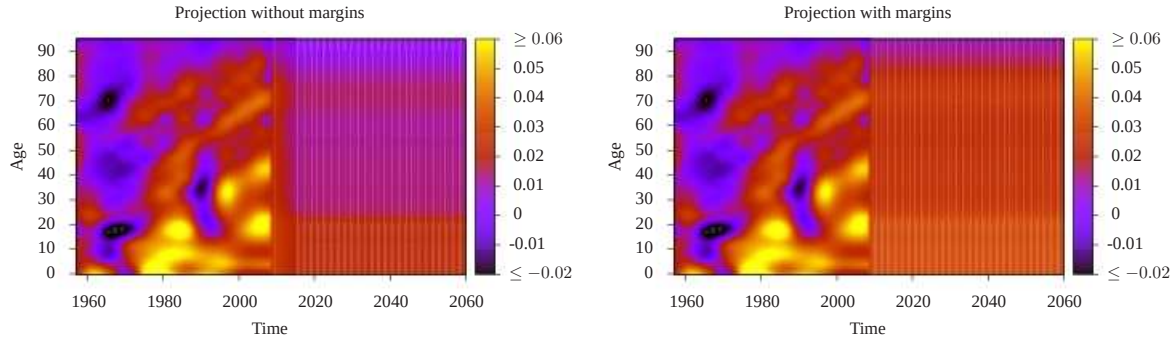

Figure 1: Standard mortality projection for German annuity business

However, regulatory requirements often permit the balancing of profits and losses from different products or product tranches. Thus, a spread of risk over a wide range of ages is often not possible which underlines the need for an adequate projection for each age and cohort individually. We have made similar observations for the corresponding projection for German females.

Our observations clearly show that there is still need and space for improved projection methodologies. The derivation of such a methodology is the goal of this paper. We base our model specification on historical data patterns thus allowing for age, period, and cohort dependent effects. All parameters have a clear interpretation and provide a decent fit to various data sets. Moreover, our model allows for a large flexibility in the level of forecast mortality improvements and we present several ideas how this level can be fixed, e.g. based on extrapolations of historical life expectancies. The flexibility also allows us to derive coherent projections. In the past, projections for males and females and/or different populations have typically been derived independently from each other which often lead to implausible results. We use demographic insights here to improve plausibility in simultaneous projections but also to improve reliability of projections for each single population.

We focus on the projection of mortality improvements here instead of the more common approach of forecasting mortality rates for two reasons: First, we do not have to model the (current) level of mortality rates but only their changes over time. This reduces the number of required parameters and improves the interpretation of the remaining parameters. Secondly, the resulting projection can be applied to a base table, i.e. most recent realized mortality rates, for basically any population for which it seems adequate. This is particularly convenient as the projection could be derived from an extensive data set, e.g. for the general population, and then be used for small subpopulations without sufficient data as well, e.g. the population of a pension fund. As such a group of pensioners is a subpopulation of the general population, the long-term mortality changes in both populations should be very similar (see, e.g., Jarner and Kryger (2011) and Cairns et al. (2010)). Significant differences in the mortality experience between the populations may still arise from the base table.

The increasing demand for assessing and managing longevity risk has provoked considerable academic research in this field – both with respect to deterministic mortality projections and stochastic mortality modeling. Nevertheless, we are convinced that our approach adds to this literature and that some of our

ideas can also be applied to improve other existing projection models. We refrain from giving a literature overview here and instead refer to a later section where we compare our projection methodology with other modeling approaches.

The remainder of this text is structured as follows: In the following section, we analyze historical mortality improvement patterns and deduce the specification of our projection model. We describe the model fitting and discuss related issues for the example of West German males. Finally, we comment on the applicability of model simplifications and on the stability of our model. The derivation of mortality projections is then discussed in Section 3. We start with the case of a single population and then show how coherent forecasts for several populations can be obtained. Here, several crucial assumptions are necessary and we provide reasoning for all assumption we make. In the subsequent section, we analyze uncertainties inherent in our projection model and describe ways to assess and account for these uncertainties. In particular, we show how basis risk can be measured in case the projection is to be applied to a population different from the one it has originally been derived for. In Section 5, we compare our model to other projection models and outline how it adds to the existing literature. Finally, Section 6 concludes.

2 Historical Mortality Improvements

As indicated in the Introduction, our projection methodology can be applied to basically any population with a sufficient data history. For illustrative purposes, here we focus on the male population of West Germany. We exclude data from East Germany as there seems to be a consensus that the reunification in 1990 has led to the East German mortality experience moving towards that of West Germany. Thus, a combined data set may be blurred by this one-off effect.

Mortality data for West Germany is available from the Human Mortality Database (2011) for years 1956 to 2008 and ages 0 to 109 (as of March 2011). However, the data is extrapolated and graduated above age 95 and therefore we limit the data set to ages up to 95. For simplicity, in the following we will refer to the West German population as the German population only.

2.1 Model Specification

Figure 2 shows raw mortality improvements

$$v(x,t) = \frac{q(x,t-1) - q(x,t)}{q(x,t-1)} = 1 - \frac{q(x,t)}{q(x,t-1)} \tag{1}$$

for German males.[2] We clearly observe vertical and diagonal structures which means that mortality improvements depend on calender year and year of birth or cohort. In the literature, mortality improvements have been shown to also depend on age which is in fact the dependency most commonly modeled

[2]Obviously, this definition of mortality improvements is only valid as long as $q(x,t-1) > 0$. For a population as large as the German one, this is always the case but for smaller populations there may well be raw mortality rates of zero. A few undefined mortality improvements are uncritical for the fitting of our model though. As we will see later on, all parameters in our model are fitted to a considerable number of data points thus omitting a few data points hardly affects the calibration.

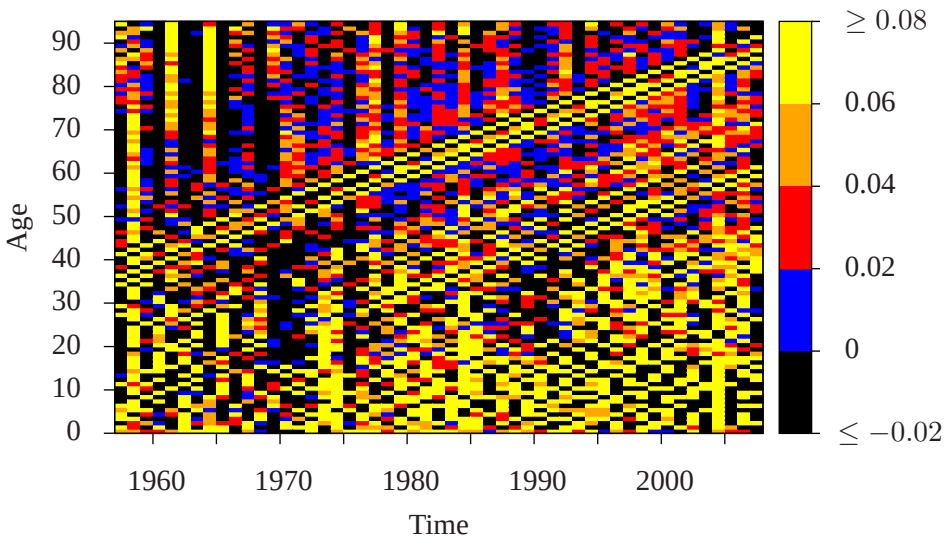

Figure 2: Raw mortality improvements for German males

(see, e.g., Figure 1 or the model of Lee and Carter (1992)). Therefore, it seems reasonable to specify mortality improvements as a combination of age, period, and cohort effects. The most simple combination is the following linear one:

$$v(x,t) = a_x + p_t + c_{t-x} + \epsilon(x,t),$$

where a_x is the age dependent component, p_t the calender year component, c_{t-x} the cohort component, and $\epsilon(x,t)$ is an error term with mean zero.[3] All parameters possess a very clear interpretation which helps understanding historical improvement patterns.

When calibrating our model two issues arise immediately. The first one is an identification problem. For instance, increasing all age parameters a_x by a fixed amount and decreasing all period parameters p_t by the same amount would yield equal fitted mortality improvements. To solve this issue, we impose the following constraints:

- $\sum_t p_t = 0$
- $\sum_{t-x} c_{t-x} = 0,$

i.e. the average period and cohort parameters are equal to zero. Thus, all "substance" in the mortality improvements is contained in the age parameters. In fact, these two constraints still do not guarantee uniqueness. This can be seen in Cairns et al. (2009) who apply a third constraint to the Age-Period-Cohort model which, for log mortality rates, has exactly the same structure as our model. However, we

[3] Obviously, other model specifications would be possible, e.g. $v(x,t) = \exp\{a_x + p_t + c_{t-x} + \epsilon(x,t)\}$, but results vary only insignificantly. Therefore, we stick to the most simple specification.

do not need their third constraint as another issue in our setup makes it superfluous. This issue is the calibration of the cohort parameters at the boundaries of the data set. The parameters for the very first and very last cohorts are fitted to only a few data points and are thus likely to take up random noise instead of modeling proper cohort effects. Therefore, in line with the second constraint, we set the first and last cohort parameters to the long-term average of zero.[4] The number of affected cohort parameters at both boundaries is rather subjective and clearly depends on the data set under consideration.

2.2 Model Calibration

In contrast to the model structure, the fitting of our model is not straightforward. The most common approach of Maximum Likelihood estimation is not possible as the distribution of the mortality improvements is fairly complex. For a sufficiently large population, deaths and mortality rates can be assumed to be approximately normally distributed. Thus, according to Equation (1), mortality improvements follow the distribution of the ratio of two normal random variables. Such a distribution can be specified (see Hinkley (1969)) but its parameters cannot be expressed by our model parameters $a_x, p_t,$ and c_{t-x}. In order to check for approximate normality in the improvements or the residuals, respectively, we also performed some statistical tests. However, for German males and also females the assumption of normality was clearly rejected by each test.[5]

Therefore, we fit our model in iteratively reweighted least squares, i.e. we minimize the expression

$$\sum_x \sum_t \frac{(v(x,t) - a_x - p_t - c_{t-x})^2}{w(x,t)^2} \longrightarrow \min,$$

where $w(x,t)$ are the weights. The need for weighting becomes obvious from Figure 2. The variability in mortality improvements differs significantly between different age groups and periods. For young ages or old ages in the earlier years of the data set, the raw mortality improvements fluctuate much stronger because the numbers of observed deaths are much lower there than elsewhere. Unfortunately, the choice of weights is not obvious. The distribution of the mortality improvements does not possess any moments in general due to its extremely fat tails. In the absence of preferable alternatives, we nevertheless apply empirical standard deviations of the residuals as a measure of variability. However, this approach requires an iterative procedure as we need to fit the model once before we can compute weights. Therefore, we start with a first run of unweighted fitting and compute the empirical standard deviation for each data point from the (up to) 81 residuals in a square around this data point. This choice for the number and location of residuals is rather subjective but for the data set of German males it provides a fairly smooth surface of weights. We then repeat the fitting with iteratively updated weights until convergence in the model parameters is reached. For German males we need six runs to get changes in all parameter values below 0.1%.

[4] Actually, as soon as we set two cohort parameters equal to zero we could also drop the second constraint. However, the overall fit of the model would hardly change and it is convenient for the projection to not have any substance in the cohort parameters. We therefore keep this constraint.

[5] Note that this also impacts the derivation of confidence intervals for the model parameters. Due to the extremely fat tails of the ratio distribution, confidence intervals based on normal approximations may be much too narrow.

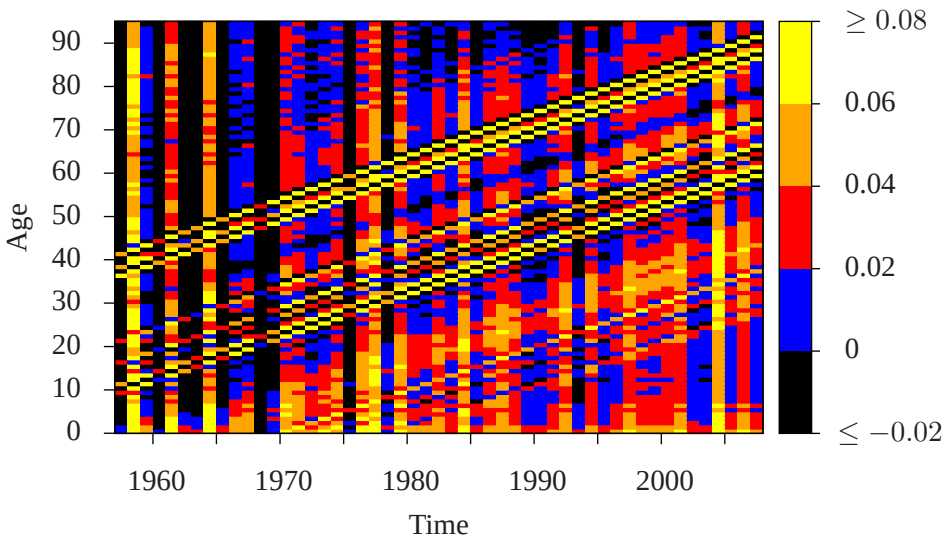

Figure 3: Fitted mortality improvements for German males

Figure 3 shows the resulting fitted mortality improvements for German males with parameters for the first 10 and last 27 cohorts set equal to zero.[6] The structure of the plot looks very similar to that in Figure 2 and our model just seems take some noise out of the raw data. This indicates that our model fits the data well which we will analyze in more detail in the following section. Figure 4 contains plots for the age, period, and cohort parameters – both as fitted to the raw data and graduated using P-splines. The graduated parameter values will be particularly relevant for the projection later on. From the plots, we observe that the period and cohort parameters lie around zero and that thus all substance is contained in the age parameters. We also see that variability in the age parameters is significantly smaller than for the other two parameter sets. This is why we were not able to detect a dependency of the mortality improvements on age in Figure 2 in the first place.

The panel with the cohort parameters contains two sets of raw parameter values. Those plotted in green stem from a preliminary model fit which included all cohorts in the data set. Here, we observe large variability in the boundaries which is most probably due to noise. Moreover, it typically takes some time until a cohort effect develops. Thus, it is questionable whether one should assume the existence of cohort effects for cohorts who are still very young today. Therefore, we decide to set the first 10 and the last 27 cohort parameters equal to zero in our final model fit. The resulting parameters are plotted in red.

[6] A reasoning for this constraint on the cohort parameters is given below.

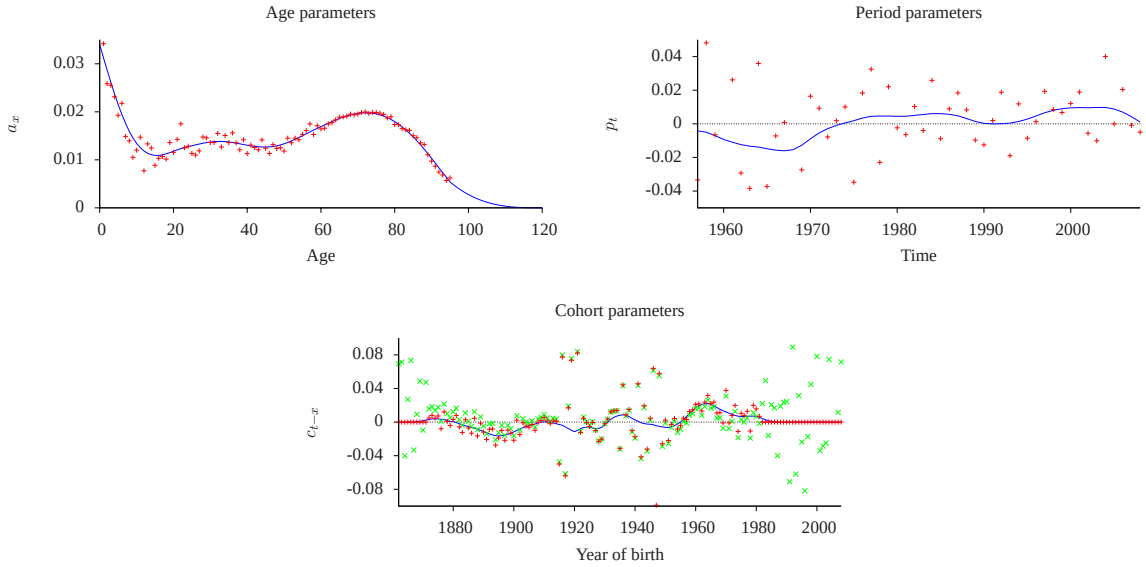

Figure 4: Fitted and smoothed parameter values for German males

2.3 Residual Analysis, Optimal Fitting Period, Model Stability, and Model Simplifications

In Figure 5, the standardized residuals for the model fit for German males are plotted.[7] There is only very little structure contained in the plot which suggests our model is capable of explaining historical mortality improvements. Significant structure only remains for calender years up to 1970 where period dependent effects seem to differ between younger and older ages sometimes. For instance, in 1968 we observe very small residuals for ages around 80 but significantly positive residuals for most younger ages. In the following year, this relation is inversed. Our model is obviously not able to allow for such effects. However, these effects seem to be only temporary and rather irrelevant for a projection. Furthermore, omitting data before 1970 would reduce the number of data points age and cohort parameters are fitted to. Therefore, it seems desirable to keep the fitting period as large as possible and to ignore the remaining period structures in the residuals. For significantly longer data sets a limitation may be appropriate though.

In order to check the stability and robustness of our model, we nevertheless additionally fitted it to data starting from 1970. We do not show a plot of the fit here but it is very similar to the one for the full data set. The values for some of the individual model parameters change but the fitted mortality improvements for both model fits only differ by 0.26% on average.[8]

As another robustness check, we omitted data for young ages from our model fitting. This is a typical setting for annuities or pensions where data on child mortality or even young adult mortality is very sparse. For the rather extreme case of omitting all ages below 60, the fitted mortality improvements

[7] We standardized the residuals using the empirical standard deviations from the fitting.

[8] Note that, regarding the differences between the model fits, we only considered data points for which cohort parameters are calibrated in both cases. Otherwise, the difference can easily be significant which again suggests using the full data set.

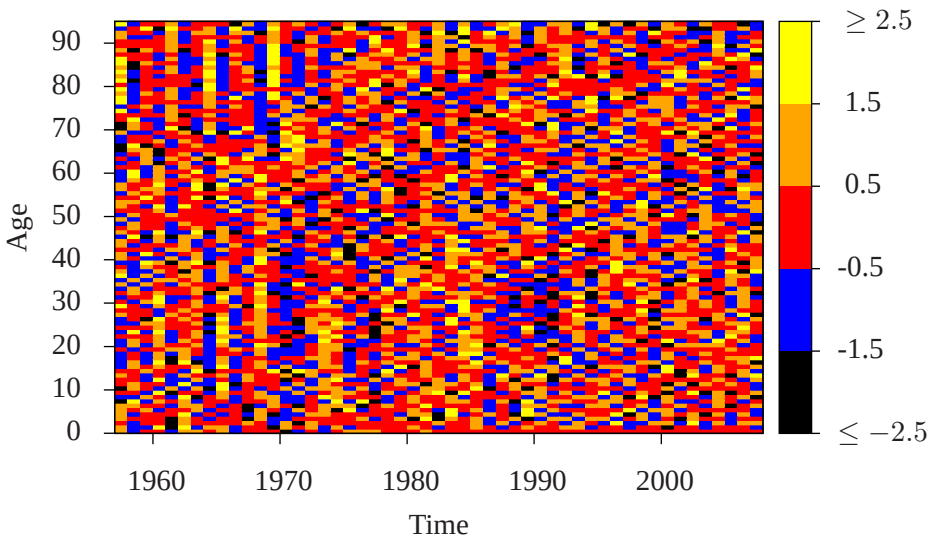

Figure 5: Standardized residuals for German males

change by 0.43% on average.[9] Most significant changes occur for years up to 1969. This is due to the sometimes opposing period effects for young and old ages in those years and the period parameters now fully taking up the effects for old ages instead of some "average". For years starting from 1970, the average change in the fitted mortality improvements is only 0.26%, as for the full age range.

Finally, we analyze whether our model can be simplified for the German data set. To this end, we fit reduced models consisting of only two of the three parameters sets. In Figure 6, we exemplarily show the fitted improvements and the standardized residuals when omitting the cohort parameters. There clearly is diagonal structure missing in the fitted values compared to Figure 2 and certainly too much diagonal structure in the residuals. Similar observations can be made when excluding the age or period parameters. The fit is significantly worse in each case and therefore, we conclude that, at least for German males, the model should not be simplified.[10] Obviously, results may be different for other populations.

3 Projection of Mortality Improvements

In the previous section, we have seen how our model can be calibrated to historical data. Now we derive projections based on the calibrated model. We start with a projection for a single population and explain different approaches for forecasting each of the three parameter sets. Then we turn to coherent projections for several populations and show which modifications should or must be applied to the

[9] Again, we disregard data points for which cohort parameters are zero for restricted model fit.

[10] A statistical test on the significance of the parameters in the full model is not straightforward. The commonly used likelihood ratio test requires normal residuals which we do not have.

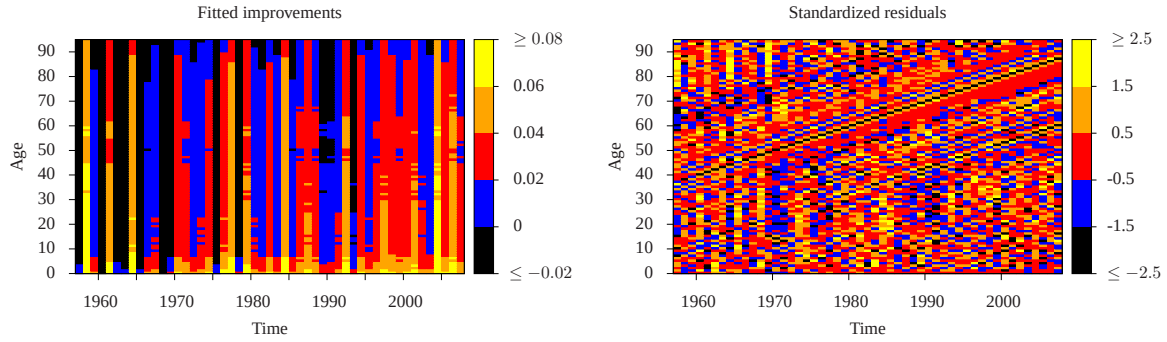

Figure 6: Results for a model fit without cohort parameters

approaches for a single population. Here, we do not try to achieve totally coherent projections for each individual mortality improvement as this is hardly possible. The goal rather is to have plausible forecasts at a more aggregated level, e.g. coherent life expectancy extrapolations.

3.1 Projection of Mortality Improvements for a Single Population

We commence with the forecast of future age parameters. The most obvious approach is to simply maintain the parameter values as calibrated to the historical data. In order to reduce the remaining noise they should be smoothed. Unfortunately, we can fit age parameters only for ages with available data history, i.e. for ages below 95 in our case. The limiting age of a new projection should be considerably larger though, e.g. 120. Thus, we need a different approach to derive parameter values for those very old ages.

An extrapolation of the existing age parameters is not obvious. Therefore, we try to extract as much information as possible from the original data set. We extrapolate mortality rates up to age 120 for each year using different curves, i.e. the laws of Gompertz and Kannisto as well as a logistic curve.[11] From the resulting artificial mortality improvements, we can then derive age parameters. For all mortality curves and different populations, we observed that the parameter values tend to zero or become even slightly negative for ages towards 120. Negative age parameters and thus ongoing mortality deterioration at very old ages does not seem plausible. But the results indicate that – based on the historical data – the existing age parameters should be extrapolated such that a value of zero is obtained at the limiting age. Our findings are in line with those of Gampe (2010) who analyzes available data for supercentenarians worldwide. In Figure 4, we applied a cubic function for the extrapolation with function value and first two derivatives at age 120 being zero and a function value at age 95 according to the graduated age parameter from the actual data.

At this point it should be mentioned that – given the steady increase in life expectancy in the past – it seems possible that mortality improvements at very old ages may increase considerably in the future. Some authors have already tried to predict such an increase using frailty models (see Jarner and Kryger

[11] For details on these mortality curves, we refer to Thatcher et al. (1998).

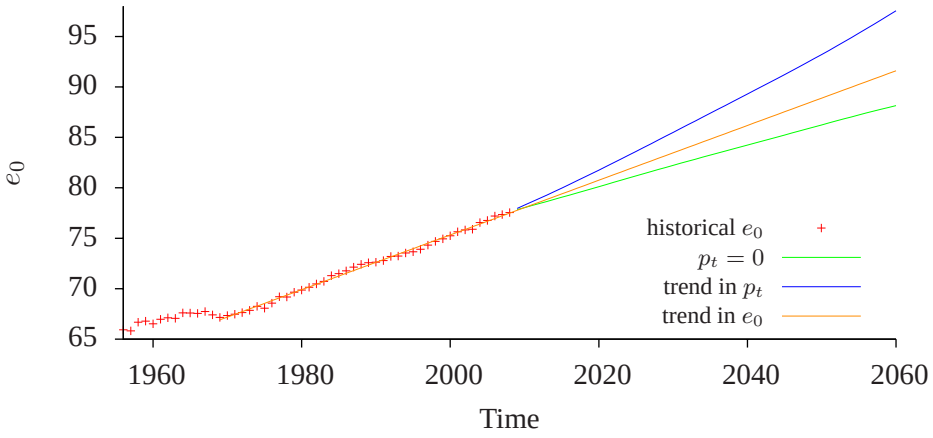

Figure 7: Life expectancy extrapolations for German males

(2011) and references therein). Therefore, it might be worthwhile adjusting some age parameters, also based on epidemiological insights, instead of solely relying on historical data extrapolations. Similar arguments may hold for other age groups of a specific population, e.g., in case historical mortality improvements have been significantly driven by one-off effects. We do not observe such effects in our data set but they could be corrected for easily in our model.

For the fitted cohort parameters, we proceed as for the age parameters by simply graduating them (see Figure 4). For future cohorts, we do not have any information available. Therefore, it is most plausible to simply set parameters for such cohorts equal to the long-term average of zero (as for the cohorts at the boundaries of the data set).

The future period parameters are most difficult to forecast. They determine the overall level of mortality improvements in the future and are thus the most crucial set of parameters. For their projection, several approaches with potentially significantly different outcomes are possible. The most simple approach is to set future period parameters to their long-term historical average of zero. However, in Figure 4, we can observe an increasing trend in the historical period parameters for German males. At least for the next decades, it is thus also plausible to forecast period parameters according to this trend. A third possibility is to calibrate future period parameters to a reasonable extrapolation of an aggregated mortality statistic like the period life expectancy at birth. All three approaches appear equally plausible and it is impossible to state which one may provide the most reasonable projection in general. In Figure 7 we see period life expectancies at birth for German males projected according to the three approaches.[12] Starting from 1969, historical life expectancies show a rather linear pattern. Therefore, a linear extrapolation seems to be a reasonable basis for deriving future period parameters.[13] The resulting parameter values must be positive because the linear trend is steeper than the slope in the life expectancies based on period

[12] Life expectancies are computed based on mortality rates up to age 120 with mortality rates for ages above 95 obtained from Kannisto extrapolations.

[13] There is an extensive literature on the question whether life expectancy can increase infinitely or whether there is some biological limit. From an actuarial perspective, we think it is dangerous to assume a limit. History tells us that previously assumed limits have been surpassed rather quickly (cf. Oeppen and Vaupel (2002)).

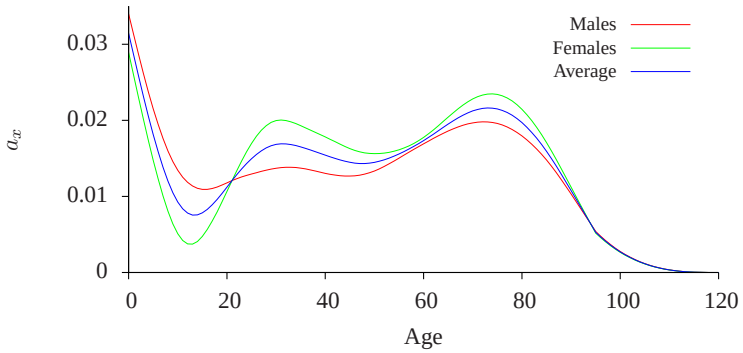

Figure 8: Age parameters for German males, females, and their averages

parameters of zero. However, linearly extrapolated life expectancies are still significantly smaller than life expectancies which are derived from a linear trend in the historical period parameters. In 2060, the difference between the three projections is already more than 9 years which clearly shows the huge uncertainty associated with mortality projections.

This uncertainty can usually be reduced by projecting mortality for different populations simultaneously. All three projections for German males look plausible on their own but they may not when related to possible projections for German females or male populations from other countries. Even if one is only interested in a projection for a single population it is thus worthwhile considering other populations as well. Data from those populations can provide valuable insights and help distinguish between sustainable and rather temporary effects in the mortality evolution. Therefore, we extend our projection framework to coherent modeling in the next subsections.

3.2 Coherent Projections of Age and Cohort Parameters

The age parameters prevail until infinity and thus differences in the age parameters for two populations yield steadily diverging mortality rates (assuming similar period and cohort parameter values for both populations). In particular for the case of males and females in the same country who are exposed to the same social, political and economic environment, such a scenario seems highly implausible in general. Mortality rates may be significantly different also in the long run but they should not diverge until infinity. Therefore, we need to impose the constraint of equal long-term age parameters for both genders on our model. This could be done by introducing a functional structure into the age parameters which interpolates between the fitted values for each gender and some kind of “average long-term value”. However, this functional structure would clearly increase complexity and thus reduce interpretability of our model. We could also fit our model to a combined set of historical data allowing for possibly different cohort and period parameters for both genders but demanding equal age parameters. Alternatively, we could simply average the individually fitted age parameters for males and females and assume the resulting parameter values for both genders in the future. Obviously, this is only valid if the structures in the age parameters for both genders are rather similar. This should in general be the case and, according to Figure 8, it is for our example of German males and females. We therefore proceed using this approach.

For populations from different countries, it is not obvious whether age parameters should be adjusted. This depends on the social, political and economic differences between the countries as well as the significance of differences in the age parameters in the first place. However, Edwards and Tuljapurkar (2005) show that the distributions of deaths often differ between countries. They also detect differences in the variances of life spans – in the level as well as in the trend of the variances. Thus, (slightly) different age parameters seem acceptable even for populations in closely related countries and we therefore refrain from adjusting age parameters based on cross-country information.

The cohort parameters describe only temporary effects. Thus, even if they differ for two populations mortality rates will not automatically diverge in the long run. MacMinn and Weber (2009) also show that cohort effects do not necessarily appear for males and females simultaneously and find no convincing evidence of correlated cohort effects in different countries. Consequently, we stick to the cohort parameters fitted to each population individually.

3.3 Coherent Projections of Period Parameters

The goal of this subsection is to derive a methodology for the calibration of period parameters such that projections for different populations become coherent at an overall level. To be able to include information from other populations, we require some flexibility in the forecasts of period parameters for each single population. Thus, of the three forecasting approaches presented in Section 3.1, the fitting to extrapolated life expectancies is most promising as the life expectancy extrapolations can basically have any shape. A linear extrapolation looked plausible in Figure 7 but in general, we can use any curve for the extrapolation. In short, we determine projections of mortality improvements by deriving coherent forecasts of period life expectancies at birth for all populations and then fitting the future period parameters for each individual population to the corresponding life expectancies.[14]

In order to explain our approach in more detail and to illustrate solutions to some caveats, we proceed with our example for German males. A reasonable set of reference populations is the set of male and female populations from Austria, Belgium, Denmark, Finland, France, Germany, Ireland, Italy, the Netherlands, Norway, Portugal, Spain, Sweden, Switzerland, and the United Kingdom. We do not take into account any Eastern European countries because their mortality experience was somewhat different in the past but seems to have been moving towards that of Western European countries since the fall of the Soviet Union (see, e.g., Li and Lee (2005)). Whether non-European countries like the US, Canada, Japan or Australia should be included in the analysis is very difficult to tell. Similarities in the mortality evolutions between those countries and Germany are not necessarily as strong as between Germany and its European "neighbors". As the number of populations under consideration is already fairly large, we have therefore decided to disregard any non-European countries here. We also disregard countries like Iceland and Luxembourg as they are too small to have a significant impact.

Figure 9 shows historical life expectancies for both males and females in the aforecited countries and the total male and female populations.[15] For both genders, we observe convergence in life expectancies

[14] Obviously, one could also use life expectancies at other ages, e.g. 65, or even annuity present values as the aggregated mortality statistic. However, the life expectancy at birth certainly is the most intuitive statistic.

[15] For each population, mortality data is obtained from the Human Mortality Database (2011) for years 1956 to 2006. We

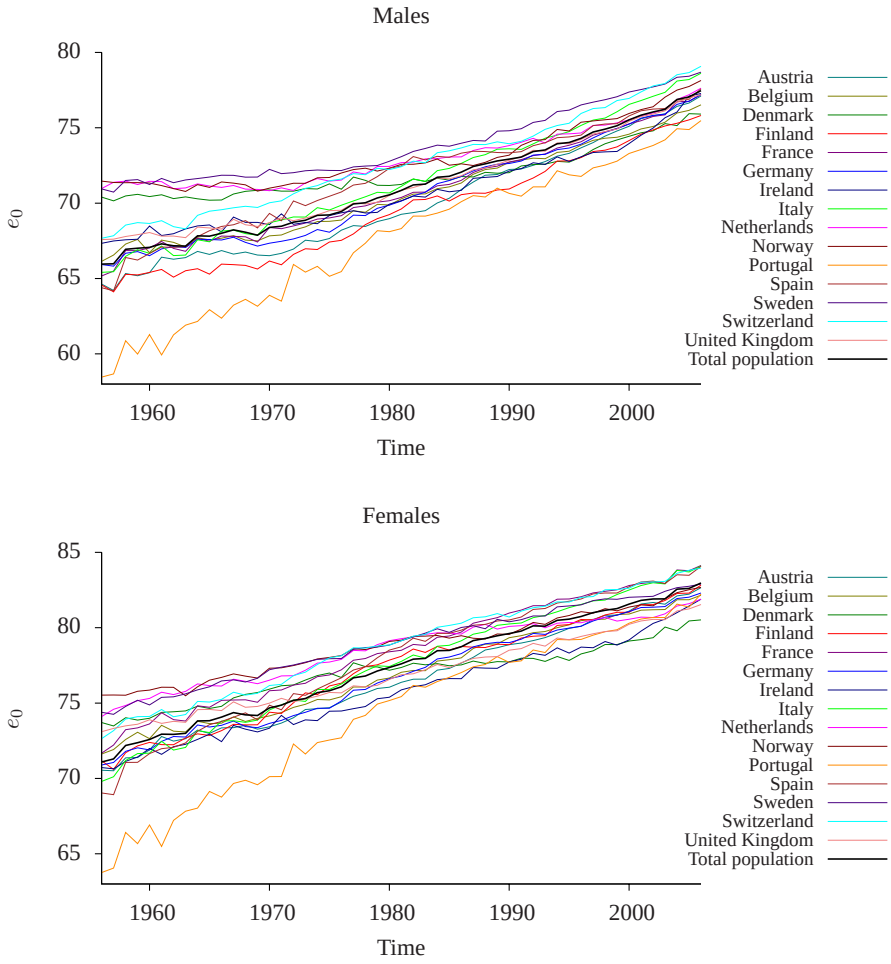

Figure 9: Historical life expectancies at birth in Europe

over time and, clearly, a common trend between countries. Clearly, extrapolations of life expectancies for each country should be related to this common trend. Therefore, we proceed by extrapolating the trends for the total populations – coherently for males and females – and by analyzing deviations between the total populations' mortality experiences and those of some exemplary single populations.

Life expectancies for the total populations have evolved pretty much linear over recent decades (for males starting from 1969) and have converged slightly. Hence, it seems obvious that coherent projections can be obtained by simply extrapolating these linear trends. Unfortunately, this is not quite the case. In Figure 10, we see historical differences in life expectancies between males and females and differences as projected by such linear extrapolations (blue line). In 2006, the extrapolations already miss the actual difference in life expectancy by more than 0.6 years. Thus, even though the extrapolation for each gender looks plausible stand-alone, in combination with the other gender it does not. Life expectancies should

restrict ourselves to this time period because, at the time of writing, only for those years data was available for all countries. For simplicity, we compute life expectancies from the HMD data up to ages 110 instead of extrapolating mortality rates for each population as we did for German males in Section 3.1. The differences should be negligible though. The life expectancies for the total population are derived from weighted averages of mortality rates for the individual populations.

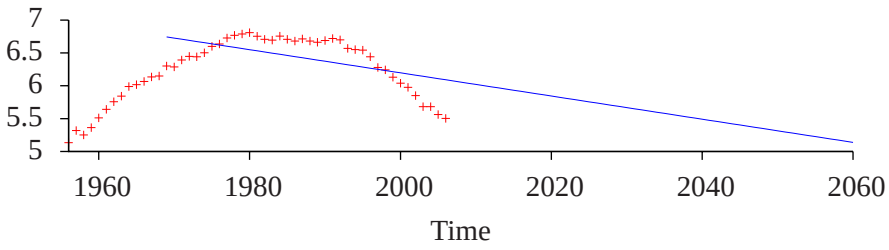

Figure 10: Difference in life expectancies between total populations

be extrapolated such that also the projected difference between genders is plausible.

Figure 10 clearly shows shrinking differences starting from the mid-1990's. However, we can hardly expect the current trend in life expectancy differences to continue until infinity. Somewhere around 2060, male life expectancy would surpass female life expectancy which seems fairly unrealistic. It is more reasonable to assume further shrinking differences for the next years which level off at a certain long-term difference above zero but below today's difference. This is in line with the convergence in lifestyles which has been observed for males and females in many European countries over the last decades. For instance, the consumption of tobacco has increased significantly for females but decreased for males thus narrowing the gap between the genders (see, e.g., European Commission (2009)). The same holds for the share of women in employment compared to the corresponding share of men (see OECD (2010)). The latter trend, in particular, is very likely to continue. So what might be a lower bound for the difference in life expectancies? Luy (2002) analyzes the mortality experience of nuns and monks who live under very similar socio-economic conditions and finds only about one year difference in remaining life expectancy in young adult ages. Thus, most of the currently observed difference for the general population seems to be related to socio-economic factors and a considerable further shrinkage may well be attainable. Nevertheless, we regard the assumption of a total convergence in lifestyles and a long-term difference of about one year as very bold. Therefore, we assume a long-term difference between European male and female life expectancies of three years in our example.

We now need to adjust the linear life expectancy extrapolations from above according to this supposed long-term difference. In fact, we need to fix a common long-term trend for males and females such that the long-term difference in life expectancies remains constant, and we need to specify how life expectancies move towards this trend and this difference over the next decades. Figure 11 shows how this can be done. The orange lines are the historical long-term trends for males (dashed) and females (solid), the green lines represent long-term asymptotes for male and female life expectancies, respectively, and the blue lines represent our actual life expectancy extrapolations. The graphs are constructed based on the following considerations:

- We assume a long-term difference in life expectancy between males and females of $\Delta = 3$ years.
- The common trend has a slope s according to the average slope of the long-term historical trends for males and females, i.e. $s = 0.2385$ as the average of 0.2473 (males) and 0.2296 (females). Alternatively, the slope could be fixed according to the historical slope for either males or females

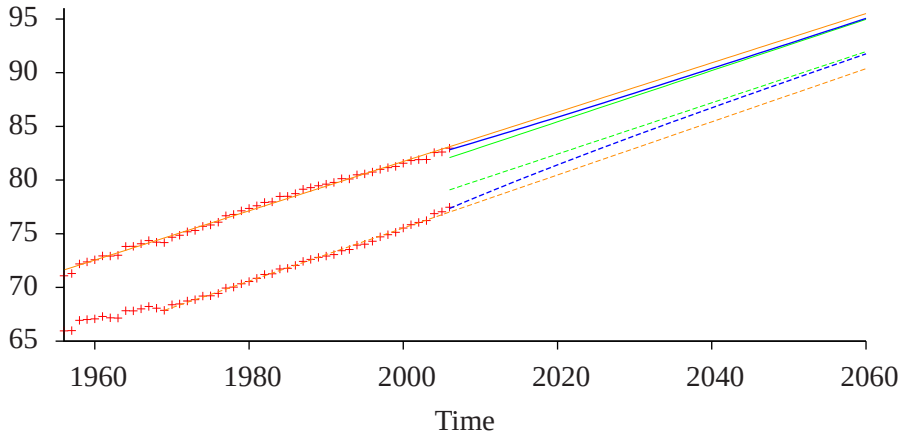

Figure 11: Coherent extrapolations of life expectancies for total populations

which might be useful for a scenario analysis. The average slope also lies right between the slopes Oeppen and Vaupel (2002) find for worldwide maximum life expectancies for males (0.222) and females (0.243) between 1840 and 2000. Even though, under our assumptions, male life expectancies in some European countries will surpass the supposed long-term maximum sometime in the (far) future, we regard our forecast as plausible. Since our assumption of a convergence between genders contrasts somewhat with the (very long-term) divergence observed by Oeppen and Vaupel (2002), a simultaneous full coherence with their extrapolations for both genders is unachievable per se.

- As the trend in life expectancy differences changes in the mid-1990's, trends for male and/or female life expectancies must (slightly) change at that time as well. Therefore, we identify the current slopes in life expectancy increases by a regression to historical life expectancies from 1995 onwards. The current trend for males is stronger than the long-term trend, i.e. the current slope is 0.3052 compared to 0.2473 in the long run. For females, the current slope is lower, i.e. 0.2099 compared to 0.2296. Our life expectancy extrapolations are set to start at the current life expectancy values and slopes for both genders.

- We assume that the extrapolated life expectancies for each gender can be written as a straight line (the long-term asymptote) plus/minus a difference term which decreases to zero exponentially with time, i.e.

$$le_m(t) = d_m + s(t - 2006) - \exp\{g_m(t - 2006) + h_m\}$$

and

$$le_f(t) = (d_m + \Delta) + s(t - 2006) + \exp\{g_f(t - 2006) + h_f\},$$

where $\cdot_m$ indicates male and $\cdot_f$ female. The asymptote for females differs from that for males only by the fixed value Δ and time is shifted simply for convenience.

- We want both life expectancy curves to converge to their asymptotes equally fast. To achieve this,

Parameters	d_m	Δ	s	$g_m = g_f$	h_m	h_f
	79.0874	3.0	0.2385	-0.0386	0.5485	-0.3008

Table 1: Parameter values for coherent life expectancy extrapolations

the slope parameters in the exponential terms, g_m and g_f, must coincide.

- These specifications and constraints leave us with a set of uniquely identifiable parameters whose values are summarized in Table 1.

After extrapolating life expectancies for the male and female total populations, we now need to specify how country specific life expectancies may evolve relative to these extrapolations. In Figure 9, we observed convergence in life expectancies for males and females across Europe which indicates that best estimate life expectancies may be equal for all countries in the long run (see also Jarner and Kryger (2011) and references therein). In that case, only transitions from current life expectancies to the common long-term life expectancies would have to be specified for each population. However, convergence seems to stop around 1980. Therefore, it is not directly clear whether the remaining variability in life expectancies is simply due to random fluctuations or whether some populations have consistently experienced longer life spans than others.

Figure 12 shows how life expectancies of selected countries have deviated from those of the total populations in the past. We have chosen these countries as we can observe significantly different patterns in their deviations which are somewhat exemplary. Regarding the question from above, the deviations for Switzerland are fairly conclusive. For both genders, they are significantly positive over the whole data period. The reason for this might be above average socio-economic conditions in Switzerland. Thus, Swiss actuaries should feel rather uncomfortable with projecting local life expectancies as being equal to the European average, even in the long run. Instead, data hints at assuming a sustainable difference of about 1.5 years and introducing a transition to that level over the next decade or so. An analogous conclusion can be drawn for Finish males where average European life expectancies seem overly conservative for a (best estimate) projection.

Opposing trends can be observed for Italy and Denmark. Italian life expectancies were below average at the beginning of the data period but have risen significantly above towards the end. Life expectancies in Denmark, on the other hand, have increased by 5 to 6 years less than the European average. Here, we see how valuable coherent projections can be. Forecasting of life expectancies according to historical trends would almost certainly yield implausible long-term projections for both countries. We would move away from the European average rapidly and continuously. Instead, it is more reasonable to assume a leveling-off in the deviations at the current level or somewhat closer to zero.

For the Netherlands, we observe a fairly linear downward trend for most of the data period. Over the last years, this trend seems to have bottomed out though – at about the European average for males and about one year below average for females. Thus, assuming sustainable differences at these levels and a long-term difference in life expectancies of three years between the genders in Europe would imply a long-term difference of only two years between Dutch males and females. This can well be possible but

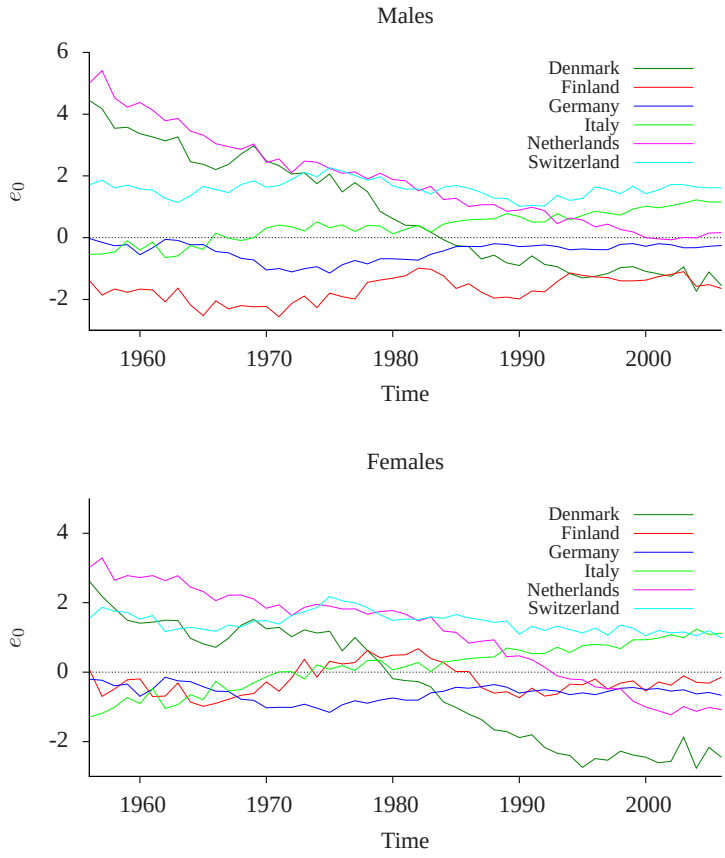

Figure 12: Differences between life expectancies of the total populations and selected countries

may also require additional demographic justification.

Finally, we have a closer look at the deviations for Germany as this is the final step to completing our example. We see in Figure 12 that, from about 1985, fluctuations become rather small around a fixed level of about -0.3 for males and -0.5 for females. Therefore, the most obvious forecast for German life expectancies is to assume the forecast for the total populations, slightly shifted downwards according to the observed deviations. We then fit the future period parameters to these life expectancies and obtain coherent projections as plotted in Figure 13.[16] The historical data is smoothed either using P-splines or our model. In the latter case, the charts also contain ages beyond 95.

We observe that our model smoothes the data more strongly than the P-spline method does. This then obviously leads to a break between the historical and the projected data in the P-spline case. In general, it is difficult to tell which level of smoothing is most appropriate. More importantly, however, the general structure in the historical data is the same for both smoothing methods.

Accepting that our model provides adequate smoothing for historical mortality improvements, the projection looks highly plausible for both males and females. In particular, all cohort dependent structures are carried forward appropriately. The very slight break in 2009 is due to the use of average age effects

[16] We show results also for German females here for completion. The projection for females has been constructed completely analogous to that for males.

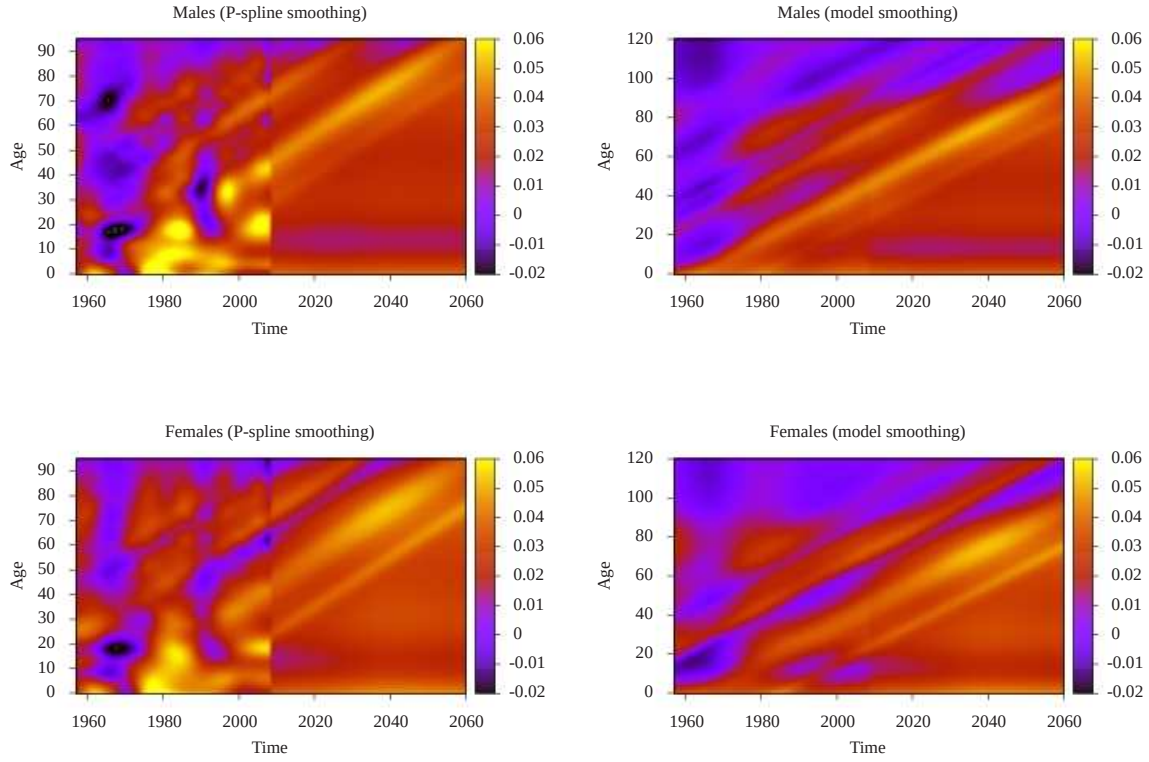

Figure 13: Coherent mortality projections for German males and females

in the projections. It is most obvious for ages around 10 where the age parameters differ most between the genders (see Figure 8) and where no cohort effects cover this difference. For males, projected mortality improvements are slightly lower than the historical data suggests and for females they are slightly larger.

4 Modeling Uncertainties and Margins

Modeling and forecasting of mortality always involves a considerable amount of uncertainty. One can never be sure whether a model's fit to the available data is adequate and whether extrapolations of historical trends into the future are appropriate. In this section, we discuss uncertainties related to our model and how they may be quantified and accounted for.

4.1 Model Uncertainty and Risk of Changes

From our point of view, the most significant model uncertainty in our projection is related to the specification of future period parameters. We calibrate these parameters based on the assumption of an ongoing (fairly linear) increase in life expectancies. But this assumption need not hold true. We have also outlined alternative approaches for specifying future period parameters and any one of those approaches may be closer to reality. However, the approach of linear life expectancy extrapolations is rather con-

servative compared to the most common approach of projecting mortality improvements based on fixed age and possibly cohort parameters. The latter approach coincides with setting future period parameters equal to zero in our setting which lead to smaller values for future life expectancies at least for German males (see Figure 7).

Model uncertainty or the risk of changes, respectively, is also inherent in the assumptions on how the mortality evolutions of different populations may relate to each other. We assume that the mortality experiences for males and females become very similar when we average age parameters for the projection. Moreover, we postulate narrowing life expectancies between both genders. The supposed long-term difference of three years is nothing else but an informed guess though. Similarly, we assume an equal long-term trend in life expectancies for all populations in Europe. In fact, life expectancies in Europe might diverge again with a significantly steeper increase for at least some countries. On the other hand, mortality experiences in Europe may become even more similar with pretty much equal age parameters for all populations in the long run. Not necessarily related to the issue of convergence or divergence is the general possibility of changes in the age dependency of mortality improvements with time. We already mentioned this uncertainty in the previous section.

We think the most effective way to account for the aforecited uncertainties is to increase the slope of assumed life expectancy gains. It can be carried out easily and induces margins for all ages increasing with time which is in line with the nature of the uncertainties. For the next years, we are rather well informed about the forthcoming mortality evolution (as long as no mortality/longevity shock occurs) but in the long run uncertainty becomes considerable. Specifying an adequately steeper slope however is not that simple. Possibly, the historical slopes for individual populations can provide some insight into the possible range of future slopes for the total population.

Another significant model uncertainty may be contained in the smoothing, in particular the smoothing of the cohort parameters. If the portfolio of a pension fund or an annuity provider consists of only a few cohorts with extraordinary mortality improvements the smoothing may hide the actual mortality experience and lead to insufficient reserves. Therefore, for very distinct portfolios it may be necessary to analyze the risk situation based on a mortality projection which is not or only partially graduated. The general inclusion of margins in a graduated projection to account for such a case does not seem imperative though.

All the highlighted issues underline the significance of model uncertainties and their increase with time – not only in our model but in any model for the projection of future mortality experience. Obviously, margins can help mitigate these uncertainties but the most effective approach certainly is to update mortality projections regularly.

4.2 Parameter Uncertainty

All parameters in our model are only fitted to a limited number of data points which are also blurred by noise to some extent. This induces parameter uncertainty which should be accounted for in a conservative projection. In general, parameter uncertainties seem rather negligible compared to the model uncertainties and the risk of changes. This holds in particular for the parameters fitted to the historical

data. For instance, if we underestimated age parameters this would be somewhat compensated for by larger period parameters in the projection given a fixed life expectancy extrapolation. However, for the sake of completeness, we still want to outline how parameter uncertainties may be quantified.
For the age and cohort parameters a bootstrap could be performed.[17] Koissi et al. (2005) and Brouhns et al. (2004) describe a residual bootstrap or a parametric bootstrap, respectively, for the Lee-Carter model which could be applied in our setting as well. The parameter uncertainty in the future period parameters stems from the uncertainty in fitting the regression line to historical life expectancies. Here confidence bounds for the regression parameters can be derived analytically.

4.3 Basis Risk

A risk not related to the construction of the projection but to its application is basis risk. Basis risk arises from the use of a projection for a population different from the one it has originally been constructed for. In this paper, we have derived a projection for the general population of German males but we might want to apply it to a population of insured or pensioners.[18]
If one population is a subpopulation of the other, as the insured are a subpopulation of the general population, future mortality evolutions should not diverge until infinity. However, over the next years mortality improvements may differ as we have observed for European populations in the previous section. Stronger improvements of the subpopulation would be critical in particular. The challenge is now to quantify this difference. Here, we need to distinguish two cases.
If no data is available for the subpopulation it is hardly possible to measure basis risk and to adjust the projection accordingly. One would have to rely on expert opinion or, possibly, information from other (sub)populations. If some data is available but not sufficient for the derivation of a full projection our setup allows to quantify basis risk. We can carry age and cohort parameters from the reference population over to the subpopulation and fit only the period parameters to the subpopulation's limited data. These period parameters are possibly more volatile than those for the typically larger reference population but the average level of both parameter sets should be very similar. Significantly different levels in the period parameters, on the other hand, would indicate the need for an adjustment to the projection according to the difference. If one questions the adequacy of the age parameters or the cohort parameters basis risk in these parameters could be measured analogously.

5 Comparison with other Projection Models

In this section we compare our projection model to alternative models and show how our approach adds to the existing literature. We focus on qualitative aspects as most existing projection models describe the future evolution of mortality rates instead of mortality improvements. Thus, a direct quantitative comparison is difficult and may be misleading. However, one general advantage of our approach in many

[17] In our case, the bootstrap would only work for age parameters up to age 95. Above, expert opinion would be required, as for the calibration of the parameters in the first place.

[18] Note that, here, we focus on the projection and thus differences in the future changes in mortality rates only. Differences in the levels of mortality rates are accounted for by using appropriate base tables.

situations is that we model mortality improvements directly. For the derivation of (standard) projections, the modeling of the level of mortality is redundant. It only binds some of the model parameters and thus, in general, requires a more complex model to provide comparable results.

A very popular projection model is the P-spline model of Currie et al. (2004) which we have already used above for smoothing mortality data. In the previous section, we saw that the fit to historical data is comparable with the fit of our model in the sense that both models detect very similar structures in the data. However, our model offers advantages in the projection of mortality improvements. We are very flexible in projecting the level of future mortality improvements which is valuable for two reasons: We can derive coherent forecasts for several populations and we can easily specify alternative scenarios, e.g. with and without margins. In the P-spline model, the level of future mortality improvements is fixed by model assumptions. This level may or may not be reasonable and cannot be adjusted according to information from other populations.

Mortality projections have also often been derived as the central trajectories of stochastic mortality models. Starting with the model of Lee and Carter (1992), a bunch of such models has been proposed over the last two decades and some of them have also been extended to yield coherent projections (see, e.g., Li and Lee (2005) and Cairns et al. (2010)). However, most of the models do not allow for cohort effects which are significant and thus highly relevant for many populations. If they do incorporate a cohort component like the Lee-Carter extension of Renshaw and Haberman (2006) the fitting often becomes unstable. Moreover, these models are generally specified as parsimonious as possible to speed up simulations. This obviously worsens the model fit and thus the quality of the projection. As an example, we cite the age dependent parameters in the Lee-Carter model which are multiplied by the time index. These parameters determine the level of mortality improvements as well as the volatility in simulated mortality rates. Thus, they are not fitted as to provide the most plausible central projection in general. Moreover, Lee and Miller (2001) show that the Lee-Carter model has tended to underpredict life expectancy gains for most countries. This is due to the assumption of linear changes in log mortality rates over time which is an assumption underlying most of the commonly used mortality models. Thus, this seems to be a general issue with those models and it is critical when deriving projections for actuarial purposes.

In the previous sections, we have already cited Jarner and Kryger (2011) who also propose a model for coherent projections. However, they allow for less variability between the populations as they derive one projection for the total population and model random fluctuations around this projection for the individual populations. Moreover, they pay less attention to details in the mortality structure by fitting a much more parsimonious model without a cohort component.

Finally, we compare our model to the Continuous Mortality Investigation (2010, CMI) mortality projection model. Both models are quite similar in structure and flexibility as they allow for age, cohort, and time dependent mortality improvements. However, we show how our model can be fully calibrated whereas the CMI leaves the derivation of parameter values to the user. In particular, we provide ideas how long-term mortality improvements can be obtained and the fitting to extrapolated life expectancies determines changes in mortality improvements over time automatically. The user of the CMI model has to decide over which time horizon age and cohort dependent mortality improvements move from their

current level to the expected long-term level. Thus, our methodology provides additional insights which can be informative for the calibration of the CMI model as well.
In conclusion, we think our model adds to the literature on mortality projections as it eliminates several drawbacks of existing projection models. Moreover, our methodology is very intuitive which clearly helps motivating the use of a particular projection.

6 Conclusion

Projections of future mortality evolutions are particularly necessary for the computation of reserves and risk management in the insurance business and for population forecasts for social security systems. The derivation of reliable projections, however, is very sophisticated and some projections which are currently used in practice seem questionable. In this paper, we develop a projection methodology which provides highly plausible extrapolations of historical mortality improvement patterns. Our model is very flexible in terms of changes in the future level of improvements and we present different ideas for fixing this level. The most promising idea is the fitting of future period parameters to extrapolated life expectancies. The extrapolation of historical life expectancies is usually much more obvious than the extrapolation of individual mortality rates or mortality improvements. At the same time, this approach provides a measure for the strength of a projection in just one aggregated statistic. But most importantly, it allows for the incorporation of information from other populations. Coherent extrapolations of life expectancies for different populations induce coherent and plausible projections for those populations at an overall level. As we have seen in Section 3, the simultaneous consideration of several populations can have a significant impact on the resulting projection for each single population.

References

Brouhns, N., Denuit, M., Van Keilegom, I., 2004. Bootstrapping the Poisson Log-Bilinear Model for Mortality Forecasting. Scandinavian Actuarial Journal, 3: 212–224.

Cairns, A., Blake, D., Dowd, K., Coughlan, G., Epstein, D., Ong, A., Balevich, I., 2009. A Quantitative Comparison of Stochastic Mortality Models Using Data from England & Wales and the United States. North American Actuarial Journal, 13: 1–35.

Cairns, A., Blake, D., Dowd, K., Coughlan, G., Khalaf-Allah, M., 2010. Bayesian Stochastic Mortality Modelling for Two Populations. Working Paper, The Pensions Institute.

Continuous Mortality Investigation (CMI), 2010. Working Paper 49 – The CMI Mortality Projections Model, CMI_2010. Available at: www.actuaries.org.uk.

Currie, I., Durban, M., Eilers, P., 2004. Smoothing and Forecasting Mortality Rates. Statistical Modelling, 4: 279–298.

Deutsche Aktuarvereinigung, 2004. Herleitung der DAV-Sterbetafel 2004 R für Rentenversicherungen. Deutsche Aktuarvereinigung e.V., Cologne.

Edwards, R., Tuljapurkar, S., 2005. Inequality in Life Spans and a New Perspective on Mortality Convergence Across Industrialized Countries. Population and Development Review, 31: 645–674.

European Commission, 2009. Data and Information on Women's Health in the European Union. Available at: http://ec.europa.eu/health/population_groups/gender/index_en.htm.

Gampe, J., 2010. Human Mortality Beyond Age 110. In Maier, H., Gampe, J., Jeune, B., Robine, J.-M., Vaupel, J., Eds.: Supercentenarians. Springer, Heidelberg.

Hinkley, D., 1969. On the Ratio of Two Correlated Normal Random Variables. Biometrika, 56: 635–639.

Human Mortality Database, 2011. University of California, Berkeley, USA, and Max Planck Institute for Demographic Research, Germany. Available at: www.mortality.org.

Jarner, S., Kryger, M., 2011. Modelling Adult Mortality in Small Populations: The SAINT Model. Working Paper, ATP Pension fund.

Koissi, M.-C., Shapiro, A., Högnäs, G., 2005. Evaluating and Extending the Lee-Carter Model for Mortality Forecasting: Bootstrap Confidence Intervals. Insurance: Mathematics and Economics, 38: 1–20.

Lee, R., Carter, L., 1992. Modeling and Forecasting US mortality. Journal of the American Statistical Association, 87: 659–671.

Lee, R., Miller, T., 1992. Evaluating the Performance of the Lee-Carter Method for Forecasting Mortality. Demography, 38: 537–549.

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

Luy, M., 2002. Warum Frauen länger leben. Erkenntnisse aus einem Vergleich von Kloster- und Allgemeinbevölkerung. Materialien zur Bevölkerungswissenschaft 106, Bundesinstitut für Bevölkerungsforschung, Wiesbaden.

MacMinn, R., Weber., F., 2009. Select Birth Cohorts. Working Paper, Ludwig-Maximilians-Universität München and Illinois State University.

OECD, 2010. OECD Factbook 2010: Economic, Environmental and Social Statistics. Available at: http://www.oecd-ilibrary.org/economics/oecd-factbook-2010_factbook-2010-en.

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

Renshaw, A., Haberman, S., 2006. A Cohort-Based Extension to the Lee-Carter Model for Mortality Reduction Factors. Insurance: Mathematics and Economics, 38: 556–570.

Thatcher, A., Kannisto, V., Vaupel, J., 1998. The Force of Mortality at Ages 80 to 120. In Odense Monographs on Population Aging 5. Odense University Press, Odense, Denmark.