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

Matthias Börger* and Marie-Christine Aleksic†

Presented at the Living to 100 Symposium

Orlando, Fla.

January 8–10, 2014

* Institute for Finance and Actuarial Sciences (ifa), Ulm & Institute of Insurance, Ulm University, Lise-Meitner-Straße 14, 89081, Ulm, Germany. Phone: +49 731 20644 257. Fax: +49 731 20644 299. Email: m.boerger@ifa-ulm.de

† Kern Mauch und Kollegen, Stuttgart & Institute of Insurance, Ulm University

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

Matthias Börger

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

Lise-Meitner-Straße 14, 89081 Ulm, Germany

Phone: +49 731 20644 257. Fax: +49 731 20644 299

Email: m.boerger@ifa-ulm.de

Marie-Christine Aleksic

Kern Mauch und Kollegen, Stuttgart & Institute of Insurance, Ulm University

March 2014

Abstract

The projection of future mortality experience constitutes a challenge for both actuaries and demographers. As we show, some of the currently used standard mortality projections have several shortcomings which might pose a serious threat to insurers, pension funds, and social security systems.

In this paper, we propose a new projection methodology which overcomes these shortcomings. We introduce a model which allows mortality improvements to depend on age, period, and cohort, and we explain how the model can be estimated and applied. In particular, we show how coherent projections for several populations, i.e. males and females of the same country and populations from closely related countries, can be derived. The basis for these projections are coherent extrapolations of historical life expectancies. As aggregated mortality statistics, life expectancies typically exhibit steady patterns which often makes forecasting rather obvious. We observe that the incorporation of information on the mortality experience of other populations can have a significant impact on the projection for a given population. A comparison with other commonly used projection models shows that our methodology provides stable and highly plausible projections. Finally, we discuss uncertainties in our projection approach and explain how they can be accounted for. 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.

1 Introduction

Longevity risk, i.e. the risk of insured/pensioners living longer than expected, is one of the most prominent 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. The potential need for additional funds 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 enormously in recent decades. Riskless yields in the financial markets have fallen considerably in many countries leaving only little 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.

For the assessment of longevity risk, e.g. for risk management or solvency purposes, stochastic mortality modeling or a scenario analysis is required. However, the resulting risk capital charges are only credible if best estimate mortality is projected adequately. If best estimate liabilities are systematically underestimated, e.g., by an inappropriate mortality projection the risk capital charges will never guarantee the desired safety level.

Thus, mortality projections do not help in quantifying longevity risk, but an adequate mortality projection can significantly reduce longevity risk. Therefore, mortality projections are extremely important for practical actuarial work, and insurers, pension funds, and social security institutions should consistently look to improve their projections. However, graphical analyses reveal that some of the 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 also be identified for many other projections, and it indicates 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

[1] The historical improvements are derived from data from the Human Mortality Database (2012) (see Appendix A for details), 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 in the table DAV 2004 R was derived from data for that population, too. The trend parameters in the projection are set to $T_1 = 10$ and $T_2 = 15$, and margins for possibly stronger mortality improvements in insured mortality have been deducted. For more details on this projection and its parameters, we refer to Deutsche Aktuarvereinigung (2004, DAV).

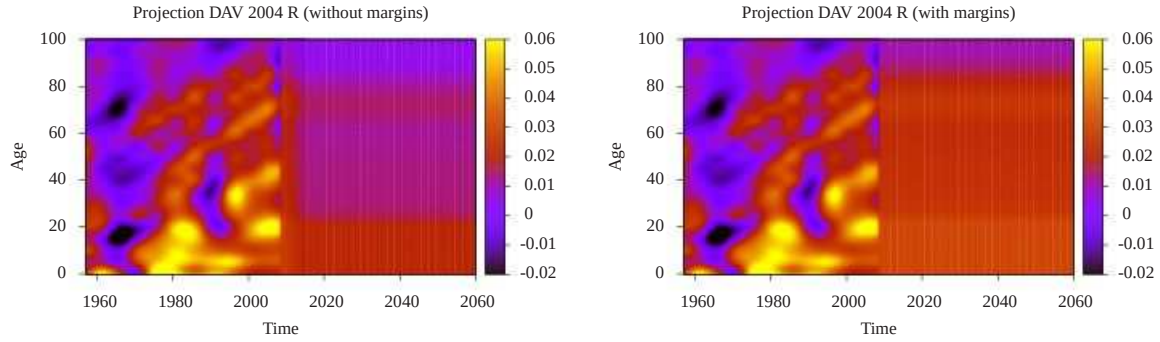

Figure 1: Standard mortality projection for German annuity business

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 table DAV 2004 R including margins. The structure of the projected mortality improvements still looks inadequate, 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 might be sufficient for a portfolio of contracts with a widely spread age distribution. However, regulatory requirements can prohibit the balancing of profits and losses from different products or product tranches. Thus, a spread of risk over a wide range of ages is not always 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 as well as standard projections in other countries.

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 propose a model structure which allows mortality improvements to depend on age, period, and cohort – in contrast to many existing projection methods which account for only one or two dependencies.[2] Nevertheless, the proposed model structure is fairly simple which facilitates a clear understanding of historical improvement patterns. We also provide a framework for estimating the model and explain how coherent projections for several populations as well as different improvement scenarios can be derived. In the past, projections for males and females and/or different populations have typically been constructed independently from each other which often lead to inconsistent long-term forecasts. As we show, taking into account data from several populations can improve the reliability of projections for each single population.

It is important to note that we do not propose a fully fixed and data driven estimation and forecasting procedure as it is typically applied for statistical projection models. Instead, we only establish a

[2]Note that, in many projection models for mortality rates which account for age, period, and cohort dependencies, the projected mortality improvements often depend on period and cohort only. The age parameters then only describe the level of mortality with age and not changes in this level over time.

framework by outlining the necessary steps for estimation and forecasting and provide examples how each of these steps can be implemented in practice. Thus, only the general procedure is fixed, while the concrete design of each step is to be customized to the population under consideration and the data available. Such a customized approach obviously requires some expert judgment which is often seen as a disadvantage compared to fully data driven extrapolation models due to the implied loss in objectivity. However, we believe that this approach also offers several advantages. In particular, it allows the user to directly implement what he regards as consistent with the historical data and coherent between different populations. When using statistical extrapolation models, this implementation is done rather indirectly. One would typically choose the one model from the bunch of available models which appears to yield the most plausible results given one's own understanding of historical mortality patterns and expectation of the future mortality evolution. In fact, statistical models are often modified or the data set is determined such that model outcomes are in line with the user's intuition. The approach proposed in this paper avoids this indirect implementation as well as being limited to the projection scenarios which available statistical extrapolation models can provide. At the same time, the fixed framework for model estimation and forecasting reduces the necessary user input to a minimum. Therefore, we believe that the proposed methodology provides a useful alternative to fully data driven approaches, in particular with respect to Solvency II requirements. Once Solvency II comes into effect, insurers will have to explain in detail how they derived their mortality assumptions and why they think these assumptions are valid. When using the framework of this paper, the reasoning comes with the derivation of a projection automatically.

Most existing projection models consider mortality rates from which mortality improvements can then be deduced. We directly forecast improvements as we see two main advantages in this approach: 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 interpretability of the remaining parameters. Secondly, some features in the underlying mortality data become more obvious when changes in mortality rates are considered instead of the mortality rates themselves. Appendix A provides an example of such a feature. A case for modeling mortality rates is often seen in the stability of mortality rates compared to the improvements which can be interpreted as the rates' first derivatives. However, from our point of view, this is more a question of specifying an adequate model structure. As mentioned above, the number of parameters should be smaller in case of a model for improvements, thus compensating for the larger fluctuations in the data.

The increasing demand for assessing and managing longevity risk has provoked considerable academic research in this field, with a clear focus on stochastic mortality modeling.[3] Starting with the model of Lee and Carter (1992), a variety of mortality rate models has been proposed over the last two decades, and some of them have also been extended to yield coherent forecasts for several populations (see, e.g., Li and Lee (2005) and Cairns et al. (2011)). Of course, projected mortality improvements can be derived from their central trajectories. However, these stochastic models are generally specified as parsimonious as possible to speed up simulations. This typically implies that certain structures in the mortality data are ignored, e.g. cohort effects. Thus, in general, the central trajectories of stochastic

[3] For an overview of stochastic approaches to modeling mortality, we refer the interested reader to Cairns et al. (2008).

mortality models are not the best choice for a deterministic mortality projection. Moreover, Lee and Miller (2001) show that the Lee-Carter model has tended to underestimate life expectancy gains for most countries. This is due to the assumption of linear changes in log mortality rates over time. Since this assumption is also embedded in most other commonly used stochastic mortality models, there seems to be another general issue which limits these models' applicability for deriving mortality projections. A model which is particularly designed for deriving deterministic forecasts can account for such issues more easily, thus providing more reasonable best estimate projections while keeping model complexity at the necessary minimum. Explicit models for mortality improvements have been proposed by Plat (2011) and Haberman and Renshaw (2012). The authors come to the conclusion that stochastic modeling of mortality improvements can be a valuable alternative to modeling mortality rates. However, they also focus on stochastic modeling and not the derivation of most plausible and adequate deterministic projections.

A very popular (deterministic) projection model is the P-spline model of Currie et al. (2004). This model is very flexible in that it can account for age, period, and cohort dependent effects, and we will see later that it discovers basically the same patterns in the historical data as our model. However, projections from the P-spline model can be quite implausible at times. We provide an example for that when we compare projections from our model and the P-spline model in Section 5. Moreover, the P-spline model does not allow for coherent projections for different populations. A model for coherent projections has been proposed by Jarner and Kryger (2011). However, in comparison to our approach, this model is rather strict as it only considers one projection for the total population and random fluctuations around this projection for the individual populations. Thus, the central projections for all populations under consideration are equal in the long run which is often not a plausible assumption. For instance, mortality rates in Switzerland have constantly been lower than mortality rates in most other European countries, and it seems rather bold to assume that this will change in the future. Hyndman at al. (2012) overcome the issue of equal long-term projections and, instead, allow for projections with constant differences between log mortality rates in the long run. Both Jarner and Kryger (2011) and Hyndman at al. (2012) project mortality rates and not improvements though and they do not account for cohort effects which limits the applicability of their models for populations where cohort effects exist. The Continuous Mortality Investigation (2010, CMI) has proposed and regularly updated a mortality projection model the structure of which is rather similar to the structure of our model. However, we provide a framework for the full calibration of our model, whereas the CMI leaves the derivation of some crucial parameter values to the user. In particular, we show how long-term mortality improvements can be obtained from life expectancy forecasts. This also automatically determines changes in mortality improvements over time. 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.

The remainder of this text is structured as follows: In the next section, we analyze historical mortality improvement patterns and deduce the specification of our projection model. In Section 3, we then establish the estimation framework, present several approaches for fitting the model to historical data, and discuss the advantages and the applicability of each of those approaches. We also show how random

noise in the model parameters can be eliminated and how the applicability of possible model simplifications can be checked. The derivation of mortality projections is then discussed in Section 4. In particular, we explain how coherent forecasts for several populations can be obtained. In the subsequent section, we perform a back test and compare our model to other projection models which are often used in practice. We then analyze uncertainties inherent in our projection methodology and describe ways to assess and account for these uncertainties. Finally, Section 7 concludes.

2 Model Specification

The mortality improvements in Figure 1 are computed as[4]

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

The historical parts in the heat charts clearly show vertical and diagonal structures, i.e. mortality improvements depending on period and cohort. Moreover, we can observe that improvements for younger ages tend to be larger on average than for older ages. Thus, there also seems to be a dependency on age. Clearly, for other data sets, only one or two of these dependencies might be relevant, but a generally applicable projection model should allow for all three. In Subsection 3.3, we explain how the relevancy of each dependency can be tested. In order to keep the model as simple as possible and to ensure a robust estimation of its parameters, we propose modeling mortality improvements according to the well-known Age-Period-Cohort (APC) model:

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

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. Obviously, other model structures which include age, period, and cohort components could also be applied. However, as we will see, this most simple model structure works very well. More sophisticated structures, as in the model of Renshaw and Haberman (2006), might provide a slightly better fit but the significantly larger number of parameters makes the model estimation more difficult. In fact, we have experienced numerical problems when trying to estimate the Renshaw-Haberman model and, therefore, do not consider this model as a reasonable alternative to the APC model. The cohort variant of the Cairns-Blake-Dowd model (see model M6 in Cairns et al. (2009)) possesses less parameters than the APC model but assumes a linear trend in the age component which is not a reasonable a priori assumption in any case.

In the following sections, we explain the general methodology for estimating the model parameters and deriving future mortality improvements. We first present three approaches for fitting the model to historical data. This provides us with age and cohort parameters which we adopt for projecting, i.e. we assume that, in the future, mortality improvements will depend on age and cohort as they did in the past. For

[4]Obviously, this definition of mortality improvements is only valid as long as $q(x,t-1) > 0$. For populations as large as the German one, this is always the case, but for smaller populations, there might be raw mortality rates of zero. A few undefined mortality improvements would be uncritical for the estimation of our model though.

the projection of future period parameters, we rely on life expectancy forecasts. Period life expectancies at birth or any other age exhibit rather steady and often almost linear patterns since they are aggregate mortality statistics. This makes forecasting rather obvious in many cases. However, any assumption on the future life expectancy evolution can be incorporated in our framework. Coherence between projections for several populations is also to be achieved at the aggregate level of life expectancies. Here, one needs to provide reasonable long-term relationships between life expectancies for the populations under consideration. For instance, life expectancy forecasts for females should be larger than those for males until infinity in general. Once life expectancies are extrapolated, the period parameters can be deduced such that the assumed life expectancy for each respective year is obtained. This approach implies that the historical period parameters from the model estimation are not required for the projection. Nevertheless, we think that they are an integral part of the model as, in the fitting to historical data, they take up a significant amount of the random year by year fluctuations and, thereby, improve the estimation of the age and cohort parameters.

In order to illustrate our methodology, we derive projections for German males and females as part of a larger reference set of European populations. We show exemplarily how each step in our rather general framework for model estimation and projection can be carried out such that it best fits the population under consideration and the available data. Note though that the projection methodology is not derived specifically for the case of Germany. It can be applied to basically any population with a sufficient data history.

For our example projections, we use historical data for West Germany from the Human Mortality Database (2012). 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 (see, e.g., Kunde and Ortmann (2011)). Thus, a combined data set may be blurred by this one-off effect. However, for simplicity, we refer to the West German population as the German population only in the following. For reasons outlined in Appendix A, we do not use the mortality rates which are tabulated in the Human Mortality Database but derive mortality rates from the also provided deaths counts and population sizes. This yields mortality rates for years 1956 to 2008 and ages 0 to 109. However, since mortality rates fluctuate extremely for very old ages or are not even defined due to nobody being alive at those ages anymore, we cap the data set at age 100.

3 Projection of Age and Cohort Parameters

As outlined above, age and cohort parameters for the projection are to be obtained from fitting the model to historical data. For this, we propose a four step approach:

Step 1: Estimation of all model parameters based on mortality improvements in the historical mortality data.

Step 2: Modification of the parameters to account for the fact that the APC model is not uniquely determined. Constraints are imposed on the model parameters ex post such that the parameters have a clear interpretation and that the parameters of different populations become comparable.

Step 3: Smoothing of the raw parameters to eliminate remaining noise, and possibly significance tests of the model components.

Step 4: If applicable, another modification of the age parameters to obtain coherent projections for the populations under consideration. Since the age parameters prevail until infinity, differing age parameters for two populations imply that – assuming rather equal period parameters – mortality rates will diverge in the long run.

These four steps are discussed in detail in the following subsections where we explain rather generally how each step could be carried out and where we provide concrete implementations for the example case of Germany.

3.1 Step 1: Model Estimation

There are several ways to calibrate the model parameters to historical mortality improvements, and we present three possible approaches here. The most important difference between them is the assumed distribution for the improvements.

Least squares estimation

The most intuitive approach is weighted least squares estimation of the model parameters. The general need for weighting becomes obvious from Figure 9. The variability in the mortality improvements differs significantly between different age groups and periods. For young ages and 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 most obvious choice of weights, i.e. the improvements' variances, are unknown. Therefore, we suggest approximating the variances from the residuals and applying these sample variances as weights. To this end, one could assume that the variances are rather constant for adjacent ages and calender years and, for each improvement, compute the sample variance from a set of surrounding data points. The appropriate number of data points clearly depends on the data set under consideration.[5] Since the sample variances can only be computed from the residuals once the model has been estimated, an iterative procedure is required. In a first run, the

[5] For the case of German males and females, we have found that a rectangle of data points with length 5 in the age direction and length 15 in the time direction is a good choice. Thus, for age 30 in 1970, for instance, the residuals for ages 28 to 32 in the years 1963 to 1977 are considered. For ages and calender years close to the boundaries of the data set, this rectangle shrinks obviously. The dimensions of the rectangles are chosen as small as possible to accommodate the assumption of equal variances for all data points in a rectangle, and as large as necessary to eliminate any significant random fluctuations in the variance estimates. However, in general, we have found that the number of data points for computing the empirical variances has a rather small impact on the model estimation results.

model has to be estimated without weights (or with weights all equal to 1); then residuals and sample variances can be derived and applied in a second weighted estimation run. This procedure of deriving weights and re-estimating model parameters should be repeated until the model estimation converges, i.e. until the model parameters hardly change from one run to the next.

By estimating in least squares, one implicitly assumes normally distributed residuals. However, this is not always an adequate assumption; in particular, the light tails of the normal distribution may be inappropriate in many cases. Given initial exposure, the number of deaths can be assumed as binomially distributed, and thus, for a sufficiently large population, mortality rates can be seen as approximately normal. This means that, according to Equation (1), the improvements follow a distribution of the ratio of two normally distributed random variables. Such a ratio distribution can be stated explicitly (see Hinkley (1969)), but its tails are so heavy that moments are not defined. Nevertheless, least squares estimation can still be a valid approach for populations with rather steady mortality patterns.

Maximum Likelihood Estimation based on Non-standardized Student's t-Distribution

In order to account for possibly heavy tails, it seems reasonable to assume the ratio distribution of normal random variables for the improvements. However, in order to fully specify this distribution, means and variances of the two normal variables, i.e. the mortality rates in our case, are required which we do not know in general and which we cannot express in terms of our model parameters. As an alternative, we can assume a non-standardized student's t-distribution for the improvements or residuals, respectively, which – depending on the chosen degrees of freedom – has significantly heavier tails than the normal distribution. The model parameters can then be estimated via the method of maximum likelihood. As for the least squares case, estimation should be carried out iteratively, with the scale parameter in the student's t-distribution updated after each run according to the sample variances. The degrees of freedom of the student's t-distribution can be derived from distributional tests; the t-distribution should be as close as possible to the empirical distribution of the residuals in the final estimation run.

Least squares Estimation based on the Logarithm of Mortality Reduction Factors

An alternative approach for estimating the model parameters is to assume a lognormal distribution for the mortality reduction factors

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

This assumption can be motivated by the fact that mortality improvements lie in the interval $(-\infty, 1]$ by definition. Thus, reduction factors range from zero to infinity, with mean typically somewhere slightly below 1. This indicates that a skewed distribution with the positive axis as support might be a reasonable choice. Moreover, the binomial distribution for the number of deaths and thus also the distribution for the mortality rates might sometimes be better approximated by the lognormal distribution compared to the normal distribution. This is particularly the case when the number of deaths is rather small and the

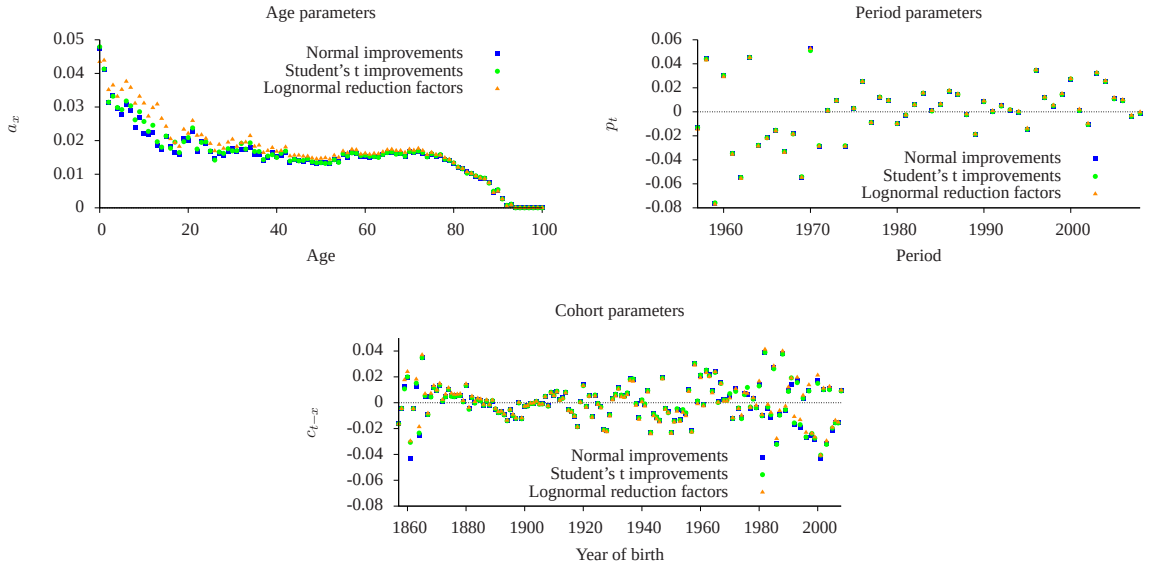

Figure 2: Estimated parameter values for German males

binomial distribution becomes significantly skewed. We then obtain

$$\log\{r(x,t)\} = \log\left\{\frac{q(x,t)}{q(x,t-1)}\right\} = \log\{q(x,t)\} - \log\{q(x,t-1)\} \sim N,$$

i.e. the log reduction factors are normally distributed, and the model can be estimated via iteratively re-weighted least squares or, equivalently, maximum likelihood on these factors. This approach appears particularly suited for small populations with rather small numbers of deaths.

Estimation Results for German Males and Females

Figure 2 shows the three parameter sets for the three estimation approaches (the degrees of freedom in the student's t-distribution are set to 30) and German males. The parameter modifications which are described in the following subsection have already been applied to the plotted parameter values, but this does not affect the conclusions we draw here. We observe that the period and the cohort parameters are almost equal for all three approaches. Moreover, the period parameters fluctuate considerably which shows the year by year random noise. In the age parameters, we observe significant differences between the estimation approaches only for young ages. There, the volatility in the raw mortality improvements is rather large, and the distributional assumptions thus have a significant impact on the estimation results. The most applicable estimation approach for a given data set can be determined by distributional tests. Figure 3 shows a combined QQ-plot of the residuals in the case of German males. Here, it seems that the lognormal assumption is least applicable which is confirmed by the smallest R^2 statistic. Table 1 contains p-values of Kolmogorov-Smirnov (KS) tests and Chi-squared tests for both males and females. For females, test results are fairly conclusive; the assumption of a student's t-distribution cannot be rejected in both tests, whereas the normal and the lognormal distribution are only slightly significant in the Chi-squared test. For males, the Chi-squared test clearly rejects the assumption of a lognormal

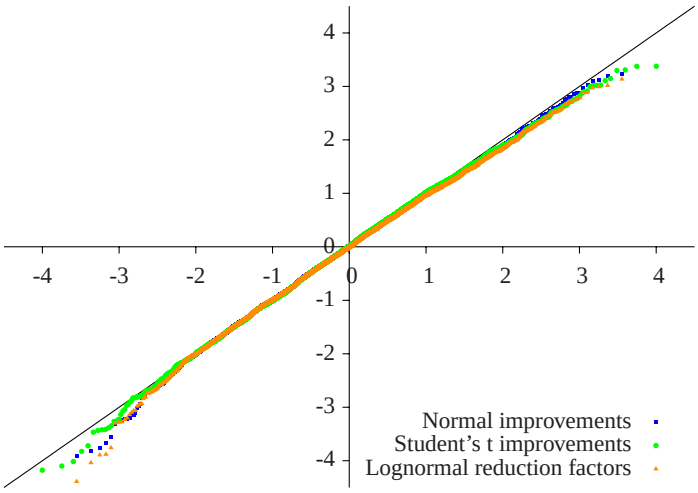

Figure 3: QQ-plot for standardized residuals in the three model estimation approaches

	Females		Males	
Distribution	**KS test**	**Chi-squared test**	**KS test**	**Chi-squared test**
Normal	0.059	0.028	0.336	0.028
Student's t	0.108	0.114	0.402	0.030
Lognormal	0.017	0.020	0.044	0.003

Table 1: P-values for Kolmogorov-Smirnov (KS) and Chi-squared distributional tests

distribution which confirms our observation from the QQ-plot. The p-values for the other two candidate distributions are very similar in both tests, with the Kolmogorov-Smirnov test clearly accepting both distributions. In conclusion, a student's t-distribution appears most reasonable for both genders, and we therefore proceed by using the parameter values from that estimation approach in our example.

3.2 Step 2: Parameter Modifications for Model Identifiability

The parameters in the APC model cannot be uniquely determined without constraints. For instance, increasing all age parameters and decreasing all period parameters by the same amount would not change the model's responses, i.e. the estimated mortality improvements. Therefore, constraints are typically imposed before the model parameters are estimated. In our estimation framework, we do not explicitly state such constraints but constrain the model implicitly such that all parameters can be easily interpreted and that coherent projections of the age parameters for different populations are intuitive (details on this issue follow in Subsection 3.4). Due to the linear structure in the APC model, sets of optimal parameter estimates only differ by linear trends in age and time (cf. Carstensen (2007) or Kuang et al. (2008a)). Thus, a switch from one set of constraints to another is equivalent to deducting a linear trend from one parameter set and adding it to the other parameters sets. The shifting of linear trends does not affect the estimated mortality improvements; in fact, such shifts do not even affect projected improvements in the APC model as long as linear trends in the period and cohort parameters are extrapolated into the future

(cf. Kuang et al. (2008b)). In line with these findings, we accept any set of "optimal" parameter estimates in Step 1 instead of constraining the parameters before estimation and then modify these estimates as follows: We determine a possible linear trend $m_c \cdot (t - x) + d_c$ in the cohort parameters and subtract this trend from the original cohort parameters.[6] The modified cohort parameters then all lie around zero as can be seen in Figure 2. To compensate for the change in the cohort parameters, the age component of the estimated trend, i.e. $-m_c \cdot x + d_c$, is added to the age parameters and the period component $m_c \cdot t + d_c$ to the period parameters. Moreover, we modify the period parameters by subtracting their mean and adding this mean to the age parameters.

These parameter modifications should be carried out for any population, not only in our example case of Germany, since they provide all parameters with clear interpretations: The cohort parameters describe the differences in mortality improvements between cohorts; a cohort parameter greater than zero means that the respective cohort experiences larger than average improvements, whereas a negative cohort parameter indicates below average improvement. The period parameters are zero on average, and we can thus identify years and/or decades with above- or below-average improvements. In Figure 2, we observe a generally increasing tendency in the period parameters which means that improvements have been increasing for German males over the last half-century. The age parameters can be interpreted as the average mortality improvement for each respective age over the whole period under consideration.

The elimination of a possible trend in the cohort parameters supports the comparison of improvements for different cohorts. Furthermore, parameters for future cohorts can then be plausibly forecast in a very simple manner, i.e. by setting them equal to zero. Since one hardly has any knowledge about the mortality improvements of cohorts still to be born, the most obvious forecast is to assume average improvements which, after the elimination of a trend, are zero improvements. If we still observed a trend in the historical cohort parameters, a reasonable projection for this trend, which would to some extent indicate increasing or decreasing improvements with time, would have to be derived.

Since the first and last cohort parameters are fitted to only a few data points, they can assume rather large values which are most likely due to random noise rather than actual cohort effects. In order to account for this, the cohort parameters should be "standardized" by multiplying them by $\sqrt{\frac{n}{m}}$ where n is the number of data points available for estimating the parameter under consideration and m is the maximum number of data points available for estimating any of the cohort parameters, i.e. the maximum of ages and periods under consideration.[7] This standardization reduces the volatility in the first and last cohort parameters in particular and makes volatility comparable for the whole data set which is necessary for the parameter smoothing in Step 3.

Finally, one needs to extrapolate the age parameters up to the desired limiting age of the projection. Here, different approaches can be applicable, depending on the data set under consideration. In our example, simple extrapolations equal to zero appear plausible for both genders. This is in line with the

[6] This trend should be obtained via weighted regression to account for the increasing volatility in the first and last cohort parameters. We suggest applying weights $\frac{n}{m}$ where n is the number of data points available for the parameter under consideration and m is the maximum number of data points available for estimating any of the cohort parameters, i.e. the maximum of ages and periods under consideration. These weights can be motivated by the fact that, in general, the variance of a parameter estimate decreases linearly with the number of data points available for estimation.

[7] These weights can be motivated by the fact that, in general, the variance of a parameter estimate decreases linearly with the number of data points available for estimation.

finding of Gampe (2010) that mortality rates for the oldest old in Europe have not improved in recent decades. Moreover, we set all negative age parameters to zero since forecasting mortality deterioration for some ages does not seem plausible in the long run when positive mortality improvements are assumed in general.

3.3 Step 3: Parameter Smoothing and Significance Tests

The estimated parameter values typically still contain some random noise which can also be seen in Figure 2. Thus, smoothing is required. In principle, any smoothing method can be applied; we have decided to use the method of Whittaker and Henderson (see Whittaker (1923) and Henderson (1924, 1925)) in our example as it is well known and widely applied in actuarial science. The method of Whittaker and Henderson requires provision of a smoothing parameter which determines how strongly the data is smoothed or – put differently – how close the smoothed data still is to the original data. We suggest using a smoothing parameter which minimizes generalized cross-validation.[8] The idea behind the concept of generalized cross-validation is that the error in predicting any one data point based on all other data points is minimized (see, e.g., Craven and Wahba (1978)). This intuition coincides with our goal of determining the "true" parameters from the noisy estimates, and thus, the smoothing parameter which minimizes generalized cross-validation seems a natural choice.[9]

Once the model parameters are smoothed, they can be used for projecting future mortality improvements. However, in order to have a model as parsimonious as possible, the significance of the age and cohort parameters may be checked.[10] To this end, likelihood ratio tests can be carried out for restricted models without age or cohort parameters. These tests should be performed on the smoothed parameters since those are the parameters used for projection. A test on the raw parameters might be misleading. For instance, if the raw cohort parameters fluctuated considerably due to random noise, a likelihood ratio test might indicate high significance since, without the cohort component, the absolute values of the residuals would increase significantly. The smoothed cohort parameters might all be (close to) zero though which means that the cohort component is not relevant for the projection at all.

Smoothing of the parameter estimates reduces the effective number of parameters or the degrees of freedom in our model. If, for instance, the smoothed parameters follow a straight line, the effective number of parameters is reduced to 2. Consequently, the likelihood ratio test statistic follows a chi-squared distribution with degrees of freedom equal to the difference between the effective numbers of parameters in the full model and the restricted model. In both cases, the effective number of parameters can be determined as the trace of the so called hat matrix in the Whittaker-Henderson smoothing (see, e.g., Garcia (2010) for details). Exemplarily, Table 2 contains the effective numbers of parameters for different model variants and p-values for according likelihood ratio tests for German males and females. We observe that smoothing significantly reduces the number of parameters in the full model, from 305 to

[8] The concept of minimizing generalized cross-validation for smoothing the parameters of a mortality model is not new. It has already been applied by Delwarde et al. (2007) for smoothing the age dependent parameters in the Lee-Carter model.

[9] The concept of generalized cross-validation assumes normally distributed errors in the noisy data which implies a slight inconsistency if the mortality improvements are not assumed normal, but we regard this issue as negligible.

[10] Obviously, also the period parameters could be checked for significance. However, since they are not used for projection in our framework, but just included for improving the estimation of the age and cohort parameters, this is counterintuitive.

Model variant	Females		Males	
	#parameters	p-value	#parameters	p-value
APC model	48.9		51.5	
AP model	21.2	$2.2 \cdot 10^{-15}$	24.0	$8.0 \cdot 10^{-12}$
PC model	31.8	$7.5 \cdot 10^{-10}$	18.8	$4.4 \cdot 10^{-2}$

Table 2: Effective numbers of parameters for model variants and p-values of likelihood ratio tests

about 50 for both genders. Moreover, the age and cohort components are significant – the age component for males only slightly though. Thus, the projection model cannot be restricted in our example.

3.4 Step 4: Parameter Modifications for Coherent Projections

With the age and cohort parameters smoothed and checked for significance, projections for individual populations could be derived. However, another adjustment may be necessary in order to guarantee coherent projections between closely related populations in the long run. The age parameters prevail until infinity, and thus, differences in the age dependencies for two populations yield steadily diverging mortality rates (assuming similar period 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 as a best estimate scenario. Mortality rates may be significantly different also in the long run, but they should not diverge until infinity. This issue can be overcome by requiring coherent long-term age parameters for both genders.

Due to the parameter modifications in Step 2, coherence can be obtained by assuming equal age parameters in the long run. The shifting of the linear trend from the cohort parameters to the other parameters and the resulting projection of the cohort parameters as zero imply that there is no age dependent trend in the cohort parameters anymore. If there were such trends in the cohort parameters for two populations and if these trends were different, equal age parameters would not imply coherence. The different age-dependent trends in the cohort parameters would yield diverging mortality rates until infinity. Thus, the parameter modifications in Step 2 do not only provide all model parameters with an intuitive interpretation but also prevent long-term divergence in projected mortality rates for related populations.

Equal long-term age parameters, e.g., for males and females in the same country could be obtained by introducing a functional structure in the age parameters which interpolates between the estimated parameters for each gender and some kind of “average long-term age parameters”. However, such a functional structure would significantly increase model complexity. Alternatively, one could fit the APC model to a combined set of historical data, allowing for possibly different cohort and period parameters for both genders but demanding equal age parameters. The simplest approach certainly is to average the individually fitted age parameters for males and females and to assume the resulting parameter averages for both genders in the future. This approach is only valid if the structures in the age parameters for both genders are rather similar which is typically the case. Otherwise, one would observe a significant structural break between historical and projected mortality improvements. In our example, we can and do proceed by using this simple 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 individually estimated age parameters. Edwards and Tuljapurkar (2005) show that the distributions of deaths often differ between countries. The authors 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 generally acceptable even for populations in closely related countries. We therefore refrain from adjusting the age parameters based on cross-country information in our example. The cohort parameters describe only temporary effects. Even if they differ significantly for two populations, mortality rates will not automatically diverge in the long run. MacMinn and Weber (2011) 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. Nevertheless, there might be reason for adjusting the cohort parameters in some cases, but for our example, we stick to the cohort parameters which we estimated for each population individually.
With the age and cohort parameters fully specified for the projection, we turn to forecasting future period parameters in the following section.

4 Projection of Period Parameters

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. Here, our projection methodology benefits from large flexibility in forecasting these parameters. Since there are no constraints on these parameters a priori, they could follow any pattern derived from basically any forecasting approach. In particular, this flexibility allows for coherent forecasting of the period parameters for several populations.
The most obvious projection approach is an extrapolation of the historical period parameters. However, these parameters fluctuate quite strongly in general which makes trend identification difficult. Moreover, it is not clear how to handle trends in the historical period parameters at all. A clearly visible and long lasting trend as in Figure 2 must be taken into account somehow when projecting but it is questionable whether such a trend can persist in the long run; mortality rates would decrease more and more rapidly every year without any limit on the mortality improvements. If one wanted to introduce such a limit for the long-term improvements, it would not be clear when and at which level this limit should come into effect.
For our projection framework, we therefore propose an alternative approach that is based on period life expectancy extrapolations. Once such an extrapolation is provided, the period parameter for each year in the future can be determined such that the forecast period life expectancy in that particular year is attained.[11] This approach has the following advantages:

- Life expectancies typically exhibit stable patterns since they are aggregated mortality statistics.

[11] Note that this approach does not imply that the parameter modifications in Subsection 3.2 impact the forecast mortality improvements. If one did not modify the parameters and projected the time trend in the original cohort parameters instead, the period parameters derived from the life expectancy extrapolation would be different, but the projected mortality improvements would be the same.

Often, they also follow clear trends which makes forecasting rather obvious. Moreover, it is typically easier to assess the plausibility and prudence of a certain life expectancy forecast than to judge whether a mortality improvement of, say, 2% for some age and year in the future is adequate.

- The forecast life expectancies can be the period life expectancies at birth or at any other age. Thus, period parameters can be derived with a focus on certain age groups, e.g. retirement ages in case the period life expectancy at, say, age 65 is forecast.
- Coherent mortality projections for several populations can be derived easily based on coherent life expectancy forecasts. We will provide an example for this later in this section and show how taking into account information from other countries can impact the projection for an individual population.
- The derivation of different projection scenarios is straightforward as only different life expectancy extrapolations need to be provided. This can be particularly helpful when building mortality projections with and without margins or when specifying a mortality/longevity stress scenario, e.g., for a (partial) internal model under Solvency II. Larger life expectancy gains imply larger period parameters, and thus, mortality improvements for all ages are increased.

Obviously, there cannot be a purely data driven standard procedure for deriving life expectancy extrapolations. Several issues need to be taken into account, e.g. the number of populations under consideration and observable patterns in the historical mortality evolution. Thus, a crucial user input is required here which certainly involves some amount of expert judgment. However, a thorough analysis of historical life expectancies and relationships between closely related populations can significantly limit the implied subjectivity. In order to illustrate this, we continue our example and show how future period parameters can be derived coherently for German males and females. Most of the recently proposed methods for coherently projecting life expectancies follow the idea of forecasting the worldwide maximum life expectancy and a life expectancy gap for each population under consideration. The worldwide maximum life expectancy has exhibited a surprisingly linear pattern for more than 160 years (see Oeppen and Vaupel (2002)) which makes extrapolation straightforward and this approach very tempting.[12] Models for the gaps between worldwide maximum life expectancy and life expectancies of particular populations have been proposed by, amongst others, Andreev and Vaupel (2006), Lee (2006), and Torri and Vaupel (2012), and, in principle, any of these models could be applied here. However, we do not follow the approach of forecasting the worldwide maximum life expectancy since it implies an inconsistency with the historical life expectancy evolution in Europe. Worldwide maximum life expectancies for males and females have diverged over the last two centuries which implies forecasting of a widening gender gap also for the future (see, e.g., Torri and Vaupel (2012)). In Europe, however, this gender gap has narrowed continuously and increasingly fast, from 6.8 years in 1980 to 5.3 years in 2008 (see also Figure 10 in Appendix B). Therefore, assuming a sudden increase of the gender gap in the future is counterintuitive.

[12] There is an extensive literature on the question whether life expectancies 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)).

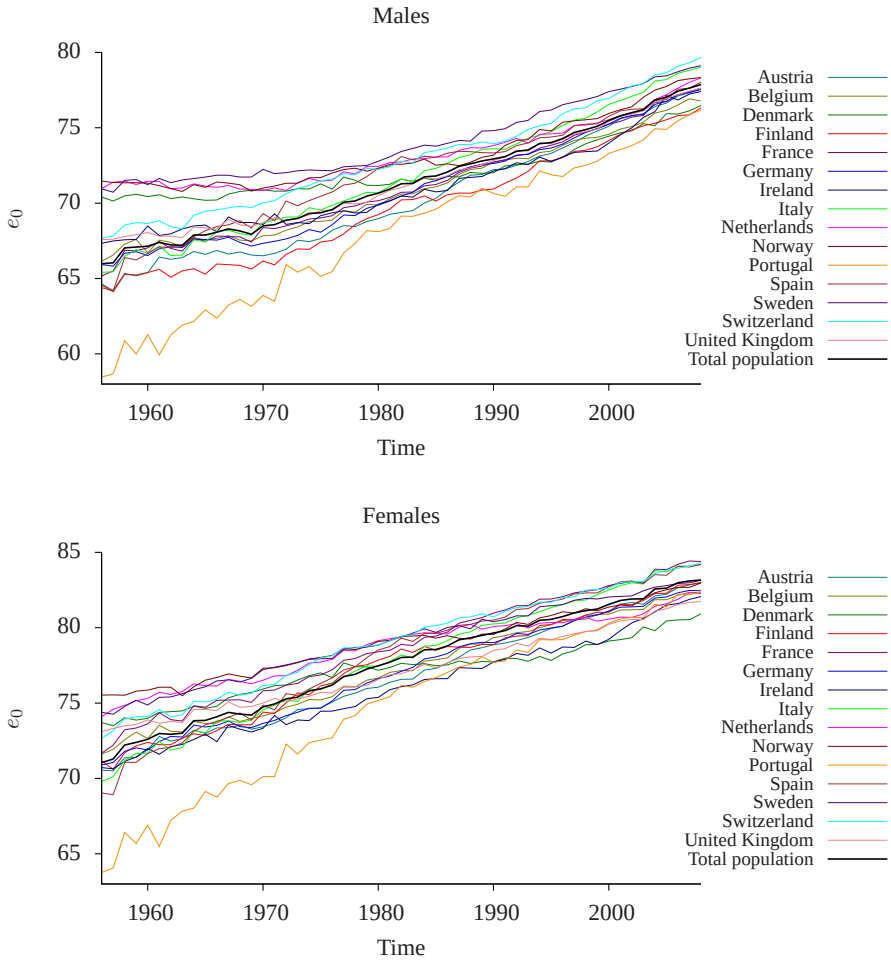

Figure 4: Historical life expectancies at birth in Europe

As can be seen in Figure 4, life expectancies in European countries have converged for both genders over the last decades. Moreover, we observe a common trend between countries, and clearly, life expectancy forecasts for any country should relate to this common trend. It can be plausible to assume above- or below-average life expectancies for one or both genders in a certain country, but the gaps between the country's life expectancies and the European average life expectancies should stay in a reasonable range. Therefore, we proceed by extrapolating the European life expectancy trend – coherently for males and females – and then analyze and forecast the life expectancy gap for Germany and some other countries. Figure 5 shows life expectancy forecasts for the male and female total populations in Europe which we regard as coherent between genders. The assumptions underlying these forecasts are explained in Appendix B. In brief, we assume that the gender gap will continue to shrink for the next decades and then level off in the long run. The long-term increase in life expectancy is fixed according to the trend in the historical data.

With life expectancy forecasts for the European populations at hand, we turn to possible deviations from these forecasts for some selected countries. In Figure 4, we observed convergence in life expectancies for males and females across Europe which indicates that best estimate life expectancies might be

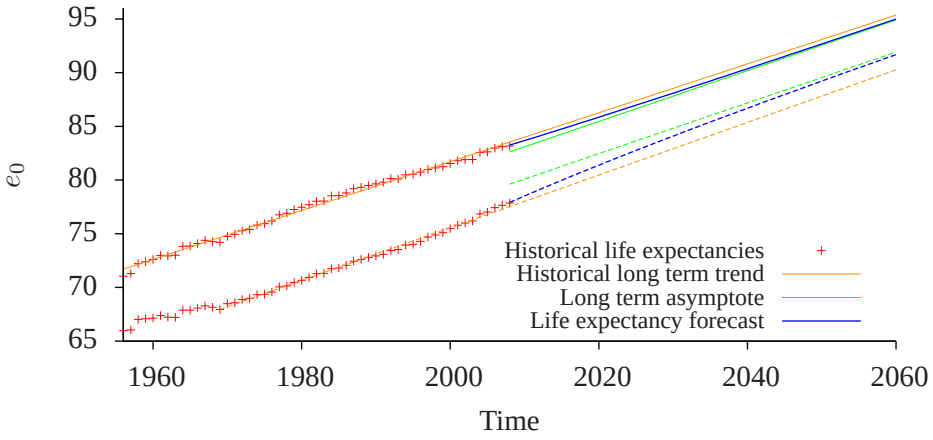

Figure 5: Coherent extrapolations of life expectancies for female (solid) and male (dashed) total populations in Europe

equal for all countries in the long run (see also Jarner and Kryger (2011) and references therein for this assumption). 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 6 shows how life expectancies in 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 may 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, the data suggests assuming a sustainable difference of about 1.5 years and introducing a smooth 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 prudent 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. In Denmark, on the other hand, the life expectancy increase has been 5 to 6 years lower than the European average increase. Here, we see how valuable coherent projections can be. Forecasting of life expectancies according to individual 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 from the European average 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

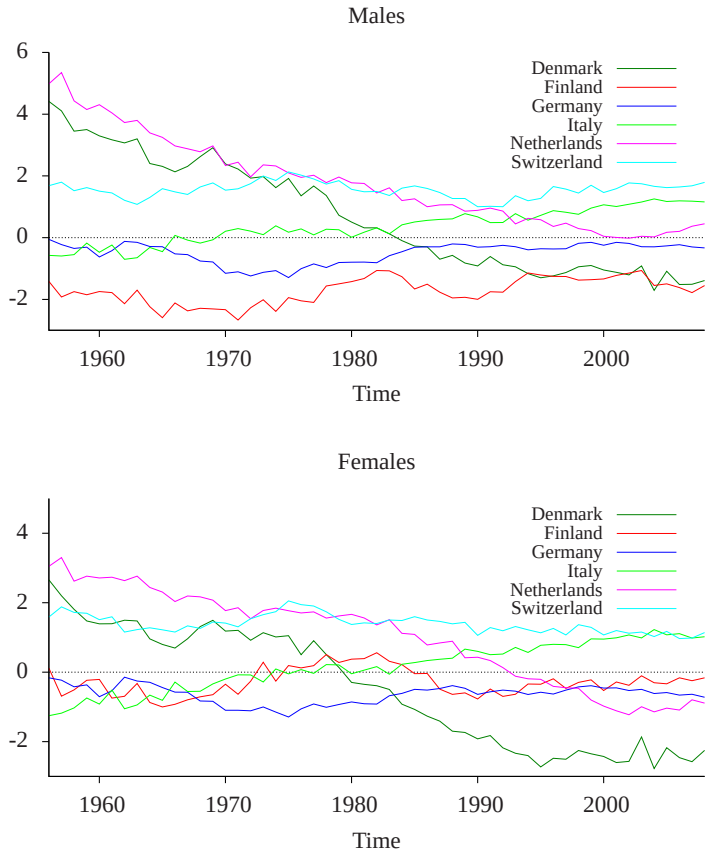

Figure 6: Differences between life expectancies of the total populations and populations in selected countries

years, this trend seems to have bottomed out though – slightly above the European average for males and about one year below average for females. Thus, assuming sustainable differences at these levels, a long-term gender gap of three years in Europe would imply a long-term gap of slightly less than two years between Dutch males and females. This can well be possible but 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 6 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 downward according to the observed deviations. We therefore fit the future period parameters to thus shifted life expectancies and obtain coherent projections as plotted in Figure 7.[13] The historical data is smoothed using either P-splines or our methodology. In the latter case, the charts also contain ages beyond 100.

We observe that our methodology smoothes the data more strongly than the P-spline method does. This then obviously leads to a slight break between the historical and the projected data in the P-spline case.

[13]Note that coherence is only achieved at an aggregate level; forecasts of individual mortality rates might not be fully coherent between genders. However, full coherence is hardly achievable in general, and slight incoherences for individual mortality rates should average out between different ages and/or periods.

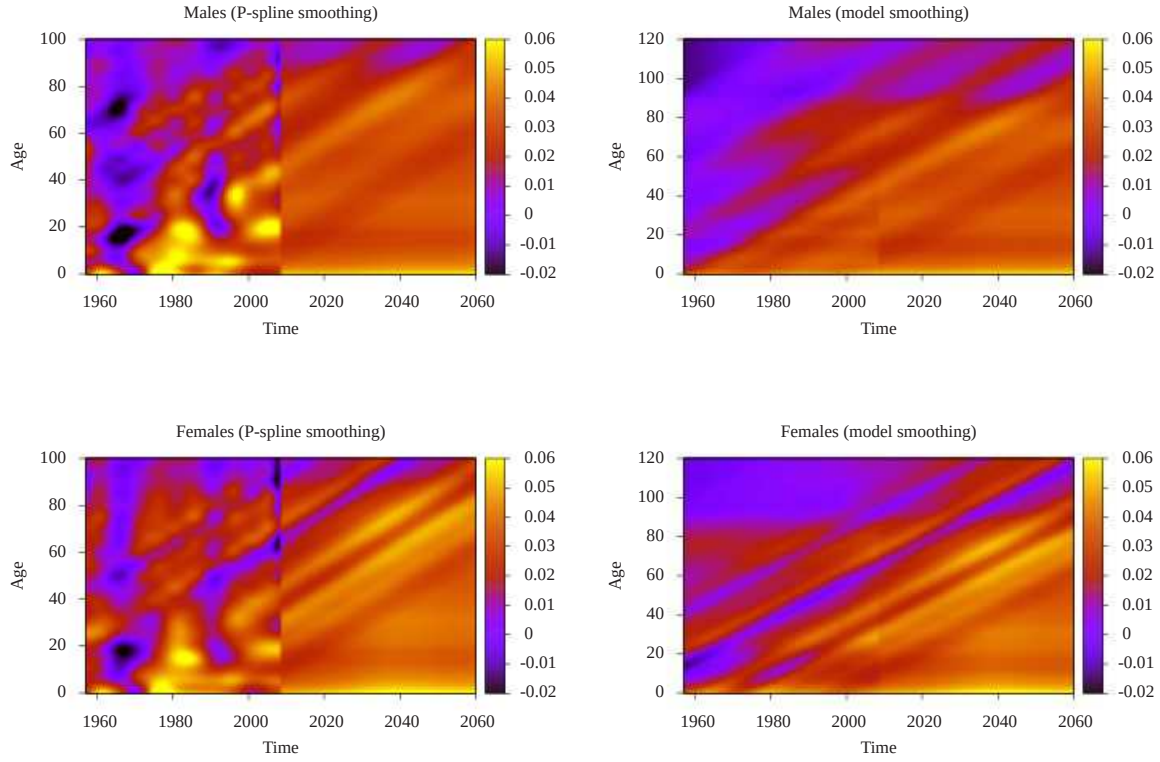

Figure 7: Coherent mortality projections for German males and females

In general, it is difficult to tell which level of smoothing is most appropriate, but we have provided reasoning for our smoothing in Subsection 3.3. More importantly, however, the general structure in the historical data is the same for both smoothing methods, and our projection extrapolates this structure nicely. Therefore, we think our projection looks highly plausible for both males and females. In particular, all cohort dependent structures are carried forward appropriately. The very slight breaks in 2009 in the graphs on the right hand side are due to the use of average age parameters in the projections.

5 Comparison with other Projection Models and Back Test

In the Introduction, we referred to the current standard projection for private annuities in Germany. In order to compare that projection to the projection which we exemplarily derived in this paper, Table 3 provides life annuity present values for males and different ages and deferment periods. For the annuity present values based on the DAV 2004 R mortality table, also percentage deviations from the annuity present values according to the new projection are provided.[14] The most striking observation is that the new projection consistently yields the largest annuity present values. This observation does not only hold for a comparison with the DAV 2004 R best estimate projection but also for a comparison with

[14]The present values are computed based on a time constant interest rate of 1.75% which is the maximum admissible interest rate for annuity reserving in Germany at the time or writing. The base mortality rates to which the projections are applied are equal in all cases, i.e. observed mortality rates in 2008.

Deferment period	Age at first payment	New projection	DAV 2004 R best estimate		DAV 2004 R incl. margins	
0 years	65	16.84	15.88	-5.7%	16.45	-2.3%
	75	10.61	10.16	-4.2%	10.42	-1.8%
	85	5.69	5.61	-1.4%	5.68	-0.1%
20 years	65	12.83	10.81	-15.8%	11.97	-6.7%
	75	7.14	6.03	-15.5%	6.87	-3.9%
	85	2.73	2.03	-25.7%	2.44	-10.6%
40 years	65	11.01	8.23	-25.2%	9.61	-12.7%
	75	7.02	4.73	-32.5%	5.90	-15.9%
	85	2.93	1.64	-44.1%	2.35	-19.9%

Table 3: Annuity present values and percentage deviations for males according to the newly constructed projection and the projection in the mortality table DAV 2004 R

the projection with margins. The differences in the annuity present values increase with the deferment period, i.e. with the time period over which the projection is applied. A discussion of the appropriateness of existing tables is not within the scope of this paper though since we focus on developing a new projection methodology here and have derived the new projection only for illustrative purposes.

In order to further illustrate the applicability of our projection methodology, we perform a back test and benchmark it to other commonly used projection models. In the back test, we only use data up to 1990 for model estimation and then project mortality improvements up to 2008, i.e. the last year for which data is available.

The age and cohort parameters in our model are derived as in the example on the full data set. The life expectancy extrapolations have to be adjusted slightly according to the historical life expectancy evolution up to 1990. Since the gender gap in European life expectancies has remained rather constant in the 1980's (see also Figure 10 in Appendix B), a parallel extrapolation of the historical life expectancies would have been a plausible and coherent forecast. The slope of these extrapolations is chosen as the average of the slopes in the historical trends for males and females. The life expectancy forecasts for Germany are derived by a downward shift of the forecasts for the European populations. Between 1985 and 1990, the gap between European and German life expectancy has been rather constant at about 0.5 for males and 0.3 for females (see Figure 6).

As alternative projection models, we consider the Lee-Carter model with cohort effects as proposed by Renshaw and Haberman (2006) and the P-spline model by Currie et al. (2004). We have chosen these models because they are widely accepted and applied in practice, because they project mortality improvements dependent on age, period, and cohort, and because they cover the full age range. These are requirements which only few of the existing mortality models comply with. For both models, we apply standard estimation and projection techniques and refer to Appendix C for details. However, we apply two variants of each model: In case of the Renshaw-Haberman model, we once use the parameter values as fitted to the historical data, and we smooth them using the Whittaker-Henderson method. In case of the P-spline model, we consider two different fits, one to data up to 1990 and one to data up to 1985 only. The reason for using two different data periods becomes clear from Figure 8.

The figure shows actual and projected cumulative mortality improvements from 1990 to 2008, i.e. $1 - q_{x,2008}/q_{x,1990}$, for ages up to 100, both genders, and the three projection methods in their respective variants. The most striking observation is that the P-spline model fitted to data up to 1990 yields highly implausible projections for both genders. For most ages, (significant) mortality deterioration is assumed which contradicts the general historical trend of decreasing mortality with time. This implausible projection is due to the P-spline model being very sensitive to the mortality trend in the last few data points. In Figure 1, we see that, at the end of the 1980's, mortality improvements for males were negative for many ages, in particular for those around age 30, and this is carried over into the projection. When these last data points are omitted, i.e. when the model is fitted to data up to 1985 only, the projections look more plausible, but generally overestimate mortality improvements. The sensitivity to only very few data points and the resulting instability of the projections question the general applicability of the P-spline model.

The Renshaw-Haberman model and our methodology provide more stable projections for both genders which, at the same time, match the actual mortality improvements significantly better in general. However, the Renshaw-Haberman variant with raw model parameters can yield questionable patterns at times, e.g. the rather extreme dent around age 30 for males.

Table 4 shows some statistics which measure the projection methods' performances. The average error is the average difference between actual and projected improvements. Thus, a positive error means that the projected improvements are, on average, too large, a negative error means that the projected improvements are too small. The root mean square (RMS) error is a measure for how well a projection method performs in matching the actual mortality improvement for each age, i.e. how similar the age structures of actual and projected improvements are. We observe that, in comparison to the other projection methods, the P-spline model variants perform poorly in terms of both average error and root mean square error. Comparing the other two methods, we see that, for females, the average errors for the Renshaw-Haberman model variants are smaller but that the root mean square errors are similar. Thus, the Renshaw-Haberman projections are closer to the average of the actual improvements, but the errors in the projection for each age are about the same for both methods. For males, our methodology performs better in terms of both average error and root mean square error. In Figure 8, we see that the Renshaw-Haberman model significantly underestimates improvements for ages 60 to 90 in particular which are the most relevant ages with respect to longevity risk. Therefore, we conclude that our methodology outperforms the alternative models in this example and that it provides a valuable alternative to existing projection models.

6 Modeling Uncertainties and Margins

The modeling and forecasting of mortality always involves a considerable amount of uncertainty. Therefore, often projections with margins are required for prudent calculations of premiums and reserves. In this section, we highlight potential sources of uncertainty and show how uncertainty can be accounted for within our methodology. Moreover, we explain how basis risk can be assessed in case a projection is applied to a population which it has not been constructed for originally.

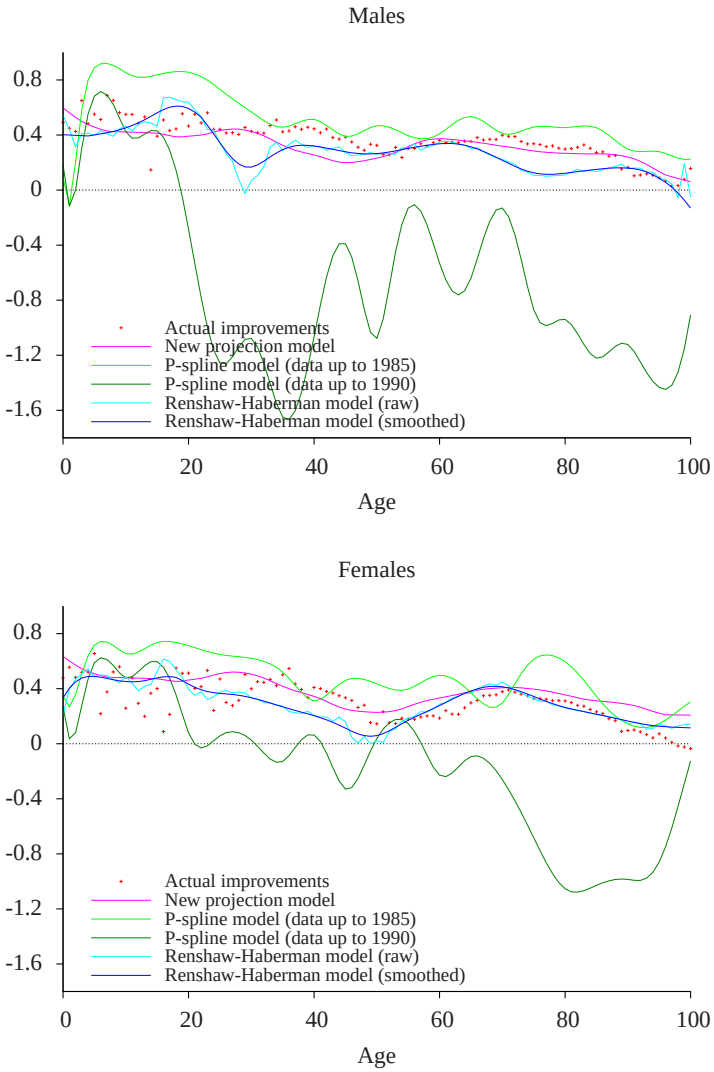

Figure 8: Cumulative mortality improvements from 1990 to 2008 for different projection methods

The largest uncertainty arises from the fact that future mortality might evolve differently from historical patterns. For our projection methodology, we see three main points where this risk of changes can materialize:

- The long-term trend in the life expectancy evolution (in our example: the long-term trend for the total population);
- The long-term relationship between the life expectancies of different populations (in our example: the long-term gender gap and the difference between life expectancies for the total population and the German population);
- The age pattern of mortality improvements.

All three issues can be accounted for by adjusting the life expectancy extrapolations, e.g. by increasing the slope of the long-term life expectancy trend. This implies a margin which increases with time, thus

Projection method	Females		Males	
	Average error	**RMS error**	**Average error**	**RMS error**
New methodology	0.074	0.126	-0.043	0.099
Renshaw-Haberman (raw)	-0.013	0.140	-0.081	0.150
Renshaw-Haberman (smoothed)	-0.013	0.127	-0.082	0.139
P-splines (data up to 1985)	0.170	0.233	0.149	0.215
P-splines (data up to 1990)	-0.485	0.660	-1.004	1.157

Table 4: Average errors and root mean square (RMS) errors for different projection methods

matching the structure of the uncertainty it is to account for. For the next years, one is generally well informed about the forthcoming mortality evolution (as long as no mortality/longevity shock occurs), but in the long run, uncertainty becomes considerable. At the same time, an increase in projected life expectancies and thus in the period parameters would imply a margin which is evenly spread over all ages. Therefore, this approach can also account for the uncertainty regarding a change in the age pattern of mortality improvement. In case one is particularly concerned about the evolution for certain ages or cohorts, the respective parameters could also be adjusted individually. Note though that the order of adjusting those parameters and estimating the future period parameters is crucial since the period parameters react to and, to some extent, compensate for increases in the other parameters.

Compared to the risk of significant changes in the future mortality evolution, the parameter uncertainty in the model estimation seems rather small. If one nevertheless wants to account for parameter uncertainty in the age and cohort parameters, confidence bounds for these parameters can either be derived analytically (depending on the estimation approach under consideration) or by bootstrapping. 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 analogously in our setting. The parameter uncertainty in the future period parameters typically stems from potential misestimation of long-term life expectancy trends. Here, confidence bounds for the regression parameters can be derived analytically.

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 which is different from the one the projection has originally been constructed for. In our example, we derived projections for German males and females which may or may not be applicable to, e.g., the particular population of a pension fund. In most cases, however, basis risk is limited in the long run since the mortality evolutions of some wider reference population, e.g. the general population, and a subpopulation, e.g. the population of a pension fund, should not diverge until infinity. Over the next years, mortality improvements might differ though. If no data is available for the subpopulation, it is virtually impossible 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 for 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 larger reference population, but the average level and possible trends of both period parameter sets should be very similar. Significant differences, on the other hand, would

indicate the potential need for an adjustment to the projection according to the observed differences. If one questions the adequacy of the age parameters or the cohort parameters, basis risk in these parameters could be measured and accounted for analogously.
The highlighted uncertainties illustrate the general difficulties in forecasting future mortality experience – not only in our methodology but in any modeling framework. Margins can help mitigating these uncertainties, but the most effective approach certainly is to update mortality projections on a regular basis.

7 Conclusion

Projections of future mortality evolutions are particularly necessary for the computation of reserves and risk management in the insurance and pension business as well as 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 for mortality improvements based on the Age-Period-Cohort (APC) model and present a general framework for parameter estimation and forecasting. We propose different approaches for fitting the APC model to historical mortality data and discuss their advantages and applicability in certain situations.
A feature which distinguishes our methodology from most other projection methods is that the future period parameters are not necessarily forecast based on the evolution of the historical period parameters. Instead, we propose using life expectancy extrapolations and deriving the period parameters such that the forecast life expectancies are obtained. Life expectancies typically evolve fairly stable since they are aggregate mortality statistics, and they often exhibit rather obvious patterns which can be easily extrapolated. Moreover, this approach allows for high flexibility in the projection which can be utilized to derive coherent forecasts for several populations. As we have shown, the simultaneous consideration of several populations can in fact have a significant impact on the projection for each individual population. Therefore, a model for projecting best estimate mortality should always take into account information provided by data from other closely related populations. The approach of forecasting life expectancies also provides an intuitive way to include margins into a projection by simply increasing the slope in the life expectancy trend. In that case, margins are spread evenly over all ages, and they increase with time which is in line with the structure of the uncertainty about the future mortality evolution.
The proposed projection methodology provides a valuable alternative to other projection methods, in particular statistical extrapolation models. The latter typically have rather rigid model structures which can limit their applicability in cases where these structures do not match the historical and/or expected mortality patterns. In contrast, the methodology proposed in this paper offers a high degree of flexibility and a way to combine data driven extrapolations with expert judgment. Moreover, in comparison to stochastic mortality models, unnecessary complexity and restrictions which are related to the stochastic simulation are omitted, thus offering a clear focus on the best estimate forecast. On the other hand, the application of our methodology is not straightforward in any case, and it requires some case specific assumptions from the user. Thus, our methodology is not suitable for repeated and quick re-estimation,

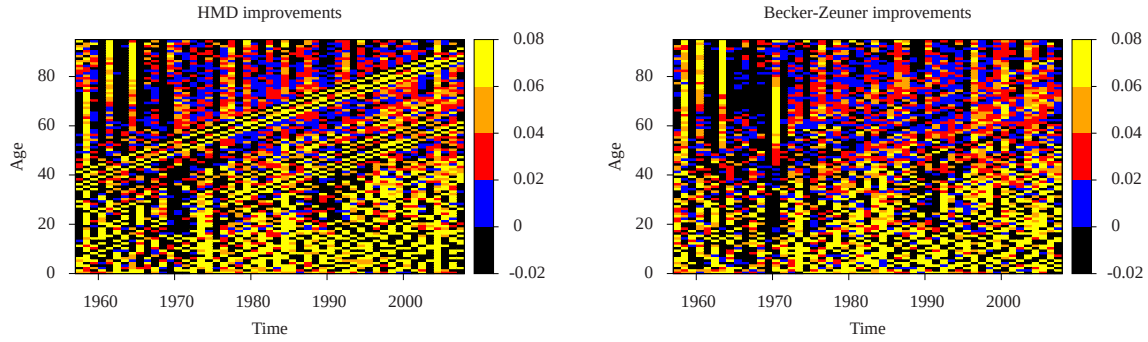

Figure 9: Raw mortality improvements based on HMD data and the Becker-Zeuner method

e.g. within simulations. However, this is hardly an issue when best estimate mortality projections are updated once in a while in practice.

In order to illustrate our methodology, we have derived projections for German males and females as part of a larger European reference population. A back test and comparison with commonly used projection models shows that our methodology provides highly plausible forecasts and confirms that it constitutes a valuable alternative to existing projection approaches.

Appendix

A Data

Throughout this paper, we use mortality data from the Human Mortality Database (2012, HMD) in order to illustrate our projection methodology. However, we do not use the mortality rates provided by the HMD but, instead, derive mortality rates from the also provided deaths counts and population sizes by applying a different methodology. The reason is that – at least for the case of West Germany – the HMD methodology implies some implausible cohort features (cf. Hiester et al. (2012)). The left panel of Figure 9 shows the raw mortality improvements for West German males as derived from the mortality rates of the HMD. The clear diagonal patterns are not surprising at first sight as we can see obvious cohort effects also in Figure 1. However, the repeated switches from large mortality improvements (yellow) to strong mortality deterioration (black) and vice versa from one cohort to the next look ominous and are hardly justifiable by demographic intuition. In fact, they are probably only an artifact of an assumption made when deriving the mortality rates. A uniform distribution of birthdays within each calender year is assumed which is typically not the case. This assumption is uncritical as long as successive cohorts are rather equal in size. For West Germany, however, this is not the case for cohorts which were born during or shortly after the world wars. Here, cohort sizes sometimes change significantly from one birth year to the next.

A method for deriving mortality rates which overcomes this critical assumption of birthdays being uni-

formly distributed is the method of Becker-Zeuner. While the HMD methodology is based on the number of deaths at a certain age in a certain calender year, the Becker-Zeuner method considers the number of deaths at a certain age for a certain cohort. For further details on this "cohort mortality method", we refer to standard textbooks. The right panel of Figure 9 shows mortality improvements which are derived from Becker-Zeuner mortality rates. The rapid switches from strong mortality improvements to mortality deterioration have disappeared, and cohort effects are hardly visible anymore. However, that does not mean that cohort effects do not exist. As we see in Figures 1 and 7, cohort effects are still present and can easily be detected in the smoothed data. However, they are somewhat obscured by random fluctuations in the raw data.

Finally, we should note that the methodology for deriving mortality rates only has a minor effect on the results presented in this paper. When projecting mortality improvements, we smooth the model parameters to eliminate random fluctuations. At that stage, also the cohort features in the HMD mortality rates would get smoothed out, resulting in mortality projections which are very similar to those derived from the Becker-Zeuner mortality rates.

B Life Expectancy Forecasts for the European Total Populations

In Section 4, we project the period parameters for German males and females in coherence with mortality evolutions in other European countries. These projections are based on life expectancy forecasts for the male and female total populations in Europe as plotted in Figure 5. In order to derive these forecasts, a couple of assumptions had to be made. Note that these assumptions particularly relate to our specific example and are not meant to be most suitable in any case. Nevertheless, the general approach described here should be valid for most populations.

First of all, we assume a long-term difference in life expectancies between males and females of $\Delta = 3$ years. The gender difference has been shrinking since 1980 in Europe, with a significant acceleration starting in the mid-1990's (cf. the red curve in Figure 10). This shrinkage has often been explained by convergence in lifestyles of males and females. 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. Luy (2002) also comes to the conclusion that the life expectancy gap must be mostly due to differences in lifestyles. He finds that the gap between life expectancies of nuns and monks, who live under very similar socio-economic conditions, is only about one year in young adult ages. Therefore, it is reasonable to forecast a further shrinkage of the gender gap in Europe in the short- to mid-term future. Since assuming a long-term gender gap of only one year seems rather bold from our point of view, we apply a gap value which lies in the middle of the current difference and this one year, i.e. 3 years. Obviously, this choice is somewhat subjective, and other choices may also be reasonable.

In order to obtain a constant gender gap in the long run, a common trend for the long-term life expectancy increase for both genders is required. This trend should reasonably extrapolate historical trends, and we therefore fix the slope s of this trend according to the average of the slopes of the long-term historical

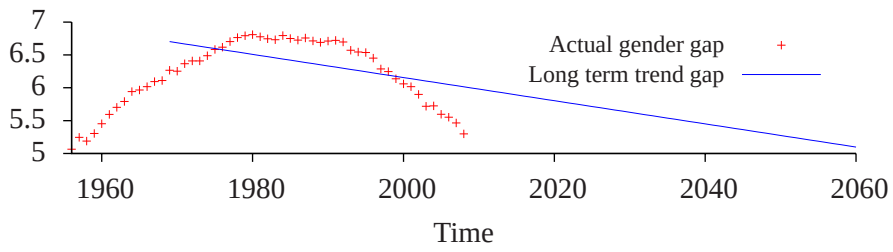

Figure 10: Differences in life expectancies between males and females in Europe

trends for males and females. In Figure 4, we see that, for males, a linear trend seems to start off in 1969, with a slope of 0.2473; for females, we observe a rather linear trend for the whole data set, with a slope of 0.2296. This implies a slope for the common long-term trend of $s = 0.2385$.

This slope is as consistent as possible with 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 long-term divergence observed by Oeppen and Vaupel (2002), a simultaneous full coherence with their extrapolations for both genders is unachievable per se.

Instead of explicitly specifying a long-term gender difference, we have also considered the rather simple case of life expectancy extrapolations for both genders according to the slopes in their respective historical data. Since life expectancies have increased stronger for males than for females, this approach also implies some (slight) convergence between genders. However, this approach does not yield plausible results. The blue line in Figure 10 shows the resulting gender differences, and we observe that the actual life expectancy gap at the end of the historical data set, i.e. in 2008, would already be misestimated by about 0.7 years.

We also see in that figure that the historical life expectancy differences have fluctuated only very little but started to shrink significantly in the mid-1990's. Thus, the life expectancy trends for males and/or females must have also changed at that time. In order to guarantee a smooth transition from historical to projected life expectancies, in the short run, we extrapolate the trends in the historical data starting in 1995. The corresponding slopes are 0.3052 for males (in comparison to 0.2473 in the long-term trend) and 0.2099 for females (in comparison to 0.2296).

Finally, we reconcile the short and long-term assumptions by assuming 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.

$$le_m(t) = d_m + s(t - 2008) - \exp\{g_m(t - 2008) + h_m\}$$

and

$$le_f(t) = (d_m + \Delta) + s(t - 2008) + \exp\{g_f(t - 2008) + h_f\},$$

Parameters	d_m	Δ	s	$g_m = g_f$	h_m	h_f
	79.6221	3.0	0.2363	-0.0388	-0.4606	0.5350

Table 5: Parameter values for coherent life expectancy extrapolations for European males and females

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. Moreover, as we want both life expectancy curves to converge to their asymptotes equally fast, we require the slope parameters in the exponential terms, g_m and g_f, to coincide. These specifications and constraints leave us with a set of uniquely identifiable parameters whose values are summarized in Table 5.

C Benchmark Models

For the back test in Section 5, we consider two alternative projection models, i.e. the Lee-Carter extension of Renshaw and Haberman (2006) and the P-spline model of Currie et al. (2004). In the Renshaw-Haberman model, one year log mortality rates are modeled as

$$\log m_{x,t} = \alpha_x + \beta_x^{(1)} \cdot \kappa_t + \beta_x^{(2)} \cdot \gamma_{t-x} + \epsilon_{x,t},$$

where α_x, $\beta_x^{(1)}$, and $\beta_x^{(2)}$ are age dependent parameters, κ_t describes the mortality evolution over time, γ_{t-x} accounts for cohort effects, and $\epsilon_{x,t}$ is random noise with mean zero. This model is typically estimated via maximum likelihood; given population sizes as exposures, the numbers of deaths for each age and calender year are assumed to be independently Poisson distributed. For further details on model estimation, we refer to, e.g., Cairns et al. (2009). The time trend κ_t is typically projected as random walk with drift, and we follow the same approach. However, since we are only interested in a deterministic best estimate projection, we only consider the central trajectory of this random walk with drift. The question which process is applicable for projecting γ_{t-x} strongly depends on the data set under consideration. Often, a mean reverting AR(1) process is applied. For simplicity, we set the parameters for new cohorts equal to the parameter of the last cohort in the historical data. Since we only project the central trajectory for just 18 years, the effect of new cohorts in our back test is negligible anyway.

In the model of Currie et al. (2004), B-splines are fitted to the surface of log mortality rates via maximum likelihood. In order to eliminate random fluctuations, these splines are penalized by subtracting a penalty function from the likelihood function. Thus, the better the fit of the splines to the noisy data, i.e. the larger the likelihood function, the larger the penalty. The trade-off between goodness of fit and smoothness is typically solved by optimizing some information criterion. We use the Bayesian Information Criterion here, as proposed by Currie et al. (2004). Future mortality rates are then projected by treating them as missing data points. In that case, the likelihood function is zero, and the splines are calibrated such that the penalty is minimized. For further details on model estimation and projection, we refer to Currie et al. (2004). In our back test, we use standard model parameters, i.e. cubic splines, a distance of five data

points between the knots of adjacent splines, and a quadratic penalty function which implies a linear forecast of the most recent historical mortality trend.

References

Andreev, K., Vaupel, J., 2006. Forecasts of cohort mortality after age 50. Working Paper, Max Planck Institute for Demographic Research, Rostock.

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., 2008. Modelling and Management of Mortality Risk: A Review. Scandinavian Actuarial Journal, 2: 79–113.

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., 2011. Bayesian Stochastic Mortality Modelling for Two Populations. ASTIN Bulletin, 41: 29–59.

Carstensen, B., 2007. Age-period-cohort models for the Lexis diagram. Statistics in Medicine, 26: 3018–3045.

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

Craven, P., Wahba, G., 1978. Smoothing noisy data with spline functions. Numerische Mathematik, 31: 377–403.

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

Delwarde, A., Denuit, M., Eilers, P., 2007. Smoothing the Lee-Carter and Poisson log-bilinear models for mortality forecasting: A penalized log-likelihood approach. Statistical Modelling, 7: 29–48.

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.

Garcia, D., 2010. Robust smoothing of gridded data in one and higher dimensions with missing values. Computational Statistics and Data Analysis, 54: 1167–1178.

Haberman, S., Renshaw, A., 2012. Parametric mortality improvement rate modelling and projecting. Insurance: Mathematics and Economics, 50: 309–333.

Henderson, R., 1924. A new method of graduation. Transactions of the Actuarial Society of America, 25: 29–40.

Henderson, R., 1925. Further remarks on graduation. Transactions of the Actuarial Society of America, 26: 52–57.

Hiester, T., Köcher, J., Mümken, B., Winter, J., 2012. Comment on: Coherent Projections of Age, Period, and Cohort Dependent Mortality Improvements, Preliminary version of November 2011. Working Paper, Allianz Lebensversicherungs-AG, Stuttgart.

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

Human Mortality Database, 2012. University of California, Berkeley, USA, and Max Planck Institute for Demographic Research, Germany. Available at: www.mortality.org.

Hyndman, R., Booth, H., Yasmeen, F., 2012. Coherent mortality forecasting: the product-ratio method with functional time series models. Demography, 50: 261–283.

Jarner, S., Kryger, E., 2011. Modelling Adult Mortality in Small Populations: The Saint Model. ASTIN Bulletin, 41: 377–418.

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.

Kuang, D., Nielsen, B., Nielsen, J., 2008a. Identification of the age-period-cohort model and the extended chain-ladder model. Biometrika, 95: 979–986.

Kuang, D., Nielsen, B., Nielsen, J., 2008b. Forecasting with the age-period-cohort model and the extended chain-ladder model. Biometrika, 95: 987–991.

Kunde, S., Ortmann, K., 2011. Zur Differenzierung der Mortalität nach geodemographischen und sozioökonomischen Merkmalen. Der Aktuar, 17: 173–176.

Lee, R., 2006. Mortality forecasts and linear life expectancy trends. In Bengtsson, T., Ed.: Perspectives on mortality forecasting. III. The linear rise in life expectancy: history and prospects. Swedish Social Insurance Agency, Stockholm.

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

Lee, R., Miller, T., 2001. 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., 2011. Select Birth Cohorts. European Actuarial Journal, 1: 395–409.

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.

Plat, R., 2011. One-year Value-at-Risk for longevity and mortality. Insurance: Mathematics and Economics, 49: 462–470.

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.

Torri, T., Vaupel, J., 2012. Forecasting life expectancy in an international context. International Journal of Forecasting, 28: 519–331.

Whittaker, E., 1923. On a new method of graduation. Proceedings of the Edinburgh Mathematical Society, 41: 63–75.