The Future of Mortality: Mortality Forecasting by Extrapolation of Deaths Curve Evolution Patterns

Matthias Börger* Martin Genz† Jochen Ruß‡

Abstract

A variety of mortality models can be used to project future mortality. However, the parameters of most of these models lack a clear demographic interpretation. Hence, it may be hard to identify forecasts which are not consistent to the most recent observed trends in the mortality evolution or that are demographically implausible. On the other hand, demographers make predictions on future mortality but typically focus on single aspects only instead of comprehensive mortality forecasts. This article aims to close the gap between these forecasting approaches.

We establish a new best estimate mortality model which is based on the extrapolation of four statistics that have a clear demographic interpretation. The four statistics are taken from the classification framework of Börger et al. (2018). Our model yields forecasts for the deaths curve which are consistent with the most recent demographic trends in the deaths curve evolution. Moreover, expert opinions on future trends in the mortality evolution can easily be incorporated, and we illustrate how the model can be used for scenario analyses.

Keywords: Mortality scenario classification, Demographic mortality trends, Demographic scenario analyses, (consistent) Mortality forecasting

*Institut für Finanz- und Aktuarwissenschaften (ifa), Lise-Meitner-Straße 14, 89081 Ulm, Germany, email: m.boerger@ifa-ulm.de

†Institut für Finanz- und Aktuarwissenschaften (ifa) and Institut für Versicherungswissenschaften, Universität Ulm, Lise-Meitner-Straße 14, 89081 Ulm, Germany, email: m.genz@ifa-ulm.de (corresponding author)

‡Institut für Finanz- und Aktuarwissenschaften (ifa) and Institut für Versicherungswissenschaften, Universität Ulm, Lise-Meitner-Straße 14, 89081 Ulm, Germany, email: j.russ@ifa-ulm.de

1 Introduction

Estimates for future mortality are important in many areas, e.g. for projections of social security systems or risk management in the private pension and life insurance sector. To derive such estimates, a variety of mortality models has been developed most of which are statistical models in the sense that they extract certain patterns from historical mortality data and extrapolate those patterns into the future.

Some of the earliest mortality models (which are also called "laws of mortality") are quite simple, as they are based on only few parameters. One of the first models in this spirit is the law of mortality by Gompertz (1825). This model has often been extended. For example, Pollard (1987) gives a comprehensive review on mortality models which have been developed from the 19th century until the 1980s. Such models have often been used for deterministic mortality forecasts. During the most recent decades, stochastic mortality models have emerged, e.g. the Lee-Carter model (Lee and Carter, 1992) or the Cairns-Blake-Dowd model (Cairns et al., 2006). An overview is given for example by Booth and Tickle (2008) or (with a more actuarial focus) in Cairns et al. (2008). While the majority of the early "laws of mortality" only take into account age-dependent patterns of mortality, some of the more recent models also allow for period and even cohort effects. This allows for model fitting to longer data series and in particular for more flexible (e.g. stochastic) extrapolations.

However, in many cases, the parameters in these models lack a clear demographic interpretation. Therefore, it is not easy for the user of the model to assess if an extrapolation of these parameters will result in plausible mortality forecasts from a demographic point of view. We will provide examples for that later on.

In the literature we also find many approaches for forecasting (certain aspects of) future mortality from a demographic point of view. For example, Manton et al. (1991) divide the perspectives on the future life expectancy in three groups - the "traditional", the "visionary", and the "empiric" perspective - and give examples for each perspective. Olshansky et al. (1990) take the "traditional" point of view by suggesting that there might be a limit to the increase in human longevity. More than one decade later, Oeppen and Vaupel (2002) analyzed the world record life expectancy and found a linearly increasing trend which they extrapolate several decades into the future. In the spirit of Manton et al. (1991) this is rather a "visionary" perspective. Sometimes demographic forecasts are also made implicitly (i.e. without an explicit extrapolation of any statistic). For example, Dong et al. (2016) "strongly suggest that the maximum lifespan of humans is fixed" and build their claim on historical trends in demographic statistics measuring the evolution of old-age mortality. In the same spirit, but more than three decades earlier Fries (1980) derived future limits for the maximum human lifespan based on demographic statistics. So demographers typically do not only use statistical observations (which are often the only input for the statistical mortality modeling), but also combine knowledge from biology, medicine, sociology, and other fields of research in their forecast. However, their forecasts often focus on

single aspects of the mortality evolution only, e.g. the life expectancy at birth or the maximum human lifespan, and they typically do not make suggestions for the entire distribution of deaths over all ages in any future year. The latter, however, is provided by mortality models (but typically without taking demographers' forecasts into account). Therefore, it seems worthwhile to combine the purely statistical approach of mortality modeling with the demographic expertise. This paper aims to close this gap. To this end, we propose a methodology that can derive the entire mortality distribution for any future point in time based on forecasts for a set of statistics that come with a clear demographic interpretation.

We build our methodology on recent work of Börger et al. (2018). They introduce a new classification framework for (observed historic) changes in mortality over time which is based on the deaths curve, i.e. the density function of the age at death. They show that four statistics, each with a clear demographic interpretation, describe the key structural elements of a deaths curve - its support and the pattern it exhibits on this support. Thus, they express changes in a deaths curve between two points in time by changes in these four statistics. In their classification framework, a uniquely defined "mortality scenario" is then assigned to every possible evolution of a deaths curve over time. Such a scenario consists of four components and each component has one of three specifications: It can 1) be "right-shifting" or "left-shifting" (or neutral in respect of "shifting mortality"), it can 2) exhibit "extension" or "compression" (or be neutral in this dimension), it can 3) show signs of "compression" or "decompression" (or be neutral in that respect), and it can 4) show "concentration" or "diffusion" (or again be neutral in that respect). In Section 2.1 we will explain this in more detail.

This classification framework allows us to check if any given mortality forecast is consistent with most recent demographic trends. A mortality forecast can be considered demographically reasonable in case the historical patterns in the four statistics are extrapolated in a demographically plausible manner.

The other way round, it can also serve as a basis to create mortality forecasts for the whole deaths curve that match a given forecast for the four statistics from the classification framework. Such a forecast can either be an extrapolation of the four statistics based on their past evolution or a forecast based on demographic insights (or a combination thereof, e.g. an extrapolation that is adjusted based on demographic insights). This results in a new type of mortality model which will be derived in this paper.

The remainder of this paper is organized as follows: In Section 2 we briefly summarize the classification framework of Börger et al. (2018) and test frequently used mortality forecasting models for consistency with the past in terms of their projection of the four key statistics. This illustrates the need for mortality forecasts with better consistency to recently observed mortality scenarios. Consequently, in Section 3 we introduce a deterministic mortality forecasting model, which is based on the four components of the aforementioned classification framework. This model uses forecasts for the four statistics (e.g. extrapolation of recent trends or some expert opinion) to project the entire deaths curve and thus allows for consistency of the mortality

forecasts with recently observed mortality scenarios and/or the experts' expectations. In Section 4 we discuss examples and potential applications for this model. Finally, Section 5 concludes.

2 Consistency of Existing Mortality Models

2.1 A Unique Classification Framework for Mortality Evolution Patterns

This section briefly summarizes the classification framework introduced by Börger et al. (2018). They uniquely assign a mortality scenario to any change of a deaths curve between two points in time. More concretely, a mortality scenario is defined as a four dimensional vector with the following components:

1. *Right-* or *left-shifting mortality* is linked to an increase or decrease in the *modal age at death* (M). This statistic is defined as the age in which most people die, i.e. the maximum of the deaths curve. Thus M can intuitively be interpreted as the position of the center of the deaths curve. So shifting mortality is linked to a shift of the deaths curve's center to the right or to the left, respectively.[1]

2. *Extension* or *contraction* is linked to an increase or decrease of the *upper bound of the deaths curve's support* (UB). This statistic is defined as the right endpoint of the deaths curve's support.[2] Intuitively it is the highest age that has been reached within the population of interest. Whenever this age moves to the right, the support of the deaths curve extends and whenever UB moves to the left the deaths curve's support contracts. Like the modal age at death, UB indicates changes in the deaths curve's position.

3. *Compression* or *decompression* is linked to an increase or a decrease of the *degree of inequality* (DoI). This statistic measures the size of the area between the observed deaths curve and a hypothetical deaths curve where the age at death is uniformly distributed. Intuitively this area (and hence the inequality of the distribution of deaths by age) will increase, if the observed deaths curve moves farther away from the uniformly distributed deaths curve. Consequently, there must be age segments, where the number of deaths increases, i.e. deaths become more compressed. If the observed deaths curve moves closer towards this uniform distribution, the area between these two curves will decrease and hence we observe decompression.

[1] Börger et al. (2018) noted that the "peak might not be unique in only rather theoretical scenarios [...]. In such a case, one might use a suitable alternative to M or modify the framework to include additional statistics."

[2] In the literature we find an extensive discussion about the existence of this age. For example Gampe (2010) or more recently Barbi et al. (2018) find evidence for a plateau in the mortality hazard after an age beyond 100. In contrast, for example Gavrilov and Gavrilova (2011) find that this is an artefact. By use of methods from extreme value theory for example Feifel et al. (2018) found significant statistical evidence for the existence of a finite lifespan for US females. Thus, like Börger et al. (2018) we assume that the right endpoint of the deaths curve's support always exists.

4. *Concentration* or *diffusion* is linked to an increase or decrease in the *number of deaths in the modal age at death* ($d(M)$). This statistic measures the deaths curve's value at age M. If more/less people die at the ages around M, $d(M)$ will increase/decrease. Intuitively this statistic indicates the relative importance of the ages around M compared to the rest of the deaths curve. Like DoI, $d(M)$ indicates changes in the deaths curve's shape.

Each component of this vector can also assume the value neutral, which is the case if the respective statistic neither (significantly) increases nor decreases between two points in time. This results in a total of $3^4 = 81$ possible different scenarios, some of which might not be relevant in practice. It is noteworthy though that any material change of the deaths curve is detected by at least one of the components. Börger et al. (2018) consider a scenario to be a pure scenario if only one of four components is different from neutral. If at least two components indicate a change, a mixed scenario prevails. They find that in practice, mixed scenarios are the rule rather than the exception (see also Genz, 2017) and conclude that the presence of mixed scenarios is not considered sufficiently in the literature.

This classification framework for observed changes of the deaths curve can be applied to deaths curves with any starting age x_0. For example, if one is interested in the evolution of old-age mortality, x_0 should be a rather old age and the deaths curve gives the age distribution at death conditional on survival to this (old) age.

We would like to emphasize once again that each of the aforementioned statistics has an intuitive interpretation. So either an extrapolation of past trends or an expert opinion on the future development can be used in forecasting. Also, it can be relatively easily assessed whether a simple statistical extrapolation of a past trend is (demographically) reasonable or not. This is much more difficult in models that are based on an extrapolation of statistics which do not allow for an intuitive interpretation. While an intuitive interpretation is not necessary for extrapolations of past trends, it is of course very helpful in case of extrapolations based on expert opinion.

An extrapolation of past trends can be based on the methodology that was developed by Börger et al. (2018) to determine (piecewise linear) trends in each of the four statistics in the past. In order to illustrate this concept, we use data from the Human Mortality Database (HMD, 2015) for Swiss females and determine the deaths curve for each calendar year between 1920 and 2014 and starting age $x_0 = 60$. For each calendar year, we estimate the four statistics of the classification framework and fit piecewise linear trends to the resulting time series (see Figure 1).[3] For the extrapolation performed in this paper, the slope of most recent trend line is relevant, whereas for the classification in Börger et al. (2018), the sign of the trend was more

[3]Between any two trends, Börger et al. (2018) allow for jumps, which can be important whenever a time series has sudden changes in its level. In Figure 1 we can see an example for such a jump in UB.

relevant.[4]

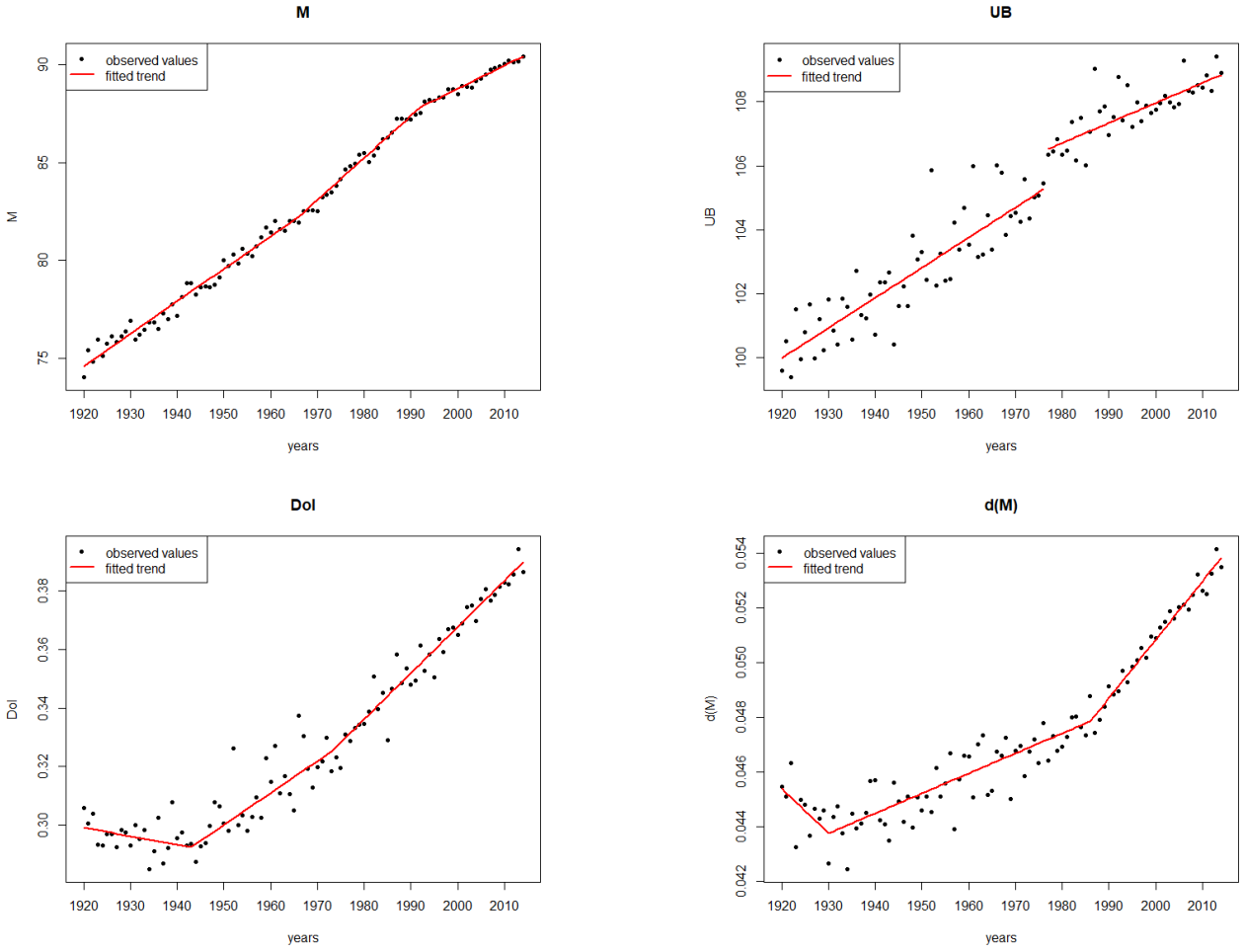

Figure 1: Observed trends in the four statistics of the classification framework for Swiss females between 1920 and 2014; top-left: M, top-right: UB, bottom-left: DoI, bottom-right: $d(M)$

2.2 Consistency Issues in Existing Forecasting Models

In this subsection, we check the demographic consistency of mortality forecasts. We fit two well-established mortality forecasting models (the Lee-Carter model (LC) and the Cairns-Blake-Dowd model (CBD)) to the mortality data from the previous subsection. From Figure 1 we can see, that the most recent trend change occurred in 1992 for M, in 1976 for UB, in 1963 for DoI, and in 1986 for $d(M)$. Since all statistics exhibit stable trends from 1992 onward, we fit both mortality models to data between 1992 and 2014 and for ages above 60. This gives both

[4]Hence, they developed a test whether any trend is significantly different from neutral (i.e. the slope of the line fitted to the time series is significantly different from zero). For any calendar year, each of the four components can then be classified as increasing, decreasing, or neutral. The resulting four-dimensional vector is the mortality scenario for this calendar year.

models the best possible chance to extrapolate historical trends in a demographically reasonable manner.[5]

Figure 2 shows the observed values of each statistic of the classification framework between 1992 and 2014 as well as extrapolated linear trends and the values forecast with the LC and the CBD model from 2015 to 2070.[6]

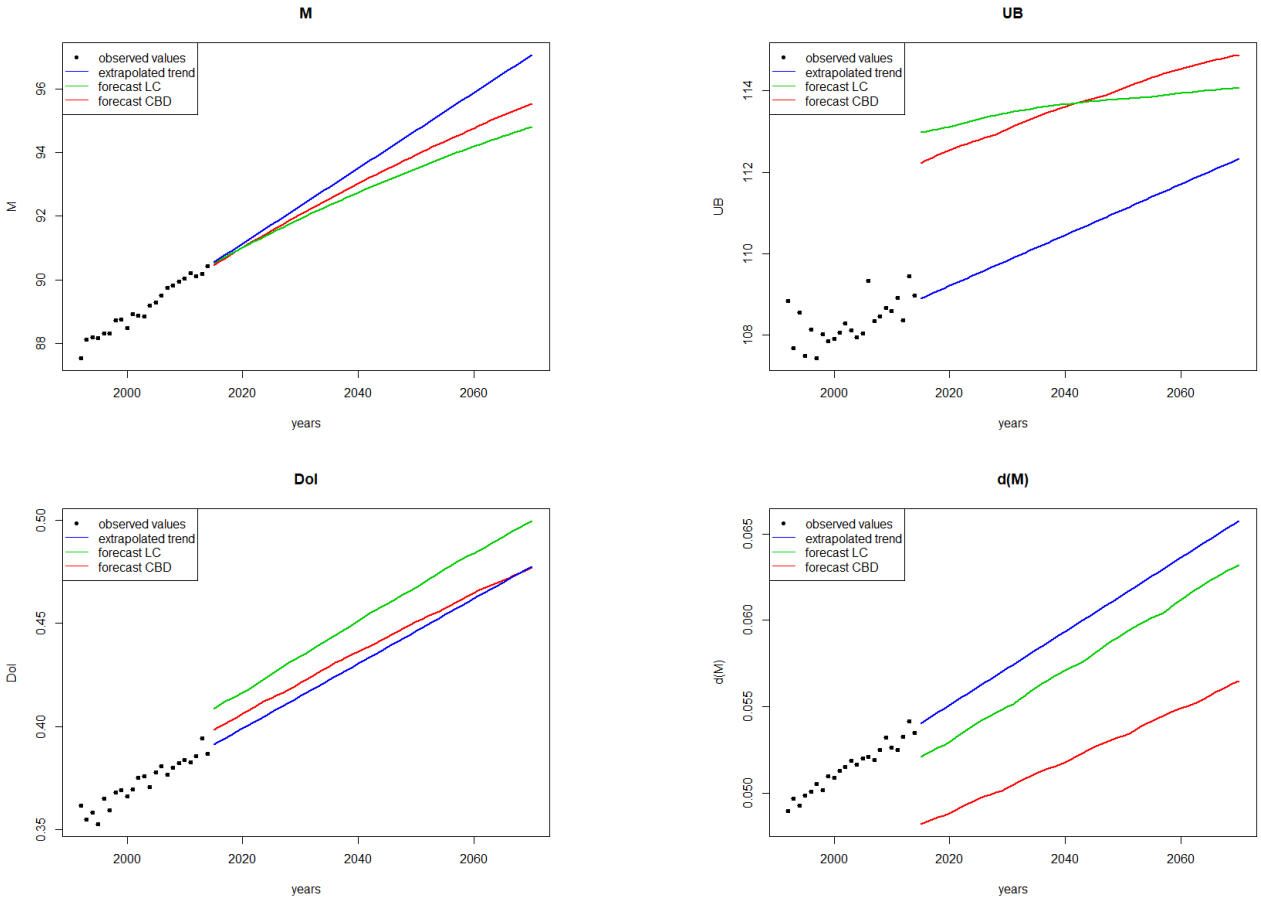

Figure 2: Observed values (black dots) between 1992 and 2014, extrapolated trends (blue lines), and model forecasts of the LC (green line) and the CBD model (red line) from 2015 to 2070; top-left: M; top-right: UB; bottom-left: DoI; bottom-right: $d(M)$

Three of the four statistics exhibit significant jumps at the transition from the calibration to the forecasting period in both mortality models. This indicates that the extrapolated deaths curves

[5]When determining the calibration period for mortality models which do not allow for trend changes, the mortality evolution during the calibration period should not exhibit any trend change. Thus for such models the most recent trend change in the classification framework of Börger et al. (2018) provides a suitable criterion for the choice of the calibration period.

[6]To calibrate the parameters of both models we use the R package StMoMo (see Villegas et al., 2018). This package also gives best estimate forecasts from which we determine estimates of the four statistics of the classification framework. The kinks in the forecast values for UB, DoI, and $d(M)$ stem from a discretization issue which is due to the fact that the models only provide mortality rates for integer ages. We ignore this minor issue since it does not affect the interpretation of the graphs.

do not exhibit smooth changes at this transition, i.e. they are not consistent with the immediate past. Thus, for both mortality models, the forecast deaths curves appear to be implausible from a demographic point of view even for the immediate future.
The projection of M on the other hand seems more plausible than those for the other three statistics since the transition from the calibration to the forecasting period does not exhibit a jump.

3 A New Best Estimate Mortality Model

In this section, we present a deterministic mortality forecasting model which can yield deaths curves that are consistent with a specific given forecast for the four statistics defined in Section 2.1. In particular, consistency with the most recent evolution of the deaths curve can be ensured by extrapolating the most recent trends in the four statistics.

3.1 Theoretical Concept

Our mortality model forecasts (continuous) deaths curves which – if scaled such that it integrates to one – can be interpreted as the density function of the age distribution at death From these deaths curves basically all other mortality statistics and in particular age dependent mortality rates can be deduced.
In order to develop the theoretical concept behind our model, we first assume that a forecast $(M_t, UB_t, DoI_t, d(M)_t)$ of the four statistics for any year t of the forecasting period is given. We then denote the space of all density functions on the (age) interval $[x_0, UB_t]$ which are differentiable at least three times by $\mathcal{D}_{t,x_0}$. It contains all functions $d_t : [x_0, UB_t] \to [0, 1]$ that fulfill the following two properties:

$$\int_{x_0}^{UB_t} d_t(x)dx = 1 \tag{1}$$

and

$$d_t(x) \geq 0 \text{ for all } x \in [x_0, UB_t]. \tag{2}$$

In order to restrict that space to deaths curves with "typical" shapes, we demand the following additional properties:

The deaths curve assumes the value of zero only at UB_t:

$$d_t(x) \begin{cases} > 0 \text{ for all } x \in [x_0, UB_t) \\ = 0,\ x = UB_t \end{cases} . \tag{3}$$

For a sufficiently large x_0, the deaths curve is unimodal, i.e. it has a unique global maximum which is also the only local maximum.[7] Consequently, we get:

$$d_t'(x) \begin{cases} > 0 \text{ for all } x \in (x_0, M_t) \\ < 0 \text{ for all } x \in (M_t, UB_t) \end{cases} . \tag{4}$$

For a sufficiently small choice of x_0, a typical deaths curve has two (unknown) inflection points, one between x_0 and M_t and one between M_t and UB_t.[8] Further, the deaths curve is concave between the inflection points and convex in both tails. Accordingly, the second derivative is positive in x_0 and UB_t, and negative in M_t:

$$d_t''(x_0) > 0, \tag{5}$$

$$d_t''(M_t) < 0,$$

$$d_t''(UB_t) > 0, \text{ and}$$

$$d_t'''(x) \begin{cases} < 0 \text{ for all } x \in [x_0, M_t) \\ > 0 \text{ for all } x \in (M_t, UB_t] \end{cases} .$$

We denote by D_{t,x_0} the subset of $\mathcal{D}_{t,x_0}$ which contains all deaths curves that fulfill these constraints. It is the set of "density functions with reasonable shape".
In the practical application of this model there might be further requirements the deaths curve should fulfill to be plausible. Therefore the set D_{t,x_0} further can be restricted. In particular, for more stable results in the left tail, it is advisable to specify the number of deaths in the starting age $d(x_0)$. This is:

[7] Besides the expected maximum at M, local maxima can typically be observed at age 0 (infant mortality) and at the peak of the so-called accident hump (typically at some age between 18 and 25). Thus, for example a choice of $x_0 = 30$ would be sufficiently large in any relevant case. In this sense, this is not a limitation if one is interested in mature- and old-age mortality.

[8] If the distance between x_0 and M becomes too small, it may happen that the deaths curve is concave at x_0 and there is no inflection point between x_0 and M_t. In practical applications, however, this is irrelevant unless an x_0 is chosen that is very close to the initial value of M and/or a strongly decreasing trend for M is assumed.

Given a forecast for the number of deaths at the starting age at time t, $d(x_0)_t$, it must hold:

$$d_t(x_0) = d(x_0)_t \tag{6}$$

for pre-specified values $d(x_0)_t$.

Since typically, the logarithm of the number of deaths at the starting age $\log d(x_0)$ follows a piecewise linear trend, we linearly extrapolate the most recent trend of this statistic in order to get reasonable and demographically consistent forecasts for $d(x_0)_t$ at any future year t.
In the last step, we consider the subset $\overline{D}_{t,x_0} \subset D_{t,x_0}$ containing only deaths curves which also match the forecasts of the statistics $(M_t, UB_t, DoI_t, d(M)_t)$, i.e. for each element $d_t : [x_0, UB_t] \to [0,1]$ of $\overline{D}_{t,x_0}$ it must hold:[9]

$$\arg\max d_t(x) = M_t, \tag{7}$$

$$\sup_{x>x_0}(d_t(x) > 0) = UB_t, \tag{8}$$

$$\frac{UB_t - x_0}{2 \cdot (UB_t - x_0) - 1} \cdot \int_{x=x_0}^{UB_t} \left| d_t(x) - \frac{1}{UB_t - x_0} \right| dx = DoI_t, \text{ and} \tag{9}$$

$$\max_{x \geq x_0}(d_t(x)) = d(M_t). \tag{10}$$

Of course, the set $\overline{D}_{t,x_0}$ might be empty or there might be more than one curve in this set. In case this set is empty, we follow that the (long-term) extrapolations of the statistics are not plausible.[10] In case there is more than one deaths curve in the set $\overline{D}_{t,x_0}$, a criterion is required to determine which deaths curve should serve as a forecast at time t. We discuss concrete examples for that in the next subsection.
The constraints from the Equations 1, 2, and 7 to 10 are mandatory for our model. A curve which does not fulfill all these constraints is either no deaths curve at all (if Equations 1 and/or 2 are not fulfilled) or not consistent with the forecasts of the future mortality scenarios (i.e.

[9] Börger et al. (2018) propose discrete estimators for the four statistics. Since we are in a continuous setting we had to adjust the estimators accordingly.

[10] Alternatively, this could also mean that the shape requirements (Equations 3 to 6) are too restrictive, i.e. a deaths curve that we regard as implausible from today's point of view because it violates one of these equations should nevertheless be considered feasible at a (probably far away) future point of time.

at least one of the Equations 7 to 10 is not fulfilled). The constraints from Equations 3 to 6 are helpful in choosing a plausible deaths curve but not necessary for an extrapolation of the deaths curve. In this sense these requirements can be extended or relaxed depending on the application of the model, the question at hand, or the user's view on what a plausible deaths curve should look like.

3.2 Basic Idea of Practical Implementation

In the previous subsection we have introduced the theoretical concept of our new model. Now we explain how it can be implemented in order to derive concrete mortality forecasts (i.e. for computing deaths curves for future years t, starting age x_0, and statistics' forecasts $(M_t, UB_t, DoI_t, d(M)_t)$). First we need to specify a functional representation of the elements of the set of "typically shaped" deaths curves D_{t,x_0}. The functional representation needs to be chosen carefully since a restrictive and inflexible representation can imply that the set $\overline{D}_{t,x_0}$ becomes empty too easily. This would for instance be the case if one considered only those deaths curve which correspond to a mortality curve according to the Gompertz law (Gompertz, 1825). In this case all deaths curves would be determined by only two parameters which makes it impossible in general to find deaths curves which fulfill Equations 7 to 10 simultaneously. Therefore, it appears more reasonable to use non-parametric representations, e.g. based on splines. In this case, the number and the degree of the splines as well as their positions obviously have an impact on whether the set $\overline{D}_{t,x_0}$ is empty or not. Here, a reasonable trade-off between tractability of the representation and variability in the deaths curves it can generate is necessary. For the example applications in the subsequent section, we use a set of 21 B-Splines of polynomial degree five for the deaths curve representation. While the high polynomial degree ensures a reasonable degree of smoothness of the deaths curves, the large number of splines allows for sufficient flexibility in the deaths curves. This representation clearly outperforms the common parametric deaths or mortality curve representations in terms of flexibility in the generated deaths curves. In Appendix A we provide details on the spline representation and its estimation in forecasting future deaths curves.

Based on this representation we can determine whether the set $\overline{D}_{t,x_0}$ is empty. In this case there is no reasonably shaped deaths curve for the year t. In the other case it usually contains infinitely many feasible deaths curves which differ only very slightly since they coincide with respect to the four statistics form Section 2.1. In order to determine a unique mortality forecast, one of the following criteria could be applied to pick a single deaths curve from $\overline{D}_{t,x_0}$:

- *Smoothness of the deaths curve*: One may pick the deaths curve from $\overline{D}_{t,x_0}$ which leads to maximal smoothness according to some smoothness measure.

- *Minimum deaths curve changes over time*: One may choose the deaths curve which is "most similar" to a previous deaths curve. The similarity can be determined by any

distance measure.

In our algorithm we implicitly use the criterion of minimum deaths curve changes over time and choose the deaths curve with minimal distance to the deaths curve of the last year where we have observed data, i.e. the last year before the projection starts (see Appendix A for more details).

4 Applications of the New Model

In this section we apply the model to the data that had also been used in Section 2, i.e. Swiss females with starting age 60. For the first example, we extrapolate the most recent linear trends for each of the four statistics to the years between 2015 and 2070 and for each calendar year fit a deaths curve to the extrapolated statistics.

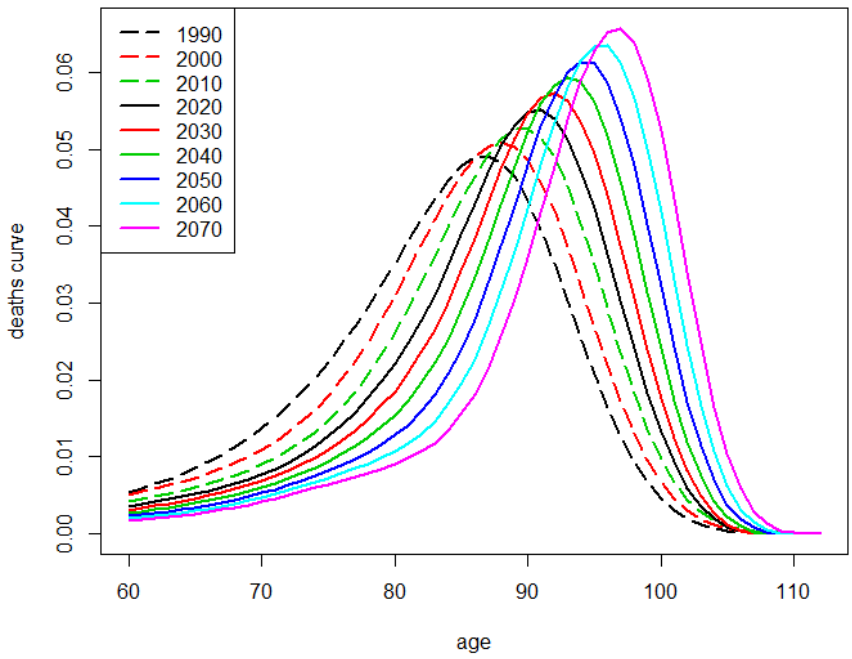

Figure 3: Deaths curves for Swiss females in the years 1990, 2000, 2010 (observed, dashed lines) and 2020, 2030, 2040, 2050, 2060, and 2070 (forecast with linear extrapolation of most recent trends, solid lines).

Figure 3 shows nine deaths curves for Swiss females between 1990 and 2070. Both for the historical deaths curves and for the extrapolated deaths curves we observe right-shifting mortality as the peaks of the deaths curves move more and more to the right. At the same time the upper bounds of the deaths curve's support also move to the right indicating extension. Also, the deaths curves from Figure 3 become more and more compressed between 1990 and 2070 since the extrapolation yields an increasing DoI implying compression. Finally, the number of deaths

in the modal age at death $d(M)$ increases over time as the peak of the deaths curves in this figure becomes ever higher.

It is particularly noteworthy that the forecast deaths curves between 2020 and 2070 smoothly continue the historical changes of the deaths curve between 1990 and 2010. The four trends (right-shifting mortality, extension, compression, concentration) are neither accelerated nor decelerated at or after the transition from the calibration to the forecasting period. Hence the mortality forecasts are consistent with the most recent observed demographic trends.

Of course, this new mortality model by definition extrapolates the four statistics it is based on in a reasonable manner. For the whole curve to be considered a reasonable forecast, also other quantities like probabilities of death should be plausibly forecast. Figure 4 shows forecasting results for $\log q_t(x)$ for selected ages from 60 to 100. We can see that these are reasonably extrapolated to the future.[11]

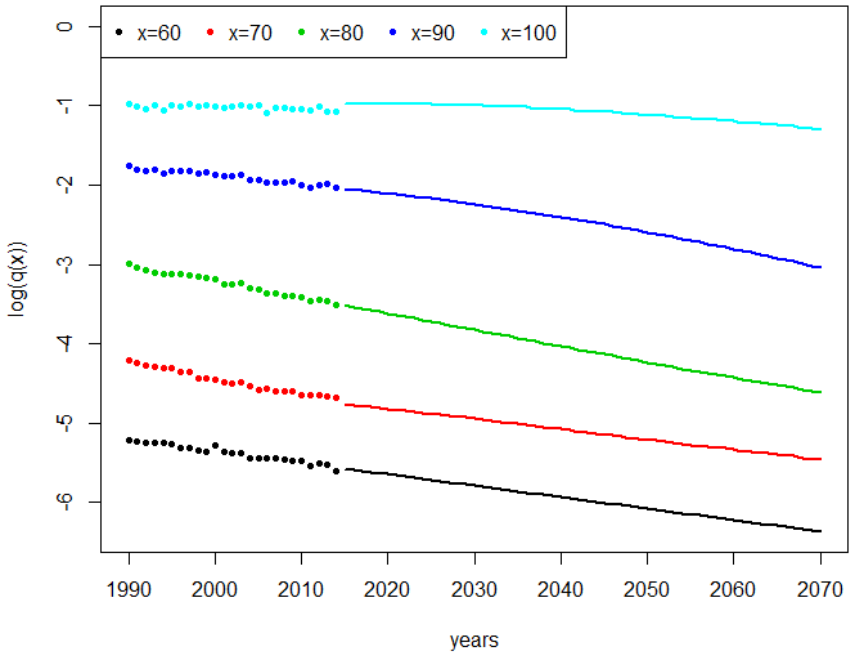

Figure 4: Logarithm of the probability of death for the ages $x = 60$, 70, 80, 90, and 100 between 1990 and 2014 (observed values) and between 2015 and 2070 (forecast values).

The demographic literature on mortality scenarios contains many different statistics which are designed to measure changes in the age distribution of deaths over time. In order to test our model for consistency also in terms of such other statistics we analyze their evolution between 1990 and 2070 (where from 2015 to 2070 we calculated the statistics based on our forecasts). As examples we considered the remaining period life expectancy at different ages ($e(60)$ in

[11] We have performed similar analyses also for the force of mortality and the probability of survival. All were plausible. For the sake of brevity, we omit details.

particular), statistics of the C-family ($C50$ in particular; see Kannisto, 2000), the Inter-Quartile Range (IQR; see Wilmoth and Horiuchi, 1999), graphical measures like the Moving Rectangle (MR; see Wilmoth and Horiuchi, 1999), and the standard deviation of the age distribution at death above the modal age at death ($SD(M+)$; see e.g. Kannisto, 2001).

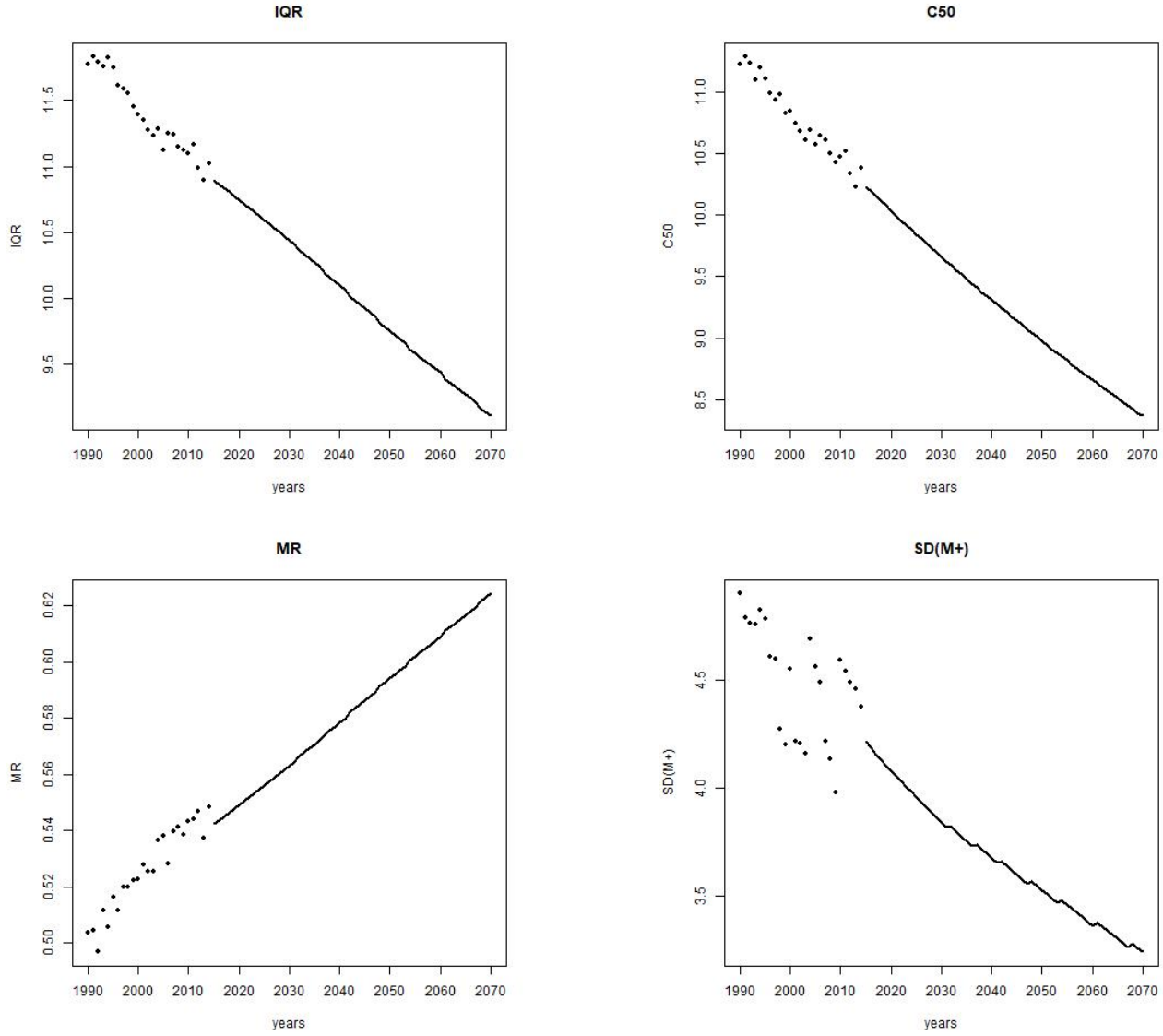

Figure 5: Alternative statistics between 1990 and 2014 (observed values) and between 2015 and 2070 (forecast values)

In Figures 5 (IQR, $C50$, MR, and $SD(M+)$) and 7 ($e(60)$ given by the black line) we illustrate some examples. Here we can see that the most recent trends in these statistics are reasonably

extrapolated until 2070.[12]

Since the four statistics used in our model allow for an intuitive interpretation, one might be interested in analyzing the effects resulting from the individual scenario components. For example the development displayed in Figure 3 is a "mixed scenario", where all four statistics of the classification framework of Börger et al. (2018) increase over time. We simultaneously observe right-shifting mortality, extension, compression, and concentration. In order to quantify the impact of each of the four scenario components separately, we consider so-called "pure scenarios" in which only one statistic changes with time (and the others remain constant).

Figure 6 displays deaths curves for each pure scenario in 2014 and the last future year where we still have reasonable deaths curves. Only in the pure right-shifting scenario we obtain deaths curves with reasonable shape until 2070. For all other scenarios the deaths curves' shapes in the pure scenarios become deformed such that before 2070 Equations 3 to 6 from Section 3.1 can not be fulfilled anymore. This is further evidence for the observation in Börger et al. (2018): pure scenarios may occur for a certain period of time, but, over a longer time period, typically only mixed scenarios can prevail.

Figure 7 shows the development of $e(60)$ in each of the pure scenarios as well as in the mixed scenario. In each pure scenario, the increase of the forecasts of $e(60)$ is slower than in the mixed scenario. The increase in $e(60)$ in the pure right-shifting scenario is faster than in any other pure scenario. This is an indication that in this example, the process of right-shifting mortality has a stronger impact on the increase of life expectancy at age 60 compared to the effects resulting from extension or compression.

As an alternative to the straightforward extrapolation of historic trends, the trends can easily be altered in our model. If, e.g., one has reason to believe that the trend in one or more of the four components will change at some point in time, one could either stop, intensify, reduce, or even reverse the increase / decrease in these components at one or more arbitrary points in time. Thus this model also allows for "what-if-analyses" for virtually any change of the mortality evolution in the future based on specifications of the four statistics. As an example for such an analysis, we use the base scenario from Figure 3 but double the intensity of right-shifting mortality and extension in 2015. Demographically this expert scenario means that the position of the deaths curve will change twice as fast in the future than recently observed, while trends in the deaths curve's shape remain unaltered. In this sense, this scenario is a stress scenario

[12] In total we considered 16 different statistics on the deaths curve: $C10$, $C50$, and $C90$ from the so-called C-family (see Kannisto, 2000), the remaining life expectancy at the ages 60, 70, 80, 90, and 100, the Fixed and Moving Rectangle, the Fastest Decline, the Sharpest Corner, the Quickest Plateau, as well as the Inter-Quartile Range (see Wilmoth and Horiuchi, 1999), the Prolate Index (see Eakin and Witten, 1995), and the standard deviation above the modal age at death (see Kannisto, 2001). We have obtained plausible results also for the statistics that are omitted in the figures for the sake of brevity.

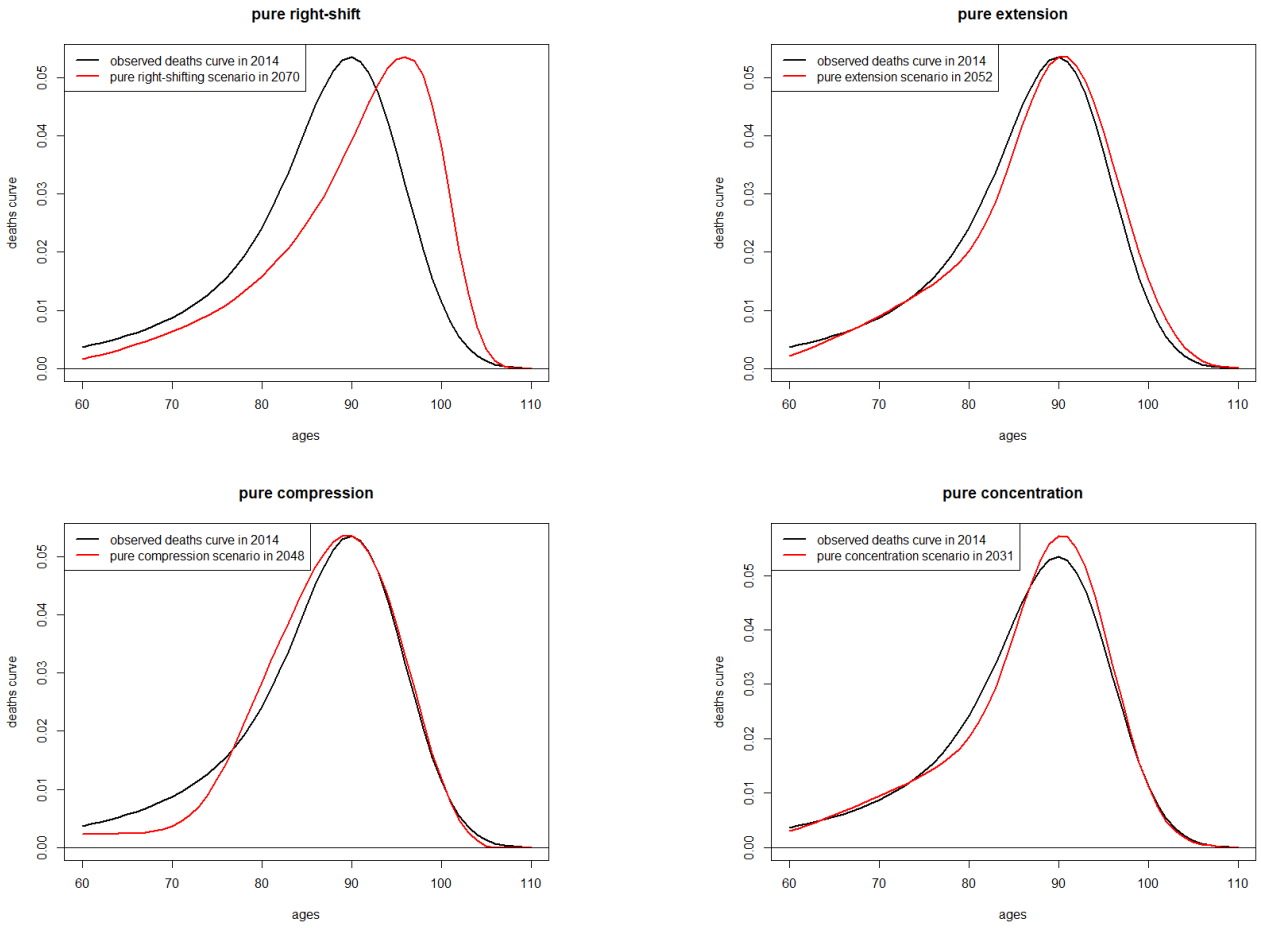

Figure 6: Observed deaths curve from 2014 and forecast deaths curve in each pure scenario; top-left: pure right-shifting scenario in 2070, top-right: pure extension scenario in 2052, bottom-left: pure compression scenario in 2048, bottom-right: pure concentration scenario in 2031.

for changes in the deaths curve's position.[13] Note that for this scenario we obtain reasonable deaths curves until 2070. Also, in this scenario, the cohort life expectancy of a 60-year old Swiss female in 2015 (i.e. with year of birth 1954) would increase from 30.4 years to 32.1 years which is an increase of more than 5.4%.

5 Conclusion

This paper discusses the plausibility of mortality forecasts from a demographic perspective. In particular, trends in key demographic figures should be reasonably extrapolated into the future.

[13] It is noteworthy that such a stress scenario is not implausible. The observed trends for M and UB in Figure 1 have also exhibited strong changes in the past: While M increased by 0.21 years per calendar year between 1966 and 1992, its increase was only 0.12 years per calendar year thereafter. For UB we historically observed a stronger increase (0.09 years per calendar year from 1920 to 1976) than in the most recent period of stable trend (0.06 years per calendar year between 1976 and 2014).

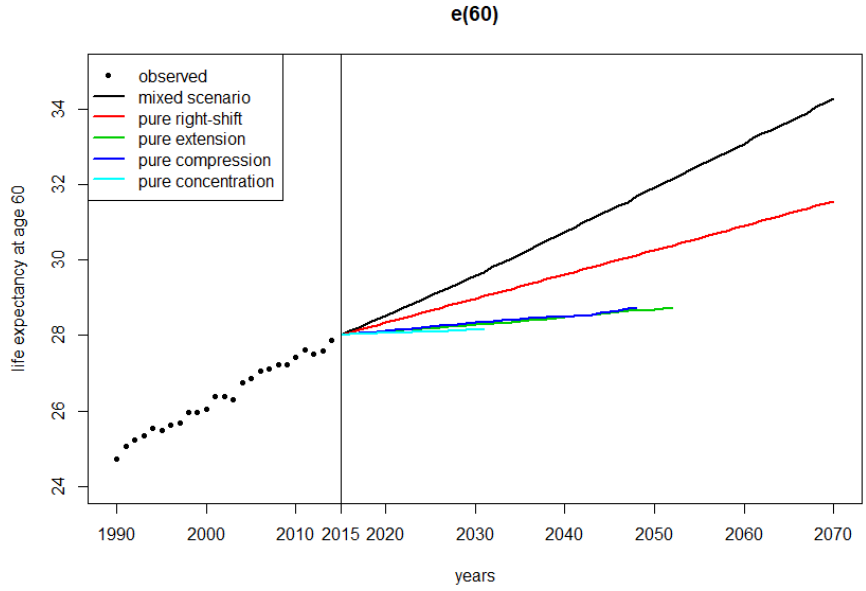

Figure 7: Observed and forecasted period life expectancy at starting age 60 in five different scenarios: mixed, pure right-shifting, pure extension, pure compression, and pure concentration scenario.

For purely statistical projection models whose parameters lack a clear demographic interpretation this is often not the case, and we provide examples for this. We apply the well-known models of Lee and Carter (LC) and Cairns, Blake, and Dowd (CBD) and analyze the resulting deaths curves in the classification framework of Börger et al. (2018). Here, deaths curves are classified based on four statistics which all have an intuitive demographic interpretation: the modal age at death M, the upper bound of the deaths curve's support UB, the degree of inequality DoI, and the number of deaths in the modal age at death $d(M)$. We find that both, the LC and the CBD model do not adequately extrapolate most recent historical trends in the four statistics. In most cases, we find jumps in the statistics at the transition from the historical to the forecast data in particular.

However, the classification framework of Börger et al. (2018) can not only be applied in order to analyze the demographic plausibility of existing mortality projections. Since in the framework two deaths curves are significantly different if and only if at least one of the four statistics differs, the framework can also be the basis for a projection model. Projections of future deaths curves (and thus essentially also any other mortality related quantity) can be derived from extrapolations of the four statistics. This ensures by construction that the mortality forecasts from this model are plausible from a demographic perspective. For the extrapolation of the statistics one may rely on their most recent observed trends or on experts' opinions on how the statistics may evolve in the future. Suitable additional constraints on the shape of forecast deaths curve need to be specified in order to ensure that the forecast deaths curves maintain a reasonable shape, and we show how this can be done in practice. We also provide a concrete

example forecast for Swiss females and illustrate its demographic plausibility by analyzing the evolution of additional statistics like the IQR over time.

The proposed projection model also allows for informative sensitivity analyses based on scenarios with a clear demographic interpretation. As (rather theoretical) examples, we consider the pure mortality scenarios in which only one of the four statistics evolves as in the most recent data and the other three statistics remain constant in the future. We find that such pure scenarios can only prevail for a certain period of time before the deaths curve's shape becomes implausible. As a possibly more realistic scenario, we also consider a 100% increase in the trends in M and UB compared to most recent historical trends (indicating increased right-shifting and extension of mortality). Here the remaining cohort life expectancy of 60-year olds increases by more than 5%. Such a scenario would be particularly critical from an annuity provider's perspective as the cohort life expectancy can be interpreted as an annuity present value with discount rate zero. Thus, the proposed projection model can be a useful risk management tool for quantifying the impact of demographic trends and its potential changes on the liabilities.

Furthermore, the proposed projection model can be a helpful addition to the toolkit of mortality projection models with respect to quantifying model risk as its structure and its forecasting approach are fundamentally different from those in established (statistical) models like the LC or the CBD model.

References

Barbi, E., Lagona, F., Marsili, M., Vaupel, J.W., and Wachter, K.W. (2018). The plateau of human mortality: Demography of longevity pioneers. *Science* 360: 1459–1461. DOI: 10.1126/science.aat3119.

Booth, H. and Tickle, L. (2008). Mortality Modelling and Forecasting: A Review of Methods. *Annals of Actuarial Science* 3(1-2): 3–43. DOI: 10.1017/S1748499500000440.

Börger, M., Genz, M., and Ruß, J. (2018). Extension, Compression, and Beyond - A Unique Classification System for Mortality Evolution Patterns. *Demography* 55(4): 1343–1361. DOI: 10.1007/s13524-018-0694-3.

Cairns, A.J.G., Blake, D., and Dowd, K. (2006). A Two-Factor Model for Stochastic Mortality with Parameter Uncertainty: Theory and Calibration. *The Journal of Risk and Insurance* 73(4): 687–718. DOI: 10.1111/j.1539-6975.2006.00195.x.

Cairns, A.J.G., Blake, D., and Dowd, K. (2008). Modelling and Management of Mortality Risk: A Review. *Scandinavian Actuarial Journal* 2008(2-3): 79–113. DOI: 10.1080/03461230802173608.

Dong, X., Milholland, B., and Vijg, J. (2016). Evidence for a limit to human lifespan. *Nature* 538(7624): 257–259.

Eakin, T. and Witten, M. (1995). How Square is the Survival Curve of a Given Species? *Experimental Gerontology* 30(1): 33–64. DOI: 10.1016/0531-5565(94)00042-2.

Feifel, J., Genz, M., and Pauly, M. (2018). *The Myth of Immortality: An Analysis of the Maximum Lifespan of US Females*. Preprint, Ulm University and Institute for Finance- and Actuarial Science, Ulm. URL: https://www.ifa-ulm.de/fileadmin/user_upload/download/forschung/2018_ifa_Feifel-etal_The-Myth-of-Immortality-An_Analysis-of-the-Maximum-Lifespan-of-US-Females.pdf.

Fries, J.F. (1980). Aging, natural death, and the compression of morbidity. *New England Journal of Medicine* 303(3): 130–135.

Gampe, J. (2010). Human Mortality Beyond Age 110. In: *Supercentenarians. Demographic Research Monographs (A series of the Max Planck Institute for Demographic Research)*. Ed. by H. Maier, J. Gampe, B. Jeune, Robine J.-M., and J.W. Vaupel. Berlin, Heidelberg: Springer: 219–230. DOI: 10.1007/978-3-642-11520-2_13.

Gavrilov, L.A. and Gavrilova, N.S. (2011). Mortality Measurement at Advanced Ages: A Study of the Social Security Administration Death Master File. *North American Actuarial Journal* 15(3): 432–447. DOI: 10.1080/10920277.2011.10597629.

Genz, M. (2017). *A Comprehensive Analysis of the Patterns of Worldwide Mortality Evolution*. Paper presented at the 2017 Living to 100 Society of Actuaries International Symposium, Orlando, FL. URL: https://www.soa.org/essays-monographs/2017-living-to-100/2017-living-100-monograph-genz-paper.pdf.

Gompertz, B. (1825). On the Nature of the Function Expressive of the Law of Human Mortality, and on a New Mode of Determining the Value of Life Contingencies. *Philosophical Transactions of the Royal Society of London* 115: 513–585.

HMD (2015). *Human Mortality Database*. University of California, Berkeley, and Max Planck Institute for Demographic Research. URL: www.mortality.org.

Kannisto, V. (2000). Measuring the compression of mortality. *Demographic Research* 3(6). DOI: 10.4054/DemRes.2000.3.6.

Kannisto, V. (2001). Mode and Dispertion of the Length of Life. *Population: An English Selection* 13(1): 159–171. URL: http://www.jstor.org/stable/3030264.

Lee, R.D. and Carter, L. (1992). Modelling and Forecasting U.S. Mortality. *Journal of the American Statistical Association* 87(419): 659–671. DOI: 10.1080/01621459.1992.10475265.

Manton, K.G., Stallard, E., and Tolley, D. (1991). Limits to Human Life Expectancy: Evidence, Prospects, and Implications. *Population and Development Review* 17(4): 603–637. DOI: 10.2307/1973599.

Oeppen, J. and Vaupel, J.W. (2002). Broken Limits to Life Expectancy. *Science* 296(5570): 1029–1031. DOI: 10.1126/science.1069675.

Olshansky, S.J., Carnes, B.A., and Cassel, C. (1990). In Search of Methuselah: Estimating the Upper Limits to Human Longevity. *Science* 250(4981): 634–640. DOI: 10.1126/science.2237414.

Pollard, J.H. (1987). Projection of Age-Specific Mortality Rates. *Population Bulletin of the United Nations* 21/22: 55–69.

Schoenberg, I.J. (1946). Contributions to the problem of approximation of equidistant data by analytic functions. *Quarterly of Applied Mathematics* 4(1–2): 45–99, 112–141.

Villegas, A.M., Kaishev, V.K., and Millossovich, P. (2018). StMoMo: An R Package for Stochastic Mortality Modeling. *Journal of Statistical Software* 84(3): 1–38. DOI: 10.18637/jss.v084.i03.

Wilmoth, J.R. and Horiuchi, S. (1999). Rectangularization Revisited: Variability of Age at Death Within Human Populations. *Demography* 36(4): 475–495. DOI: 10.2307/2648085.

Appendices

A Aspects of the Practical Implementation of the Model

A.1 Deaths curve representation by B-splines

For the examples in this paper we use a B-Spline representation for the deaths curve (Schoenberg, 1946). The deaths curve in any year t is described as a linear combination of 21 polynomial spline functions $b_t^{(j)}(x)$, $j \in \{1, ..., 21\}$ of degree five:

$$\widehat{d}_t(x) = \sum_{j=1}^{21} a_t^{(j)} \cdot b_t^{(j)}(x) = B_t(x) * a_t, \tag{11}$$

where $a_t = (a_t^{(1)}, ..., a_t^{(21)})^T$ is the vector of spline weights and the vector-valued function $B_t : \mathbb{R} \to \mathbb{R}^{21}$ gives the value of each spline at any age $x \in [x_0, UB_t]$. Each spline $b_t^{(j)}(x)$ is centered at its so-called knot $k_t^{(j)}$, $j \in \{1, ..., 21\}$ and, since the polynomial degree is odd, symmetric around its knot. Furthermore, it is different from zero only on a certain interval:

$$b_t^{(j)}(x) \begin{cases} > 0 \text{ for all } x \in (\max\left\{x_0, k_t^{(j-3)}\right\}, \min\left\{UB_t, k_t^{(j+3)}\right\}) \\ = 0 \text{ for all } x \notin (\max\left\{x_0, k_t^{(j-3)}\right\}, \min\left\{UB_t, k_t^{(j+3)}\right\}) \end{cases} . \tag{12}$$

In our forecasting model, deaths curves and thus also the spline functions need to be continuously differentiable at least three times (see Section 3.1), but differentiability in higher order is desirable. The polynomial degree of five therefore is a reasonable choice in our examples, but may of course be altered for other applications of the model.[14]

The number of splines essentially determines the flexibility of the deaths curve representation. The larger the number of splines, the more likely we are to find a deaths curve which matches a concrete forecast for the four statistics M_t, UB_t, DoI_t, and $d(M)_t$. The specific number of 21 splines is a result of the chosen positioning of the spline knots within the interval $[x_0, UB_t]$ (see below). We found that a number in that range is necessary to have a sufficiently flexible representation. Note that we fix the number of splines independently of the length of the deaths curve's support. Thus in case of extension or contraction (i.e. when UB increases or decreases), we reposition the spline knots over time, but do not change the number of splines.

In order to specify a concrete spline representation of a deaths curve, the spline knot positions $k_t^{(j)}$, $j \in \{1, ..., 21\}$ and the spline weights $a_t = (a_t^{(1)}, ..., a_t^{(21)})^T$ need to be determined such that Equations 1 to 10 are fulfilled. In what follows, we explain these two steps in more detail.

[14] We also considered the more commonly applied third degree polynomials, but found that deaths curves can become slightly "wavy".

A.2 Positioning of spline knots

Given a forecast of the four statistics $(M_t, UB_t, DoI_t, d(M)_t)$ for any year t, we determine the positions of the spline knots as follows: First, we set one knot at each endpoint of the deaths curve's support, i.e. at x_0 and at UB_t, and at the modal age at death M_t. Between x_0 and M_t we add nine equidistant knots and between M_t and UB_t we add five equidistant knots.[15] For sufficient flexibility also at the boundaries of the deaths curve's support we need two additional knots left of x_0 and right of UB_t. These knots are positioned such that all knots left or right of M_t, respectively, are equidistant. Thus, the total number of splines is $3 + 9 + 5 + 2 \cdot 2 = 21$. Figure 8 illustrates the 21 splines before weighting (see next paragraph) and their knot positions. The two splines at each boundary whose maxima are not yet visible have their knots outside the support $[x_0, UB_t]$.

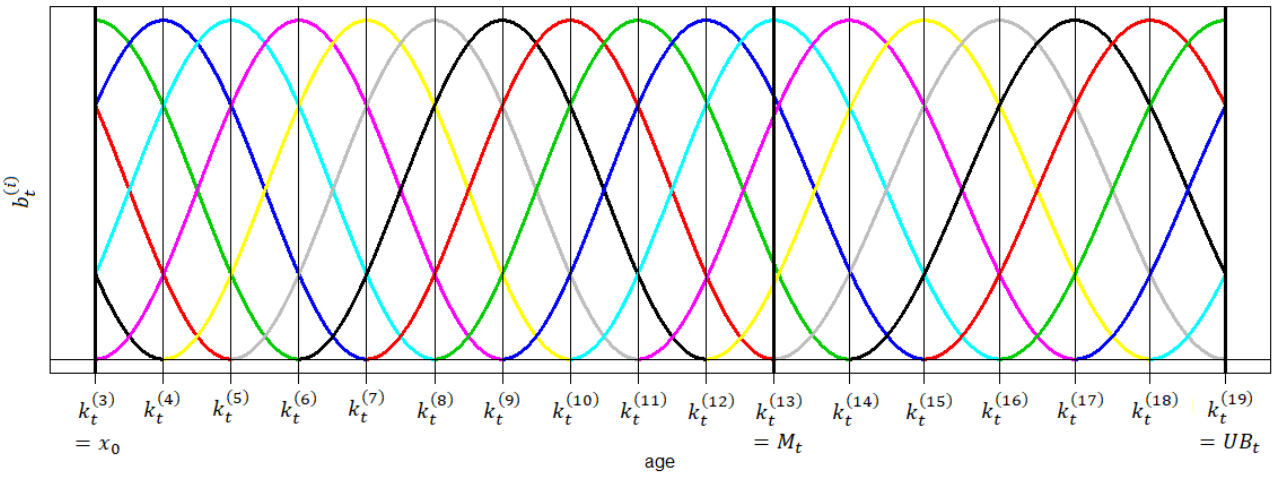

Figure 8: Knots and polynomial functions of degree five for 21 splines.

A.3 Determination of spline weights

Now, the spline weights need to be chosen for any future year t such that the resulting linear combination of the 21 polynomial spline functions (see Equation 11) results in a deaths curve that matches the given four statistics $(M_t, UB_t, DoI_t, d(M)_t)$. For the determination of the vector of spline weights a_t we use numerical methods: For any t the Equations 1 and 6 to 10 from Section 3.1 can be translated into a system of non-linear, but convex equalities:

$$A_t^{(1)}(a_t) = y_t^{(1)}, \tag{13}$$

where $A_t^{(1)} : \mathbb{R}^{21} \to \mathbb{R}^6$ is a corresponding function in the spline weights and $y_t^{(1)}$ is a vector of the target values. Analogously, the Equations 2 to 5 from Section 3.1 define a system of linear

[15] It might be reasonable to modify the respective number of knots (nine or five) if the distances between x_0 and M_t or between M_t and UB_t are rather small or extremely large.

(and thus also convex) inequalities:

$$A_t^{(2)} * a_t < y_t^{(2)}. \tag{14}$$

These constraints are continuous in the sense that they refer to each (also non-integer) age in certain intervals of the support $[x_0, UB_t]$. For the numerical implementation we need to "discretize" the constraints by evaluating them only at a finite number of ages. It turned out to be sufficient to only evaluate the constraints at the spline knots.
In order to obtain starting values for the algorithm described below, we fit a spline representation to the observed deaths curve of the most recent year t_0 for which data is available. To this end we perform a least square estimation where we require the constraints from Equations 1 to 5 to be fulfilled.
The basic idea behind the algorithm is that – for any future year t – we search for spline weights a_t which solve the Equations 13 and 14 and differ as little as possible from the weights a_{t_0}. Thus, in terms of picking a single deaths curve from $\overline{D}_{t,x_0}$ we opt for minimum deaths curve changes over time (see Section 3).[16] The algorithm proceeds as follows:

1. In a first step we ensure that the curve is an element of the space $\mathcal{D}_{t,x_0}$. To this end, we scale the vector of spline weights, i.e.

$$\widetilde{a}_t = \frac{1}{\int_{x_0}^{UB_t} B_t(x) * a_{t_0} dx} \cdot a_{t_0},$$

   such that Equation 1 holds.

2. In the second step, we alter the vector of spline weights such that the deaths curve is an element of the set D_{t,x_0}. To this end, we check if $\widetilde{a}_t$ solves Equation 14. If it does, we proceed with the third step. Otherwise, we carry out the following steps a to c which build on the method of gradient descent:

   (a) Based on Equation 14 we define a function f_t of a vector of weights a:

$$f_t(a) = \sum_{j=1}^{m} w_j \cdot A_t^{(2,j)} * a,$$

   where m is the number of rows of the matrix $A_t^{(2)}$, $A_t^{(2,j)}$ is the j^{th} row of the matrix

[16] By this algorithm we find deaths curves with minimal distance from the observed deaths curve in the year t_0. The algorithm would also work iteratively, i.e. by using the weights a_{t-1} as starting values for year t. However, if one was only interested in a forecast for year t, one would have to forecast deaths curves for all years up to year t in that case.

$A_t^{(2)}$, and the weights w_j are given by[17]

$$w_j = \begin{cases} 1 \text{ if } A_t^{(2,j)} * a < 0 \\ 10^6 \text{ if } A_t^{(2,j)} * a \geq 0 \end{cases} .$$

(b) We determine the gradient $\nabla f_t(a)$ of the function $f_t(a)$ and, following the method of gradient descent, update the vector of weights $\widetilde{a}_t^{(\nu)}$ via

$$\widetilde{a}_t^{(\nu)} = \widetilde{a}_t^{(\nu-1)} - \theta \cdot \nabla f_t(\widetilde{a}_t^{(\nu-1)}).$$

The step size θ is determined by minimizing the difference $f_t(\widetilde{a}_t^{(\nu-1)}) - f_t(\widetilde{a}_t^{(\nu)})$ under the constraint that no additional component of the vector $A_t^{(2)} * \widetilde{a}_t^{(\nu)}$ becomes positive. In the first iteration we set $\widetilde{a}_t^{(0)} = \widetilde{a}_t$ and $\nu = 1$.

(c) We update $\widetilde{a}_t$ as long as there are positive entries in the vector $A_t^{(2)} * \widetilde{a}_t$. After the final update we set $\widetilde{a}_t = \widetilde{a}_t^{(\nu)}$.

3. In the third step, we modify the vector of spline weights such that the deaths curve is an element of the set $\overline{D}_{t,x_0}$. Similarly to step 2.a, we define a function g_t of the vector of weights a,

$$g_t(a) = \sum_{j=1}^{k} \omega_j \cdot (A_t^{(1,j)}(a) - y_t^{(1,j)})^2,$$

where k is the number of equality constraints, ω_j is a scalar weight for the j^th equality constraint, $A_t^{(1,j)}(a)$ is the left hand side value of Equation 13 for the j^{th} equality constraint, and $y_t^{(1,j)}$ is the right hand side value of Equation 13 for the j^{th} equality constraint. In order to determine the root of the function g_t, we apply the method of gradient descent analogously to the optimization of the function f_t. Starting with the vector of spline weights $\widetilde{a}_t$ from Step 2, we search for the vector of spline weights a_t for which the value of the function g_t assumes its minimum value zero.[18]

The resulting deaths curve $\widehat{d}_t(x) = B_t(x) * a_t$ is an element of the set $\overline{D}_{t,x_0}$ and thus a valid forecast of our model.

[17]The constant 10^6 in the following formula stresses the constraints which are not yet fulfilled and therefore accelerates convergence in the subsequent steps. We found 10^6 to be in a reasonable range.

[18]For numerical reasons the algorithm terminates as soon as $g_t(\widetilde{a}_t) < e^{-7}$.