Drivers of Mortality Dynamics: Identifying Age/Period/Cohort Components of Historical U.S. Mortality Improvements

Johnny Siu-Hang Li, Rui Zhou, Yanxin Liu, George Graziani, Dale Hall, Jennifer Haid, Andrew Peterson, and Laurence Pinzur

Abstract

The goal of this paper is to obtain an Age/Period/Cohort (A/P/C) decomposition of historical U.S. mortality improvement. Two different routes to achieving this goal are considered. In the first route, the desired components are obtained by fitting an A/P/C model directly to historical mortality improvement rates. In the second route, an A/P/C model is estimated to historical crude death rates and the desired components are then obtained by differencing the estimated model parameters. For each route, various possible A/P/C model structures are experimented, and are evaluated on the basis of their robustness to several factors (e.g., changes in the calibration window) and their ability to explain historical changes in mortality improvement. Based on the evaluation results, an A/P/C decomposition for each gender is recommended. The decomposition will be examined in a follow-up project, in which the linkages between the A/P/C components and certain intrinsic factors will be identified.

Keywords: Age-period-cohort models; Longevity risk; U.S. mortality.

1 Introduction

While mortality experience studies have been at the core of the research efforts of the Society of Actuaries (SOA) since its inception, wide-ranging mortality and related longevity studies have received increasing attention within the Society over the past few years. Indeed, "Mortality and Longevity" has been selected as one of the SOA's five Strategic Research Programs for 2018. And as the technology, tools, and data available to better understand and model mortality patterns has been evolving at an ever quickening pace, the Longevity Advisory Group (LAG) has been given a mandate to keep the SOA membership apprised of major developments in these fields.

One of the first projects LAG initiated was an inventory of mortality projection techniques used for various applications across each of the SOA disciplines that require estimates of future mortality rates. Without any overarching methodology for these projections, it is not surprising that a number of different approaches evolved for a procedure that – at its core – should produce similar results. In reaction to the wide variety of techniques currently in use by the membership, LAG decided that it would be worthwhile to assess the possibility of developing a more consistent

framework for mortality projection. Such a framework would be comprehensive enough to accommodate most applications but also flexible enough to reflect different underlying populations and different provisions for conservatism.

The SOA and LAG anticipate that any model that emerges from such a consistent framework would ultimately produce two-dimensional (age and calendar year) gender-specific mortality projection scales, similar in format to those produced by the SOAs Retirement Plans Experience Committee for retirement plans (e.g., Scale MP-2017) and by the Continuous Mortality Investigation (e.g., CMI Mortality Projections Model: CMI_2017). The development of these and other two-dimensional projection scales typically begins with a decomposition of historical mortality experience for the covered group into separate age, period (time-related), and cohort (year-of-birth-related) (A/P/C) components.

The historical A/P/C decomposition process can be performed in a variety of ways using a number of different underlying mortality models. Therefore, the first step in the larger "consistent framework" project has been to determine which A/P/C decomposition methodologies are most robust when applied to U.S. population mortality experience. In 2017, the SOA commissioned a study called "Components of Historical Mortality Improvement," which aims to compare and contrast methodologies for allocating historical gender-specific mortality improvement experience in the U.S. into A/P/C components. This paper documents the research work performed for the study, with an emphasis on explaining how and why a certain model structure is identified (from a large collection of candidate model structures) as the best suited for deriving an A/P/C decomposition of historical U.S. mortality improvement experience. While this paper does identify one methodology that proves especially robust for the U.S. population, it also provides useful background for those who are interested in gaining a better understanding of the range of approaches currently being used by leading mortality model researchers around the world and the techniques they use to assess their efficacy for individual populations.

Subsequent phases of the "consistent framework" project will research how effectively the historical A/P/C components identified in this paper can be attributed to specific mortality drivers. A short list of the potential mortality drivers includes:

- Medical/pharmacological advances; e.g., gene therapy and statins
- Government programs; e.g., Medicare/Medicaid and anti-smoking campaigns
- Societal trends; e.g., increased incidence of obesity and opioid use

Assuming that a number of the more important causal relationships can be identified, LAG will then be in a better position to assess how reliably the evolution of these drivers can be projected into the future. In conjunction with this research project, LAG is in the process of organizing a forum of experts from a broad range of disciplines (including those in various medical professions, demographers, futurists, and actuaries) to discuss past, present and anticipated future drivers of mortality. The expectation is that the results of these research projects and the knowledge gained

from the forum of experts will play important roles in shaping the SOA's ultimate mortality improvement framework.

In the search for the most effective A/P/C decomposition methodology, the project team relies on the following three criteria:

1. Adequacy

   The resulting A/P/C decomposition should adequately capture the observable patterns in the historical mortality improvement experience. That being said, after filtering the historical mortality improvement experience with an A/P/C decomposition, the residuals left behind should appear to be reasonably random.

2. Robustness

   The resulting A/P/C decomposition should remain largely unchanged when there are small changes to the data input or model set-up. This property is important because, for instance, if an additional year of data becomes available and the model is updated accordingly, there should not be significant, unexplainable changes to the A/P/C components.

3. Smoothness

   It is anticipated that several peaks and troughs will be found in the resulting A/P/C components. However, on top of that, the resulting A/P/C components should not contain any noise which carries no possible demographic meaning.

The search for the most effective methodology is conducted in two pathways, namely, Route A and Route B. As explained in the following paragraphs, the difference between these two pathways lies in the quantity being fed into the candidate models.

In Route A, the quantity being modeled is *mortality improvement rate*, defined as the percentage reduction in the conditional probability of death at a certain age and in a certain calendar year. All of the Route A candidate models are fitted to historical mortality improvement rates directly. Given how Route A models are defined, the parameters in an estimated Route A model can be regarded as an A/P/C decomposition of historical mortality improvement. Although the approach of modeling mortality improvement rates directly is not very common in the literature, it has been considered in several recent papers including those authored by Plat (2011) and Hunt and Villegas (2017).

In Route B, the quantity being modeled is either *death rate* or *death probability*. This approach is in line with the well-known Lee-Carter (Lee and Carter, 1992) and Cairn-Blake-Dowd (Cairns et al., 2006) methods for stochastic mortality modeling. As the Route B candidate models are applied to death rates/probabilities instead of mortality improvement rates, a transformation (first differencing) has to be performed for some of the parameters in a Route B model before an A/P/C decomposition of historical mortality improvement can be obtained.

For both pathways, we 'horse race' a large number of candidate model structures. We assess how well they meet the adequacy and robustness criteria using the backtesting methods outlined

in the work of Cairns et al. (2009) and Dowd et al. (2010b). The smoothness criterion is taken into account by means of a roughness penalty in the estimation process, and is assessed by visually inspecting the resulting A/P/C decompositions.

We emphasize that the scope of this paper is confined to analyzing *historical* mortality improvement experience. As such, no attempt has been made in this paper to project future U.S. mortality. The SOA believes that a comprehensive mortality projection model requires additionally input from experts in related fields such as demography and medical science, and, as previously mentioned, has planned to invite such experts to participate in follow-up projects that will draw from the results presented in this paper. Since forecasting is beyond the scope of this paper, we have not considered attaching stochastic processes to the identified A/P/C components, and have not performed any ex-post evaluation (Cairns et al., 2011; Dowd et al., 2010a) of the mortality density forecasts produced by the candidate models.

The remainder of this paper is organized as follows. Section 2 describes the data sets on which the "Component of Historical Mortality Improvement" study is based. Section 3 defines the notation that is used throughout the paper. Section 4 provides an overview of the model selection process, which is then detailed in Section 5 (for Route A) and Section 6 (for Route B). Finally, Section 7 concludes the paper with a recommendation. It also discusses the caveats of our work, and how the results of this paper may be used in anticipated future SOA projects.

2 Data

2.1 The Data Sets Used

We consider two data sets, which are provided by the Human Mortality Database (HMD) and the U.S. Social Security Administration (SSA), respectively. Both data sets have the same age range (20 to 95) and the same sample period (1968 to 2014).

For ages 65 and over, the origins of the two data sets are different. In particular, the SSA data set is based on the death and exposure counts from the Medicare-enrolled population, whereas the HMD data set is based on deaths reported by states and census population (exposure) estimates. Goss et al. (2015) argue that the SSA data set is more reliable for the following reasons:

1. age accuracy (Medicare requires proof of age when enrolling);
2. representation of almost the entire Social Security area population;
3. both death and exposure counts are obtained from a single, consistent source.

In addition, according to the HMD documentation, the HMD data at higher ages are not 'raw'.[1]

[1] Above age 80, population estimates in the HMD data set are derived by the method of extinct generations for all cohorts that are extinct and by the survivor ratio method for non-extinct cohorts who are older than age 90 at the end of the observation period.

For these reasons, the conclusions drawn in this study are based primarily on the SSA data set. The HMD data set is used as a benchmark only.

2.2 Exploratory Data Analyses

Before performing the modeling work, we examine the historical mortality improvement rates and identify the features that we expect the candidate models to capture.

Figures 1 and 2 display the heat maps of the historical mortality improvement rates for U.S. males and females, respectively.[2] The labels in the diagram represent the ages at which the peaks and troughs of mortality improvement are located. For both genders, the following observations are made:

- Variation of colors along the vertical dimension

  Each row in the diagram represents mortality improvement rates for a specific age. The variation of colors along the vertical dimension therefore represents the age effect of mortality improvement.

- Variation of colors along the horizontal dimension

  Each column in the diagram represents mortality improvement rates for a specific calendar year. The variation of colors along the horizontal dimension therefore represents the period (time-related) effect of mortality improvement.

- Variation of colors among diagonals

  Each diagonal (from lower-left to upper-right) in the diagram represents mortality improvement rates for a specific year of birth. As such, the variation of colors among diagonals represents the cohort (year-of-birth-related) effect of mortality improvement.

The chosen model should capture all of these three observed effects, leaving residuals that exhibit no observable pattern.

Finally, we note that the historical mortality improvement rates from the two data sets are very similar. Hence, the differences between the two data sets do not seem to be a big concern.

3 Notation

The following notation is used throughout the rest of this paper:

- x represents the age of an individual;

[2] To facilitate identification of patterns, the mortality improvement rates are calculated from mortality rates that have been smoothed by a two-dimensional (age/cohort) P-Spline (Currie et al., 2004). The mortality improvement rate used here is defined in equation (2) in Section 4.

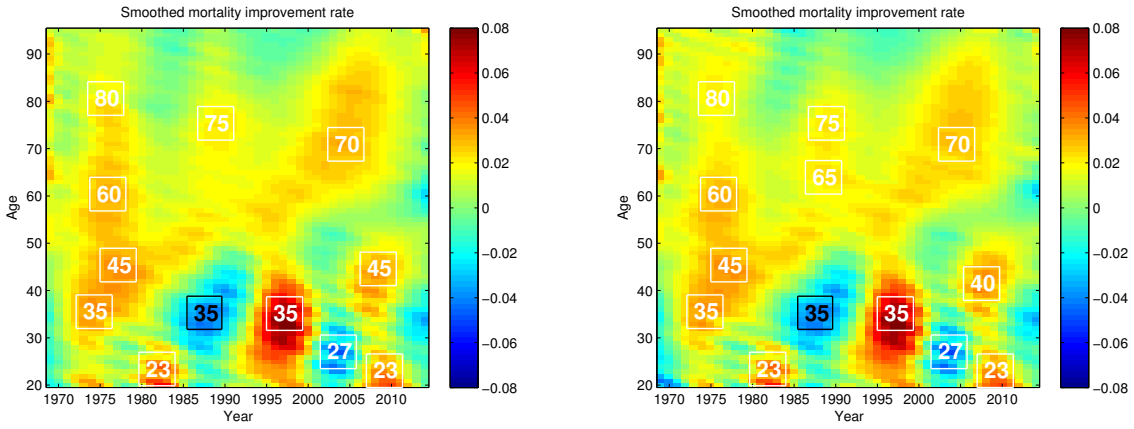

Figure 1: Heat maps of the smoothed mortality improvement rates for U.S. males, calculated from the HMD data set (left panel) and the SSA data set (right panel).

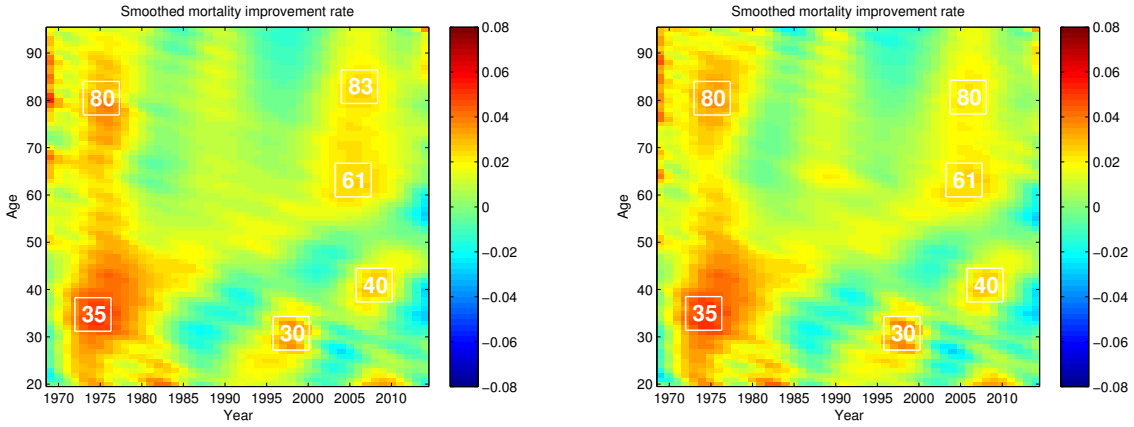

Figure 2: Heat maps of the smoothed mortality improvement rates for U.S. females, calculated from the HMD data set (left panel) and the SSA data set (right panel).

- t represents calendar year (running from time point $t-1$ to t);
- $[x_0, x_1]$ is the sample age range;
- $[t_0, t_1]$ is the sample period (i.e., the calibration window);
- $n_a = x_1 - x_0 + 1$ is the number of ages covered by the sample age range;
- $n_y = t_1 - t_0 + 1$ is the length of the sample period;
- $c = t - x$ is the year of birth; note that within the data sample c ranges from $t_0 - x_1$ to $t_1 - x_0$;
- $m_{x,t}$ is the central rate of death at age x and in year t;
- $q_{x,t}$ is the (conditional) probability that an individual who has survived to age x exact at the beginning of year t dies during year t;
- $\bar{x} = (x_0 + x_1)/2$ is the mid-point of the sample age range;
- $\bar{t} = (t_0 + t_1)/2$ is the mid-point of the sample period;
- $(\bar{x} - x)_+ = \max(\bar{x} - x, 0)$ is the maximum of $\bar{x} - x$ and 0.

In some calculations, we need a relationship between $m_{x,t}$ and $q_{x,t}$. Unless otherwise specified, we use the following approximation:

$$q_{x,t} \approx 1 - \exp(-m_{x,t}), \tag{1}$$

which holds exact if the force of mortality between two consecutive integer ages and time points is constant.

Finally, in describing the A/P/C model structures, the following conventions are used:

- $\beta_x^{(i)}$, $i = 1, 2, 3$, are age-specific parameters:
  - $\beta_x^{(1)}$ is a stand-alone age-specific parameter,
  - $\beta_x^{(2)}$ is an age-specific parameter that interacts with a time-varying parameter,
  - $\beta_x^{(3)}$ is an age-specific parameter that interacts with a cohort-related parameter;
- $\kappa_t^{(i)}$, $i = 1, 2, 3$, are time-varying parameters:
  - $\kappa_t^{(1)}$ is a stand-alone time-varying parameter;
  - $\kappa_t^{(2)}$ is a time-varying parameter that interacts with an age-specific parameter or a linear function of age;
  - $\kappa_t^{(3)}$ is a time-varying parameter that interacts with a non-linear function of age;
- γ_c is a cohort-related parameter;
- $e_{x,t}$ is the error term.

4 An Overview of the Methodology

4.1 The Two Pathways

We consider two pathways to obtaining A/P/C decompositions of historical mortality improvements: Route A and Route B.

In Route A, various candidate A/P/C models are fitted to the historical *mortality improvement rates* directly. More precisely, the candidate models in Route A are fitted to the observed values of

$$Z_{x,t} = 1 - \frac{q_{x,t}}{q_{x,t-1}}, \tag{2}$$

for $x = x_0, \ldots, x_1$ and $t = t_0, \ldots, t_1$, a quantity representing the reduction in the conditional probability of death for age x from year $t-1$ to year t.[3] Given how Route A models are defined, the parameters in an estimated Route A model can be regarded as an A/P/C decomposition of historical mortality improvement. Further details concerning Route A are provided in Section 5.

In Route B, a number of candidate age-period-cohort models are applied to the *death rates* ($m_{x,t}$) or *death probabilities* ($q_{x,t}$), for $x = x_0, \ldots, x_1$ and $t = t_0, \ldots, t_1$. Since Route B models are applied to death rates/probabilities instead of mortality improvement rates, a transformation (first differencing) has to be performed for some of the parameters in a Route B model before an A/P/C decomposition of historical mortality improvement can be obtained. Further details concerning Route B are provided in Section 6.

4.2 The Model Selection Process

For both pathways, we use the following multi-step procedure to identify the most effective model for the data sets under consideration.

1. A large number of candidate models are fitted to the historical mortality improvement rates (Route A) or death rates/probabilities (Route B).

2. Robustness tests are performed for each candidate model. On the basis of the robustness test results, a smaller number of models are shortlisted for further consideration.

3. The standardized residuals produced by the shortlisted models are carefully examined. For each pathway, the most effective model is the one that yields standardized residuals with the least unexplained patterns.

Finally, we compare the A/P/C decompositions from the most effective models in Routes A and B, and make a recommendation.

[3] We calculate the observed value of $Z_{x,t}$ using equation (1) with $m_{x,t}$ being set to the ratio of the death count to the exposure count at age x in year t.

5 Route A

This section details the modeling work associated with Route A. We first define the Route A candidate models, and explain how they are estimated. These are followed by a collection of robustness tests which allow us to shortlist a smaller number of model structures for further consideration. We then perform a residual analysis to identify the model structure that is best suited for decomposing historical U.S. mortality improvements.

5.1 The Candidate Models

We consider seven candidate APC models, including M2, M3, M6, M7, M8, the full Plat model and the simplified Plat model, in Route A. These models were originally developed to model death rates or probabilities ($m_{x,t}$'s or $q_{x,t}$'s), but here we use them to model mortality improvement rates ($Z_{x,t}$'s) instead.

Using the notation defined in Section 3, we have the following definitions of the seven candidate models in Route A.

- Model M2 – The Renshaw-Haberman Model (Renshaw and Haberman, 2006)

$$Z_{x,t} = \beta_x^{(1)} + \beta_x^{(2)}\kappa_t^{(2)} + \beta_x^{(3)}\gamma_c + e_{x,t}$$

- Model M3 – The Age-Period-Cohort Model (Osmond, 1985)

$$Z_{x,t} = \beta_x^{(1)} + \kappa_t^{(1)} + \gamma_c + e_{x,t}$$

- Model M6 – The CBD Model with a Cohort Effect (Cairns et al., 2009)

$$Z_{x,t} = \kappa_t^{(1)} + \kappa_t^{(2)}(x - \bar{x}) + \gamma_c + e_{x,t}$$

- Model M7 – The CBD Model with Quadratic Age and Cohort Effects (Cairns et al., 2009)

$$Z_{x,t} = \kappa_t^{(1)} + \kappa_t^{(2)}(x - \bar{x}) + \kappa_t^{(3)}((x - \bar{x})^2 - \hat{\sigma}_x^2) + \gamma_c + e_{x,t}$$

- Model M8 – The CBD Model with an Age-Dependent Cohort Effect (Cairns et al., 2009)

$$Z_{x,t} = \kappa_t^{(1)} + \kappa_t^{(2)}(x - \bar{x}) + \gamma_c(x_c - x) + e_{x,t}$$

- The Full Plat model (Plat, 2009):

$$Z_{x,t} = \beta_x^{(1)} + \kappa_t^{(1)} + \kappa_t^{(2)}(\bar{x} - x) + \kappa_t^{(3)}(\bar{x} - x)_+ + \gamma_c + e_{x,t}$$

- The Simplified Plat model (Plat, 2009):

$$Z_{x,t} = \beta_x^{(1)} + \kappa_t^{(1)} + \kappa_t^{(2)}(\bar{x} - x) + \gamma_c + e_{x,t}$$

Model	**Identifiability constraints**
Model M2	$\sum_{x=x_0}^{x_1}\beta_x^{(1)}=0,\ \sum_{x=x_0}^{x_1}\beta_x^{(2)}=1,\ \sum_{x=x_0}^{x_1}\beta_x^{(3)}=1,$ $\sum_{c=t_0-x_1}^{t_1-x_0} n_c\gamma_c=0$
Model M3	$\sum_{x=x_0}^{x_1}\beta_x^{(1)}=0,\ \sum_{c=t_0-x_1}^{t_1-x_0}\gamma_c=0,\ \sum_{c=t_0-x_1}^{t_1-x_0} c\gamma_c=0$
Model M6	$\sum_{c=t_0-x_1}^{t_1-x_0}\gamma_c=0,\ \sum_{c=t_0-x_1}^{t_1-x_0} c\gamma_c=0.$
Model M7	$\sum_{c=t_0-x_1}^{t_1-x_0}\gamma_c=0,\ \sum_{c=t_0-x_1}^{t_1-x_0} c\gamma_c=0,\ \sum_{c=t_0-x_1}^{t_1-x_0} c^2\gamma_c=0$
Model M8	$\sum_{c=t_0-x_1}^{t_1-x_0} n_c\gamma_c=0$
Full Plat	$\sum_{x=x_0}^{x_1}\beta_x^{(1)}=0,\ \sum_{t=t_0}^{t_1}\kappa_t^{(2)}=0,\ \sum_{t=t_0}^{t_1}\kappa_t^{(3)}=0,$ $\sum_{c=t_0-x_1}^{t_1-x_0}\gamma_c=0,\ \sum_{c=t_0-x_1}^{t_1-x_0} c\gamma_c=0,\ \sum_{c=t_0-x_1}^{t_1-x_0} c^2\gamma_c=0$
Simplified Plat	$\sum_{x=x_0}^{x_1}\beta_x^{(1)}=0,\ \sum_{t=t_0}^{t_1}\kappa_t^{(2)}=0,\ \sum_{c=t_0-x_1}^{t_1-x_0}\gamma_c=0,$ $\sum_{c=t_0-x_1}^{t_1-x_0} c\gamma_c=0,\ \sum_{c=t_0-x_1}^{t_1-x_0} c^2\gamma_c=0$

Table 1: The baseline identifiability constraints used in estimating model parameters

We do not consider Model M1 (the Lee-Carter model), Model M4 (the P-splines regression) and Model M5 (the original Cairns-Blake-Dowd model), because Models M1 and M5 do not incorporate any cohort effect while Model M4 does not give an explicit A/P/C decomposition.

All of the model structures under consideration are subject to an identifiability problem (that is, the problem that the optimal parameters are not unique). For example, if $\beta_x^{(1)}$, $\beta_x^{(2)}$, $\beta_x^{(3)}$, $\kappa_t^{(2)}$ and γ_c, for $x=x_0,\ldots,x_1$, $t=t_0,\ldots,t_1$, $c=t_0-x_1,\ldots,t_1-x_0$, are parameters for Model M2, then

$$\beta_x^{(1,*)}=\beta_x^{(1)}-\frac{a_1}{a_2}\beta_x^{(2)}-\frac{a_4}{a_3}\beta_x^{(3)},\quad \beta_x^{(2,*)}=\frac{1}{a_2}\beta_x^{(2)},\quad \beta^{(3,*)}=\frac{1}{a_3}\beta_x^{(3)},$$

$$\kappa_t^{(2,*)}=a_1+a_2\kappa_t^{(2)}\quad\text{and}\quad \gamma_c^*=a_3\gamma_c+a_4,$$

where a_1, a_2, a_3 and a_4 are arbitrary constants, are equivalent parameters for Model M2, because

$$\begin{aligned}&\beta_x^{(1,*)}+\beta_x^{(2,*)}\kappa_t^{(2,*)}+\beta_x^{(3,*)}\gamma_c^*\\ =\ &\left(\beta_x^{(1)}-\frac{a_1}{a_2}\beta_x^{(2)}-\frac{a_4}{a_3}\beta_x^{(3)}\right)+\frac{1}{a_2}\beta_x^{(2)}(a_1+a_2\kappa_t^{(2)})+\frac{1}{a_3}\beta_x^{(3)}(a_3\gamma_c+a_4)\\ =\ &\beta_x^{(1)}+\beta_x^{(2)}\kappa_t^{(2)}+\beta_x^{(3)}\gamma_c.\end{aligned}$$

To stipulate parameter uniqueness, we need to impose some identifiability constraints. The identifiability constraints used for each candidate model are shown in Table 1. Note that n_c in the table stands for the number of data points associated with year-of-birth c.[4]

[4]Our data sample spans ages 20 to 95 and years 1968 to 2014. For the earliest cohort ($c=1968-95=1873$), we have $n_{1873}=1$ as there is only one data point (which corresponds to age 95 and year 1968) associated with this cohort. Likewise, for the second earliest cohort $c=1874$, we have $n_{1874}=2$ as there are two data points (one of which corresponds to age 94 and year 1968 and the other of which corresponds to age 95 and year 1969) associated with this cohort. The values of n_c for other cohorts can be deduced in a similar manner.

The identifiability problem reflects the fact that there are infinitely many possible A/P/C decompositions, even if the model for decomposition purposes is fixed. For any model, the age, period and cohort parameters have no meaning in isolation.

In particular, the cohort and period components are defined with reference to each other. For all candidate models (except M2 and M8), the chosen identifiability constraints ensure that the cohort component will fluctuate around zero with no observable trend.[5] Therefore, for each of these models, the identified period component is the one that is conditioned on a cohort component that exhibits no trend, and the identified cohort component is the one that is conditioned on a period component that absorbs the entire trend.

5.2 Estimation

We consider two methods for estimating Route A candidate models. It is found that the two methods yield very similar parameter estimates.[6]

5.2.1 A Two-Staged Method

In the first method, we first calculate smoothed historical mortality improvement rates by applying a two-dimensional (age/cohort) P-splines regression to the raw central rates of death. The smoothed mortality improvement rate at age x and in year t is denoted by $\tilde{Z}_{x,t}$. Then, we obtain parameter estimates by minimizing the sum of squared differences between $\tilde{Z}_{x,t}$ and the parameterized structure, subject to the identifiability constraints shown in Table 1; for example, for Model M3, parameter estimates are obtained by minimizing

$$\sum_{x=x_0}^{x_1}\sum_{t=t_0}^{t_1}(\tilde{Z}_{x,t}-\beta_x^{(1)}-\kappa_t^{(1)}-\gamma_c)^2,$$

subject to $\sum_{x=x_0}^{x_1}\beta_x^{(1)}=0$, $\sum_{c=t_0-x_1}^{t_1-x_0}\gamma_c=0$, and $\sum_{c=t_0-x_1}^{t_1-x_0}c\gamma_c=0$. The minimization is solved by using an iterative Newton's method, in which parameters are updated one at a time until convergence is achieved.[7] We remark that this estimation method is also used in Working Papers 38 and 39 produced by the CMI (CMI, 2009a,b).

5.2.2 The Penalized Least Squares Method

The second method (penalized least squares) is extended from the work of Delwarde et al. (2007). It integrates parameter smoothing and estimation in one single process. In this method, model

[5] We refer interested readers to Cairns et al. (2009) for an explanation as to why the identifiability constraints we use ensure that the resulting cohort component will fluctuate around zero and will exhibit no linear and/or quadratic trend.

[6] We refer interested readers to Section 8 of Volume 1 of the full report of this study (Li et al., 2018a) for a comparison between the two methods.

[7] At the end of each iteration (when all of the parameters are updated), the parameters are rescaled so that the applicable identifiability constraints are satisfied.

parameters are estimated by minimizing an objective function, which is formulated as the sum of the squared errors plus the roughness penalty for each parameter series. The sum of squared errors measures the goodness-of-fit, whereas the roughness penalty for a parameter series measures the jaggedness of the parameter series.

We now explain the method of penalized least squares by way of an example. Let us consider Model M3, which encompasses three parameter series: $\{\beta_x^{(1)}\}$, $\{\kappa_t^{(1)}\}$ and $\{\gamma_c\}$. When the method of penalized least squares is applied to Model M3, estimates of the model parameters are obtained by minimizing the following objective function:

$$\sum_{x=x_0}^{x_1}\sum_{t=t_0}^{t_1}(Z_{x,t}-\beta_x^{(1)}-\kappa_t^{(1)}-\gamma_c)^2+(\boldsymbol{\beta}^{(1)})'\mathbf{P}_\beta(\boldsymbol{\beta}^{(1)})+(\boldsymbol{\kappa}^{(1)})'\mathbf{P}_\kappa(\boldsymbol{\kappa}^{(1)})+(\boldsymbol{\gamma})'\mathbf{P}_\gamma(\boldsymbol{\gamma}),$$

subject to the identifiability constraints stated in Section 5.1. In the objective function, $\boldsymbol{\beta}^{(1)}$, $\boldsymbol{\kappa}^{(1)}$ and $\boldsymbol{\gamma}$ are the vectors of $\beta_x^{(1)}$'s, $\kappa_t^{(1)}$'s and γ_c's, respectively,

$$\mathbf{P}_\beta=\pi_\beta\boldsymbol{\Delta}_\beta'\boldsymbol{\Delta}_\beta,\quad\text{with}\quad\boldsymbol{\Delta}_\beta=\begin{pmatrix}1&-2&1&0&\cdots&0\\0&1&-2&1&\cdots&0\\\vdots&\ddots&\ddots&\ddots&\ddots&\vdots\\0&\cdots&1&-2&1&0\\0&\cdots&0&1&-2&1\end{pmatrix}_{(n_a-2)\times n_a}$$

is the roughness penalty for the age dimension,

$$\mathbf{P}_\kappa=\pi_\kappa\boldsymbol{\Delta}_\kappa'\boldsymbol{\Delta}_\kappa,\quad\text{with}\quad\boldsymbol{\Delta}_\kappa=\begin{pmatrix}1&-2&1&0&\cdots&0\\0&1&-2&1&\cdots&0\\\vdots&\ddots&\ddots&\ddots&\ddots&\vdots\\0&\cdots&1&-2&1&0\\0&\cdots&0&1&-2&1\end{pmatrix}_{(n_y-2)\times n_y}$$

is the roughness penalty for the time dimension,

$$\mathbf{P}_\gamma=\pi_\gamma\boldsymbol{\Delta}_\gamma'\boldsymbol{\Delta}_\gamma,\quad\text{with}\quad\boldsymbol{\Delta}_\gamma=\begin{pmatrix}1&-2&1&0&\cdots&0\\0&1&-2&1&\cdots&0\\\vdots&\ddots&\ddots&\ddots&\ddots&\vdots\\0&\cdots&1&-2&1&0\\0&\cdots&0&1&-2&1\end{pmatrix}_{(n_y+n_a-3)\times(n_y+n_a-1)}$$

is the roughness penalty for cohort dimension, and π_β, π_κ and π_γ are the smoothing parameters for $\{\beta_x^{(1)}\}$, $\{\kappa_t^{(1)}\}$ and $\{\gamma_c\}$, respectively. For fixed values of π_β, π_κ and π_γ, the objective function can be minimized using an iterative Newton's method.

The specifications of the $\boldsymbol{\Delta}$ matrices indicate that the degree of roughness is measured by the sum of squared second-order differences (e.g., the sum of $(\beta_{x+2}^{(1)}-2\beta_{x+1}^{(1)}+\beta_x^{(1)})^2$). In effect, we

are using a second-order polynomial as the standard of smoothness.[8]

The tradeoff between goodness-of-fit and smoothness is determined by the smoothing parameters π_β, π_κ and π_γ. Following Delwarde et al. (2007), the smoothing parameters are carefully selected by a leave-one-out cross-validation, which involves the following steps:

(i) Given the values of the smoothing parameters π_β, π_κ and π_γ, re-estimate the age, period and cohort components to a pseudo data set with one observation being intentionally left out; use the re-estimated model to 'predict' the corresponding missing observation; record the 'prediction error':

$$(e^-_{x,t})^2 = (Z_{x,t} - \hat{\beta}^{(1,-)}_x - \hat{\kappa}^{(1,-)}_t - \hat{\gamma}^-_c)^2,$$

where $\hat{\beta}^{(1,-)}_x$, $\hat{\kappa}^{(1,-)}_t$ and $\hat{\gamma}^-_c$ respectively represent the estimated values of $\beta^{(1)}_x$, $\kappa^{(1)}_t$ and γ_c that are calculated from the pseudo data set (in which the observations at age x and in year t are excluded)

(ii) Repeat step (i) by leaving out one other observation at a time until all observations in the full data set have been considered. The sum of squared prediction errors (SPE) can be calculated by

$$\text{SPE}(\pi_\beta, \pi_\kappa, \pi_\gamma) = \sum_{x=x_0}^{x_1} \sum_{t=t_0}^{t_1} (e^-_{x,t})^2$$

(iii) Repeat steps (i) and (ii) for different possible values π_β, π_κ and π_γ; choose the values of π_β, π_κ and π_γ that lead to the smallest sum of squared prediction errors.

5.3 Robustness Tests

In this sub-section, we perform five robustness tests on the Route A candidate models, and shortlist a subset of them for further consideration.

5.3.1 The Robustness Tests Performed

Robustness Relative to Changes in the Tolerance Value

To obtain an A/P/C decomposition, the objective function specified in Section 5.2 is minimized with an iterative Newton's method. The iterations stop when the change in the value of the objective function is smaller than a pre-specified tolerance value. In this test, we examine how the estimation results may change when the tolerance value is varied. Three tolerance values are considered: 10^{-6}, 10^{-8}, and 10^{-10}.

[8] In principle, a higher order polynomial can be used as the standard of smoothness instead. We have experimented higher order polynomials, and found that the resulting age/period/cohort decompositions are similar.

Model	Identifiability constraints
Model M2	$\sum_{x=x_0}^{x_1} \beta_x^{(1)} = 0,\ \sum_{x=x_0}^{x_1} \beta_x^{(2)} = 1,\ \sum_{x=x_0}^{x_1} \beta_x^{(3)} = 1,$ $\sum_{c=t_0-x_1}^{t_1-x_0} \gamma_c = 0$
Model M3	$\sum_{x=x_0}^{x_1} \beta_x^{(1)} = 0,\ \sum_{c=t_0-x_1}^{t_1-x_0} \gamma_c = 0,\ \sum_{c=t_0-x_1}^{t_1-x_0} n_c c\gamma_c = 0$
Model M6	$\sum_{c=t_0-x_1}^{t_1-x_0} \gamma_c = 0,\ \sum_{c=t_0-x_1}^{t_1-x_0} n_c c\gamma_c = 0.$
Model M7	$\sum_{c=t_0-x_1}^{t_1-x_0} \gamma_c = 0,\ \sum_{c=t_0-x_1}^{t_1-x_0} n_c c\gamma_c = 0,\ \sum_{c=t_0-x_1}^{t_1-x_0} n_c c^2\gamma_c = 0$
Model M8	$\sum_{c=t_0-x_1}^{t_1-x_0} \gamma_c = 0$
Full Plat	$\sum_{x=x_0}^{x_1} \beta_x^{(1)} = 0,\ \sum_{t=t_0}^{t_1} \kappa_t^{(2)} = 0,\ \sum_{t=t_0}^{t_1} \kappa_t^{(3)} = 0,$ $\sum_{c=t_0-x_1}^{t_1-x_0} \gamma_c = 0, \sum_{c=t_0-x_1}^{t_1-x_0} n_c c\gamma_c = 0, \sum_{c=t_0-x_1}^{t_1-x_0} n_c c^2\gamma_c = 0$
Simplified Plat	$\sum_{x=x_0}^{x_1} \beta_x^{(1)} = 0,\ \sum_{t=t_0}^{t_1} \kappa_t^{(2)} = 0,\ \sum_{c=t_0-x_1}^{t_1-x_0} n_c\gamma_c = 0,$ $\sum_{c=t_0-x_1}^{t_1-x_0} n_c c\gamma_c = 0,\ \sum_{c=t_0-x_1}^{t_1-x_0} n_c c^2\gamma_c = 0$

Table 2: The alternative identifiability constraints used in estimating model parameters

Robustness Relative to Changes in the Calibration Window

In this test, we examine how the estimation results may change when the calibration window becomes different. We consider the three calibration windows, which have the same length (37 years) but different starting/ending years: 1968-2004, 1973-2009, 1978-2014. This set-up mimics the situation when the models are updated every five years.

Robustness Relative to Changes in the Age Range

In this test, we examine how the estimation results may change when the age range used is altered. The following three age ranges are considered: 20-95 (76 ages), 30-85 (56 ages), and 40-75 (36 ages).

Robustness Relative to Changes in the Identifiability Constraints

In this test, we examine how the estimation results may change when the identifiability constraints are slightly modified. We consider two sets of identifiability constraints: the baseline constraints shown in Table 1 and the alternative constraints presented in Table 2.

The difference between the baseline and alternative constraints lies in the inclusion/exclusion of n_c, which represents the number of data points that are related to year-of-birth c. By including n_c, the cohorts about which we have more information are weighted more heavily in the parameter constraints.

We emphasize that, as discussed in Section 5.1, there are many other combinations of constraints that can be used to stipulate parameter uniqueness. When very different constraints are used, the patterns of the resulting A/P/C components may be very different and may carry different interpretations. Our goal here is to examine the impact of small changes in the constraints

used only.

Robustness Relative to the Inclusion/Exclusion of the Oldest and Newest Cohorts

In this test, we examine how the estimation results may change when the oldest and newest cohorts in the data set are included/excluded. The following three situations are considered.

1. All available data are used.
2. The oldest and youngest five cohorts in the data sample are excluded.
3. The oldest and youngest ten cohorts in the data sample are excluded.

5.3.2 The Robustness Measure

We define the following measure of robustness for the candidate models in Route A:

$$\text{robustness} = \frac{\max_i(\text{maximum absolute change in the } i\text{-th model term})}{\max_{x,t}(\tilde{Z}_{x,t}) - \min_{x,t}(\tilde{Z}_{x,t})}, \tag{3}$$

where $\max_{x,t}(\tilde{Z}_{x,t})$ and $\min_{x,t}(\tilde{Z}_{x,t})$ represent the maximum and minimum values of the smoothed mortality improvement rates in the dataset, respectively, and 'the i-th model term' refers to the i-th term on the right-hand-side of the model structure (excluding the error term). The denominator $\max_{x,t}(\tilde{Z}_{x,t}) - \min_{x,t}(\tilde{Z}_{x,t})$ 'standardizes' the robustness measure by considering the variability of the data being fed into the model.

To illustrate, let us consider testing the robustness of Model M6 relative to changes in the calibration window. We consider three calibration windows: (a) 1968-2004, (b) 1973-2009, and (c) 1978-2014. Recall that Model M6 has the following structure:

$$Z_{x,t} = \kappa_t^{(1)} + \kappa_t^{(2)}(x - \bar{x}) + \gamma_c + e_{x,t},$$

which contains three model terms, namely $\kappa_t^{(1)}$, $\kappa_t^{(2)}(x - \bar{x})$ and γ_c. The following steps are taken to obtain the relevant robustness measure:

1. Since the maximum and minimum values of the smoothed mortality improvement rates over year 1968 to 2014 and age 20 to 95 are 0.0860 and -0.0530, respectively, the denominator of equation (7) is calculated as $0.0860 - (-0.0530) = 0.1363$.
2. Fit Model M6 using calibration window (a) and obtain estimates of the three model terms: $\kappa_t^{(1,\text{a})}$, $\kappa_t^{(2,\text{a})}(x - \bar{x})$, and $\gamma_c^{(\text{a})}$.
3. Re-fit Model M6 using calibration window (b) and obtain new estimates of the three model terms: $\kappa_t^{(1,\text{b})}$, $\kappa_t^{(2,\text{b})}(x - \bar{x})$, and $\gamma_c^{(\text{b})}$.

4. Compare the estimation results from calibration windows (a) and (b), and calculate the maximum change in each of the three model terms for this comparison:

$$\begin{aligned}
\max_{t=1974,\ldots,2004} |\kappa_t^{(1,\mathrm{a})} - \kappa_t^{(1,\mathrm{b})}| &= 0.0017; \\
\max_{x=20,\ldots,95; t=1974,\ldots,2004} |\kappa_t^{(2,\mathrm{a})}(x-\bar{x}) - \kappa_t^{(2,\mathrm{b})}(x-\bar{x})| &= 0.0036; \\
\max_{c=1879,\ldots,1984} |\gamma_c^{(\mathrm{a})} - \gamma_c^{(\mathrm{b})}| &= 0.0040.
\end{aligned}$$

5. Re-fit Model M6 using calibration window (c) and obtain new estimates of the three model terms: $\kappa_t^{(1,\mathrm{c})}$, $\kappa_t^{(2,\mathrm{c})}(x-\bar{x})$, and $\gamma_c^{(\mathrm{c})}$.

6. Compare the estimation results from calibration windows (a) and (c), and calculate the maximum change in each of the three model terms for this comparison:

$$\begin{aligned}
\max_{t=1979,\ldots,2004} |\kappa_t^{(1,\mathrm{a})} - \kappa_t^{(1,\mathrm{c})}| &= 0.0027; \\
\max_{x=20,\ldots,95; t=1979,\ldots,2009} |\kappa_t^{(2,\mathrm{a})}(x-\bar{x}) - \kappa_t^{(2,\mathrm{c})}(x-\bar{x})| &= 0.0045; \\
\max_{c=1884,\ldots,1984} |\gamma_c^{(\mathrm{a})} - \gamma_c^{(\mathrm{c})}| &= 0.0071.
\end{aligned}$$

7. Compare the estimation results from calibration windows (b) and (c), and calculate the maximum change in each of the three model terms for this comparison:

$$\begin{aligned}
\max_{t=1979,\ldots,2009} |\kappa_t^{(1,\mathrm{b})} - \kappa_t^{(1,\mathrm{c})}| &= 0.0018; \\
\max_{x=20,\ldots,95; t=1979,\ldots,2009} |\kappa_t^{(2,\mathrm{b})}(x-\bar{x}) - \kappa_t^{(2,\mathrm{c})}(x-\bar{x})| &= 0.0030; \\
\max_{c=1884,\ldots,1989} |\gamma_c^{(\mathrm{b})} - \gamma_c^{(\mathrm{c})}| &= 0.0059.
\end{aligned}$$

8. Calculate the overall maximum change in each of the model term:

$$\begin{aligned}
\text{maximum change in } \kappa_t^{(1)} &= \max(0.0017, 0.0027, 0.0018) = 0.0027; \\
\text{maximum change in } \kappa_t^{(2)} &= \max(0.0036, 0.0045, 0.0020) = 0.0045; \\
\text{maximum change in } \gamma_c &= \max(0.0040, 0.0071, 0.0059) = 0.0071.
\end{aligned}$$

9. Set the maximum of the three values obtained in Step (8) as the numerator in equation (7), and calculate the robustness:

$$\frac{\max(0.0027, 0.0045, 0.0071)}{0.1363} = 5.2\%.$$

In the following robustness tests, we rate the robustness of the candidate models using the following criteria:

- High robustness: $0 \leq$ robustness measure $\leq 10\%$
- Medium robustness: $10\% <$ robustness measure $\leq 20\%$
- Low robustness: robustness measure $> 20\%$

5.3.3 Test Results

The test results are reported in Table 3. To demonstrate the differences among the three levels of robustness, we show in Figures 3, 4 and 5 the estimation results for models that exhibit low, medium and high levels of robustness relative to changes in the tolerance value, respectively. For the model with a high level of robustness, we observe that the three tolerance values lead to almost identical parameter estimates. For the model with a low level of robustness, the parameter estimates based on different levels of tolerance are highly different.

Robustness Test	Model M2	Model M3	Model M6	Model M7	Model M8	Full Plat	Simplified Plat
SSA males							
Tolerance value	19.8% Medium	0.0% High	0.1% High	0.2% High	0.4% High	6.2% High	0.0% High
Calibration window	136.3% Low	11.1% Medium	5.2% High	24.9% Low	5.8% High	24.2% Low	10.8% Medium
Age range	80.3% Low	11.1% Medium	11.8% Medium	13.0% Medium	14.1% Medium	27,4% Low	11.1% Medium
Parameter constraints	24.0% Low	1.6% High	4.9% High	11.9% Medium	0.4% High	7.8% High	6.7% High
Exclusion of cohorts	38.7% Low	4.1% High	4.7% High	12.1% Medium	2.7% High	8.3% High	7.9% High
SSA females							
Tolerance value	131.1% Low	0.0% High	0.1% High	0.2% High	21.1% Medium	10.5% Medium	0.0% High
Calibration window	62.0% Low	9.8% High	9.2% High	12.9% Medium	31.6% Low	75.8% Low	9.4% High
Age range	120.0% Low	16.4% Medium	13.9% Medium	19.4% Medium	75.4% Low	60.3% Low	13.9% Medium
Parameter constraints	129.5% Low	4.0% High	1.3% High	6.1% High	12.3% Medium	13.6% Medium	3.0% High
Exclusion of cohorts	144.8% Low	4.1% High	6.1% High	8.7% High	15.4% Medium	15.3% Medium	7.1% High

Table 3: The results of the robustness tests performed for Route A candidate models

We remark here that the low robustness of Models M2 and M8 to changes in the tolerance value is due possibly to the convergence problem. The optimization for Models M2 and M8 often converges very slowly. For instance, when applying M2 to the SSA female dataset, it takes 7,276 iterations to reach the tolerance value of 10^{-8}.[9] In addition, we have to bound the absolute value of the cohort component γ_c in these models, or otherwise the optimization may never converge.

From Table 3 we observe that Model M2, Model M7, Model M8 and the full Plat model show low robustness in some tests. Therefore, we only consider Model M3, Model M6 and the simplified

[9]Typically, it takes less than 100 iterations to reach convergence.

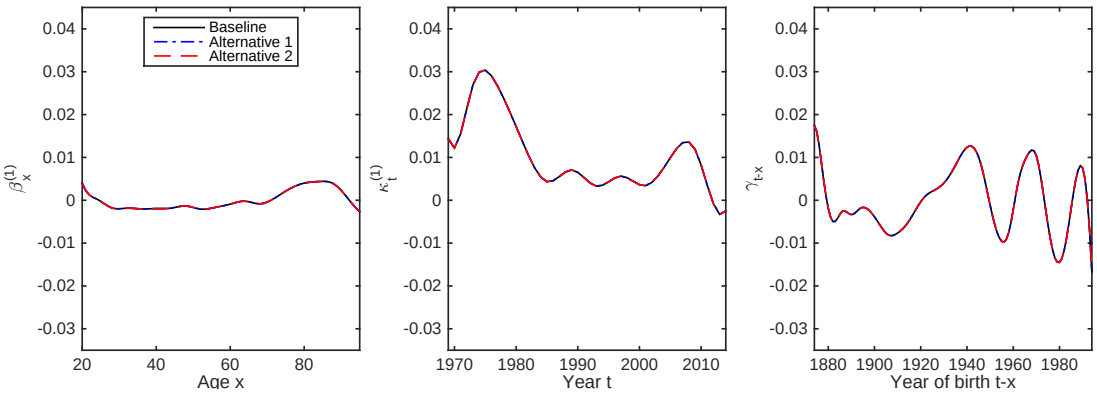

Figure 3: An example of high robustness to the tolerance value used: The estimated parameters in the Route A Model M3 fitted to SSA female data, using tolerance values of 10^{-6}, 10^{-8} and 10^{-10}.

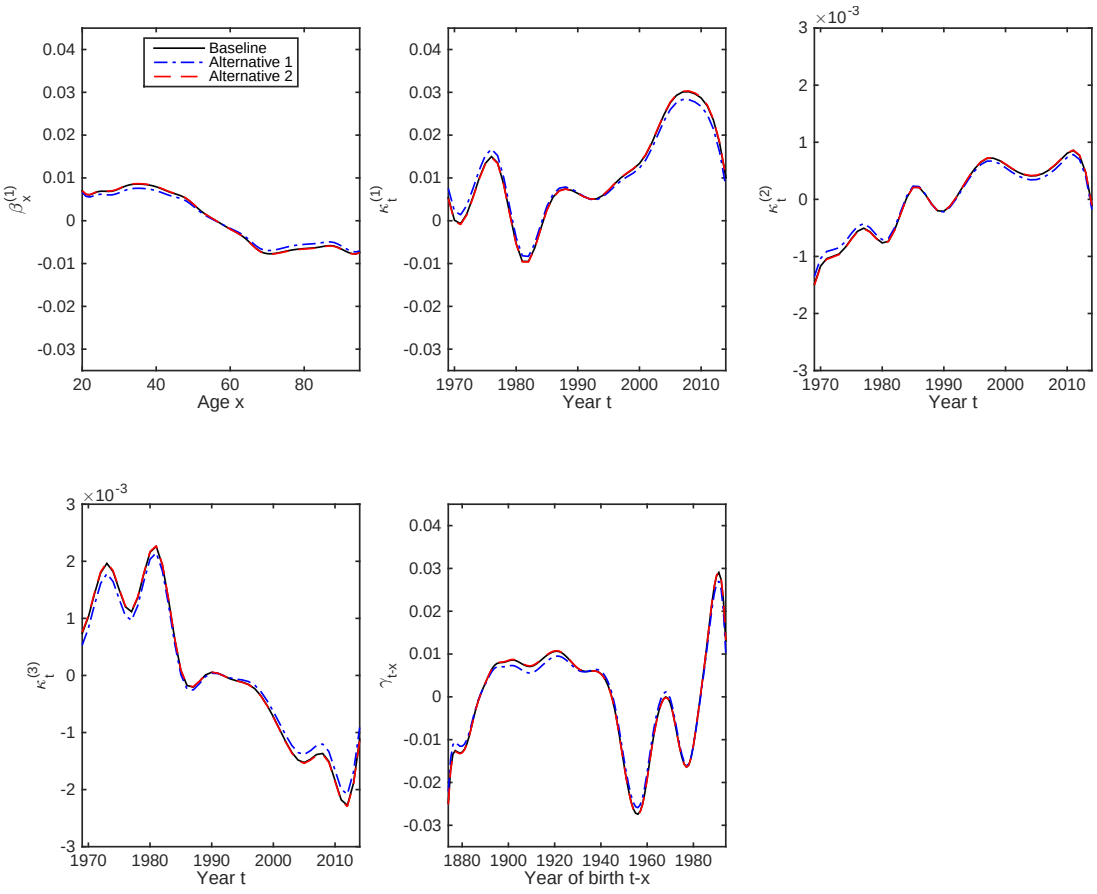

Figure 4: An example of medium robustness to the tolerance value used: The estimated parameters in the Route A full Plat model fitted to SSA female data, using tolerance values of 10^{-6}, 10^{-8} and 10^{-10}.

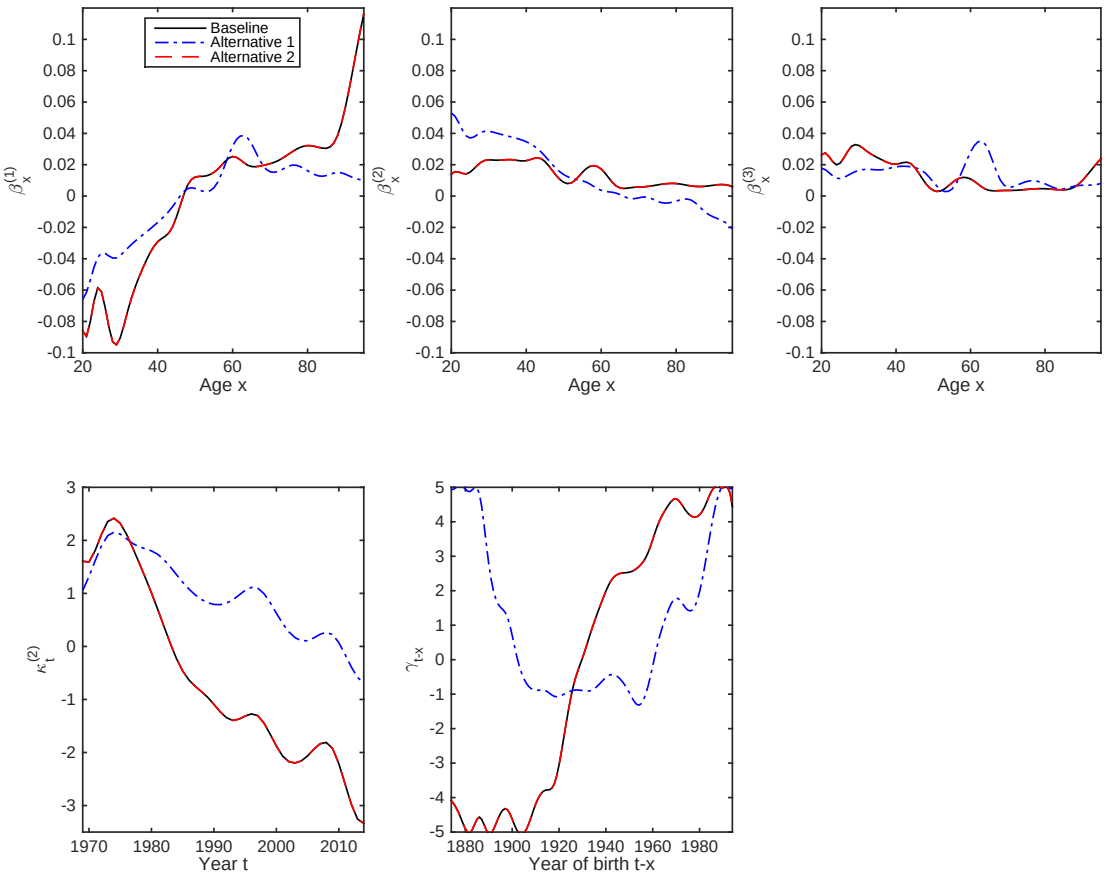

Figure 5: An example of low robustness to the tolerance value used: The estimated parameters in the Route A Model M2 fitted to SSA female data, using tolerance values of 10^{-6}, 10^{-8} and 10^{-10}.

Plat model in the next stage of the model selection process.

5.4 Analyzing the Standardized Residuals for the Shortlisted Models

We now analyze the standardized residuals produced by Model M3, Model M6 and the simplified Plat model, which show high to medium levels of robustness in the tests performed in the previous sub-section. For models in Route A, the standardized residual for age x and year t is defined as

$$\frac{\hat{e}_{x,t}}{\text{s.d.}(\hat{e}_{x,t})},$$

where s.d.$(\hat{e}_{x,t})$ is the sample standard deviation of $\hat{e}_{x,t}$ over the entire age range and sample period, and $\hat{e}_{x,t}$ represents the estimated (non-standardized) residual for age x and year t.

First, we perform the Anderson-Darling test, of which the null hypothesis is that standardized residuals come from a standard normal distribution. Based on the p-values reported in Table 4, the following conclusions are drawn.

1. The standardized residuals from Model M6 for SSA males and females do not pass the test at a 5% significance level.[10]

2. The p-values indicate that the standardized residuals from the simplified Plat model are more 'normal' than those from Model M3.

Next, we calculate the standardized residuals' descriptive statistics (mean, standard deviation, skewness and kurtosis). If the standardized residuals follow a standard normal distribution, their mean should be close to 0, standard deviation should be close to 1, skewness should be close to 0, and kurtosis should be close to 3. The values of the descriptive statistics (reported in Table 4) also suggest that the residuals from the simplified Plat model are more in line with the standard normal distribution compared to those from Model M3.

Then, we consider the Q-Q plots of the standardized residuals from the three shortlisted models (Figures 6, 7 and 8). A Q-Q plot is a plot of the quantiles of two distributions against each other. In each of the Q-Q plots shown, we compare the empirical distribution of the standardized residuals against the standard normal distribution. If the standardized residuals follow the standard normal distribution, then the points in the Q-Q plot should lie on the 45° line. The Q-Q plots indicate that the distribution of the standardized residuals from Model M3 is quite different from the standard normal distribution.

Finally, we examine the heat maps of the standardized residuals from the three shortlisted models (Figure 9). If an estimated model is adequate, then its standardized residuals should show a random pattern. The heat maps suggest that Model M3 seems to be adequate for U.S. females, but not quite for U.S. males. In Model M3's residuals heat map for U.S. males, we observe three

[10] At a 5% significance level, we reject the null hypothesis if p-value is smaller than 0.05; that is, we conclude that the standardized residuals do not follow the standard normal distribution if the p-value is less than 0.05.

	Dataset	Model M3	Model M6	Simplified Plat
p-value	SSA males	0.0890	0.0069	0.3886
	SSA females	0.0680	0.0258	0.4436
Mean	SSA males	-0.0008	0.0282	-0.0010
	SSA females	-0.0013	0.0293	-0.0005
Standard deviation	SSA males	0.9896	1.0070	0.9902
	SSA females	0.9898	1.0256	0.9904
Skewness	SSA males	-0.0094	-0.0202	-0.0519
	SSA females	0.0182	-0.0275	-0.0337
Kurtosis	SSA males	2.8572	2.7054	2.8789
	SSA females	3.4885	2.9974	3.1917

Table 4: The p-value (for the Anderson-Darling test), mean, standard deviation, skewness, and kurtosis of the standardized residuals: Model M3, Model M6 and the simplified Plat model.

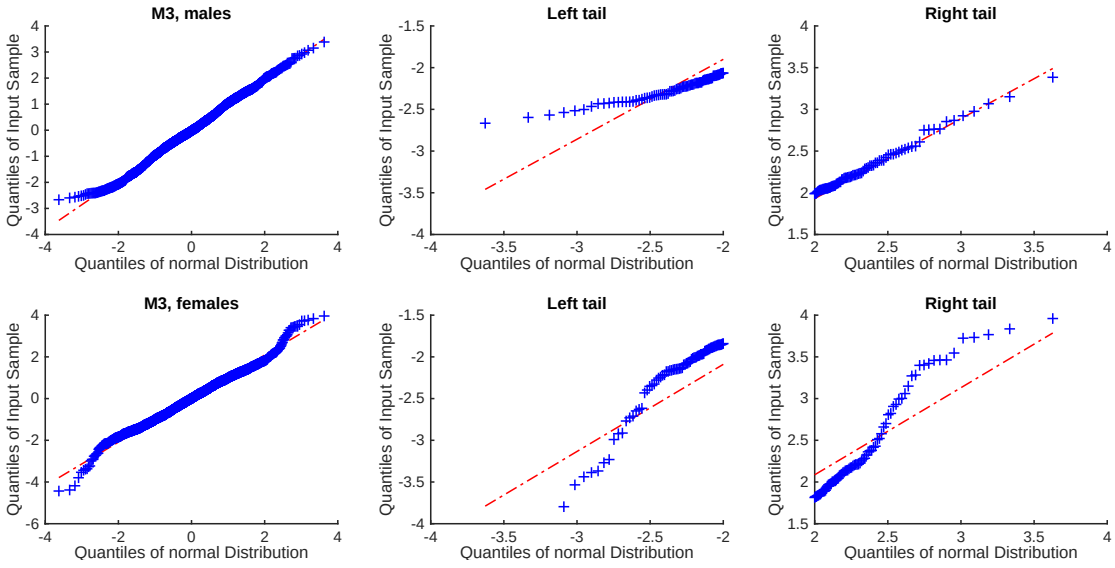

Figure 6: The Q-Q plots of standardized residuals obtained from the Route A Model M3 fitted to the SSA dataset.

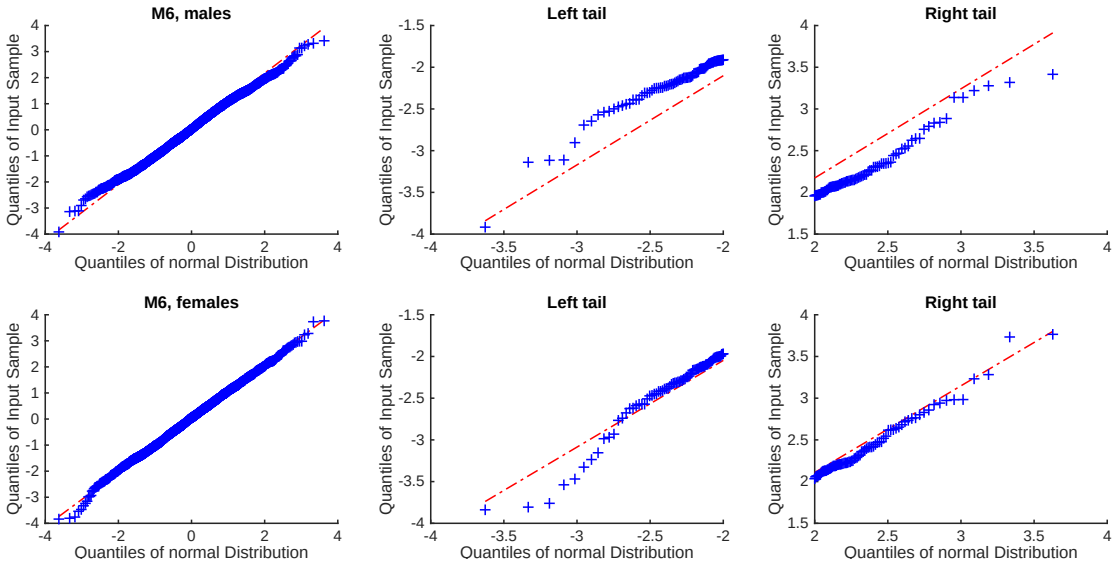

Figure 7: The Q-Q plots of standardized residuals obtained from the Route A Model M6 fitted to the SSA dataset.

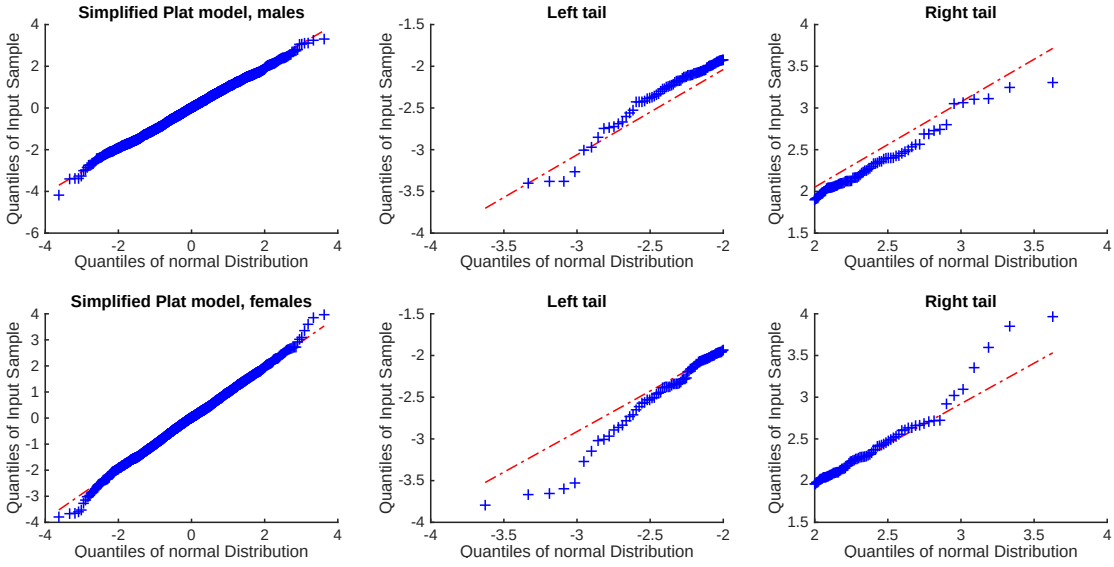

Figure 8: The Q-Q plots of standardized residuals obtained from the Route A simplified Plat model fitted to the SSA dataset.

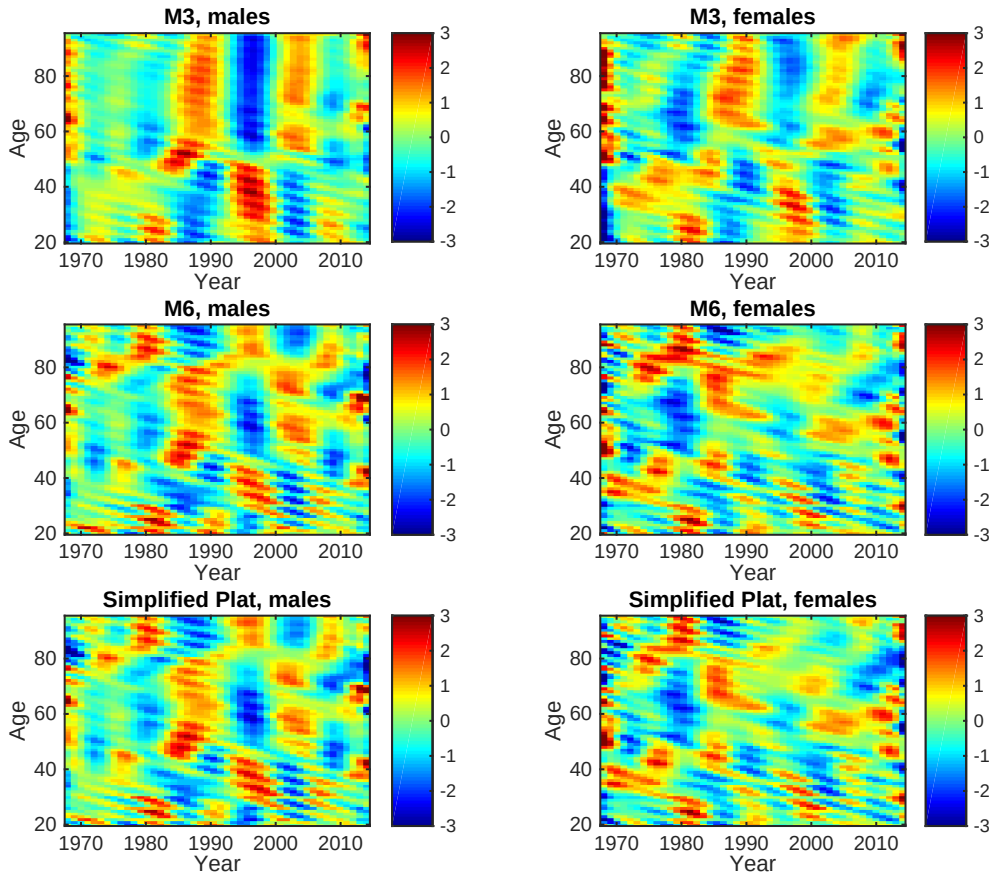

Figure 9: Heatmaps of the standardized residuals produced by the shortlisted Route A models.

large vertical bands over the period of 1985 to 2005. These residual clusters suggest that some characteristics that are specific to the historical mortality of U.S. males have not been picked up by Model M3. This problem is much less significant in the residual heat maps from the simplified Plat model for U.S. males.

Summing up, Model M3 and the simplified Plat model are the only two models that pass the Anderson-Darling normality test, but Model M3 does not perform well in terms of the descriptive statistics, residual heatmaps and Q-Q plots. We therefore conclude that the simplified Plat model is the most effective model in Route A.

6 Route B

This section is devoted to the modeling work associated with Route B. We first define the Route B candidate models (which are applied to death rates/probabilities instead of mortality improvement rates), and derive the A/P/C components of historical mortality improvements implied by each of the candidate models. We then describe the estimation procedure, the robustness tests (for shortlisting the candidate models), and the residual analysis (for identifying the most effective

model among the shortlisted models).

6.1 The Candidate Models

In Route B, we consider eight candidate models, which are applied to either $m_{x,t}$'s or $q_{x,t}$'s. Using the notation defined in Section 3, the Route B candidate models are defined as follows.

- M2 – The Renshaw-Haberman Model (Renshaw and Haberman, 2006)

$$\ln m_{x,t} = \beta_x^{(1)} + \beta_x^{(2)}\kappa_t^{(2)} + \beta_x^{(3)}\gamma_c$$

- M3 – The Age-Period-Cohort Model (Osmond, 1985)

$$\ln m_{x,t} = \beta_x^{(1)} + \kappa_t^{(1)} + \gamma_c$$

- M6 – The CBD Model with a Cohort Effect (Cairns et al., 2009)

$$\ln \frac{q_{x,t}}{1-q_{x,t}} = \kappa_t^{(1)} + \kappa_t^{(2)}(x-\bar{x}) + \gamma_c$$

- M7 – The CBD Model with Quadratic Age and Cohort Effects (Cairns et al., 2009)

$$\ln \frac{q_{x,t}}{1-q_{x,t}} = \kappa_t^{(1)} + \kappa_t^{(2)}(x-\bar{x}) + \kappa_t^{(3)}((x-\bar{x})^2 - \hat{\sigma}_x^2) + \gamma_c$$

- M8 – The CBD Model with an Age-Dependent Cohort Effect (Cairns et al., 2009)

$$\ln \frac{q_{x,t}}{1-q_{x,t}} = \kappa_t^{(1)} + \kappa_t^{(2)}(x-\bar{x}) + \gamma_c(x_c - x)$$

- The Full Plat model (Plat, 2009):

$$\ln m_{x,t} = \beta_x^{(1)} + \kappa_t^{(1)} + \kappa_t^{(2)}(\bar{x}-x) + \kappa_t^{(3)}(\bar{x}-x)_+ + \gamma_c$$

- The Simplified Plat model (Plat, 2009):

$$\ln m_{x,t} = \beta_x^{(1)} + \kappa_t^{(1)} + \kappa_t^{(2)}(\bar{x}-x) + \gamma_c$$

- The APCI model (CMI, 2017a,b,c):

$$\ln(m_{x,t}) = \beta_x^{(1)} + \beta_x^{(2)}(t-\bar{t}) + \kappa_t^{(1)} + \gamma_c.$$

We remark that the APCI model is the model recommended by the Continuous Mortality Investigation (CMI) Bureau of the Institute and Faculty of Actuaries in their Working Paper 97 for decomposing historical mortality improvements in the United Kingdom.

The Route B candidate models also require identifiability constraints to stipulate unique parameters. For the APCI model, the following identifiability constraints are used:

$$\sum_{c=t_0-x_1}^{t_1-x_0} \gamma_c = 0, \sum_{c=t_0-x_1}^{t_1-x_0} c\gamma_c = 0, \sum_{c=t_0-x_1}^{t_1-x_0} c^2\gamma_c = 0, \sum_{t=t_0}^{t_1} \kappa_t^{(1)} = 0 \quad \text{and} \quad \sum_{t=t_0}^{t_1} t\kappa_t^{(1)} = 0.$$

For the other Route B candidate models, the identifiability constraints used are the same as those for their Route A counterparts (see Table 1).

In contrast to the Route A models, the definitions of the Route B models do not contain an error term. The absence of the error term is because a Poisson error structure is used in estimating the Route B models (see Section 6.3), and in this estimation framework the estimation error is captured by the difference between the actual and expected (Poisson) death counts.

Parameters $\beta_x^{(i)}$, $\kappa_t^{(i)}$ and γ_c have different interpretations in the two pathways. In Route A, as the models are fitted to historical mortality improvement rates directly, parameters $\beta_x^{(i)}$, $\kappa_t^{(i)}$ and γ_c can be interpreted to mean an A/P/C decomposition of historical *mortality improvement rates*. In Route B, as the models are fitted to $m_{x,t}$'s or $q_{x,t}$'s, parameters $\beta_x^{(i)}$, $\kappa_t^{(i)}$ and γ_c should be regarded as the A/P/C components of historical *mortality rates*, and some transformations are needed in order to obtain an A/P/C decomposition of historical *mortality improvement rates*. The required transformations are explained in the next sub-section.

6.2 The Implied Mortality Improvement Rates

The following symbols are defined to represent the A/P/C components of historical *mortality improvement rates* implied by the Route B candidate models:

- $B_x^{(i)}$, $i = \cdot, K, G$ are the age components of the mortality improvement rates:
  - $B_x^{(\cdot)}$ is a stand-alone age component,
  - $B_x^{(K)}$ is an age component that interacts with a period component,
  - $B_x^{(G)}$ is an age component that interacts with a cohort component;
- $K_t^{(i)}$, $i = 1, 2, 3$, are period components of the mortality improvement rates:
  - $K_t^{(1)}$ is a stand-alone period component;
  - $K_t^{(2)}$ is a period component that interacts with an age component or a linear function of age;
  - $K_t^{(3)}$ is a period component that interacts with a non-linear function of age.
- G_c is the cohort component of the mortality improvement rates.

To obtain neat and interpretable parameterizations, two versions of mortality improvement are used in Route B.

The first definition is based on the change in log central death rates, and is given by

$$MI_{x,t} = \ln m_{x,t-1} - \ln m_{x,t}. \tag{4}$$

We apply this definition of mortality improvement to Model M2, Model M3, the full Plat model, the simplified Plat model and the APCI model, as they are all constructed to model $\ln m_{x,t}$. The annual mortality improvement implied by these models can be expressed as follows.

- Model M2:

  $$MI_{x,t} = B_x^{(K)} K_t^{(2)} + B_x^{(G)} G_c,$$

  where $B_x^{(K)} = \beta_x^{(2)}$, $K_t^{(2)} = \kappa_{t-1}^{(2)} - \kappa_t^{(2)}$, $B^{(G)} = \beta_x^{(3)}$ and $G_c = \gamma_{c-1} - \gamma_c$.

  Comment: Model M2 in Route B implies that mortality improvement is driven by a period component, a cohort component, and two age components which interact with the period and cohort components, respectively.

- Model M3:

  $$MI_{x,t} = K_t^{(1)} + G_c,$$

  where $K_t^{(1)} = \kappa_{t-1}^{(1)} - \kappa_t^{(1)}$ and $G_c = \gamma_{c-1} - \gamma_c$.

  Comment: Model M3 in Route B implies that mortality improvement is driven by the sum of a period component and a cohort component. There is no age component.

- The Full Plat model:

  $$MI_{x,t} = K_t^{(1)} + K_t^{(2)}(\bar{x} - x) + K_t^{(3)}(\bar{x} - x)_+ + G_c,$$

  where $K_t^{(1)} = \kappa_{t-1}^{(1)} - \kappa_t^{(1)}$, $K_t^{(2)} = \kappa_{t-1}^{(2)} - \kappa_t^{(2)}$, $K_t^{(3)} = \kappa_{t-1}^{(3)} - \kappa_t^{(3)}$ and $G_c = \gamma_{c-1} - \gamma_c$.

  Comment: The full Plat model in Route B implies that mortality improvement is driven by three period components and a cohort component. Among the period components, one is stand-alone, one interacts with a linear function of age , and one interacts with a non-linear function of age.

- The Simplified Plat model:

  $$MI_{x,t} = K_t^{(1)} + K_t^{(2)}(\bar{x} - x) + G_c,$$

  where $K_t^{(1)} = \kappa_{t-1}^{(1)} - \kappa_t^{(1)}$, $K_t^{(2)} = \kappa_{t-1}^{(2)} - \kappa_t^{(2)}$ and $G_c = \gamma_{c-1} - \gamma_c$.

  Comment: The simplified Plat model in Route B implies that mortality improvement is driven by two period components and a cohort component. Among the period components, one is stand-alone and one interacts with a linear function of age.

- The APCI model:

  $$MI_{x,t} = B_x^{(\cdot)} + K_t^{(1)} + G_c,$$

  where $B_x^{(\cdot)} = -\beta_x^{(2)}$, $K_t^{(1)} = \kappa_{t-1}^{(1)} - \kappa_t^{(1)}$ and $G_c = \gamma_{c-1} - \gamma_c$.

  Comment: The APCI model in Route B implies that mortality improvement is driven by a (stand-alone) age component, a period component and a cohort component. There is neither age-period nor age-cohort interaction.

The second definition of mortality improvement is based on the change in logit-transformed[11] conditional death probabilities:

$$MI_{x,t} = \ln \frac{q_{x,t-1}}{1 - q_{x,t-1}} - \ln \frac{q_{x,t}}{1 - q_{x,t}}. \tag{5}$$

We use this definition of mortality improvement for Models M6, M7 and M8, because all of these models are constructed to model $\ln(q_{x,t}/(1 - q_{x,t}))$. The annual mortality improvement implied by these models can be expressed as follows:

- Model M6:

  $$MI_{x,t} = K_t^{(1)} + K_t^{(2)}(x - \bar{x}) + G_c,$$

  where $K_t^{(1)} = \kappa_{t-1}^{(1)} - \kappa_t^{(1)}$, $K_t^{(2)} = \kappa_{t-1}^{(2)} - \kappa_t^{(2)}$ and $G_c = \gamma_{c-1} - \gamma_c$.

  Comment: Model M6 in Route B implies that mortality improvement is driven by two period components and one cohort component. One period component is stand-alone, while the other period component interacts with a linear function of age.

- Model M7:

  $$MI_{x,t} = K_t^{(1)} + K_t^{(2)}(x - \bar{x}) + K_t^{(3)}((x - \bar{x})^2 - \hat{\sigma}_x^2) + G_c,$$

  where $K_t^{(1)} = \kappa_{t-1}^{(1)} - \kappa_t^{(1)}$, $K_t^{(2)} = \kappa_{t-1}^{(2)} - \kappa_t^{(2)}$, $K_t^{(3)} = \kappa_{t-1}^{(3)} - \kappa_t^{(3)}$ and $G_c = \gamma_{c-1} - \gamma_c$.

  Comment: Model M7 in Route B implies that mortality improvement is driven by three period components and one cohort component. Among the three period components, one is stand-alone, one interacts with a linear function of age, and one interacts with a quadratic function of age.

- Model M8:

  $$MI_{x,t} = K_t^{(1)} + K_t^{(2)}(x - \bar{x}) + G_c(x_c - x)$$

  where $K_t^{(1)} = \kappa_{t-1}^{(1)} - \kappa_t^{(1)}$, $K_t^{(2)} = \kappa_{t-1}^{(2)} - \kappa_t^{(2)}$ and $G_c = \gamma_{c-1} - \gamma_c$.

  Comment: Model M8 in Route B implies that mortality improvement is driven by two period components and one cohort component. Among the two period components, one is stand-alone and one interacts with a linear function of age. The cohort component interacts with a linear function of age.

[11] The logit transform of a quantity y is defined as $\ln(y/(1 - y))$.

6.3 Estimation

6.3.1 The Objective Function

We estimate and smooth parameters in each Route B candidate model simultaneously by minimizing the following objective function

$$\text{Objective} = \text{Deviance} + \sum_i \text{Penalty}(\beta_x^{(i)}) + \sum_j \text{Penalty}(\kappa_t^{(j)}) + \text{Penalty}(\gamma_c), \tag{6}$$

subject to the applicable identifiability constraints. The minimization is accomplished by using an iterative Newton's method.

The deviance in the objective function measures goodness-of-fit. The smaller the deviance is, the better the goodness-of-fit is. Let $d_{x,t}$ be the observed number of deaths at age x and in year t, and $E_{x,t}$ be the corresponding number of exposures-to-risk. Assuming that $d_{x,t}$ is a realization of a Poisson distribution with a mean of $E_{x,t}m_{x,t}$, the deviance can be calculated with the following formula:

$$\text{Deviance} = 2\sum_{x=x_0}^{x_1}\sum_{t=t_0}^{t_1}\left[d_{x,t}\ln\frac{d_{x,t}}{E_{x,t}m_{x,t}} - (d_{x,t} - E_{x,t}m_{x,t})\right].$$

The other three terms in the objective function are roughness penalty terms. The more jagged a parameter series is, the higher the penalty term for the parameter series is. The penalty terms can be calculated with the following formulas:

$$\text{Penalty}(\beta_x^{(i)}) = \lambda_{\beta_x^{(i)}}\sum_{x=x_0}^{x_1}(\beta_x^{(i)} - 3\beta_{x-1}^{(i)} + 3\beta_{x-2}^{(i)} - \beta_{x-3}^{(i)})^2,$$

$$\text{Penalty}(\kappa_t^{(j)}) = \lambda_{\kappa_t^{(j)}}\sum_{t=t_0}^{t_1}(\kappa_t^{(j)} - 2\kappa_{t-1}^{(j)} + \kappa_{t-2}^{(j)})^2$$

and

$$\text{Penalty}(\gamma_c) = \lambda_{\gamma_c}\sum_{c=t_0-x_1}^{t_1-x_0}(\gamma_c - 3\gamma_{c-1} + 3\gamma_{c-2} - \gamma_{c-3})^2,$$

where $\lambda_{\beta_x^{(i)}}$, $\lambda_{\kappa_t^{(j)}}$ and λ_{γ_c} are smoothing parameters which determine the degrees of smoothness of the parameters.

6.3.2 Choosing the Smoothing Parameters

The following procedure is used to select the smoothing parameter for $\beta_x^{(1)}$:

1. Set all smoothing parameters except $\lambda_{\beta_x^{(1)}}$ to zero.

2. Plot the estimated values of $\beta_x^{(1)}$ for $\lambda_{\beta_x^{(1)}} = 10^0, 10^1, 10^2, \ldots, 10^{10}$.

3. The optimal value of $\lambda_{\beta_x^{(1)}}$ should be the one that removes the jaggedness and largely keeps the shape of the unsmoothed series.

Similar procedures are used to select the smoothing parameters for other parameter series. The smoothing parameters selected for each Route B candidate model are shown in Table 5.

Penalty parameter	$\lambda_{\beta_x^{(1)}}$	$\lambda_{\beta_x^{(2)}}$	$\lambda_{\beta_x^{(3)}}$	$\lambda_{\kappa_t^{(1)}}$	$\lambda_{\kappa_t^{(2)}}$	$\lambda_{\kappa_t^{(3)}}$	λ_{γ_c}
M2	10^7	10^9	10^8	-	10^3	-	10^3
M3	10^6	-	-	10^6	-	-	10^7
M6	-	-	-	10^6	10^9	-	10^7
M7	-	-	-	10^6	10^9	10^{12}	10^7
M8	-	-	-	10^6	10^9	-	10^9
Full Plat	10^6	-	-	10^6	10^9	10^8	10^7
Simplified Plat	10^6	-	-	10^6	10^9	-	10^7
APCI	10^6	10^7	-	-	-	10^6	10^6

Table 5: The selected smoothing parameters for each Route B candidate model

As an example, let us consider the choice of the smoothing parameter for $\kappa_t^{(1)}$ in Model M3. The following describes how this smoothing parameter is chosen.

- When $\lambda_{\kappa_t^{(1)}}$ is set to 10^5 (the left panel of Figure 10), the pattern of $\kappa_t^{(1)}$ still exhibit some jaggedness, indicating that the parameter series is under-smoothed.
- When $\lambda_{\kappa_t^{(1)}}$ is set to 10^7 (the right panel of Figure 10), some features in the unsmoothed series (represented by circles in the diagram) between 1980 and 2000 are being smoothed out. The over-smoothing problem can be more easily discerned in the pattern of $K_t^{(1)}$ (the period component in the historical mortality improvement rates), shown in Figure 11. When the smoothing parameter is raised to 10^7, the locations of the peaks and troughs of $K_t^{(1)}$ become quite different.
- Therefore, we choose 10^6 as the optimal value for $\lambda_{\kappa_t^{(1)}}$ (see the middle panels of Figures 10 and 11).

We have experimented optimizing the smoothing parameters using a cross validation that is similar to the one described in Section 5.2.2. Despite being more rigorous, this method does not yield reasonable empirical results for Route B. One possible explanation for this problem is that optimally smoothed A/P/C parameters in the *mortality rate model* do not necessarily guarantee that the implied A/P/C components of historical *mortality improvements* are adequately smooth.

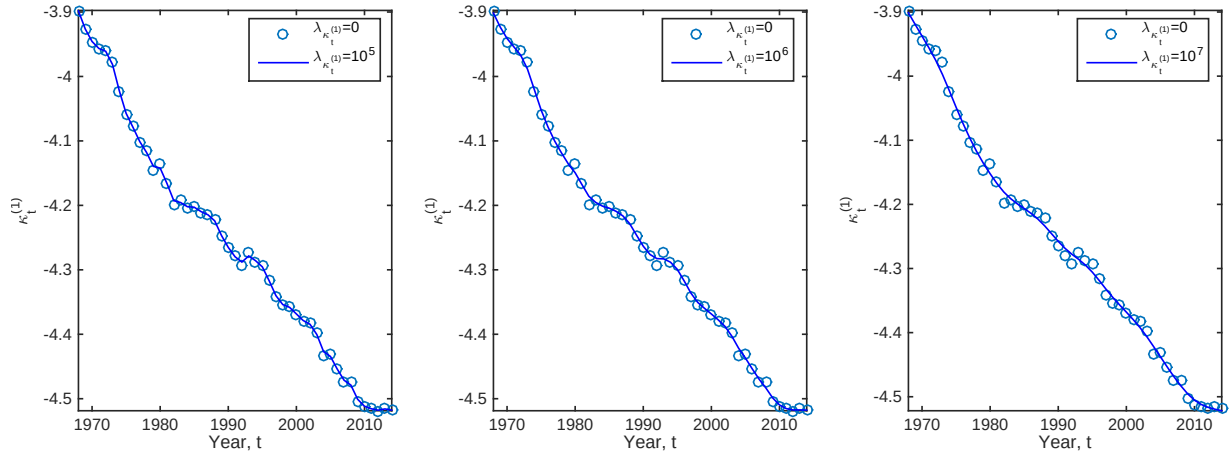

Figure 10: The estimates of $\kappa_t^{(1)}$ in the Route B Model M3 for SSA males when different smoothing parameters are used.

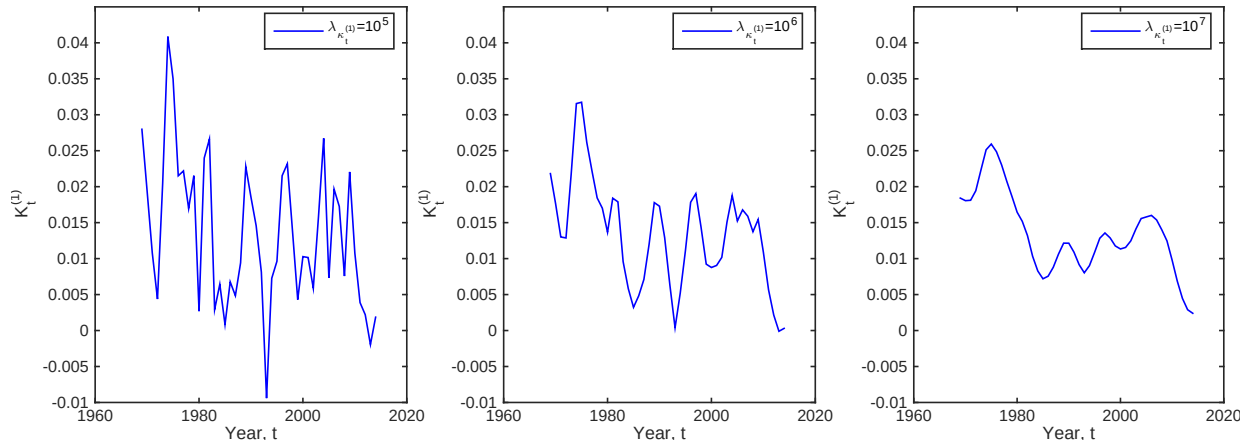

Figure 11: The values of $K_t^{(1)}$ (the period component of the mortality improvement rates) implied by the Route B Model M3 for SSA males when different smoothing parameters are used.

6.4 Robustness Tests

6.4.1 The Robustness Tests Performed

We test the robustness of the Route B models relative to (1) changes in the tolerance value used in optimizing the model parameters, (2) changes in the calibration window, (3) changes in the age range, (4) changes in the identifiability constraints, and (5) the inclusion/exclusion of the oldest and newest cohorts.

The set-up of the robustness tests is identical to that for the Route A models (see Section 5.3.1), except the following:

- When testing the robustness relative to changes in the tolerance value for Models M2 and M8, the tolerance values considered are 10^{-4}, 10^{-6} and 10^{-8} (instead of 10^{-6}, 10^{-8} and 10^{-10}). This special treatment is made because Models M2 and M8 do not converge when a tolerance value of 10^{-10} is used.
- For the APCI model, the following two sets of identifiability constraints are considered:
  1. $\sum_{c=t_0-x_1}^{t_1-x_0} \gamma_c = 0$, $\sum_{c=t_0-x_1}^{t_1-x_0} c\gamma_c = 0$, $\sum_{c=t_0-x_1}^{t_1-x_0} c^2\gamma_c = 0$, $\sum_{t=t_0}^{t_1} \kappa_t^{(1)} = 0$ and $\sum_{t=t_0}^{t_1} t\kappa_t^{(1)} = 0$
  2. $\sum_{c=t_0-x_1}^{t_1-x_0} n_c\gamma_c = 0$, $\sum_{c=t_0-x_1}^{t_1-x_0} n_c c\gamma_c = 0$, $\sum_{c=t_0-x_1}^{t_1-x_0} n_c c^2\gamma_c = 0$, $\sum_{t=t_0}^{t_1} \kappa_t^{(1)} = 0$ and $\sum_{t=t_0}^{t_1} t\kappa_t^{(1)} = 0$

6.4.2 The Robustness Measure

The robustness measure for Route B is identical to that for Route A, except that $MI_{x,t}$ is used instead of $Z_{x,t}$. More specifically, the robustness measure for the candidate models in Route B is defined as follows:

$$\text{robustness} = \frac{\max_i(\text{maximum absolute change in the } i\text{-th model term})}{\max_{x,t}(MI_{x,t}) - \min_{x,t}(MI_{x,t})}, \tag{7}$$

where $\max_{x,t}(MI_{x,t})$ and $\min_{x,t}(MI_{x,t})$ represent the maximum and minimum values of the historical mortality improvement rates in the dataset, respectively.[12] Here, 'the i-th model term' refers to the i-th term on the right-hand-side of the equation specifying $MI_{x,t}$. For instance, for Model M3 in Route B, the first model term of $MI_{x,t}$ is $K_t^{(1)}$ and the second model term of $MI_{x,t}$ is G_c. To calculate this robustness measure, a procedure similar to that described in Section 5.3.2 is used.

As before, we rate the robustness of the candidate models using the following criteria:

[12] For Model M2, Model M3, the full Plat model, the simplified Plat model and the APCI model, we calculate the historical mortality improvement rates as $\ln m_{x,t-1} - \ln m_{x,t}$; for Models M6, M7 and M8, the historical mortality improvement rates are calculated as $\ln \frac{q_{x,t-1}}{1-q_{x,t-1}} - \ln \frac{q_{x,t}}{1-q_{x,t}}$. This arrangement maximizes consistency with the way in which $MI_{x,t}$ is defined for the models.

- High robustness: $0 \leq$ robustness measure $\leq 10\%$
- Medium robustness: $10\% <$ robustness measure $\leq 20\%$
- Low robustness: robustness measure $> 20\%$

6.4.3 Test Results

The test results are reported in Table 6. Model M3, Model M6, Model M7, the full Plat model, the simplified Plat model and the APCI model exhibit medium to high levels of robustness in all of the tests performed. These six models are therefore shortlisted for further consideration.

Robustness Test	Model M2	Model M3	Model M6	Model M7	Model M8	Full Plat	Simplified Plat	APCI
SSA males								
Tolerance value	29.3% Low	0.0% High	0.0% High	0.0% High	0.0% High	0.0% High	0.0% High	0.0% High
Calibration window	7.1% High	4.4% High	6.2% High	9.1% High	8.2% High	4.3% High	4.1% High	5.1% High
Age range	38.5% Low	5.2% High	12.7% Medium	7.8% High	26.9% Low	7.8% High	5.7% High	2.3% High
Parameter constraints	10.5% Medium	0.4% High	0.7% High	1.1% High	0.0% High	1.8% High	1.4% High	0.1% High
Exclusion of cohorts	35.0% Low	2.9% High	7.6% High	8.9% High	9.3% High	12.2% Medium	6.9% High	5.2% High
SSA females								
Tolerance value	25.1% Low	0.0% High	0.0% High	0.0% High	0.0% High	0.0% High	0.0% High	0.0% High
Calibration window	8.8% High	3.4% High	11.3% Medium	6.5% High	8.5% High	12.9% Medium	5.2% High	3.8% High
Age range	27.2% Low	4.9% High	10.2% Medium	14.1% Medium	59.8% Low	10.0% Medium	5.6% High	3.2% High
Parameter constraints	23.5% Low	0.2% High	0.0% High	1.1% High	18.5% Medium	1.4% High	0.7% High	0.7% High
Exclusion of cohorts	23.4% Low	3.7% High	6.3% High	6.7% High	11.7% Medium	11.0% Medium	6.8% High	3.2% High

Table 6: The results of the robustness tests performed for each Route B candidate model

6.5 Analyzing the Standardized Residuals from the Shortlisted Models

We now analyze the standardized residuals produced by the shortlisted models. Given how Route B models are estimated, the standardized residuals for Route B models are calculated using the

following formula:

$$\frac{d_{x,t} - E_{x,t}\hat{m}_{x,t}}{\sqrt{E_{x,t}\hat{m}_{x,t}}},$$

where $\hat{m}_{x,t}$ represents the fitted value of $m_{x,t}$ produced by the model being analyzed.[13]

Because Route B models are estimated using a Poisson death count assumption, their standardized residuals are not necessarily normally distributed even if the model is adequate. As such, we cannot apply the 'normal q-q plot' and the normality test which we used when we pursue Route A. However, we can still access the adequacy of a Route B model by examining the heat map of its standardized residuals. If a model is adequate, then the pattern of its standardized residuals should be random with little clustering.

Figure 12 compares the heatmaps of standardized residuals obtained from shortlisted Route B models. The following observations are made:

- Models M6 and M7 perform the worst in the analysis of residuals. Large horizontal clusters are found in the heat maps produced from these models, indicating that age effect is not adequately captured.
- Significant clustering is observed in the residuals heatmaps produced by Model M3 and the APCI model for males. Large vertical clusters are observed between mid-80s and mid-90s.
- The full Plat model performs the best in the analysis of residuals. The standardized residuals produced from the full Plat model appear to be the most random, although there still exist some gentle diagonal patterns.
- The simplified Plat model is the second best performing model in the analysis of residuals. Compared to those from full Plat model, the standardized residuals from the simplified Plat model exhibit slightly more clustering.

The heat maps of standardized residuals suggest that the Plat models (both full and simplified) are reasonable choices. We recommend the simplified Plat model for use in Route B for the two reasons. First, in the analysis of residuals, the full Plat model only wins the simplified Plat model by a narrow margin. Second, the simplified Plat model outperforms the full Plat model in the robustness tests.

[13]Because the standardized residuals for Route A and Route B models are calculated in different manners, the standardized residuals presented in this section cannot be directly compared and contrasted with those shown in Section 5.4.

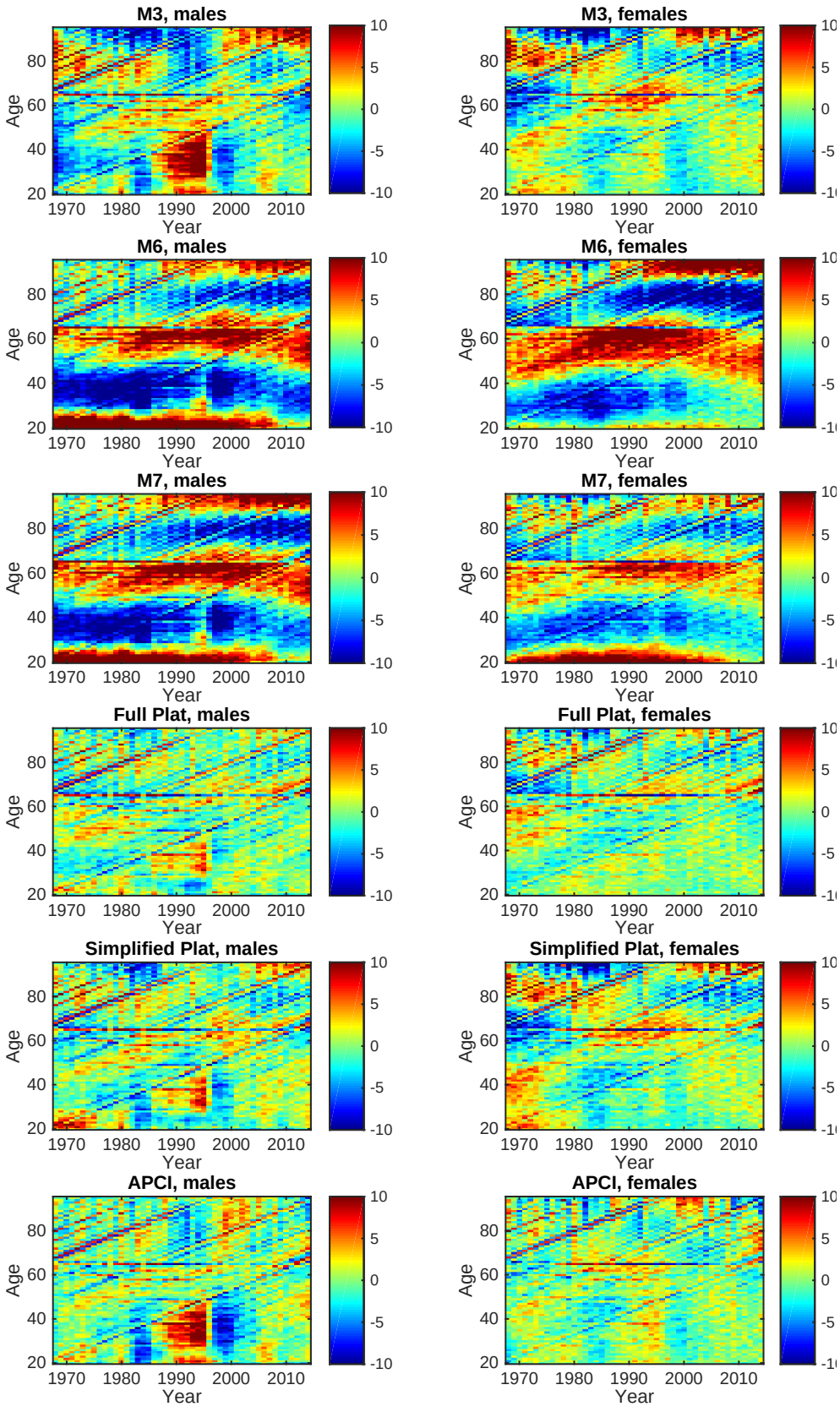

Figure 12: Heatmaps of the standardized residuals produced by the shortlisted Route B models.

7 Conclusion

7.1 A Summary of the Modeling Work Performed

We have considered two pathways (Route A and Route B) to obtaining an A/P/C decomposition of historical mortality improvement rates.

The following summarizes the modeling work entailed in Route A.

- Estimate all candidate models to historical *mortality improvement rates*

  The candidate models considered are Models M2, M3, M6, M7 and M8, the full Plat model and the simplified Plat model.

- Perform robustness tests

  Model M3, Model M6 and the simplified Plat model exhibit medium to high levels of robustness, and are therefore shortlisted for further consideration.

- Analyze the standardized residuals produced by the shortlisted models.

  Among the shortlisted models, the simplified Plat model performs the best in the analysis of residuals.

In Route A, the most effective model (the simplified Plat model) implies that mortality improvement is driven by (1) a stand-alone age component, (2) a stand-alone period component, (3) a period component that interacts with (a linear function) of age, and (4) a stand-alone cohort component.

The following summarizes the modeling work involved in Route B.

- Estimate all candidate models to historical *death rates/probabilities*

  The candidate models considered are Models M2, M3, M6, M7 and M8, the full Plat model, the simplified Plat model, and the APCI model.

- Perform robustness tests

  Model M3, Model M6, Model M7, the full Plat model, the simplified Plat model, and the APCI model exhibit medium to high levels of robustness, and are therefore shortlisted for further consideration.

- Analyze the standardized residuals produced by the shortlisted models.

  Among the shortlisted models, the simplified Plat model is chosen on the basis of its performance in the analysis of residuals and robustness tests.

In Route B, the most effective model (the simplified Plat model) implies that mortality improvement is driven by (1) a stand-alone period component, (2) a period component that interacts with (a linear function) of age, and (3) a stand-alone cohort component.

Figures 13 and 14 provide a side-by-side comparison between the A/P/C components of historical mortality improvements derived from the two pathways. Note that the 'stand-alone age component' does not apply to Route B.

7.2 Recommendation and Caveats

We conclude that the simplified Plat model in Route A is the most appropriate model for identifying the A/P/C components of the U.S. gender-specific historical mortality improvements. The conclusion is drawn on the following bases.

- We have argued previously that the simplified Plat model performs the best in both Routes A and B. From Figures 13 and 14 we observe that the A/P/C components derived from the two pathways are not too different, but those from Route B are noticeably less smooth. The lack of smoothness is a concern, in part because the underlying A/P/C components should in principle be free of noise (the noise should be captured by the residual component) and in part because it would be more challenging (for the researchers of the follow-up project) to link the A/P/C components to intrinsic factors if they are jagged.
- One technical limitation of Route B is that the smoothing parameters for Route B models are chosen with a rather subjective approach. Therefore, from a technical viewpoint, the simplified Plat model in Route A is more preferred than its Route B counterpart.

We emphasize that the conclusion drawn is data-dependent, applicable to the data set under consideration (U.S. gender-specific, ages 20-95, years 1968-2014) only. Although the model selection procedures (Routes A and B) stand for any data set, the conclusion drawn can be (very) different when a different population is considered.

We also emphasize that the conclusion might be different if we focus on a particular age range (say, pensionable ages) only. We have repeated the entire analysis using data for ages above 55 only, and found that for this age range the advantage of the simplified Plat model over the other candidate models becomes less apparent.[14]

Further, the selected model (the simplified Plat model in Route A) is only the best currently. While the robustness test results indicate that the conclusion is unlikely to change in the next few years, it is entirely possible that the conclusion will change when a substantial volume of new data is added to the analyses, for reasons such as structural changes in the underlying A/P/C components.

Finally, we acknowledge that this study focuses only on the overall U.S. population, without considering the possible differences in mortality improvements due to, for example, differences in socioeconomic status. It is warranted to revisit this study using more granular data when such data becomes available in the future.

[14]The additional analyses are documented in the full report of this study (Section 7, Li et al., 2018a; Section 5, Li et al., 2018b).

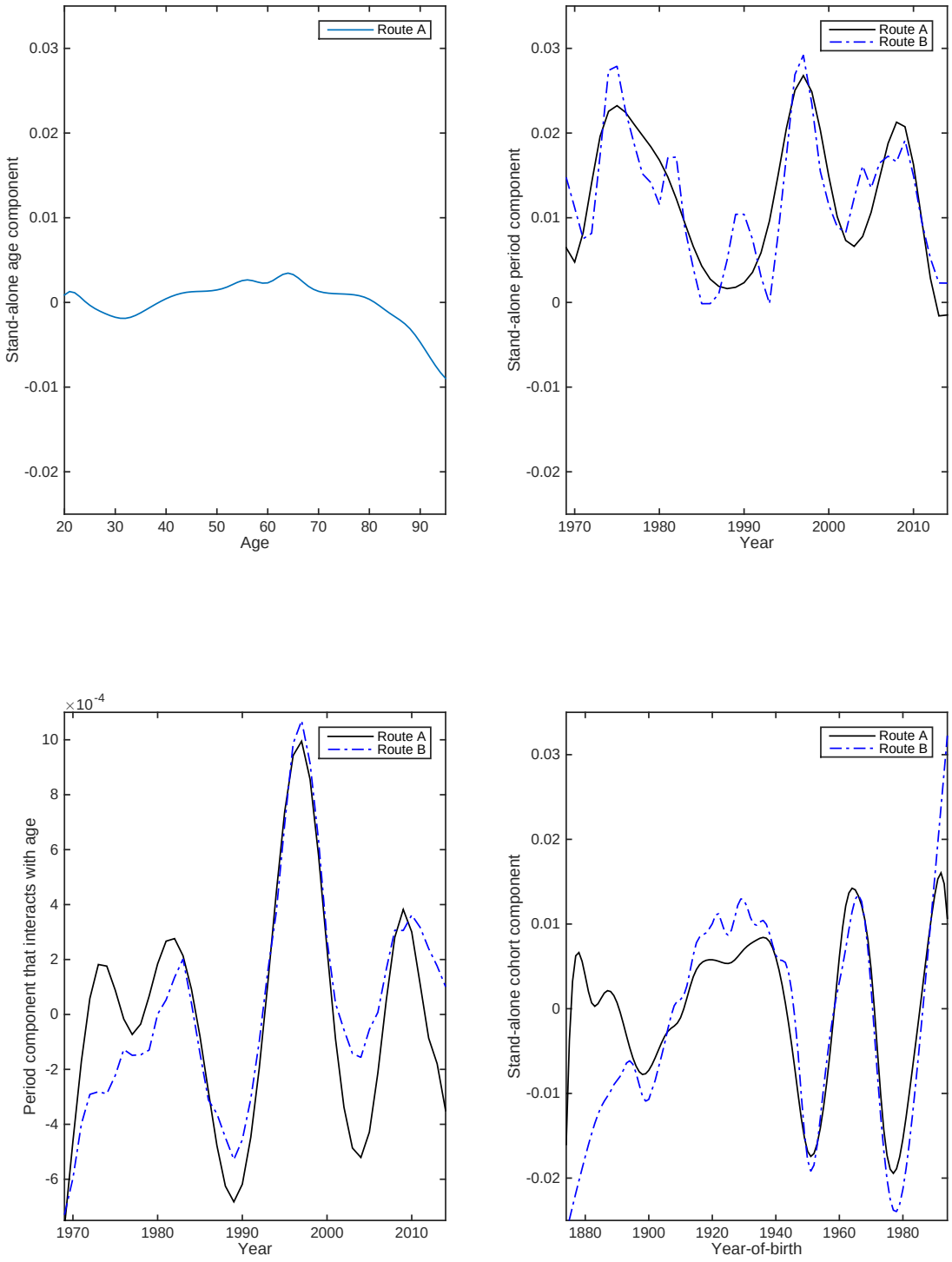

Figure 13: A comparison between the A/P/C components of historical mortality improvements obtained from the most effective Route A and Route B models, U.S. males.

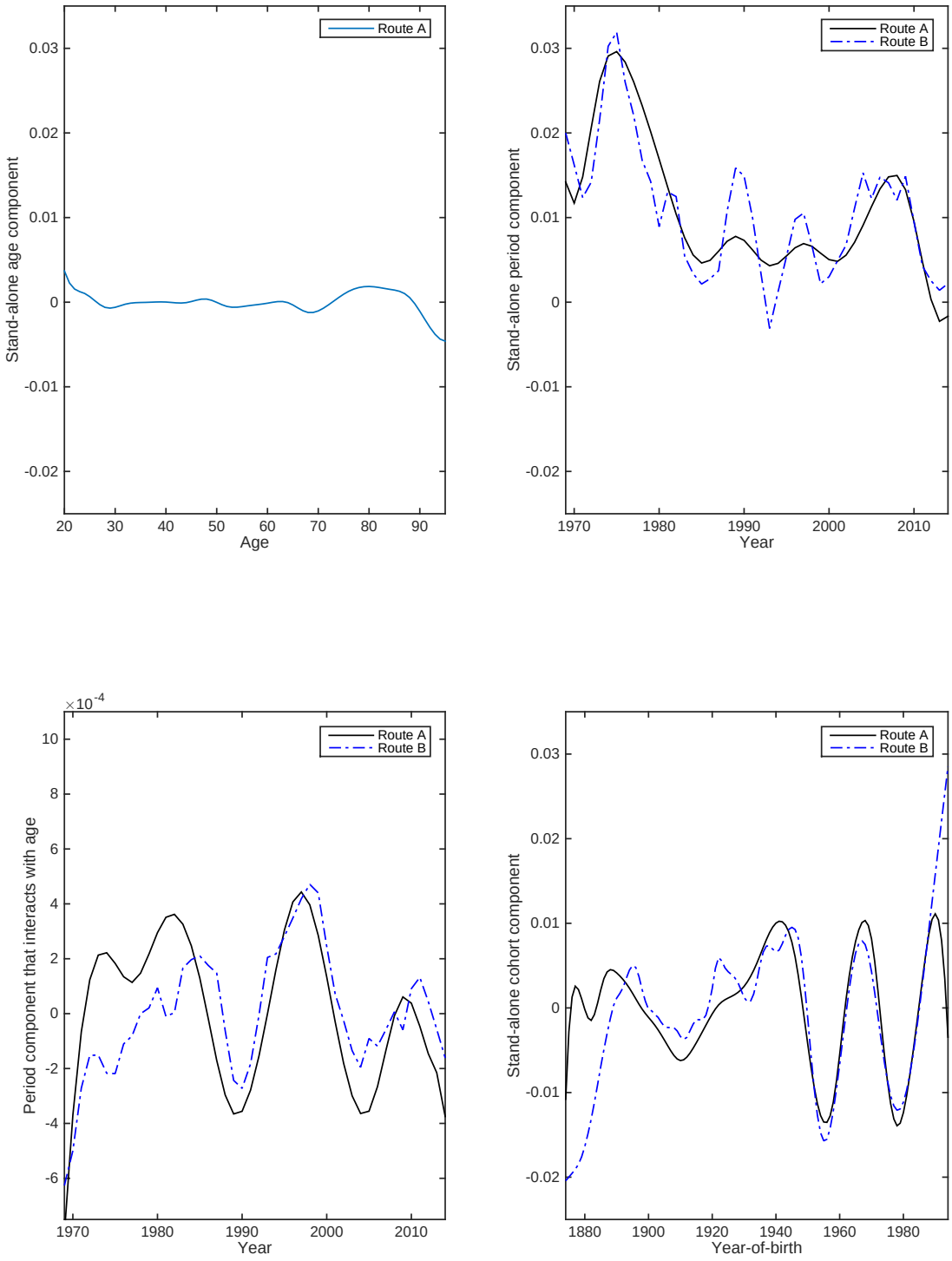

Figure 14: A comparison between the A/P/C components of historical mortality improvements obtained from the most effective Route A and Route B models, U.S. females.

7.3 The Next Steps

The mortality models analyzed in this paper and the final conclusions regarding the A/P/C decomposition of historical US population mortality experience have provided LAG with a solid foundation upon which to build subsequent phases of its longer-term research objectives. In particular, the A/P/C components identified in this paper are essential to the next project phase, which will be researching degrees of correlation between significant historical US population mortality trends and specific mortality drivers.

References

Cairns, A.J.G., Blake, D., and Dowd, K. (2006). A two-factor model for stochastic mortality with parameter uncertainty: Theory and calibration. *Journal of Risk and Insurance* 73: 687-718.

Cairns, A.J.G., Blake, D., Dowd, K., Coughlan, G.D., Epstein, D., and Khalaf-Allah, M. (2011). Mortality Density Forecasts: An Analysis of Six Stochastic Mortality Models. *Insurance: Mathematics and Economics* 48: 355-367.

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

Continuous Mortality Investigation Bureau (2017a). *CMI Mortality Projections Model: CMI_2016.* Mortality Projections Committee Working Paper 97.

Continuous Mortality Investigation Bureau (2017b). *CMI Mortality Projections Model: Methods.* Mortality Projections Committee Working Paper 98.

Continuous Mortality Investigation Bureau (2017c). *CMI Mortality Projections Model: Software user guide.* Mortality Projections Committee Working Paper 99.

Continuous Mortality Investigation Bureau (2009a). Working Paper 38, A Prototype Mortality Projections Model, Part One – An Outline of the Proposed Approach. Available at www.actuaries.org.uk.

Continuous Mortality Investigation Bureau (2009b). Working Paper 39, A Prototype Mortality Projections Model, Part Two – Detailed Analysis. Available at www.actuaries.org.uk.

Currie I.D., Durban, M., and Eilers, P.H.C. (2004). Smoothing and forecasting mortality rates. *Statistical Modelling* 4: 279-298.

Delwarde, A., M. Denuit, and P. Eilers (2007). Smoothing the Lee-Carter and Poisson Log-Bilinear Models for Mortality Forecasting: A Penalized Log-Likelihood Approach. *Statistical Modelling* 7(1): 29-48.

Dowd, K., Cairns, A.J.G., Blake, D., Coughlan, G.D., Epstein, D., and Khalaf-Allah, M. (2010a). Backtesting Stochastic Mortality Models: An Ex-Post Evaluation of Multi-Period-Ahead Density Forecasts. *North American Actuarial Journal* 14: 281-298.

Dowd, K., Cairns, A.J.G., Blake, D., Coughlan, G.D., Epstein, D., and Khalaf-Allah, M. (2010b). Evaluating the Goodness of Fit of Stochastic Mortality Models. *Insurance: Mathematics and Economics* 47: 255-265.

Goss, S., Wade, A., Glenn, K., Morris, M., and Bye, M. (2015). Accuracy of Mortality Projections in Trustees Reports. Actuarial Note No. 156. Social Security Administration, Office of the Chief Actuary, Baltimore, Maryland.

Hunt, A. and Villegas, A.M. (2017). Mortality Improvement Rates: Modeling and Parameter Uncertainty. Paper presented at the Society of Actuaries Living to 100 Symposium, Orlando, Florida, January 46, 2017.

Li, J.S.-H., Zhou, R., and Liu, Y. (2018a). Components of Historical Mortality Improvement Final Report Volume 1 – Background and Mortality Improvement Rate Modeling. Schaumburg, IL: Society of Actuaries.

Li, J.S.-H., Zhou, R., and Liu, Y. (2018b). Components of Historical Mortality Improvement Final Report Volume 2 – Mortality Rate Modeling and Conclusions. Schaumburg, IL: Society of Actuaries.

Osmond, C. (1985). Using Age, Period and Cohort Models to Estimate Future Mortality Rates. *International Journal of Epidemiology* 14: 124-129.

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

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

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