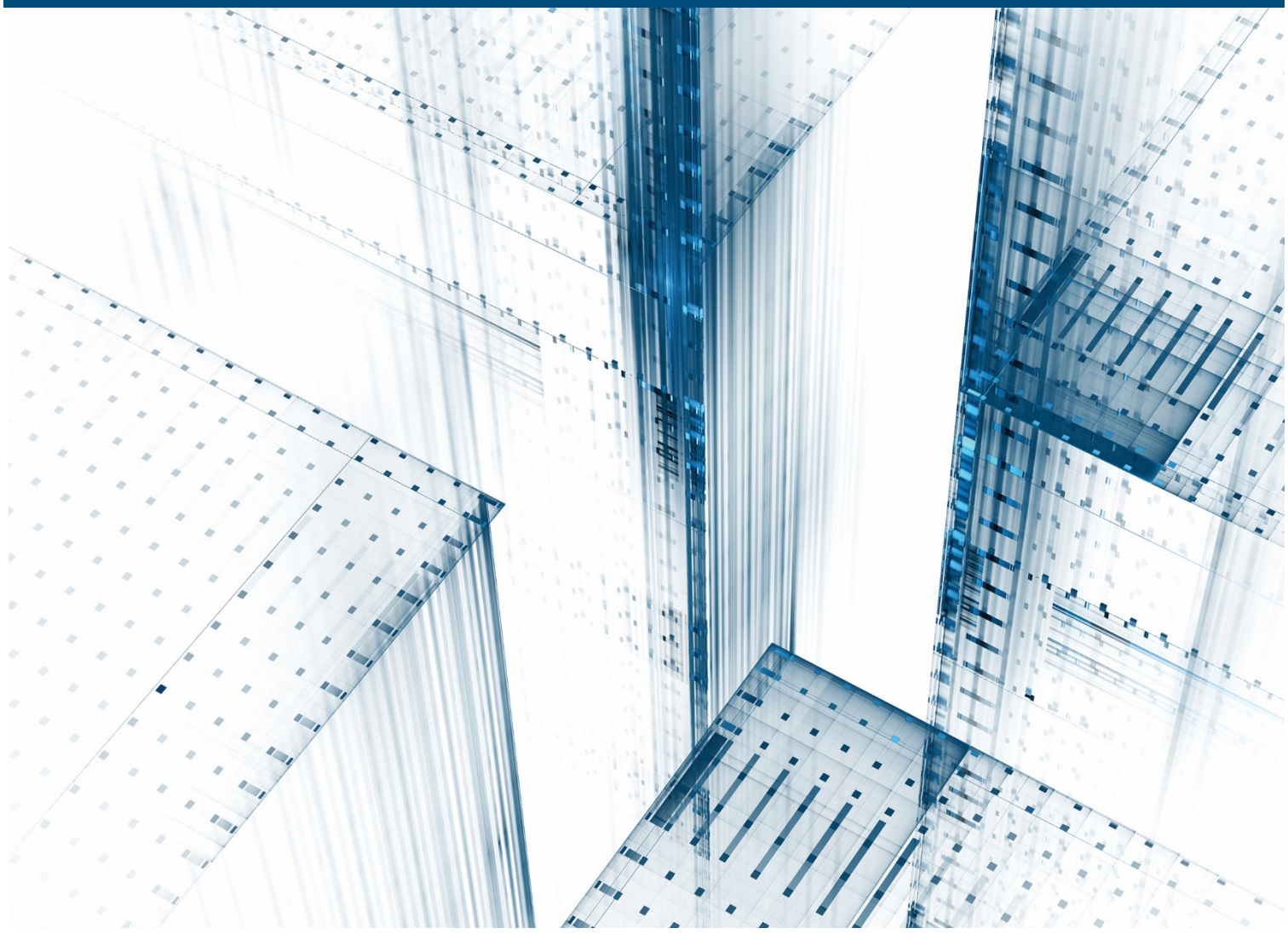
SOCIETY OF ACTUARIES

Components of Historical Mortality Improvement

Volume 1 — Background and Mortality Improvement Rate Modeling



September 2017

Components of Historical Mortality Improvement

Volume 1 – Background and Mortality Improvement Rate Modeling

AUTHOR Johnny S.-H. Li, PhD, FSA
Rui Zhou, PhD, FSA
Yanxin Liu, PhD

Caveat and Disclaimer

Executive Summary

This report documents the "Components of Historical Mortality Improvement" project commissioned by the Longevity Advisory Group (LAG) of the Society of Actuaries. The purpose of the project is to compare and contrast methodologies for allocating historical gender-specific mortality improvement (or deterioration) experience in the U.S. into four components (age, period, cohort, and residual), drawing from the methodologies developed previously by the Continuous Mortality Investigation (CMI) of the Institute and Faculty of Actuaries (IFoA).

The deliverables of this project include the following:

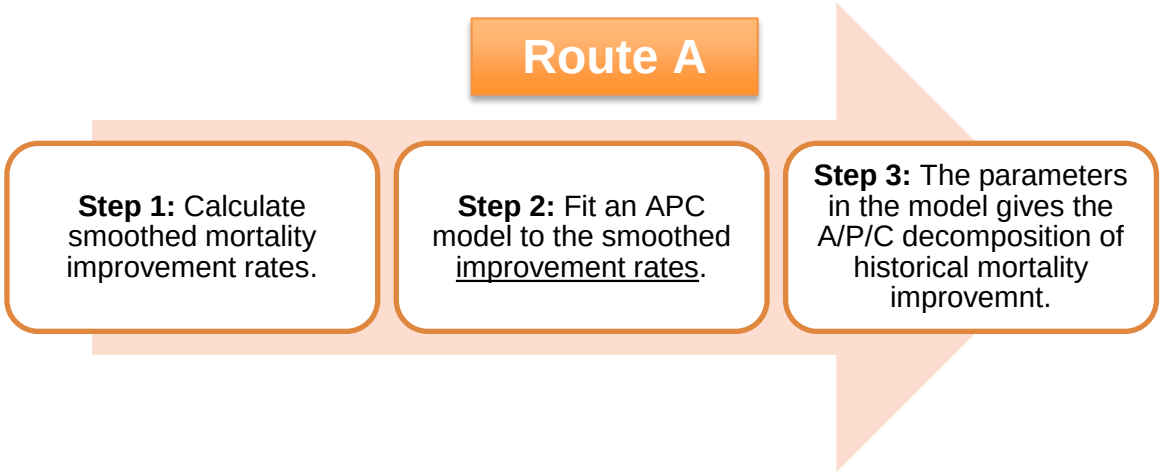
1. A summary of the methodologies used by the CMI to allocate historical mortality improvement experience into gender-specific age, period, cohort (year-of-birth), and residual components.

2. Specific suggestions of other APC allocation methodologies that deserve consideration.

3. An assessment of the overall effectiveness of all of the methodologies under consideration with respect to (1) the adequacy in capturing the features possessed by the US gender-specific historical mortality improvement experience and (2) the robustness relative to different factors (e.g., confirming that relatively small changes to the selected observation period does not produce disproportionately large changes in the underlying A/P/C components).

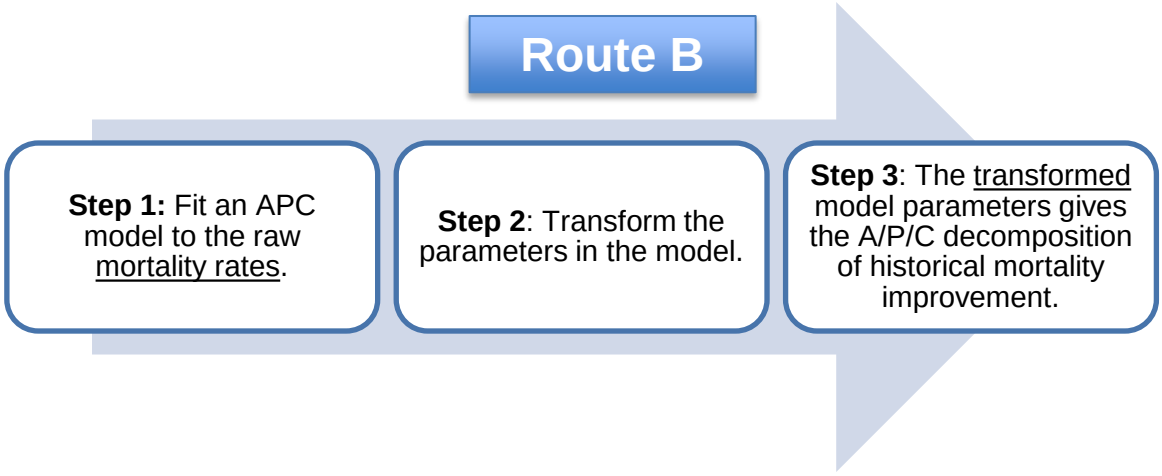
4. Identify the most effective A/P/C allocation method, based on the gender-specific US population mortality data from 1950 through current, and for ages 20 through 95.

The project team divides the methodologies under consideration into two broad categories, Route A and Route B. In Route A, the APC models are fitted to smoothed mortality improvement rates, and the parameters in the estimated models give A/P/C decompositions of historical mortality improvement. The steps involved in Route A are shown in the flow chart below:



Route A includes the older CMI method (thereafter called the CMI-09 method), which is documented in the CMI Working Papers 38 and 39 (CMI, 2009a,b).

In Route B, APC models are fitted to mortality rates and the desired A/P/C decomposition of mortality improvement experience is obtained by transforming the parameters in the chosen APC model. The steps involved in Route B are shown in the flow chart below:



Route B encompasses the newer CMI method (thereafter referred to as the CMI-17 method), which is documented in the CMI Working Papers 97, 98 and 99 (CMI, 2017a,b,c).

The project report is divided into two volumes. Volume 1 (this volume) provides the background information and presents the modeling work associated with Route A. Volume 2 documents the modeling work related to Route B, and concludes with a recommendation.

Section 1 of this volume describes the two data sets used, which are respectively provided by the Human Mortality Database (HMD) and the U.S. Social Security Administration (SSA):

Data set	Source	Age Range	Sample Period
(i)	HMD	20 to 95	1968 to 2014
(ii)	SSA	20 to 95	1968 to 2014

The possible limitations of the data sets are noted. Sections 2 and 3 define the notation and summarizes the CMI-09 method.

Section 4 details the implementation of the CMI-09 method using the U.S. data sets. The implementation was successful, but a major drawback of the method was identified. When applied to the U.S. male data sets, large vertical clusters are found in the heat maps of the standardized residuals. This outcome in an indication that the APC model structure used in the CMI-09 method is unable to pick up some features that are specific to the U.S. historical mortality improvement experience. It is found that adding an age breakpoint ameliorates the problem, but using the use of an age breakpoint may result in inconsistencies between mortality projections for younger and older ages.

Section 5 presents the robustness tests that we performed on the CMI-09 method. We test the robustness of the resulting age, period and cohort components to (1) changes in the calibration window, (2) changes in the age range, (3) changes in the parameter constraints used, and (4) inclusion/exclusion

of the oldest/newest cohorts. It is found that the CMI-09 method is reasonably robust, but the problem of residuals clustering for U.S. male still remains even if calibration window or age range is somewhat modified.

Section 6 studies if the CMI-09 method may be improved by considering alternative, more sophisticated APC model structures for decomposing the smoothed mortality improvement rates. Seven candidate model structures are examined: M2, M3 (the model used in the CMI-09 method), M6, M7, M8, the full Plat model and the simplified Plat model.[1] We first perform a range of robustness tests to shortlist a smaller number of model structures that merit further consideration. Then, we analyze the standardized residuals produced by the shortlisted models using statistical tools such as the Anderson-Darling test. The analyses suggest that the simplified Plat model is the most suitable for the U.S. smoothed mortality improvement rates. This model also eliminates the need for an age breakpoint.

Section 7 studies a more statistically rigorous method for estimating the Route A APC models. The method integrates smoothing and estimation into one single process by introducing a roughness penalty to the objective function from which optimized parameters are derived. A cross-validation is used to determine the best tradeoff between smoothness and goodness-of-fit. It is found that this alternative estimation/smoothing method yields fairly similar A/P/C decompositions.

Section 8 repeats the analyses using the data for ages 55 to 95 only. The results suggest that the simplified Plat model is still the most effective Route A model even when the data for younger ages are discarded.

Section 9 draws the following conclusion: When Route A is used, the simplified Plat model (without any age breakpoint) is the most effective model for decomposing the U.S. historical mortality improvement experience into A/P/C components.

The next volume describes and evaluates the Route B models. It also compares the decomposition results from the two routes, and make a final recommendation.

The project team is grateful to all members of the Project Oversight Group, formed by Jennifer Haid (Chair), Jean-Marc Fix, Zach Granovetter, George Graziani, Alla Kleyner, Dale Hall, Bob Howard, Al Klein, Andy Peterson, Larry Pinzur, Erika Schulty, and Larry Stern, for their guidance and insightful comments.

[1] We exclude M1 (the Lee-Carter model) and M5 (the Cairns-Blake-Dowd model), because these models do not incorporate cohort effects. We also exclude M4 as it does not explicitly decompose historical mortality into APC components. Details concerning the models under consideration can be found in the papers by Cairns et al. (2009) and Plat (2009).

Acknowledgements

The SOA would like to extend their gratitude to the Project Oversight Group:

- Jennifer A. Haid, FSA (Chair)
- Jean-Marc Fix, FSA, MAAA
- Zachary Granovetter, FSA
- George A. Graziani, FSA, FCIA
- Robert C. W. Howard, FSA, FCIA
- Allen M. Klein, FSA, MAAA
- Alla Kleyner, FSA
- Laurence Pinzur, FSA
- Larry N. Stern, FSA, MAAA

The authors would like to thank the SOA staff for their leadership and coordination of the project:

- R. Dale Hall, FSA, MAAA, CERA
- Andrew J. Peterson, FSA, EA, MAAA, FCA
- Erika Schulty

Contents

1 The Data Sets Used

We consider two data sets, which have the same age range and the same age sample period. Data set (i) is composed of the raw death and exposure counts provided by the Human Mortality Database (HMD), whereas data set (ii) is based on the raw death and exposure counts provided by the U.S. Social Security Adminstration (SSA). The following table summarizes the two data sets used in this report:

Data set	Source	Age Range	Sample Period
(i)	HMD	20 to 95	1968 to 2014
(ii)	SSA	20 to 95	1968 to 2014

It should be noted that the two data sets are based on different sources. For the SSA data set, the death and exposure counts for age 65 and over are based on experience derived from the Medicare-enrolled population. In contrast, 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'.[2] 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] 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.

2 Notation

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

- x represents the age of an individual;
- t represents calendar year;
- $[x_0, x_1]$ is the sample age range (i.e., the calibration window);
- $[t_0, t_1]$ is 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 $t_1 - x_0$;
- $m_{x,t}$ and $\tilde{m}_{x,t}$ represent the raw and smoothed central rates of death, respectively;
- $q_{x,t}$ and $\tilde{q}_{x,t}$ represent the raw and smoothed conditional death probabilities, respectively;
- $Z_{x,t} = 1 - q_{x,t}/q_{x,t-1}$ is the raw mortality improvement rate for age x and year t;
- $\tilde{Z}_{x,t} = 1 - \tilde{q}_{x,t}/\tilde{q}_{x,t-1}$ is the smoothed mortality improvement rate for age x and year t;

3 A Summary of the CMI-09 Method

The CMI-09 decomposition method includes two stages. In the first stage, a two-dimensional (age/cohort) P-Spline (Currie et al., 2004) is applied to the crude mortality rates. In the second stage, the smoothed mortality improvement rates are decomposed into age/period/cohort/residual components using an additive APC model. The CMI-09 decomposition method can be summarized as follows:

- Stage 1: Two dimensional P-spline smoothing
  - Step 1.1: Obtain the smoothed mortality rate $\tilde{m}_{x,t}$ by fitting a two-dimensional P-spline to the crude mortality rates $m_{x,t}$'s.
  - Step 1.2: Connect the smoothed mortality rate $\tilde{m}_{x,t}$ with the smoothed death probability $\tilde{q}_{x,t}$ using the UDD (uniform distribution of deaths between integer ages) assumption:

$$\tilde{q}_{x,t} = \tilde{m}_{x,t}/(1 + 0.5\tilde{m}_{x,t}).$$

- Stage 2: Age/Period/Cohort/Residual Components Decomposition
  - Step 2.1: Define the smoothed mortality improvement $\tilde{Z}_{x,t}$ as the fall in the smoothed mortality rates from $\tilde{q}_{x,t-1}$ to $\tilde{q}_{x,t}$ (CMI, 2009b, pp.35).

$$\tilde{Z}_{x,t} = 1 - \tilde{q}_{x,t}/\tilde{q}_{x,t-1}$$

– Step 2.2: Fit the smoothed mortality improvement to a simple additive APC model,

$$\tilde{Z}_{x,t} = a_x + k_t + g_c + e_{x,t}, \tag{1}$$

where a_x, k_t, g_c and $e_{x,t}$ are the age, period, cohort and residual components, respectively. The age, period and cohort components are obtained by minimizing the residual sum of squares (CMI, 2009b, pp.36):

$$\sum_{x=x_0}^{x_1} \sum_{t=t_0}^{t_1} e_{x,t}^2.$$

Three constraints,

$$\sum_{x=x_0}^{x_1} a_x = 0, \quad \sum_{c=t_0-x_1}^{t_1-x_0} g_c = 0, \quad \sum_{c=t_0-x_1}^{t_1-x_0} c \cdot g_c = 0,$$

are used to stipulate parameter uniqueness.[3]

– Step 2.3: For each age/period/cohort components, we use basis-splines with 4-year knot spacing (consistent with the knot-spacing in the P-Spline models applied to the raw population data) to ensure reasonable smoothness.

[3]The first two constraints are provided in the CMI documentations. The third constraint is provided by Jon Palin from the CMI.

4 Implementing the CMI-09 Method

In this section, we apply the CMI-09 method to the U.S. mortality data. Possible limitations of the CMI-09 method are also identified.

4.1 The Smoothed Death Rates

In the first stage of the CMI method, a two-dimensional (age/cohort) P-Spline is applied to the crude mortality rates. Following Currie et al. (2006), the project team has applied the following settings for the two-dimensional P-Spline:

- Dimensions: age-cohort
- Knot spacing: one knot every four years up to a maximum of 40 knots
- Type of splines: cubic
- Penalty degree: quadratic

Figures 1 and 2 show the raw and smoothed central death rates (in log scale) across ages for three specific years-of-birth: 1930, 1950 and 1970. It should be noted that not all rates in the age range of 20 to 95 are available. For instance, as the data sets cover calendar years 1968 to 2014 only, for year-of-birth 1930 the available central death rates span age 38 ($= 1968 - 1930$) to age 84 ($= 2014 - 1930$) only. For the same reason, we can only show the central death rates up to age 64 ($= 2014 - 1950$) for year-of-birth 1950 and up to age 44 ($= 2014 - 1970$) for year of birth 1970. It can be observed that the P-splines effectively removed the jaggedness in the raw rates, and that the rates from the HMD and SSA data sets are quite similar.

Figures 3 and 4 show the raw and smoothed central death rates (in log scale) across years-of-birth for three specific ages: 35, 65 and 95. As the data sets cover calendar years 1968 to 2014, the available mortality rates for age 35 begin in year-of-birth 1933 ($= 1968 - 35$) and end in year-of-birth 1979 ($= 2014 - 35$). Likewise, the available mortality rates for age 65 span years-of-birth 1903 ($= 1968 - 65$) to 1949 ($= 2014 - 65$) and those for age 95 cover years-of-birth 1873 ($= 1968 - 95$) to 1919 ($= 2014 - 95$). The rates from the HMD and SSA data sets are broadly in line, but at age 95 the SSA rates are notably higher.

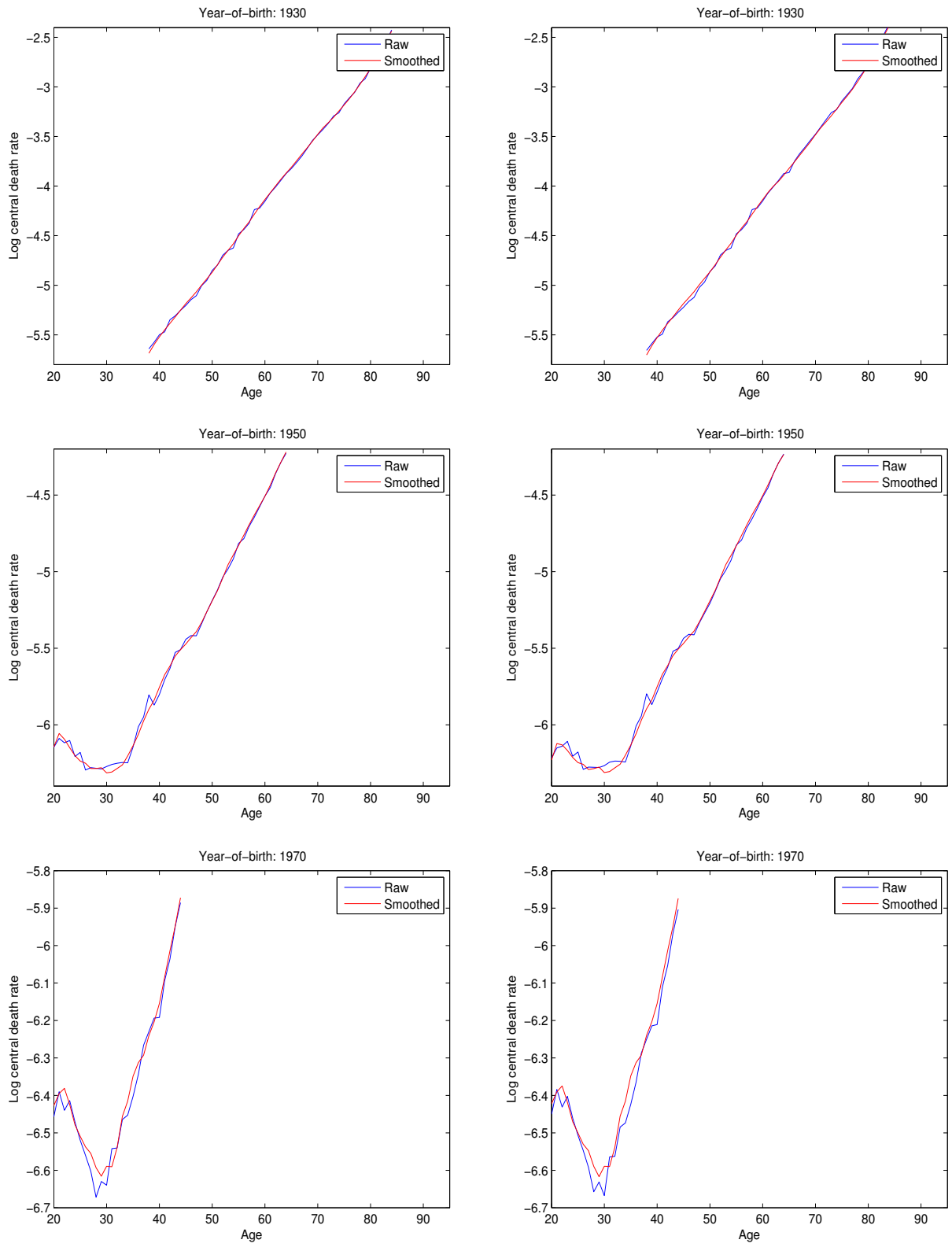


Figure 1: The raw and smoothed mortality rates of U.S. **males** born in year 1930 (upper panels), year 1950 (middle panels) and year 1970 (lower panels).
- Left panels: Data set (i) (HMD, ages 20-95, years 1968-2014).
- Right panels: Data set (ii) (SSA, ages 20-95, years 1968-2014).

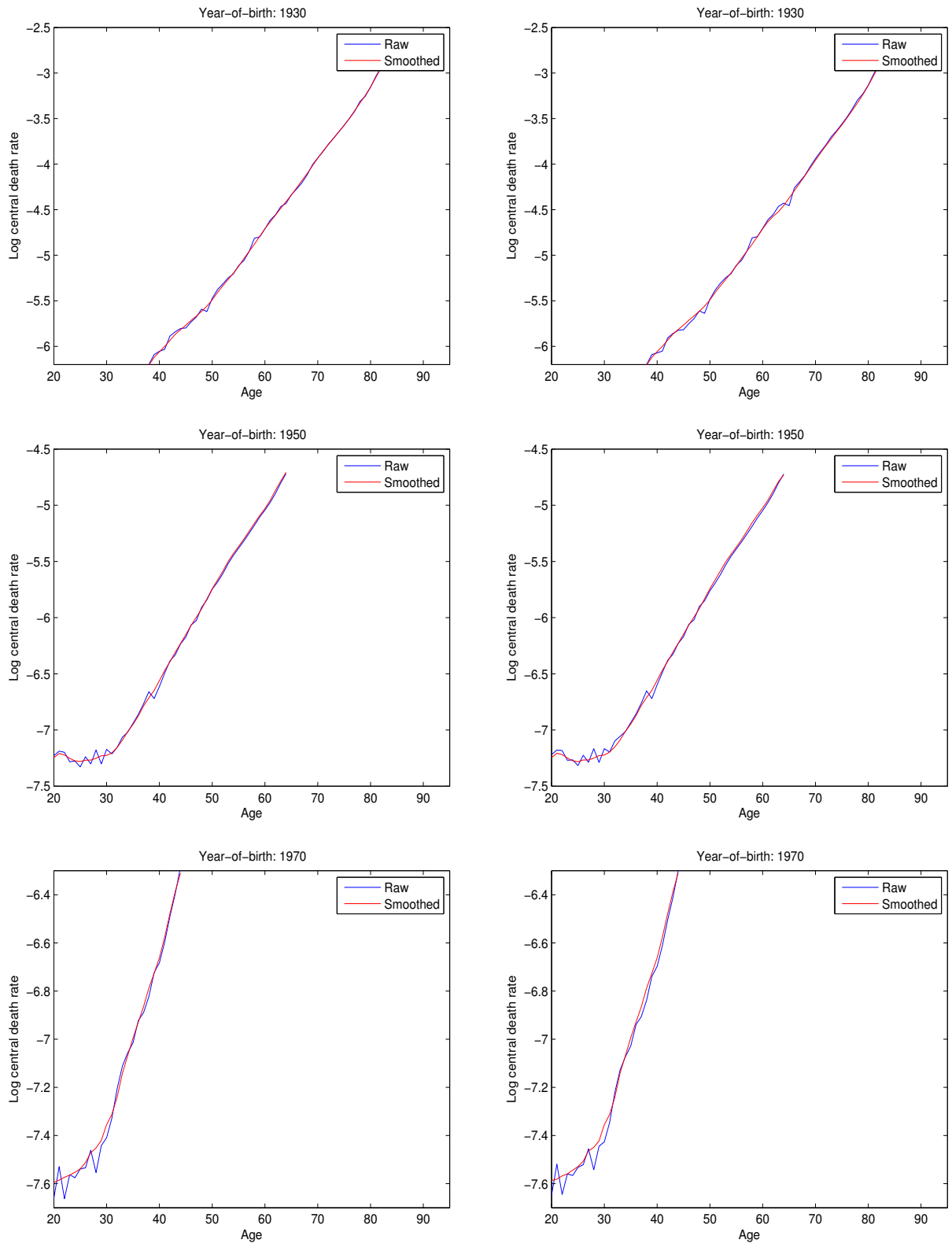


Figure 2: The raw and smoothed mortality rates of U.S. **females** born in year 1930 (upper panels), year 1950 (middle panels) and year 1970 (lower panels).
- Left panels: Data set (i) (HMD, ages 20-95, years 1968-2014).
- Right panels: Data set (ii) (SSA, ages 20-95, years 1968-2014).

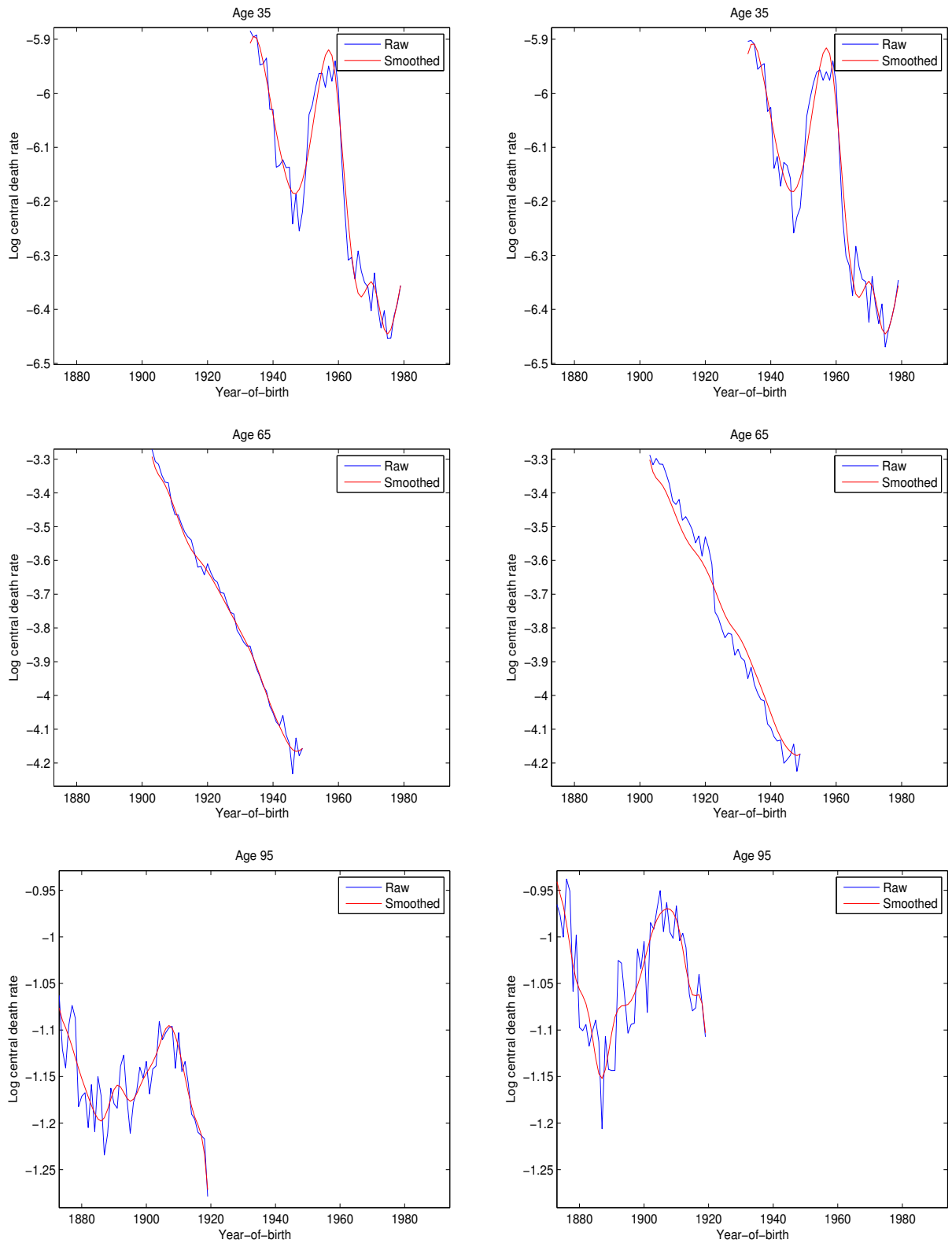


Figure 3: The raw and smoothed mortality rates of U.S. **males** at age 35 (upper panels), age 65 (middle panels) and age 95 (lower panels).
- Left panels: Data set (i) (HMD, ages 20-95, years 1968-2014).
- Right panels: Data set (ii) (SSA, ages 20-95, years 1968-2014).

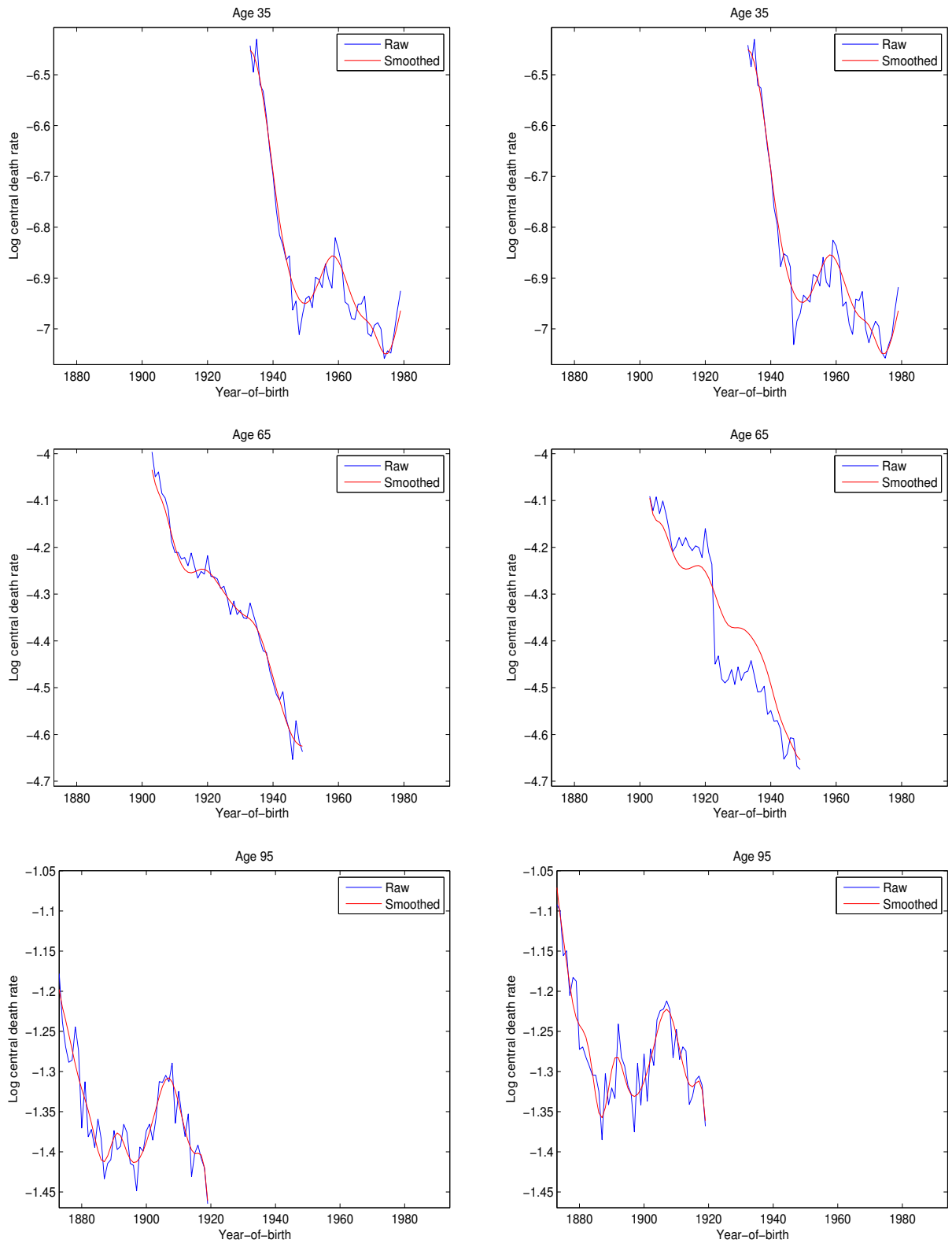


Figure 4: The raw and smoothed mortality rates of U.S. **females** at age 35 (upper panels), age 65 (middle panels) and age 95 (lower panels).
- Left panels: Data set (i) (HMD, ages 20-95, years 1968-2014).
- Right panels: Data set (ii) (SSA, ages 20-95, years 1968-2014).

4.2 The Smoothed Mortality Improvement Rates

The input for the second stage of the CMI-09 method is the smoothed mortality improvement rates. Figures 5 and 6 provide the heat maps for the smoothed mortality improvement rates for U.S. males and females, respectively. The smoothed mortality improvement rates from the two data sets are very similar. For the readers' information, we indicate in the heat maps the ages at which the peaks and troughs are located.

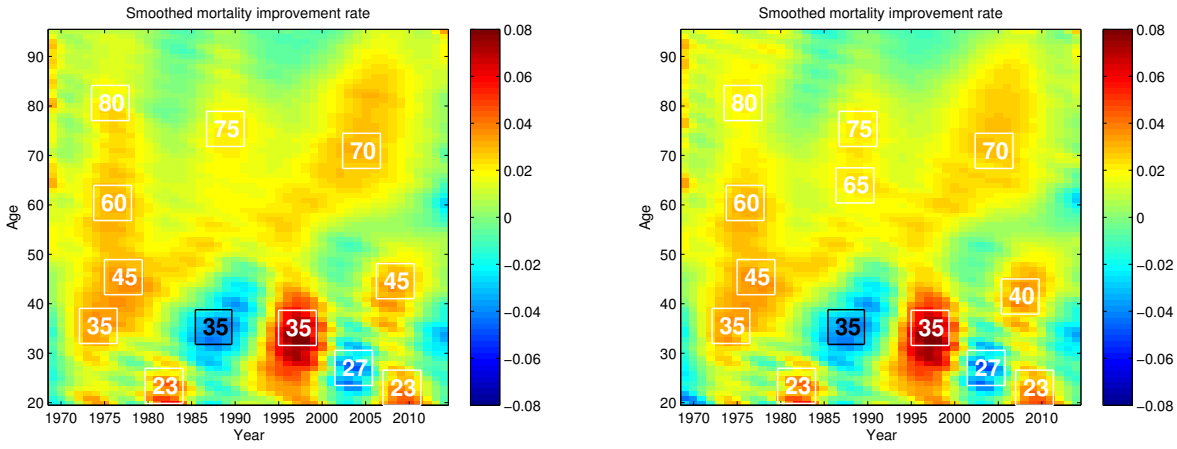


Figure 5: Heat maps of the smoothed mortality improvement rates $\tilde{Z}_{x,t}$ for the U.S. **males**.
- Left panel: Data set (i) (HMD, ages 20-95, years 1968-2014).
- Right panel: Data set (ii) (SSA, ages 20-95, years 1968-2014).

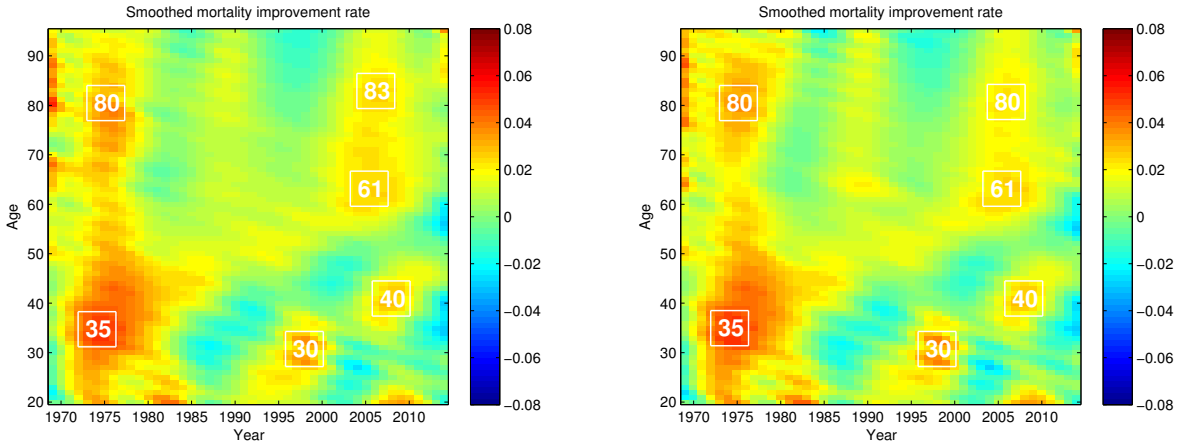


Figure 6: Heat maps of the smoothed mortality improvement rates $\tilde{Z}_{x,t}$ for the U.S. **females**.
- Left panel: Data set (i) (HMD, ages 20-95, years 1968-2014).
- Right panel: Data set (ii) (SSA, ages 20-95, years 1968-2014).

4.3 The Age/Period/Cohort Components in the CMI-09 Method

In the second stage of the CMI-09 method, the following additive APC model is fitted to the smoothed mortality improvement rates $\tilde{Z}_{x,t}$:

$$\tilde{Z}_{x,t} = a_x + k_t + g_c + e_{x,t},$$

where a_x, k_t, g_c and $e_{x,t}$ are the age, period, cohort and residual components, respectively. Figures 7 and 8 show the estimated age, period and cohort components based on the two data sets for U.S. males and females, respectively. The two data sets yield very similar age, period and cohort decompositions.

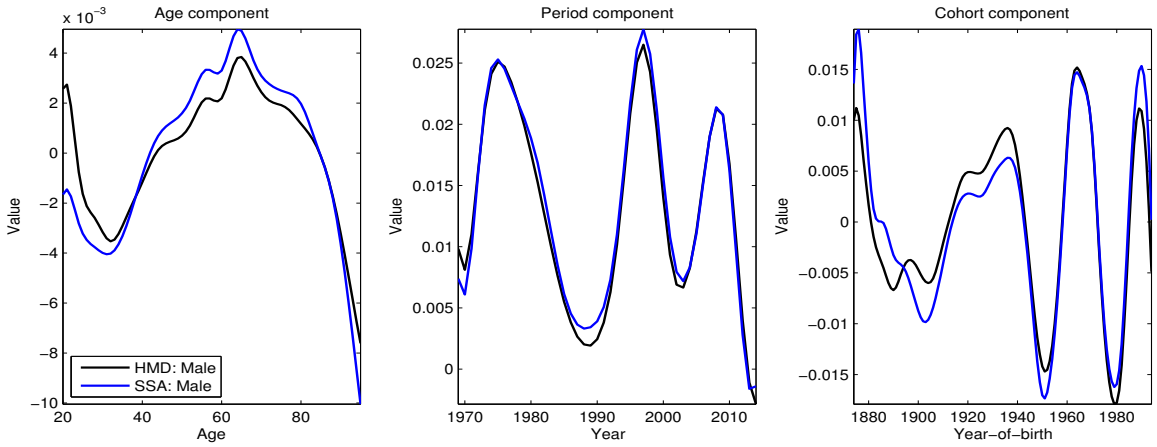


Figure 7: The estimated Age/Period/Cohort components obtained from the CMI-09 method, where the smoothed mortality improvement rates of U.S. **males** are fitted to an additive APC model.

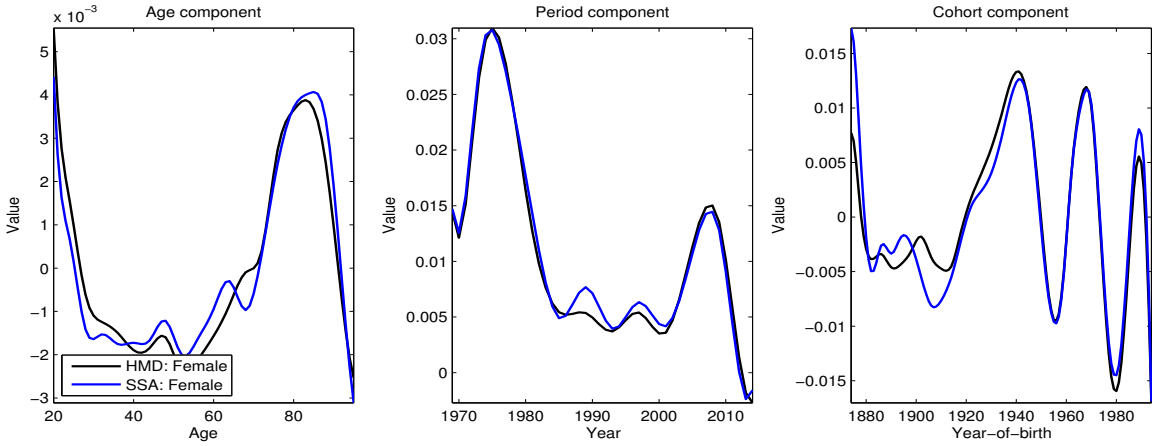


Figure 8: The estimated Age/Period/Cohort components obtained from the CMI-09 method, where the smoothed mortality improvement rates of U.S. **females** are fitted to an additive APC model.

4.4 The Residual Component in the CMI-09 Method

Figures 9 and 10 show the heat maps of the standardized residuals

$$\frac{e_{x,t}}{\text{s.d.}(e_{x,t})}$$

resulting from the two estimated additive APC models for U.S. males and females, respectively. For the readers' information, we indicate in the heat maps the ages at which the peaks and troughs are located.

If an estimated model is adequate, then the standardized residuals should exhibit a random pattern. The heat maps suggest that the CMI-09 APC model seems to be adequate for U.S. females, but not quite for U.S. males. In the residuals heat map for U.S. males, we observe three 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 the CMI-09 APC model. This problem may be fixed by adapting the CMI-09 APC model or by considering a more sophisticated APC model structure. These possible methods are examined in later parts of this report.

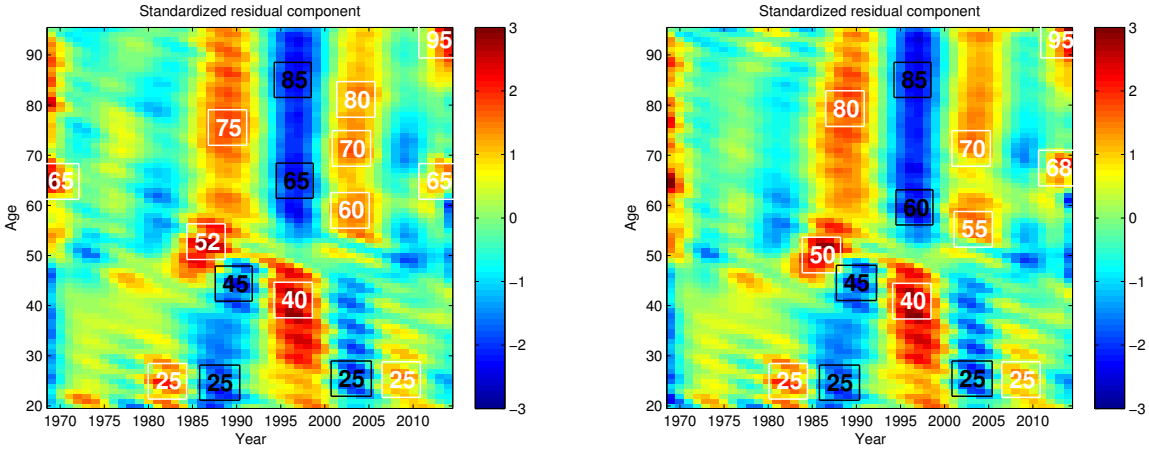


Figure 9: Heat maps of the standardize residuals calculated from the CMI-09 APC model $\tilde{Z}_{x,t} = a_x + k_t + g_c + e_{x,t}$, where the smoothed mortality improvement rates are fitted to the U.S. **males**.
- Left panel: Data set (i) (HMD, ages 20-95, years 1968-2014).
- Right panel: Data set (ii) (SSA, ages 20-95, years 1968-2014).

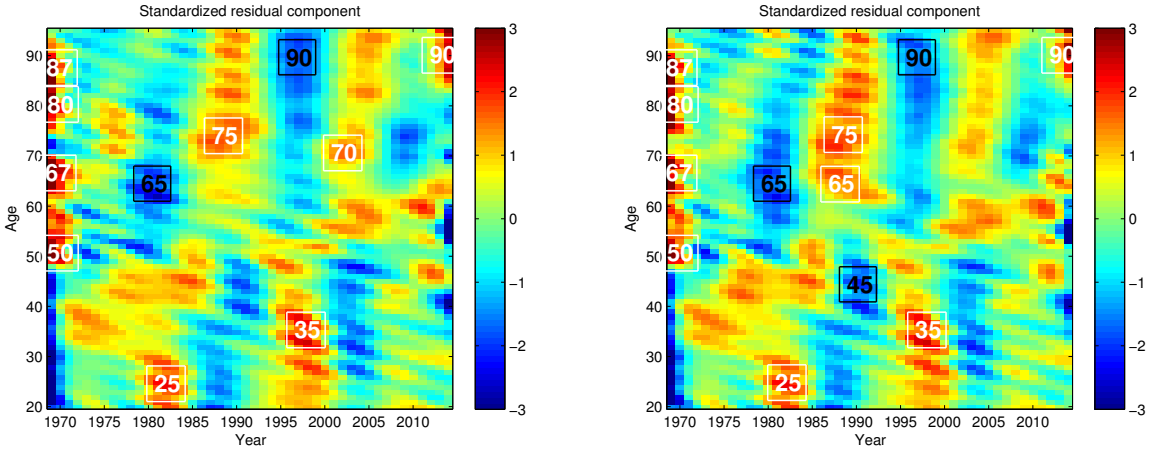


Figure 10: Heat maps of the standardized residuals calculated from the CMI-09 APC model $\tilde{Z}_{x,t} = a_x + k_t + g_c + e_{x,t}$, where the smoothed mortality improvement rates are fitted to the U.S. **females**.
- Left panel: Data set (i) (HMD, ages 20-95, years 1968-2014).
- Right panel: Data set (ii) (SSA, ages 20-95, years 1968-2014).

4.5 The Impact of Different Age Breakpoints on the Residual Component

One possible way to ameliorate the problem of residuals clustering is to introduce an age breakpoint. In this sub-section, we study the impact of introducing an age breakpoint on the residual component. Six age breakpoints, 45, 50, 55, 60, 65 and 70, are considered. For example, when the age breakpoint is set to 45, we fit one APC model to the smoothed improvement rates over ages 20 to 45 and another APC model to the smoothed improvement rates over ages 46 to 95. The residual component is then recalculated.

The results are presented in Figures 11 and 12 (for males) and Figures 13 and 14 (for females). In all figures, the top panels show the residual component when no age breakpoint is used (i.e., what was shown in Figures 9 and 10), while the other panels show the residual components when various age breakpoints (indicated by the black dashed lines) are used.

For males, the use of an age breakpoint dampens the patterns found in the original residuals. Using the Anderson-Darling test, it is found that the distribution of the standardized residuals is the closest to a standard normal distribution when the age breakpoint is set to 55.[4] This optimal age breakpoint is in line with what the heat maps shown in Figure 9, where we observe during 1996-1998 the standardized residuals change sign at age 55. For females, the effect of using an age breakpoint is minimal, no matter where the age breakpoint is located.

[4]Further analyses on the normality of standardized residuals are presented in Section 6.3.

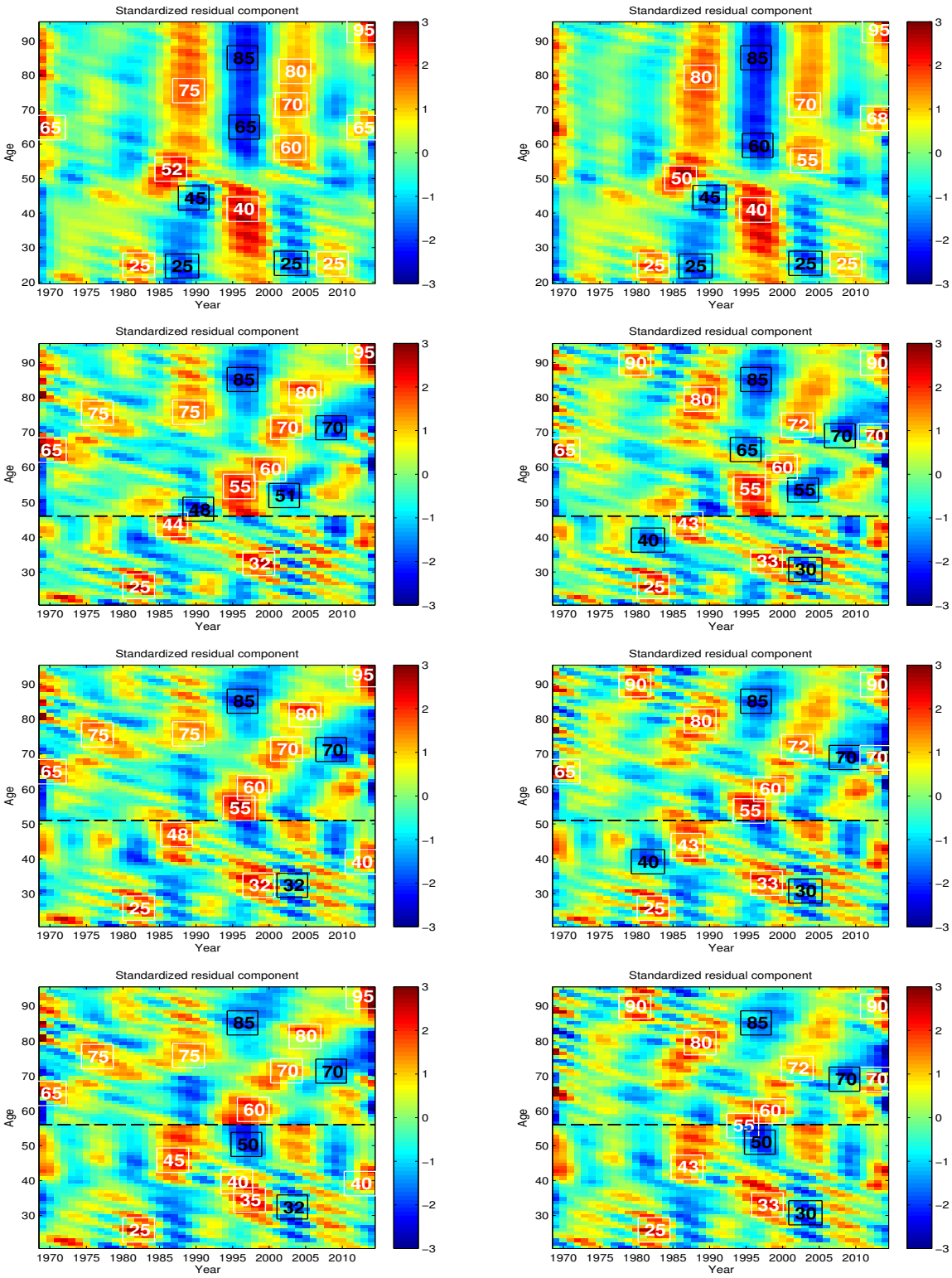


Figure 11: Heat maps of the standardized residuals calculated from the CMI-09 APC model $\tilde{Z}_{x,t} = a_x + k_t + g_c + e_{x,t}$, where the smoothed mortality improvement rates are fitted to the U.S. **males**. The dotted lines represent the "age breakpoints".
- Left panel: Data set (i) (HMD, ages 20-95, years 1968-2014).
- Right panel: Data set (ii) (SSA, ages 20-95, years 1968-2014).

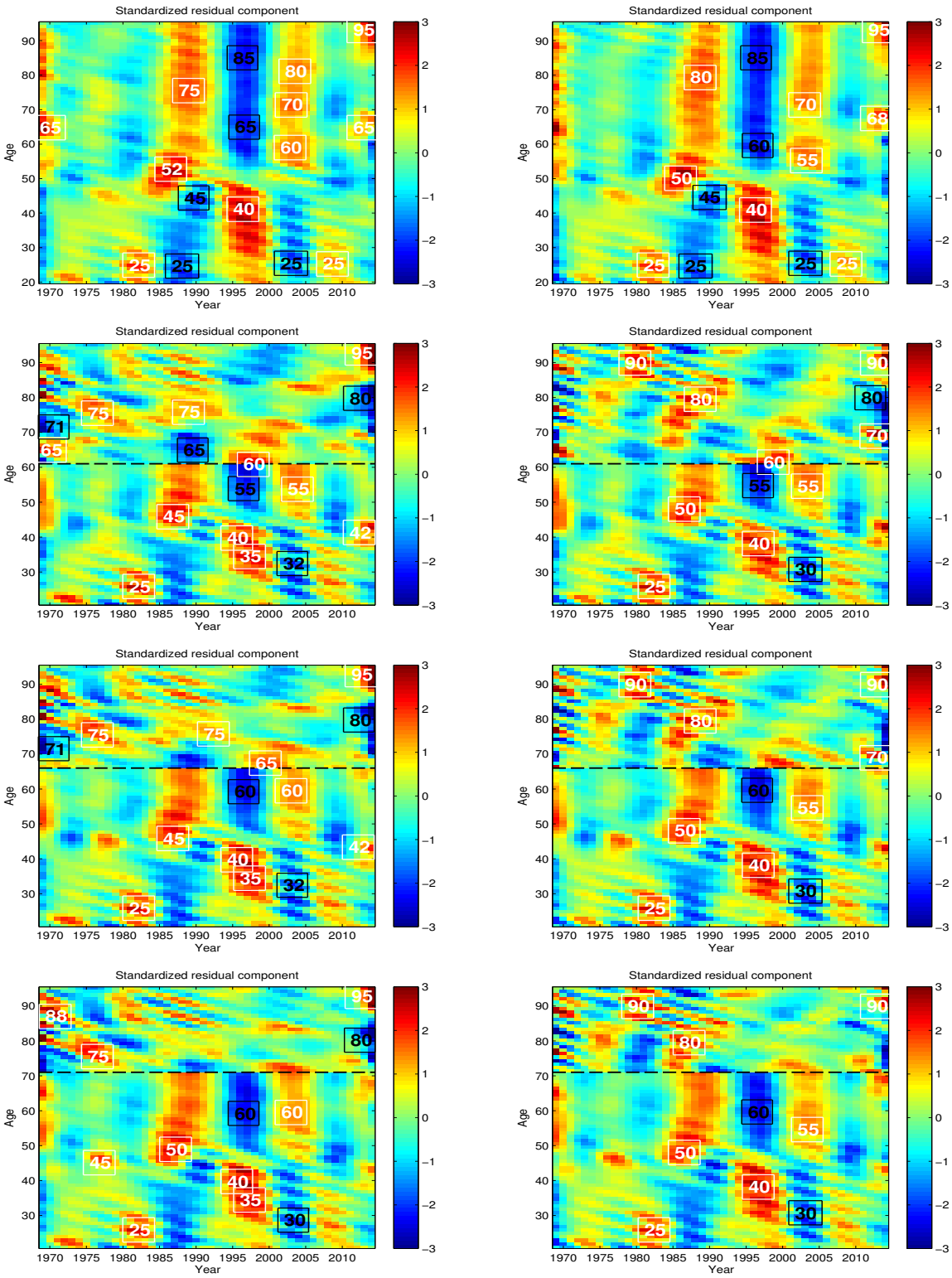


Figure 12: Heat maps of the standardized residuals calculated from the CMI-09 APC model $\tilde{Z}_{x,t} = a_x + k_t + g_c + e_{x,t}$, where the smoothed mortality improvement rates are fitted to the U.S. **males**. The dotted lines represent the "age breakpoints".
- Left panel: Data set (i) (HMD, ages 20-95, years 1968-2014).
- Right panel: Data set (ii) (SSA, ages 20-95, years 1968-2014).

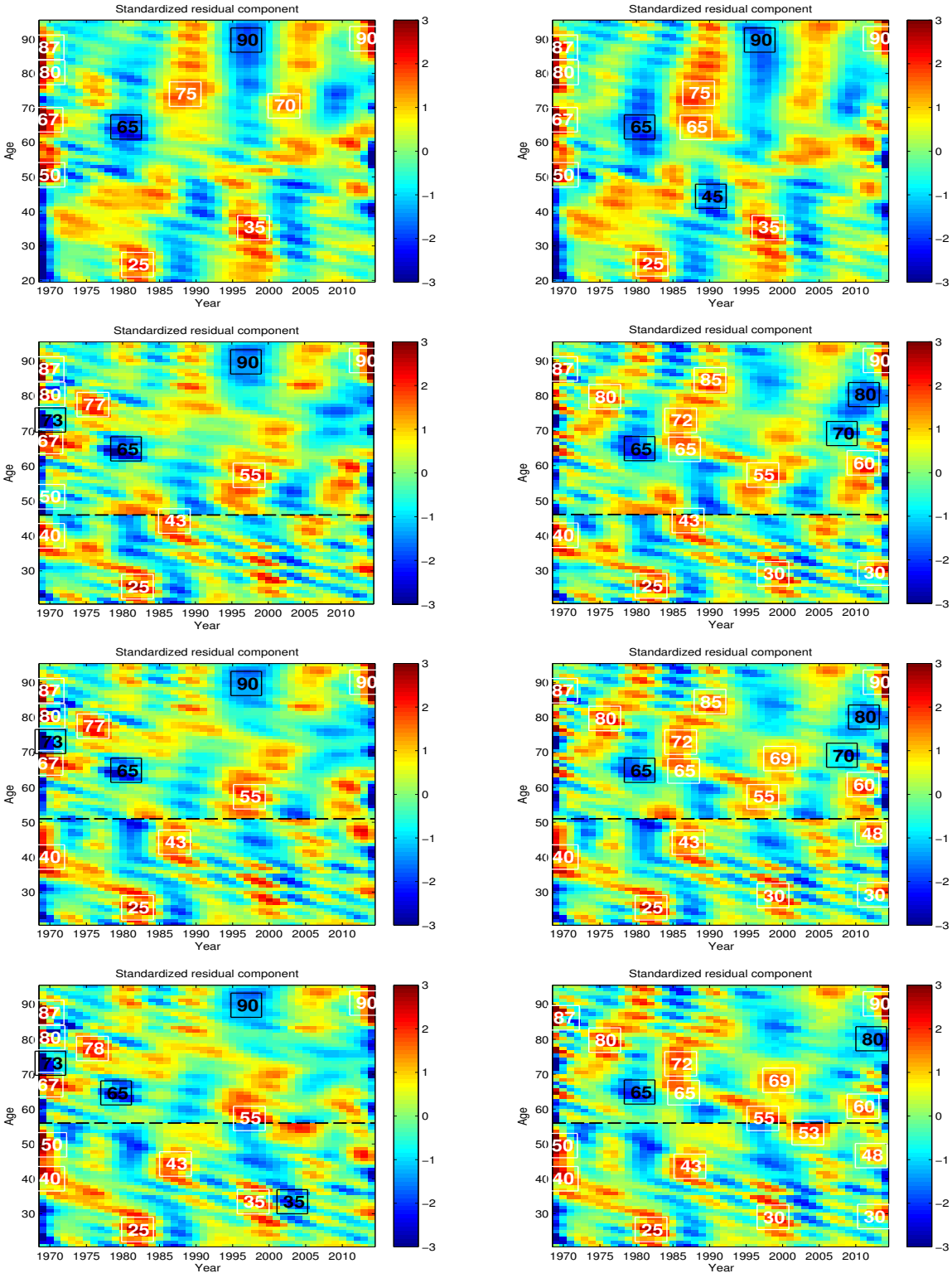


Figure 13: Heat maps of the standardized residuals calculated from the CMI-09 APC model $\tilde{Z}_{x,t} = a_x + k_t + g_c + e_{x,t}$, where the smoothed mortality improvement rates are fitted to the U.S. **females**. The dotted lines represent the "age breakpoints".
- Left panel: Data set (i) (HMD, ages 20-95, years 1968-2014).
- Right panel: Data set (ii) (SSA, ages 20-95, years 1968-2014).

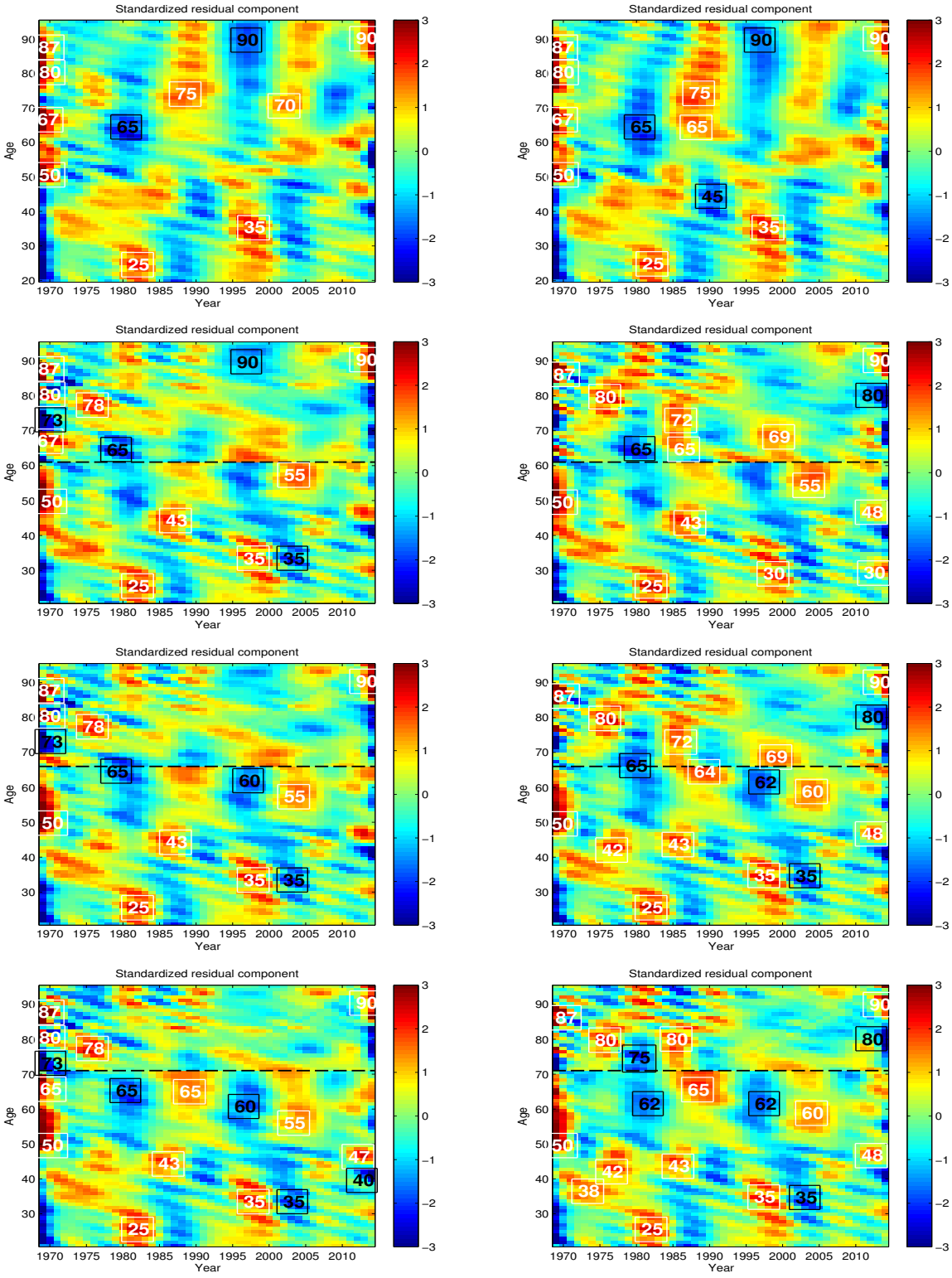


Figure 14: Heat maps of the standardized residuals calculated from the CMI-09 APC model $\tilde{Z}_{x,t} = a_x + k_t + g_c + e_{x,t}$, where the smoothed mortality improvement rates are fitted to the U.S. **females**. The dotted lines represent the "age breakpoints".
- Left panel: Data set (i) (HMD, ages 20-95, years 1968-2014).
- Right panel: Data set (ii) (SSA, ages 20-95, years 1968-2014).

4.6 Concluding Remark

When the CMI-09 approach is applied to U.S. male mortality experience, large vertical residual clusters are observed in 1996 to 1998. The clusters indicate some sort of 'regime-switching': the standardized residuals before and after age 55 are of opposite signs. This represents an indication that the CMI-09 APC model is unable to pick up some features that are specific to U.S. male mortality experience.

To confirm this conjecture, here we apply the CMI-09 decomposition method to data from the population of England and Wales (E&W) over the same age range and sample period, and examine if similar clusters are found in the residual heat maps for the E&W models. The results are shown in Figures 15 (males) and 16 (females). The following observations are made:

- Of our biggest concern are the large vertical clusters that are found in the plot for U.S. males during 1996-1998. These vertical clusters are not found in the residual heat maps for the E&W models, suggesting that they are due to data-specific characteristics that have not been picked up by the APC structure used.
- The age breakpoint (at around age 55) found in the plot for U.S. males is not observed in the plots derived from the E&W data, suggesting that it is also a data-specific characteristic.
- Some obvious diagonal bands are observed in the residuals heat maps for the E&W models. These bands suggest that E&W mortality is subject to highly significant cohort effects, which have not been adequately captured by the APC structure (even though the structure contains a cohort effect term).

The results shown in Figures 15 and 16 further confirm that we need to adapt the CMI-09 APC model to better suit U.S. historical mortality. Introducing an age breakpoint (discussed in the previous sub-section) ameliorates the problem of residual clustering, but may result in inconsistencies between mortality projections for younger and older ages. Another possible direction is to consider alternative, more sophisticated APC model structures. Research along this direction is presented in Section 6 of this report.

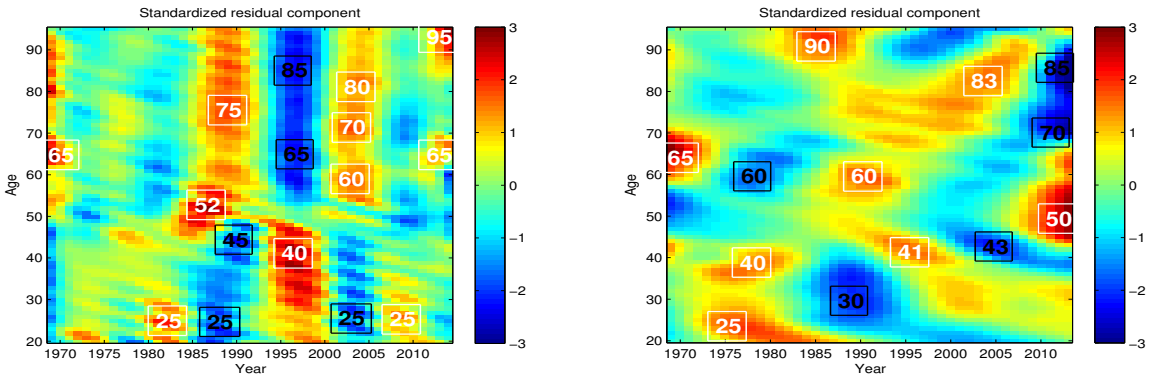


Figure 15: Heat maps of the standardized residuals of the CMI-09 APC model $\tilde{Z}_{x,t} = a_x + k_t + g_c + e_{x,t}$ fitted to the smoothed mortality improvement rates for **U.S. males** (left panel) and **E&W males** (right panel).

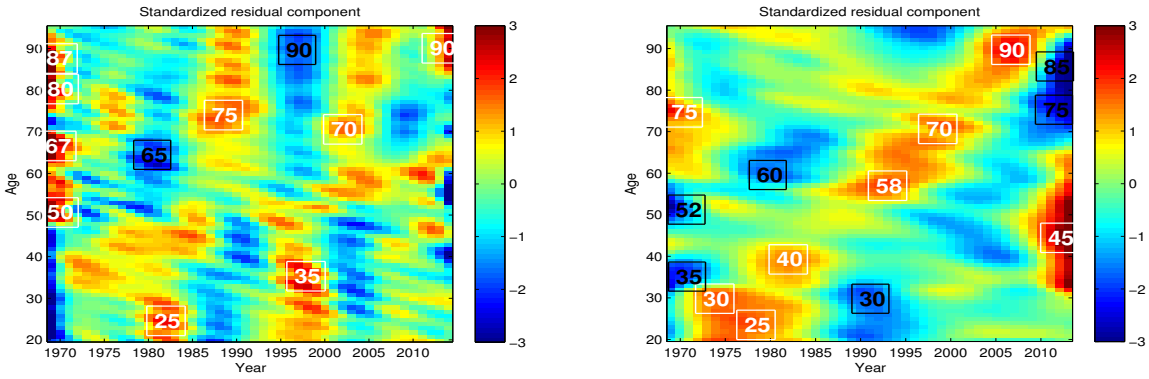


Figure 16: Heat maps of the standardized residuals of the CMI-09 APC model $\tilde{Z}_{x,t} = a_x + k_t + g_c + e_{x,t}$ fitted to the smoothed mortality improvement rates for **U.S. females** (left panel) and **E&W females** (right panel).

5 Testing the Robustness of the CMI-09 Method

It is important to ensure that the resulting A/P/C components remain largely unchanged when there are small changes to the data input or model set-up. In this section, we perform several robustness tests on the CMI-09 decomposition method. Specifically, we test the robustness of the decomposition result to (1) changes in the calibration window, (2) changes in the age range, (3) the choice of parameter constraints, and (4) the exclusion of the oldest/newest cohorts.

5.1 Changes in the Calibration Window

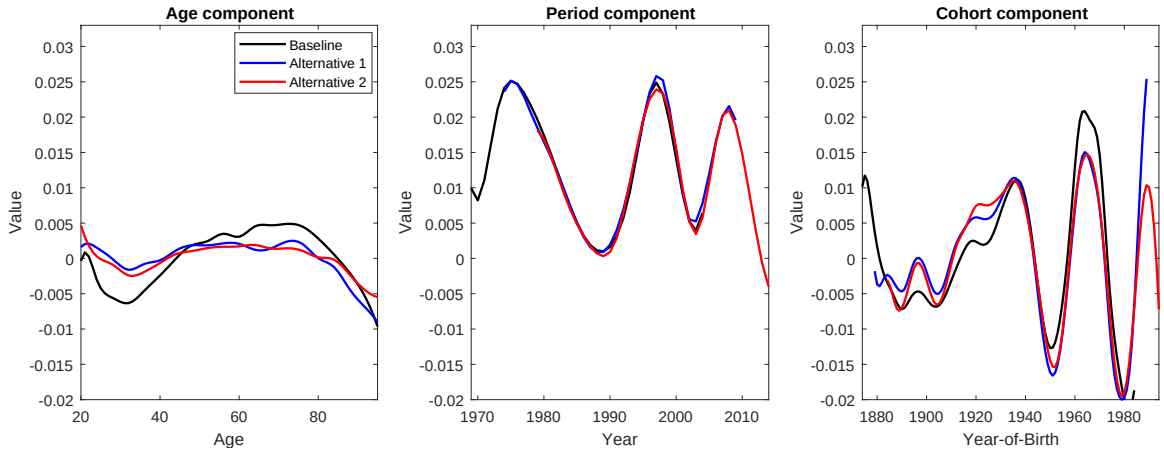
In this sub-section, we apply the CMI-09 approach to three calibration windows, which have the same length but different starting/ending years. This set-up is to mimic the situation when the model is updated every 5 years. The following table summarizes the calibration windows under consideration:

	Starting Age	Ending Age	Starting Year	Ending Year	Length of Calibration Window
Baseline	20	95	1968	2004	37 years
Alternative 1			1973	2009	37 years
Alternative 2			1978	2014	37 years

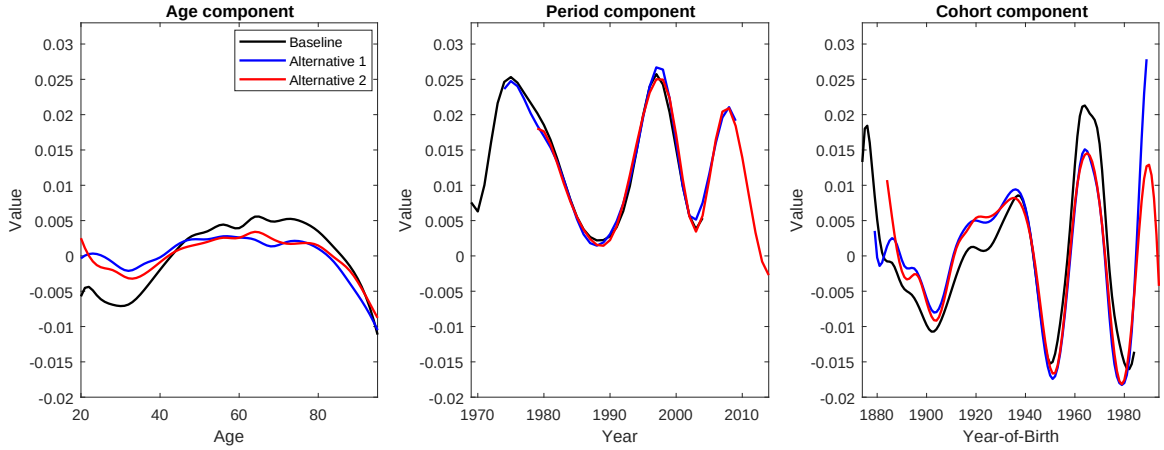
Figures 17 (males) and 18 (females) show the age, period and cohort components estimated from the CMI-09 approach when the different calibration windows are used. The following observations are made:

- The lengths of the period and cohort components remain the same, as the length of the calibration window is always 37 years.
- As the calibration window is changed, the values of the age, period and cohort components change, but the overall shapes of the components remain fairly the same.
- Among all three components, the age component appears to be the most sensitive to the change in calibration window.
- No matter what calibration window is used, the cohort component always fluctuates around zero because of the imposed parameter constraints.

Figures 19 (males) and 20 (females) show the (standardized) residual components when different calibration windows are used. The peaks and valleys in the residual plots remain in approximately the same locations as the calibration window is changed, indicating that the residual component is fairly insensitive to changes in the calibration window.

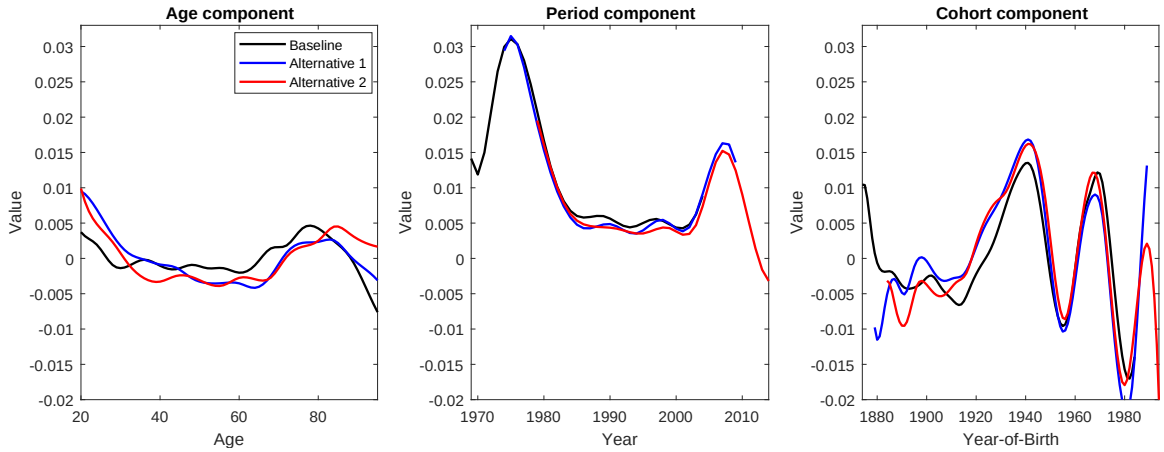


(a) U.S. Males: Dataset (i), HMD data

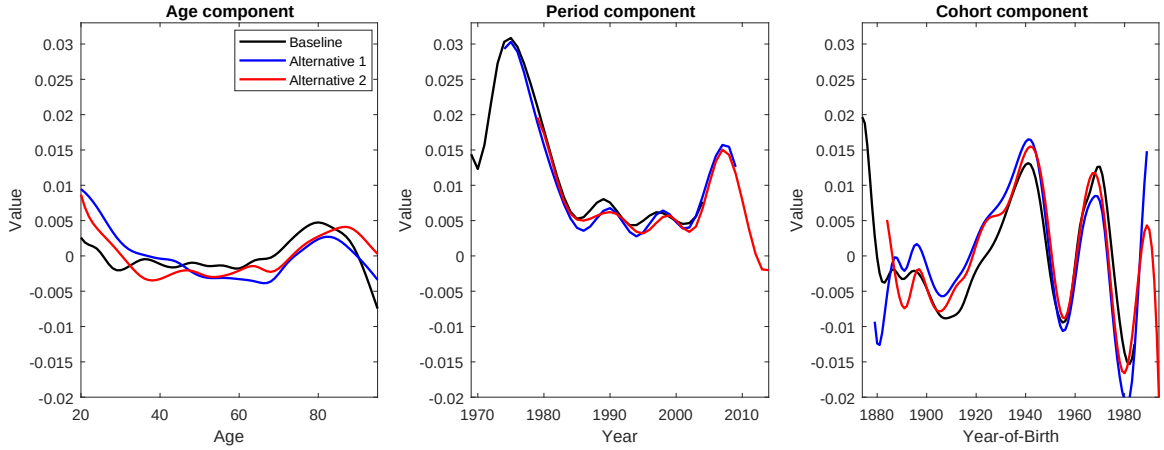


(b) U.S. Males: Dataset (ii), SSA data

Figure 17: The estimated Age/Period/Cohort components obtained from the CMI-09 method when three different calibration windows are used, U.S. **males**.
- Baseline: ages 20 to 95, years 1968 to 2004 (covering years-of-birth 1873 to 1984).
- Alternative 1: ages 20 to 95, years 1973 to 2009 (covering years-of-birth 1878 to 1989).
- Alternative 2: ages 20 to 95, years 1978 to 2014 (covering years-of-birth 1883 to 1994).



(a) U.S. Females: Dataset (i), HMD data



(b) U.S. Females: Dataset (ii), SSA data

Figure 18: The estimated Age/Period/Cohort components obtained from the CMI-09 method when three different calibration windows are used, U.S. **females**.
- Baseline: ages 20 to 95, years 1968 to 2004 (covering years-of-birth 1873 to 1984).
- Alternative 1: ages 20 to 95, years 1973 to 2009 (covering years-of-birth 1878 to 1989).
- Alternative 2: ages 20 to 95, years 1978 to 2014 (covering years-of-birth 1883 to 1994).

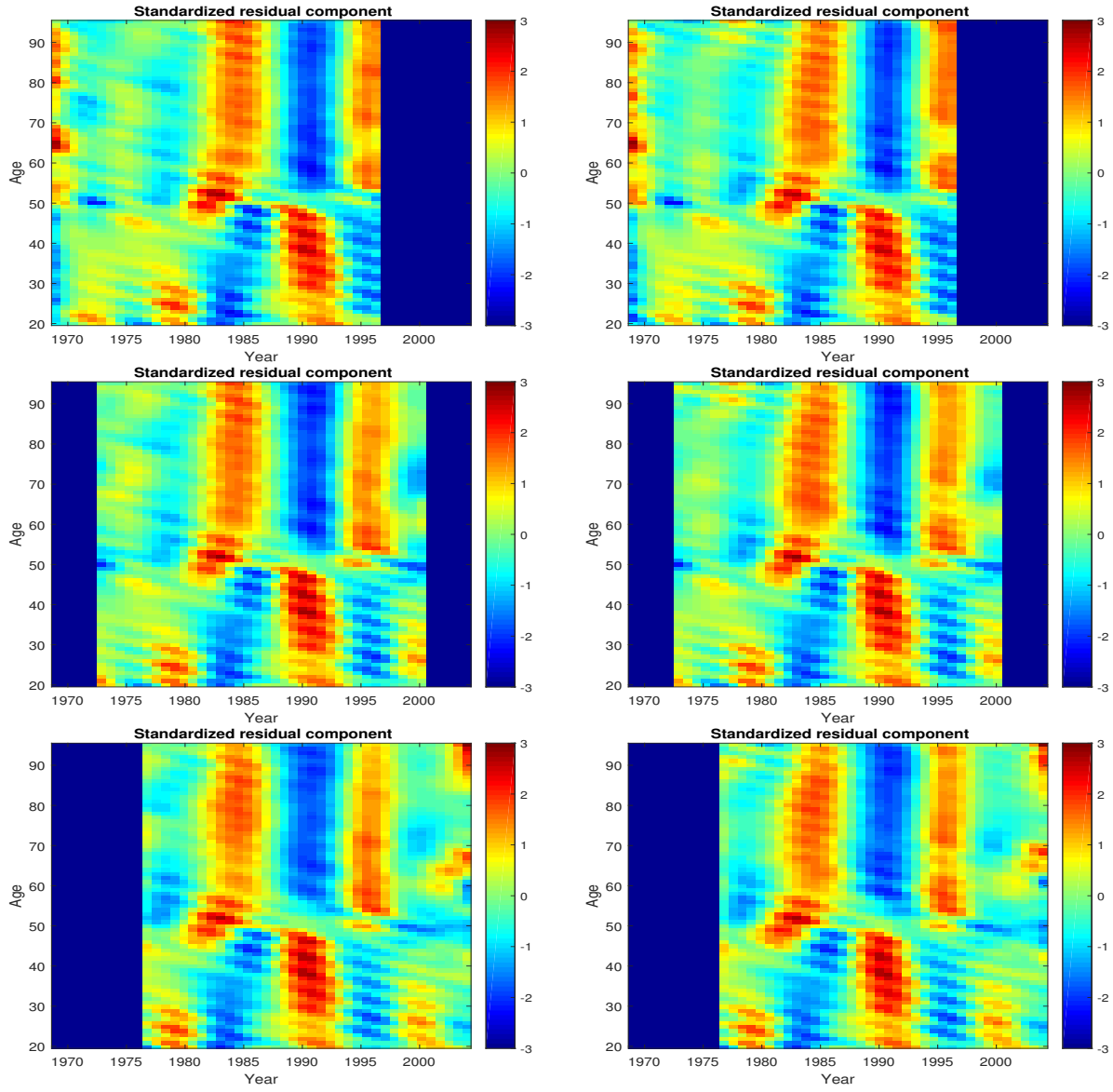


Figure 19: Heat maps of the standardized residuals of the CMI-09 APC model $\tilde{Z}_{x,t} = a_x + k_t + g_c + e_{x,t}$ fitted to the smoothed mortality improvement rates for U.S. **males**, when different calibration windows are used.
- Left panels: Data set (i) (HMD).
- Right panels: Data set (ii) (SSA).
- Top panels: Baseline, ages 20-95, years 1968-2004.
- Middle panels: Alternative 1, ages 20 to 95, years 1973 to 2009.
- Lower panels: Alternative 2, ages 20 to 95, years 1978 to 2014.

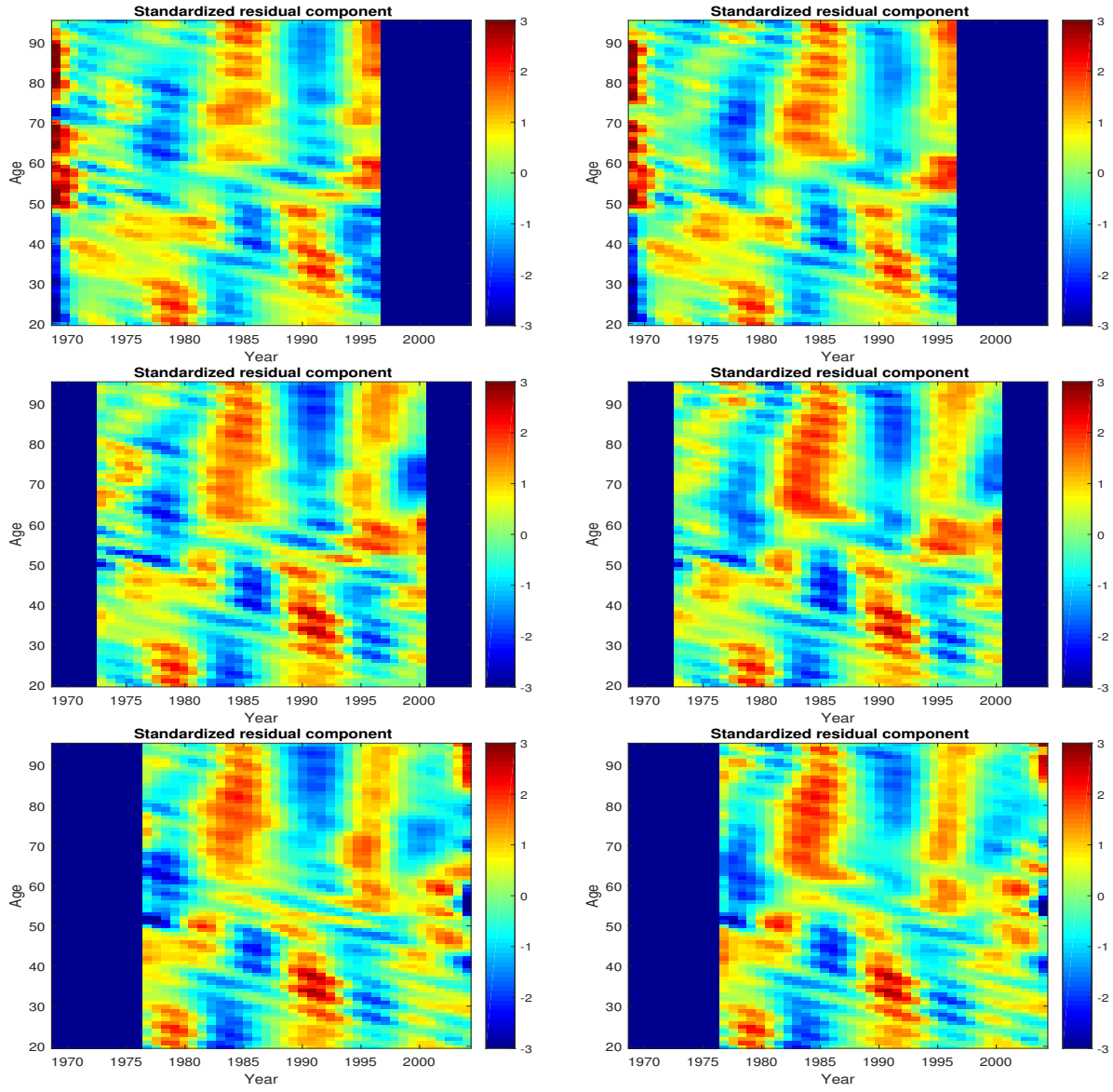


Figure 20: Heat maps of the standardized residuals of the CMI-09 APC model $\tilde{Z}_{x,t} = a_x + k_t + g_c + e_{x,t}$ fitted to the smoothed mortality improvement rates for U.S. **females**, when different calibration windows are used.
- Left panels: Data set (i) (HMD).
- Right panels: Data set (ii) (SSA).
- Top panels: Baseline, ages 20-95, years 1968-2004.
- Middle panels: Alternative 1, ages 20 to 95, years 1973 to 2009.
- Lower panels: Alternative 2, ages 20 to 95, years 1978 to 2014.

5.2 Changes in the Age Range

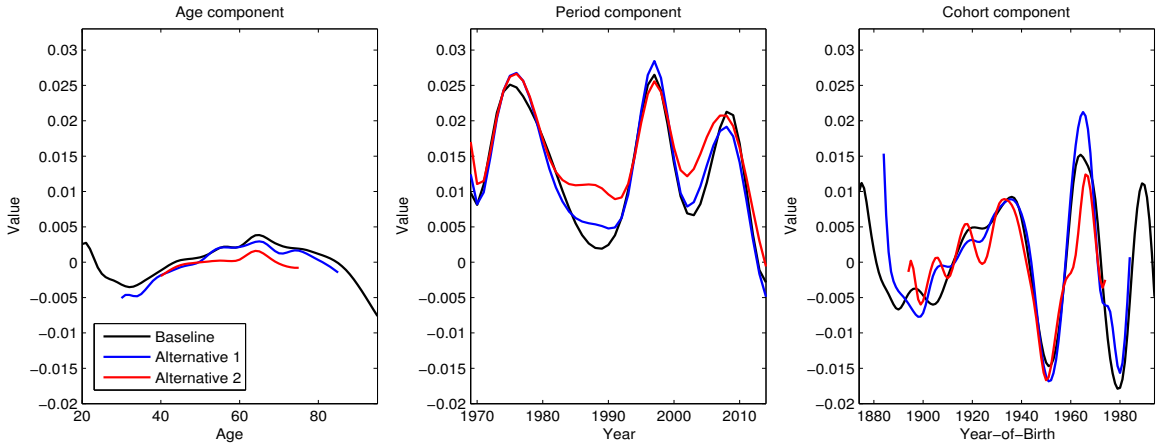
In this sub-section, we apply the CMI-09 approach to different age ranges:

	Starting Age	Ending Age	Starting Year	Ending Year	Number of Ages
Baseline	20	95			76
Alternative 1	30	85	1968	2014	56
Alternative 2	40	75			36

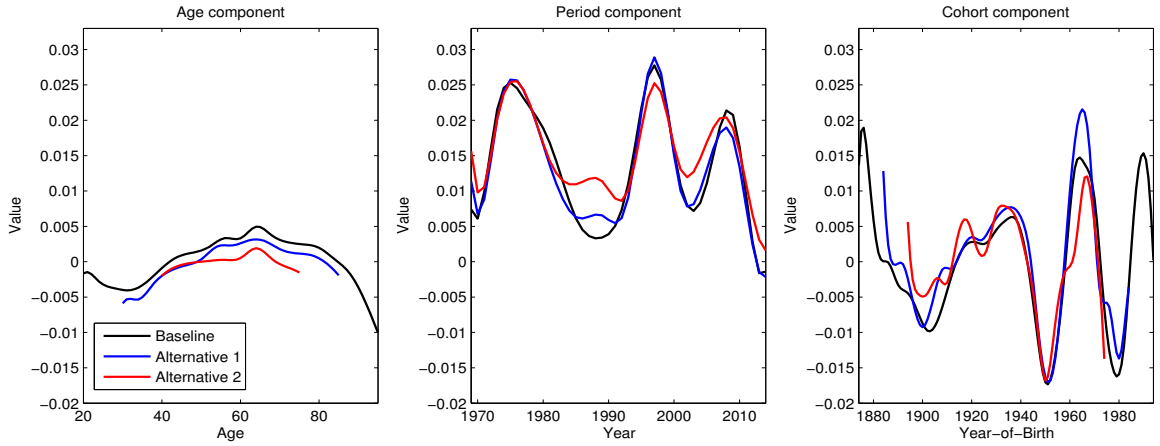
Figures 21 (males) and 22 (females) show the age, period and cohort components estimated from the CMI-09 approach when different age ranges are used. The following observations can be made:

- The lengths of the age and cohort components become smaller as the age range is shortened.
- As the age range window is shortened, the values of the age, period and cohort components change, but the overall shapes of the components remain fairly the same.
- No matter what age range is used, the cohort component always fluctuates around zero because of the imposed parameter constraints.

Figures 23 (males) and 24 (females) show the residual components when different age ranges are used. It is found that the residual component is rather insensitive to changes in the age range used.

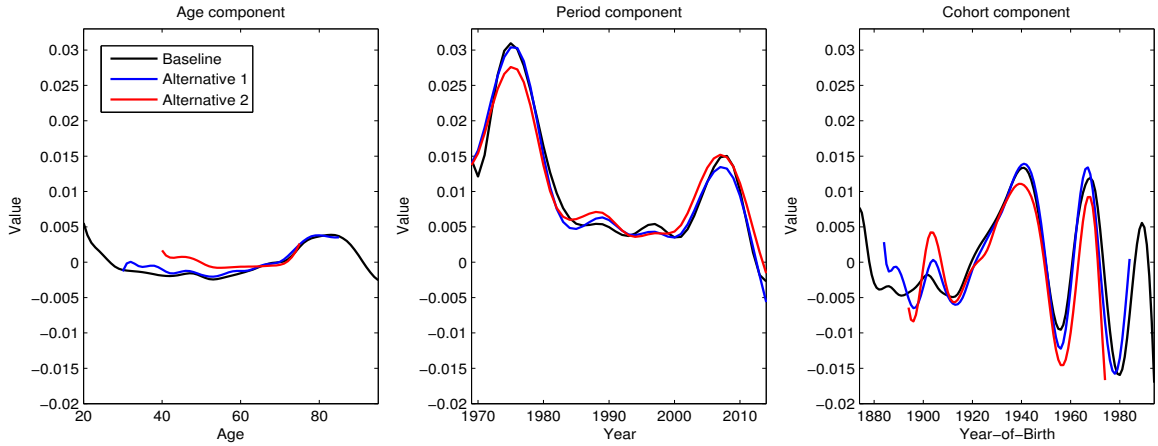


(a) U.S. Males: Dataset (i), HMD data

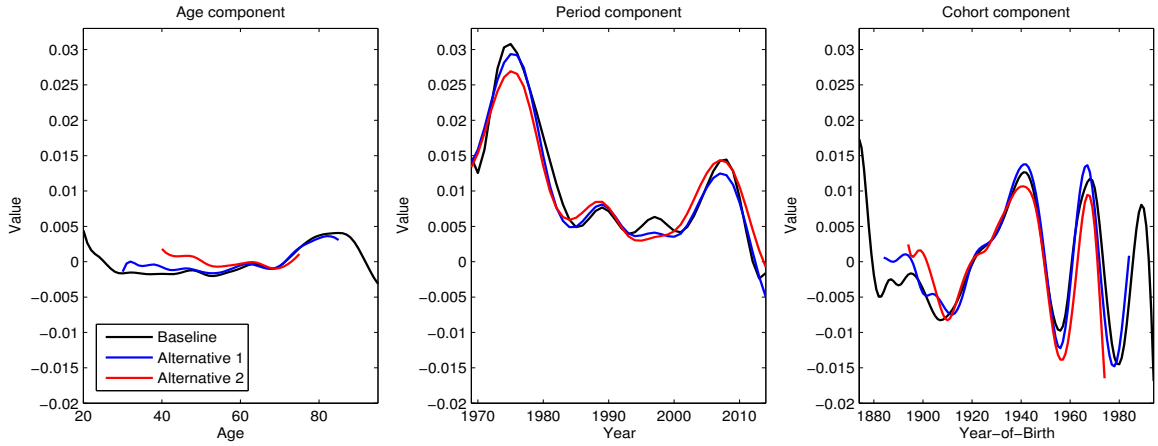


(b) U.S. Males: Dataset (ii), SSA data

Figure 21: The estimated Age/Period/Cohort components obtained from the CMI-09 method when different age ranges are used, U.S. **males**.
- Baseline: ages 20 to 95, years 1968 to 2014 (covering years-of-birth 1873 to 1994).
- Alternative 1: ages 30 to 85, years 1968 to 2014 (covering years-of-birth 1883 to 1984).
- Alternative 2: ages 40 to 75, years 1968 to 2014 (covering years-of-birth 1893 to 1974).



(a) U.S. Females: Dataset (i), HMD data



(b) U.S. Females: Dataset (ii), SSA data

Figure 22: The estimated Age/Period/Cohort components obtained from the CMI-09 method when different age ranges are used, U.S. **females**.
- Baseline: ages 20 to 95, years 1968 to 2014 (covering years-of-birth 1873 to 1994).
- Alternative 1: ages 30 to 85, years 1968 to 2014 (covering years-of-birth 1883 to 1984).
- Alternative 2: ages 40 to 75, years 1968 to 2014 (covering years-of-birth 1893 to 1974).

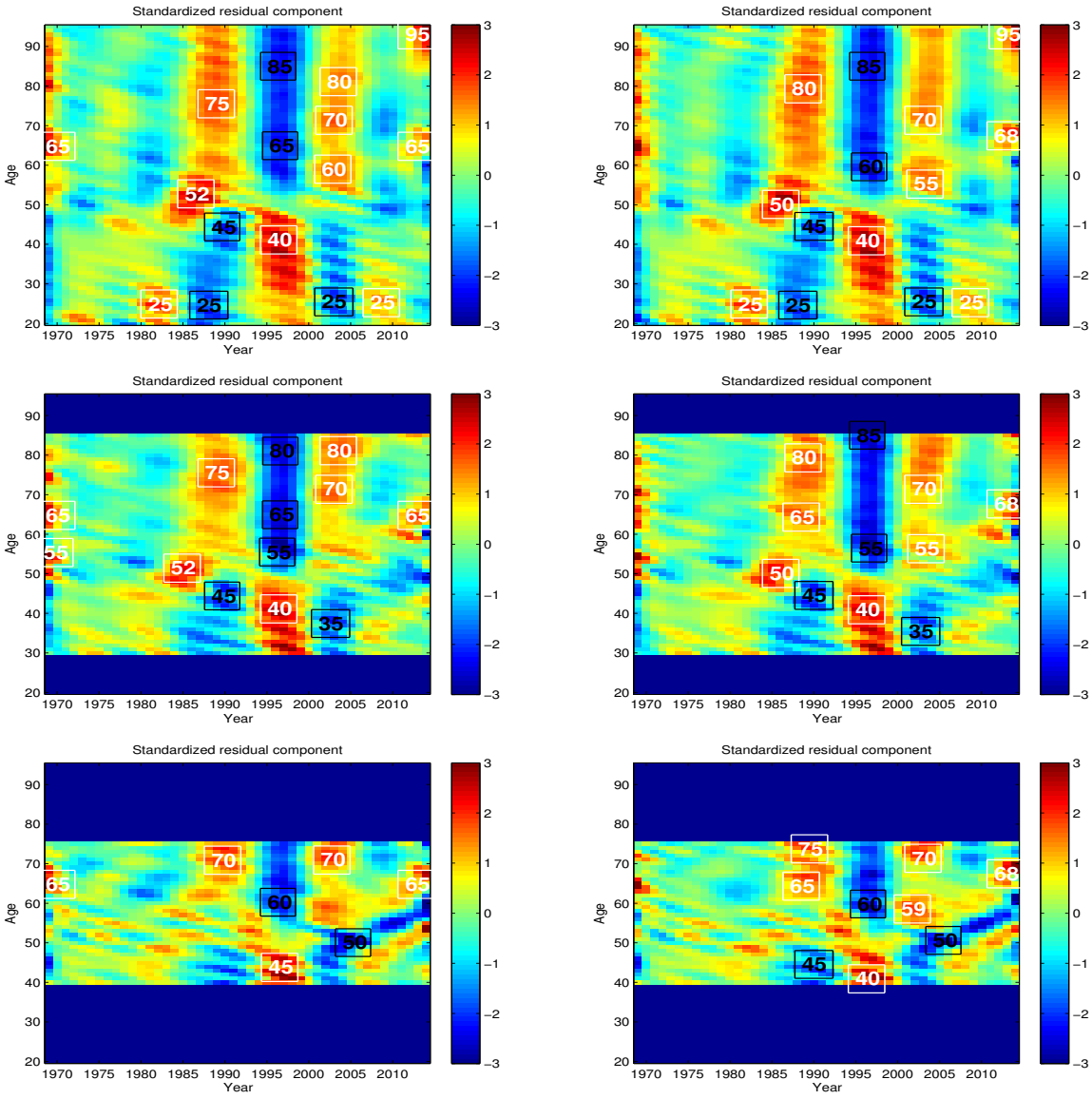


Figure 23: Heat maps of the standardized residuals of the CMI-09 APC model $\tilde{Z}_{x,t} = a_x + k_t + g_c + e_{x,t}$ fitted to the smoothed mortality improvement rates for U.S. **males**, when different age ranges are used.
- Left panels: Data set (i) (HMD).
- Right panels: Data set (ii) (SSA).
- Top panels: Baseline, ages 20 to 95, years 1968 to 2014.
- Middle panels: Alternative 1, ages 30 to 85, years 1968 to 2014
- Lower panels: Alternative 2, ages 40 to 75, years 1968 to 2014.

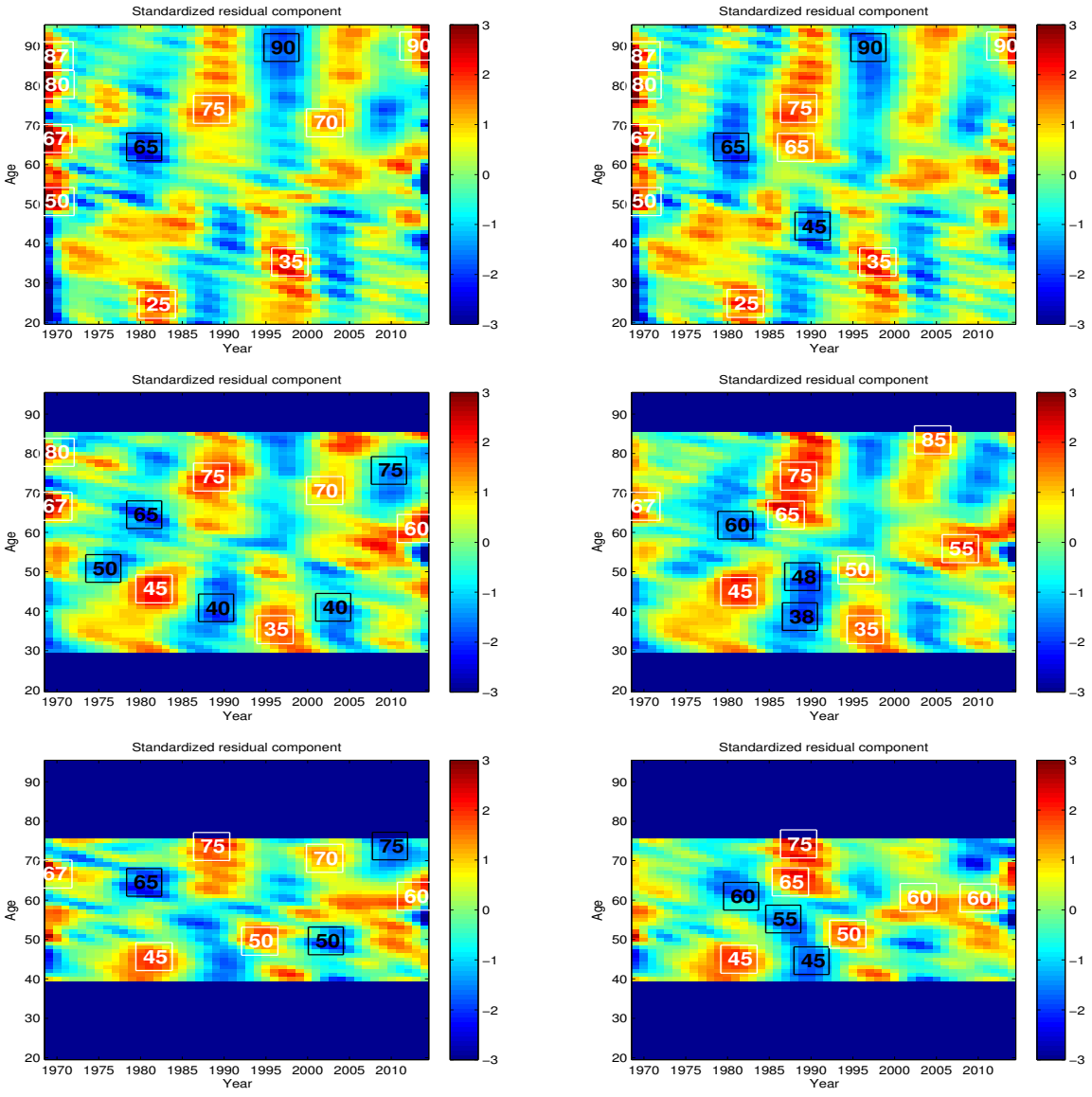


Figure 24: Heat maps of the standardized residuals of the CMI-09 APC model $\tilde{Z}_{x,t} = a_x + k_t + g_c + e_{x,t}$ fitted to the smoothed mortality improvement rates for U.S. **females**, when different age ranges are used.
- Left panels: Data set (i) (HMD).
- Right panels: Data set (ii) (SSA).
- Top panels: Baseline, ages 20 to 95, years 1968 to 2014.
- Middle panels: Alternative 1, ages 30 to 85, years 1968 to 2014
- Lower panels: Alternative 2, ages 40 to 75, years 1968 to 2014.

5.3 The Choice of Parameter Constraints

In the CMI-09 method, three parameter constraints are used to stipulate the parameter uniqueness. In this section, we examine how the estimation results may change when these constraints are altered. The following three sets of constraints are considered:

- **Baseline:**
$$\sum_{x=x_0}^{x_1} a_x = 0, \quad \sum_{c=t_0-x_1}^{t_1-x_0} g_c = 0, \quad \sum_{c=t_0-x_1}^{t_1-x_0} c \cdot g_c = 0.$$
These constraints ensures that both the age and cohort components fluctuate around zero, and that the cohort component exhibits no linear trend.

- **Alternative 1:**
$$\sum_{t=t_0}^{t_1} k_t = 0, \quad \sum_{c=t_0-x_1}^{t_1-x_0} n_c \cdot g_c = 0, \quad \sum_{c=t_0-x_1}^{t_1-x_0} n_c \cdot c \cdot g_c = 0.$$
In the above, n_c stands for the number of data points associated with year-of-birth c.[5] By including n_c, the cohorts for which we have more information about play a more significant role in the parameter constraints. Using these constraints, the period and (weighted) cohort components will fluctuate around zero, and the weighted cohort component will exhibit no linear trend. We remark here that this collection of constraints is also considered in the paper by Hunt and Blake (2015).

- **Alternative 2:**
$$\sum_{t=t_0}^{t_1} k_t = 0, \quad \sum_{c=t_0-x_1}^{t_1-x_0} g_c = 0, \quad \sum_{c=t_0-x_1}^{t_1-x_0} c \cdot g_c = 0.$$
The same as alternative 1, except that n_c is not involved in the second and third constraints. These constraints ensure that both the period and cohort components will fluctuate around zero and that cohort component will exhibit no linear trend.

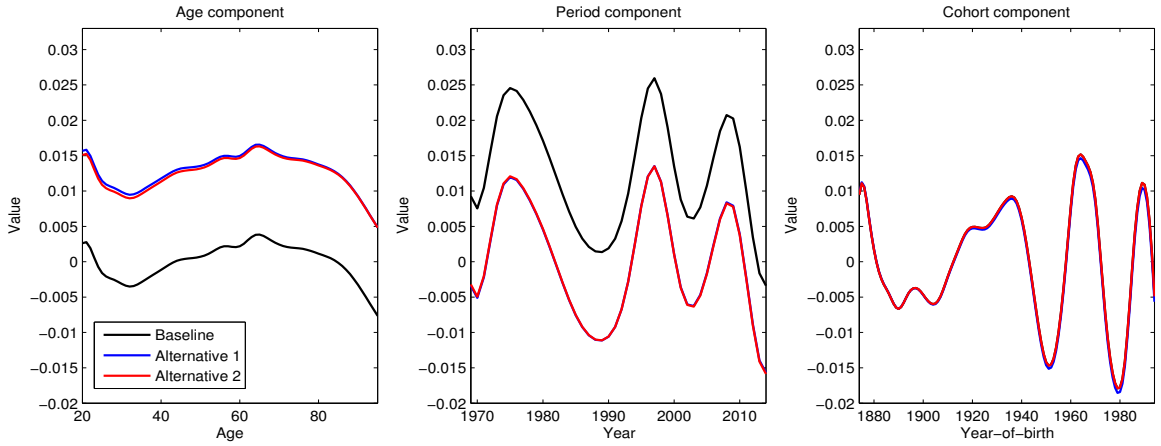
Figures 25 (males) and 26 (females) show the estimated age, period, and cohort components when different parameter constraints are used. The following observations are made:

- As we switch from the baseline to alternative 2, the cohort component remains the same, whereas the age and period components shift in a parallel manner. In fact, it can be proven mathematically that the increase in the magnitude of the age component is exactly the same as the reduction in the magnitude of the period component.

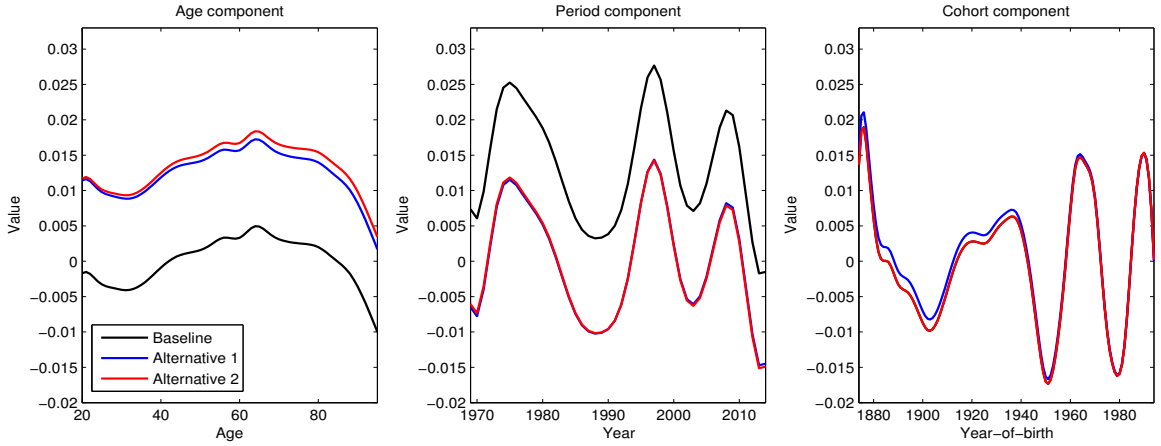
[5] Our data sample spans ages 20 to 95 and years 1968 to 2014. For the earliest cohort ($c = 1968 - 95 = 1873$), we have $n_{1873} = 1$ as there is only one data point (which corresponds to age 95 and year 1968) is associated with this cohort. Likewise, for the second earliest cohort $c = 1874$, we 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.

- Alternatives 1 and 2 yield very similar estimates, indicating that the weights (n_c's) have only modest impact on the resulting estimates.

We remark here that because the cohort component is very insensitive to the choice of constraints, the blue and red lines in the right panels of Figure 25 are very close to each other (and difficult to distinguish).

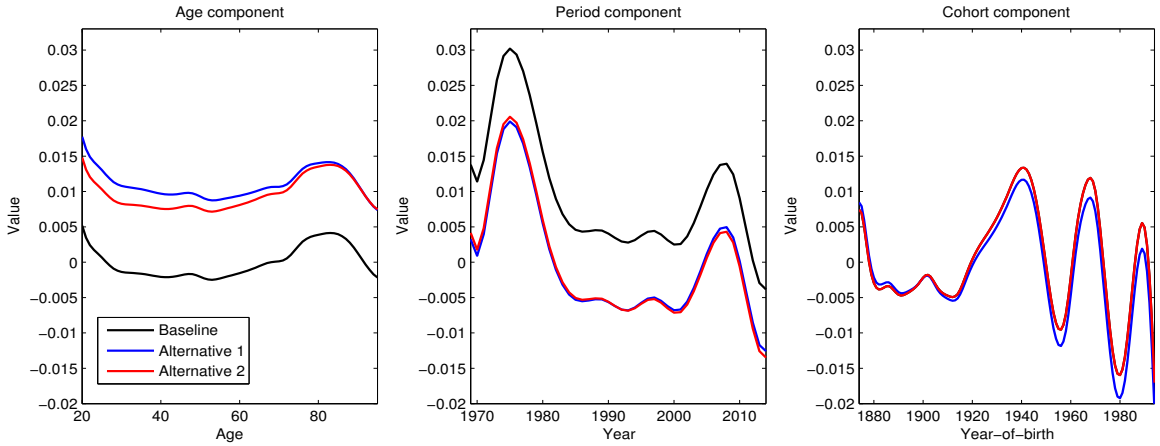


(a) U.S. Males: Dataset (i), HMD data

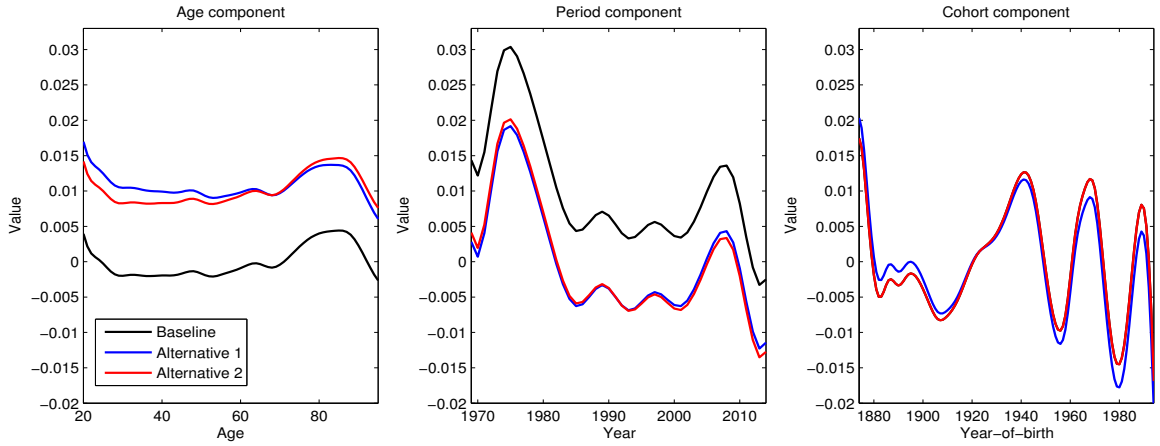


(b) U.S. Males: Dataset (ii), SSA data

Figure 25: The estimated Age/Period/Cohort components obtained from the CMI-09 method when different parameter constraints are used, U.S. **males**.
- Baseline: $\sum_x a_x = 0$, $\sum_c g_c = 0$ and $\sum_c c \cdot g_c = 0$.
- Alternative 1: $\sum_t k_t = 0$, $\sum_c n_c \cdot g_c = 0$ and $\sum_c n_c \cdot c \cdot g_c = 0$.
- Alternative 2: $\sum_t k_t = 0$, $\sum_c g_c = 0$ and $\sum_c c \cdot g_c = 0$.



(a) U.S. Females: Dataset (i), HMD data



(b) U.S. Females: Dataset (ii), SSA data

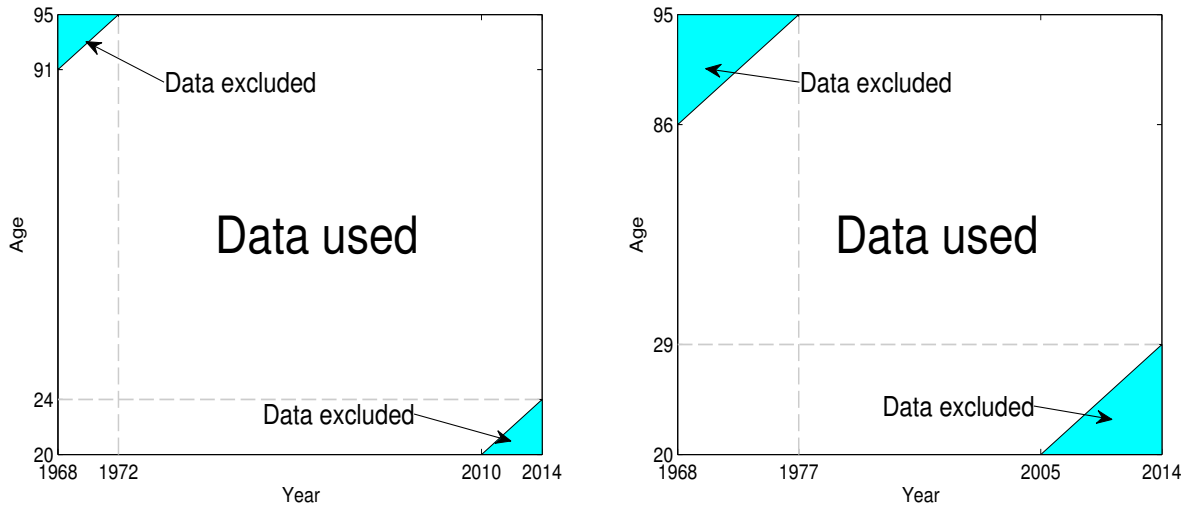
Figure 26: The estimated Age/Period/Cohort components obtained from the CMI-09 method when different parameter constraints are used, U.S. **females**.
- Baseline: $\sum_x a_x = 0$, $\sum_c g_c = 0$ and $\sum_c c \cdot g_c = 0$.
- Alternative 1: $\sum_t k_t = 0$, $\sum_c n_c \cdot g_c = 0$ and $\sum_c n_c \cdot c \cdot g_c = 0$.
- Alternative 2: $\sum_t k_t = 0$, $\sum_c g_c = 0$ and $\sum_c c \cdot g_c = 0$.

5.4 Exclusion of the Oldest/Newest Cohorts

There are only a few observations associated with the oldest/newest cohorts covered by the data sample. If the model is not stable, including these cohorts in the estimation process may distort the resulting age, period and cohort components. In this sub-section, we examine how the resulting age, period and cohort components may change when the oldest and newest cohorts are excluded. The following three situations are considered:

- **Baseline:** All available data are used.
- **Alternative 1:** The oldest and youngest five cohorts in the data sample are excluded.
- **Alternative 2:** The oldest and youngest ten cohorts in the data sample are excluded.

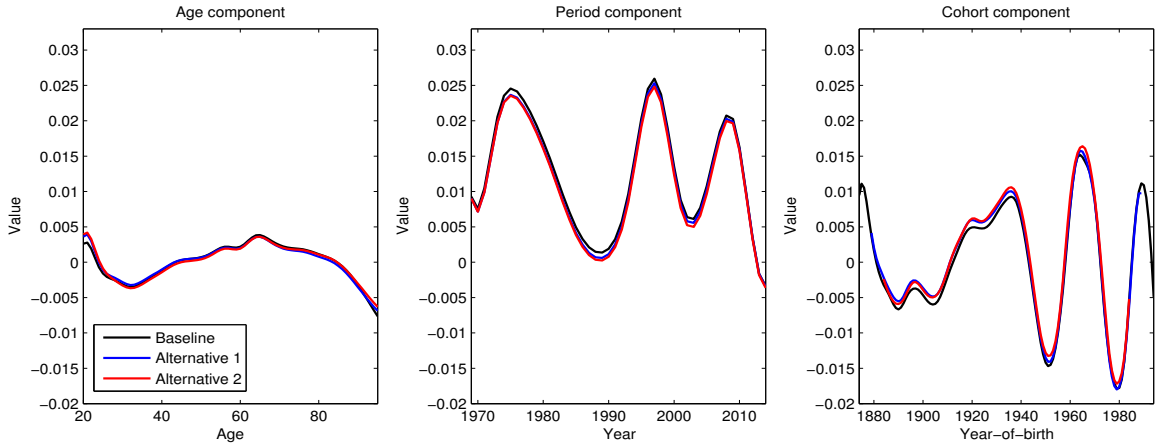
Alternatives 1 and 2 are illustrated in the following lexis diagrams.



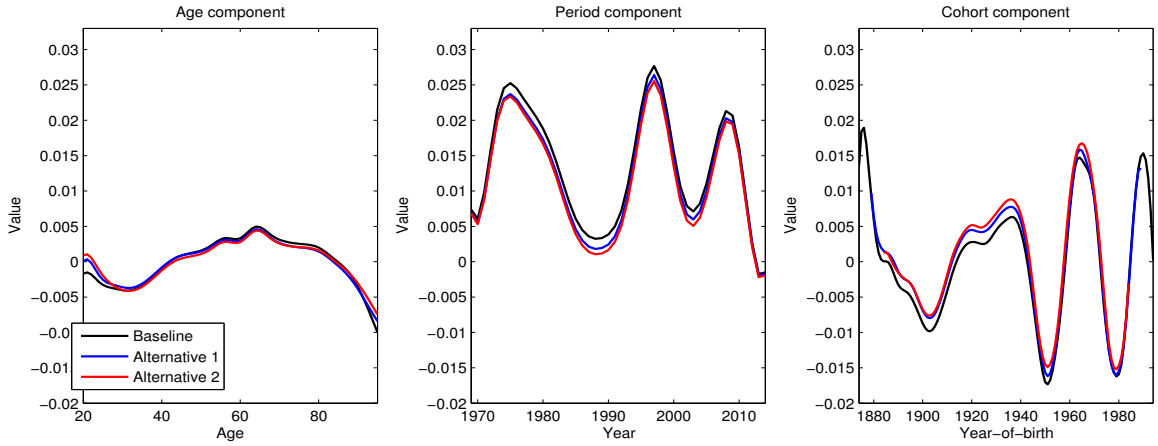
The resulting age/period/cohort components are shown in Figures 27 (males) and 28 (females). It is found that the exclusion of the oldest and the newest cohorts has only a modest impact on the estimation results.

5.5 Concluding Remark

The results of the robustness tests suggest that the CMI-09 approach is reasonably robust relative to small changes in the data set and model set-up. However, the problem of residual clustering remains significant and must be addressed. In the next section, we investigate if the problem can be mitigated by using a different APC structure.

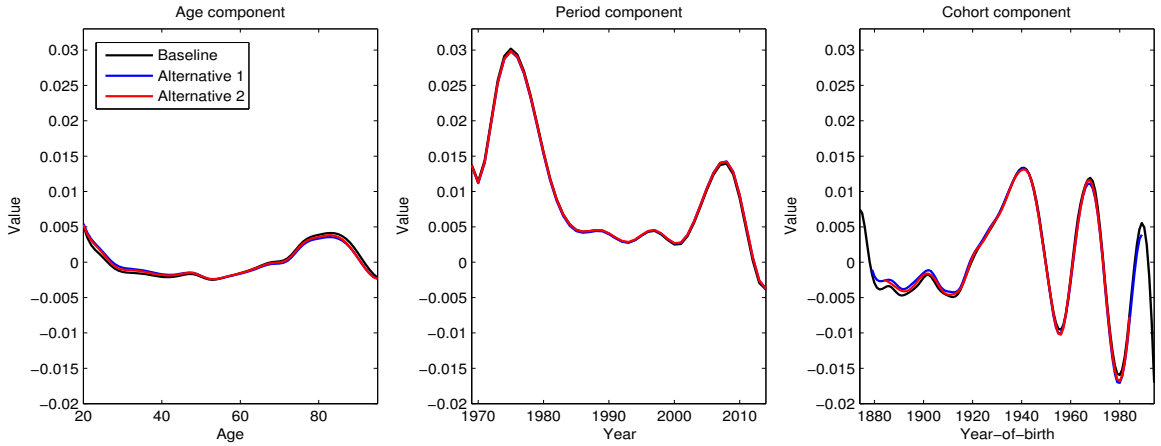


(a) U.S. Males: Dataset (i), HMD data

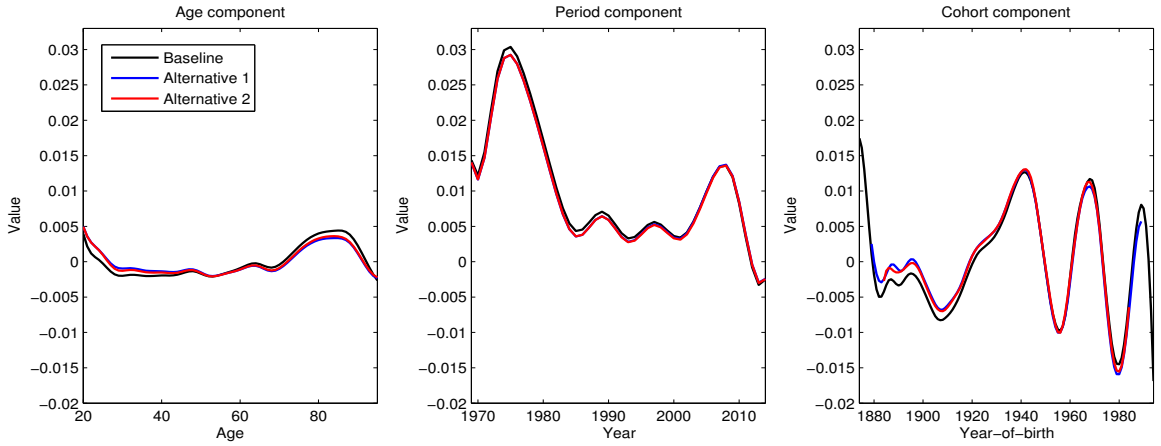


(b) U.S. Males: Dataset (ii), SSA data

Figure 27: The estimated Age/Period/Cohort components obtained from the CMI-09 method, with and without exclusion of the youngest/oldest cohorts, U.S. **males**.
- Baseline: All available data are used.
- Alternative 1: The oldest and youngest five cohorts in the data sample are excluded.
- Alternative 2: The oldest and youngest ten cohorts in the data sample are excluded.



(a) U.S. Females: Dataset (i), HMD data



(b) U.S. Females: Dataset (ii), SSA data

Figure 28: The estimated Age/Period/Cohort components obtained from the CMI-09 method, with and without exclusion of the youngest/oldest cohorts, U.S. **females**.
- Baseline: All available data are used.
- Alternative 1: The oldest and youngest five cohorts in the data sample are excluded.
- Alternative 2: The oldest and youngest ten cohorts in the data sample are excluded.

6 Improving the CMI-09 Method

In this section, we investigate if the problem of residuals clustering can be mitigated by using a different APC structure. Seven candidate model structures are considered. In what follows, we first define the candidate model structures. The definitions are followed by a collection of robustness tests which allow us to shortlist a smaller number of model structures for further consideration. Finally, we perform a residual analysis to identify the model structure that is best suited for U.S. historical mortality experience.

6.1 Definitions of the Candidate APC Model Structures

We consider seven candidate APC models, including M2, M3, M6, M7, M8, the full Plat model and the simplified Plat model. These models were originally developed to model mortality rates or probabilities ($m_{x,t}$'s or $q_{x,t}$'s), but in the context of this investigation, we use them to model the smoothed mortality improvement rates ($\tilde{Z}_{x,t}$'s). Note that M3 is identical to the APC model used in the CMI-09 approach.

In describing the model structures, the following conventions are used:

- $[x_0, x_1]$ is the sample age range;
- $[t_0, t_1]$ is 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 $t_1 - x_0$;
- n_c is the number of data points associated with year-of-birth c;
- $\bar{x} = (x_0 + x_1)/2$ is the mean age over the sample age range;
- $(\bar{x} - x)_+ = \max(\bar{x} - x, 0)$ is the maximum of $\bar{x} - x$ and 0;
- $\tilde{Z}_{x,t}$ is the smoothed mortality improvement rate for age x and year t;
  - $\beta_x^{(1)}$ 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.

The mathematical definitions of the seven candidate models are provided below. For all APC model structures, identifiability constraints are needed to stipulate parameter uniqueness. The identifiability constraints used for each candidate model are also given below.

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

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

  Identifiability constraints:
  $\sum_{x=x_0}^{x_1} \beta_x^{(1)} = 0$, $\sum_{x=x_0}^{x_1} \beta_x^{(2)} = 1$, $\sum_{x=x_0}^{x_1} \beta_x^{(3)} = 1$, and $\sum_{c=t_0-x_1}^{t_1-x_0} n_c\gamma_c = 0$.

- M3 – The Age-Period-Cohort Model (Osmond, 1985; used in the CMI-09 decomposition method)

$$\tilde{Z}_{x,t} = \beta_x^{(1)} + \kappa_t^{(2)} + \gamma_c + e_{x,t}$$

  Identifiability constraints: $\sum_{x=x_0}^{x_1} \beta_x^{(1)} = 0$, $\sum_{c=t_0-x_1}^{t_1-x_0} \gamma_c = 0$, and $\sum_{c=t_0-x_1}^{t_1-x_0} c\gamma_c = 0$.

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

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

  Identifiability constraints: $\sum_{c=t_0-x_1}^{t_1-x_0} \gamma_c = 0$ and $\sum_{c=t_0-x_1}^{t_1-x_0} c\gamma_c = 0$.

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

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

  Identifiability constraints: $\sum_{c=t_0-x_1}^{t_1-x_0} \gamma_c = 0$, $\sum_{c=t_0-x_1}^{t_1-x_0} c\gamma_c = 0$, and $\sum_{c=t_0-x_1}^{t_1-x_0} c^2\gamma_c = 0$

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

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

  Identifiability constraints: $\sum_{c=t_0-x_1}^{t_1-x_0} n_c\gamma_c = 0$.

- The Full Plat model (Plat, 2009):

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

  Identifiability constraints:
  $\sum_{x=x_0}^{x_1} \beta_x^{(1)} = 0$, $\sum_{t=t_0}^{t_1} \kappa_t^{(2)} = 0$, $\sum_{t=t_0}^{t_1} \kappa_t^{(3)} = 0$, $\sum_{c=t_0-x_1}^{t_1-x_0} \gamma_c = 0$, $\sum_{c=t_0-x_1}^{t_1-x_0} c\gamma_c = 0$, and $\sum_{c=t_0-x_1}^{t_1-x_0} c^2\gamma_c = 0$.

- The Simplified Plat model (Plat, 2009):

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

  Identifiability constraints:
  $\sum_{x=x_0}^{x_1} \beta_x^{(1)} = 0$, $\sum_{t=t_0}^{t_1} \kappa_t^{(2)} = 0$, $\sum_{c=t_0-x_1}^{t_1-x_0} \gamma_c = 0$, $\sum_{c=t_0-x_1}^{t_1-x_0} c\gamma_c = 0$, and $\sum_{c=t_0-x_1}^{t_1-x_0} c^2\gamma_c = 0$.

All of the models are estimated by minimizing the residual sum of squares; that is, we minimize the following objective function:

$$\sum_{x=x_0}^{x_1} \sum_{t=t_0}^{t_1} e_{x,t}^2, \tag{2}$$

subject to the applicable identifiability constraints.

6.2 Testing the Robustness of the Alternative Model Structures

6.2.1 Defining the Robustness Measure

We define the following quantitative measure of robustness:

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

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 M6 to changes in the calibration window. We consider three calibration windows, which have the same length but different starting years (1968, 1978, and 1988, respectively). The following table summarizes the calibration windows under consideration:

	Starting Age	**Ending Age**	**Starting Year**	**Ending Year**	**Length of Calibration Window**
Baseline	20	95	1968	2004	37 years
Alternative 1			1973	2009	37 years
Alternative 2			1978	2014	37 years

Recall that M6 has the following model structure:

$$\tilde{Z}_{x,t} = \kappa_t^{(1)} + \kappa_t^{(2)}(x - \bar{x}) + \gamma_c + e_{x,t}.$$

The following steps are taken to obtain the relevant robustness measure:

(1) 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; the denominator of equation (3) is calculated as $0.0860 - (-0.0530) = 0.1363$;

(2) fit M6 using the baseline setting and obtain estimates of the three model terms: $\kappa_t^{(1)}$, $\kappa_t^{(2)}(x - \bar{x})$, and γ_c;

(3) re-fit M6 using the alternative 1 setting and obtain new estimates of the three model terms: $\tilde{\kappa}_t^{(1)}$, $\tilde{\kappa}_t^{(2)}(x - \bar{x})$, and $\tilde{\gamma}_{t-x}$;

(4) compare the baseline and alternative 1 setting, calculate the maximum change in each of the three model terms:

$$\begin{aligned}
\max_{t=1974,\ldots,2004} |\kappa_t^{(1)} - \tilde{\kappa}_t^{(1)}| &= 0.0017 \\
\max_{x=20,\ldots,95;t=1974,\ldots,2004} |\kappa_t^{(2)}(x - \bar{x}) - \tilde{\kappa}_t^{(2)}(x - \bar{x})| &= 0.0036 \\
\max_{c=1879,\ldots,1984} |\gamma_c - \tilde{\gamma}_c| &= 0.0040
\end{aligned}$$

(5) re-fit M6 using the alternative 2 setting and obtain new estimates of the three model terms: $\hat{\kappa}_t^{(1)}$, $\hat{\kappa}_t^{(2)}(x-\bar{x})$, and $\hat{\gamma}_{t-x}$;

(6) compare the baseline and alternative 2 setting, calculate the maximum change in each of the three model terms:

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

(7) compare the alternative 1 and alternative 2 setting, calculate the maximum change in each of the three model terms:

$$\begin{aligned} \max_{t=1979,\ldots,2009} |\hat{\kappa}_t^{(1)} - \tilde{\kappa}_t^{(1)}| &= 0.0018 \\ \max_{x=20,\ldots,95;t=1979,\ldots,2009} |\hat{\kappa}_t^{(2)}(x-\bar{x}) - \tilde{\kappa}_t^{(2)}(x-\bar{x})| &= 0.0030 \\ \max_{c=1884,\ldots,1989} |\hat{\gamma}_c - \tilde{\gamma}_c| &= 0.0059 \end{aligned}$$

(8) calculate the 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 (3), 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 $> 20\%$

6.2.2 Changes in the Tolerance Value Used in Optimizing Model Parameters

To obtain parameter estimates, the objective function specified in equation (2) is minimized using 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 section, we robustness test the candidate models with respect to changes in the tolerance value. Three tolerance values are considered:

- Baseline: tolerance value = 10^{-8}.
- Alternative 1: tolerance value = 10^{-6}.
- Alternative 2: tolerance value = 10^{-10}.

The result of this robustness test is summarized in the following table:

Model	SSA male		SSA female	
	Robustness measure	Category	Robustness measure	Category
M2	19.8%	Medium	131.1%	Low
M3	0.0%	High	0.0%	High
M6	0.1%	High	0.1%	High
M7	0.2%	High	0.2%	High
M8	0.4%	High	21.1%	Medium
Full Plat	6.2%	High	10.5%	Medium
Simplified Plat	0.0%	High	0.0%	High

Figures 29, 30 and 31 illustrate high, medium and low levels of robustness to changes in the tolerance value, respectively. In each diagram, the black, blue and red lines represent parameter estimates obtained using different tolerance values. For the M3 fitted to SSA female data (high robustness), the three colored lines overlap one another. For the full Plat model fitted to SSA female data (medium robustness), the three colored lines are slightly different, and it appears that the fourth term $\kappa_t^{(3)}(\bar{x} - x)_+$ contributes the most to the robustness measure. Finally, for the M2 fitted to SSA female data (low robustness), the black and red lines overlap; however, for females, the blue line in each panel is significantly different from the other two lines. This indicates that reducing the tolerance value to 10^{-6} leads to a very different A/P/C decomposition.

For M2 and M8, the low robustness to changes in the tolerance value is due possibly to the convergence problem. The optimization for 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}.[6] 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.

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

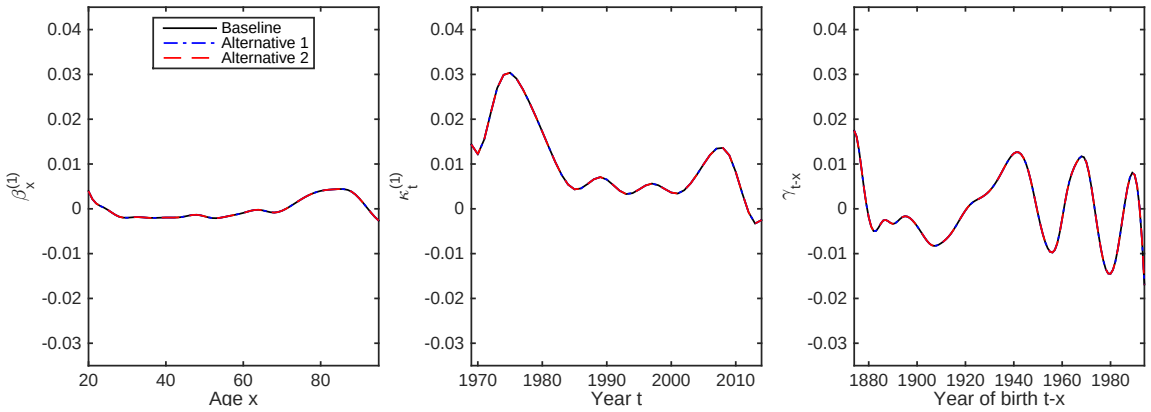


Figure 29: An example of **high robustness** to the tolerance value used: The estimated A/P/C components obtained from the Route A M3 model fitted to SSA female data.
- Baseline: tolerance value = 10^{-8}. Alternative 1: tolerance value = 10^{-6}. Alternative 2: tolerance value = 10^{-10}.

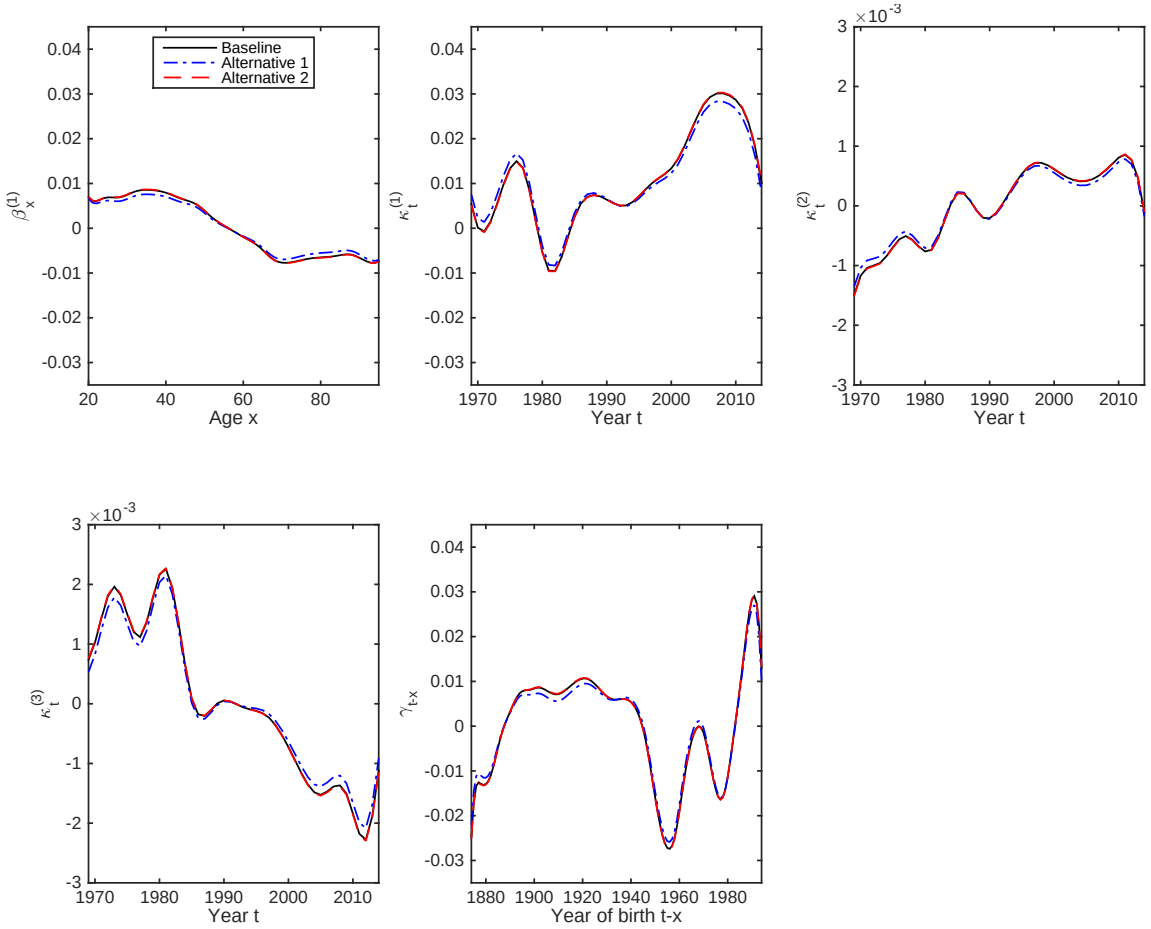


Figure 30: An example of **medium robustness** to the tolerance value used: The estimated A/P/C components obtained from the Route A full Plat model fitted to SSA female data.
- Baseline: tolerance value = 10^{-8}. Alternative 1: tolerance value = 10^{-6}. Alternative 2: tolerance value = 10^{-10}.

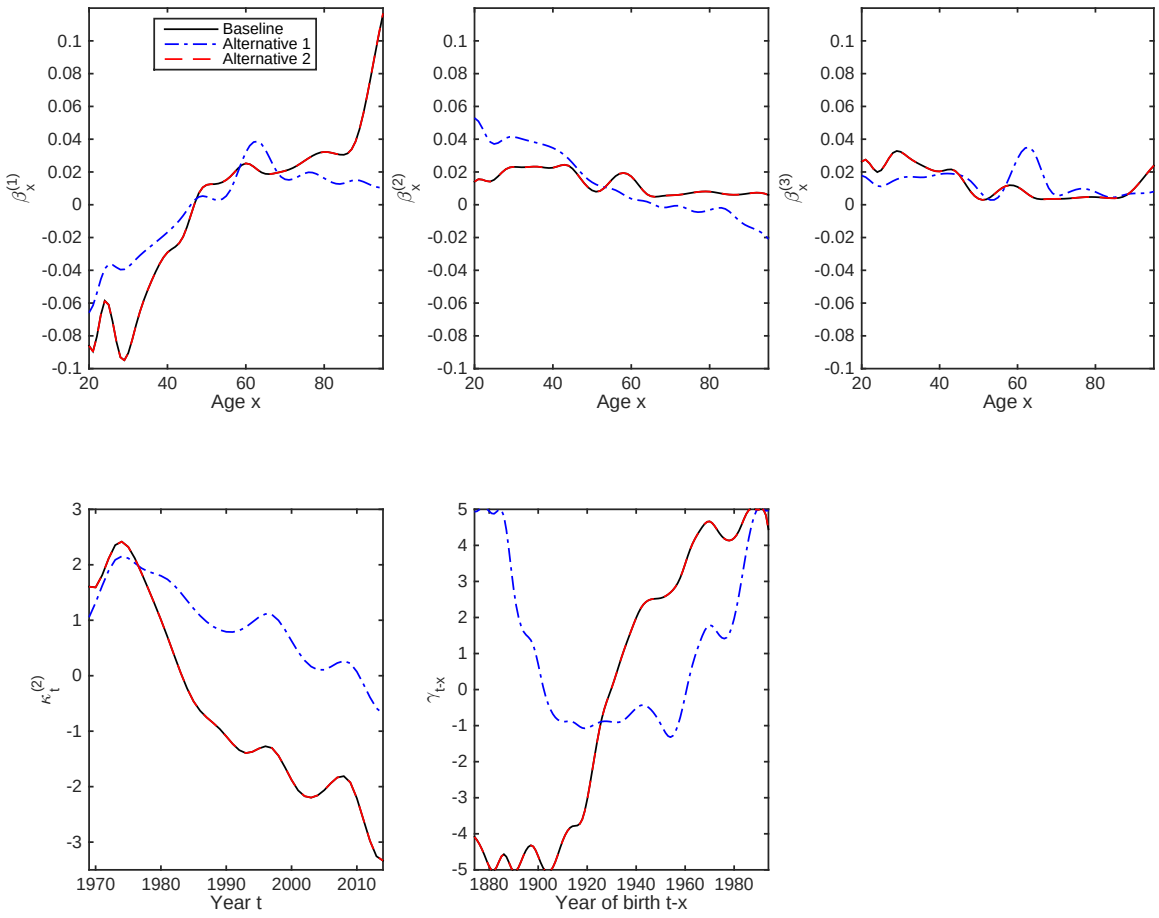


Figure 31: An example of **low robustness** to the tolerance value used: The estimated A/P/C components obtained from the Route A M2 model fitted to SSA female data.
- Baseline: tolerance value = 10^{-8}. Alternative 1: tolerance value = 10^{-6}. Alternative 2: tolerance value = 10^{-10}.

6.2.3 Changes in the Calibration Window

Here we test the robustness to changes in the calibration window. We consider three calibration windows, which have the same length but different starting years (1968, 1978, and 1988, respectively). This set-up mimics the situation when the models are updated every five years. The following table summarizes the calibration windows under consideration:

	Starting Age	**Ending Age**	**Starting Year**	**Ending Year**	**Length of Calibration Window**
Baseline	20	95	1968	2004	37 years
Alternative 1			1973	2009	37 years
Alternative 2			1978	2014	37 years

The result of this robustness test is summarized in the table below:

Model	SSA male		SSA female	
	Robustness measure	Category	Robustness measure	Category
M2	136.3%	Low	62.0%	Low
M3	11.1%	Medium	9.8%	High
M6	5.2%	High	9.2%	High
M7	24.9%	Low	12.9%	Medium
M8	5.8%	High	31.6%	Low
Full Plat	24.2%	Low	75.8%	Low
Simplified Plat	10.8%	Medium	9.4%	High

Figures 32, 33 and 34 illustrate high, medium, and low levels robustness to changes in the calibration window, respectively.

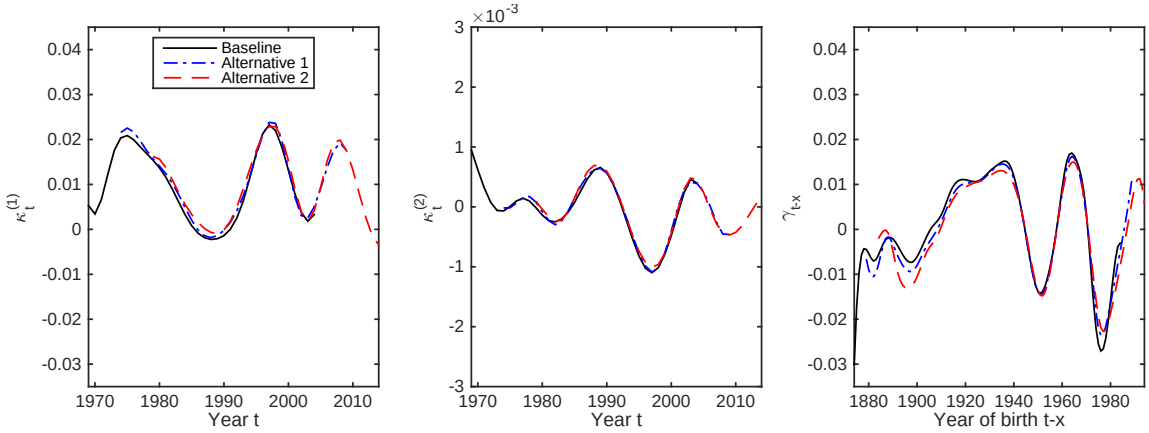


Figure 32: An example of **high robustness** to the calibration window used: The estimated A/P/C components obtained from the Route A M6 model fitted to SSA male data.
- Baseline: ages 20 to 95, years 1968 to 2004. Alternative 1: ages 20 to 95, years 1973 to 2009. Alternative 2: ages 20 to 95, years 1978 to 2014.

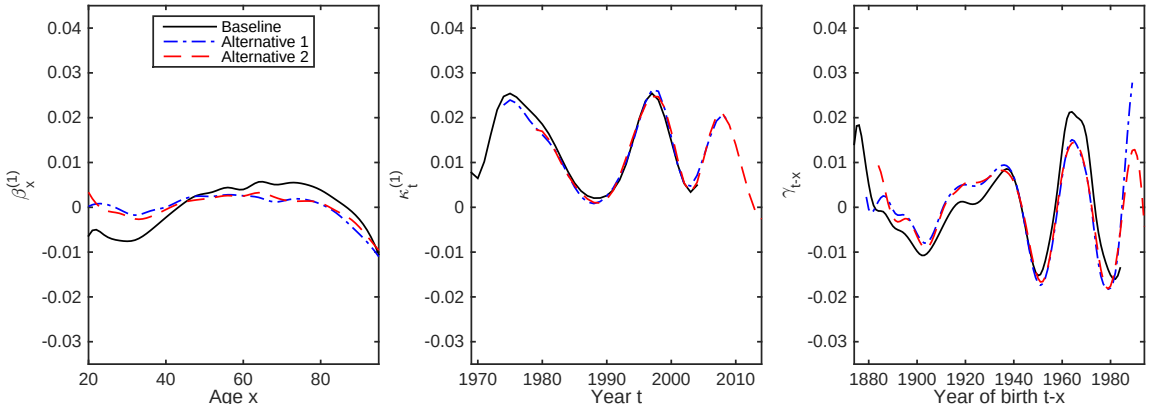


Figure 33: An example of **medium robustness** to the calibration window used: The estimated A/P/C components obtained from the Route A M3 model fitted to SSA male data.
- Baseline: ages 20 to 95, years 1968 to 2004. Alternative 1: ages 20 to 95, years 1973 to 2009. Alternative 2: ages 20 to 95, years 1978 to 2014.

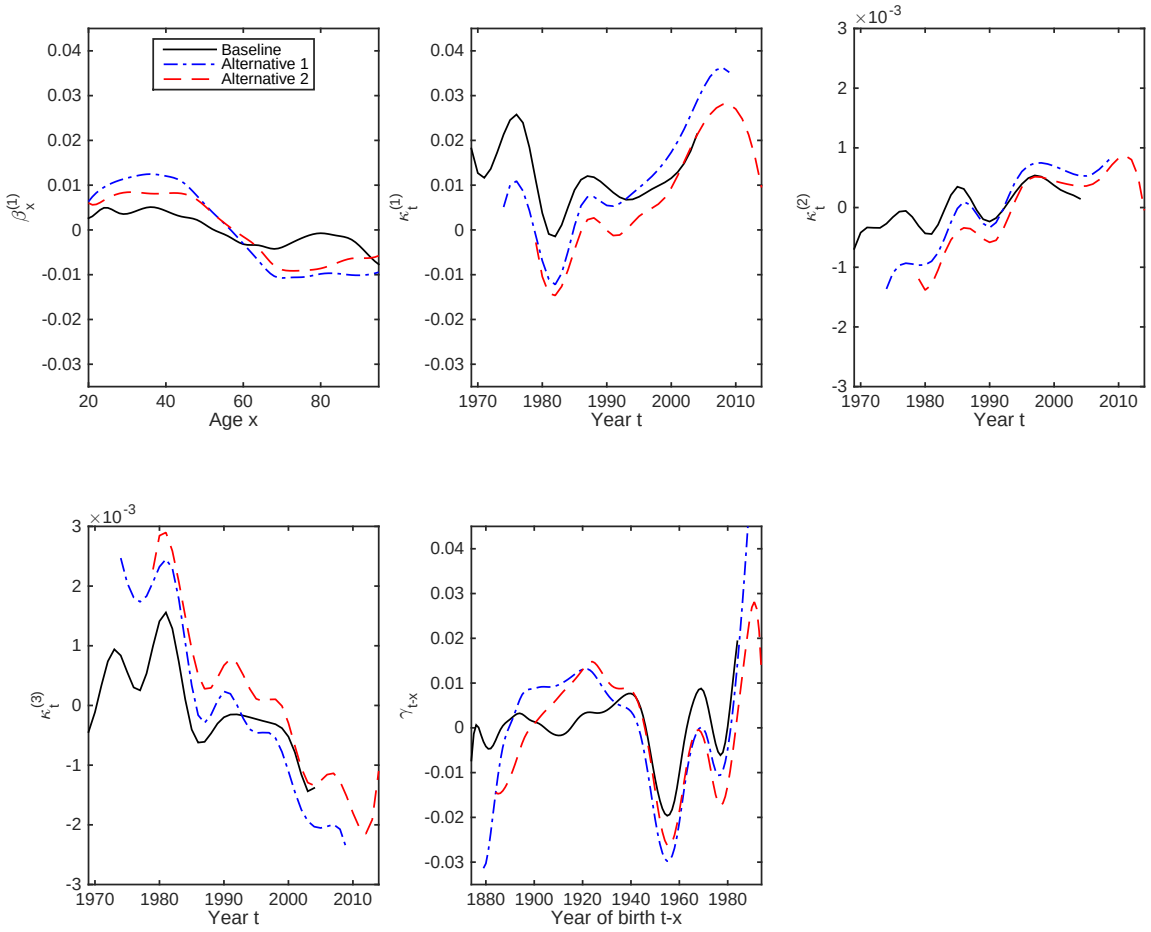


Figure 34: An example of **low robustness** to the calibration window used: The estimated A/P/C components obtained from the Route A full Plat model fitted to SSA female data.
- Baseline: ages 20 to 95, years 1968 to 2004. Alternative 1: ages 20 to 95, years 1973 to 2009. Alternative 2: ages 20 to 95, years 1978 to 2014.

6.2.4 Changes in the Age Range

Here we test the robustness to changes in age ranges. The following age ranges are considered:

	Starting Age	Ending Age	Starting Year	Ending Year	Number of Ages
Baseline	20	95	1968	2014	76
Alternative 1	30	85			56
Alternative 2	40	75			36

The result of this robustness test is summarized in the table below:

Model	SSA male		SSA female	
	Robustness measure	Category	Robustness measure	Category
M2	80.3%	Low	120.0%	Low
M3	11.1%	Medium	16.4%	Medium
M6	11.8%	Medium	13.9%	Medium
M7	13.0%	Medium	19.4%	Medium
M8	14.1%	Medium	75.4%	Low
Full Plat	27.4%	Low	60.3%	Low
Simplified Plat	11.1%	Medium	13.9%	Medium

Figures 35 and 36 illustrate medium and low levels of robustness to changes in the age ranges, respectively. Note that none of the seven candidate models is classified as 'high robustness' in this test.

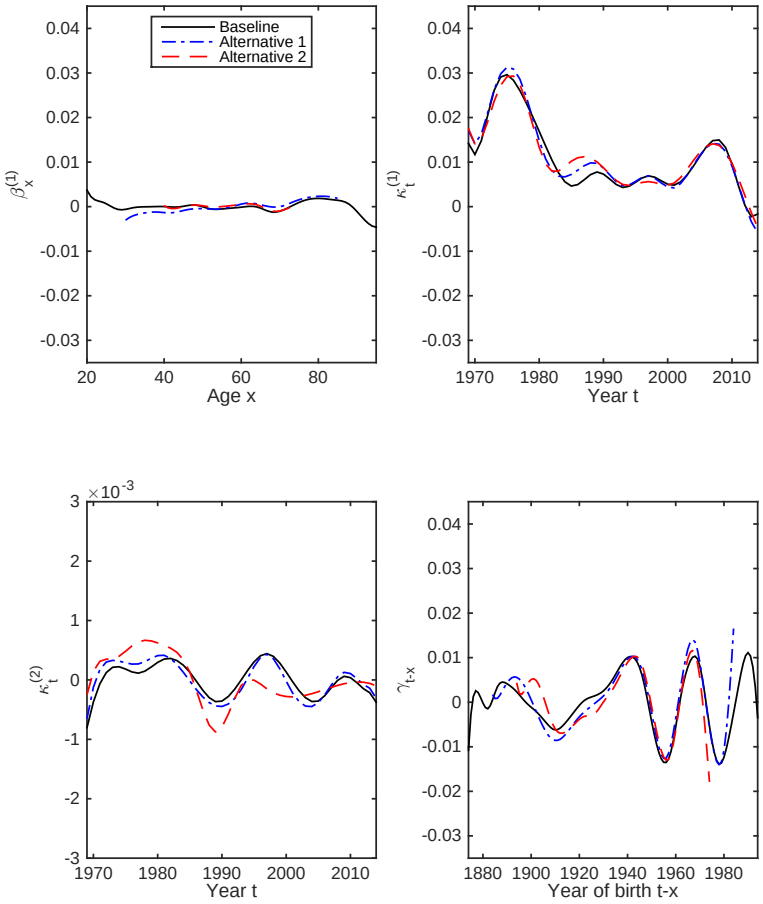


Figure 35: An example of **medium robustness** to the age range used: The estimated A/P/C components obtained from the Route A simplified Plat model fitted to SSA female data.
- Baseline: ages 20 to 95, years 1968 to 2014. Alternative 1: ages 30 to 85, years 1968 to 2014. Alternative 2: ages 40 to 75, years 1968 to 2014.

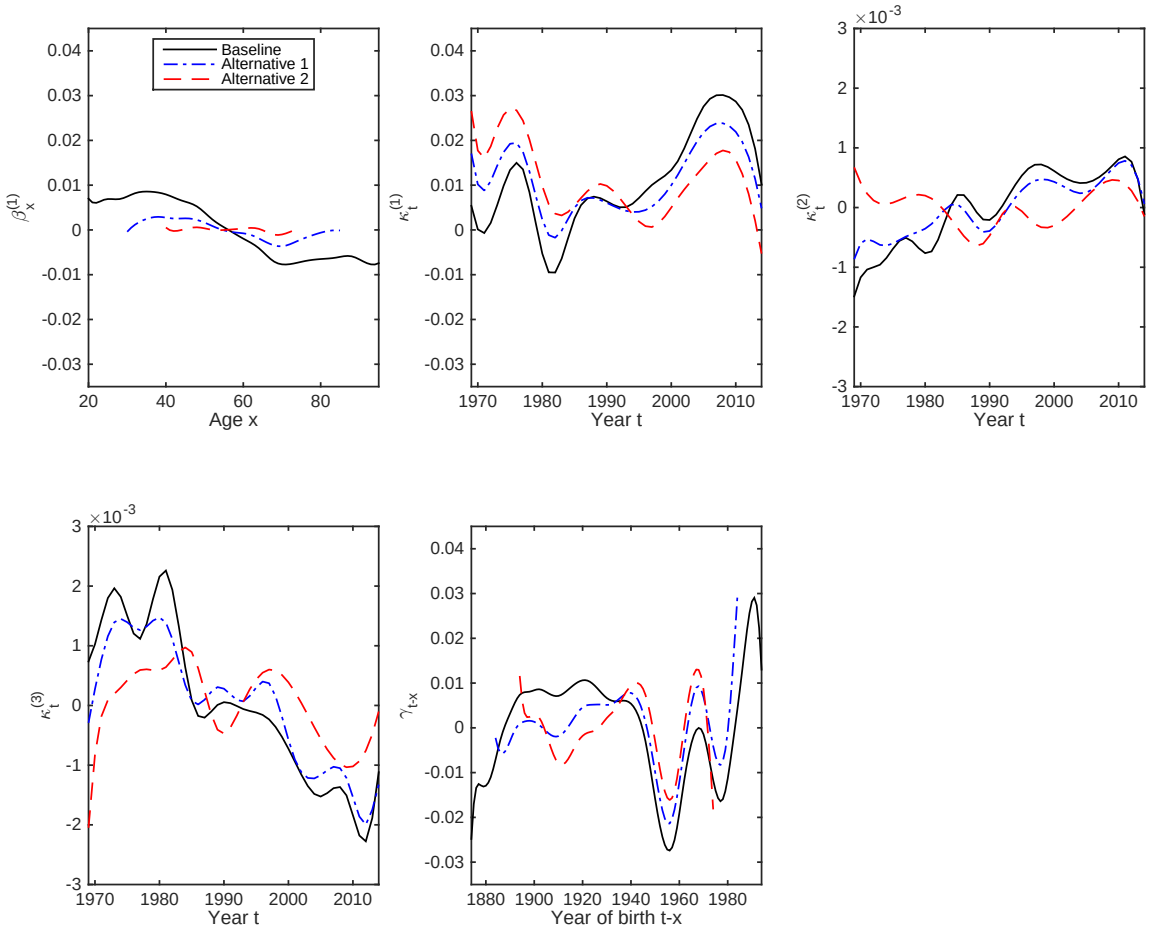


Figure 36: An example of **low robustness** to the age range used: The estimated A/P/C components obtained from the Route A full Plat model fitted to SSA female data.
- Baseline: ages 20 to 95, years 1968 to 2014. Alternative 1: ages 30 to 85, years 1968 to 2014. Alternative 2: ages 40 to 75, years 1968 to 2014.

6.2.5 Choice of Parameter Constraints

Here we examine the robustness to the choice of identifiability constraints. The baseline and alternative constraints for each candidate model are listed below.

- M2
  Model structure: $\tilde{Z}_{x,t} = \beta_x^{(1)} + \beta_x^{(2)}\kappa_t^{(2)} + \beta_x^{(3)}\gamma_{t-x} + e_{x,t}$
  Baseline constraints: $\sum_x \beta_x^{(1)} = 0$, $\sum_x \beta_x^{(2)} = 1$, $\sum_x \beta_x^{(3)} = 1$, $\sum_c n_c\gamma_c = 0$
  Alternative constraints: $\sum_x \beta_x^{(1)} = 0$, $\sum_x \beta_x^{(2)} = 1$, $\sum_x \beta_x^{(3)} = 1$, $\sum_c \gamma_c = 0$

- M3
  Model structure: $\tilde{Z}_{x,t} = \beta_x^{(1)} + \kappa_t^{(2)} + \gamma_{t-x} + e_{x,t}$
  Baseline constraints: $\sum_x \beta_x^{(1)} = 0$, $\sum_c \gamma_c = 0$, $\sum_c c\gamma_c = 0$
  Alternative constraints: $\sum_x \beta_x^{(1)} = 0$, $\sum_c n_c\gamma_c = 0$, $\sum_c n_c c\gamma_c = 0$

- M6
  Model structure: $\tilde{Z}_{x,t} = \kappa_t^{(1)} + \kappa_t^{(2)}(x - \bar{x}) + \gamma_{t-x} + e_{x,t}$
  Baseline constraints: $\sum_c \gamma_c = 0$, $\sum_c c\gamma_c = 0$
  Alternative constraints: $\sum_c n_c\gamma_c = 0$, $\sum_c n_c c\gamma_c = 0$

- M7
  Model structure: $\tilde{Z}_{x,t} = \kappa_t^{(1)} + \kappa_t^{(2)}(x - \bar{x}) + \kappa_t^{(3)}((x - \bar{x})^2 - \hat{\sigma}_x^2) + \gamma_{t-x} + e_{x,t}$
  Baseline constraints: $\sum_c \gamma_c = 0$, $\sum_c c\gamma_c = 0$, $\sum_c c^2\gamma_c = 0$
  Alternative constraints: $\sum_c n_c\gamma_c = 0$, $\sum_c n_c c\gamma_c = 0$, $\sum_c n_c c^2\gamma_c = 0$

- M8
  Model structure: $\tilde{Z}_{x,t} = \kappa_t^{(1)} + \kappa_t^{(2)}(x - \bar{x}) + \gamma_{t-x}(x_c - x) + e_{x,t}$
  Baseline constraints: $\sum_c n_c\gamma_c = 0$.
  Alternative constraints: $\sum_c \gamma_c = 0$

- The full Plat model
  Model structure: $\tilde{Z}_{x,t} = \beta_x^{(1)} + \kappa_t^{(1)} + \kappa_t^{(2)}(\bar{x} - x) + \kappa_t^{(3)}(\bar{x} - x)_+ + \gamma_{t-x} + e_{x,t}$
  Baseline constraints: $\sum_x \beta_x^{(1)} = 0$, $\sum_t \kappa_t^{(2)} = 0$, $\sum_t \kappa_t^{(3)} = 0$, $\sum_c \gamma_c = 0$, $\sum_c c\gamma_c = 0$, $\sum_c c^2\gamma_c = 0$.
  Alternative constraints: $\sum_x \beta_x^{(1)} = 0$, $\sum_t \kappa_t^{(2)} = 0$, $\sum_t \kappa_t^{(3)} = 0$, $\sum_c n_c\gamma_c = 0$, $\sum_c n_c c\gamma_c = 0$, $\sum_c n_c c^2\gamma_c = 0$

- The simplified Plat model
  Model structure: $\tilde{Z}_{x,t} = \beta_x^{(1)} + \kappa_t^{(1)} + \kappa_t^{(2)}(\bar{x} - x) + \gamma_{t-x} + e_{x,t}$
  Baseline constraints: $\sum_x \beta_x^{(1)} = 0$, $\sum_t \kappa_t^{(2)} = 0$, $\sum_c \gamma_c = 0$, $\sum_c c\gamma_c = 0$, $\sum_c c^2\gamma_c = 0$.
  Alternative constraints: $\sum_x \beta_x^{(1)} = 0$, $\sum_t \kappa_t^{(2)} = 0$, $\sum_c n_c\gamma_c = 0$, $\sum_c n_c c\gamma_c = 0$, $\sum_c n_c c^2\gamma_c = 0$

The difference between the baseline and alternative constraints lies in the inclusion/exclusion of n_c, which represents the number of data points 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 there are many other combinations of constraints that can be used to stipulate parameter uniqueness. When very different constraints are used, the resulting A/P/C components may be very different. Our goal here is to examine the impact of small changes in the constraints used only.

The result of this robustness test is summarized in the table below:

Model	SSA male		SSA female	
	Robustness measure	Category	Robustness measure	Category
M2	24.0%	Low	129.5%	Low
M3	1.6%	High	4.0%	High
M6	4.9%	High	1.3%	High
M7	11.9%	Medium	6.1%	High
M8	0.4%	High	12.3%	Medium
Full Plat	7.8%	High	13.6%	Medium
Simplified Plat	6.7%	High	3.0%	High

Figures 37, 38 and 39 illustrate high, medium and low levels of robustness to the choice of parameter constraints, respectively.

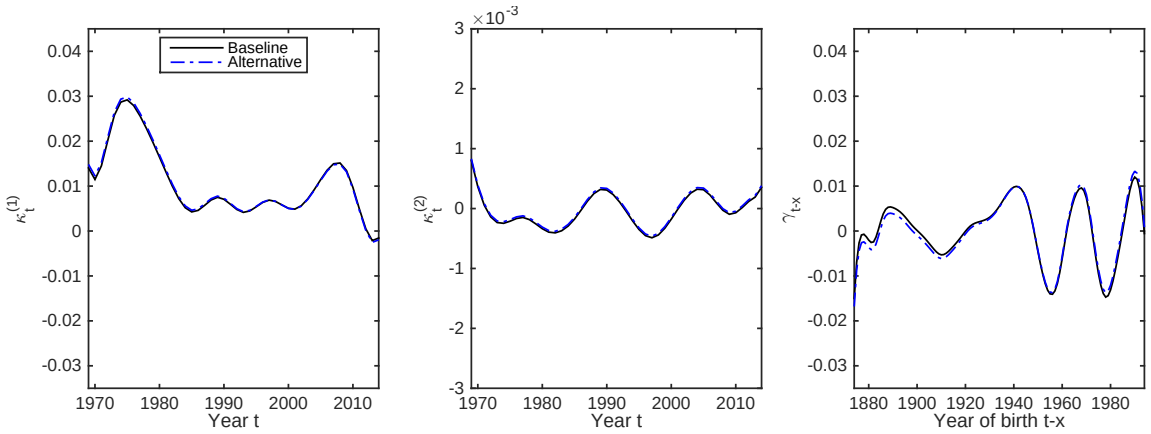


Figure 37: An example of **high robustness** to the parameter constraints used: The estimated A/P/C components obtained from the Route A M6 model fitted to SSA female data.
- Baseline: $\sum_c \gamma_c = 0$, $\sum_c c\gamma_c = 0$, $\sum_c c^2\gamma_c = 0$
- Alternative 1: $\sum_c n_c\gamma_c = 0$, $\sum_c n_c c\gamma_c = 0$, $\sum_c n_c c^2\gamma_c = 0$

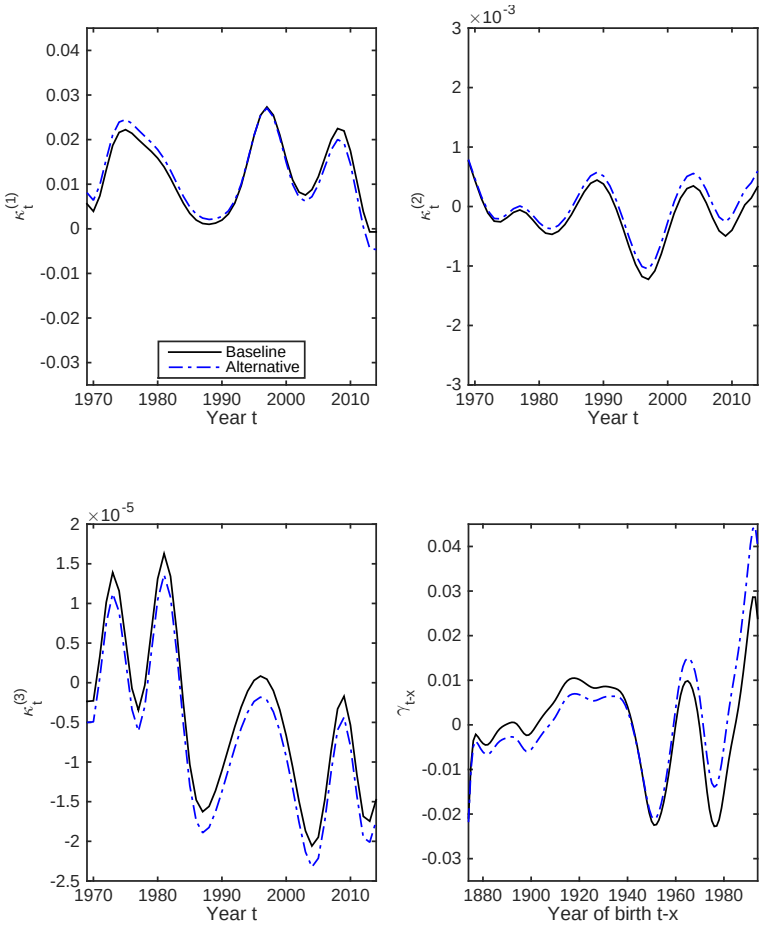


Figure 38: An example of **medium robustness** to the parameter constraints used: The estimated A/P/C components obtained from the Route A M7 model fitted to SSA male data.
- Baseline: $\sum_c \gamma_c = 0$, $\sum_c c\gamma_c = 0$, $\sum_c c^2\gamma_c = 0$
- Alternative 1: $\sum_c n_c\gamma_c = 0$, $\sum_c n_c c\gamma_c = 0$, $\sum_c n_c c^2\gamma_c = 0$

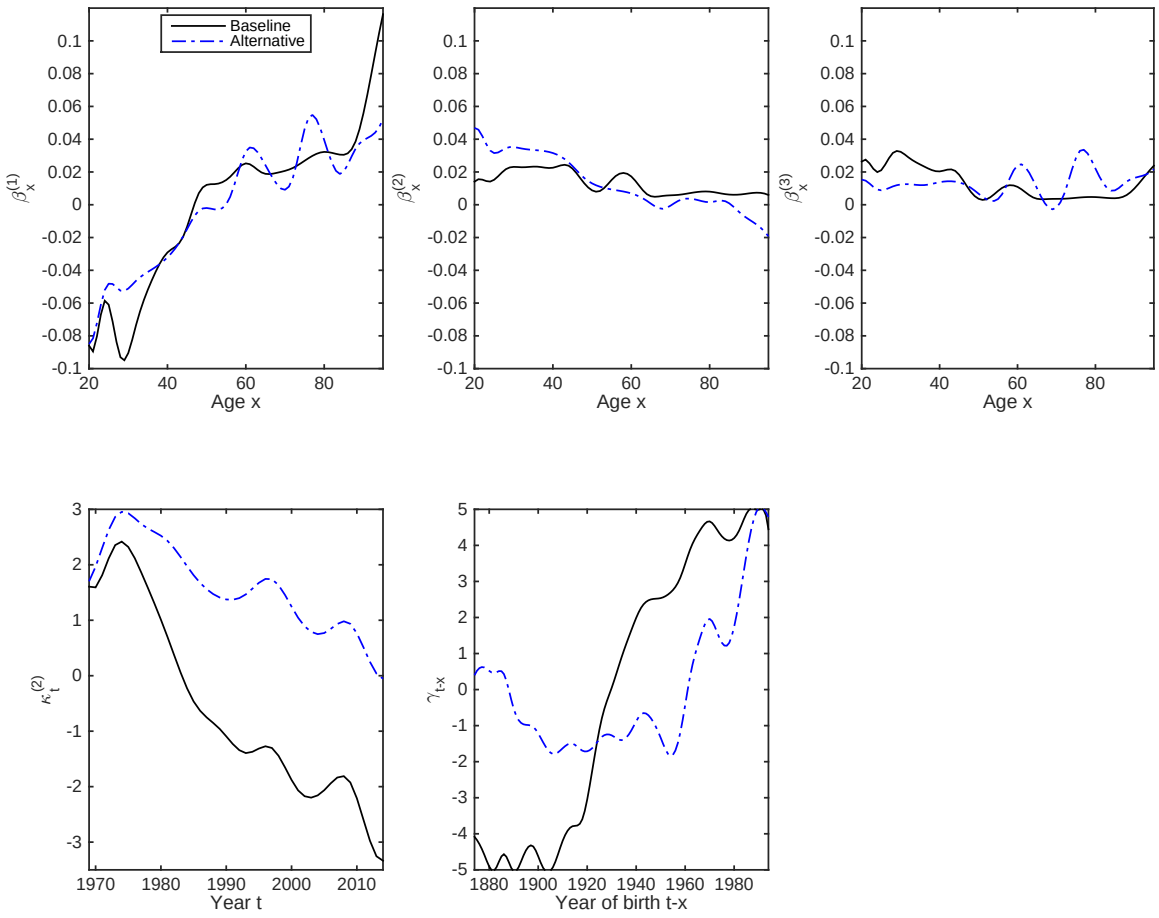


Figure 39: An example of **low robustness** to the parameter constraints used: The estimated A/P/C components obtained from the Route A M2 model fitted to SSA female data.
- Baseline: $\sum_c \gamma_c = 0$, $\sum_c c\gamma_c = 0$, $\sum_c c^2\gamma_c = 0$
- Alternative 1: $\sum_c n_c\gamma_c = 0$, $\sum_c n_c c\gamma_c = 0$, $\sum_c n_c c^2\gamma_c = 0$

6.2.6 Exclusion of the oldest/newest cohorts

Here we examine the robustness of the A/P/C components of each candidate model to the inclusion/exclusion of the oldest and newest cohorts. The following three situations are considered:

- Baseline: All available data are used.
- Alternative 1: The oldest and youngest five cohorts in the data sample are excluded.
- Alternative 2: The oldest and youngest ten cohorts in the data sample are excluded.

The result of this robustness test is summarized in the table below:

Model	SSA male		SSA female	
	Robustness measure	Category	Robustness measure	Category
M2	38.7%	Low	144.8%	Low
M3	4.1%	High	4.1%	High
M6	4.7%	High	6.1%	High
M7	12.1%	Medium	8.7%	High
M8	2.7%	High	15.4%	Medium
Full Plat	8.3%	High	15.3%	Medium
Simplified Plat	7.9%	High	7.1%	High

Figures 40, 41 and 42 demonstrate high, medium and low levels of robustness to the inclusion/exclusion of the youngest/oldest cohorts, respectively.

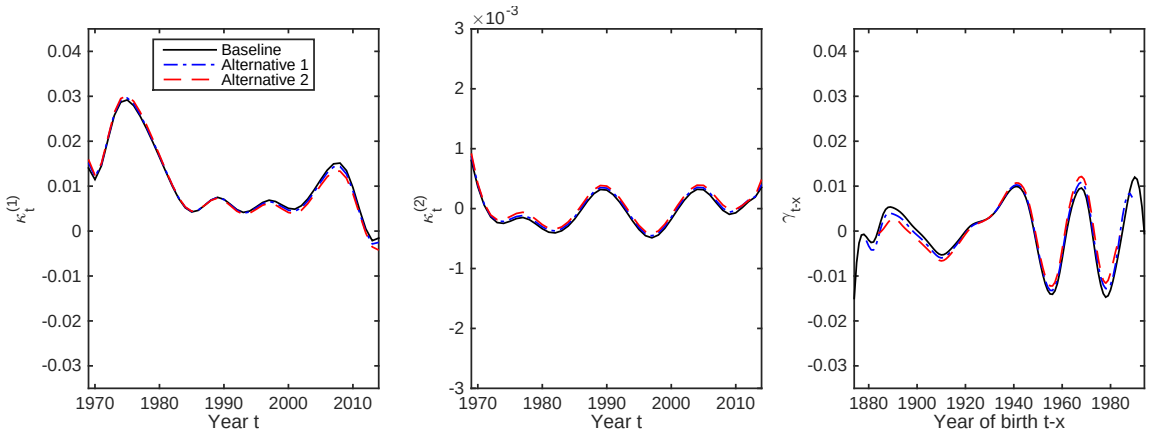


Figure 40: An example of **high robustness** to the inclusion/exclusion of the youngest/oldest cohorts: The estimated A/P/C components obtained from the Route A M6 model fitted to SSA female data.
- Baseline: All available data are used.
- Alternative 1: The oldest and youngest five cohorts in the data sample are excluded.
- Alternative 2: The oldest and youngest ten cohorts in the data sample are excluded.

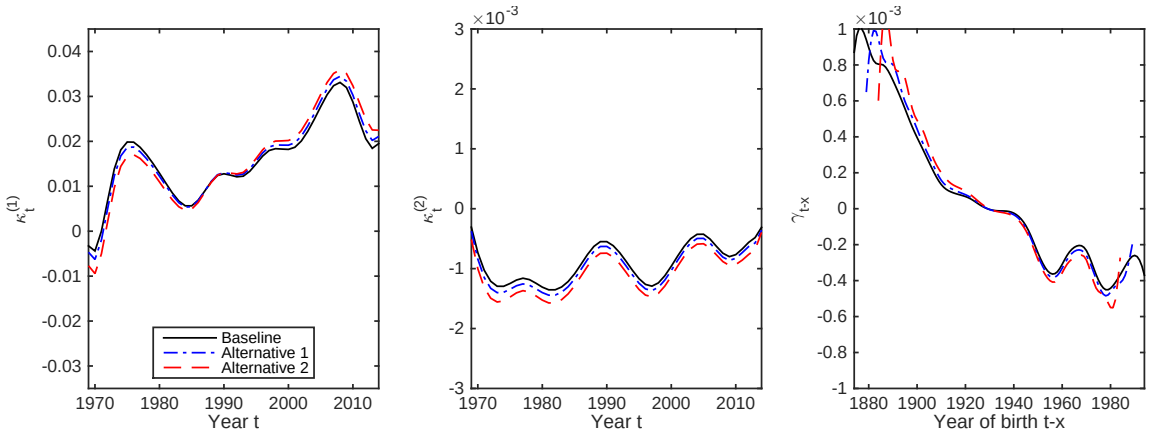


Figure 41: An example of **medium robustness** to the inclusion/exclusion of the youngest/oldest cohorts: The estimated A/P/C components obtained from the Route A M8 model fitted to SSA female data.
- Baseline: All available data are used.
- Alternative 1: The oldest and youngest five cohorts in the data sample are excluded.
- Alternative 2: The oldest and youngest ten cohorts in the data sample are excluded.

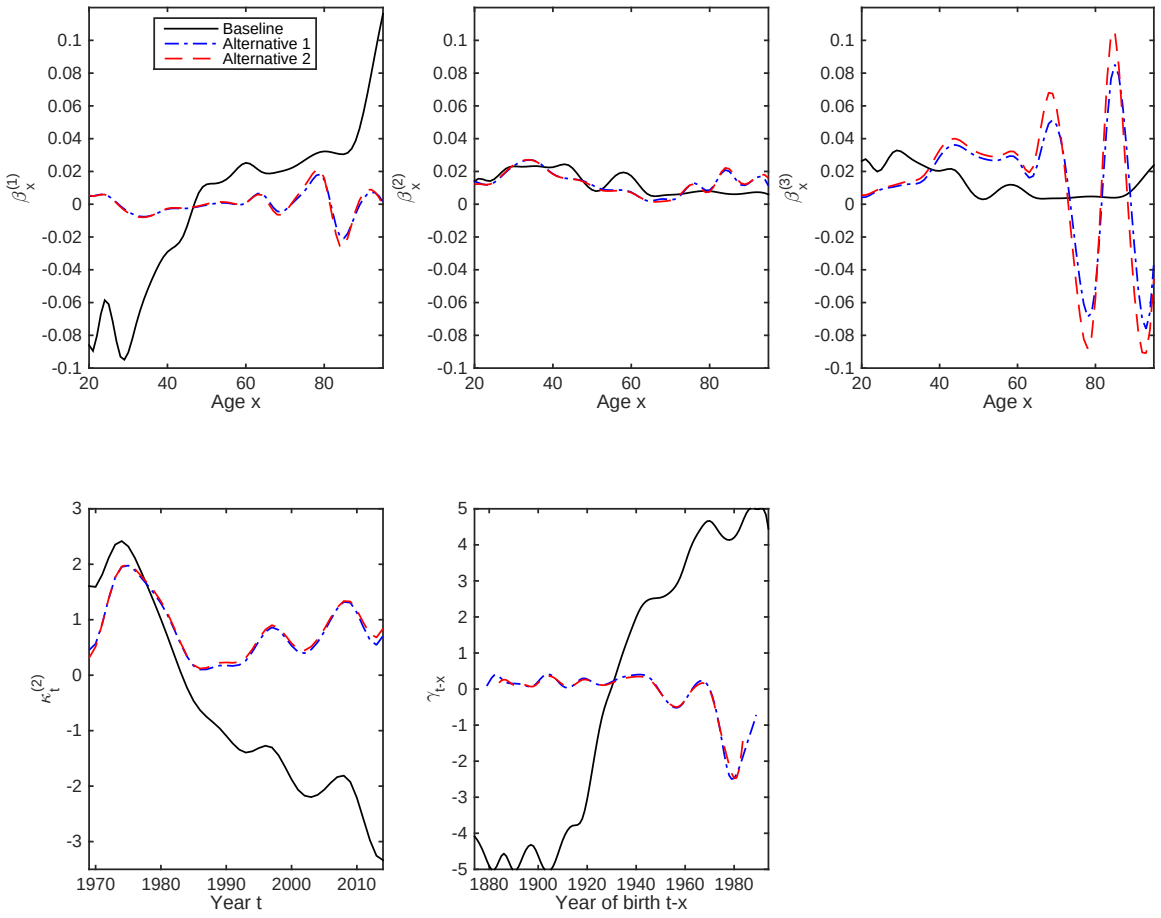


Figure 42: An example of **low robustness** to the inclusion/exclusion of the youngest/oldest cohorts: The estimated A/P/C components obtained from the Route A M2 model fitted to SSA female data.
- Baseline: All available data are used.
- Alternative 1: The oldest and youngest five cohorts in the data sample are excluded.
- Alternative 2: The oldest and youngest ten cohorts in the data sample are excluded.

6.3 Analyzing the Standardized Residuals Produced by the Shortlisted Models

The table below summarizes the results of all robustness tests we performed.

Robustness Test	M2	M3	M6	M7	M8	Full Plat	Simplified Plat
SSA males							
Tolerance value	Medium	High	High	High	High	High	High
Calibration window	Low	Medium	High	Low	High	Low	Medium
Age range	Low	Medium	Medium	Medium	Medium	Low	Medium
Parameter constraints	Low	High	High	Medium	High	High	High
Exclusion of cohorts	Low	High	High	Medium	High	High	High
SSA females							
Tolerance value	Low	High	High	High	Medium	Medium	High
Calibration window	Low	High	High	Medium	Low	Low	High
Age range	Low	Medium	Medium	Medium	Low	Low	Medium
Parameter constraints	Low	High	High	High	Medium	Medium	High
Exclusion of cohorts	Low	High	High	High	Medium	Medium	High

M2, M7, M8 and the full Plat model show low robustness in some tests. These models are therefore not given further consideration.

We now analyze the standardized residuals produced by M3, M6 and the simplified Plat model, which show high to medium levels robustness in tests performed in the previous section.

First, we perform the Anderson-Darling test, of which the null hypothesis is that standardized residuals come from a standard normal distribution. The resulting p-values are presented in Table 1. 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. The following conclusions are drawn:

1. The standardized residuals from M6 for SSA males and females do not pass the test.
2. The p-values indicate that the standardized residuals from the simplified Plat model are more 'normal' than those from M3 (the model used in the CMI-09 decomposition method).

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 are provided in Table 1. They also suggest that the simplified Plat model is more in line with the standard normal distribution, compared to M3 (the model used in the CMI-09 decomposition method).

Finally, we consider the Q-Q plots of the standardized residuals from M3, M6 and the simplified Plat models. A Q-Q plot is a plot of the quantiles of two distributions against each other. Here, we

	Dataset	M3	M6	Simplified Plat	M3 with an age break at 55
p-value	SSA males	0.0890	0.0069	0.3886	0.6873
	SSA females	0.0680	0.0258	0.4436	0.1136
Mean	SSA males	−0.0008	0.0282	−0.0010	−0.0021
	SSA females	-0.0013	0.0293	−0.0005	0.0003
Standard deviation	SSA males	0.9896	1.0070	0.9902	0.9913
	SSA females	0.9898	1.0256	0.9904	0.9905
Skewness	SSA males	−0.0094	−0.0202	−0.0519	0.1116
	SSA females	0.0182	−0.0275	−0.0337	0.0759
Kurtosis	SSA males	2.8572	2.7054	2.8789	3.2062
	SSA females	3.4885	2.9974	3.1917	3.6635

Table 1: The normality test p-value, mean, standard deviation, skewness, kurtosis of the standardized residuals: M3, M6, the simplified Plat model, and M3 with an age break point at 55.

compare the empirical distribution of the standardized residuals against the standard normal distribution. If the standardized residuals follow the standard normal distribution, the points in the Q-Q plot should fall on the 45° line.

The left panels of Figure 43, 44, and 45 show the full picture of the Q-Q plots, while the middle and right panels show the left and right tails of the Q-Q plots. The Q-Q plots indicate that the distribution of the standardized residuals from M3 (the model used in the CMI-09 decomposition method) is quite different from the standard normal distribution. The simplified Plat model performs the best for SSA dataset.

Summing up, M3 and the simplified Plat model are the only two models that pass the Anderson-Darling normality test, but M3 does not perform well in terms of the descriptive statistics and Q-Q plots. We remark here that M3 does not perform as well as the simplified Plat model even if it contains an age breakpoint at age 55 (see the last column of Table 1).

6.4 Concluding Remark

Before we conclude our work for Route A, let us revisit the heatmaps of standardized residuals. Figure 46 compares the heatmaps of standardized residuals obtained from M3 (the model used in the CMI-09 decomposition method), M3 with an age break at 55 and the simplified Plat model.

When M3 is used, large vertical clusters are found in the residual heat maps for U.S. males. This problem is mitigated when an age breakpoint at 55 is introduced to M3. More importantly, it can also be mitigated by using the simplified Plat model even if no age breakpoint is used. This outcome is possibly because the simplified Plat model can pick up the interaction between age and period effect, which cannot be captured by M3.

Overall, we conclude that when Route A is pursued, the simplified Plat model is the best suited for modeling the smoothed mortality improvement rates for the U.S. population.

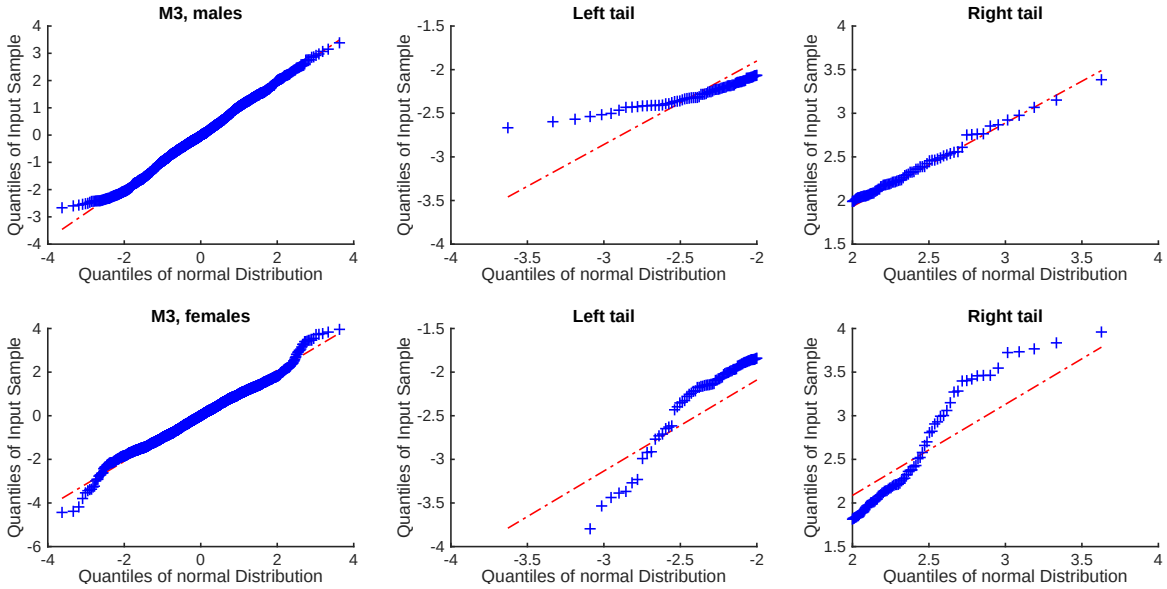


Figure 43: The Q-Q plots of standardized residuals obtained from the Route A M3 model fitted to the SSA dataset.

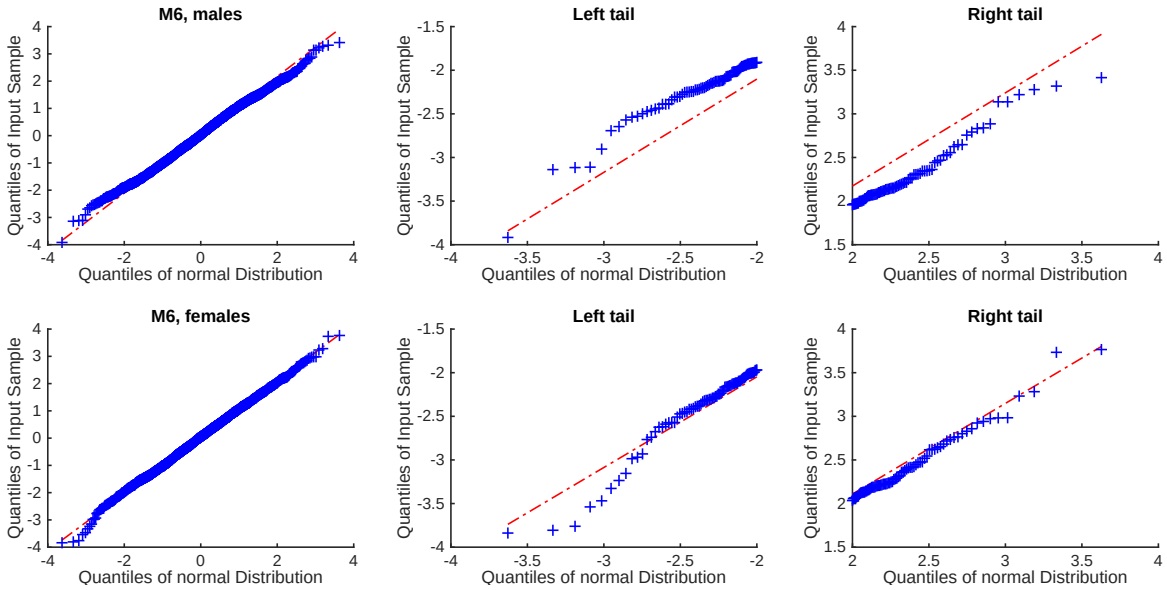


Figure 44: The Q-Q plots of standardized residuals obtained from the Route A M6 model fitted to the SSA dataset.

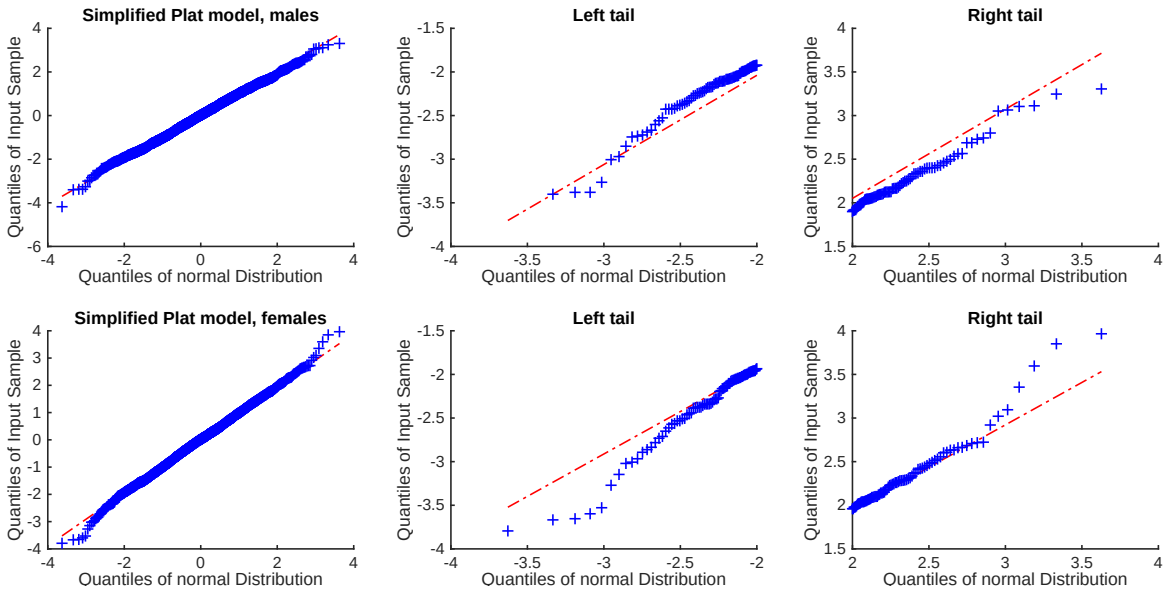


Figure 45: The Q-Q plots of standardized residuals obtained from the Route A simplified Plat model fitted to the SSA dataset.

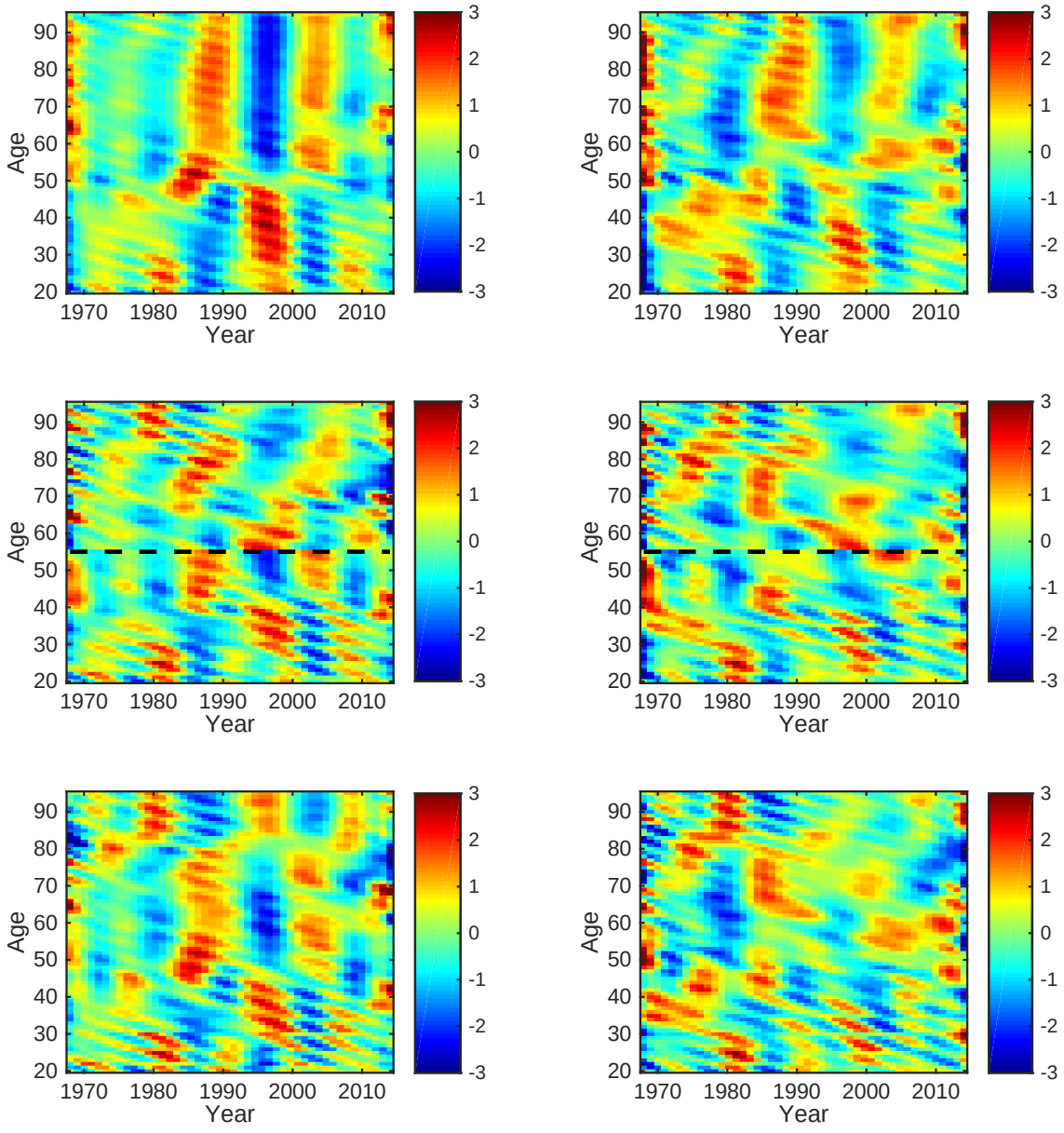


Figure 46: Heatmaps of the standardized residuals obtained from three Route A models: M3 (upper panels), M3 with an age breakpoint at 55 (middle panels) and the simplified Plat model (lower panels)
- Left panels: SSA males. Right panels: SSA females.

7 Repeating the Analyses Using Data for Ages 55 to 95 Only

For the readers' information, in this section we repeat the analyses using the data for ages 55 to 95 only. The table below presents the results of the robustness tests for each of the Route A models. The baseline and alternative settings in these robustness tests are the same as before, except that when testing robustness changes in age ranges we consider ages 55-95 (baseline), ages 60-90 (alternative 1) and ages 65-85 (alternative 2).

Robustness Test	M2	M3	M6	M7	M8	Full Plat	Simplified Plat
SSA males							
Tolerance value	Medium	High	High	High	Low	Medium	High
Calibration window	Low	Medium	Low	High	Low	Low	Medium
Age range	Low	Low	Medium	Medium	Medium	Low	Medium
Parameter constraints	High	Medium	Medium	High	Medium	High	Medium
Exclusion of cohorts	Low	Medium	Medium	Medium	Low	Low	Medium
SSA females							
Tolerance value	High	High	High	High	High	High	High
Calibration window	Low	Medium	Low	Low	Medium	Low	Low
Age range	Low	Low	Medium	Low	Low	Low	Medium
Parameter constraints	High	Medium	High	Low	High	Low	Medium
Exclusion of cohorts	Low	Medium	Medium	Low	Medium	Low	Medium

We observe that all of the candidate models have at least one low robustness rating. The simplified Plat model performs the best in the sense that it has only one low robustness rating. M3 and M6 are the second best performers. We therefore shortlist these three models for further consideration.

Figure 47 compares the heatmaps of the standardized residuals obtained from the three shortlisted models for ages 55 to 95. Residual clustering is observed in all heatmaps. However, the residual clusters in the heatmaps for the simplified Plat model seem to be the smallest.

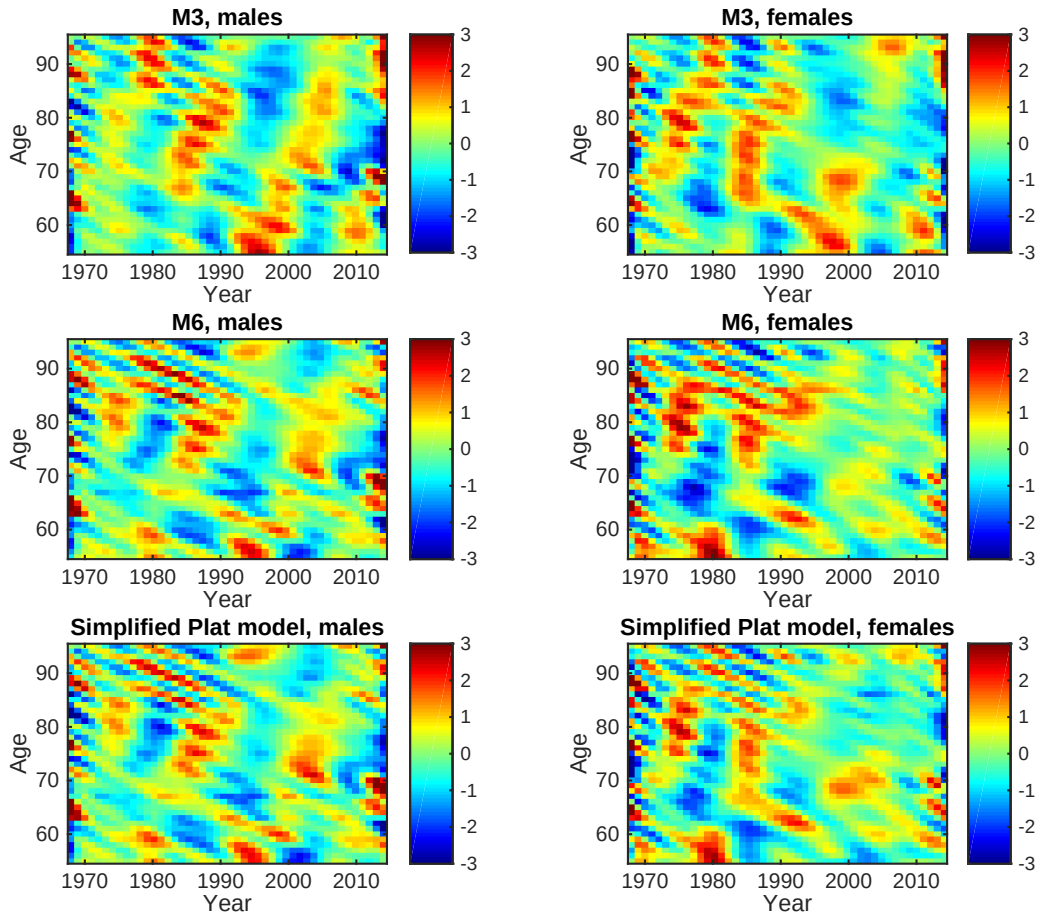


Figure 47: Heatmaps of the standardized residuals produced by the shortlisted Route A models for ages 55 to 95.

8 Modifying the Smoothing Method

8.1 Motivation

The CMI-09 decomposition method involves two stages: (1) smoothing, and (2) estimation of an APC model. A more rigorous approach is to integrate smoothing and estimation into one single process. This may be achieved by the roughness penalty approach (Delwarde et al., 2007). Note that the new CMI decomposition approach (proposed in the CMI Working Paper 90 released on June 22, 2016) is also based on an integrated process of smoothing and estimation.

8.2 The Roughness Penalty Approach

We now examine how the age, period and cohort components may be different if the single-stage roughness penalty approach is used. In the roughness penalty approach, a penalty term is applied to each component of the APC model for the crude mortality improvement rates. When applied to the APC model used in the CMI-09 method (equation (1)), the age, period and cohort components are obtained by minimizing the following penalized sum of squared errors:

$$\sum_{x=20}^{95}\sum_{t=1968}^{2014}(Z_{x,t}-a_x-k_t-g_c)^2+\vec{a}'P_a\vec{a}+\vec{k}'P_k\vec{k}+\vec{g}'P_g\vec{g}$$

where $\vec{a}$, $\vec{k}$ and $\vec{g}$ are the vectors of a_x's, k_t's and g_c's,

$$P_a=\pi_a\Delta_a'\Delta_a,\quad \text{with}\quad \Delta_a=\begin{pmatrix}1&-2&1&0&\cdots&0\\0&1&-2&1&\cdots&0\\\vdots&\ddots&\ddots&\ddots&\ddots&\vdots\\0&\cdots&1&-2&1&0\\0&\cdots&0&1&-2&1\end{pmatrix}_{74\times 76}$$

is the roughness penalty for the age dimension,

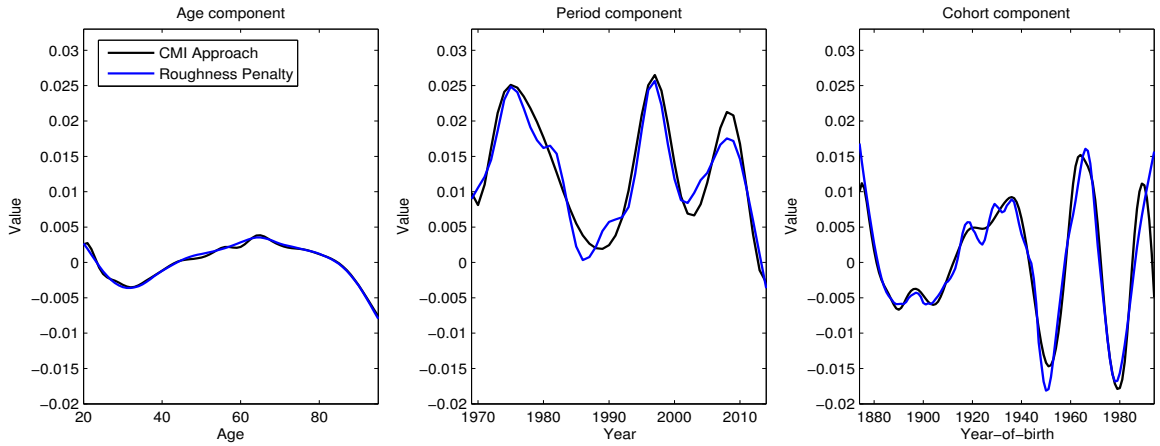
$$P_k=\pi_k\Delta_k'\Delta_k,\quad \text{and}\quad \Delta_k=\begin{pmatrix}1&-2&1&0&\cdots&0\\0&1&-2&1&\cdots&0\\\vdots&\ddots&\ddots&\ddots&\ddots&\vdots\\0&\cdots&1&-2&1&0\\0&\cdots&0&1&-2&1\end{pmatrix}_{45\times 47}$$

is the roughness penalty for the time dimension, and

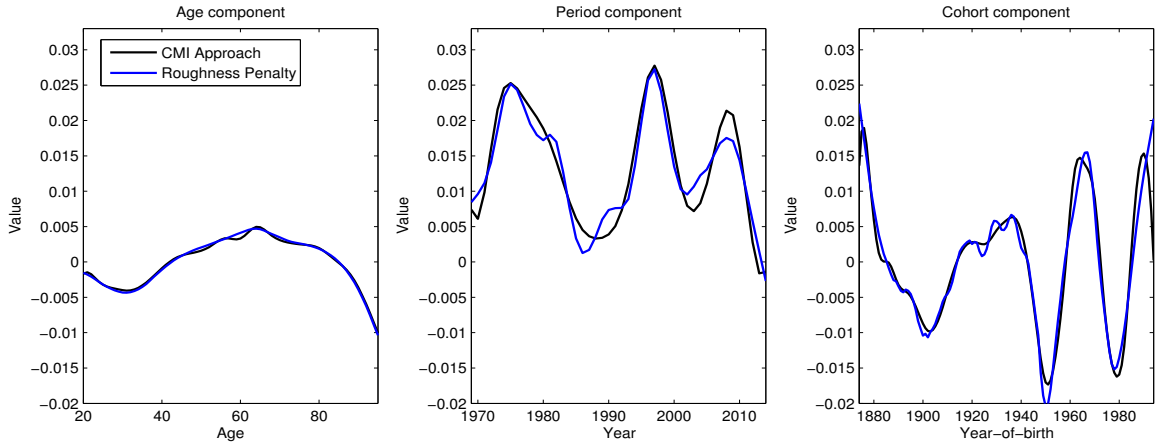
$$P_g=\pi_g\Delta_g'\Delta_g,\quad \text{and}\quad \Delta_g=\begin{pmatrix}1&-2&1&0&\cdots&0\\0&1&-2&1&\cdots&0\\\vdots&\ddots&\ddots&\ddots&\ddots&\vdots\\0&\cdots&1&-2&1&0\\0&\cdots&0&1&-2&1\end{pmatrix}_{118\times 120}$$

is the roughness penalty for cohort dimension. In the above, π_a, π_k and π_g are the smoothing parameters for a_x, k_t and g_c, respectively. The specification of Δ indicates that the degree of roughness is measured by the sum of squared second-order difference (e.g., $a_{x+2} - 2a_{x+1} + a_x$). The optimal smoothing parameters are selected by a cross validation as described by Delwarde et al. (2007). Further details concerning the smoothing procedure are provided in the Technical Appendix.

Figures 48 (males) and 49 (females) compare the estimated A/P/C components when different smoothing methods are used. The two smoothing methods yield similar A/P/C components. However, the positions of the troughs of the period component are changed if we replace the original method with the roughness penalty approach.

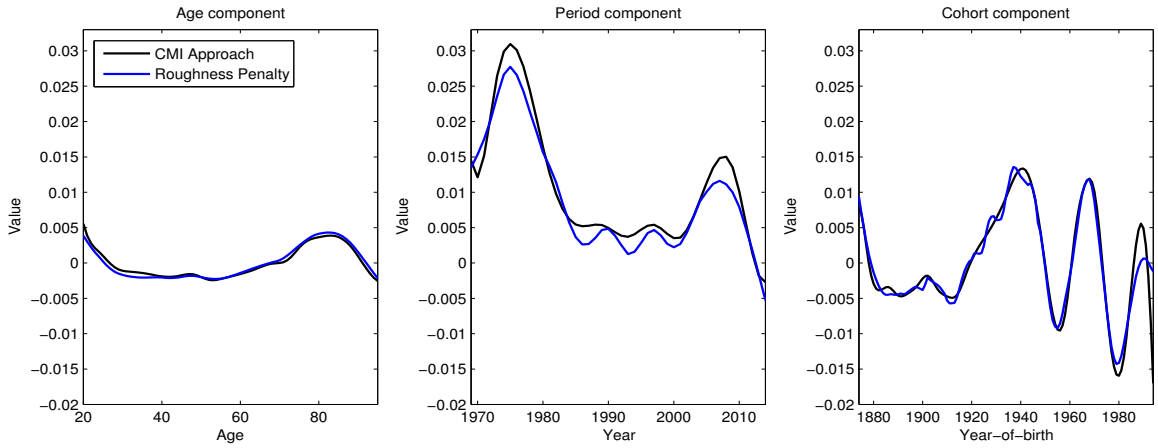


(a) U.S. Males: Dataset (i), HMD data

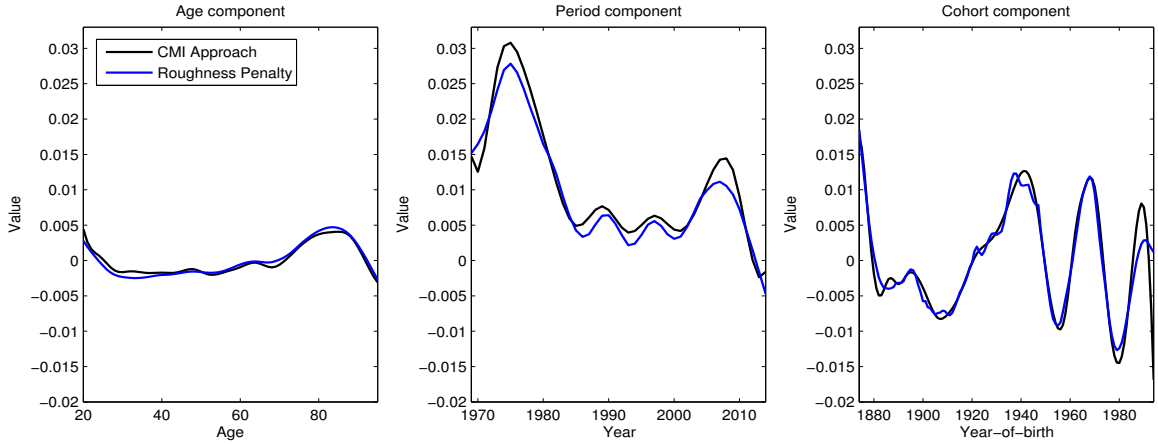


(b) U.S. Males: Dataset (ii), SSA data

Figure 48: The estimated Age/Period/Cohort components obtained from the CMI-09 method when different smoothing methods are used, U.S. **males**.



(a) U.S. Females: Dataset (i), HMD data



(b) U.S. Females: Dataset (ii), SSA data

Figure 49: The estimated Age/Period/Cohort components obtained from the CMI-09 method when different smoothing methods are used, U.S. **females**.

9 Conclusion

The following are the major conclusions drawn in this volume of the project report:

1. The CMI-09 method may be applied to the U.S. data sets, but there is a major drawback. When applied to the U.S. male data sets, large vertical clusters are found in the heat maps of the standardized residuals. This outcome in an indication that the APC model structure used in the CMI-09 method is unable to pick up some features that are specific to the U.S. historical mortality experience. It is found that adding an age breakpoint ameliorates the problem, but the use of an age breakpoint may result in inconsistencies between mortality projections for younger and older ages.

2. The CMI-09 method is fairly insensitive to (1) changes in the calibration window, (2) changes in the age range, (3) changes in the parameter constraints used, and (4) inclusion/exclusion of the oldest/newest cohorts.

3. Among all seven Route A candidate model structures, the simplified Plat model is the most suitable for modeling the U.S. smoothed mortality improvement rates in terms of both robustness and goodness-of-fit. The simplified model also eliminates the need for an age breakpoint.

4. A more statistically rigorous approach is to integrate smoothing and estimation into one single process. When using the roughness penalty approach to combine smoothing and estimation, the resulting age, period and cohort components remain roughly the same.

We conclude that if Route A is chosen, then it is the most appropriate to use the simplified Plat model (without any age breakpoint) to model the smoothed mortality improvement rates for the U.S. population. The A/P/C decompositions resulting from the Route A simplified Plat model are shown in Figure 50 (males) and Figure 51 (females).

The next volume of the project report focuses on Route B. It also compares the optimal modeling results from Routes A and B, and makes a final recommendation.

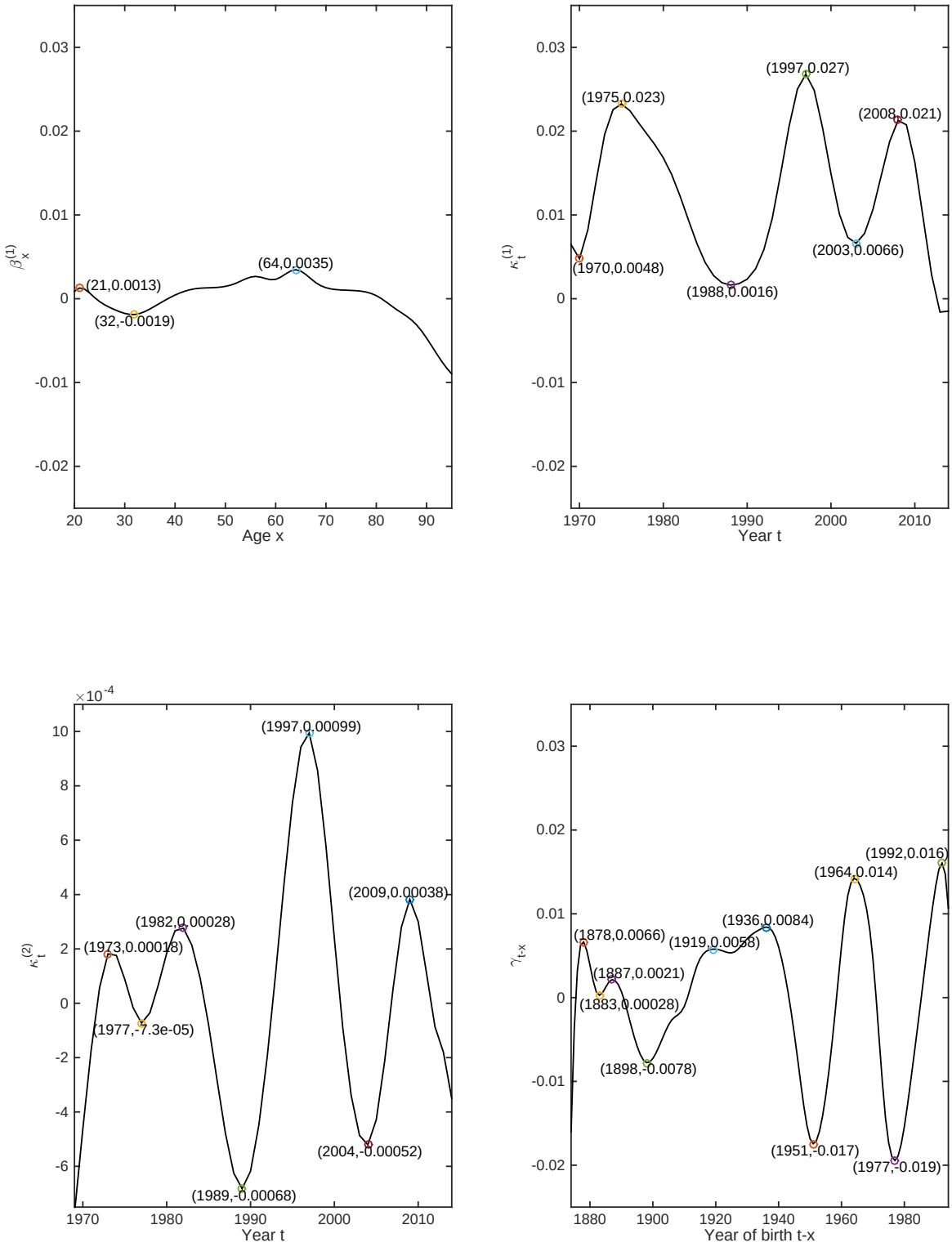


Figure 50: The estimated Age/Period/Cohort components obtained from the Route A simplified Plat model, U.S. **males**.

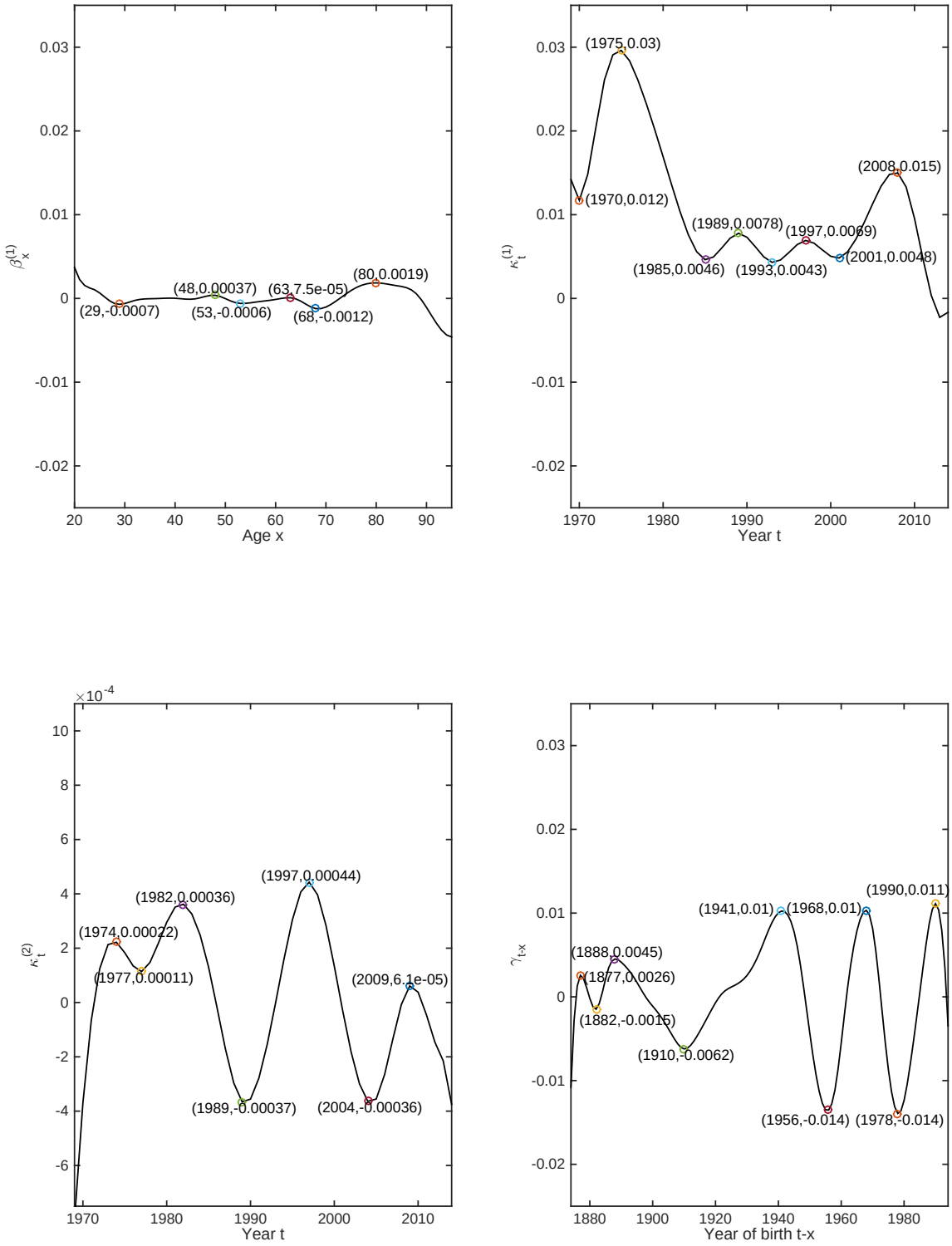


Figure 51: The estimated Age/Period/Cohort components obtained from the Route A simplified Plat model, U.S. **females**.

References

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.

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 Blake, D. (2015). Identifiability in Age/Period/Cohort Mortality Models. Pensions Institute Technical Report PI 15-09, Cass Business School, City University of London.

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

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.

Technical Appendix

This appendix provides further details on the roughness penalty approach used in Section 8.

In implementing the roughness penalty approach, we have specified Δ_a, Δ_k and Δ_g in such a way that the degree of roughness is measured by the sum of squared second-order differences of the parameters. In effect, we are using a second-order polynomial as a standard of smoothness. The same standard of smoothness is also adopted in the new (2017) CMI decomposition method. Although, in principle, a third- or even fourth-order polynomial can be used instead, we choose to use a second-order polynomial to maximize consistency with the new CMI method.

The degrees of smoothness in the estimates of the A/P/C components are determined by the smoothing parameters π_a, π_k and π_g. When the smoothing parameters are zero, the A/P/C components are not smoothed at all and are identical to the ordinary least squares estimates. In the other extreme when $\pi_a/\pi_k/\pi_g \to \infty$, the estimates of the A/P/C components follow perfectly linear patterns. 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_a, π_k and π_g, re-estimate the A/P/C 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':

$$(e^{-}_{x,t})^2 = (Z_{x,t} - \hat{a}^{-}_x - \hat{k}^{-}_t - \hat{g}^{-}_c)^2,$$

where $\hat{a}^{-}_x$, $\hat{k}^{-}_t$ and $\hat{g}^{-}_c$ respectively represent the estimated values of a_x, k_t and g_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

$$\text{SPE}(\pi_a, \pi_k, \pi_g) = \sum_{x=20}^{95} \sum_{t=1968}^{2014} (e^{-}_{x,t})^2$$

(iii) Repeat steps (i) and (ii) for different possible values π_a, π_k and π_g; choose the value of π_a, π_k and π_g that leads to the smallest sum of squared prediction errors.

About The Society of Actuaries

The Society of Actuaries (SOA), formed in 1949, is one of the largest actuarial professional organizations in the world dedicated to serving more than 27,000 actuarial members and the public in the United States, Canada and worldwide. In line with the SOA Vision Statement, actuaries act as business leaders who develop and use mathematical models to measure and manage risk in support of financial security for individuals, organizations and the public.

The SOA supports actuaries and advances knowledge through research and education. As part of its work, the SOA seeks to inform public policy development and public understanding through research. The SOA aspires to be a trusted source of objective, data-driven research and analysis with an actuarial perspective for its members, industry, policymakers and the public. This distinct perspective comes from the SOA as an association of actuaries, who have a rigorous formal education and direct experience as practitioners as they perform applied research. The SOA also welcomes the opportunity to partner with other organizations in our work where appropriate.

The SOA has a history of working with public policymakers and regulators in developing historical experience studies and projection techniques as well as individual reports on health care, retirement and other topics. The SOA's research is intended to aid the work of policymakers and regulators and follow certain core principles:

Objectivity: The SOA's research informs and provides analysis that can be relied upon by other individuals or organizations involved in public policy discussions. The SOA does not take advocacy positions or lobby specific policy proposals.

Quality: The SOA aspires to the highest ethical and quality standards in all of its research and analysis. Our research process is overseen by experienced actuaries and nonactuaries from a range of industry sectors and organizations. A rigorous peer-review process ensures the quality and integrity of our work.

Relevance: The SOA provides timely research on public policy issues. Our research advances actuarial knowledge while providing critical insights on key policy issues, and thereby provides value to stakeholders and decision makers.

Quantification: The SOA leverages the diverse skill sets of actuaries to provide research and findings that are driven by the best available data and methods. Actuaries use detailed modeling to analyze financial risk and provide distinct insight and quantification. Further, actuarial standards require transparency and the disclosure of the assumptions and analytic approach underlying the work.

Society of Actuaries
475 N. Martingale Road, Suite 600
Schaumburg, Illinois 60173
www.SOA.org