Testing Gender Disparity in Mortality Improvement Trends in Asia-Pacific Countries: Implications on the Life Insurance Industry

NOVEMBER| 2023

Mortality and Longevity

Aging and Retirement

LIVING TO 100
SOA INTERNATIONAL SYMPOSIUM

SOA Research INSTITUTE

Testing Gender Disparity in Mortality Improvement Trends in Asia-Pacific Countries: Implications on the Life Insurance Industry

AUTHOR

Johnny S.-H. Li, FSA, Ph.D.
Department of Statistics and Actuarial Science
University of Waterloo

Kenneth Q. Zhou, FSA, Ph.D.
School of Mathematical and Statistical Science
Arizona State University

Xiaobai Zhu
School of Insurance
Southwestern University of Finance and Economics

Wai-Sum Chan, FSA, Ph.D.
Department of Finance
The Chinese University of Hong Kong

SPONSORS

Mortality and Longevity Strategic Research Program Steering Committee

Aging and Retirement Strategic Research Program Steering Committee

Give us your feedback!
Take a short survey on this report.
Click Here

Caveat and Disclaimer
The opinions expressed and conclusions reached by the authors are their own and do not represent any official position or opinion of the Society of Actuaries Research Institute, Society of Actuaries, or its members. The Society of Actuaries Research Institute makes no representation or warranty to the accuracy of the information.

Copyright © 2023 Society of Actuaries Research Institute

Testing gender disparity in mortality improvement trends in Asia-Pacific countries: Implications on the life insurance industry

Johnny S.-H. Li[1], Kenneth Q. Zhou[2], Xiaobai Zhu[3,4] and Wai-Sum Chan[4]

[1]Department of Statistics and Actuarial Science, University of Waterloo, Canada
[2]School of Mathematical and Statistical Science, Arizona State University, USA
[3]School of Insurance, Southwestern University of Finance and Economics, China
[4]Department of Finance, The Chinese University of Hong Kong, Hong Kong

Abstract: In this project, we study the issue of gender disparity in mortality improvement trends. We first develop a statistical test to examine the extent of such disparity for different age groups, and then apply the proposed test to mortality data from various Asia-Pacific countries. Our preliminary results indicate that there exists a significant long-term divergence in mortality improvement trends between genders, and the extent of such a divergence depends on age and geographical location. For example, in Japan gender disparity is more significant at retirement ages, but in China working age groups are experiencing stronger gender disparity. We further develop an adapted Lee-Carter model that captures the gender disparity found in the statistical test. The proposed model is then used to investigate the impact of gender disparity in mortality improvement trends on life insurers in the region, particularly that concerning gender-neutral pricing of life insurance and annuity products.

Keywords: *Gender disparity; Asia-Pacific countries; Mortality improvement trends; Gender-neutral pricing*

1 Introduction

Gender is one important factor in life insurance pricing and valuation. Starting in 2013, life insurers in European Union (EU) member states are prohibited to factor gender into life insurance and annuity premiums, and need to charge the same price to male and female for the same insurance products.[1] This rule is known as gender-neutral pricing in insurance, and may have a number of implications on actuarial practices. For example, a gender-neutral premium on a life insurance could cause female to pay a higher rate than male, while the lower rate for male could make this life insurance more attractive to male, which in turn leads to adverse selection. A similar gender-neutral pricing rule also exists in the auto insurance industry in California, USA.[2]

Related to the practice of gender-neutral pricing, gender disparity in mortality rates and mortality improvement trends is another influential matter in life insurance pricing and valuation. It is well-known that female tends to have a lower mortality rate than male at the same age. However, this difference in mortality

[1]Factsheet: EU rules on gender-neutral pricing in insurance
[2]California Prohibits Auto Insurance Companies From Considering Gender When Setting Prices

rate between male and female might be changing over time, and more importantly the rate of change could be age-dependent and population-specific. Consequently, life insurance products issued to different age groups and populations might be affected differently by the practice of gender-neutral pricing. In this paper, we study the implications of gender-neutral pricing and gender disparity in the Asia-Pacific life insurance industries.

We aim to achieve three objectives. The first objective is to develop a method to test the statistical significance of gender disparity in mortality improvement trends. Considering mortality data from various Asia-Pacific countries, our preliminary results reveal that there exists a significant long-term divergence in mortality improvement trends between genders, and the extent of such a divergence depends on age and geographical location. The test results suggest that existing stochastic mortality models for forecasting mortality of males and females simultaneously should be adapted; in particular, the usual assumption of coherence (non-divergence) must be reconsidered.

The second objective is to construct a modified version of the Li-Lee model for gender-specific mortality modeling, on the basis of the gender disparity in mortality improvements that is identified through the mentioned statistical test. We make the modeling framework flexible enough so that the projected mortality improvement trends for males and females can be diverging for some age groups but non-diverging for the others. To achieve this goal, we will consider incorporating a piece-wise age function and estimate the proposed model under the Bayesian paradigm.

The final objective is to investigate the impact of gender disparity in mortality improvement trends on the life insurance market. Specifically, we considered six life insurance products with different age ranges, terms, gender disparity and gender mixes. On the basis of a gender-neutral regulation scenario, we examined how the funding position of various life insurance portfolios is affected by gender disparity and gender-neutral pricing.

The rest of this paper is outlined as follows. Section 2 provides a visual analysis and a statistical analysis to examine the existence and level of gender disparity in mortality. Section 3 develops a modified version of the Li-Lee model for gender-specific mortality modeling along with the model's estimation and projection results. Section 4 provides a numerical analysis on the impact of gender-neutral pricing and gender disparity in life insurance products. Section 5 concludes the paper.

2 Gender disparity in mortality

2.1 Visual analysis

The data used in this paper include three Asia-Pacific populations, namely, Mainland China, Taiwan and Japan. Our goal is to explore the degree of disparity in the difference of mortality improvement trends between genders. We start with visually examining the observed mortality rates of both genders of these three populations in different years.

Figure 1 shows the log-scale central death rates of both genders, ages 0-99 and years 1981, 2000 and 2010 for the total population of China, Japan and Taiwan, respectively. We focus on years 1981, 2000 and 2010 because only these three years have full age range mortality data collected from Chinese censuses. We

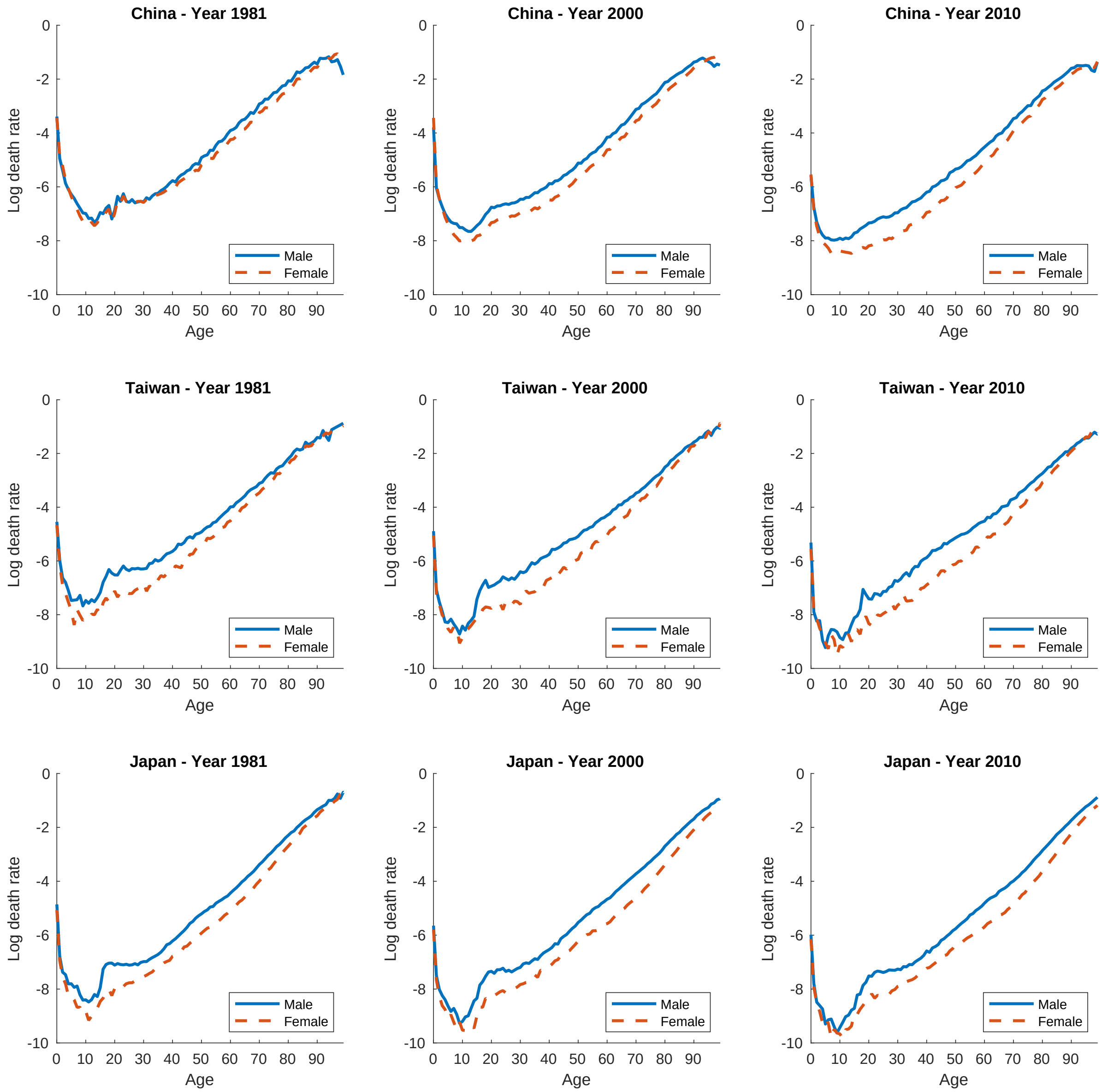

Figure 1: Observed log-scale central death rates of males (blue solid lines) and females (red dash lines) for China (top row), Taiwan (middle row) and Japan (bottom row) in years 1981, 2000 and 2010.

can draw the following empirical conclusions from Figure 1:

- The width of the gap in central death rates between males and females are different in the three years shown. For example, the difference in central death rates between males and females is not significant in China in year 1981, whereas the same difference in year 2010 is very obvious. The same observation of a widening gap can also be made for Japan and Taiwan. Thus, we say that gender disparity in mortality experience is changing over time for the populations under consideration.
- The widening gap in central death rates between males and females is observed in different age ranges among the three populations under consideration. In particular, the widening gap is more obviously

observed in young ages (around ages 20-50) in China, middle ages (around ages 50-80) in Taiwan, and old ages (around ages 80-100) in Japan. Thus, we conclude that gender disparity in mortality experience depends on age.

- Based on the previous observation, we can also say that, depending on the population under consideration, the extent of gender disparity in mortality experience might be varying for different age groups. The young working class in China has been experiencing strong gender disparity in mortality experience in the recent decades, while the same situation is experienced in Taiwan by the old working class and young retirees. For Japan, gender disparity in mortality experience exists mostly in very old ages.

To provide a better visual assessment of the above three observations, Figure 2 shows the difference in log-scale death rates between males and females for Mainland China, Taiwan and Japan in years 1981, 2000 and 2010. The difference is calculated as the log-scale death rate of the males minus that of the females. Because female death rates are in general lower than male death rates, we observed that most of the values shown in Figure a2 are positive. If the values are becoming more positive (i.e., increasing) over time, then we say that the difference in death rates between males and females is increasing or widening, and there exists gender disparity in mortality experience.

In Figure 2, it is obvious that gender disparity in mortality experience exists, and more importantly, depending on the age groups and population (geographical location), the level of gender disparity is different. For some age groups in a certain population, the difference is widening over time, indicating that there is gender disparity in mortality improvement trends. For example, in Japan, the difference is gradually increasing over time for ages 80-100, while in China, the difference is dramatically increasing over time for ages 20-50. These empirical results again suggest that there is clear gender disparity in the three population but at different age groups. Furthermore, by comparing the changes of the difference over time, it is doubtful that the diverging mortality improvement trend has ever slowed down. In term of the coherent modeling assumption, it is hard to believe that this difference in death rates between genders for all ages should be modeled by a stationary process, as in, for example, the original Li and Lee model (Li and Lee, 2005).

We thus conclude in this subsection that there exists a significant level of gender disparity in mortality improvement experiences, and the extent of such disparity depends on age and geographical location in the Asia-Pacific region. To verify this empirical conjecture, we develop a statistical method to test the significance of gender disparity in mortality improvement experiences in the next subsection. Lastly, we remark that in many existing mortality modeling studies for Asia-Pacific populations, the idea of "borrowing information" from a more developed country (e.g., Japan) to aid the mortality forecasts of a less developed country (e.g., China) is not supported by our empirical findings here.

2.2 Statistical analysis

2.2.1 Setup

To establish a foundation for the statistical method to be used for verifying gender disparity, let us first consider the common factor model proposed in Li and Lee (2005) for jointly modeling the central death

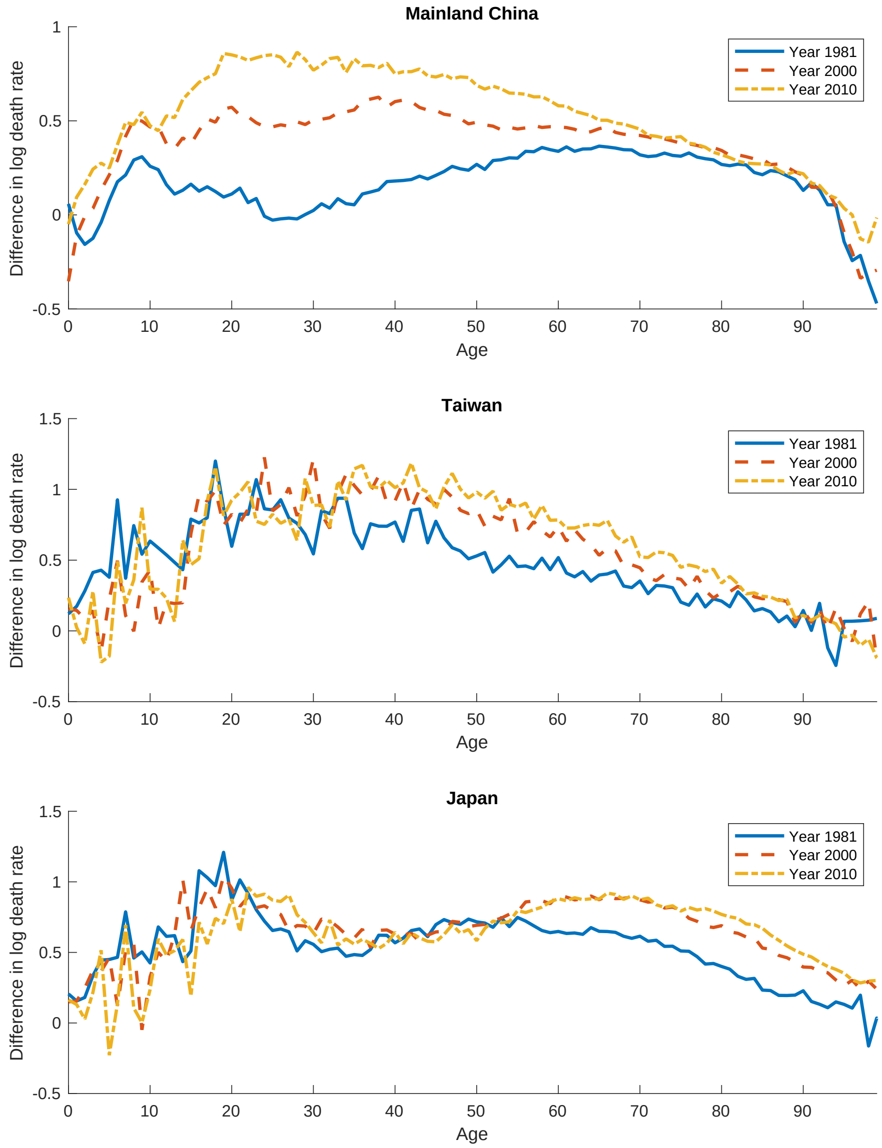

Figure 2: Difference in the observed log death rates between males and females for Mainland China (top panel), Taiwan (middle panel) and Japan (bottom panel) for years 1981, 2000 and 2010. The difference is calculated as the log-scale death rate of the males minus that of the females.

rates of both males (denoted by $m_{x,t,m}$) and females (denoted by $m_{x,t,f}$):

$$\ln m_{x,t,m} = \alpha_{x,m} + B_x K_t + \epsilon_{x,t,m}$$
$$\ln m_{x,t,f} = \alpha_{x,f} + B_x K_t + \epsilon_{x,t,f}$$

where $\alpha_{x,m}$ and $\alpha_{x,f}$ are the gender-specific age effect, B_x is the common age effect, K_t is the common period effect, and $\epsilon_{x,t,m}$ and $\epsilon_{x,t,f}$ are the gender-specific residuals. Since $m_{x,t,m}$ and $m_{x,t,f}$ might be obtained from different sample sizes (exposures) at different age x and time t (especially for China where the mortality data could be collected from census, 1% sampling survey, or 0.1% sampling survey), we assume that $\epsilon_{x,t,g} = \sigma_{x,t,g} z_{x,t,g}$, $g = \{m, f\}$, and $z_{x,t,g}$ follows a i.i.d. standard normal distribution for all x, t and g. We discuss how $\sigma_{x,t,g}$ can be estimated in Section 2.2.3.

To capture any short-term deviation from the common trend $B_x K_t$, Li and Lee (2005) further proposed the augmented common factor model, which is specified as

$$\ln m_{x,t,m} = \alpha_{x,m} + B_x K_t + \beta_{x,m}\kappa_{t,m} + \epsilon_{x,t,m}$$
$$\ln m_{x,t,f} = \alpha_{x,f} + B_x K_t + \beta_{x,f}\kappa_{t,f} + \epsilon_{x,t,f},$$

where $\beta_{x,f}$ and $\beta_{x,f}$ are two additional gender-specific age effects, $\kappa_{t,m}$ and $\kappa_{t,f}$ are two gender-specific period effects, and the remaining model parameters maintain their original definition. To ensure that the difference in central death rates between male and female does not diverge indefinitely over time (i.e., maintain coherence in gender-specific mortality forecasting), it is assumed that K_t follows a random walk with drift, while $\kappa_{t,m}$ and $\kappa_{t,f}$ follow two independent stationary AR(1) processes.

The empirical results analyzed in the previous subsection suggest that the gender difference in central death rates might be widening and hence diverging over time. Our goal is to test whether this empirical finding is statistically sound. To do so, we first define the the gender difference in central death rates as

$$\begin{aligned}\eta_{x,t} &:= \ln m_{x,t,m} - \ln m_{x,t,f} \\ &= (\alpha_{,m} - \alpha_{x,f}) + (\beta_{x,m}\kappa_{t,m} - \beta_{x,f}\kappa_{t,f}) + (\epsilon_{x,t,m} - \epsilon_{x,t,f})\end{aligned}$$

and further denote that

$$\zeta_{x,t} := \beta_{x,m}\kappa_{t,m} - \beta_{x,f}\kappa_{t,f}.$$

It can be shown that if $\kappa_{t,m}$ and $\kappa_{t,f}$ follow two independent stationary AR(1) processes, then, for each x, $\zeta_{x,t}$ follows a stationary ARMA(2,1) process and $\eta_{x,t}$ follows a stationary ARMA(2,2) process.[3]

[3] Assume that $\kappa_{t,f}$ and $\kappa_{t,m}$ follow two independent stationary AR(1) processes. For $g = \{m, f\}$, let us denote $\kappa_{x,t,g} := \beta_{x,g}\kappa_{t,g} = \theta_g + \phi_g \kappa_{x,t-1,g} + \omega_{t,g}$, where we omitted x in the subscript of process parameters for simplicity. Then, we have

$$\begin{aligned}\zeta_{x,t} &= \kappa_{x,t,m} - \kappa_{x,t,f} \\ &= \theta_m - \theta_f - \theta_m\phi_f + \theta_f\phi_m + (\phi_m + \phi_f)\zeta_{x,t-1} - \phi_m\phi_f\zeta_{x,t-2} + \omega_{t,m} - \omega_{t,f} - \phi_f\omega_{t-1,m} + \phi_m\omega_{t-1,f} \\ &= \tilde{\theta} + \tilde{\phi}_1\zeta_{x,t-1} + \tilde{\phi}_2\zeta_{x,t-2} + \tilde{\omega}_t\end{aligned}$$

which is an ARMA(2,1) process. We also have

$$\begin{aligned}\eta_{x,t} &= (\alpha_{x,m} - \alpha_{x,f}) + \zeta_{x,t} + (\epsilon_{x,t,m} - \epsilon_{x,t,f}) \\ &= \tilde{\alpha}_x + \zeta_{x,t} + \tilde{\epsilon}_{x,t} \\ &= \tilde{\alpha}_x(1 - \tilde{\phi}_1 - \tilde{\phi}_2) + \tilde{\epsilon}_{x,t} - \tilde{\phi}_1\tilde{\epsilon}_{x,t-1} - \tilde{\phi}_2\tilde{\epsilon}_{x,t-2} + \tilde{\phi}_1\eta_{x,t-1} + \tilde{\phi}_2\eta_{x,t-2} + \tilde{\omega}_t\end{aligned}$$

which is an ARMA(2,2) process.

2.2.2 Our testing procedure

To verify whether $\eta_{x,t}$ follows a stationary process, we conduct the Kwiatkowski–Phillips–Schmidt–Shin (KPSS) test for each x.[4] We begin with our initial belief that the gender differences will converge ($\eta_{x,t}$ is a stationary process) for all ages, unless strong evidence against our null hypothesis.

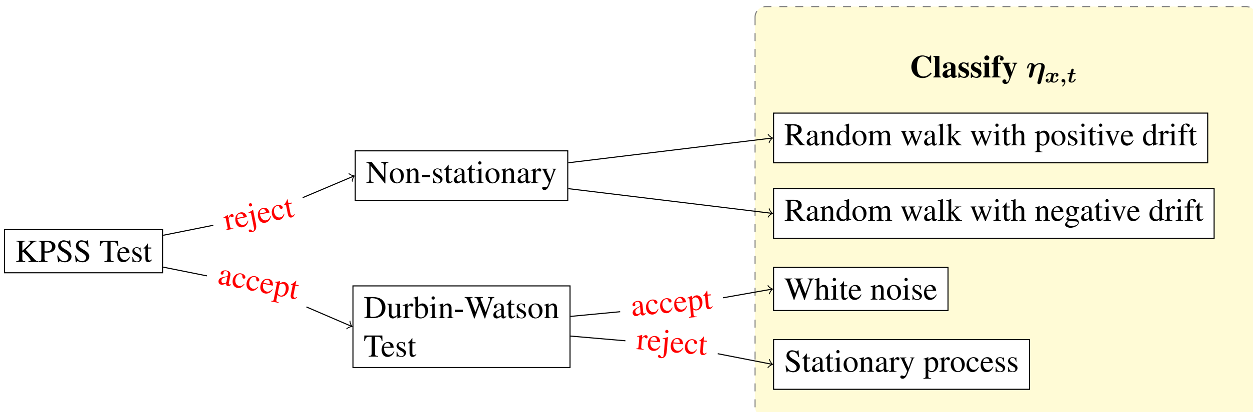

Figure 3: Workflow of our testing procedure

Aside from the previous setup on a mortality model, the KPSS test is conducted by assuming that for each age x, the process of $\eta_{x,t}$ is

$$\eta_{x,t} = u_{x,t} + \delta_x \times t + e_{x,t}$$
$$u_{x,t} = u_{x,t-1} + \sigma_x \epsilon_{x,t}$$

where $\epsilon_{x,t}$ is a white noise and $e_{x,t}$ is a stationary process. The KPSS test tests over the null $H_0 : \sigma_x = 0$ versus the alternative $H_a : \sigma_x > 0$. If we have significant evidence against the stationary null for a particular age x, then $\eta_{x,t}$ is non-stationary, and we will assume it follows a random walk with drift. We need to be careful about the sign of the drift, as a positive drift represents a diverging trend whereas a negative drift represents an initial converging trend until the mortality rates, $\ln m_{x,t,m}$ and $\ln m_{x,t,f}$, have a cross-over. If there is insufficient evidence to reject the null hypothesis, then we are further interested in whether the gender difference has already reached its converged state. Therefore, the Durbin-Watson test is performed to verify if $\eta_{x,t}$ is simply a white noise. If we have sufficient evidence to reject the Durbin-Watson test, then $\eta_{x,t}$ will be modeled by an AR(1) process. The entire testing procedure is displayed in Figure 3.

2.2.3 Adjustment for time and age effects

Because of the age- and time-varying sample sizes of the observed death rates, we have previously assumed that the residuals $\epsilon_{x,t,m}$ and $\epsilon_{x,t,f}$ have age- and time-dependent variance, that is $\epsilon_{x,t,g} = \sigma_{x,t,g} z_{x,t,g}$ for $g = \{m, f\}$. It follows that the residuals of $\eta_{x,t}$ (i.e., $\epsilon_{x,t,m} - \epsilon_{x,t,f}$) will also have age- and time-dependent

[4] Other stationary tests, such as the ADF test, Hadri-Larsson test, and Levin-Lin test, can also be considered.

variance, which we denote as $\sigma^2_{x,t}$. Consequently, we cannot directly apply the KPSS test to $\eta_{x,t}$, but instead need to perform adjustments first.

In developed countries, σ^2 is often assumed to be both age- and time-independent. However, for developing countries, this assumption can not hold as the data may obtained from different sources. For example, depending on the sample year t, the Chinese mortality data could be collected from either census, 1% sampling survey, or 0.1% sampling survey. In addition, due to limited data for Asia-Pacific countries, we will aggregate $\sigma^2_{x,t}$ on the age-dimension, such that $\sigma^2_{x,t} = \sigma^2_t$, and σ^2_t can take three different values depending on year t.

To empirically estimate σ^2_t, we need to remove the time and age effects first. The time effect is captured by defining $\bar{\eta}_t$ as

$$\bar{\eta}_t = \frac{1}{|\mathcal{X}_t|} \sum_{x \in \mathcal{X}_t} \eta_{x,t},$$

where $\mathcal{X}_t$ as the set of ages with data at time t (due to missing data), and $|\cdot|$ is the cardinality of the set (i.e., the number of elements). $\bar{\eta}_t$ measures the overall gender disparity at time t. Next, the age effect is captured by defining $\bar{\eta}_x$ as

$$\bar{\eta}_x = \frac{1}{\sum_{t \in \mathcal{T}} \mathbb{1}_{x \in \mathcal{X}_t}} \sum_{t \in \mathcal{T}} \mathbb{1}_{x \in \mathcal{X}_t} \times (\eta_{x,t} - \bar{\eta}_t),$$

where $\mathcal{T}$ is the set contains all years. Then, we may define the adjusted residuals $\hat{\eta}_{x,t}$ such that both age and time effects are removed,

$$\hat{\eta}_{x,t} = \eta_{x,t} - \bar{\eta}_t - \bar{\eta}_x.$$

Now, let $\mathcal{G}_1$, $\mathcal{G}_2$ and $\mathcal{G}_3$ be three groups of sample years that contain mortality data collected from census year, 1% survey year, and 0.1% survey year, respectively. The empirical variance can be estimated via $\hat{\eta}_{x,t}$ as

$$\hat{\sigma}^2_{x,t} = \hat{\sigma}^2_t = \frac{1}{\sum_{s \in \mathcal{G}_i} |\mathcal{X}_s| - 1} \sum_{s \in \mathcal{G}_i} \sum_{y \in \mathcal{X}_s} \hat{\eta}^2_{y,s}, \qquad t \in \mathcal{G}_i.$$

The standardized residuals $\frac{\hat{\eta}_{x,t}}{\hat{\sigma}_{x,t}}$ are approximately white noises with variance of one. To retain the same overall volatility of the original residuals, we define the empirical standard deviation for all residuals $\hat{\sigma}^2$ as

$$\hat{\sigma}^2 = \frac{1}{\sum_{s \in \mathcal{T}} |\mathcal{X}_s| - 1} \sum_{s \in \mathcal{T}} \sum_{y \in \mathcal{X}_s} \hat{\eta}^2_{y,s},$$

and the transformed residuals $\frac{\hat{\eta}_{x,t}}{\hat{\sigma}_{x,t}} \times \hat{\sigma}$ will preserve approximately same overall volatility as the original $\eta_{x,t}$. Next, by adding back the age and time effects to the transformed residuals, we are able to construct a new process $\tilde{\eta}_{x,t}$ that retains the same age and time effects as in the original $\eta_{x,t}$,

$$\tilde{\eta}_{x,t} = \frac{\hat{\eta}_{x,t}}{\hat{\sigma}_{x,t}} \times \hat{\sigma} + \bar{\eta}_t + \bar{\eta}_x.$$

Since the residuals of $\tilde{\eta}_{x,t}$ are white noises, the stationary tests can be directly applied.

2.2.4 Test results

We now apply the KPSS test to $\tilde{\eta}_{x,t}$, the standardized gender difference in central death rates, to verify whether it is stationary and to determine whether gender disparity exists for a particular age x over time. To focus on insurance applications, we shorten the age range from 0-99 to 30-90 and use all available data after year 1981 (the first data available year of China). We also add the population of Hong Kong into our statistical analysis here, which has mortality data from year 1986 and onward. The standardization process of $\eta_{x,t}$ for Hong Kong is similar to that of Japan and Taiwan but with a shorter sample period.

Figure 4 summarizes the test results from applying the KPSS test to $\tilde{\eta}_{x,t}$ for Japan, Taiwan, Hong Kong and Mainland China. The KPSS test is individually applied to $\tilde{\eta}_{x,t}$ for each age x, for $x = 30, \ldots, 90$, over the sample period. The test result is either a value of 1 indicating a rejection of the null hypothesis at a significance level of 5%, or a value of 0 indicating the opposite. So, if the test result reported in Figure 4 is 1 for a certain age x and population, then we say that $\tilde{\eta}_{x,t}$ is non-stationary and gender disparity is present. Furthermore, if the KPSS test rejects the null hypothesis, then we use a symbol of $\triangledown$ to indicate that the gender difference is getting larger, and use a symbol of $\triangle$ to represent that the gender difference is becoming smaller.

If there is not enough evidence to reject the stationary null hypothesis, then we additionally perform the Durbin-Watson test to check whether the standardized gender difference $\tilde{\eta}_{x,t}$ follows an autoregressive process or simply a white noise without drift. The null hypothesis of the Durbin-Watson test assumes that $\tilde{\eta}_{x,t}$ is a serially uncorrelated process. We use a symbol of $\bigcirc$ to indicate that the null hypothesis of the Durbin-Watson test is rejected and $\tilde{\eta}_{x,t}$ has autocorrelation at lag 1, and a symbol of $\otimes$ to indicate that there is not enough evidence to reject the null hypothesis.

We can observe from Figure 4 that different populations produce different test results. For Japan, ages higher than 60 unanimously show $\triangledown$, indicating that the gender difference is getting larger over time for these ages. This conclusion coincides with the empirical observations made in the previous subsection, for example, in Figure 2. For ages between 40 and 60, the KPSS test indicates that the gender difference is mostly getting smaller, while for ages lower than 40, a rejection of the stationary null hypothesis cannot be made for some ages.

For Taiwan, almost all ages between 30 and 90 show $\triangledown$ indicating that gender difference is widening for the entire insurance age range. For Hong Kong, gender disparity is getting larger for ages higher than 65, while for the other ages, the test result varies among the four cases ($\triangledown$, $\triangle$, $\bigcirc$ and $\otimes$).

Finally, for China, a situation that is unlike the other three populations is observed. For ages younger than 60, we see that a test result of $\triangledown$ is produced for most ages, while for ages older than 60, a test result of $\bigcirc$ is mostly observed. The conclusion that ages higher than 60 in China are not experiencing gender

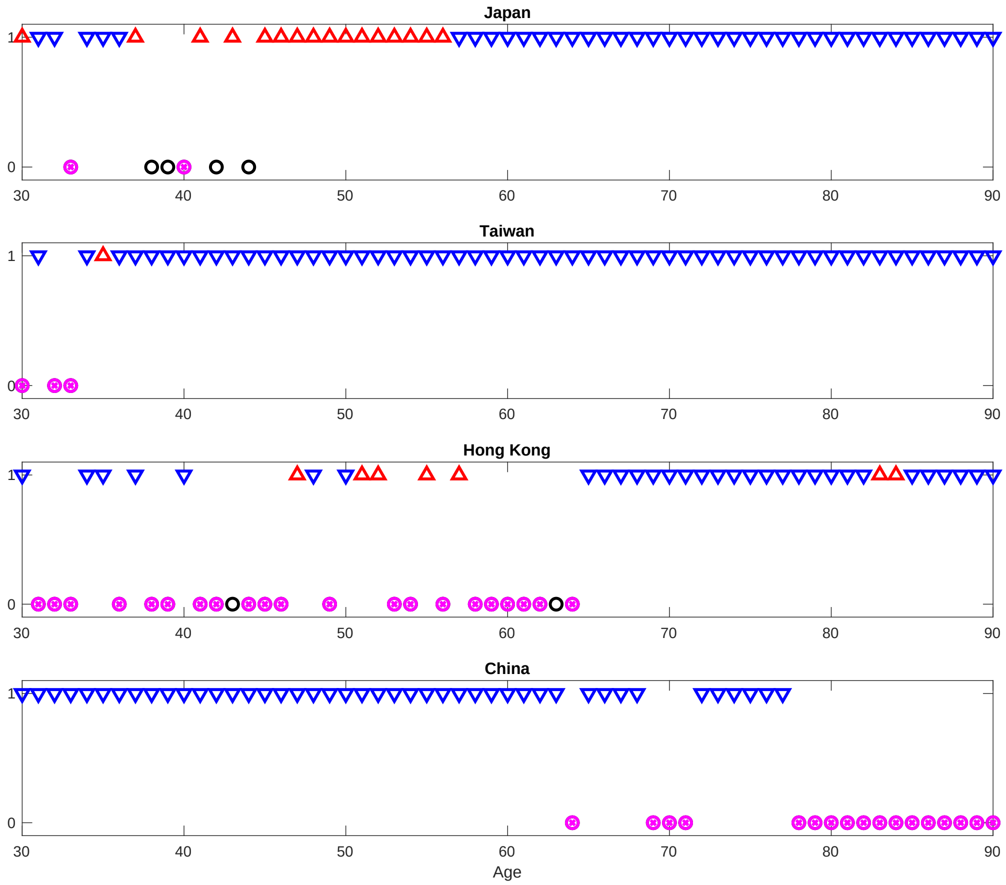

Figure 4: Test results from applying the KPSS test to the standardized gender difference in central death rates $\tilde{\eta}_{x,t}$, $x = 30, \ldots, 90$, for Japan, Taiwan, Hong Kong and Mainland China. Note: $\triangledown$ indicates that gender difference is widening, $\triangle$ indicates that gender difference is narrowing, $\bigcirc$ indicates that gender difference is stationary with autocorrelation, and $\otimes$ indicates that gender difference is stationary with no autocorrelation.

disparity is unique compared to the other three populations. For ages lower than 60, gender disparity is present for China in most ages, which is similar to Taiwan but different than Japan and Hong Kong.

In conclusion, the KPSS test results confirm our previous conjecture that there exists a significant level of gender disparity in mortality improvement trends, and more importantly the extent of such disparity depends on age and geographical location in the Asia-Pacific region. Moreover, when gender disparity is not present, there may or may not be a need to use an autoregressive process to capture any short-term gender differences as in the Li and Lee model.

3 Modeling mortality with gender disparity

The conclusions made in the previous section will have implications for an life insurer who is practicing gender-neutral pricing. To properly capture this age-varying phenomenon of gender disparity in mortality forecasts and other relevant actuarial practices, one needs a tailor-made stochastic mortality model. We develop in this section a tailored stochastic mortality model based on the original Li and Lee model for incorporating age-specific gender disparity. We estimate the proposed model using the Bayesian method to not only address the uncertain indication of gender disparity over the age range, but also solve the data-related issues of Asia-Pacific populations. At the end, we apply the proposed model and estimation method to the gender-specific mortality data sets of Japan, Taiwan, Hong Kong and Mainland China, and analyze the estimation results.

3.1 The modified Li and Lee model

Both empirical and statistical results from the previous section suggest that the level and type of gender disparity vary among different ages. Recall that the gender difference in central death rates from the original Li and Lee model is specified as

$$\eta_{x,t} = (\alpha_{x,m} - \alpha_{x,f}) + (\beta_{x,m}\kappa_{t,m} - \beta_{x,f}\kappa_{t,f}) + (\epsilon_{x,t,m} - \epsilon_{x,t,f}),$$

where the two stochastic components $\kappa_{t,m}$ and $\kappa_{t,f}$ are both stationary and age-independent. Thus, as an 'all-age' coherent mortality model, the original Li and Lee model is unable to capture this behavior in mortality modeling and forecasting.

To address the issue of age-dependent gender disparity, we consider an age-dependent stochastic component to replace the term $\beta_{x,m}\kappa_{t,m} - \beta_{x,f}\kappa_{t,f}$ in the original Li and Lee model. It is obviously infeasible to assume one stochastic process for each single age (i.e., simply considering an age- and time-specific period effect). However, as observed in Figure 4 from the previous section, adjacent ages in a population may share the same type of gender disparity. We thus propose a piece-wise structure for the age-dependent stochastic component, and cluster the individual ages into four groups, in which the gender disparity is assumed to be widening (non-stationary), diminishing (non-stationary), auto-regressive (stationary) and stable (stationary).

Denote

- $\mathcal{D}_1$ as the set of ages with non-stationary $\eta_{x,t}$ and widening gender disparity,
- $\mathcal{D}_2$ as the set of ages with non-stationary $\eta_{x,t}$ and diminishing gender disparity,
- $\mathcal{D}_3$ as the set of ages with stationary $\eta_{x,t}$ following an AR(1) process, and
- $\mathcal{D}_4$ as the set of ages with stationary $\eta_{x,t}$ following a white noise process.

It is reasonable to assume that an individual age must belong to one and only one of the above four sets. In other words, we assume that $\mathcal{D}_1$, $\mathcal{D}_2$, $\mathcal{D}_3$ and $\mathcal{D}_4$ are disjoint sets and the union of $\mathcal{D}_1$, $\mathcal{D}_2$, $\mathcal{D}_3$ and $\mathcal{D}_4$ is the full age range under consideration. We further denote that $\mathcal{D} = \mathcal{D}_1 \cup \mathcal{D}_2$ as the set of all ages, where $\eta_{x,t}$ follows a non-stationary process.

The proposed stochastic component for capturing the age-dependent gender disparity (which replaces the term $\zeta_{x,t} = \beta_{x,m}\kappa_{t,m} - \beta_{x,f}\kappa_{t,f}$ in the original Li and Lee model) is specified as

$$\zeta_{x,t} = \beta_x \times \mathbf{1}_x' \times \boldsymbol{\kappa}_t,$$

where β_x is an age-specific parameter capturing the sensitivity of gender disparity to $\boldsymbol{\kappa}_t$,

$$\mathbf{1}_x = \begin{pmatrix} \mathbb{1}_{x\in\mathcal{D}_1} \\ \mathbb{1}_{x\in\mathcal{D}_2} \\ \mathbb{1}_{x\in\mathcal{D}_3} \end{pmatrix} \quad \text{and} \quad \boldsymbol{\kappa}_t = \begin{pmatrix} \kappa_t^{\mathcal{D}_1} \\ \kappa_t^{\mathcal{D}_2} \\ \kappa_t^{\mathcal{D}_3} \end{pmatrix}.$$

Since $\mathcal{D}_1$, $\mathcal{D}_2$ and $\mathcal{D}_3$ are disjoint, only one of the indicator functions in $\mathbf{1}_x$ will produce a value of 1, indicating which of the period effects in $\boldsymbol{\kappa}_t$ should be used. If an age below to $\mathcal{D}_4$ (i.e., $\eta_{x,t}$ follows a white noise process), then no period effect is needed in modeling gender disparity and we have $\mathbf{1}_x$ becoming a vector of zeros.

To model gender disparity with the right age-dependent features, we consider the following stochastic process for $\boldsymbol{\kappa}_t$:

$$\begin{pmatrix} \kappa_t^{\mathcal{D}_1} \\ \kappa_t^{\mathcal{D}_2} \\ \kappa_t^{\mathcal{D}_3} \end{pmatrix} = \begin{pmatrix} \theta_{\mathcal{D}_1} \\ \theta_{\mathcal{D}_2} \\ \theta_{\mathcal{D}_3} \end{pmatrix} + \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & \phi \end{pmatrix} \begin{pmatrix} \kappa_{t-1}^{\mathcal{D}_1} \\ \kappa_{t-1}^{\mathcal{D}_2} \\ \kappa_{t-1}^{\mathcal{D}_3} \end{pmatrix} + \begin{pmatrix} \omega_t^{\mathcal{D}_1} \\ \omega_t^{\mathcal{D}_2} \\ \omega_t^{\mathcal{D}_3} \end{pmatrix}$$

where $\theta_{\mathcal{D}_1}$, $\theta_{\mathcal{D}_2}$ and $\theta_{\mathcal{D}_3}$ are the drift/offset terms, ϕ is the AR coefficient, and $\omega_t^{\mathcal{D}_1}$, $\omega_t^{\mathcal{D}_2}$ and $\omega_t^{\mathcal{D}_3}$ are the innovations at time t following a normal distribution with a mean of zero and a variance of $\sigma^2_{\mathcal{D}_1}$, $\sigma^2_{\mathcal{D}_2}$ and $\sigma^2_{\mathcal{D}_3}$, respectively. To capture the widening gender disparity, we should have $\theta_{\mathcal{D}_1} > 0$ (the gender difference is continuing to drift upward and becoming more positive), while to the diminishing gender disparity, we should have $\theta_{\mathcal{D}_2} < 0$ (the gender difference is approaching zero). Lastly, to have a stationary process for $\eta_{x,t}$, we should have $0 < \phi < 1$.

For the common period effect K_t, we continue to assume that it follows a random walk with drift as in the original Li and Lee model; that is,

$$K_t = K_{t-1} + \theta + \omega_t,$$

where θ is the constant drift term, and ω_t is the innovation at time t following a normal distribution with a mean of zero and a variance of σ^2.

The original Li and Lee model is subject to identifiability issues. The modification made in this section inherits the problem, and thus needs model constraints in the estimation process. We consider the following set of constraints for the modified Li and Lee model:

$$\sum_x B_x = 1, \sum_{x\in\mathcal{D}_1} \beta_x = 1, \sum_{x\in\mathcal{D}_2} \beta_x = 1, \sum_{x\in\mathcal{D}_3} \beta_x = 1$$

$$K_{t_0} = 0, \kappa_{t_0}^{\mathcal{D}_1} = 0, \kappa_{t_0}^{\mathcal{D}_2} = 0, \kappa_{t_0}^{\mathcal{D}_3} = 0$$

where t_0 is the first year of the sample period.

3.2 Model estimation

To estimate the proposed model, we follow the Bayesian estimation work of Pedroza (2006) and Li et al. (2019), who used Gibbs sampling methods to obtain the joint posterior distribution of all model parameters with parameter uncertainty taken into account.

Let n_a and n_y denote the number of ages and years in the data set, respectively. The number of parameters in the modified Li and Lee model can be calculated by

$$n_p = 3n_a + |\mathcal{D}_1| + |\mathcal{D}_2| + |\mathcal{D}_3| + (1 + \mathbb{1}_{|\mathcal{D}_1|>0} + \mathbb{1}_{|\mathcal{D}_2|>0} + \mathbb{1}_{|\mathcal{D}_3|>0})(n_y + 2) + 4 + \mathbb{1}_{|\mathcal{D}_3|>0}.$$

If the Chinese population is being modeled, we have four additional parameters for different sampling uncertainties. Denote Θ as the set of all of the parameters in the proposed model, and use $\Theta_{-\theta_j}, j = 1, \cdots, \theta_{n_p}$ to represent Θ excluding its $j$the entry. In each iteration of Gibbs sampling, a sample of θ_j is drawn from its full conditional, and the process is repeated after a large number of iterations. For the state variables (K_t and $\kappa_t^{\mathcal{D}_i}$ for $i = 1, 2, 3$), the full conditionals are obtained with sequential Kalman Filter (Koopman and Durbin (2000)). For all of the model parameters, improper prior distributions are assumed.

We summarize the estimation details of each parameter as follows:

- To apply Kalman Filter directly, we rewrite the model in a more compact state-space form:

$$\begin{aligned} \text{observation equation:} \quad & \boldsymbol{y}_t = \tilde{\boldsymbol{\alpha}} + \tilde{\boldsymbol{B}} \times \tilde{\boldsymbol{K}}_t + \tilde{\boldsymbol{\epsilon}}_t, \qquad \tilde{\boldsymbol{\epsilon}}_t \sim \text{MVN}(0, \Sigma_y) \\ \text{state equation:} \quad & \tilde{\boldsymbol{K}}_t = \tilde{\boldsymbol{\theta}} + \tilde{\boldsymbol{\Phi}} \times \tilde{\boldsymbol{K}}_{t-1} + \tilde{\boldsymbol{\omega}}_t, \quad \tilde{\boldsymbol{\omega}}_t \sim \text{MVN}(0, \Sigma_K) \end{aligned}$$

  where

$$\begin{aligned} \boldsymbol{y}_t &= (\ln m_{x_0,t,m}, \cdots, \ln m_{x_0+n_a-1,t,m}, \ln m_{x_0,t,f}, \cdots, \ln m_{x_0+n_a-1,t,f})^T, \\ \tilde{\boldsymbol{\alpha}} &= (\alpha_{x_0,m}, \cdots, \alpha_{x_0+n_a-1,m}, \alpha_{x_0,f}, \cdots, \alpha_{x_0+n_a-1,f})^T, \\ \tilde{\boldsymbol{B}} &= ((B_{x_0}, \cdots, B_{x_0+n_a-1}, B_{x_0}, \cdots, B_{x_0+n_a-1})^T, \tilde{\boldsymbol{\beta}}), \\ \tilde{\boldsymbol{\beta}} &= \left[\tilde{\boldsymbol{\beta}}\right]_{ij} = \begin{cases} \beta_i \times \mathbb{1}_{i\in\mathcal{D}_j} & i = 1, \cdots, n_a, \qquad\qquad j = 1, 2, 3 \\ 0 & i = n_a + 1, \cdots, 2 \times n_a, \quad j = 1, 2, 3, \end{cases} \\ \tilde{\boldsymbol{K}}_t &= (K_t, \kappa_t^{\mathcal{D}_1}, \kappa_t^{\mathcal{D}_2}, \kappa_t^{\mathcal{D}_3})^T, \\ \tilde{\boldsymbol{\theta}} &= (\theta, \theta^{\mathcal{D}_1}, \theta^{\mathcal{D}_2}, \theta^{\mathcal{D}_3})^T, \\ \tilde{\boldsymbol{\Phi}} &= \text{diag}(1, 1, 1, \phi), \\ \Sigma_y &= \text{diag}(s_{t,m}^2, \cdots, s_{t,m}^2, s_{t,f}^2, \cdots, s_{t,f}^2), \\ \Sigma_K &= \text{diag}(\sigma^2, \sigma_{\mathcal{D}_1}^2, \sigma_{\mathcal{D}_2}^2, \sigma_{\mathcal{D}_3}^2). \end{aligned}$$

  The model constraints on $\tilde{\boldsymbol{K}}_{t_0}$ are incorporated by setting $\text{E}[\tilde{\boldsymbol{K}}_{t_0}] = (0, 0, 0, 0)^T$ and $\text{Cov}(\tilde{\boldsymbol{K}}_{t_0}) = \text{diag}(0, 0, 0, 0)$. Based on the above state-space form, the filtering process and the smoothing process are then used to draw samples for the state variable $\tilde{\boldsymbol{K}}_t$. We refer the interested reader to Li et al. (2019) for the technical detail.

- The full conditional of $\alpha_{x,g}$ is $\mathrm{N}(\mu_\alpha, \Sigma_\alpha)$, where

$$\mu_\alpha = \Sigma_\alpha \times \left(\frac{\sum_{t\in\mathcal{T}} \ln m_{x,t,g} - B_x \times K_t - \beta_x \times \mathbf{1}'_x \times \boldsymbol{\kappa}_t}{s^2_{t,g}} \right),$$
$$\Sigma_\alpha = \left(\sum_{t\in\mathcal{T}} \frac{1}{s^2_{t,g}} \right)^{-1}.$$

- The model constraints on $\tilde{\boldsymbol{B}}$ (i.e., the sum on B_x and β_x) are incorporated into their priors. Using B_x as an example, notice that the constraint can be written as $B_{x_0} = 1 - \sum_{x>x_0} B_x$. Let $\boldsymbol{B}_{-x_0}$ be the vector of B_x except x_0. Then, the full conditional of $\boldsymbol{B}_{-x_0}$ is $\mathrm{MVN}(\boldsymbol{\mu}_B, \boldsymbol{\Sigma}_B)$, where

$$\boldsymbol{\mu}_B = \boldsymbol{\Sigma}_B \times \left(\sum_{g\in\{m,f\}} \sum_{t\in\mathcal{T}} \frac{\boldsymbol{H}_B^T \times \tilde{\boldsymbol{y}}_{t,g} \times K_t}{s^2_{t,g}} \right)$$
$$\boldsymbol{\Sigma}_B = \left(\sum_{g\in\{m,f\}} \sum_{t\in\mathcal{T}} \frac{\boldsymbol{H}_B^T \times \boldsymbol{H}_B \times K_t^2}{s^2_{t,g}} \right)^{-1}$$

  with

$$\tilde{\boldsymbol{y}}_{\boldsymbol{t,g}} = (\ln m_{x_0,t,g} - \alpha_{x_0,g} - \beta_{x_0}\mathbf{1}'_{x_0}\boldsymbol{\kappa}_t, \cdots, \ln m_{x_0+n_a-1,t,g} - \alpha_{x_0+n_a-1,g} - \beta_{x_0+n_a-1}\mathbf{1}'_{x_0+n_a-1}\boldsymbol{\kappa}_t)^T,$$

  $\boldsymbol{H}_B = \begin{pmatrix} -1 \times \mathbf{1}^T_{n_a-1} \\ \boldsymbol{I}_{n_a-1} \end{pmatrix}$, $\mathbf{1}_{n_a-1}$ is a $(n_a-1)\times 1$ vector of ones, and $\boldsymbol{I}_{n_a-1}$ is an $(n_a-1)\times(n_a-1)$ identity matrix. The full conditional for β_x can be derived similarly by dividing ages into different age groups according to $\mathcal{D}_i$, $i = 1, 2, 3$, and replacing K_t by $\kappa_t^{\mathcal{D}_i}$.

- The full conditional of $s^2_{t,g}$, for all $t \in \mathcal{G}_i$, is an inverse gamma with a shape parameter of a_s and a rate parameter of b_s, where

$$a_s = \frac{\sum_{t\in\mathcal{G}_i} |\mathcal{X}_t|}{2},$$
$$b_s = \frac{\sum_{t\in\mathcal{G}_i} \sum_{x\in\mathcal{X}_t} (\ln m_{x,t,g} - \alpha_{x,g} - B_x \times K_t - \beta_x \times \mathbf{1}_x \times \kappa_t)}{2}.$$

- The full conditional of σ^2 is an inverse gamma with a shape parameter a_σ and a rate parameter of b_σ, where

$$a_\sigma = \frac{|\mathcal{T}| - 1}{2},$$
$$b_\sigma = \frac{\sum_{t\in\mathcal{T}\setminus t_0} (K_t - \theta - K_{t-1})^2}{2}.$$

  The full conditional of $\sigma^2_{\mathcal{D}_i}$ can be derived by replacing K with $\kappa_t^{\mathcal{D}_i}$ and θ with $\theta_{\mathcal{D}_i}$. The only exception is $\sigma^2_{\mathcal{D}_3}$ where the numerator of b_σ is modified to subtracting $\phi \times \kappa_{t-1}^{\mathcal{D}_3}$ instead of $\kappa_{t-1}^{\mathcal{D}_3}$.

- The full conditional of θ is $\mathrm{N}(\mu_\theta, \Sigma_\theta)$, where

$$\mu_\theta = \frac{\sum_{t\in\mathcal{T}\setminus t_0} K_t - \theta - K_{t-1}}{|\mathcal{T}| - 1},$$
$$\Sigma_\theta = \frac{\sigma^2}{|\mathcal{T}| - 1}.$$

  The same modification for σ^2 can be adopted to get the full conditionals for $\theta_{\mathcal{D}_i}$.

- For parameter ϕ, using a prior of $\pi(\phi) \sim U(0,1)$, the full conditional of ϕ is $\mathrm{N}(\mu_\phi, \Sigma_\phi)$ truncated between $(0,1)$, where

$$\mu_\phi = \left(\sum_{t\in\mathcal{T}\setminus t_0} \left(\kappa_{t-1}^{\mathcal{D}_3}\right)^2\right)^{-1} \times \left(\sum_{t\in\mathcal{T}\setminus t_0} (\kappa_t^{\mathcal{D}_3} - \theta_{\mathcal{D}_3}) \times \kappa_{t-1}^{\mathcal{D}_3}\right),$$
$$\Sigma_\phi = \frac{\sigma^2_{\mathcal{D}_t}}{\sum_{t\in\mathcal{T}\setminus t_0} \left(\kappa_{t-1}^{\mathcal{D}_3}\right)^2}.$$

It is well known that samples drawn from Gibbs sampling exhibit auto-correlation, and hence thinning must be applied to avoid bias in assessing the posterior distribution. We use the method of Effective Sample Size (ESS) to decide the number of effective samples for our final Bayesian analysis. In addition, both the single-chain convergence test (Geweke (1991)) and the multi-chain convergence test (Gelman and Rubin (1992)) are applied to verify the convergence of the effective samples.

3.3 Estimation results

In this subsection, we present the empirical posterior distribution of the model parameters from fitting the modified Li and Lee model proposed in Section 3.1 using the Bayesian method described in Section 3.2. We considered mortality data from the four populations, namely, Japan, Taiwan, Hong Kong and Mainland China, which we have analyzed in Section 2. To initialize the Bayesian estimation procedure, we need to first decide the age groups, $\mathcal{D}_1$, $\mathcal{D}_2$, $\mathcal{D}_3$ and $\mathcal{D}_4$, for each population under consideration. We admit that this modeling component of grouping ages is not a trivial task, and may require subjective knowledge and judgment. The statistical tests conducted in Section 2.2 can help us to somewhat facilitate this subjective modeling component.

Based on the test results reported in Figure 4, we choose to group individual ages of each population as follows. For Japan, there are three age groups for the different types of gender disparity. For ages 30-40, gender disparity is not diverging but has short-term effect, and thus the gender difference follows an AR(1) process (i.e., $\mathcal{D}_3$). For ages 41-60, gender disparity is diminishing over time (i.e., $\mathcal{D}_2$), while gender disparity is widening over time (i.e., $\mathcal{D}_1$) for ages 61-90. For Taiwan, most ages have gender disparity widening over time (i.e., $\mathcal{D}_1$). For only ages 30-34, there is no gender disparity (i.e., $\mathcal{D}_4$). For Hong Kong and Mainland China, there are two age groups, namely $\mathcal{D}_1$ and $\mathcal{D}_4$, but the age range is different for them. For Hong Kong, old ages 65-90 will have a widening gender difference (i.e., $\mathcal{D}_1$), whereas the same type of gender disparity happens to young ages 30-60 in Mainland China. The rest of the individual ages of these two population will have no gender disparity (i.e., $\mathcal{D}_4$). The above description is summarized in Table 1.

Population	$\mathcal{D}_1$	$\mathcal{D}_2$	$\mathcal{D}_3$	$\mathcal{D}_4$
Japan	$\{61,\ldots,90\}$	$\{41,\ldots,60\}$	$\{30,\ldots,40\}$	$\emptyset$
Taiwan	$\{35,\ldots,90\}$	$\emptyset$	$\emptyset$	$\{30,\ldots,34\}$
Hong Kong	$\{65,\ldots,90\}$	$\emptyset$	$\emptyset$	$\{30,\ldots,64\}$
China	$\{30,\ldots,60\}$	$\emptyset$	$\emptyset$	$\{61,\ldots,90\}$

Table 1: Age grouping results for Japan, Taiwan, Hong Kong and Mainland China.

Using the age groups shown in Table 1, we fit the gender-specific mortality data of each population to the modified Li and Lee model using the Bayesian estimation method. We focus on analyzing the estimation results of Japan and China, and report the empirical posterior distribution of model parameters for Japan and China in Figures 5 and 6, respectively.

Figure 5 shows the empirical posterior distribution of model parameters obtained for Japan. We first focus on parameters $\alpha_{x,g}$, K_t and B_x, shown in the first row of Figure 5. The age- and gender-specific parameter $\alpha_{x,g}$, as the average log central death rate for age x and gender g, is showing a clear gap between the two genders and reflecting the difference in the level of death rates between genders. The common time-varying parameter K_t is showing a clear downward trend reflecting the persistent mortality improvement rate of both genders over, while the common age-specific parameter B_x is measuring the sensitivity of death rates to K_t for both genders. Since the mortality data of Japan is complete and large in exposure size, the level of parameter uncertainty in these parameters is low, as reflected by the width of the fan charts.[5]

For Japan, we have used three age groups $\mathcal{D}_1$, $\mathcal{D}_2$ and $\mathcal{D}_3$, which in turn require three time-varying parameters $\kappa_t^{\mathcal{D}_1}$, $\kappa_t^{\mathcal{D}_2}$ and $\kappa_t^{\mathcal{D}_3}$, each following a different stochastic process. The empirical posterior distributions of these parameters are shown in the middle row of Figure 5. It is clear that $\kappa_t^{\mathcal{D}_1}$ is following a random walk with a positive drift, driving gender disparity to be more severe over time for old ages. For $\kappa_t^{\mathcal{D}_2}$, we see a gradual downward trend over time, especially in the recent years, indicating that gender disparity is diminishing for middle ages. Lastly, $\kappa_t^{\mathcal{D}_3}$ follows an AR(1) process capturing the short-term gender disparity for young ages. The AR coefficient ϕ is displayed in the left middle panel of Figure 5, while parameter β_x for capturing the age-specific sensitivity is shown in the right middle panel of Figure 5.

The bottom row of Figure 5 shows parameters s^2, θ and σ^2. The means of the empirical posterior distribution of the drift/offset term θ vary significantly among age groups (and thus gender disparity types). We can clearly observe that for group $\mathcal{D}_1$ the mean is positive and its confidence interval does not include zero, suggesting a very strong force of widening gender disparity, while for group $\mathcal{D}_2$, the mean is negative and include zero indicating a weaker force of diminishing gender disparity. For $\mathcal{D}_3$, the offset term θ for the AR(1) process is centered around, indicating that in the long-term gender difference is converging to zero.[6] The empirical posterior distribution of the variance of the error term s^2, shown in the left bottom panel of

[5]Each fan chart shows the 10% confidence interval with the heaviest shading, surrounded by the 20%, 30%, . . . , 90% confidence intervals with progressively lighter shadings.

[6]Note that we used constraint $\kappa_{t_0}^{\mathcal{D}_3} = 0$ and assume an AR(1) process to $\kappa_t^{\mathcal{D}_3}$. It is thus reasonable to have an empirical posterior distribution of θ centered around zero.

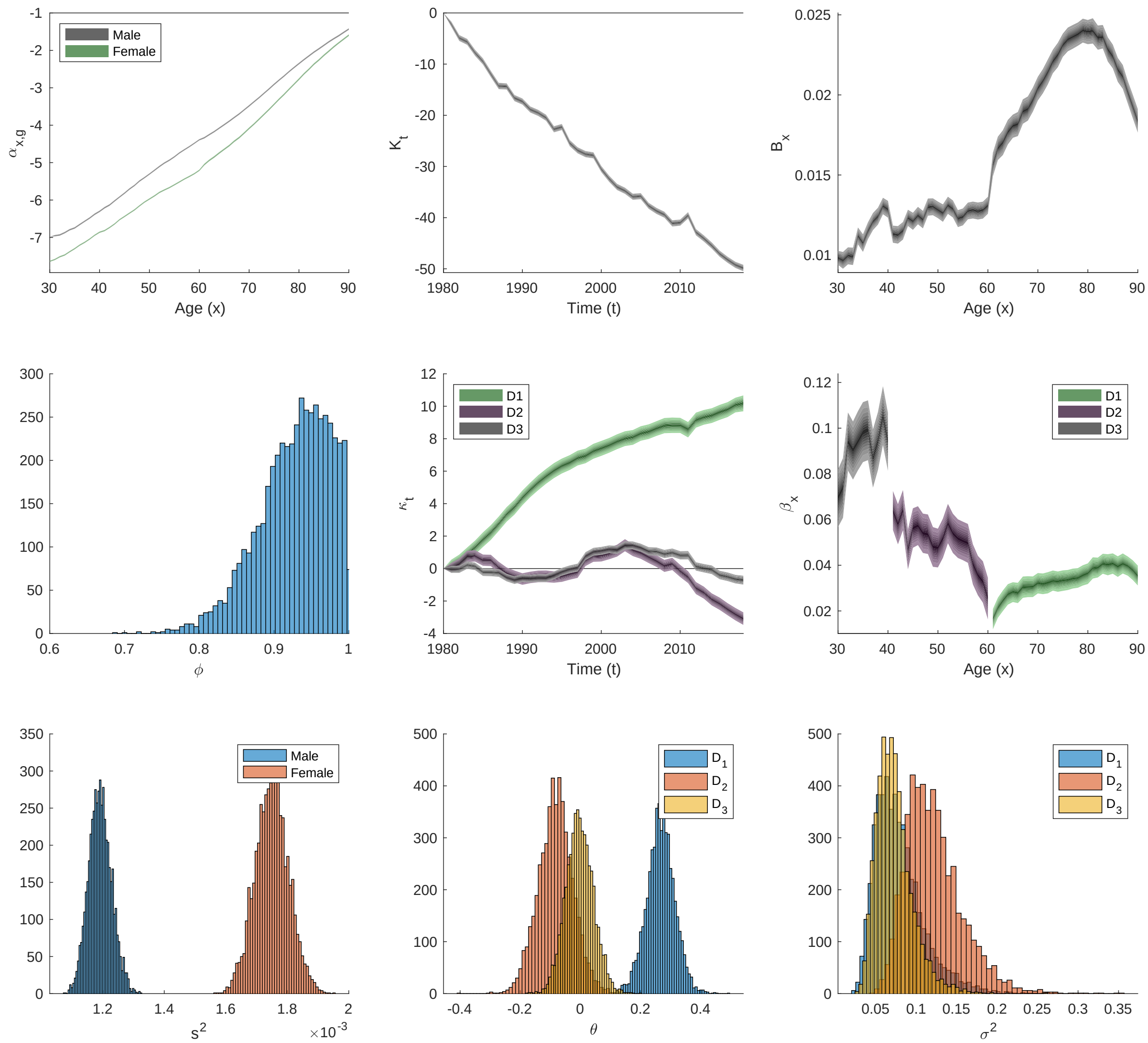

Figure 5: The empirical posterior distribution of model parameters obtained for Japan.

Figure 5, varies between the two genders, while the empirical posterior distribution of the variance of the innovation term σ^2 is also different among the three age groups.

Figure 6 shows the empirical posterior distribution of model parameters obtained for China. For $\alpha_{x,g}$, K_t and B_x, the observations are similar to that for Japan. However, because Chinese mortality data are subject to missing values and inconsistent exposure sizes, the obtained empirical posterior distributions, as shown by fan charts, are jagged over both ages and years. For years that have no data points, the width of the fan charts is significant, reflecting a higher level of parameter uncertainty. We refer the interested reader to Li et al. (2019) for a more detailed analysis of the data-related problems and modeling solutions for China.

Because we do not assume an AR(1) process to any period effect for China (i.e., there is no age group $\mathcal{D}_3$), there is no parameter ϕ to report in Figure 6. However, since we have ages 35-60 belonging to $\mathcal{D}_1$ for China (i.e., gender disparity is widening for the working ages in China), we have one time-varying period effect $\kappa_t^{\mathcal{D}_1}$ capturing the dynamics of gender difference for ages 30-60 over time, while the age-specific

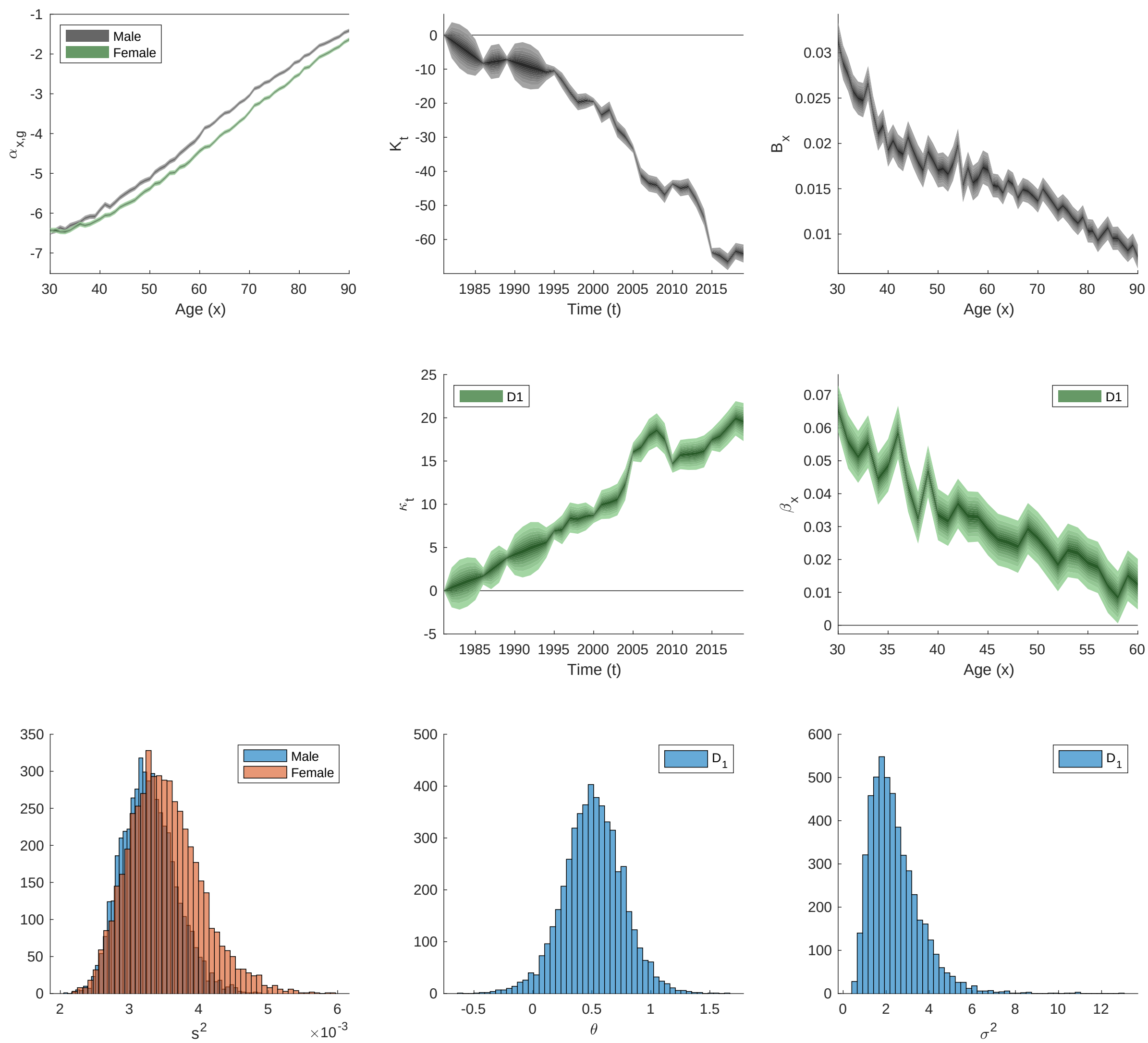

Figure 6: The empirical posterior distribution of model parameters obtained for China.

parameter β_x captures the sensitivity to $\kappa_t^{\mathcal{D}_1}$ for each age from 30 to 60. It is clear from the middle row of Figure 6 that $\kappa_t^{\mathcal{D}_1}$ has a upward trend over time indicating a diverging gender disparity for ages 30-60 in China.

The diverging gender disparity is confirmed by the empirical posterior distribution of $\theta_{\mathcal{D}_1}$ shown in the bottom row of Figure 6, which has a mean around 0.5 and low density for negative values. The empirical posterior distributions of s^2 and σ^2 are also reported in the bottom row of Figure 6, but are both unremarkable.

3.4 Projection results

We now discuss the mortality projection results produced by our Bayesian estimated modified Li and Lee model for Japan and Mainland China.

Figure 7 shows the predictive intervals of log-scaled central death rates for both genders in Japan from three different age groups. In particular, the left panel shows the mean forecast and 95% predictive interval for age 75, which is from Group $\mathcal{D}_1$ (widening gender difference) for Japan. We can observe that the mean forecasts over time for male and female death rates are diverging over time. The gender difference in death rates (in log scale) is less than 1 unit in 2019, while the difference increases to more than 1.5 units in 2065.

The middle panel shows the mean forecast and 95% predictive interval for age 55, which is from Group $\mathcal{D}_1$ (diminishing gender difference). It is clear that the mean forecasts between the two genders are approaching each other over time. Note that, as the gender difference diminishing over time, even at the end of the forecast period (year 2068), the relationship that females have lower mortality rates remain to hold.

The right panel shows the mean forecast and 95% predictive interval for age 35, which is from Group $\mathcal{D}_3$ (short-term stationary gender difference). We can observe that, in the long run, the death rates from the two genders are coherent over time, while in the short run, there is a short-term trend in gender difference governed by the estimated AR(1) process for $\kappa_t^{\mathcal{D}_3}$. Overall, it is clear that our proposed model is able to produce different types of gender disparity that were detected in our empirical and statistical analyses.

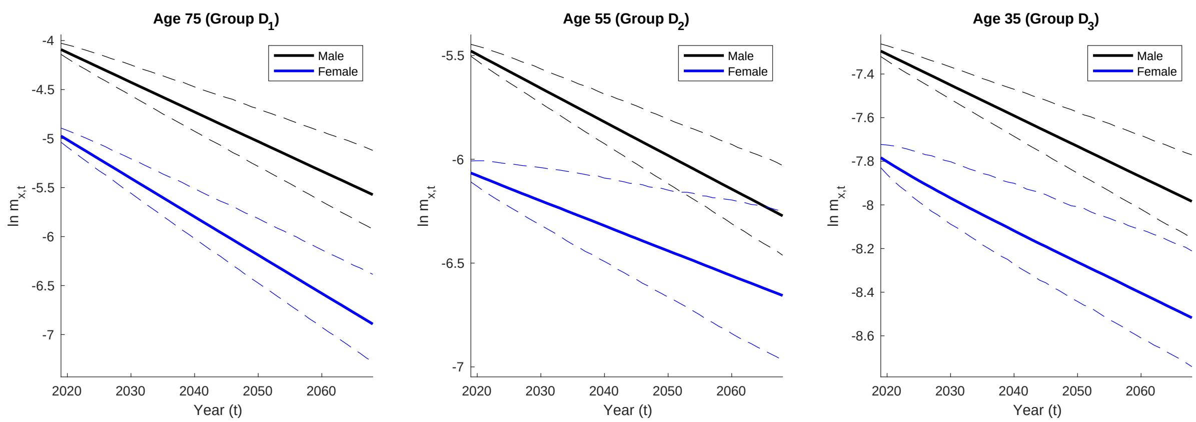

Figure 7: Predictive intervals of central death rates for both genders at ages 75 (Group $\mathcal{D}_1$), 55 (Group $\mathcal{D}_2$) and 35 (Group $\mathcal{D}_3$) in Japan.

Figure 8 shows the predictive intervals of log-scaled central death rates for both genders in Mainland China. We have assumed only two age groups for China, namely, Group $\mathcal{D}_1$ and Group $\mathcal{D}_4$. The left panel shows the mean forecast and 95% predictive interval for age 35 from Group $\mathcal{D}_1$, while the right panels shows the same results for age 75 from Group $\mathcal{D}_4$. For age 35, it is clear that the mean forecasts for the two genders are diverging over time. The gender difference in log-scaled death rates increases from 1 unit in year 2020 to about 2 units in year 2069. For age 75, since Group $\mathcal{D}_4$ has stable gender difference and no additional period effect is used, we see that the gap between male and female death rates is a constant over time. In other words, there is no gender disparity in mortality improvement trends for age 75 in China.

Lastly, we examine the projected period life expectancy at age 30 and 60 for both genders. Figure 9 and Figure 10 show the predictive intervals of period life expectancy in Japan and Mainland China, respectively. Although age groups from a population may have different underlying time-varying effects ($\kappa_t^{\mathcal{D}_1}$, $\kappa_t^{\mathcal{D}_2}$ and $\kappa_t^{\mathcal{D}_3}$) driven a diverging or converging process, we observe from Figure 9 and Figure 10 that the projected life expectancy at both ages seems to be coherent between the two genders. The values for male and female are clearly converging in the long run, and the difference is slowly narrowing over time. Comparing Japan

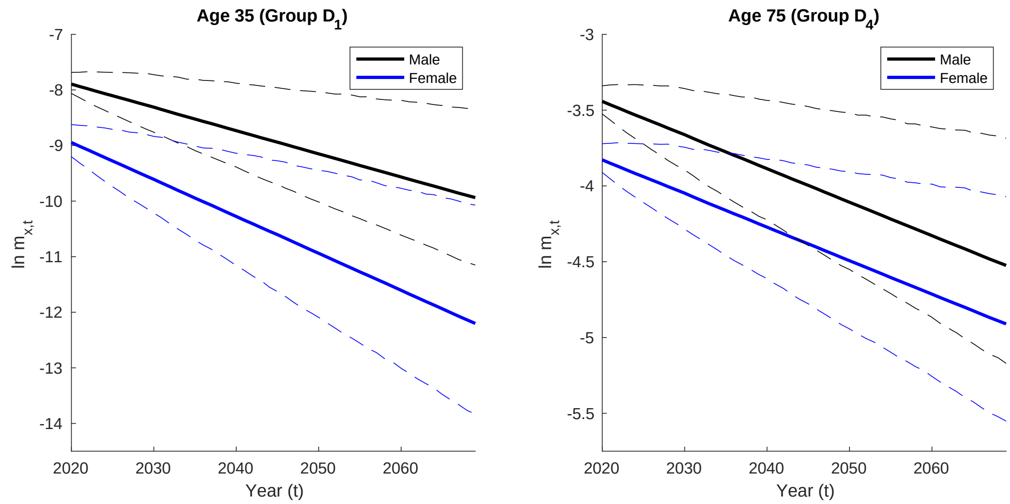

Figure 8: Predictive intervals of central death rates for both genders at ages 35 (Group $\mathcal{D}_1$) and 75 (Group $\mathcal{D}_4$) in Mainland China.

and China, we see that the projected values for China have higher volatility because of the data problems in Chinese mortality, while the mean values of Japan is clearly higher than that for China.

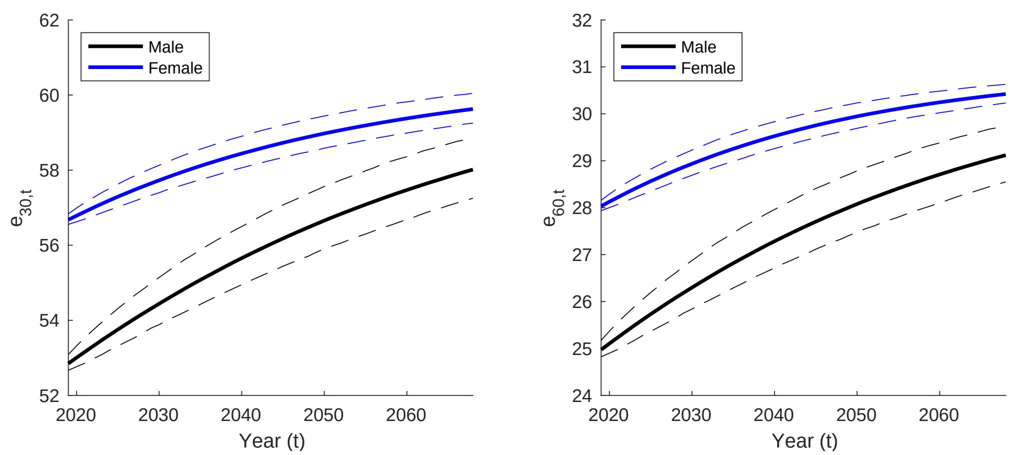

Figure 9: Predictive intervals of period life expectancy for both genders at age 30 (left panel) and age 60 (right panel) in Japan.

4 Gender-neutral pricing

In this section, we investigate the impact of the gender disparity discovered in Section 2 and modeled in Section 3. In particular, we considered both life insurance and life annuity products with various age, term and gender settings, and examine the effect of gender-neutral pricing on the funding position of life insurance portfolios. We illustrate our numerical results using the Japanese population, which has three different types of gender disparity.

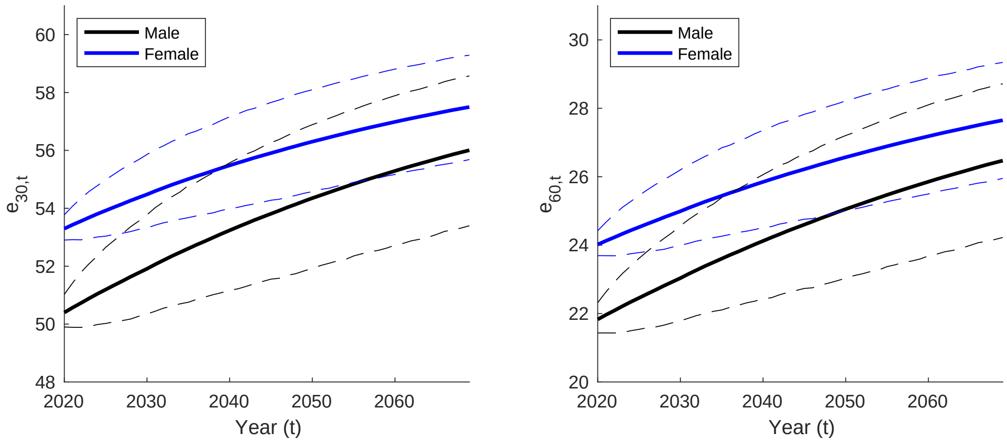

Figure 10: Predictive intervals of period life expectancy for both genders at age 30 (left panel) and age 60 (right panel) in Mainland China.

4.1 Setup

The Bayesian estimated model in Section 3 has provided the mean forecasts of age-specific central death rates with different gender disparity assumptions for both genders of a national population. We assume that an equally weighted average of these mean forecasts between the two genders is used by the insurance regulator of the population to create a gender-neutral life table. We further assume that the life insurer has to use this gender-neutral life table to price their life insurance products regardless of the insured's gender mix.

To explore the impact of gender-neutral pricing on the products of a life insurer with different gender disparity, we consider the following three life insurance products:

- Insurance 1: A 10-year term life insurance issued to an individual age 30,
- Insurance 2: A 20-year term life insurance issued to an individual age 40,
- Insurance 3: A 30-year term life insurance issued to an individual age 60,

and three life annuity products:

- Annuity 1: A 10-year term life annuity issued to an individual age 30,
- Annuity 2: A 20-year term life annuity issued to an individual age 40,
- Annuity 3: A 30-year term life annuity issued to an individual age 60.

Note that the age and term of these products are selected to match the age groups of the three types of gender disparity reported in Table 1 for Japan. For each of these products, we further consider a gender mix of

- Mix 1: 25% males and 75% females,
- Mix 2: 50% males and 50% females, and
- Mix 3: 75% males and 25% females.

In other words, we create three hypothetical portfolios (of one unit of the product) with a different gender mix for each of the products under consideration.

4.2 Baseline results

Figure 11 shows the projected survival probabilities for an individual aged 30, 40 and 60 for a survival period of 10, 20 and 30 years, respectively. The survival probabilities for male and female are obtained from our Bayesian estimated model, while the ones for the gender-neutral case are calculated using equal weights on central death rates between the two genders. For the three ages considered, it is clear that the survival probabilities for female are higher than that for male, while the gender-neutral ones roughly sits in the middles of the male and females cases. Note that the gender-neutral values will be used to produce a gender-neutral price for each of the three life insurance and three annuity products.

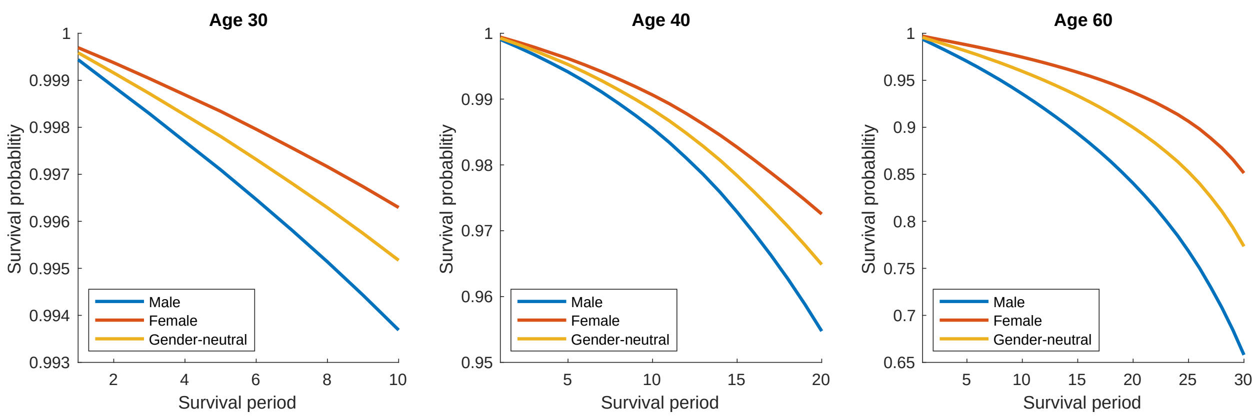

Figure 11: Projected survival probabilities for an individual aged 30 (left panel), 40 (middle panel) and 60 (right panel) for a survival period of 10, 20 and 30 years, respectively, for male, female and the gender-neutral case.

Figure 12 shows the empirical distribution of present values of the projected liability cash flows of each of the six products under consideration. These present values are calculated using a constant discount rate of 3%, and obtained from the predictive distribution of the estimated model. Note that, without any assumption on gender-neutral pricing, these values are directly derived from the projection results of the estimated model. It is not surprising that for annuities, the actuarial present values for female are higher than that for male, while the opposite is true for life insurances. In the next subsection, we will use these values to construct portfolios with different gender mixes.

4.3 The impact of gender-neutral pricing

We now examine the impact of gender-neutral pricing on the liability position of the life insurance and annuity portfolios. We assume that a gender-neutral premium is calculated using the gender-neutral life table for each of the six products under consideration. Then, for each product, we use the projection results shown in Figure 12 to form a portfolio using each of the assumed gender mixes. Lastly, the net liability is obtained from subtracting the present value of the portfolio's projected liabilities by the gender-neutral premium. Here, the present values are randomly simulated by our estimated model, while the gender-neutral premium is calculated using the model's mean forecast.

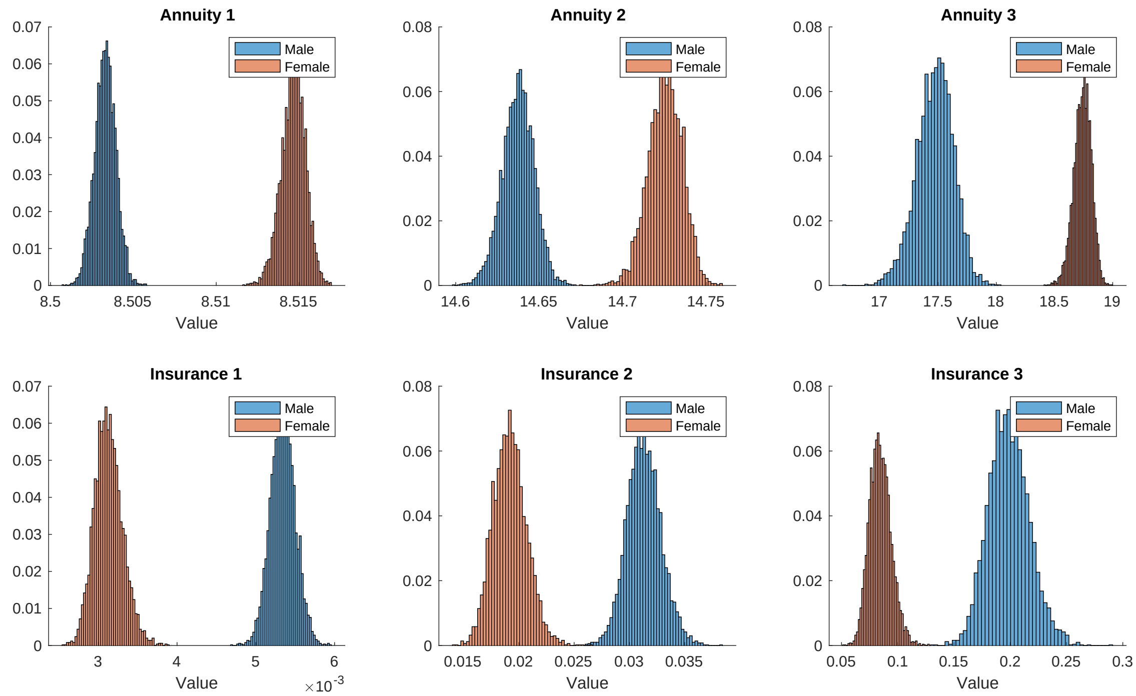

Figure 12: Empirical distributions of the present values of the six life products under consideration for both genders.

Figure 13 shows the empirical distribution of the net liability position (in percentage of the gender-neutral premium) for each of the three annuity products and three gender mixes. In Figure 13, the first, second and third rows correspond to gender mix 1, 2 and 3, respectively, while the first, second and third columns correspond to annuity 1, 2 and 3, respectively. Comparing the distributions in the first row with that in the second row, we see that when there are more females in the portfolio (Mix 1), the net liability position will be positive and centered above zero, indicating that the gender-neutral premium is not enough to cover all projected liabilities. The opposite is true when we compare the third row (Mix 3), where there are more males in the portfolio, with the second row (Mix 2). It is reasonable to observe these results because we observe from Figure 12 that female annuities have a higher present value of liabilities than male annuities. A gender-neutral premium is thus insufficient to cover liabilities when there are more females in the portfolio, and is excessive when there are more males in the portfolio.

Comparing the distributions among the three columns, we find that the conclusion made regarding gender-neutral pricing applies to all three annuities regardless the annuitant age, annuity term and gender disparity type. However, the magnitude of the change in net liability position is very small for annuity 1 (10-year term life annuity for age 30), while the magnitude for annuity 3 (30-year term life annuity for age 60) is at most 5%. We will further examine the impact of gender disparity in the next subsection.

Figure 14 shows the empirical distribution of the net liability position (in percentage of the gender-neutral premium) for each of the three insurance products and three gender mixes. The layout of Figure 14 is similar to Figure 13, where the first, second and third rows correspond to gender mix 1, 2 and 3, and the first, second and third columns correspond to annuity 1, 2 and 3, respectively. However, the conclusion

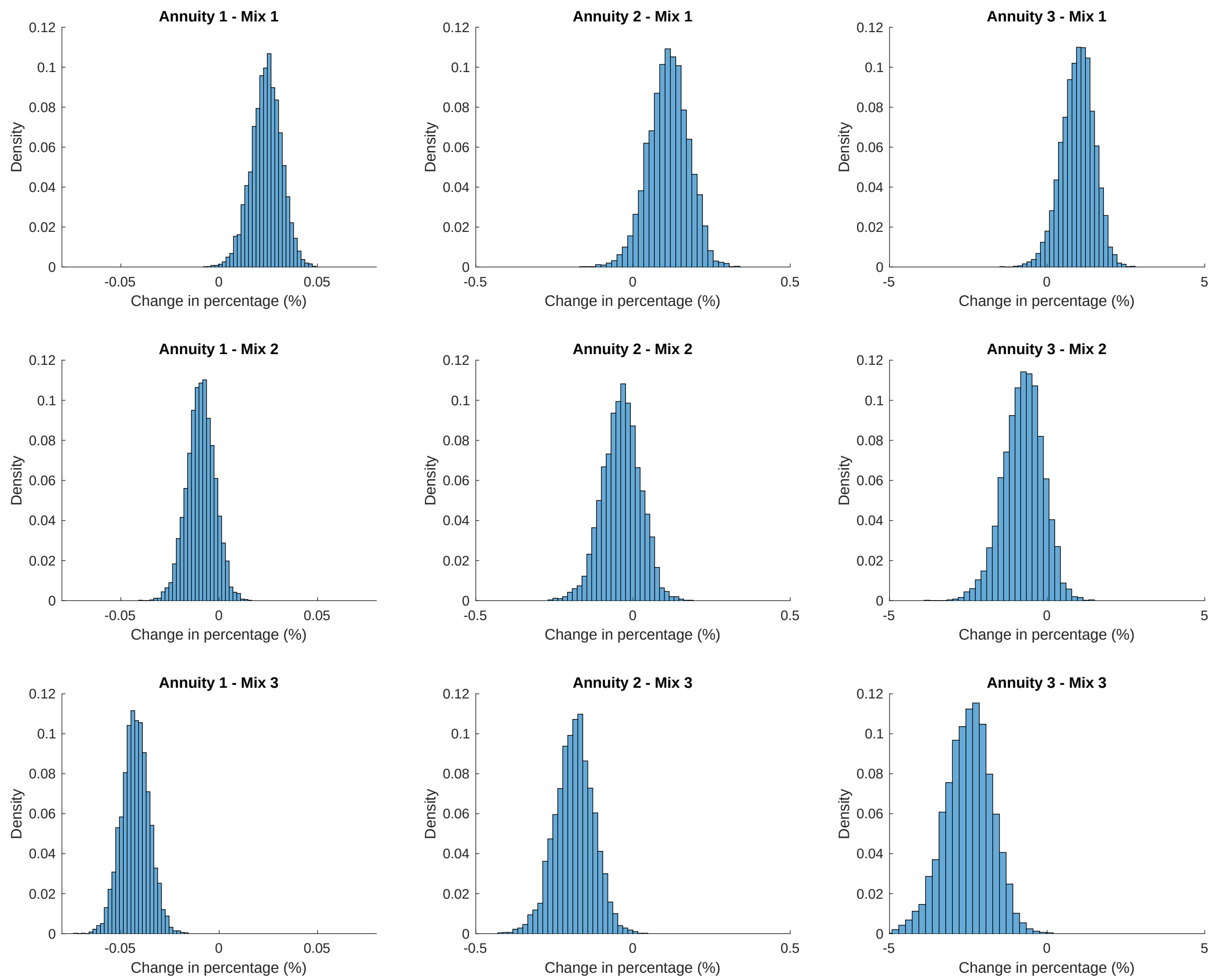

Figure 13: Empirical distribution of the net liabilities in percentage of the gender-neutral premium for nine different annuity portfolios.

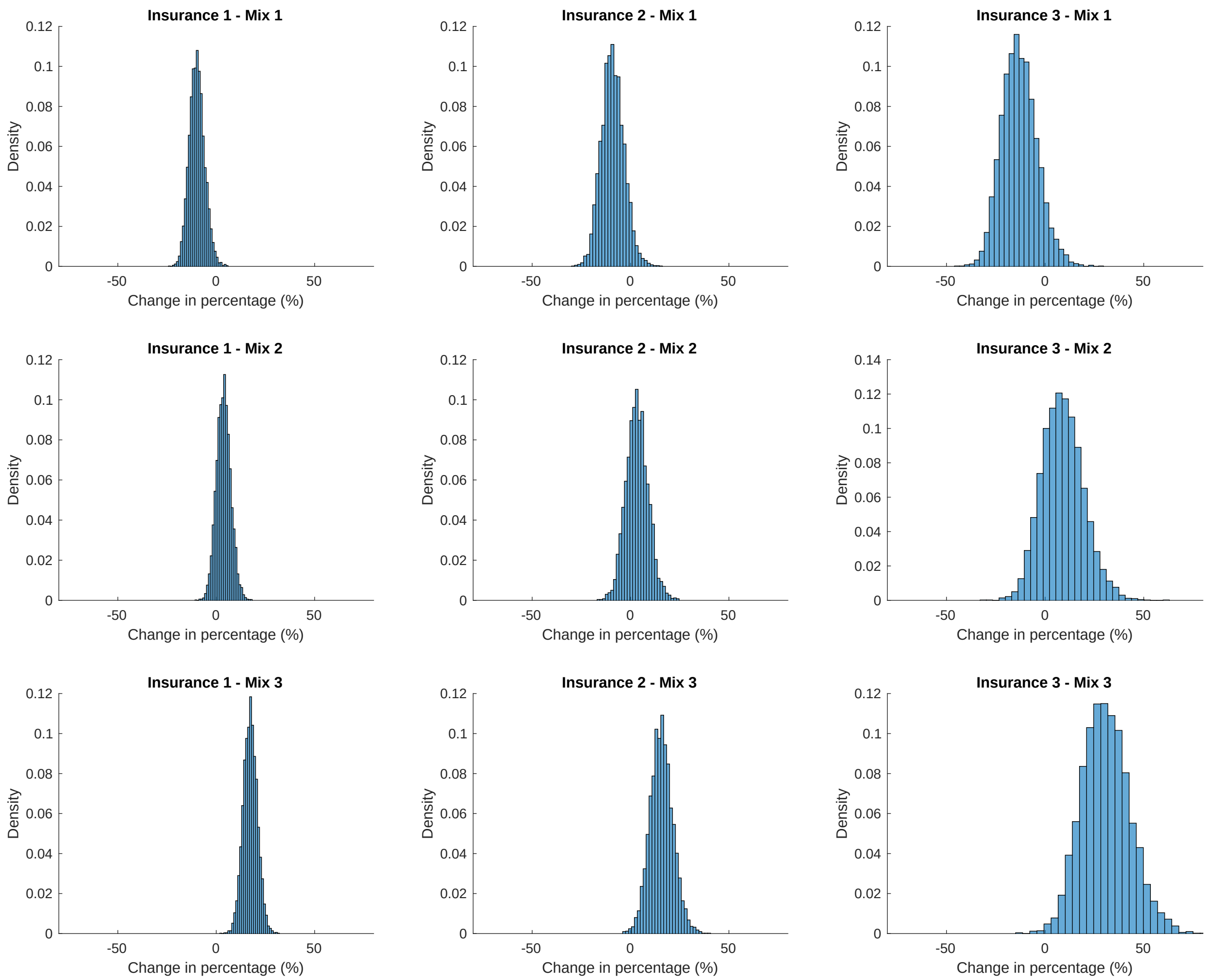

Figure 14: Empirical distribution of the net liabilities in percentage of the gender-neutral premium for nine different insurance portfolios.

regarding gender-neutral pricing on life insurance is reversed if compared with life annuity. Particularly, the gender-neutral premium is excessive for an insurance portfolio when there are more females, but not enough when there are more males in the portfolio. This result is expected because we know that life insurance issued to female has a lower present value of liabilities than the same product issued to male. The magnitude of the change in net liability position is also significant larger than that for annuities. We can thus say that insurance products are more sensitive to gender-neutral pricing than annuity products.

4.4 The impact of gender disparity

To examine the impact of gender disparity, we generate a new set of projections without using the gender-specific period effects in the estimated model. Without the gender-specific period effects, the resulting model is similar to the common factor model (Li and Lee, 2005), which is an 'all-age' coherent mortality model. We then compare the net liability positions between our model and the reduced model for all of the six products and three genders mixes. The results for insurance and annuity are reported in Figure 15 and Figure 16, respectively.

For life insurance, we observe in Figure 15 that the impact of gender disparity is more significant to insurance 3 (30-year term life insurance for age 60) than the other two insurance products. In particular, for insurance 3 and mix 3 (75% males and 25% females), not accounting for gender disparity (widening between genders for the underlying range group) will underestimate the effect of gender-neutral pricing on the liability position. Here, the gender-neutral premium undercharges the portfolio and the gender disparity amplifies the undercharging situation because the gender difference in mortality is widening over time. For insurance 2 and mix 3, where the underlying age group experiences diminishing gender disparity, the situation is reversed and not accounting for gender disparity will overestimate the effect of gender-neutral pricing. This is because the diminishing gender disparity will slightly alleviate the undercharging situation resulted from mix 3. For insurance 1, where the age group is experiencing stationary gender disparity, the impact of gender disparity is not very obvious. Lastly, for mix 1 (25% males and 75% females), the opposite results are observed because the gender-neutral premium overcharges the portfolios.

For life annuity, we observe in Figure 16 that the impact of gender disparity is not very significant to most annuities. For instance, under mix 1, although the gender-neutral premium undercharges the three portfolios, the gender disparity is not substantially changing the undercharging situation.

5 Conclusion

In this paper, we studied gender disparity in mortality for Asia-Pacific populations. We first visually examined the difference in historical mortality rates between the two genders for three Asia-Pacific populations. The empirical results revealed that the difference changes over time, and more importantly the rate of change varies among age groups and populations. To verify this observation, we developed a statistical method to test the significance of gender disparity in mortality. The test results suggest that existing stochastic mortality models for forecasting mortality of males and females simultaneously should be adapted for the phenomenon of gender disparity in mortality.

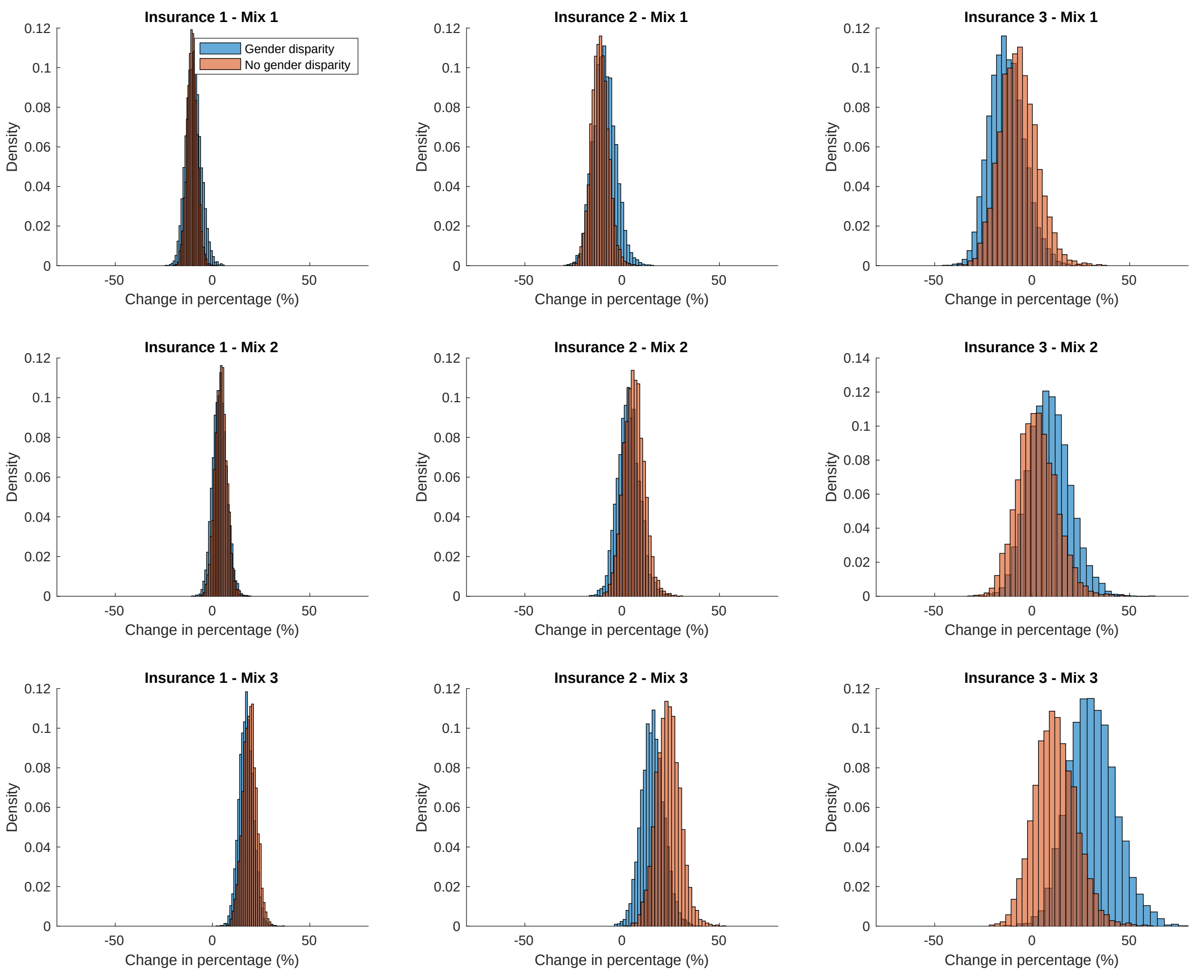

Figure 15: Empirical distribution of the net liabilities in percentage of the gender-neutral premium for nine different insurance portfolios under the proposed model (with gender disparity) and the reduced model (without gender disparity).

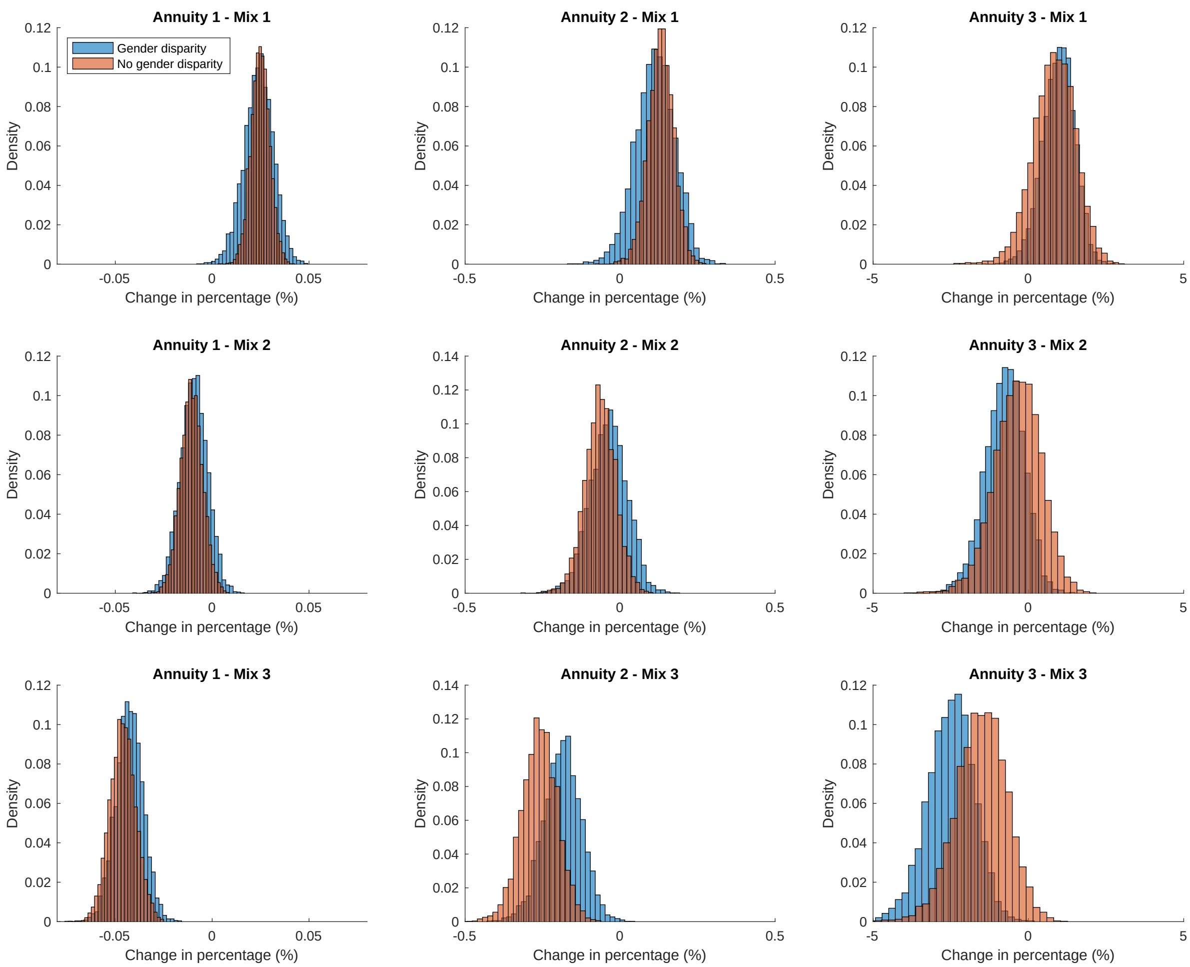

Figure 16: Empirical distribution of the net liabilities in percentage of the gender-neutral premium for nine different annuity portfolios under the proposed model (with gender disparity) and the reduced model (without gender disparity).

To capture the gender disparity in mortality modeling and forecasting, we constructed a modified version of the Li-Lee model. The proposed model allows the projected mortality rates for males and females to be diverging for some age groups but non-diverging for the others. We estimate the proposed model using the Bayesian method, which enable us to overcome some data problems in the mortality data of China. The Bayesian estimated model can also produce stochastic projections of future mortality rates of both genders simultaneously, while consider the age-specific gender disparity situation. We provided a detailed analysis of the estimation and projection results of our proposed model.

Based on the projection results of the proposed model, we further studied the impact of gender disparity in mortality on life insurance products. Specifically, we considered six different products with different age ranges, duration, gender disparity and gender mixes. Lastly, on the basis of a gender-neutral regulation scenario, we examined how the funding position of various life insurance portfolios is affected by gender disparity and gender-neutral pricing.

References

Gelman, A. and Rubin, D. B. (1992). Inference from iterative simulation using multiple sequences. *Statistical science*, 7(4):457–472.

Geweke, J. F. (1991). *Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments*, volume 148. Federal Reserve Bank of Minneapolis, Research Department Minneapolis, MN.

Koopman, S. J. and Durbin, J. (2000). Fast filtering and smoothing for multivariate state space models. *Journal of Time Series Analysis*, 21(3):281–296.

Li, J. S.-H., Zhou, K. Q., Zhu, X., Chan, W.-S., and Chan, F. W.-H. (2019). A bayesian approach to developing a stochastic mortality model for china. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 182(4):1523–1560.

Li, N. and Lee, R. (2005). Coherent mortality forecasts for a group of populations: An extension of the lee-carter method. *Demography*, 42(3):575–594.

Pedroza, C. (2006). A bayesian forecasting model: predicting us male mortality. *Biostatistics*, 7(4):530–550.

About The Society of Actuaries Research Institute

Serving as the research arm of the Society of Actuaries (SOA), the SOA Research Institute provides objective, data-driven research bringing together tried and true practices and future-focused approaches to address societal challenges and your business needs. The Institute provides trusted knowledge, extensive experience and new technologies to help effectively identify, predict and manage risks.

Representing the thousands of actuaries who help conduct critical research, the SOA Research Institute provides clarity and solutions on risks and societal challenges. The Institute connects actuaries, academics, employers, the insurance industry, regulators, research partners, foundations and research institutions, sponsors and non-governmental organizations, building an effective network which provides support, knowledge and expertise regarding the management of risk to benefit the industry and the public.

Managed by experienced actuaries and research experts from a broad range of industries, the SOA Research Institute creates, funds, develops and distributes research to elevate actuaries as leaders in measuring and managing risk. These efforts include studies, essay collections, webcasts, research papers, survey reports, and original research on topics impacting society.

Harnessing its peer-reviewed research, leading-edge technologies, new data tools and innovative practices, the Institute seeks to understand the underlying causes of risk and the possible outcomes. The Institute develops objective research spanning a variety of topics with its strategic research programs: aging and retirement; actuarial innovation and technology; mortality and longevity; diversity, equity and inclusion; health care cost trends; and catastrophe and climate risk. The Institute has a large volume of topical research available, including an expanding collection of international and market-specific research, experience studies, models and timely research.

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

Copyright © 2023 Society of Actuaries Research Institute