Grouped Multivariate and Functional Time Series Forecasting: An Application to Annuity Pricing

Han Lin Shang
Research School of Finance, Actuarial Studies and Statistics, Australian National University
Steven Haberman
Cass Business School, City, University of London

Presented at the Living to 100 Symposium
Orlando, Fla.
January 4–6, 2017

GROUPED MULTIVARIATE AND FUNCTIONAL TIME SERIES FORECASTING: AN APPLICATION TO ANNUITY PRICING

Han Lin Shang* and Steven Haberman

Abstract

Age-specific mortality rates are often disaggregated by different attributes, such as sex, state, ethnic group and socioeconomic status. In making social policies and pricing annuities at national and sub-national levels, not only is it important to forecast mortality accurately, but also, forecasts at sub-national levels should add up to the forecasts at the national level. This motivates recent developments of grouped functional time series methods (Shang, Han and Rob Hyndman., 2017) to reconcile age-specific mortality forecasts. We extend these grouped functional time series forecasting methods to multivariate time series and apply them to produce point forecasts of mortality rates at older ages, from which fixed-term annuities for different ages and maturities can be priced. Using the regional age-specific mortality rates in Japan obtained from the Japanese Mortality Database, we investigate the one-step-ahead to 15-step-ahead point forecast accuracy between the independent and grouped forecasting methods. The grouped forecasting methods are shown not only to be useful for reconciling forecasts of age-specific mortality rates at national and sub-national levels, but also to enjoy improved forecast accuracy. The improved forecast accuracy of mortality rates would be of great interest to the insurance and pension industries for estimating annuity prices, in particular at the level of population subgroups defined by key factors such as gender, region and socioeconomic grouping.

Keywords: forecast reconciliation; hierarchical time series; bottom-up method; optimal combination method; Lee-Carter method; Japanese Mortality Database.

*Corresponding author. Address: Research School of Finance, Actuarial Studies and Statistics, Level 4, Building 26C, Australian National University, Kingsley Street, Canberra, ACT 2601, Australia; Telephone: +61(2) 612 50535; Fax: +61(2) 612 50087; Email: hanlin.shang@anu.edu.au. The authors are grateful for the comments and suggestions received from the seminar participants at the Cass Business School, City, University of London and ESRC Centre for Population Change, University of Southampton. The research is funded by a research school grant from the Australian National University.

1 Introduction

In many developed countries such as Japan, increases in longevity and an aging population have led to concerns regarding the sustainability of pensions and health and aged-care systems (see, for example, Coulmas, Florian., 2007; Organisation for Economic Co-operation and Development, 2013). These concerns have resulted in a surge of interest among government policy makers and planners in accurately modeling and forecasting age-specific mortality rates. Any improvements in the forecast accuracy of mortality would be beneficial for annuity providers, corporate pension funds and governments (see, for example, Koissia, Marie-Claire., 2006; Denuit, Michel, Pierre Devolder and Anne-Cécile Goderniaux., 2007; Hanewald, Katja, Thomas Post and Helmut Gründl., 2011)—in particular, for determining age of retirements and allocating pension benefits at the national and sub-national levels.

Several authors have proposed new approaches for forecasting age-specific mortality at the national level using statistical models (see Booth, Heather., 2006; Booth, Heather and Leonie Tickle., 2008, for reviews). These models can be categorized into three main streams: explanation, expectation and extrapolation approaches. Among the extrapolation methods, a significant milestone in demographic forecasting was the work of Lee, Ronald and Lawrence Carter. (1992), which has since received widespread attention. This model has been extensively studied and widely used for forecasting mortality rates in various countries (see Shang, Han, Heather Booth and Rob Hyndman., 2011, and references herein).

The strengths of the Lee-Carter method are its simplicity and robustness in situations where age-specific log mortality rates have linear trends (Booth, Heather, Rob Hyndman, Leonie Tickle and Piet De Jong., 2006). The main weakness of the Lee-Carter method is that it attempts to capture the patterns of mortality rates using only one principal component and its scores. To rectify this deficiency, the Lee-Carter model has been extended and modified in several directions; see, for example, Brouhns, Natacha, Michel Denuit and Jeroen Vermunt. (2002), Renshaw, A and S Haberman. (2003), Currie, Iain, Maria Durban and Paul Eilers. (2004), Renshaw, A and S Haberman. (2006), Hyndman and Ullah (2007) and Pitacco, Ermanno, Michel Denuit, Steven Haberman and Annamaria Olivieri. (2009).

Although mortality forecasts at the national level are comparably accurate, mortality forecasts at the sub-national level often suffer from relatively poor data quality and/or missing data. However, sub-national forecasts of age-specific mortality are valuable for informing policy within local regions and allow us to appreciate the heterogeneity in the population and understand individuals' differences. If one can better understand individual characteristics, an assurer can better price an annuity for annuitants.

In insurance and pension companies, one is typically interested in forecasting age-specific mortality for multiple subpopulations that often obey a hierarchical (unique) or group (non-unique) structure. Let us consider a simple group structure, where total age-specific mortality rates can be disaggregated by sex. If we forecast female, male and total age-specific mortality independently, the forecast female and male mortality may not add up to the forecast total mortality. This is known as the problem of forecast reconciliation, which has been considered in economics for balancing national accounts (see Stone, Richard, D Champernowne and J Meade., 1942, for an example), in statistics for forecasting tourism demand (Hyndman, Rob, Roman Ahmed, George Athanasopoulos and Han Shang., 2011), and in demography for forecasting age-specific mortality rates (Shang, Han., 2017; Shang, Han and Rob Hyndman., 2017). To the best of our knowledge, forecast reconciliation of age-specific mortality has not been considered in actuarial studies to date, and it is our goal to fill this methodological gap.

As two forecasting techniques, we apply the Lee-Carter method and functional time series method of Hyndman and Ullah (2007) to a large set of multivariate or functional time series with rich structure, respectively. We put forward two statistical methods—namely, bottom-up and optimal combination methods—to reconcile point forecasts of age-specific mortality and potentially improve the point forecast accuracy. In turn, this may lead to more accurate forecasts of mortality and conditional life expectancy, and thus better estimates of annuity prices. The "bottom-up" method involves forecasting each of the disaggregated series and then using simple aggregation to obtain forecasts for the aggregated series (Kahn, Kenneth., 1998). This method works well when the bottom-level series have a high signal-to-noise ratio. For highly disaggregated series, this does not work well, as the series become too noisy. This motivated the development of an optimal combination method (Hyndman, Rob, Roman Ahmed, George Athanasopoulos and Han Shang., 2011), where forecasts are obtained independently for all series at all levels of disaggregation and then a linear regression is used with an ordinary least-squares or a generalized least-squares estimator to optimally combine and reconcile these forecasts.

Using the national and sub-national Japanese age-specific mortality rates from 1975 to 2013, we compare the point forecast accuracy among the independent (base) forecasting, bottom-up and optimal combination methods. The independent forecasts can be produced from any univariate or multivariate time series forecasting method, such as exponential smoothing and autoregressive integrated moving average. These independent forecasts are generally not reconciled according to the group structure. To evaluate the point forecast accuracy, we consider the mean absolute forecast and root mean squared forecast errors, and found that the bottom-up method performs the best among these three methods in our data set.

The rest of this paper is structured as follows: In Section 2, we describe the motivating data set, which is Japanese national and sub-national age-specific mortality rates. In Section 3, we briefly revisit the Lee-Carter and functional time series methods for producing point forecasts, then introduce two grouped forecasting methods in Section 4. Using the forecast error criteria in Section 5.2, we first evaluate and compare point forecast accuracy between the Lee-Carter and functional time series methods, and then between independent and grouped forecasting methods in Section 5.3. In Section 6, we apply the independent and grouped forecasting methods to estimate the fixed-term annuity prices for different ages and maturities. Conclusions are presented in Section 7, along with some reflections on how the methods presented here can be further extended.

2 Data

We study Japanese age-specific mortality rates from 1975 to 2013, obtained from the Japanese Mortality Database (Japanese Mortality Database, 2016). Since our focus is on actuarial applications, we consider ages from 60 to 99 in a single year of age, and the last age group is the age at and beyond 100. The structure of the data is displayed in Table 1, where each row denotes a level of disaggregation.

Table 1. *Hierarchy of Japanese Mortality Rates*

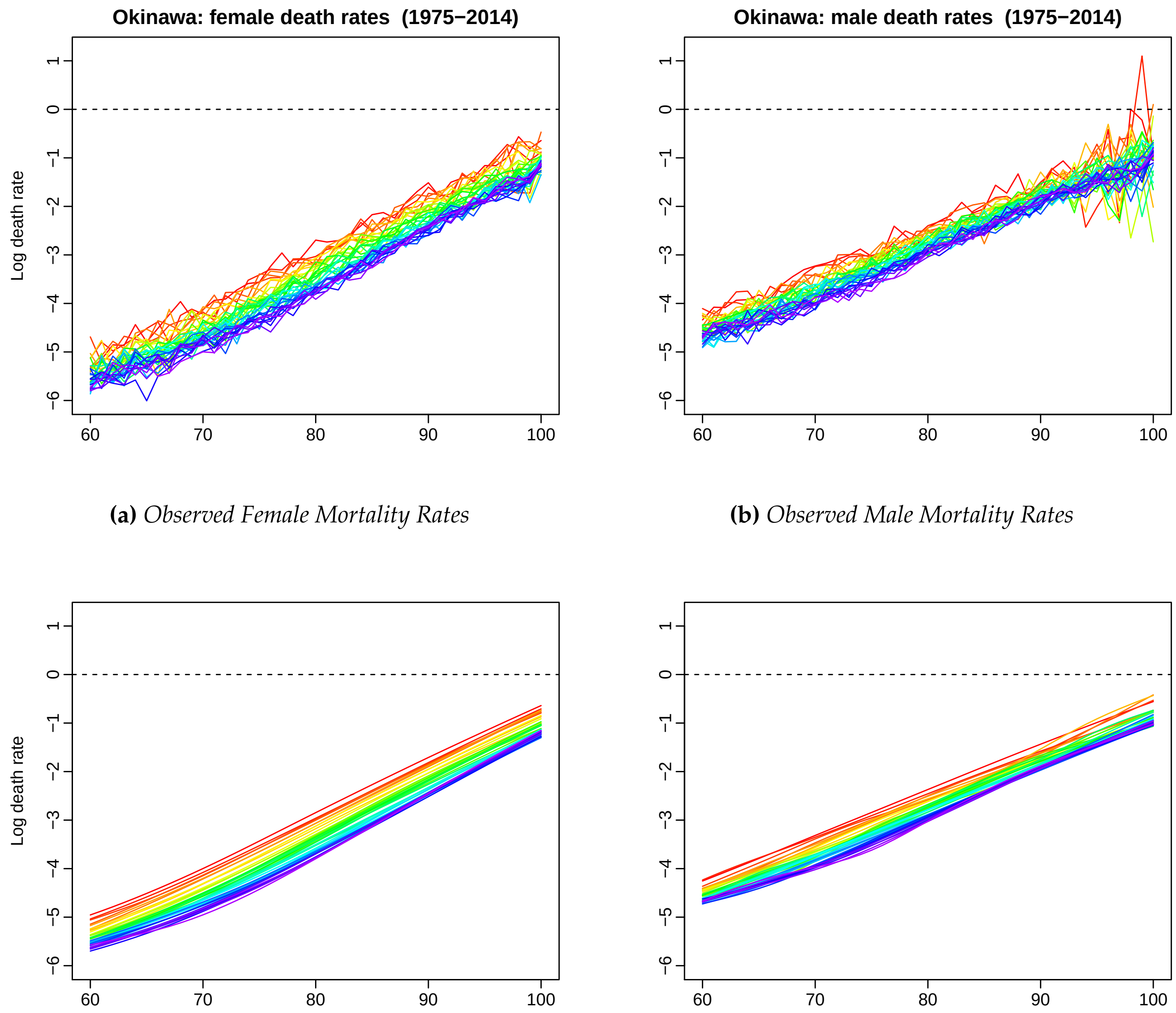
Group Level	Number of Series
Japan	1
Sex	2
Region	8
Region × Sex	16
Prefecture	47
Prefecture × Sex	94
Total	168

At the top level, we have total age-specific mortality rates for Japan. We can split these total mortality rates by sex, by region or by prefecture. Japan is divided into eight regions, which contain a total of 47 prefectures. The most disaggregated data arise when we consider the mortality rates for each combination of prefecture and sex, giving a total of $47 \times 2 = 94$ series. In total, across all levels of disaggregation, there are 168 series.

2.1 Rainbow Plots

Figure 1 shows rainbow plots of the female and male age-specific log mortality rates in prefecture Okinawa from 1975 to 2013. The time ordering of the curves follows the color order of a rainbow, where curves from the distant past are shown in red and the more recent curves are shown in purple (Hyndman, Rob and Han Shang., 2010). The figures show typical age-specific mortality curves with gradually increasing mortality rates as age increases.

From Figures 1a and 1b, the observed mortality rates are not smooth across age. Due to observational noise, male mortality rates in some years are above 1 (when log mortality rates are above 0). To obtain smooth functions and deal with possible missing values, we consider a penalized regression spline smoothing with monotonic constraint, described in Section 3.2. It incorporates the shape of log mortality curves (see also Hyndman and Ullah, 2007; D'Amato, Valeria, Gabriella Piscopo and Maria Russolillo., 2011).



(a) *Observed Female Mortality Rates*

(b) *Observed Male Mortality Rates*

(c) *Smoothed Female Mortality Rates*

(d) *Smoothed Male Mortality Rates*

Fig. 1. ***Functional Time Series Graphical Displays***

Figures 1c and 1d display smooth age-specific mortality rates for Okinawa females and males, but we apply smoothing to all series at different levels of disaggregation. We have developed a Shiny app (Chang, Winston, Joe Cheng, J. Allaire, Yihui. Xie and Jonathan McPherson., 2017) in R (R Core Team, 2017) to allow interactive exploration of the smoothing of all the data series; this is available in the online supplementary material.

2.2 Image Plots

Another visual perspective of the data is the image plot of Shang, Han and Rob Hyndman. (2017). In Figure 2, we plot the log of the ratio of mortality rates for each prefecture to mortality rates for Japan, as this facilitates relative mortality comparison. A divergent color palette is used, with blue representing positive values and orange denoting negative values. The prefectures are ordered geographically from north (Hokkaido) to south (Okinawa).

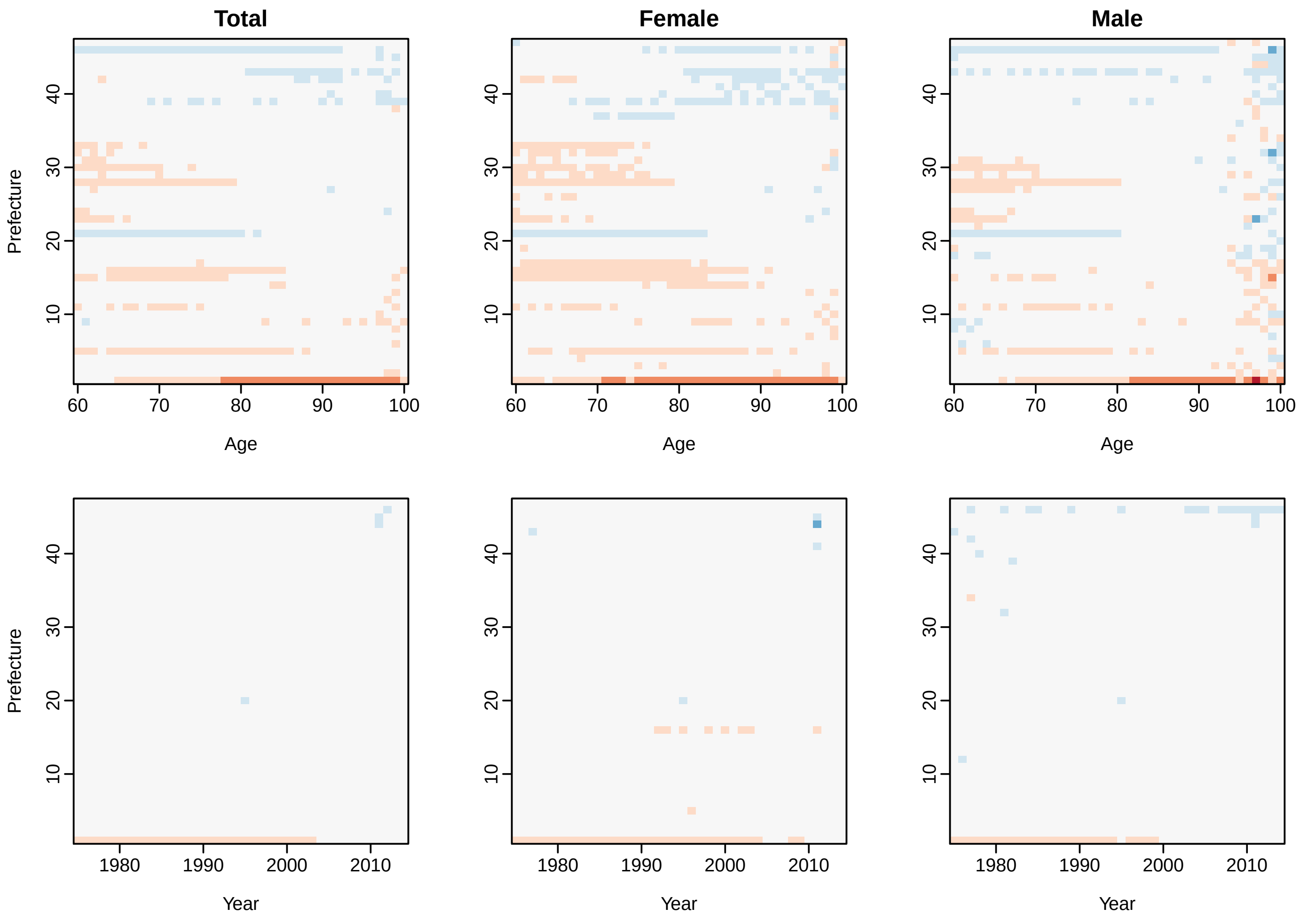


Fig. 2. ***Image Plots Showing Log Ratios of Mortality Rates***

Note: The top panels show mortality rates averaged over years, while the bottom panels show mortality rates averaged over ages.

The top row of panels shows mortality rates for each prefecture and age, averaged over years. There are strong differences between the prefectures for the elderly; this is possibly due to differences in socioeconomic status and accessibility of health services. The most southerly prefecture of Okinawa has very low mortality rates and thus extreme longevity for the elderly (see, for example, Takata, Hajime, Toshiharu Ishii, Makoto Suzuki, Susumu Sekiguchi and Hisami Iri., 1987; Suzuki, Makoto, Bradley Willcox and Craig Willcox., 2004; Willcox, Craig, Bradley J Willcox, Sanae Shimajiri, Sayuri Kurechi and Makoto Suzuki., 2007).

The bottom row of panels shows mortality rates for each prefecture and year, averaged over all ages. We found three abnormalities. In 2011, in prefecture 44 (Miyagi) and 45 (Iwate), there was an abnormally large increase in mortality compared with other prefectures. These are northern coastal regions, and the inflated relative mortality is due to the tsunami of March 11, 2011 (Shang, Han and Rob Hyndman., 2017). In 1995, there was an abnormal increase in mortality for prefecture 20 (Hyōgo), which corresponds to the Kobe (Great Hanshin) earthquake of Jan. 17, 1995. In prefecture Okinawa, the residents enjoyed relatively low mortality rates until 2000 and even beyond, especially for females; in recent years, the comparably lower mortality rates have become less evident.

3 Forecasting Methods

We revisit the Lee-Carter and functional time series models for forecasting age-specific mortality that are compared in the present study. The Lee-Carter model considers age as a discrete variable (see, for example, Li, Nan and Ronald Lee., 2005), while the functional time series model treats age as a continuous variable (see, for example, D'Amato, Valeria, Gabriella Piscopo and Maria Russolillo., 2011; Shang, Han., 2016). To stabilize the high variance associated with high age-specific mortality rates, it is necessary to transform the raw data by taking the natural logarithm. We denote by $m_{x,t}$ the observed mortality rate at age x in year t, calculated as the number of deaths aged x in calendar year t, divided by the corresponding midyear population aged x. The models are all expressed in log scale.

3.1 The Lee-Carter (LC) Method

The original formulation of the Lee-Carter (LC) model is given by

$$\ln(m_{x,t}) = a_x + b_x \kappa_t + \varepsilon_{x,t}, \tag{1}$$

where a_x is the age pattern of the log mortality rates averaged across years; b_x is the first principal component reflecting relative change in the log mortality rate at each age; κ_t is the first set of principal component scores at year t and measures the general level of the log mortality rates; and $\varepsilon_{x,t}$ is the residual at age x and year t.

The LC model in (1) is over-parameterized in the sense that the model structure is invariant under the following transformations:

$$\{a_x, b_x, \kappa_t\} \mapsto \{a_x, b_x/c, c\kappa_t\},$$
$$\{a_x, b_x, \kappa_t\} \mapsto \{a_x - cb_x, b_x, \kappa_t + c\}$$

To ensure the model identifiability, Lee, Ronald and Lawrence Carter. (1992) imposed two constraints, given as

$$\sum_{t=1}^{n} \kappa_t = 0, \qquad \sum_{x=x_1}^{x_p} b_x = 1,$$

where n denotes the number of years and p denotes the number of ages in the observed data set.

The LC method adjusts κ_t by refitting the total number of deaths. The adjustment gives more weight to high rates (Shang, Han, Heather Booth and Rob Hyndman., 2011). The adjusted κ_t is then extrapolated using autoregressive integrated moving average (ARIMA) models. Lee, Ronald and Lawrence Carter. (1992) used a random walk with drift model, which can be expressed as

$$\kappa_t = \kappa_{t-1} + d + e_t,$$

where d is known as the drift parameter and measures the average annual change in the series, and e_t is an uncorrelated error. Based on the forecast of principal component scores, the forecast age-specific log mortality rates are obtained using the estimated mean function $\widehat{a}_x$ and estimated first principal component $\widehat{b}_x$ in (1).

3.2 A Functional Time Series Method

The Lee-Carter model considers age as a discrete variable, while the functional time series model treats age as a continuous variable. One advantage of the functional time series model is that a nonparametric smoothing technique can be incorporated into the modeling procedure, in order to obtain smoothed principal components. Among many possible nonparametric smoothing techniques, we use penalized regression spline with a partial monotonic constraint, where the smoothed log mortality rates can be expressed as

$$m_t(x_i) = f_t(x_i) + \sigma_t(x_i)\varepsilon_{t,i}, \qquad i = 1, \dots, p, \quad t = 1, \dots, n,$$

where $m_t(x_i)$ denotes the log of the observed mortality rate for age x_i in year t; $\sigma_t(x_i)$ allows the amount of noise to vary with x_i in year t, and $\varepsilon_{t,i}$ is an independent and identically distributed standard normal random variable.

The smoothed log mortality curves $\boldsymbol{f}(x) = \{f_1(x), \dots, f_n(x)\}$ are treated as realizations of a stochastic process. Using functional principal component analysis, these smoothed log mortality curves are decomposed into

$$f_t(x) = a(x) + \sum_{j=1}^{J} b_j(x)k_{t,j} + e_t(x), \tag{2}$$

where $a(x)$ denotes the mean function, $\{b_1(x), \dots, b_J(x)\}$ denotes a set of functional principal components, $\{k_{t,1}, \dots, k_{t,J}\}$ denotes a set of principal component scores in year t, $e_t(x)$ is the error function with mean 0, and $J < n$ is the number of principal components retained. Decomposition (2) facilitates dimension reduction, as the first J terms often provide a reasonable approximation to the infinite sums, and thus the information contained in $\boldsymbol{f}(x)$ can be adequately summarized by the J-dimensional vector $(\boldsymbol{b}_1, \dots, \boldsymbol{b}_J)$. In contrast to the Lee-Carter model, the other advantage of the functional time series model is that more than one component may be used.

Conditioning on the observed data $\mathcal{I} = \{m_1(x), \dots, m_n(x)\}$ and the set of functional principal components $\boldsymbol{B} = \{b_1(x), \dots, b_J(x)\}$, the h-step-ahead forecast of $m_{n+h}(x)$ can be obtained by

$$\begin{aligned}\widehat{m}_{n+h|n}(x) &= \text{E}[m_{n+h}(x)|\mathcal{I}, \boldsymbol{B}] \\ &= \widehat{a}(x) + \sum_{j=1}^{J} b_j(x)\widehat{k}_{n+h|n,j},\end{aligned}$$

where $\widehat{k}_{n+h|n,j}$ denotes the h-step-ahead forecast of $k_{n+h,j}$ using a univariate or multivariate time series model (see Hyndman, Rob and Han Shang., 2009; Aue, Alexander, Diogo Norinho and Siegfried. Hörmann., 2015, for more details). Here, we consider a univariate time series forecasting method and implement the automatic algorithm of Hyndman, Rob and Yeasmin Khandakar. (2008) for selecting optimal orders in the ARIMA model.

To select J, we determine the value of J as the minimum number of components that reaches a certain level of the proportion of total variance explained by the J leading components such that

$$J = \underset{J:J\geq 1}{\text{argmin}} \left\{ \sum_{j=1}^{J} \widehat{\lambda}_j \Big/ \sum_{j=1}^{\infty} \widehat{\lambda}_j \mathbb{1}\{\widehat{\lambda}_j > 0\} \geq \delta \right\},$$

where $\delta = 99\%$ and $\mathbb{1}\{\cdot\}$ denotes the binary indicator function, which excludes possible 0 eigenvalues.

4 Grouped Forecasting Methods

For ease of explanation, we will introduce the notation using the Japanese example given in Section 2. The Japanese data follow a three-level geographical hierarchy coupled with a sex grouping variable. The geographical hierarchy is shown in Figure 3. Japan can be split into eight regions from north to south, which in turn can be split into 47 prefectures.

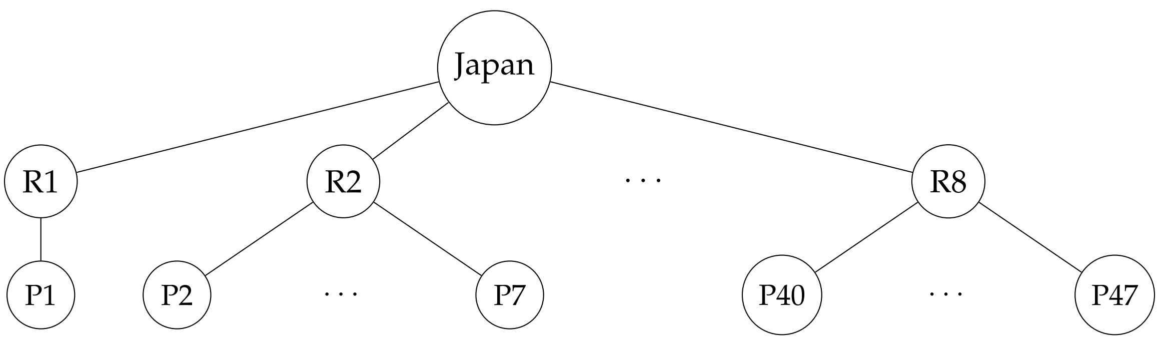


Fig. 3. *The Japanese Geographical Hierarchy Tree Diagram, With Eight Regions and 47 Prefectures*

The data can also be split by sex. Each of the nodes in the geographical hierarchy can also be split into males and females. We refer to a particular disaggregated series using the notation X $*$ S, meaning the geographical area X and the sex S, where X can take the values shown in Figure 3 and S can take values M (males), F (females) or T (total). For example, R1 $*$ F denotes females in Region 1; P1 $*$ T denotes all females and males in Prefecture 1; and Japan $*$ M denotes all males in Japan.

Denote $\text{E}_{\text{X}*\text{S},t}(x)$ as the exposure-to-risk for series $\text{X}*\text{S}$ in year t and age x, and let $\text{D}_{\text{X}*\text{S},t}(x)$ be the number of deaths for series $\text{X}*\text{S}$ in year t and age x. Then the age-specific mortality rate is given by $\text{R}_{\text{X}*\text{S},t}(x) = \text{D}_{\text{X}*\text{S},t}(x)/\text{E}_{\text{X}*\text{S},t}(x)$. To simplify expressions, we will drop the age argument (x). Then, for a given age, we can write

$$
\underbrace{\begin{bmatrix}
\text{R}_{\text{Japan*T},t} \\
\text{R}_{\text{Japan*F},t} \\
\text{R}_{\text{Japan*M},t} \\
\text{R}_{\text{R1*T},t} \\
\text{R}_{\text{R2*T},t} \\
\vdots \\
\text{R}_{\text{R8*T},t} \\
\text{R}_{\text{R1*F},t} \\
\text{R}_{\text{R2*F},t} \\
\vdots \\
\text{R}_{\text{R8*F},t} \\
\text{R}_{\text{R1*M},t} \\
\text{R}_{\text{R2*M},t} \\
\vdots \\
\text{R}_{\text{R8*M},t} \\
\text{R}_{\text{P1*T},t} \\
\text{R}_{\text{P2*T},t} \\
\vdots \\
\text{R}_{\text{P47*T},t} \\
\text{R}_{\text{P1*F},t} \\
\text{R}_{\text{P1*M},t} \\
\text{R}_{\text{P2*F},t} \\
\text{R}_{\text{P2*M},t} \\
\vdots \\
\text{R}_{\text{P47*F},t} \\
\text{R}_{\text{P47*M},t}
\end{bmatrix}}_{\boldsymbol{R}_t}
=
\underbrace{\begin{bmatrix}
\frac{\text{E}_{\text{P1*F},t}}{\text{E}_{\text{Japan*T},t}} & \frac{\text{E}_{\text{P1*M},t}}{\text{E}_{\text{Japan*T},t}} & \frac{\text{E}_{\text{P2*F},t}}{\text{E}_{\text{Japan*T},t}} & \frac{\text{E}_{\text{P2*M},t}}{\text{E}_{\text{Japan*T},t}} & \frac{\text{E}_{\text{P3*F},t}}{\text{E}_{\text{Japan*T},t}} & \frac{\text{E}_{\text{P3*M},t}}{\text{E}_{\text{Japan*T},t}} & \cdots & \frac{\text{E}_{\text{P47*F},t}}{\text{E}_{\text{Japan*T},t}} & \frac{\text{E}_{\text{P47*M},t}}{\text{E}_{\text{Japan*T},t}} \\
\frac{\text{E}_{\text{P1*F},t}}{\text{E}_{\text{Japan*F},t}} & 0 & \frac{\text{E}_{\text{P2*F},t}}{\text{E}_{\text{Japan*F},t}} & 0 & \frac{\text{E}_{\text{P3*F},t}}{\text{E}_{\text{Japan*F},t}} & 0 & \cdots & \frac{\text{E}_{\text{P47*F},t}}{\text{E}_{\text{Japan*F},t}} & 0 \\
0 & \frac{\text{E}_{\text{P1*M},t}}{\text{E}_{\text{Japan*M},t}} & 0 & \frac{\text{E}_{\text{P2*M},t}}{\text{E}_{\text{Japan*M},t}} & 0 & \frac{\text{E}_{\text{P3*M},t}}{\text{E}_{\text{Japan*M},t}} & \cdots & 0 & \frac{\text{E}_{\text{P47*M},t}}{\text{E}_{\text{Japan*M},t}} \\
\frac{\text{E}_{\text{P1*F},t}}{\text{E}_{\text{R1,T},t}} & \frac{\text{E}_{\text{P1*M},t}}{\text{E}_{\text{R1,T},t}} & 0 & 0 & 0 & 0 & \cdots & 0 & 0 \\
0 & 0 & \frac{\text{E}_{\text{P2*F},t}}{\text{E}_{\text{R2,T},t}} & \frac{\text{E}_{\text{P2*M},t}}{\text{E}_{\text{R2,T},t}} & \frac{\text{E}_{\text{P3*F},t}}{\text{E}_{\text{R2,T},t}} & \frac{\text{E}_{\text{P3*M},t}}{\text{E}_{\text{R2,T},t}} & \cdots & 0 & 0 \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \cdots & \vdots & \vdots \\
0 & 0 & 0 & 0 & 0 & 0 & \cdots & \frac{\text{E}_{\text{P47*F},t}}{\text{E}_{\text{R8,T},t}} & \frac{\text{E}_{\text{P47*M},t}}{\text{E}_{\text{R8,T},t}} \\
\frac{\text{E}_{\text{P1*F},t}}{\text{E}_{\text{R1,F},t}} & 0 & 0 & 0 & 0 & 0 & \cdots & 0 & 0 \\
0 & 0 & \frac{\text{E}_{\text{P2*F},t}}{\text{E}_{\text{R2,F},t}} & 0 & \frac{\text{E}_{\text{P3*F},t}}{\text{E}_{\text{R2,F},t}} & 0 & \cdots & 0 & 0 \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \cdots & \vdots & \vdots \\
0 & 0 & 0 & 0 & 0 & 0 & \cdots & \frac{\text{E}_{\text{P47*F},t}}{\text{E}_{\text{R8,F},t}} & 0 \\
0 & \frac{\text{E}_{\text{P1*M},t}}{\text{E}_{\text{R1,M},t}} & 0 & 0 & 0 & 0 & \cdots & 0 & 0 \\
0 & 0 & 0 & \frac{\text{E}_{\text{P2*M},t}}{\text{E}_{\text{R2,M},t}} & 0 & \frac{\text{E}_{\text{P3*M},t}}{\text{E}_{\text{R2,M},t}} & \cdots & 0 & 0 \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \cdots & \vdots & \vdots \\
0 & 0 & 0 & 0 & 0 & 0 & \cdots & 0 & \frac{\text{E}_{\text{P47*M},t}}{\text{E}_{\text{R8,M},t}} \\
\frac{\text{E}_{\text{P1*F},t}}{\text{E}_{\text{P1,T},t}} & \frac{\text{E}_{\text{P1*M},t}}{\text{E}_{\text{P1,T},t}} & 0 & 0 & 0 & 0 & \cdots & 0 & 0 \\
0 & 0 & \frac{\text{E}_{\text{P2*F},t}}{\text{E}_{\text{P2,T},t}} & \frac{\text{E}_{\text{P2*M},t}}{\text{E}_{\text{P2,T},t}} & 0 & 0 & \cdots & 0 & 0 \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \cdots & \vdots & \vdots \\
0 & 0 & 0 & 0 & 0 & 0 & \cdots & \frac{\text{E}_{\text{P47*F},t}}{\text{E}_{\text{P47,T},t}} & \frac{\text{E}_{\text{P47*M},t}}{\text{E}_{\text{P47,T},t}} \\
1 & 0 & 0 & 0 & 0 & 0 & \cdots & 0 & 0 \\
0 & 1 & 0 & 0 & 0 & 0 & \cdots & 0 & 0 \\
0 & 0 & 1 & 0 & 0 & 0 & \cdots & 0 & 0 \\
0 & 0 & 0 & 1 & 0 & 0 & \cdots & 0 & 0 \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \cdots & \vdots & \vdots \\
0 & 0 & 0 & 0 & 0 & 0 & \cdots & 1 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & \cdots & 0 & 1
\end{bmatrix}}_{\boldsymbol{S}_t}
\underbrace{\begin{bmatrix}
\text{R}_{\text{P1*F},t} \\
\text{R}_{\text{P1*M},t} \\
\text{R}_{\text{P2*F},t} \\
\text{R}_{\text{P2*M},t} \\
\vdots \\
\text{R}_{\text{P47*F},t} \\
\text{R}_{\text{P47*M},t}
\end{bmatrix}}_{\boldsymbol{b}_t}
$$

or $\boldsymbol{R}_t = \boldsymbol{S}_t\boldsymbol{b}_t$, where $\boldsymbol{R}_t$ is a vector containing all series at all levels of disaggregation, $\boldsymbol{b}_t$ is a vector of the most disaggregated series, and $\boldsymbol{S}_t$ shows how the two are connected.

Hyndman, Rob, Roman Ahmed, George Athanasopoulos and Han Shang. (2011) considered four hierarchical forecasting methods for univariate time series—namely, the top-down, bottom-up, middle-out and optimal combination methods. Among the four, the top-down and middle-out methods rely on a unique hierarchy for assigning disaggregation weights from a higher-level series to a lower-level series. In contrast, the bottom-up and optimal combination methods are suitable for forecasting a non-unique group structure. These two methods are reviewed in Sections 4.1 and 4.2; their point forecast accuracy comparison with the independent forecasting method are presented in

Section 5.3.

4.1 Bottom-Up Method

As the simplest grouped forecasting method, the bottom-up method first generates independent forecasts for each series at the most disaggregated level, and then aggregates these to produce all required forecasts. For example, let us revert to the Japanese data. We first generate h-step-ahead independent forecasts for the most disaggregated series, namely, $\widehat{\boldsymbol{b}}_{n+h} = \left[\widehat{\mathrm{R}}_{\mathrm{P1}*\mathrm{F},n+h}, \widehat{\mathrm{R}}_{\mathrm{P1}*\mathrm{M},n+h}, \ldots, \widehat{\mathrm{R}}_{\mathrm{P47}*\mathrm{F},n+h}, \widehat{\mathrm{R}}_{\mathrm{P47}*\mathrm{M},n+h}\right]$

The observed ratios that form the $\boldsymbol{S}_t$ summing matrix are forecast using an automatic ARIMA algorithm of Hyndman, Rob and Yeasmin Khandakar. (2008), when age $x = 60$. For age above 60, we assume the exposure-to-risk of age $x+1$ in year $t+1$ will be the same as the exposure-to-risk of age x in year t. For example, let $p_t = \mathrm{E}_{\mathrm{P1}*\mathrm{F},t}/\mathrm{E}_{\mathrm{Japan}*\mathrm{T},t}$ be a nonzero element of $\boldsymbol{S}_t$. Given that we observe $\{p_1, \ldots, p_n\}$, an h-step-ahead forecast $\widehat{p}_{n+h}$ can be obtained. These are then used to form the matrix $\boldsymbol{S}_{n+h}$. Thus, we obtain forecasts for all series as

$$\overline{\boldsymbol{R}}_{n+h} = \boldsymbol{S}_{n+h}\widehat{\boldsymbol{b}}_{n+h},$$

where $\overline{\boldsymbol{R}}_{n+h}$ denotes reconciled forecasts.

The bottom-up method performs well when the bottom-level series have a strong signal-to-noise ratio. However, it may lead to inaccurate forecasts of the top-level series, particularly when there are missing or noisy data at the bottom level.

4.2 Optimal Combination Method

Instead of considering only the bottom-level series, Hyndman, Rob, Roman Ahmed, George Athanasopoulos and Han Shang. (2011) proposed the optimal combination method in which independent forecasts for all series are computed independently and then the resultant forecasts are reconciled so that they satisfy the aggregation constraints via the summing matrix. The optimal combination method combines the independent forecasts through linear regression by generating a set of revised forecasts that are as close as possible to the independent forecasts but that also aggregate consistently within the group. The method is derived by expressing the independent forecasts as the response variable of the linear regression

$$\widehat{\boldsymbol{R}}_{n+h} = \boldsymbol{S}_{n+h}\boldsymbol{\beta}_{n+h} + \boldsymbol{\varepsilon}_{n+h},$$

where $\widehat{\boldsymbol{R}}_{n+h}$ is a matrix of h-step-ahead independent forecasts for all series, stacked in the same order as for the original data; $\boldsymbol{\beta}_{n+h} = \mathrm{E}[\boldsymbol{b}_{n+h}|\boldsymbol{R}_1, \ldots, \boldsymbol{R}_n]$ is the unknown mean of the independent forecasts of the most disaggregated series; and $\boldsymbol{\varepsilon}_{n+h}$ represents the reconciliation errors.

To estimate the regression coefficient, Hyndman, Rob, Roman Ahmed, George Athanasopoulos and Han Shang. (2011) and Hyndman, Rob, Alan Lee and Earo Wang. (2016) proposed a weighted least squares solution:

$$\widehat{\boldsymbol{\beta}}_{n+h} = \left(\boldsymbol{S}_{n+h}^{\top}\boldsymbol{W}_h^{-1}\boldsymbol{S}_{n+h}\right)^{-1}\boldsymbol{S}_{n+h}^{\top}\boldsymbol{W}_h^{-1}\widehat{\boldsymbol{R}}_{n+h},$$

where $\boldsymbol{W}_h$ is a diagonal matrix. Assuming that $\boldsymbol{W}_h = k_h\boldsymbol{I}$ and $\boldsymbol{I}$ denotes an identical matrix, the revised forecasts are given by

$$\overline{\boldsymbol{R}}_{n+h} = \boldsymbol{S}_{n+h}\widehat{\boldsymbol{\beta}}_{n+h} = \boldsymbol{S}_{n+h}\left(\boldsymbol{S}_{n+h}^{\top}\boldsymbol{S}_{n+h}\right)^{-1}\boldsymbol{S}_{n+h}^{\top}\widehat{\boldsymbol{R}}_{n+h},$$

where k_h is a constant. These reconciled forecasts are aggregate consistent and involve a combination of all the independent forecasts. They are unbiased as $\widehat{\boldsymbol{\beta}}_{n+h} \rightarrow \boldsymbol{\beta}_{n+h}$ and $\mathrm{E}[\overline{\boldsymbol{R}}_{n+h}] = \boldsymbol{S}_{n+h}\boldsymbol{\beta}_{n+h}$. Alternatively, assuming that $\boldsymbol{W}_h = k_h\boldsymbol{W}_1$, we approximate $\boldsymbol{W}_1$ by its diagonal in-sample one-step-ahead forecast errors. This leads to the weighted least-squares estimate.

5 Results of the Point Forecasts

5.1 Functional Time Series Model Fitting

For the national and sub-national mortality rates, we examine the goodness of fit of the functional time series model to the smoothed data. The number of retained components in the functional principal component decomposition is determined by explaining at least 99 percent of the total variation; we present and attempt to interpret the first two components for the female mortality series in Hokkaido as an illustration.

In the first column of Figure 4, we present the average of female log mortality rates. In the first row of Figure 4, we also present the first two functional principal components, which account for 99.4 percent of the total variation. Each functional principal component models different movements in mortality rates. By inspecting the peaks, the first functional principal component models the mortality at around age 80, while the second functional principal component models the mortality mainly at around age 90. Since the principal component scores are surrogates of the original functional time series, they are forecast to decrease over the next 20 years.

In Figure 5, we present the functional time series model fit to the smoothed data. The difference between the fitted and smoothed data (i.e., residuals) is highlighted in a filled contour plot in Figure 5.

In addition to the graphical display, we measure goodness of fit via a functional version of the R^2 criterion. It is given as

$$R^2 = 1 - \frac{\int_{x\in\mathcal{I}}\sum_{t=1}^{n}\left[\exp^{f_t(x)} - \exp^{\widehat{f}_t(x)}\right]^2 dx}{\int_{x\in\mathcal{I}}\sum_{t=1}^{n}\left[\exp^{f_t(x)} - \exp^{\overline{f}(x)}\right]^2 dx}, \tag{3}$$

where $f_t(x)$ denotes the smoothed age-specific log mortality rates and $\widehat{f}_t(x)$ denotes the fitted age-specific log mortality rates. The larger the R^2 value is, the better the goodness of fit by the functional time series model.

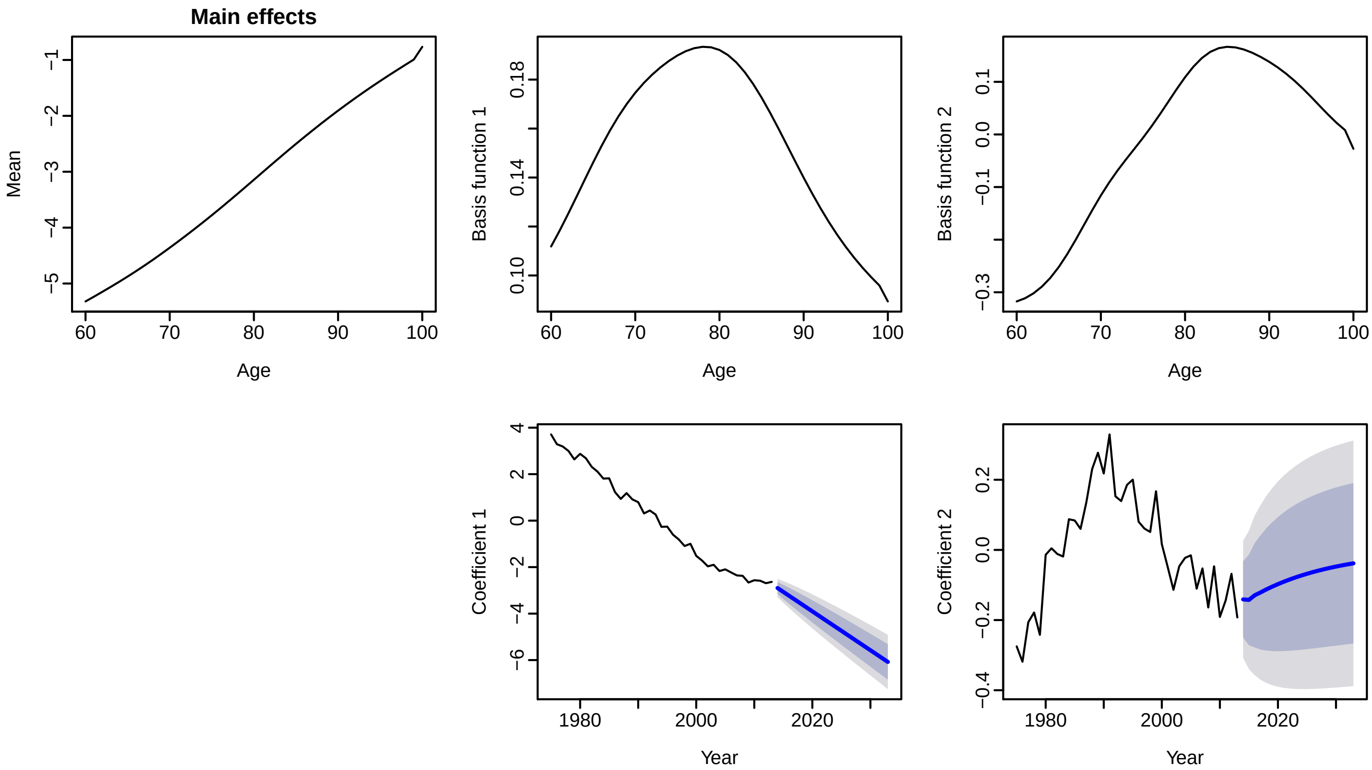
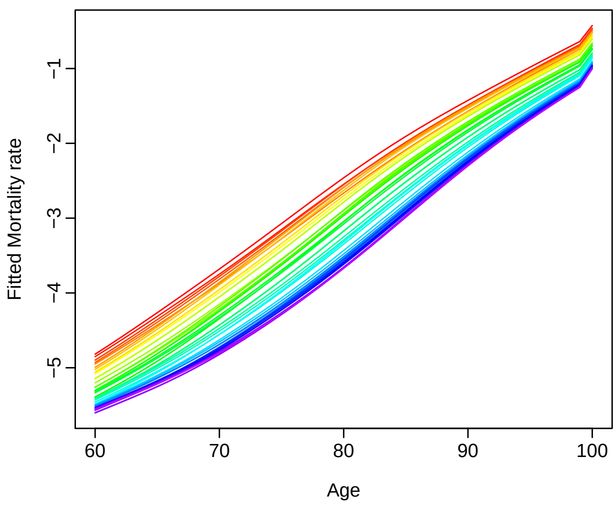
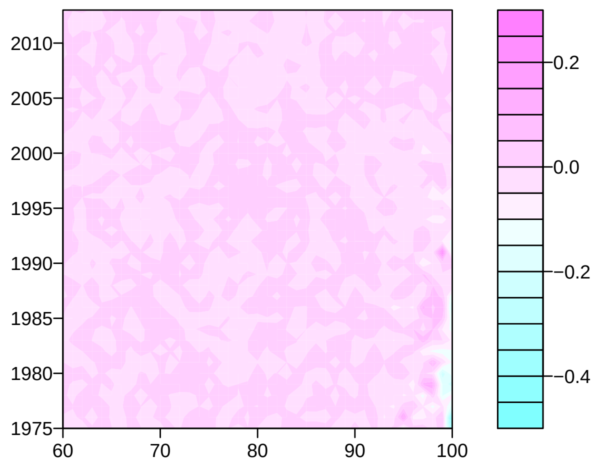


Fig. 4. ***Functional Principal Component Decomposition for the Female Mortality Data in Hokkaido***
Note: In the bottom panels, the solid blue line represents the point forecasts of scores, and the dark- and light-gray regions represent the 80% and 95% pointwise prediction intervals, respectively.



(a) *Fitted Values*



(b) *Residual Values*

Fig. 5. ***Functional Time Series Model Fitting and Residual Filled Contour Plots***

Based on the historical mortality from 1975 to 2013, we produce the point forecasts of age-specific mortality rates from 2014 to 2033. As shown in Figure 6, the mortality rates are continuing to decline, especially for the population over 60.

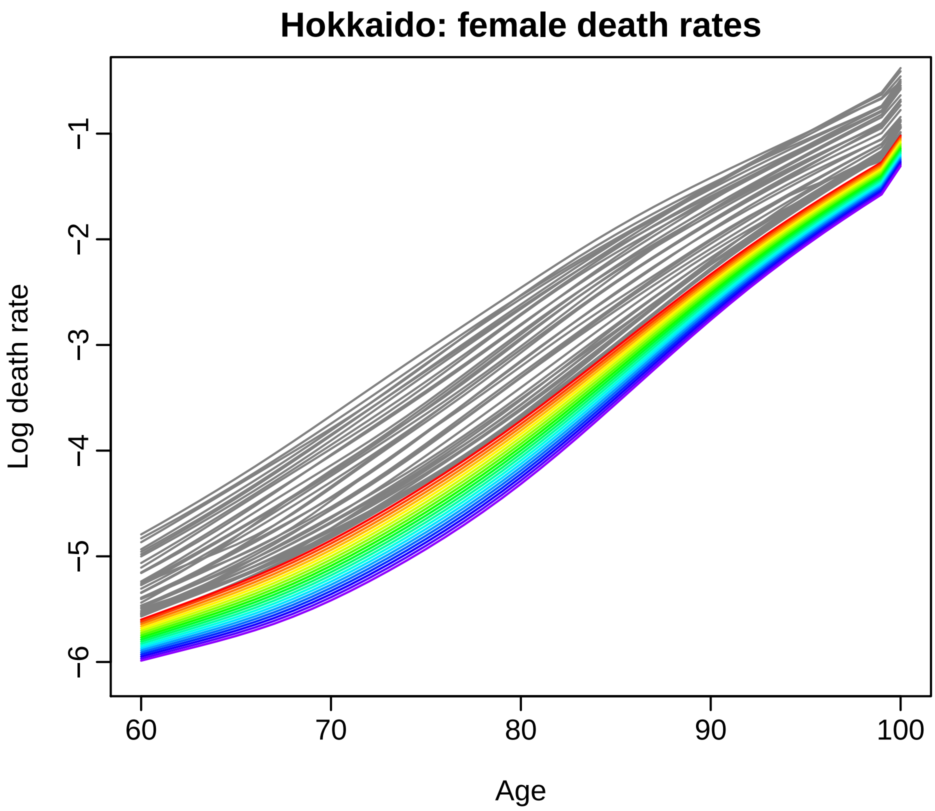


Fig. 6. ***Point Forecasts of Age-Specific Mortality Rates From 2014 to 2033***
Note: While the historical functional time series is shown in gray, the forecasts are highlighted in rainbow colors.

Due to the limited space, we cannot show the functional time series model fitting and forecasts for each sub-national group. However, in Table 2, we report the retained number of components determined by explaining at least 99 percent of total variation, and the goodness of fit of the functional time series model as measured by the R^2 criterion in (3). The retained number of components is used in the functional time series model for fitting each sub-national age-specific mortality.

5.2 Point Forecast Evaluation

An expanding window analysis of a time series model is commonly used to assess model and parameters stabilities over time, as well as prediction accuracy. The expanding window analysis assesses the constancy of a model's parameter by computing parameter estimates and their resultant forecasts over a rolling window of a fixed size through the sample (see Zivot, Eric and Jiahui Wang., 2006, Chapter 9 for details). Using the first 24 observations from 1975 to 1998 in the Japanese age-specific mortality rates, we produce one- to 15-step-ahead point forecasts. Through a rolling window approach, we re-estimate the parameters in the time series forecasting models, using the first 25 observations from 1975 to 1999. Forecasts are then produced for one- to 14-step-ahead from the

Table 2. *Number of Retained Functional Principal Components J and the Goodness of Fit as Measured by the R^2 for Each National and Sub-national Female, Male and Total Mortality Rates in Japan*

	Female		**Male**		**Total**			**Female**		**Male**		**Total**	
Prefecture	J	R^2	J	R^2	J	R^2	**Prefecture**	J	R^2	J	R^2	J	R^2
Japan	1	0.98	2	0.98	2	0.99	Mie	2	0.96	4	0.97	3	0.93
Hokkaido	2	0.96	3	0.97	2	0.96	Shiga	2	0.94	3	0.99	3	0.96
Aomori	2	0.95	3	0.96	2	0.95	Kyoto	2	0.96	3	0.97	3	0.99
Iwate	2	0.95	4	0.97	3	0.99	Osaka	1	0.95	3	0.98	3	0.97
Miyagi	2	0.91	3	0.98	2	0.94	Hyogo	1	0.97	3	0.98	2	0.94
Akita	3	0.99	4	0.98	3	0.98	Nara	2	0.96	4	0.99	3	0.97
Yamagata	2	0.95	3	0.97	2	0.99	Wakayama	2	0.98	5	0.98	3	0.99
Fukushima	2	0.95	3	0.98	2	0.97	Tottori	2	0.96	4	0.97	3	0.98
Ibaraki	2	0.95	4	0.97	3	0.97	Shimane	2	0.94	3	0.99	3	0.99
Tochigi	2	0.96	3	0.99	3	0.99	Okayama	2	0.96	4	0.98	3	0.97
Gunma	2	0.96	4	0.98	3	0.96	Hiroshima	2	0.95	4	0.98	2	0.97
Saitama	2	0.98	3	0.97	3	0.97	Yamaguchi	2	0.98	4	0.98	3	0.99
Chiba	1	0.98	4	0.99	3	0.96	Tokushima	2	0.93	3	0.99	3	0.99
Tokyo	1	0.94	3	0.98	2	0.98	Kagawa	2	0.94	4	0.98	3	0.98
Kanagawa	1	0.96	3	0.97	2	0.96	Ehime	2	0.97	4	0.98	3	0.99
Niigata	2	0.95	3	0.98	2	0.97	Kochi	2	0.98	3	0.99	2	0.99
Toyama	2	0.96	4	0.99	2	0.95	Fukuoka	1	0.98	4	0.98	3	0.95
Ishikawa	2	0.96	4	0.96	2	0.95	Saga	2	0.99	3	0.98	3	0.99
Fukui	2	0.93	3	0.99	3	0.99	Nagasaki	2	0.97	3	0.98	2	0.97
Yamanashi	2	0.97	4	0.98	3	0.98	Kumamoto	2	0.99	3	0.98	2	0.99
Nagano	2	0.97	3	0.98	2	0.96	Oita	2	0.99	3	0.98	3	0.99
Gifu	2	0.95	4	0.98	3	0.97	Miyazaki	2	0.99	4	0.97	2	0.98
Shizuoka	2	0.99	4	0.97	3	0.94	Kagoshima	2	0.95	4	0.97	2	0.96
Aichi	2	0.98	4	0.97	3	0.95	Okinawa	2	0.98	4	0.99	3	0.99

estimated models. We iterate this process by increasing the sample size by one year until reaching the end of the data period in 2013. This process produces 15 one-step-ahead forecasts, 14 two-step-ahead forecasts, ..., and one 15-step-ahead forecast. We compare these forecasts with the holdout samples to determine the out-of-sample point forecast accuracy.

To evaluate the point forecast accuracy, we consider the mean absolute forecast error (MAFE) and root mean squared forecast error (RMSFE). They measure how close the forecasts are in comparison to the actual values of the variable being forecast, regardless the direction of forecast errors. For each series k, they can be written as

$$\text{MAFE}_k(h) = \frac{1}{41 \times (16-h)} \sum_{\varsigma=h}^{15} \sum_{j=1}^{41} |m_{n+\varsigma}^k(x_j) - \widehat{m}_{n+\varsigma}^k(x_j)|$$

$$\text{RMSFE}_k(h) = \sqrt{\frac{1}{41 \times (16-h)} \sum_{\varsigma=h}^{15} \sum_{j=1}^{41} \left[m_{n+\varsigma}^k(x_j) - \widehat{m}_{n+\varsigma}^k(x_j)\right]^2},$$

where $m_{n+\varsigma}^k(x_j)$ denotes the actual holdout sample for the jth age and ςth curve in the kth series, while $\widehat{m}_{n+\varsigma}^k(x_j)$ denotes the point forecasts for the holdout sample.

By averaging $\text{MAFE}_k(h)$ and $\text{RMSFE}_k(h)$ across the number of series within each level of disaggregation, we obtain an overall assessment of the point forecast accuracy for each level within the collection of series, denoted by $\text{MAFE}(h)$ and $\text{RMSFE}(h)$. They are defined as

$$\text{MAFE}(h) = \frac{1}{m_k} \sum_{k=1}^{m_k} \text{MAFE}_k(h),$$

$$\text{RMSFE}(h) = \frac{1}{m_k} \sum_{k=1}^{m_k} \text{RMSFE}_k(h),$$

where m_k denotes the number of series at the kth level of disaggregation, for $k = 1, \dots, K$.

For 15 different forecast horizons, we consider two summary statistics to evaluate overall point forecast accuracy among the methods for national and sub-national mortality forecasts. The summary statistics chosen are the mean and median values, due to their suitability for handling squared and absolute errors (Gneiting, Tilmann., 2011). They are given by

$$\text{Mean (RMSFE)} = \frac{1}{15} \sum_{h=1}^{15} \text{RMSFE}(h),$$

$$\text{Median (MAFE)} = \text{MAFE}[8],$$

where $[8]$ denotes the eighth term after ranking $\text{MAFE}(h)$ for $h = 1, \dots, 15$ from smallest to largest.

5.3 Comparison of Point Forecast Accuracy

Averaging over all series at each level of a hierarchy, Table 3 presents $\text{MAFE}(h)$ and $\text{RMSFE}(h)$ between the Lee-Carter and functional time series methods. As measured by the MAFE, the functional time series method produces more accurate point forecasts than the ones obtained by the Lee-Carter method, at each level of the hierarchy. As measured by the RMSFE, the functional time series method outperforms the Lee-Carter method for all levels of the hierarchy, except when total mortality rates are disaggregated by regions. The superior forecast accuracy of the functional time series model over the Lee-Carter model stems from two sources. First, a smoothing technique is implemented to remove any noise in the data series, especially at older ages. Second, more than one component is used to achieve improved model fitting.

Since the functional time series method outperforms the Lee-Carter method, we evaluate and compare MAFE(h) and RMSFE(h) between the independent and grouped functional time series forecasting methods for each level within the Japanese data hierarchy. From Table 4, the optimal combination method produces most accurate point forecast accuracy in the short-term forecast horizon ($h = 1$). As the forecast horizon increases, the bottom-up method generally gives the most accurate forecasts at the national and sub-national levels for the total series. Based on averaged forecast errors, the bottom-up method performs the best at each level of the hierarchy. Averaged over all levels of the hierarchy, it is advantageous to use the bottom-up method, followed by the optimal combination method.

6 Application to Annuity Pricing

One main use of mortality forecasts for the elderly (at, say, ages over 60) is in pension and insurance industries, whose profitability relies on accurate mortality forecasts to appropriately hedge longevity

Table 3. ***MAFEs and RMSFEs*** $(\times 100)$ ***in the Holdout Sample Between the Functional Time Series and Lee-Carter Methods Applied to the Japanese Mortality Rates***

	MAFE					RMSFE				
Level	$h = 1$	**5**	**10**	**15**	**Median**	**1**	**5**	**10**	**15**	**Mean**
Functional Time Series Method										
Total	**0.27**	**0.52**	**0.78**	**0.96**	**0.68**	**0.52**	**0.95**	**1.42**	**1.54**	**1.15**
Sex	**0.34**	**0.58**	**0.85**	**0.96**	**0.73**	**0.70**	**1.15**	**1.69**	**1.75**	**1.39**
Region	**0.38**	**0.57**	**0.83**	**0.92**	**0.73**	0.85	1.16	1.61	**1.63**	1.35
Region + Sex	**0.51**	**0.67**	**0.93**	**1.02**	**0.81**	**1.18**	**1.45**	**1.95**	**1.99**	**1.69**
Prefecture	**0.58**	**0.70**	**0.89**	**0.96**	**0.82**	**1.29**	**1.40**	**1.68**	**1.70**	**1.53**
Prefecture + Sex	**0.94**	**0.97**	**1.12**	**1.17**	**1.06**	**2.21**	**2.15**	**2.35**	**2.32**	**2.26**
Lee-Carter Method										
Total	0.30	0.54	0.84	0.97	0.71	0.56	0.96	1.49	1.59	1.19
Sex	0.35	0.61	0.95	1.05	0.79	0.73	1.17	1.80	1.90	1.46
Region	0.39	0.58	0.87	1.00	0.74	**0.80**	**1.11**	**1.59**	1.74	**1.33**
Region + Sex	0.54	0.74	1.08	1.26	0.92	1.24	1.61	2.25	2.61	1.93
Prefecture	0.66	0.81	1.10	1.27	0.97	1.48	1.73	2.27	2.70	2.04
Prefecture + Sex	1.19	1.43	1.87	2.18	1.67	3.01	3.57	4.56	5.32	4.13

Note: The bold entries highlight the method that gives the most accurate forecasts for each level of the hierarchy. The forecast errors have been multiplied by 100, in order to keep 2 decimal places.

risks. When one reaches retirement, an optimal way (as demonstrated by Yaari, Menahem., 1965) of guaranteeing an individual's financial income in retirement and of ensuring that an individual does not outlive their financial assets is to purchase an annuity. An annuity is a contract offered by insurers guaranteeing a steady stream of payments for either a fixed term or the lifetime of the annuitants in exchange for an initial premium fee (Roy, Amlan., 2012).

We apply the mortality forecasts to the calculation of a fixed-term annuity (see Dickson, David, Mary Hardy and Howard Waters., 2009, p114), and we take a cohort approach to the calculation of the survival probabilities. The τ-year survival probability of a person aged x currently at $t = 0$ is determined by

$$_\tau p_x = \prod_{j=1}^{\tau} {}_1p_{x+j-1} = \prod_{j=1}^{\tau} e^{-m_{x+j-1,j-1}},$$

which is a random variable, since mortality rates for $j = 1, \dots, \tau$ are forecasts obtained by the

Table 4. ***MAFEs and RMSFEs*** $(\times 100)$ ***in the Holdout Sample Among the Independent (Base) Forecasting, Bottom-Up and Optimal Combination Methods***

		MAFE					RMSFE				
Level	**Method**	$h = 1$	**5**	**10**	**15**	**Median**	**1**	**5**	**10**	**15**	**Mean**
Total	Base	**0.27**	0.52	0.78	0.96	0.68	**0.52**	0.95	1.42	1.54	1.15
	BU	0.31	**0.43**	**0.53**	**0.47**	**0.45**	0.62	**0.82**	**1.00**	**0.94**	**0.87**
	OLS	**0.27**	0.44	0.60	0.68	0.52	**0.52**	0.84	1.18	1.24	0.98
Sex	Base	0.34	0.58	0.85	0.96	0.73	0.70	1.15	1.69	1.75	1.39
	BU	0.35	**0.54**	**0.80**	**0.93**	**0.67**	0.69	**1.01**	**1.52**	**1.62**	**1.25**
	OLS	**0.31**	0.58	0.93	1.14	0.77	**0.63**	1.09	1.72	1.92	1.40
Region	Base	0.38	0.57	0.83	0.92	0.73	0.85	1.16	1.61	1.63	1.35
	BU	0.39	0.49	**0.60**	**0.57**	**0.52**	0.83	**0.96**	**1.17**	**1.13**	**1.03**
	OLS	**0.36**	**0.48**	0.65	0.72	0.57	**0.74**	0.97	1.30	1.35	1.12
Region + Sex	Base	0.51	0.67	0.93	1.02	0.81	1.18	1.45	1.95	1.99	1.69
	BU	0.50	**0.64**	**0.89**	**1.01**	**0.77**	1.10	**1.29**	**1.76**	**1.85**	**1.53**
	OLS	**0.47**	0.67	0.99	1.19	0.83	**1.04**	1.35	1.92	2.11	1.65
Prefecture	Base	**0.58**	0.70	0.89	0.96	0.82	**1.29**	1.40	1.68	1.70	1.53
	BU	0.61	**0.65**	**0.72**	**0.67**	0.66	1.35	**1.36**	**1.45**	**1.35**	**1.39**
	OLS	**0.58**	**0.65**	0.77	0.78	0.72	**1.29**	**1.36**	1.55	1.53	1.44
Prefecture + Sex	Base	0.94	**0.97**	**1.12**	**1.17**	**1.06**	2.21	**2.15**	**2.35**	**2.32**	**2.26**
	BU	0.94	**0.97**	**1.12**	**1.17**	**1.06**	2.21	**2.15**	**2.35**	**2.32**	**2.26**
	OLS	**0.93**	0.99	1.20	1.30	1.11	**2.18**	2.18	2.47	2.50	2.35

functional time series method. Here, we assume the central mortality rates are constant throughout a one-year period.

The price of an annuity with maturity T years, written for a x-year-old with benefit \$1 per year and conditional on the path, is given by

$$a_x^T(\boldsymbol{m}_{1:T}^x) = \sum_{\tau=1}^{T} B(0,\tau)\mathrm{E}(1_{T_x>\tau}|\boldsymbol{m}_{1:\tau}^x)$$
$$= \sum_{\tau=1}^{T} B(0,\tau)_\tau p_x(\boldsymbol{m}_{1:\tau}^x),$$

where $B(0,\tau)$ is the τ-year bond price, $\boldsymbol{m}_{1:\tau}^x$ is the first τ elements of $\boldsymbol{m}_{1:T}^x$, and $_\tau p_x(\boldsymbol{m}_{1:\tau}^x)$ denotes the survival probability given a random $\boldsymbol{m}_{1:\tau}^x$ (see also Fung, M, Gareth Peters and Pavel Shevchenko., 2015). It is vital to produce an accurate forecast of the survival curve $_\tau p_x$ that best captures the mortality experience of a portfolio for pricing purposes and risk management.

In Table 5, we compare the best estimate of the annuity prices for different ages and maturities between the three forecasting methods for a female policyholder residing in Region 2, as an example. We assume a constant interest rate at $\tau = 3\%$, and hence $B(0,\tau) = e^{-\tau}$. Although the annuity price difference might appear to be small, any mispricing can have a significant risk when considering a large annuity portfolio. For an annuity portfolio that consists of N policies where the benefit per year is B, any underpricing of $\gamma\%$ of the actual annuity price will result in a shortfall of $NBa_x^\top\gamma/100$, where $a_x^\top$ is the estimated annuity price being charged with benefit \$1 per year. For example, we have $\gamma = 0.1\%$, $N = 10,000$ policies written to 85-year-old policyholders with maturity $\tau = 15$ years, and \$20,000 benefit per year will result in a shortfall of $10,000 \times 20,000 \times 8.2478 \times 0.1\% = 1.65$ million.

Table 5. ***Estimates of Annuity Prices With Different Ages and Maturities (T) for Female Policyholder Residing in Region 2***

	$T=5$	$T=10$	$T=15$	$T=20$	$T=25$	$T=30$
			Age = 60			
Base	4.5214	8.3227	11.4984	14.1234	16.2410	17.8464
BU	4.5250	8.3362	11.5243	14.1563	16.2664	17.8457
OLS	4.5269	8.3435	11.5426	14.1925	16.3307	17.9514
			Age = 65			
Base	4.5071	8.2723	11.3848	13.8954	15.7990	17.0569
BU	4.5122	8.2867	11.4028	13.9010	15.7708	16.9868
OLS	4.5142	8.2982	11.4325	13.9615	15.8784	17.1489
			Age = 70			

Continued on next page

	$T = 5$	$T = 10$	$T = 15$	$T = 20$	$T = 25$	$T = 30$
Base	4.4892	8.2000	11.1934	13.4630	14.9627	15.7340
BU	4.4916	8.1997	11.1725	13.3976	14.8446	15.5933
OLS	4.4974	8.2226	11.2284	13.5068	15.0168	15.8112
			Age = 75			
Base	4.4564	8.0512	10.7767	12.5778	13.5039	NA
BU	4.4506	8.0187	10.6893	12.4260	13.3245	NA
OLS	4.4601	8.0589	10.7867	12.5946	13.5457	NA
			Age = 80			
Base	4.3786	7.6984	9.8922	11.0203	NA	NA
BU	4.3584	7.6206	9.7421	10.8396	NA	NA
OLS	4.3773	7.6952	9.8942	11.0510	NA	NA
			Age = 85			
Base	4.1866	6.9532	8.3758	NA	NA	NA
BU	4.1513	6.8510	8.2478	NA	NA	NA
OLS	4.1850	6.9588	8.4180	NA	NA	NA
			Age = 90			
Base	3.8122	5.7725	NA	NA	NA	NA
BU	3.7776	5.7320	NA	NA	NA	NA
OLS	3.8192	5.8284	NA	NA	NA	NA
			Age = 95			
Base	3.2376	NA	NA	NA	NA	NA
BU	3.2768	NA	NA	NA	NA	NA
OLS	3.2991	NA	NA	NA	NA	NA

Note: These estimates are based on forecast mortality rates from 2014 to 2054. We only consider contracts with maturity so that age + maturity $\leq$ 100. If age + maturity $>$ 100, NA will be shown in the table.

7 Conclusions

Using the national and sub-national Japanese mortality data, we evaluate and compare the point forecast accuracy between the Lee-Carter and functional time series methods. Based on the forecast accuracy criteria, we found that the functional time series method outperforms the Lee-Carter method.

The superiority of the functional time series method is driven by the use of nonparametric smoothing techniques in order to deal with noisy mortality rates at older ages, in particular for males, and more than one component is used to achieve improved model fitting.

By using the functional time series method to produce base forecasts, we consider the issue of forecast reconciliation by applying two grouped functional time series forecasting methods, namely the bottom-up and optimal combination methods. The bottom-up method models and forecasts data series at the most disaggregated level, and then aggregates the forecasts using the summing matrix constructed on the basis of forecast exposure to risk.

Using the Japanese data, we compare the one-step-ahead to 15-step-ahead forecast accuracy between the independent functional time series forecasting method and two proposed grouped functional time series forecasting methods. We found that the grouped functional time series forecasting methods produce more accurate point forecasts than those obtained by the independent functional time series forecasting method, averaged over all levels of the hierarchy. In addition, the grouped functional time series forecasting methods produce forecasts that obey the natural group structure, thus giving forecast mortalities at the sub-national levels that add up to the forecast mortality rates at the national level. Between the two grouped functional time series forecasting methods, the bottom-up method is recommended.

We apply the independent functional time series and two grouped functional time series methods to forecast age-specific mortality rates from 2014 to 2054. Then we calculate the cumulative survival probability and obtain the fixed-term annuity prices. We found that the cumulative survival probability has a pronounced impact on annuity prices. Although annuity prices do not differ much for the mortality forecasts obtained by the three methods, mispricing could have a dramatic impact for a portfolio, especially when the yearly benefit is much larger than $1 per year.

There are a few ways in which this paper can be further extended, and we briefly mention three. First, we will compare interval forecast accuracy between the methods to assess forecast uncertainties associated with best estimates of annuity prices. Second, subject to the availability of data, the hierarchy can be disaggregated finer by considering different causes (Murray, Christopher and Alan Lopez., 1997; Gaille, Séverine and Michael Sherris., 2015) or socioeconomic status (Bassuk, Shari, Lisa Berkman and Benjamin. Amick III., 2002; Singh, Gopal, Romuladus Azuine, Mohammad Siahpush and Michael Kogan., 2013; Villegas, Andrés and Steven Haberman., 2014). Third, the methodology can be applied to calculate other types of annuity price, such as whole life immediate annuity or deferred annuity.

Supplementary Material

R package for functional time series forecasting The R package *ftsa* containing code to produce point forecasts from the Lee-Carter and functional time series forecasting methods described in the article. The R package can be obtained from CRAN (https://cran.r-project.org/web/packages/ftsa/index.html).

Code for grouped functional time series forecasting The R code to produce point forecasts from the two grouped functional time series forecasts described in the article (https://www.researchgate.net/publication/317088999_Code_for_grouped_functional_time_series_forecasting).

Code for shiny application in statistical software R The R code to produce a shiny user interface (https://www.researchgate.net/publication/317089428_Code_for_shiny_application_in_the_grouped_functional_time_series_forecasting_paper) for plotting every series in the Japanese data hierarchy.

References

Aue, Alexander, Diogo Norinho and Siegfried. Hörmann. 2015. On the prediction of stationary functional time series. *Journal of the American Statistical Association*, 110(509):378–392.

Bassuk, Shari, Lisa Berkman and Benjamin. Amick III. 2002. Socioeconomic Status and Mortality Among the Elderly: Findings From Four US Communities. *American Journal of Epidemiology*, 155(6):520–533.

Booth, Heather. 2006. Demographic Forecasting: 1980 to 2005 in Review. *International Journal of Forecasting*, 22(3):547–581.

Booth, Heather and Leonie Tickle. 2008. Mortality Modelling and Forecasting: A Review of Methods. *Annals of Actuarial Science*, 3(1–2):3–43.

Booth, Heather, Rob Hyndman, Leonie Tickle and Piet De Jong. 2006. Lee-Carter Mortality Forecasting: A Multi-country Comparison of Variants and Extension. *Demographic Research*, 15(9):289–310.

Brouhns, Natacha, Michel Denuit and Jeroen Vermunt. 2002. A Poisson Log-Bilinear Regression Approach to the Construction of Projected Lifetables. *Insurance: Mathematics and Economics*, 31(3):373–393.

Chang, Winston, Joe Cheng, J. Allaire, Yihui. Xie and Jonathan McPherson. 2017. *Shiny: Web Application Framework for R*. R package version 1.0.3.

Coulmas, Florian. 2007. *Population Decline and Ageing in Japan—the Social Consequences*. New York: Routledge.

Currie, Iain, Maria Durban and Paul Eilers. 2004. Smoothing and Forecasting Mortality Rates. *Statistical Modelling*, 4(4):279–298.

D'Amato, Valeria, Gabriella Piscopo and Maria Russolillo. 2011. The Mortality of the Italian Population: Smoothing Technique on the Lee-Carter Model. *Annals of Applied Statistics*, 5(2A):705–724.

Denuit, Michel, Pierre Devolder and Anne-Cécile Goderniaux. 2007. Securitization of Longevity Risk: Pricing Survivor Bonds With Wang Transform in the Lee-Carter Framework. *Journal of Risk and Insurance*, 74(1):87–113.

Dickson, David, Mary Hardy and Howard Waters. 2009. *Actuarial Mathematics for Life Contingent Risks*. Cambridge: Cambridge University Press.

Fung, M, Gareth Peters and Pavel Shevchenko. 2015. A State-Space Estimation of the Lee-Carter Mortality Model and Implications for Annuity Pricing. In Weber, T., McPhee, M. J., and Anderssen,

R. S., editors, *MODSIM2015, 21st International Congress on Modelling and Simulation*, pp. 952–958, Modelling and Simulation Society of Australia and New Zealand.

Gaille, Séverine and Michael Sherris. 2015. Causes-of-Death Mortality: What Do We Know on Their Dependence? *North American Actuarial Journal*, 19(2):116–128.

Gneiting, Tilmann. 2011. Making and evaluating point forecasts. *Journal of the American Statistical Association*, 106(494):746–762.

Hanewald, Katja, Thomas Post and Helmut Gründl. 2011. Stochastic Mortality, Macroeconomic Risks and Life Insurer Solvency. *Geneva Papers on Risk and Insurance: Issues and Practice*, 36:458–475.

Hyndman, R. and Ullah, M. 2007. Robust Forecasting of Mortality and Fertility Rates: A Functional Data Approach. *Computational Statistics and Data Analysis*, 51(10):4942–4956.

Hyndman, Rob, Alan Lee and Earo Wang. 2016. Fast computation of reconciled forecasts for hierarchical and grouped time series. *Computational Statistics and Data Analysis*, 97:16–32.

Hyndman, Rob and Han Shang. 2009. Forecasting functional time series (with discussions). *Journal of the Korean Statistical Society*, 38(3):199–221.

Hyndman, Rob and Han Shang. 2010. Rainbow plots, bagplots, and boxplots for functional data. *Journal of Computational and Graphical Statistics*, 19(1):29–45.

Hyndman, Rob and Yeasmin Khandakar. 2008. Automatic Time Series Forecasting: The Forecast Package for R. *Journal of Statistical Software*, 27(3).

Hyndman, Rob, Roman Ahmed, George Athanasopoulos and Han Shang. 2011. Optimal Combination Forecasts for Hierarchical Time Series. *Computational Statistics and Data Analysis*, 55:2579–2589.

Japanese Mortality Database 2016. *National Institute of Population and Social Security Research*. Available at http://www.ipss.go.jp/p-toukei/JMD/index-en.html (data downloaded on July/18/2015).

Kahn, Kenneth. 1998. Revisiting top-down versus bottom-up forecasting. *Journal of Business Forecasting*, 17(2):14–19.

Koissia, Marie-Claire. 2006. Longevity and Adjustment in Pension Annuities, With Application to Finland. *Scandinavian Actuarial Journal*, 4:226–242.

Lee, Ronald and Lawrence Carter. 1992. Modeling and Forecasting U.S. Mortality. *Journal of the American Statistical Association*, 87(419):659–671.

Li, Nan and Ronald Lee. 2005. Coherent Mortality Forecasts for a Group of Populations: An Extension of the Lee-Carter Method. *Demography*, 42(3):575–594.

Murray, Christopher and Alan Lopez. 1997. Alternative Projections of Mortality and Disability by Cause 1990–2020: Global Burden of Disease Study. *Lancet*, 349(9064):1498–1504.

Organisation for Economic Co-operation and Development 2013. Pensions at a Glance 2013: OECD and G20 Indicators. Technical report, OECD Publishing. Retrieved from http://dx.doi.org/10.1787/pension_glance-2013-en at January 20, 2016.

Pitacco, Ermanno, Michel Denuit, Steven Haberman and Annamaria Olivieri. 2009. *Modelling Longevity Dynamics for Pensions and Annuity Business*. Oxford: Oxford University Press.

R Core Team 2017. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Renshaw, A and S Haberman. 2003. Lee-Carter Mortality Forecasting With Age-Specific Enhancement. *Insurance: Mathematics and Economics*, 33(2):255–272.

Renshaw, A and S Haberman. 2006. A Cohort-Based Extension to the Lee-Carter Model for Mortality Reduction Factors. *Insurance: Mathematics and Economics*, 38(3):556–570.

Roy, Amlan. 2012. Innovative Approaches to Managing Longevity Risk in Asia: Lessons From the West. Working paper, Asian Development Bank Institute.

Shang, Han. 2016. Mortality and Life Expectancy Forecasting for a Group of Populations in Developed Countries: A Multilevel Functional Data Method. *Annals of Applied Statistics*, 10(3):1639–1672.

Shang, Han. 2017. Reconciling Forecasts of Infant Mortality Rates at National and Sub-national Levels: Grouped Time-Series Methods. *Population Research and Policy Review*, 36(1):55–84.

Shang, Han and Rob Hyndman. 2017. Grouped functional time series forecasting: An application to age-specific mortality rates. *Journal of Computational and Graphical Statistics*, 26(2):330–343.

Shang, Han, Heather Booth and Rob Hyndman. 2011. Point and Interval Forecasts of Mortality Rates and Life Expectancy: A Comparison of Ten Principal Component Methods. *Demographic Research*, 25(5):173–214.

Singh, Gopal, Romuladus Azuine, Mohammad Siahpush and Michael Kogan. 2013. All-Cause and Cause-Specific Mortality Among US Youth: Socioeconomic and Rural-Urban Disparities and International Patterns. *Journal of Urban Health*, 90(3):388–405.

Stone, Richard, D Champernowne and J Meade. 1942. The precision of national income estimates. *Review of Economic Studies*, 9(2):111–125.

Suzuki, Makoto, Bradley Willcox and Craig Willcox. 2004. Successful Aging: Secrets of Okinawan Longevity. *Geriatrics and Gerontology International*, 4(s1):S180–S181.

Takata, Hajime, Toshiharu Ishii, Makoto Suzuki, Susumu Sekiguchi and Hisami Iri. 1987. Influence of major histocompatibility complex region genes on human longevity among Okinawan-Japanese centenarians and nonagenarians. *Lancet*, 330(8563):824–826.

Villegas, Andrés and Steven Haberman. 2014. On the Modeling and Forecasting of Socioeconomic Mortality Differentials: An Application to Deprivation and Mortality in England. *North American Actuarial Journal*, 18(1):168–193.

Willcox, Craig, Bradley J Willcox, Sanae Shimajiri, Sayuri Kurechi and Makoto Suzuki. 2007. Aging Gracefully: A Retrospective Analysis of Functional Status in Okinawan Centenarians. *American Journal of Geriatric Psychiatry*, 15(3):252–256.

Yaari, Menahem. 1965. Uncertain lifetime, life insurance, and the theory of the consumer. *Review of Economic Studies*, 32(2):137–150.

Zivot, Eric and Jiahui Wang. 2006. *Modeling Financial Time Series With S-PLUS*. New York: Springer.