Forecasting multiple functional time series in a group structure: an application to mortality

Han Lin Shang*
Research School of Finance, Actuarial Studies and Statistics
Australian National University

Steven Haberman
Cass Business School
City, University of London

January 11, 2020

Abstract

When modeling sub-national mortality rates, we should consider three features: (1) how to incorporate any possible correlation among sub-populations to potentially improve forecast accuracy through multi-population joint modeling; (2) how to reconcile sub-national mortality forecasts so that they aggregate adequately across various levels of a group structure; (3) among the forecast reconciliation methods, how to combine their forecasts to achieve improved forecast accuracy. To address these issues, we introduce an extension of grouped univariate functional time series method. We first consider a multivariate functional time series method to jointly forecast multiple related series. We then evaluate the impact and benefit of using forecast combinations among the forecast reconciliation methods. Using the Japanese regional age-specific mortality rates, we investigate one-step-ahead to 15-step-ahead point and interval forecast accuracies of our proposed extension and make recommendations.

Keywords: forecast reconciliation; multivariate functional principal component analysis; bottom-up method; optimal-combination method; Japanese mortality database.

JEL code: C53; C55

---

*Postal address: Research School of Finance, Actuarial Studies and Statistics, Level 4, Building 26C, Kingsley Street, Australian National University, Acton, Canberra, ACT 2601, Australia; Phone number: +61(2) 6125 0535; Fax number: +61(2) 6125 0087; Email: hanlin.shang@anu.edu.au

1 Introduction

Increases in longevity and an ageing population have led to concerns regarding the sustainability of pensions, healthcare, and aged-care systems in many developed nations. These concerns have led to an increasing interest among government policymakers and planners to engage in the development of more accurate modeling and forecasting age-specific mortality rates. Further, annuity and pension products depend crucially on the forecast accuracy of mortality rates or their associated survival probabilities. The survival probability has been consistently underestimated in the last few decades. As a consequence, pension funds and insurance companies face longevity risk. Longevity risk is a potential risk arising from the increasing life expectancy of policyholders, which can eventually result in higher payout ratios than expected.

Many statistical methods have been proposed for modeling and forecasting age-specific mortality rates at the national level (see, e.g., Shang et al., 2011, for earlier reviews). Of these, a significant milestone in demographic forecasting was the work by Lee and Carter (1992). They implemented a principal component analysis to model age-specific mortality and extracted a single time-varying index of the level of mortality, from which the forecasts were obtained by a random walk with drift. In the demographic literature, many extensions and modifications of the Lee-Carter method are collated in Shang et al. (2011).

Modeling mortality at the sub-national level is not only important but also challenging. On the one hand, the sub-national mortality rates often suffer from relatively poor data quality with possible missing data; on the other hand, sub-national forecasts of age-specific mortality rates are useful for informing regional policy and understanding the heterogeneity in the whole population. Further, improved understanding of individual characteristics will enable insurers to price more accurately annuity products for annuitants, as in the growing market for "enhanced annuities" in the United Kingdom (see, e.g., Olivieri and Pitacco, 2016).

To our knowledge, there are few papers that model national and sub-national mortality together with respect to a group structure. Shang and Haberman (2017) and Shang and Hyndman (2017) proposed the bottom-up and optimal combination based on ordinary least squares for reconciling forecasts in a group structure. Li et al. (2019) applied the optimal combination based on generalized least squares of Wickramasuriya et al. (2019) to reconcile cause-specific mortality forecasts in a three-level hierarchy. In all these works, a demographic model, such as the Lee-Carter model or the functional time series model of Hyndman and Ullah (2007), is used to forecast each series within each level of a group structure. Similar to these papers, we also consider the forecast reconciliation methods to reconcile forecasts of age-specific mortality rates, and improve the forecast accuracy. In a novel approach that differs from these papers, we jointly model and forecast sub-national age-specific mortality rates at each level of a group structure to capture correlation among the series and further improve forecast accuracy. We consider a multivariate functional time series

2

Also, we introduce a forecast combination approach among the grouped forecasting methods to potentially improve the forecast accuracy.

Multiple-population modeling and forecasting have attracted increasing attention in actuarial science (see, e.g., Hatzopoulos and Haberman, 2013) and demography (see, e.g., Li and Lee, 2005). Our extension links multiple-population forecasting with grouped functional time series forecasting. The underlying intuition is that when multiple sub-populations are correlated, the proposed multivariate functional time-series method can capture correlation among the multiple series, and in turn, can lead to improved forecast accuracy.

The idea of forecast combination has been studied in statistics, dating back to the seminal work by Bates and Granger (1969). The underlying intuition is that when various grouped forecasting methods are combined, forecast combination may reduce bias, variance and uncertainty because of different assumptions, model structures and degrees of model complexity.

The remainder of this paper is structured as follows. In Section 2, we describe the Japanese national and sub-national mortality observed from 1975 to 2016. In Section 3, we introduce a multivariate functional time-series forecasting method for forecasting multiple series at each level of a group structure. The key technique in our multivariate functional time-series forecasting method is multivariate functional principal component analysis. In Section 4, we introduce two grouped forecasting methods and their forecast combination. In Section 5, we compare the forecast accuracy in two ways: (1) between two reconciliation methods and their forecast combination; (2) between univariate and multivariate functional time-series forecasting methods. In Section 6, we introduce an actuarial application and apply the most accurate forecasting method to estimate the temporary life annuity prices for different ages and maturities. Conclusions are drawn in Section 7, along with some reflections on how the methods presented here can be further extended.

2   Japanese age-specific mortality rates

We study Japanese age-specific mortality rates from 1975 to 2016, obtained from the Japanese Mortality Database (Japanese Mortality Database, 2019). Given that our focus is on the pricing of annuities, 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 aggregation of the last age group is to avoid the missing data issue of those at the older ages. The structure of the data is presented in Table 1, where each row denotes a level of disaggregation.

At the top level, we have total age-specific mortality rates for entire Japan. We can split these total mortality rates by various attributes such as sex, region or prefecture. For this data set, there are eight regions in Japan, 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

Table 1: *Group structure of Japanese mortality rates.*

Group level	Number of series
Japan	1
Sex	2
Region	8
Region $\times$ Sex	16
Prefecture	47
Prefecture $\times$ Sex	94
Total	168

total of $47 \times 2 = 94$ series (Shang and Haberman, 2017). All in all, across all levels of disaggregation, there are 168 series. Note that the order of the disaggregation is not unique in any group structure, as we may first disaggregate series by region. The possibility of different disaggregation orders may impact forecast accuracy and we are investigating this issue in a separate project.

3 Multivariate functional time-series forecasting

Joint modeling mortality for two or more populations simultaneously is paramount, as it allows one to model the correlations among two or more populations, distinguish between long-term and short-term effects in the mortality evolution, and explore the additional information contained in the experience of other populations to further improve forecast accuracy. These populations can be grouped by sex, geography, ethnicity, socioeconomic status and other attributes.

3.1 Multivariate functional principal component analysis

Let $y_t^{(j)}(x_i)$ be the log central mortality rates observed at the beginning of each year $t = 1, 2, \ldots, n$ at observed ages $(x_1, x_2, \ldots, x_p)$ where x is a continuous age variable, p denotes the number of ages, and superscript $^{(j)}$ represents j^{th} series. By applying a penalized regression spline smoothing, we obtain smoothed series, $f_t^{(j)}(x)$ that is observed at discrete data points with errors.

As in the case of multiple subpopulations, the multivariate functional time series are combined in a vector with

$$\boldsymbol{f}(x) = \left[f^{(1)}(x), \ldots, f^{(\omega)}(x) \right] \in R^{\omega}. \tag{1}$$

These multivariate functions are defined over the same domain Γ. The common domain Γ must be compact sets in R^{ω}, $\omega \in N$ with finite measure and with each element $f^{(j)}(x)$ assumed to be a square-integrable function $\mathcal{L}^2(\Gamma)$, where $j = 1, \ldots, \omega$. For mathematical convenience, we let

$$\boldsymbol{\mu}(x) := \mathbb{E}\left[\boldsymbol{f}(x)\right] = \left\{ \mathbb{E}\left[f^{(1)}(x) \right], \ldots, \mathbb{E}\left[f^{(\omega)}(x) \right] \right\} = \boldsymbol{0}, \tag{2}$$

where $\boldsymbol{0}$ denotes a vector of zeros. For $x, z \in \Gamma$, the cross-covariance function is defined with

elements

$$\mathcal{K}_{lj}(x,z) := \mathbb{E}\left[f^{(l)}(x)f^{(j)}(z)\right] = \text{Cov}\left[f^{(l)}(x), f^{(j)}(z)\right]. \tag{3}$$

From the cross-covariance function, we can deduce the eigenfunction as

$$(\mathcal{K}\phi)^{(l)}(z) = \sum_{j=1}^{\omega} \int_{\Gamma} \mathcal{K}_{lj}(x,z)\phi^{(j)}(x)dx, \tag{4}$$

where $\mathcal{K}$ induces the kernel of an integral operator, $\phi \mapsto \mathcal{K}\phi$ is a square-integrable function.

From the cross-covariance function, there exists an orthonormal sequence (ϕ_k) of continuous functions in $\mathcal{L}^2(\Gamma)$ and a non-increasing sequence, λ_k, of positive numbers, such that

$$\mathcal{K}_{lj}(x,z) = \sum_{k=1}^{\infty} \lambda_k \phi_k^{(l)}(x)\phi_k^{(j)}(z), \qquad x,z \in \mathcal{I}. \tag{5}$$

By functional principal component analysis, a de-centered stochastic process $f_t^{(j)}(x)$ can be expressed as

$$f_t^{(j)}(x) = \sum_{k=1}^{\infty} \beta_{t,k}^{(j)} \phi_k^{(j)}(x) \approx \sum_{k=1}^{K} \beta_{t,k}^{(j)} \phi_k^{(j)}(x), \tag{6}$$

where $\{\phi_1^{(j)}(x), \ldots, \phi_K^{(j)}(x)\}$ is a set of the first K functional principal components for the jth sub-population; $\beta_1^{(j)} = (\beta_{1,1}^{(j)}, \ldots, \beta_{n,1}^{(j)})^{\top}$ and $\{\beta_1^{(j)}, \ldots, \beta_K^{(j)}\}$ denotes a set of principal component scores and $\beta_k^{(j)} \sim N(0, \lambda_k^{(j)})$ where $\lambda_k^{(j)}$ is the kth eigenvalue of the covariance function for the jth subpopulation in (3); and $K < n$ is the retained number of functional principal components. Expansion (6) facilitates dimension reduction as the first K terms often provide a good approximation to the infinite sum, and thus the information contained in $\boldsymbol{f}^{(j)}(x) = [f_1^{(j)}(x), \ldots, f_n^{(j)}(x)]$ can be adequately summarized by the K-dimensional vector $(\beta_1^{(j)}, \ldots, \beta_K^{(j)})$.

The optimal value of K can be selected by a ratio method:

$$\underset{1 \leq K \leq n-1}{\text{argmin}} \left| \frac{\widehat{\lambda}_{K+1}}{\widehat{\lambda}_K} \right| \tag{7}$$

or the optimal value of K can be selected by explaining at least 90% of total variation:

$$\underset{1 \leq K \leq n}{\text{argmin}} \left\{ \frac{\sum_{k=1}^{K} \widehat{\lambda}_K}{\sum_{k=1}^{n} \widehat{\lambda}_K} \geq 0.9 \right\}. \tag{8}$$

We take the maximum of the K values obtained from (7) and (8).

The matrix formulation of (6) is

$$\boldsymbol{f}_t(x) \approx \boldsymbol{\beta}_t \boldsymbol{\Phi}^{\top}, \tag{9}$$

where $\boldsymbol{f}_t(x) = \left[f_t^{(1)}(x), f_t^{(2)}(x), \ldots, f_t^{(\omega)}(x)\right], \boldsymbol{\beta}_t = \left[\beta_{t,1}^{(1)}, \ldots, \beta_{t,K}^{(1)}, \beta_{t,1}^{(2)}, \ldots, \beta_{t,K}^{(2)}, \ldots, \beta_{t,1}^{(\omega)}, \ldots, \beta_{t,K}^{(\omega)}\right]$

being the vector of the basis expansion coefficients, and

$$
\boldsymbol{\Phi}(x) = \begin{pmatrix}
\phi_1^{(1)}(x) & \cdots & \phi_K^{(1)}(x) & 0 & \cdots & 0 & 0 & \cdots & 0 \\
0 & \cdots & 0 & \phi_1^{(2)}(x) & \cdots & \phi_K^{(2)}(x) & 0 & \cdots & 0 \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
0 & \cdots & 0 & 0 & \cdots & 0 & \phi_1^{(\omega)}(x) & \cdots & \phi_K^{(\omega)}(x)
\end{pmatrix}_{\omega \times \ell}, \quad (10)
$$

where $\ell = \omega \times K$.

The advantage of our multivariate functional time-series forecasting method is that the correlations among sub-populations can be captured in the cross-covariance structure described in (3). The disadvantage of our proposal is that we implicitly assume that all series share the same retained number of functional principal components, and consequently we may lose forecast accuracy for a particular series.

Based on the estimated covariance function, we can extract empirical functional principal component functions $\boldsymbol{\mathcal{B}} = \{\widehat{\phi}_1^{(j)}(x), \ldots, \widehat{\phi}_K^{(j)}(x)\}$ using functional principal component analysis. Conditioning on the smoothed functions $\boldsymbol{f}^{(j)}(x) = \{f_1^{(j)}(x), \ldots, f_n^{(j)}(x)\}$ and the estimated functional principal components $\boldsymbol{\mathcal{B}}$, the h-step-ahead point forecast of $f_{n+h}^{(j)}(x)$ can be obtained as

$$
\widehat{y}_{n+h|n}^{(j)}(x) = \widehat{f}_{n+h|n}^{(j)}(x) = \mathbb{E}\left[f_{n+h}^{(j)}(x)\,|\,\boldsymbol{f}^{(j)}(x), \boldsymbol{\mathcal{B}}\right] = \sum_{k=1}^{K} \widehat{\beta}_{n+h|n,k}^{(j)} \widehat{\phi}_k^{(j)}(x), \quad (11)
$$

where $\widehat{\beta}_{n+h|n,k}^{(j)}$ represents the time-series forecasts of the kth principal component scores for the jth subpopulation, which can be obtained by using autoregressive integrated moving average models.

4 Grouped forecasting methods

4.1 Notation

We introduce the grouped forecasting methods using the Japanese age-specific mortality rates provided in Section 2. The Japanese data follow a three-level geographic group structure, coupled with a sex-grouping variable (Shang and Haberman, 2017). The geographical group structure is presented in Figure 1. Japan can be split into eight regions from north to south, which is then divided into 47 prefectures (Shang and Haberman, 2017).

The data can also be split by sex. Each of the nodes in the geographical group structure can also be split into males and females. We refer to a particular disaggregated series using the notation $X * S$, referring to the geographic area X and the sex S, where X can take the values shown in Figure 1 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; Japan $*$ M denotes all males in Japan.
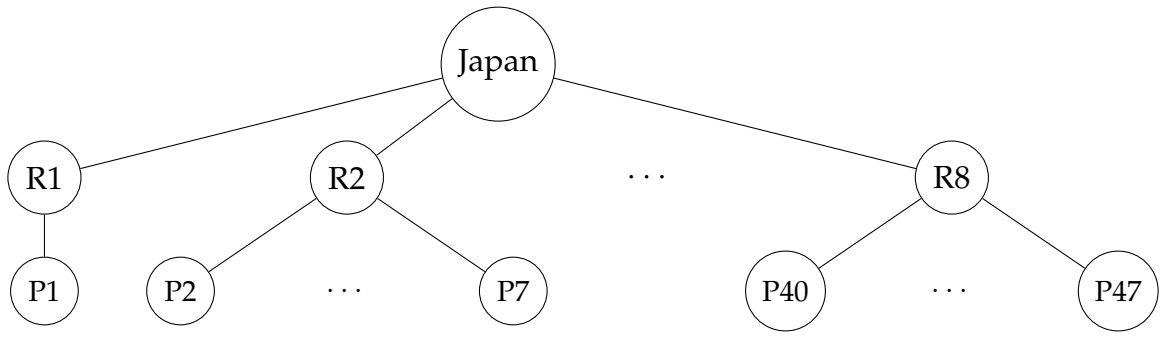
Figure 1: *The Japanese geographical group structure tree diagram, with eight regions and 47 prefectures – each node has the female, male and total age-specific mortality rates.*

Denote $E_{X*S,t}(x)$ as the exposure-to-risk for series $X*S$ in year t and age x, and let $D_{X*S,t}(x)$ be the number of deaths for series $X*S$ in year t and age x. The age-specific mortality rate is $R_{X*S,t}(x) = D_{X*S,t}(x)/E_{X*S,t}(x)$.

To simplify expressions, we will drop the age argument (x). Then for a given age, we can write

$$
\underbrace{\begin{bmatrix}
R_{Japan*T,t} \\
R_{Japan*F,t} \\
R_{Japan*M,t} \\
R_{R1*T,t} \\
R_{R2*T,t} \\
\vdots \\
R_{R8*T,t} \\
R_{R1*F,t} \\
R_{R2*F,t} \\
\vdots \\
R_{R8*F,t} \\
R_{R1*M,t} \\
R_{R2*M,t} \\
\vdots \\
R_{R8*M,t} \\
R_{P1*T,t} \\
R_{P2*T,t} \\
\vdots \\
R_{P47*T,t} \\
R_{P1*F,t} \\
R_{P1*M,t} \\
R_{P2*F,t} \\
R_{P2*M,t} \\
\vdots \\
R_{P47*F,t} \\
R_{P47*M,t}
\end{bmatrix}}_{R_t}
=
\underbrace{\begin{bmatrix}
\frac{E_{P1*F,t}}{E_{Japan*T,t}} & \frac{E_{P1*M,t}}{E_{Japan*T,t}} & \frac{E_{P2*F,t}}{E_{Japan*T,t}} & \frac{E_{P2*M,t}}{E_{Japan*T,t}} & \frac{E_{P3*F,t}}{E_{Japan*T,t}} & \frac{E_{P3*M,t}}{E_{Japan*T,t}} & \cdots & \frac{E_{P47*F,t}}{E_{Japan*T,t}} & \frac{E_{P47*M,t}}{E_{Japan*T,t}} \\
\frac{E_{P1*F,t}}{E_{Japan*F,t}} & 0 & \frac{E_{P2*F,t}}{E_{Japan*F,t}} & 0 & \frac{E_{P3*F,t}}{E_{Japan*F,t}} & 0 & \cdots & \frac{E_{P47*F,t}}{E_{Japan*F,t}} & 0 \\
0 & \frac{E_{P1*M,t}}{E_{Japan*M,t}} & 0 & \frac{E_{P2*M,t}}{E_{Japan*M,t}} & 0 & \frac{E_{P3*M,t}}{E_{Japan*M,t}} & \cdots & 0 & \frac{E_{P47*M,t}}{E_{Japan*M,t}} \\
\frac{E_{P1*F,t}}{E_{R1*T,t}} & \frac{E_{P1*M,t}}{E_{R1*T,t}} & 0 & 0 & 0 & 0 & \cdots & 0 & 0 \\
0 & 0 & \frac{E_{P2*F,t}}{E_{R2*T,t}} & \frac{E_{P2*M,t}}{E_{R2*T,t}} & \frac{E_{P3*F,t}}{E_{R2*T,t}} & \frac{E_{P3*M,t}}{E_{R2*T,t}} & \cdots & 0 & 0 \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \cdots & \vdots & \vdots \\
0 & 0 & 0 & 0 & 0 & 0 & \cdots & \frac{E_{P47*F,t}}{E_{R8*T,t}} & \frac{E_{P47*M,t}}{E_{R8*T,t}} \\
\frac{E_{P1*F,t}}{E_{R1*F,t}} & 0 & 0 & 0 & 0 & 0 & \cdots & 0 & 0 \\
0 & 0 & \frac{E_{P2*F,t}}{E_{R2*F,t}} & 0 & \frac{E_{P3*F,t}}{E_{R2*F,t}} & 0 & \cdots & 0 & 0 \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \cdots & \vdots & \vdots \\
0 & 0 & 0 & 0 & 0 & 0 & \cdots & \frac{E_{P47*F,t}}{E_{R8*F,t}} & 0 \\
0 & \frac{E_{P1*M,t}}{E_{R1*M,t}} & 0 & 0 & 0 & 0 & \cdots & 0 & 0 \\
0 & 0 & 0 & \frac{E_{P2*M,t}}{E_{R2*M,t}} & 0 & \frac{E_{P3*M,t}}{E_{R2*M,t}} & \cdots & 0 & 0 \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \cdots & \vdots & \vdots \\
0 & 0 & 0 & 0 & 0 & 0 & \cdots & 0 & \frac{E_{P47*M,t}}{E_{R8*M,t}} \\
\frac{E_{P1*F,t}}{E_{P1*T,t}} & \frac{E_{P1*M,t}}{E_{P1*T,t}} & 0 & 0 & 0 & 0 & \cdots & 0 & 0 \\
0 & 0 & \frac{E_{P2*F,t}}{E_{P2*T,t}} & \frac{E_{P2*M,t}}{E_{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{E_{P47*F,t}}{E_{P47*T,t}} & \frac{E_{P47*M,t}}{E_{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}}_{S_t}
\underbrace{\begin{bmatrix}
R_{P1*F,t} \\
R_{P1*M,t} \\
R_{P2*F,t} \\
R_{P2*M,t} \\
\vdots \\
R_{P47*F,t} \\
R_{P47*M,t}
\end{bmatrix}}_{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.

4.2   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 of the required forecasts (Shang and Haberman, 2017). For example, reverting to the Japanese data, we first generate h-step-ahead independent forecasts for the most disaggregated series, namely $\widehat{b}_{n+h} = \left[\widehat{R}_{\text{P1}*\text{F},n+h}, \widehat{R}_{\text{P1}*\text{M},n+h}, \ldots, \widehat{R}_{\text{P47}*\text{F},n+h}, \widehat{R}_{\text{P47}*\text{M},n+h} \right]^{\top}$. Then, we obtain forecasts for all series as

$$\overline{\boldsymbol{R}}_{n+h}^{\text{BU}} = \boldsymbol{S}_{n+h} \widehat{\boldsymbol{b}}_{n+h}, \tag{12}$$

where $\overline{\boldsymbol{R}}_{n+h}^{\text{BU}}$ denotes the reconciled forecasts obtained from the bottom-up method.

The bottom-up method performs well when the bottom-level series have a high signal-to-noise ratio. In contrast, the bottom-up method may lead to inaccurate forecasts of the top-level series, in particular when there are missing or noisy data at the bottom level (see, e.g., Shang and Haberman, 2017; Shang and Hyndman, 2017).

4.3   Optimal-combination method

Instead of considering only the bottom-level series, Hyndman et al. (2011) have proposed the optimal-combination method where 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

$$\widehat{\boldsymbol{R}}_{n+h} = \boldsymbol{S}_{n+h} \boldsymbol{\beta}_{n+h} + \boldsymbol{\epsilon}_{n+h}, \tag{13}$$

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; $\boldsymbol{\beta}_{n+h} = \mathbb{E}[\boldsymbol{b}_{n+h} | \boldsymbol{R}_1, \ldots, \boldsymbol{R}_n]$ is the unknown mean of the independent forecasts of the most disaggregated series; and $\boldsymbol{\epsilon}_{n+h}$ represents the reconciliation errors.

To estimate the regression coefficient, Hyndman et al. (2011) have proposed an ordinary least-squares solution,

$$\widehat{\boldsymbol{\beta}}_{n+h} = \left(\boldsymbol{S}_{n+h}^{\top} \boldsymbol{S}_{n+h} \right)^{-1} \boldsymbol{S}_{n+h}^{\top} \widehat{\boldsymbol{R}}_{n+h}. \tag{14}$$

The revised forecasts are given by

$$\overline{\boldsymbol{R}}_{n+h}^{\text{OLS}} = \boldsymbol{S}_{n+h}\widehat{\boldsymbol{\beta}}_{n+h} = \boldsymbol{S}_{n+h}\left(\boldsymbol{S}_{n+h}^{\top}\boldsymbol{S}_{n+h}\right)^{-1}\boldsymbol{S}_{n+h}^{\top}\widehat{\boldsymbol{R}}_{n+h}. \tag{15}$$

4.4 Forecast combination

Forecast combination involves the computation of weighted means for the forecast. The averaged forecast for horizon h is computed as the weighted mean:

$$\overline{\boldsymbol{R}}_{n+h}^{\text{comb}} = \sum_{g=1}^{G} w_g \overline{\boldsymbol{R}}_{n+h}^{g}, \tag{16}$$

where $\overline{\boldsymbol{R}}_{n+h}^{g}$ denotes forecasts obtained from a grouped forecasting method, such as the bottom-up or optimal-combination method, and $\overline{\boldsymbol{R}}_{n+h}^{\text{comb}}$ represents the model-averaged forecast; and $(w_1, w_2, \ldots, w_G)$ are weights that sum to 1.

The crux of the problem lies in the selection of weights. Recent studies found that past performance information enables the use of unequal weighting combinations, but simple combination methods have been shown to be robust in many settings. Here, we consider two heuristics for combining the midpoint (i.e., point forecast) and endpoints (i.e., interval forecasts) (see Gaba et al., 2017, for other heuristics). First, we combine point and interval forecasts obtained from two grouped forecasting methods with equal weighting. Second, we consider an envelope of prediction intervals and simple averaging of midpoints.

Average (Av). From the viewpoint of combining interval forecasts, $\overline{\boldsymbol{R}}_{n+h,L}^{\text{comb, Av}} = \frac{1}{G}\sum_{g=1}^{G}\overline{\boldsymbol{R}}_{n+h,L}^{g}$ and $\overline{\boldsymbol{R}}_{n+h,U}^{\text{comb, Av}} = \frac{1}{G}\sum_{g=1}^{G}\overline{\boldsymbol{R}}_{n+h,U}^{g}$, where $[\overline{\boldsymbol{R}}_{n+h,L}^{g}, \overline{\boldsymbol{R}}_{n+h,U}^{g}]$ for $g = 1, \ldots, G$ denote the lower and upper bounds for a random variable $\overline{\boldsymbol{R}}_{n+h}$, and G denotes the number of grouped forecasting methods. This heuristic takes a simple average of the midpoint and endpoints. For combining point and interval forecasts, simple averages are often considered as the benchmark because of their simplicity, good performance, and robustness. When the lower and upper bounds are symmetric, Av combination method corresponds to averaging quantiles, which bridges the gap between interval forecast combination and quantile averaging (see, e.g, Gaba et al., 2017).

Envelope (En) of endpoints and simple averaging of midpoints. The individual method tends to be overconfident, so an envelope method combines the prediction intervals by taking an extreme viewpoint. The combined $(\overline{\boldsymbol{R}}_{n+h,L}^{\text{comb, En}}, \overline{\boldsymbol{R}}_{n+h,U}^{\text{comb, En}})$ are obtained by $\overline{\boldsymbol{R}}_{n+h,L}^{\text{comb, En}} = \min(\overline{\boldsymbol{R}}_{n+h,L}^{1}, \ldots, \overline{\boldsymbol{R}}_{n+h,L}^{G})$ and $\overline{\boldsymbol{R}}_{n+h,U}^{\text{comb, En}} = \max(\overline{\boldsymbol{R}}_{n+h,U}^{1}, \ldots, \overline{\boldsymbol{R}}_{n+h,U}^{G})$. This heuristic is conservative but can overcome the overconfident issue, as each method only presents a partial view of reality. The simple average of the lower and upper bounds obtained from G number of

methods provides a combined point forecast, where

$$\overline{\boldsymbol{R}}_{n+h}^{\text{comb, AvInt}} = \frac{1}{G} \sum_{g=1}^{G} \frac{\overline{\boldsymbol{R}}_{n+h,L}^{g} + \overline{\boldsymbol{R}}_{n+h,U}^{g}}{2}. \tag{17}$$

4.5 Forecast exposure-to-risk

Since the bottom-level forecasts are mortality rates, we ought to take into account forecast exposure-to-risk in order to reconcile death counts with respect to the group structure. The observed ratios that form the S_t summing matrix are forecast using the automatic ARIMA algorithm of Hyndman and 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 (see also Shang and Haberman, 2017). For example, let $p_t = \text{E}_{\text{P1}*\text{F},t} / \text{E}_{\text{Japan}*\text{T},t}$ be a non-zero element of S_t. Given that we have observed $\{p_1, \ldots, p_n\}$, an h-step-ahead forecast $\widehat{p}_{n+h}$ can be obtained. The forecasts of ratios between any two exposure-to-risks are used to form the matrix $\boldsymbol{S}_{n+h}$. To ensure summability to 1 in each row of the group structure, every non-zero ratio is normalized by dividing the sum of ratios in each row.

The potential improvement in forecast accuracy of the reconciliation methods partially relies on the accurate forecast of the S matrix. Recall that the S matrix includes ratios of forecast exposure-at-risk. Our cohort assumption is reasonable because it allows us to forecast ratios and populate the S matrix. Shang and Haberman (2017) and Shang and Hyndman (2017) compare point forecast and interval forecast accuracies between the reconciliation methods, with the forecast S matrix and actual holdout S matrix, and found that it is advantageous to use the forecast S matrix.

5 Results

5.1 Multivariate functional time-series model fitting

For the national and sub-national mortality rates, we examine the goodness-of-fit of the proposed multivariate functional time-series method to the observed data. Because the mortality rates for subpopulations may have different mean and variance terms, we standardized age-specific mortality rates by subtracting the mean function and dividing the standard deviation function before implementing a functional principal component analysis to a stacked data matrix of size $n \times (p \times \omega)$. For the Hokkaido data, the selected number of components is one (this is effectively a Lee-Carter type model).

In the first column of Figure 2, we present the mean functions of the female and male smoothed log mortality rates. In the second and third columns, we present the first functional principal component, which accounts for around 88% of the total variation in the joint female and male subpopulations. Given that the principal component scores are surrogates for the original functional

time series, they are forecast to continue to decrease over the next 20 years. We note that the first functional principal component models the female and male mortality data at younger and older ages.
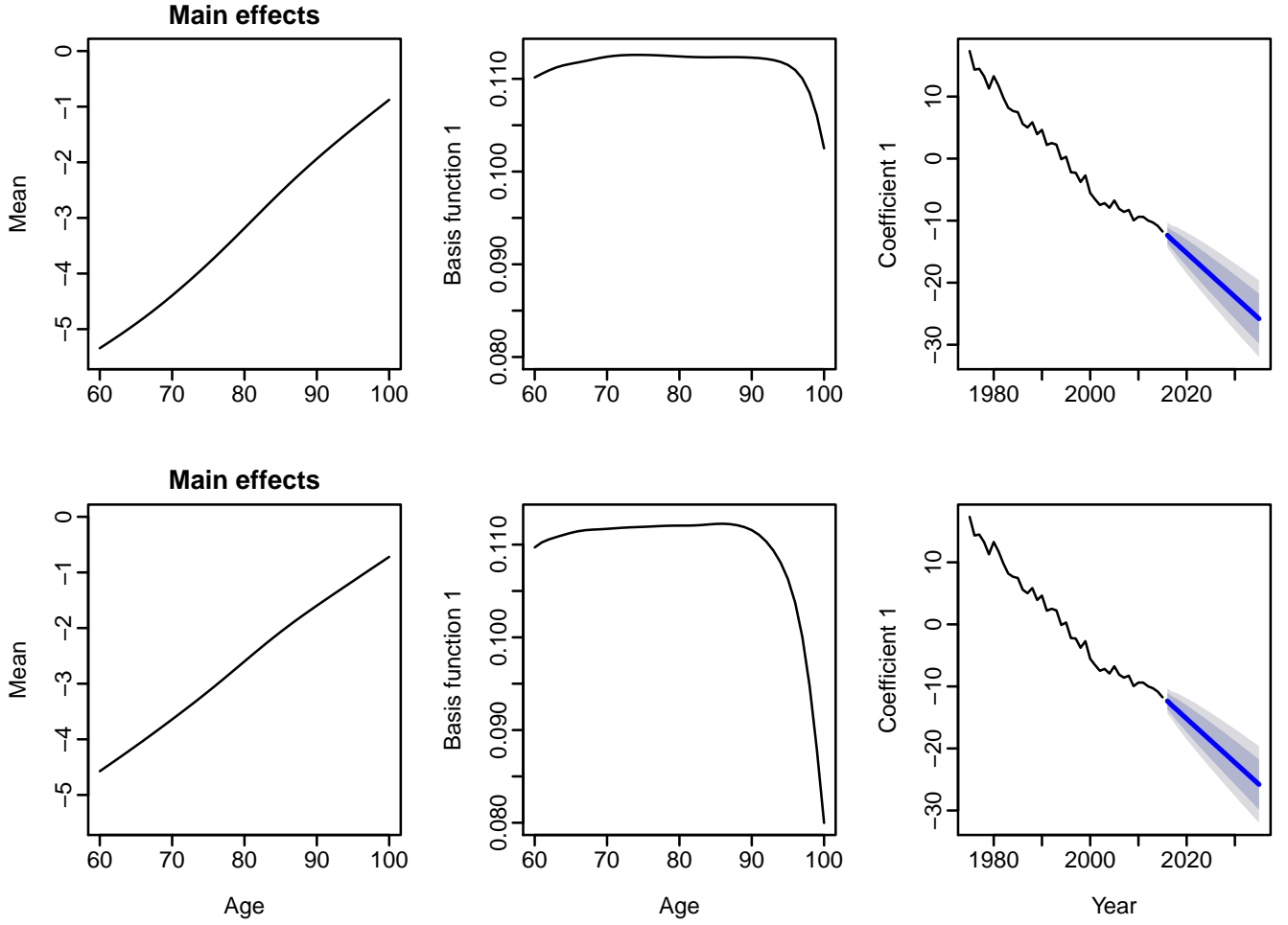

Figure 2: *In the top and bottom panels, multivariate functional principal component decomposition for the female and male smoothed log mortality rates in Hokkaido, respectively. In the third column, the solid blue line represents the point forecasts of principal component scores, where the dark and light grey regions represent the 80% and 95% pointwise prediction intervals, respectively. Standardization was applied to each subpopulation, prior to implementing the multivariate functional principal component analysis.*

We measure goodness-of-fit via a functional version of the R^2 criterion. It is given as

$$R^2 = 1 - \frac{\sum_{t=1}^{n} \int_{x \in \mathcal{I}} \left[\exp^{y_t^{(j)}(x)} - \exp^{\widehat{f}_t^{(j)}(x)} \right]^2 dx}{\sum_{t=1}^{n} \int_{x \in \mathcal{I}} \left[\exp^{y_t^{(j)}(x)} - \exp^{\bar{y}^{(j)}(x)} \right]^2 dx}, \tag{18}$$

where $y_t^{(j)}(x)$ denotes the observed age-specific log mortality rates at each t for the jth subpopulation, $\widehat{f}_t^{(j)}(x)$ denotes the fitted age-specific log mortality rates. The larger the R^2 value is, the better is the goodness-of-fit by a model. It is possible for the R^2 criterion to take negative values. A negative R^2 value implies that the fitted model may not well explain the raw data that are likely to contain a large amount of measurement errors. From a negative R^2 value, we can quantify the amount of measurement errors exhibited in a data set and the degree of smoothing that the raw

mortality data require.

In Table 2, we report the goodness-of-fit of the univariate and multivariate functional time-series methods, as measured by the R^2 criterion defined in (18). For both female and male series, the multivariate functional time-series method generally achieves a better goodness-of-fit result than the univariate functional time-series method. The superiority of the multivariate functional time series method is because it can incorporate correlation between multiple series. This correlation provides additional information that is not possessed by the univariate functional time series method.

Table 2: *Goodness-of-fit as measured by the R^2 criterion for each national and sub-national female and male age-specific mortality rates in Japan. Let FTS denote functional time series. For each series and sex, we highlight in bold the method with a higher R^2 value.*

Series	Univariate FTS Female	Male	Multivariate FTS Female	Male	Series	Univariate FTS Female	Male	Multivariate FTS Female	Male
Japan	0.939	0.802	**0.964**	**0.902**	Mie	0.399	0.110	**0.453**	**0.208**
Hokkaido	0.624	0.262	**0.625**	**0.369**	Shiga	0.250	-0.044	**0.330**	**-0.023**
Aomori	0.172	0.006	**0.202**	**0.040**	Kyoto	0.518	0.121	**0.577**	**0.162**
Iwate	0.309	0.082	**0.364**	**0.160**	Osaka	0.692	0.246	**0.740**	**0.288**
Miyagi	**0.242**	0.051	0.203	**0.087**	Hyogo	0.638	0.213	**0.649**	**0.327**
Akita	**0.269**	-0.126	0.231	**-0.124**	Nara	0.296	-0.013	**0.329**	**-0.013**
Yamagata	0.283	**-0.015**	**0.310**	-0.022	Wakayama	0.428	**-0.051**	**0.494**	-0.061
Fukushima	0.370	0.050	**0.404**	**0.079**	Tottori	0.215	-0.073	**0.278**	**-0.073**
Ibaraki	0.526	0.094	**0.559**	**0.209**	Shimane	0.413	-0.006	**0.473**	**0.032**
Tochigi	0.398	**0.017**	**0.447**	-0.005	Okayama	0.500	-0.011	**0.520**	**0.050**
Gunma	0.330	0.044	**0.332**	**0.081**	Hiroshima	0.624	0.164	**0.710**	**0.268**
Saitama	0.477	0.114	**0.497**	**0.129**	Yamaguchi	0.412	0.060	**0.473**	**0.129**
Chiba	0.568	**0.079**	**0.597**	-0.006	Tokushima	0.410	0.086	**0.420**	**0.144**
Tokyo	0.615	0.330	**0.675**	**0.430**	Kagawa	0.276	0.022	**0.293**	**0.120**
Kanagawa	0.548	0.134	**0.589**	**0.251**	Ehime	0.329	0.099	**0.375**	**0.205**
Niigata	0.530	0.172	**0.636**	**0.179**	Kochi	0.443	0.035	**0.483**	**0.090**
Toyama	0.271	0.081	**0.280**	**0.085**	Fukuoka	0.577	0.198	**0.603**	**0.332**
Ishikawa	0.159	-0.009	**0.232**	**0.009**	Saga	0.254	-0.046	**0.294**	**-0.017**
Fukui	0.317	-0.088	**0.363**	**-0.057**	Nagasaki	0.371	-0.003	**0.395**	**-0.001**
Yamanashi	0.339	**-0.009**	**0.352**	-0.061	Kumamoto	0.550	0.116	**0.581**	**0.213**
Nagano	0.374	0.128	**0.453**	**0.151**	Oita	0.413	0.159	**0.424**	**0.246**
Gifu	0.446	0.146	**0.500**	**0.214**	Miyazaki	0.377	0.113	**0.407**	**0.169**
Shizuoka	0.554	0.077	**0.629**	**0.236**	Kagoshima	0.526	0.126	**0.580**	**0.248**
Aichi	0.510	0.170	**0.572**	**0.219**	Okinawa	0.488	0.123	**0.551**	**0.143**

Based on the observed mortality rates from 1975 to 2016, we produce the 20-years-ahead point

forecasts of female and male age-specific mortality rates from 2017 to 2036. As shown in Figure 3 for the case of Hokkaido, the age-specific mortality rates are continuing to decline, and the forecast female mortality rates are likely to be lower than the male mortality rates.
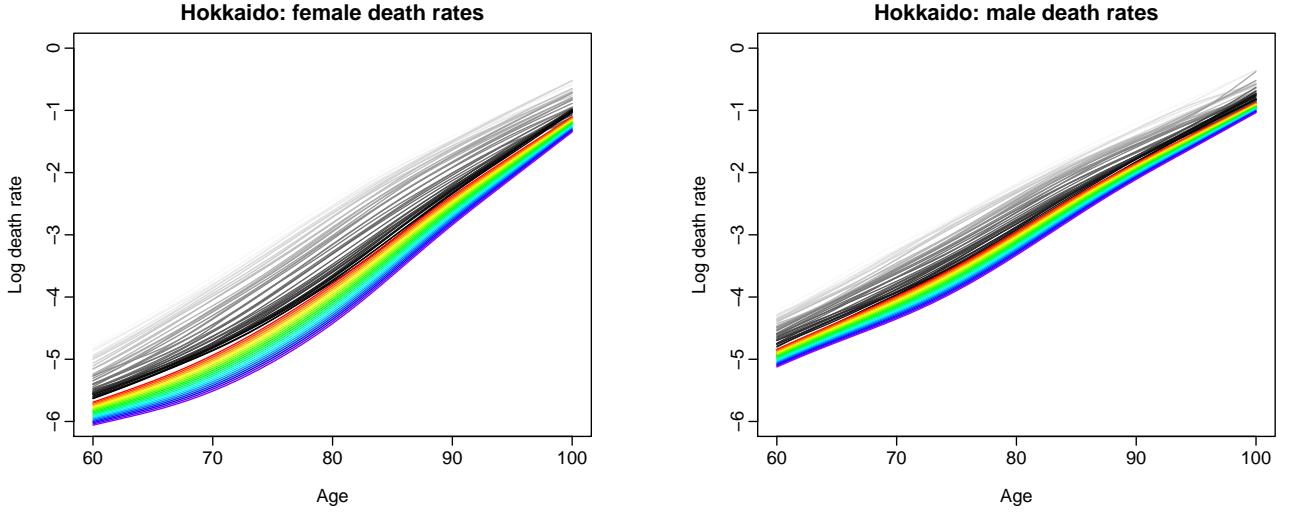

Figure 3: *Point forecasts of age-specific female and male log mortality rates from 2017 to 2036. The observed functional time series is shown in gray color palette, and the forecasts are highlighted in rainbow color palette.*

5.2 Point forecast evaluation

Using the first 27 observations from 1975 to 2001 in the Japanese age-specific mortality rates, we produce one- to 15-step-ahead point forecasts. Through an expanding window approach, we re-estimate the parameters in the time series forecasting models using the first 28 observations from 1975 to 2002. Point forecasts from the estimated models are then produced for one- to 14-step-ahead. We repeat this process by increasing the sample size by one year until reaching the end of the data period in 2016. This process produces 15 one-step-ahead forecasts, 14 two-step-ahead forecasts, . . . , and one 15-step-ahead forecast. By comparing these forecasts with the holdout samples, we evaluate the out-of-sample point forecast bias and accuracy.

To assess the point forecast bias, we consider the mean forecast error (MFE). For each series j, MFE can be written as

$$\text{MFE}_j(h) = \frac{1}{41 \times (16 - h)} \sum_{\zeta=h}^{15} \sum_{i=1}^{41} \left[y_{n+\zeta}^{(j)}(x_i) - \widehat{y}_{n+\zeta}^{(j)}(x_i) \right], \tag{19}$$

where h denotes forecast horizon, $y_{n+\zeta}^{(j)}(x_i)$ denotes the actual holdout sample for the ith age and ζ^{th} curve in the jth series, while $\widehat{y}_{n+\zeta}^{(j)}(x_i)$ denotes the point forecasts for the holdout sample. To assess the point forecast accuracy, we use the mean absolute forecast error (MAFE) defined as

$$\text{MAFE}_j(h) = \frac{1}{41 \times (16 - h)} \sum_{\zeta=h}^{15} \sum_{i=1}^{41} \left| y_{n+\zeta}^{(j)}(x_i) - \widehat{y}_{n+\zeta}^{(j)}(x_i) \right|. \tag{20}$$

13

Since the number of series is different across different levels of disaggregation, we obtain an overall assessment of point forecast bias and accuracy by taking the simple average of the error measures across the number of series at each level. With 15 different forecast horizons, we consider the mean and median values to evaluate overall point forecast bias and accuracy between the two functional time-series methods for national and sub-national mortality forecasts.

5.3 Comparisons of point forecast bias and accuracy

Averaging over all the series at each level of the group structure, Figure 4 compares MFE(h) between the univariate and multivariate functional time-series methods. As measured by the MFE, the multivariate functional time-series method generally produces more accurate point forecast bias than the ones obtained using the univariate functional time-series method. Using the multivariate functional time-series method, the smallest bias can be achieved by using the forecast combination method with simple averaging of midpoints at each level of the group structure. The superior forecast accuracy of the multivariate functional time-series method over the univariate functional time-series method stems from two sources: (1) the joint modeling of age-specific mortality patterns among multiple subpopulations; and (2) the joint forecasting of age-specific mortality rates among multiple subpopulations. The advantage of this forecast combination is that it can reduce bias.
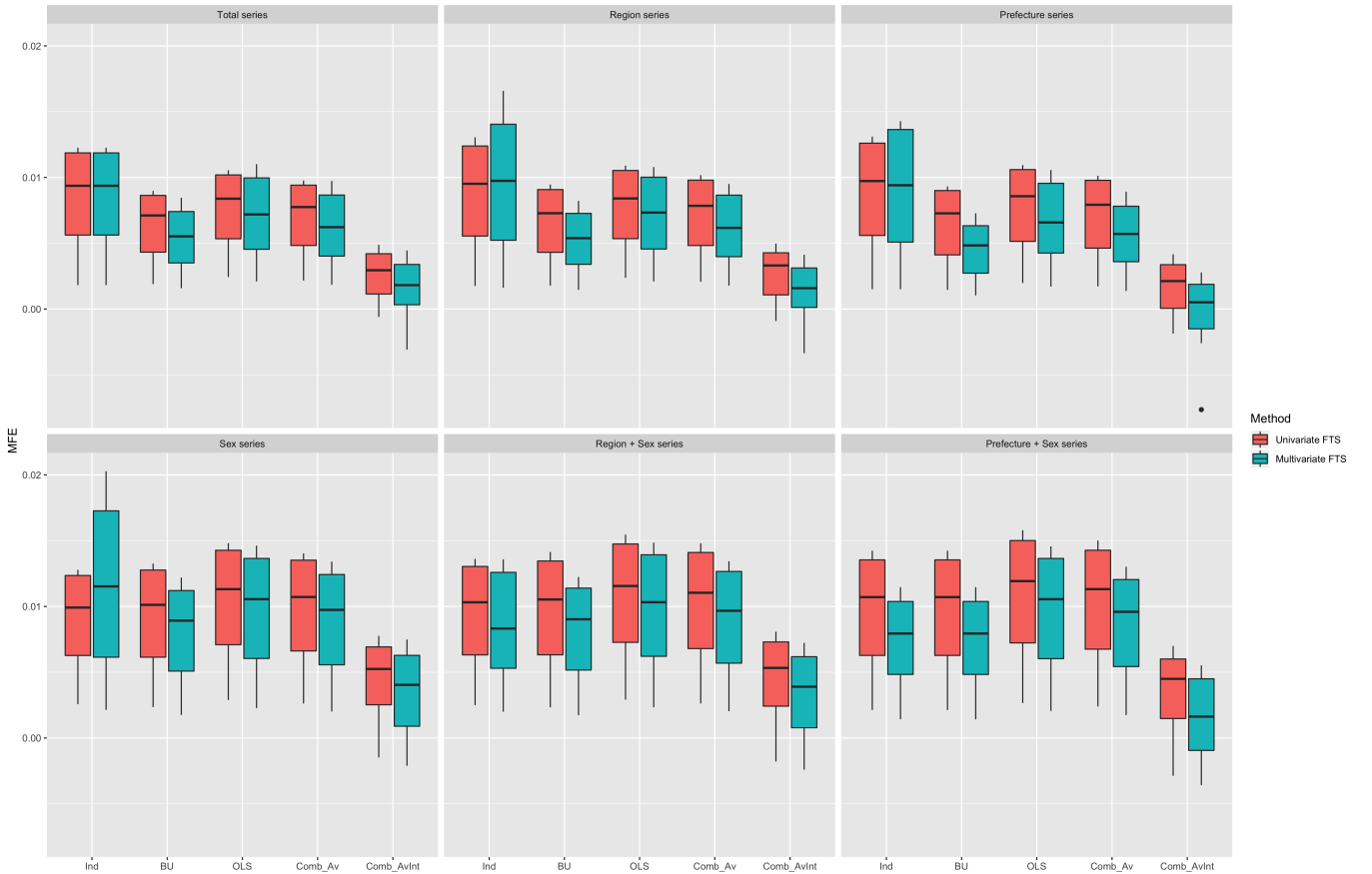

Figure 4: *MFE in the holdout sample between the univariate and multivariate functional time-series methods applied to the Japanese age-specific mortality rates.*

Figure 5 similarly compares MAFE(*h*) between the univariate and multivariate functional time-series methods. The multivariate functional time-series method generally produces smaller point forecast errors than the ones obtained using the univariate functional time-series method for almost all levels of the group structure. Based on the two summary statistics of the forecast errors, the multivariate functional time-series method coupled with the forecast combination method (AvInt) performs the best across all levels of the group structure, while it reconciles point forecasts taking account of the group structure. The advantage of this forecast combination is that it can reduce point forecast errors.
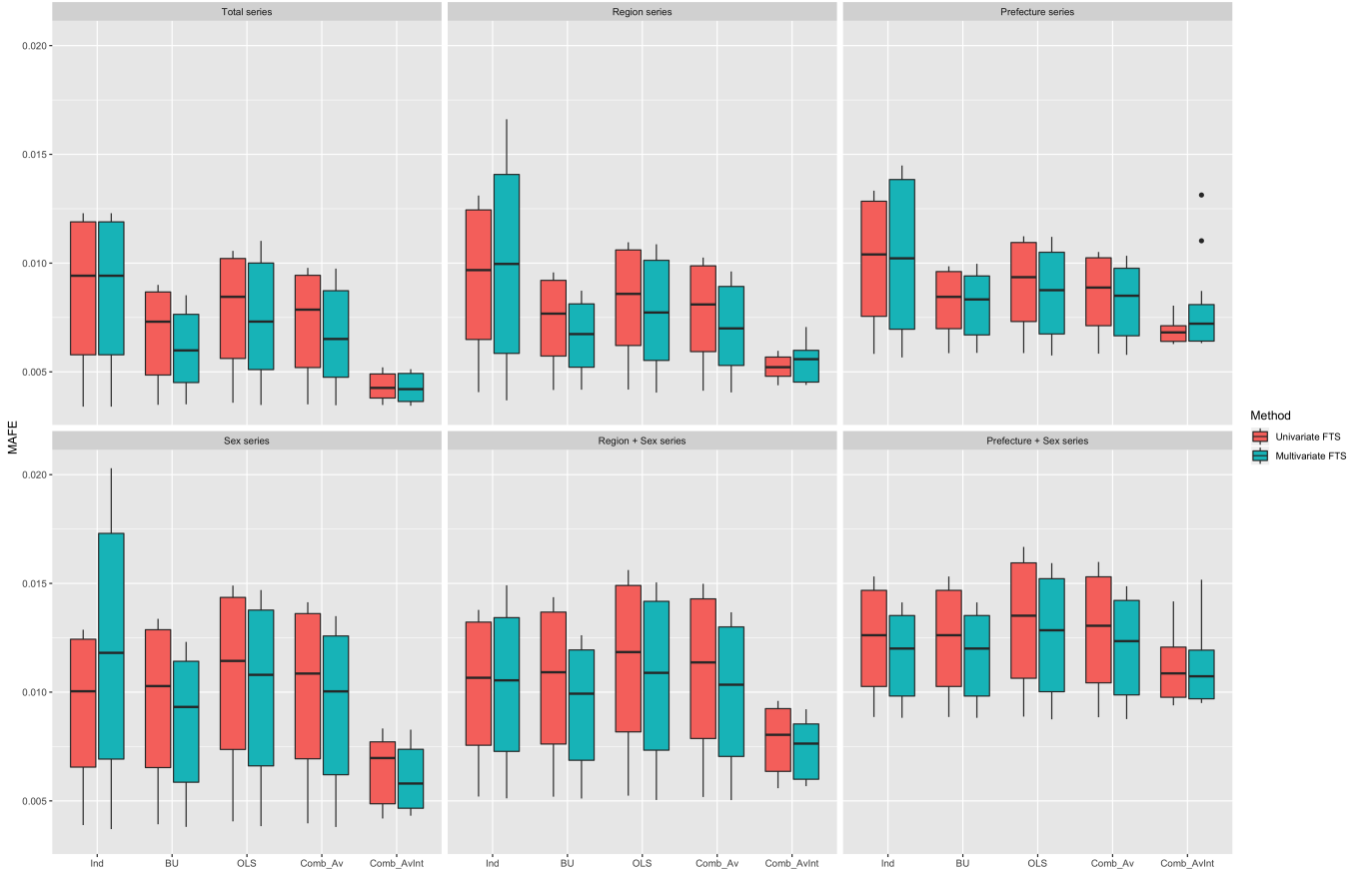

Figure 5: *MAFE in the holdout sample between the univariate and multivariate functional time-series methods applied to the Japanese age-specific mortality rates.*

5.4 Interval forecast construction and evaluation

We apply the method of Shang (2020) to construct pointwise prediction intervals. To evaluate pointwise interval forecast accuracy, we first construct lower and upper bounds of a prediction interval at the $100(1 - \alpha)\%$ nominal coverage probability, where α denotes a significance level (see, Shang and Haberman, 2017, for details). Then, we utilize the interval score of Gneiting and Raftery (2007). In the literature, extensive works are available on interval score and will not be reiterated here. The optimal interval score is achieved when there is almost $100(1 - \alpha)\%$ of times that the holdout data lie between the upper and lower bounds of the prediction interval, and the distance between the upper and lower bounds is minimal.

15

Since the number of series is different across different levels of disaggregation, we obtain an overall assessment of interval forecast accuracy by taking the simple average of the interval scores across the number of series at each level. Also, for 15 different forecast horizons, we consider the mean and median values to evaluate overall interval forecast accuracy among the methods for national and sub-national mortality forecasts.

5.5 Comparison of interval forecast accuracy

Averaging over all the series at each level of the group structure, Figure 6 presents the mean interval scores $\overline{S}_\alpha(h)$ between the univariate and multivariate functional time-series methods. Based on the averaged summary statistics of $\overline{S}_\alpha(h)$, the independent forecasting method generally performs the best because it fits each series without the constraint of a hierarchy. At the prefecture level, the optimal-combination method outperforms the independent functional time series forecasting method, which demonstrates the improved interval-forecast accuracy of the optimal-combination method while reconciling interval forecasts. Between the univariate and multivariate functional time-series methods, there is a slight advantage to use the multivariate functional time-series method at each level of the group structure. The forecast combination method with equal weighting produces mean interval scores that lie between the ones of the bottom-up and optimal-combination methods. The forecast combination method with simple averaging of midpoints produces the largest mean interval scores and thus this method is conservative.
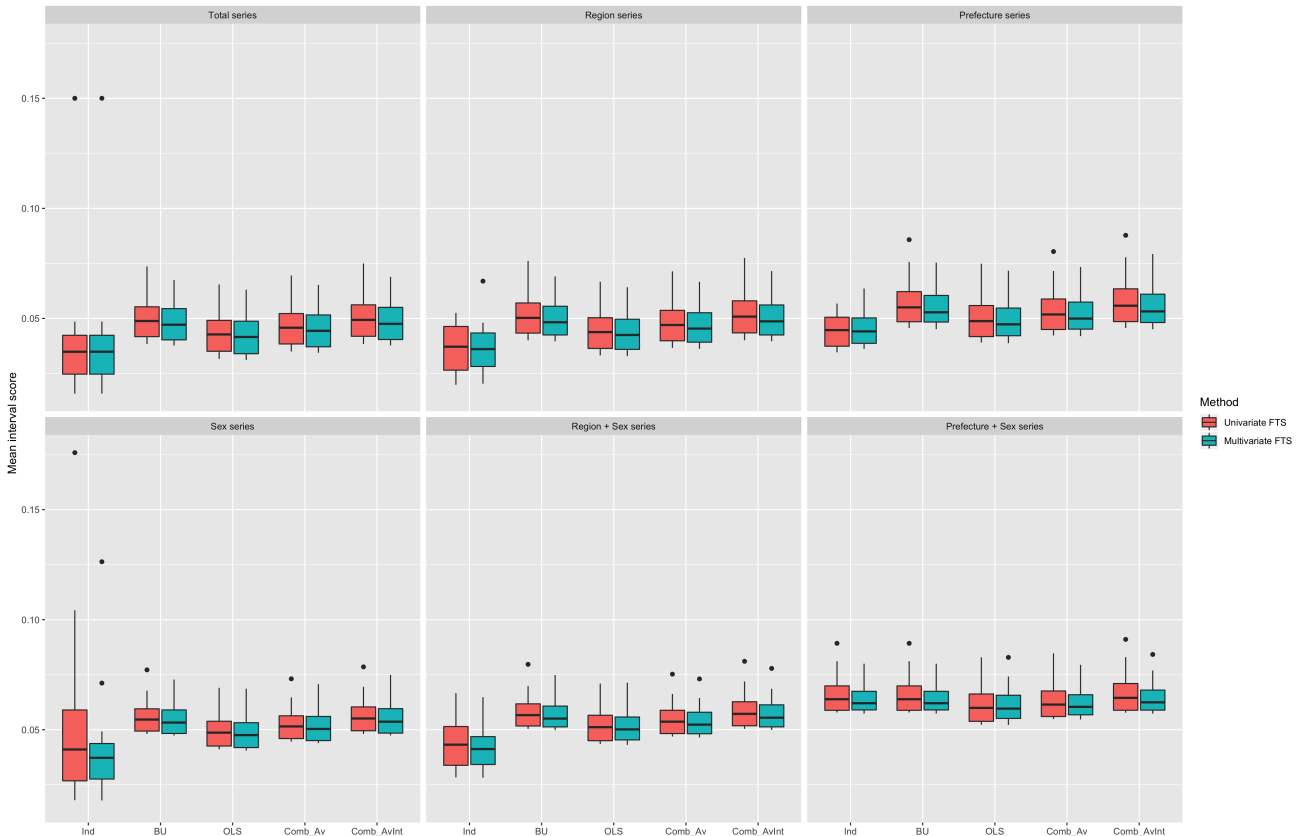

Figure 6: *Mean interval score in the holdout sample between the univariate and multivariate functional time-series methods applied to the Japanese age-specific mortality rates.*

6 Application to the pricing of temporary life annuities

An important use of mortality forecasts for elderly is in the pension and insurance industries, whose profitability and solvency rely on accurate mortality forecasts so that longevity risk can be appropriately hedged and pensions and annuities can be accurately valued. When a person reaches retirement age, an optimal way of guaranteeing one individual's financial income in retirement is to purchase an annuity (as demonstrated by Yaari, 1965). An annuity is a contract offered by insurers guaranteeing a steady stream of payments for the lifetime of the annuitant in exchange for an initial premium fee.

Lifetime immediate annuities, where rates are locked in for life, have been shown to deliver poor value for money (i.e., they may be expensive for the purchaser: see for example Cannon and Tonks, 2008, Chapter 6). In many countries selling annuities, sales of temporary annuities greatly exceed that of lifetime annuities, so this is where a major interest lies in terms of application. These temporary life annuities pay a pre-determined and guaranteed level of income which is often higher than the level of income provided by a lifetime annuity for a similar premium. Temporary annuities offer an alternative to lifetime annuities and allow the purchaser the option of also buying a deferred annuity at a later date.

We apply the mortality forecasts to the calculation of a temporary life annuity (see Dickson et al., 2009, p.114), and we adopt a cohort approach to the calculation of the survival probabilities. For a single cohort, the τ year survival probability of a person aged x currently at $t = 0$ (or year 2016) is determined by

$$\tau p_x = \prod_{\varpi=1}^{\tau} {}_1 p_{x+\varpi-1} \tag{21}$$

$$= \prod_{\varpi=1}^{\tau} \exp^{-m_{x+\varpi-1,\varpi-1}} . \tag{22}$$

The survival probability is a random variable given that age-specific mortality rates for $\varpi = 1, \ldots, \tau$ are forecasts obtained by the multivariate functional time-series method. Here, we assume that the central mortality rates are constant throughout each one-year period (see also Shang and Haberman, 2017).

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

$$a_x^T(\boldsymbol{m}_{1:T}^x) = \sum_{\tau=1}^{T} B(0,\tau) \mathbb{E}\left(1_{T_x > \tau} | \boldsymbol{m}_{1:\tau}^x\right) \tag{23}$$

$$= \sum_{\tau=1}^{T} B(0,\tau) {}_\tau p_x\left(\boldsymbol{m}_{1:\tau}^x\right), \tag{24}$$

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 et al., 2017). For the purposes of pricing and risk management, it is vital to produce an accurate forecast of the survival curve $_\tau p_x$ that best captures the mortality experience of a portfolio.

In Table 3, to provide an example of the annuity calculations, we compare the best estimate of the annuity prices for different ages and maturities produced by the three forecasting methods for a female policyholder residing in Region 2. We assume a constant interest rate at $\eta = 3\%$ so that the τ-year bond price is given by $B(0, \tau) = \exp^{-\eta\tau}$.

Table 3: *Estimates of temporary life annuity prices with different ages and maturities (T) for a female policyholder residing in Region 2. These estimates are based on the forecast age-specific mortality rates from 2017 to 2057, obtained from the independent and grouped multivariate functional time-series forecasting methods. We consider contracts with different ages and maturities, so that age + maturity $\leq$ 100.*

Method	$T = 5$	$T = 10$	$T = 15$	$T = 20$	$T = 25$	$T = 30$
			age = 60			
Base	4.5185	8.2736	11.3152	13.6552	15.2718	16.1779
BU	4.5192	8.2976	11.3969	13.8402	15.6092	16.6864
OLS	4.5244	8.3238	11.4710	14.0044	15.9177	17.1773
Comb, Av	4.5218	8.3107	11.4339	13.9218	15.7614	16.9261
Comb, AvInt	4.5272	8.3383	11.5087	14.0769	15.8899	16.5369
			age = 65			
Base	4.4672	8.0856	10.8693	12.7925	13.8705	14.2941
BU	4.4873	8.1681	11.0698	13.1707	14.4500	15.0201
OLS	4.5011	8.2297	11.2310	13.4978	14.9900	15.7458
Comb, Av	4.4942	8.1988	11.1499	13.3321	14.7136	15.3698
Comb, AvInt	4.5089	8.2600	11.2984	13.4434	14.2089	14.4340
			age = 70			
Base	4.3903	7.7677	10.1012	11.4091	11.9231	12.0420
BU	4.4284	7.9194	10.4470	11.9862	12.6720	12.8564
OLS	4.4590	8.0483	10.7592	12.5437	13.4476	13.7413
Comb, Av	4.4437	7.9835	10.6010	12.2581	13.0452	13.2779
Comb, AvInt	4.4716	8.0939	10.6509	11.5635	11.8318	11.9313

We highlight that some of the annuity prices are very sensitive to the mortality forecasts obtained from the forecasting method. For example, the annuity prices varies from 11.4091 to 12.5437 for 20-year annuity at age 70. Thus, it is important to compare forecast accuracy of various forecasting methods and provide recommendations as stated in the conclusion.

To measure forecast uncertainty, we obtain the bootstrapped forecasts of the age-specific mortality rates, derive the survival probabilities and calculate the corresponding annuity prices associated with different ages and maturities. For instance, we construct one-step-ahead to 15-step-ahead bootstrapped forecasts of the age-specific mortality rates, derive the bootstrap survival probabilities and calculate the bootstrap prices of temporary life annuities. In Table 4, we present the 95% pointwise prediction intervals of the prices of temporary life annuities for different ages and maturities, where age + maturity $\leq$ 75.

Table 4: *The 95% pointwise prediction intervals of temporary life annuity prices with different ages and maturities (T) for female policyholder residing in Region 2, for example. These estimates are based on the one-step-ahead to 15-step-ahead forecast mortality rates from 2017 to 2031, obtained from the independent and grouped multivariate functional time-series method. We only consider contracts with maturity so that age + maturity $\leq$ 75. If age + maturity > 75, NA will be shown in the table.*

Method	$T = 5$	$T = 10$	$T = 15$
	age = 60		
Base	(4.5214, 4.5343)	(8.3083, 8.3722)	(11.4147, 11.6110)
BU	(4.5174, 4.5371)	(8.2961, 8.3802)	(11.3923, 11.6242)
OLS	(4.5192, 4.5360)	(8.3037, 8.3775)	(11.4118, 11.6219)
Comb, Av	(4.5183, 4.5365)	(8.2999, 8.3789)	(11.4021, 11.6231)
Comb, En	(4.5169, 4.5370)	(8.2939, 8.3800)	(11.3854, 11.6260)
	age = 65		
Base	(4.4916, 4.5271)	(8.1760, 8.3476)	NA
BU	(4.4885, 4.5289)	(8.1662, 8.3518)	NA
OLS	(4.4918, 4.5288)	(8.1809, 8.3536)	NA
Comb, Av	(4.4901, 4.5288)	(8.1735, 8.3527)	NA
Comb, En	(4.4873, 4.5289)	(8.1602, 8.3544)	NA
	age = 70		
Base	(4.4256, 4.5189)	NA	NA
BU	(4.4226, 4.5191)	NA	NA
OLS	(4.4300, 4.5212)	NA	NA
Comb, Av	(4.4263, 4.5201)	NA	NA
Comb, En	(4.4193, 4.5217)	NA	NA

Accuracy in pricing is not the focus here as we recognize that variations exist in the approaches adopted to pricing in practice. In particular, the assumption made concerning adverse selection and the mortality experience of purchasers of annuities is critical (see Cannon and Tonks, 2008, for a fuller discussion). For example, O'Meara and Bruhn (2013) assumed, in an Australian context, that for voluntary purchase of annuities, annuitant mortality was assumed to be 30% lower than population mortality at age 60 and 20% lower than population mortality from age 90 onwards, with linear interpolation between these ages, based on the advice from life insurance actuarial consultants.

7 Conclusion

Using the national and sub-national Japanese age-specific mortality rates, we evaluate and compare the point forecast bias and accuracy between the univariate and multivariate functional time-series methods. Based on the forecast bias and accuracy criteria, we show that the proposed multivariate functional time-series method outperforms the univariate functional time-series method used in Shang and Haberman (2017). The superiority of the multivariate functional time-series method is primarily driven by the ability to incorporate correlation among the subpopulations.

We compare the one-step-ahead to 15-step-ahead point forecast accuracy between the independent and the two grouped univariate and multivariate functional time-series forecasting methods. By using the multivariate functional time-series method to produce base forecasts, we consider forecast reconciliation by applying two grouped functional time-series forecasting methods, namely the bottom-up and optimal-combination methods (see also Shang and Haberman, 2017). Through a forecast combination approach, we consider two weight selection methods. The grouped multivariate functional time-series forecasting methods produce more accurate point forecasts than those obtained by the independent multivariate functional time-series forecasting method, averaged over all levels of the group structure. Also, the grouped multivariate functional time-series forecasting methods produce forecasts that obey the natural group structure, thus giving forecast mortality rates at the sub-national levels that add up to the forecast mortality rates at the national level. Between the two grouped multivariate functional time-series forecasting methods and their forecast combination methods, the forecast combination with simple averaging of midpoints is recommended for producing point forecasts while the optimal-combination method is recommended for producing interval forecasts at the prefecture level for the data that we have considered.

We also apply the independent and the two grouped multivariate functional time-series methods to forecast age-specific mortality rates from 2016 to 2056. We then calculate the cumulative survival probability and obtain the prices of temporary life annuities. As expected, we find that the cumulative survival probability has a pronounced impact on annuity prices. Although temporary life annuity prices do not differ significantly for the mortality forecasts obtained by the four methods, mispricing could have a dramatic effect on a portfolio of annuity contracts. To assess forecast uncertainty, we obtain one-step-ahead to 15-step-ahead forecasts of age-specific mortality rates, derive their survival probabilities and calculate their annuity prices for various ages and maturities.

There are several ways in which this paper can be extended, and we briefly outline seven:

1) Subject to the availability of data, the group structure can be disaggregated more finely by considering different causes of death (Gaille and Sherris, 2015) or socioeconomic status (Villegas and Haberman, 2014).

2) We may consider other multiple population forecasting methods, in particular non-linear forecasting methods, such as neural networks (Richman and Wüthrich, 2020*a,b*).

3) In the Japanese data example, the female and male series are jointly modeled and forecast within each region or each prefecture. The total series are jointly modeled and forecast among regions or prefectures. It may be interesting to explore other combinations for modeling and forecasting multiple subpopulations.

4) A weighted least squares method could be used to estimate the regression coefficient in the optimal combination method (see, e.g., Wickramasuriya et al., 2019).

5) In the forecast combination approach, other grouped forecasting methods and choices of weight selections may be considered.

6) While the functional principal component analysis extracts latent component on the basis of explained variance, we may consider other dimension reduction methods on the basis of autocorrelation, such as the maximum autocorrelation factor and predictive factor decompositions.

7) Finally, the methodology can be applied to calculate prices for other types of annuity product, such as the whole-life immediate annuity or deferred annuity.

Acknowledgments

SUPPLEMENTARY MATERIAL

Code for grouped multivariate functional time-series forecasting The R code to produce point and interval forecasts from the independent and the two grouped univariate and multivariate functional time-series forecasting methods described in the article. (R_code.R)

Code for Shiny application The R code to produce a Shiny user interface for plotting every series in the Japanese data group structure. (shiny.zip)

References

Bates, J. M. and Granger, C. W. J. (1969), 'The combination of forecasts', *Operational Research Quarterly* **20**(4), 451–468.

Cannon, E. and Tonks, I. (2008), *Annuity Markets*, Oxford University Press, Oxford.

Dickson, D. C. M., Hardy, M. R. and Waters, H. R. (2009), *Actuarial Mathematics for Life Contingent Risks*, Cambridge University Press, Cambridge.

Fung, M. C., Peters, G. W. and Shevchenko, P. V. (2017), 'A unified approach to mortality modelling using state-space framework: Characterisation, identification, estimation and forecasting', *Annals of Actuarial Science* **11**(2), 343–389.

Gaba, A., Tsetlin, I. and Winkler, R. L. (2017), 'Combining interval forecasts', *Decision Analysis* **14**(1), 1–74.

Gaille, S. A. and Sherris, M. (2015), 'Causes-of-death mortality: What do we know on their dependence?', *North American Actuarial Journal* **19**(2), 116–128.

Gneiting, T. and Raftery, A. E. (2007), 'Strictly proper scoring rules, prediction and estimation', *Journal of the American Statistical Association* **102**(477), 359–378.

Hatzopoulos, P. and Haberman, S. (2013), 'Common mortality modelling and coherent forecasts - an empirical analysis of worldwide mortality data', *Insurance: Mathematics and Economics* **52**(2), 320–337.

Hyndman, R. J., Ahmed, R. A., Athanasopoulos, G. and Shang, H. L. (2011), 'Optimal combination forecasts for hierarchical time series', *Computational Statistics and Data Analysis* **55**(9), 2579–2589.

Hyndman, R. J. and Khandakar, Y. (2008), 'Automatic time series forecasting: the forecast package for R', *Journal of Statistical Software* **27**(3).

Hyndman, R. J. and Ullah, M. S. (2007), 'Robust forecasting of mortality and fertility rates: A functional data approach', *Computational Statistics & Data Analysis* **51**(10), 4942–4956.

Japanese Mortality Database (2019), *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, 2018.

Lee, R. D. and Carter, L. R. (1992), 'Modeling and forecasting U.S. mortality', *Journal of the American Statistical Association* **87**(419), 659–671.

Li, H., Li, H., Lu, Y. and Panagiotelis, A. (2019), 'A forecast reconciliation approach to cause-of-death mortality modeling', *Insurance: Mathematics and Economics* **86**, 122–133.

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.

Olivieri, A. and Pitacco, E. (2016), 'Frailty and risk classification for life annuity portfolios', *Risks* **4**(4), 39.

O'Meara, T. and Bruhn, A. (2013), 'Compulsory Annuitisation: A policy option for Australia?', *Australasian Accounting Business and Finance Journal* **7**(3), 5–30.

Richman, R. and Wüthrich, M. V. (2020*a*), 'A neural network extension of the Lee-Carter model to multiple populations', *Annals of Actuarial Science* **in press**.

Richman, R. and Wüthrich, M. V. (2020*b*), Lee and Carter go machine learning: Recurrent neural networks, Working paper, ETH Zurich. Available at SSRN: https://papers.ssrn.com/abstract=3441030.

Shang, H. L. (2020), 'Dynamic principal component regression for forecasting functional time series in a group structure', *Scandinavian Actuarial Journal* **in press**.

Shang, H. L., Booth, H. and Hyndman, R. J. (2011), 'Point and interval forecasts of mortality rates and life expectancy: A comparison of ten principal component methods', *Demographic Research* **25**(5), 173–214.

Shang, H. L. and Haberman, S. (2017), 'Grouped multivariate and functional time series forecasting: An application to annuity pricing', *Insurance: Mathematics and Economics* **75**, 166–179.

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

Villegas, A. M. and Haberman, S. (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.

Wickramasuriya, S. L., Athanasopoulos, G. and Hyndman, R. J. (2019), 'Optimal forecast reconciliation for hierarchical and grouped time series through trace minimization', *Journal of the American Statistical Association: Theory and Methods* **114**(526), 804–819.

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