Journal of the Royal Statistical Society Series A: Statistics in Society, 2025, **00**, 1–24
https://doi.org/10.1093/jrsssa/qnaf144

ROYAL STATISTICAL SOCIETY
DATA | EVIDENCE | DECISIONS
Journal of the Statistics Society
Series A
Statistics in Society

Original Article

Forecasting high-dimensional functional time series with dual-factor structures

Chen Tang[1,†], Han Lin Shang[2], Yanrong Yang[1] and Yang Yang[3]

[1]Research School of Finance, Actuarial Studies and Statistics, Australian National University, Canberra, ACT 2601, Australia
[2]Department of Actuarial Studies and Business Analytics, Macquarie University, Macquarie Park, NSW 2109, Australia
[3]School of Information and Physical Sciences, The University of Newcastle, Callaghan, NSW 2308, Australia

Address for correspondence: Chen Tang, Research School of Finance, Actuarial Studies and Statistics, Australian National University, Level 4, Building 26C, Kingsley Street, Acton, Canberra, ACT 2601, Australia.
Email: chen.tang@anu.edu.au

Abstract

We propose a dual-factor model for high-dimensional functional time series (HDFTS) that considers multiple populations. The HDFTS is first decomposed into a collection of functional time series (FTS) in a lower dimension and a group of population-specific basis functions. The system of basis functions describes cross-sectional heterogeneity, while the reduced-dimension FTS retains most of the information common to multiple populations. The low-dimensional FTS is further decomposed into a product of common functional loadings and a matrix-valued time series that contains the most temporal dynamics embedded in the original HDFTS. The proposed general-form dual-factor structure is connected to several commonly used functional factor models. We demonstrate the finite-sample performances of the proposed method in recovering cross-sectional basis functions and extracting common features using simulated HDFTS. An empirical study shows that the proposed model produces more accurate point and interval forecasts for subnational age-specific mortality rates in Japan. The financial benefits associated with the improved mortality forecasts are translated into a life annuity pricing scheme.

Keywords: age-specific mortality forecasting, factor models, functional panel data, functional principal component analysis, multilevel functional data

1 Introduction

The increase in human life expectancy poses risks to governments, healthcare systems, and the life insurance industry. Government stakeholders and industry professionals require accurate and easily understandable mortality forecasting methods to assess the risks associated with extended longevity. Modelling and forecasting age-specific mortality rates have become an ongoing research endeavour, particularly noting the seminal work of (Lee & Carter, 1992), which advances the field of research. However, the Lee-Carter model and many of its extensions and modifications (see, e.g. Li et al., 2013; Renshaw & Haberman, 2003; Wiśniowski et al., 2015) consider only a single population. A common disadvantage of such univariate mortality modelling methods is that they tend to produce divergent forecasts for a group of related populations over relatively long forecast horizons. As a remedy, coherent mortality forecasting methods (see, e.g. Li, 2013; Li & Lee, 2005 ; Pampel, 2005) that simultaneously consider multiple populations have gained popularity in recent years. This is because jointly modelling multiple closely related mortality series

† RingGold ID: 2219

Received: May 20, 2024. **Revised:** September 4, 2025. **Accepted:** September 5, 2025

facilitates the pooling of common information from the considered populations, resulting in improved forecasts (see, e.g. Jiménez-Varón et al., 2024, 2025; Shang, 2016).

However, the aforementioned multivariate mortality modelling methods cannot handle high-dimensional multi-population mortality data. When working with mortality data in which the number of populations (cross-sectional units) is comparable to, or even more significant than, the observation years (sample size of FTS), many conventional statistical tools tend to fail, and high-dimensional data analysis techniques are needed (see, e.g. Bühlmann & van de Geer, 2011; Cai & Chen, 2010). In addition, various populations (e.g. different countries) may report mortality rates over different time intervals. Limiting mortality observations to the same time is often necessary before applying conventional matrix-based multivariate mortality models (e.g. the extended Lee-Carter model of Li et al., 2013).

To address these issues, we propose modelling multi-population mortality time series within a functional data analysis framework, utilizing high-dimensional data analysis techniques. In this context of functional data analysis, age-specific mortality rates in any given year can be considered a curve. Hence, the curves relate to any particular population and are recorded sequentially according to time form FTS (see, e.g. Hörmann & Kokoszka, 2010; Ramsay & Silverman, 2006). In the analysis of FTS, dimension-reduction techniques, such as functional principal component analysis (FPCA), are necessary to collapse the infinite-dimensional curves into a series of orthonormal basis functions and their corresponding matrix-valued coefficients. To handle the high dimensionality of data related to a large number of populations, factor analysis is often used to reduce the variation contained in multi-population data to a smaller number of uncorrelated factors (see, e.g. Bai & Ng, 2008). When applying our proposed method to multi-population mortality data, we employ both dimension-reduction techniques to solve the two-fold 'curse of dimensionality' (Bellman, 1966).

Early attempts at jointly modelling multivariate FTS in the literature include, for example, the work of Di et al. (2009), which introduced the multilevel FPCA that performs the conventional FPCA on intra- and inter-subject geometric components of multilevel functional data. Shang (2016) applied multilevel FPCA to model the mortality rates of multiple populations. To consider functional data that exhibit significant heterogeneity, Tang et al. (2022) proposed a clustering method capable of grouping multiple FTS into homogeneous sets before jointly modelling the curves in each set. Gao et al. (2019) introduced a two-step approach to reduce HDFTS to a small number of latent factors with corresponding factor loadings and basis functions. To obtain a global FPCA projection for multi-population FTS, Hallin et al. (2023) and Tavakoli et al. (2023) arranged HDFTS into a panel structure and performed an eigendecomposition of the stacked FTS. Zhou and Dette (2023) derived Gaussian and multiplier bootstrap approximations for sums of HDFTS.

Parallel to the functional framework, Wang et al. (2019) proposed a factor model for matrix-valued high-dimensional time series, which is also capable of modelling multi-population mortality data. After arranging the considered ages and labels of the specific populations as row variables and column variables, respectively, the matrix input across columns and rows reveals information on the dependence from two distinct perspectives. The correlation of mortality rates in the age direction and among the cross-sectional units can then be summarized via factor analysis into low-dimensional front-loading and back-loading matrices. All temporal information from the original data is retained in a low-dimensional matrix-valued time series in the middle of the two loading matrices. This method, developed by Wang et al. (2019), is conceptually similar to our dual-factor structure for HDFTS, but considers conventional discrete-valued time series data.

In this work, we propose a dual-factor structure for modelling and forecasting HDFTS, which allows us to effectively decompose HDFTS into the product of a vector of common loading (front loading) functions, a low-dimensional matrix-valued time series, and a finite collection of factor loading (backloading) functions. By doing this, the common age characteristics within the high-dimensional mortality data are preserved in the front-loadings, and the region-specific features over age are reflected in the back-factor loadings. In addition, the temporal dynamics of the original HDFTS are captured by the low-dimensional matrix-valued time series that facilitates forecasting. The proposed dual factor structure in general form covers several commonly used functional factor models as special cases. By selecting the appropriate front and back-loading functions, the proposed general dual-factor model structure can be extended to several existing models in the literature.

Our proposed model has several advantages over existing methods.

1. Our dual-factor model for HDFTS considers mortality observations as finite realizations of an underlying continuous process with age as a continuum. In this functional framework, we can better capture mortality dependence between various ages for a particular population than can the matrix-valued time series factor model of Wang et al. (2019).
2. Compared to existing models for HDFTS, our proposed two-fold dimension reduction algorithm can effectively isolate homogeneous features for all data and specific heterogeneous basis functions for each population, with most of the temporal dependence of the original HDFTS retained in a series of low-dimensional factor matrices. The separation of homogeneity and heterogeneity in HDFTS yields accurate forecasts for multi-population mortality data. On the one hand, the heterogeneous basis functions accommodate the unique characteristics of each population, reducing information loss in dimensionality reduction. On the other hand, the extracted homogenous features contribute to nondivergent forecasts for closely related populations.
3. We employ FPCA based on long-term covariance (Rice & Shang, 2017; Shang, 2020) for dimensionality reduction to adequately capture the temporal dependence of HDFTS. To our knowledge, this is the preferred FTS feature extraction method for forecast purposes (Gao et al., 2019; Yang et al., 2022). The long-run covariance-based FPCA method considered in this paper is analogous to dynamic FPCA (see, e.g. Hörmann et al., 2015; Panaretos & Tavakoli, 2013a), which employs functional spectral analysis, but offers a more straightforward implementation.
4. Our proposed model is highly flexible and can be compared with several existing models in the literature. Selecting appropriate front- and backloading matrices enables our proposed model to offer performances comparable to those of popular multi-population FTS models. The structural flexibility of our proposed model is illustrated in Section 2.2.3.

Chang et al. (2024) recently proposed a two-step procedure for modelling high-dimensional HDFTS (TSHDFTS) that initially performs independent component analysis of functional observations to achieve segmentations. Any two curves after transformation are uncorrelated, both cross-sectionally and temporally. The second step involves a latent decomposition to identify a finite-dimensional dynamic representation for each transformed curve. This modelling framework of Chang et al. (2024) can be viewed as a reduced form of (2.4) where the front-loading function is simplified to a constant matrix. Information loss is inevitable in this step of reducing functions to scalars. In contrast, our proposed dual-factor model can better capture the age-varying features of mortality rates via the front-loading functional factor. The empirical results in Section 6 verify that our proposed method produces more accurate point and interval forecasts than the TSHDFTS model.

Jiménez-Varón et al. (2024) considered a two-way functional ANOVA model for HDFTS that decomposes observations into a deterministic component capturing differences of the functional means and a time-varying residual component gauging the dynamic structure of the curves. Compared to the two-way functional ANOVA method, our proposed dual-factor model has a more flexible form, imposing fewer constraints on the functional means. Moreover, our dual-factor model facilitates a decomposition of time-varying functional observations into common factors and their associated back-loading functions. The back-loading functions improve the interpretability of our results, while the common factors summarize time trends in human mortality functions.

The remainder of the paper is organized as follows. Section 2 details our factor model and draws connections with several existing models. Section 3 outlines our estimation procedures. Section 4 describes our forecasting method based on the factor model for HDFTS. Section 5 presents Monte Carlo simulation studies to show the finite sample performance of the proposed method to recover the homogeneous and heterogeneous features of HDFTS. Section 6 presents an empirical analysis of the Japanese subnational age-specific mortality data. Section 7 concludes the paper, highlighting how the methodology can be extended.

2 Factor model for HDFTS

We propose a new dual-factor model for HDFTS, which separates homogeneity and heterogeneity within HDFTS. The homogeneity represents the similarity that is shared among all cross sections, while the heterogeneity represents the dissimilarity. In doing so, we reduce the dimensions of the HDFTS, including summarizing the temporal information of the original data into a low-dimensional matrix-valued time series. Doing so addresses the curse of dimensionality and makes forecasting feasible.

2.1 Functional panel data structure

Let $\{\mathcal{X}_t(u), 1 \le t \le T, u \in [a, b]\}$ be an N-dimensional FTS, where N is the number of FTS processes, and T is the sample size, i.e. the number of functions within each set of FTS. The high-dimensional functional time series (HDFTS) can be organized into a single matrix format such that

$$\mathcal{X}(u) = \begin{bmatrix} \mathcal{X}_1^{(1)}(u) & \mathcal{X}_1^{(2)}(u) & \dots & \mathcal{X}_1^{(N)}(u) \\ \mathcal{X}_2^{(1)}(u) & \mathcal{X}_2^{(2)}(u) & \dots & \mathcal{X}_2^{(N)}(u) \\ \vdots & \vdots & \ddots & \vdots \\ \mathcal{X}_{T-1}^{(1)}(u) & \mathcal{X}_{T-1}^{(2)}(u) & \dots & \mathcal{X}_{T-1}^{(N)}(u) \\ \mathcal{X}_T^{(1)}(u) & \mathcal{X}_T^{(2)}(u) & \dots & \mathcal{X}_T^{(N)}(u) \end{bmatrix},$$

where $\mathcal{X}(u)$ is a $T \times N$ matrix, each entry in the matrix, $\mathcal{X}_t^{(i)}$ for $i \in \{1, \dots, N\}$, is defined in $\mathcal{H} := \mathcal{L}^2(\mathcal{I})$, a Hilbert space of square-integrable functions on a real interval $\mathcal{I} \in [a, b]$, with the inner product $\langle f, g\rangle = \int_{\mathcal{I}} f(u)g(u)du$ and the norm $\|\cdot\| = \langle \cdot, \cdot \rangle^{\frac{1}{2}}$.

Denote $\mathcal{Y}_t^{(i)}(u) = \mathcal{X}_t^{(i)}(u) - \mu^{(i)}(u)$ as the centered FTS, where $\mu^{(i)}(u)$ is the mean function of i^{th} FTS, denoted by $\{\mathcal{X}_1^{(i)}(u), \dots, \mathcal{X}_T^{(i)}(u)\}$. Let us consider the following factor model for the centered HDFTS

$$\mathcal{Y}_t^{(i)}(u) = [\boldsymbol{\phi}(u)]^\top \boldsymbol{F}_t \boldsymbol{\lambda}^{(i)}(u) + \epsilon_t^{(i)}(u), \tag{2.1}$$

where $\boldsymbol{\phi}(u)$ is a column vector of length r, with each entry being function-valued, that is, the features common to all sets of FTS. To maintain consistency, all vectors in this paper are defined as column vectors unless explicitly stated otherwise. The $\boldsymbol{F}_t$ in (2.1) is a matrix of $r \times k$. The $\boldsymbol{\lambda}^{(i)}(u)$ is a vector of length k, with each $i \in \{1, 2, \dots, N\}$ entry being function-valued, that is, reflecting the specific characteristic (heterogeneity) of each FTS. Finally, $\epsilon_t^{(i)}(u)$ is the error component with temporal dependence along the time t and cross-sectional dependence across the section i. We refer to $\boldsymbol{\phi}(u)$ as front loading and $\boldsymbol{\lambda}^{(i)}(u)$ as back loading, respectively (see also Wang et al., 2019). Note that the factor loadings in the literature mentioned above are scalars. In contrast, the factor loadings in our proposed model are function-valued, which is generalized to accommodate complex data.

In our setting, the i^{th} column of $\mathcal{X}(u)$, $\mathcal{X}^{(i)}(u) = [\mathcal{X}_1^{(i)}(u), \dots, \mathcal{X}_T^{(i)}(u)]^\top$, represents an FTS consisting of serial dependence. The t^{th} row of $\mathcal{X}(u)$, $\mathcal{X}_t(u) = [\mathcal{X}_t^{(1)}(u), \dots, \mathcal{X}_t^{(N)}(u)]^\top$, consists of functions of all cross sections N at a specific time t. It is noteworthy that these are correlated. Let us consider the age-specific mortality rate as an example. Each entry in $\mathcal{X}(u)$ represents the age-specific mortality rate curve[1] of a specific region at a specific time. The i^{th} column of $\mathcal{X}(u)$ represents the mortality rate curves of the region i throughout the sample period, and the t^{th} row of $\mathcal{X}(u)$ represents the mortality rate curves of all countries at time t.

2.2 Model interpretation

The proposed model can be interpreted in the following ways. We use the back-loadings to characterize the heterogeneity of the cross sections. Separation of heterogeneity reduces the original

[1] Under the functional data framework, it can be assumed that there is an underlying continuous and smooth function. For the mortality data, we smoothed the mortality rates using weighted penalized regression splines proposed by Hyndman and Ullah (2007).

data into a multivariate FTS of smaller dimensions. These common multivariate FTS characterize the homogeneity of the original data. After reducing the dimensions of these common multivariate FTS, the temporal information is captured by low-dimensional matrix-valued time series, which facilitates forecasting.

2.2.1 *Dimension reduction along cross sections*

We consider a functional factor representation given by

$$\mathcal{F}_t^{\top}(u) = \boldsymbol{\phi}(u)\boldsymbol{F}_t, \tag{2.2}$$

where the common functional loading $\boldsymbol{\phi}(u)$ and the functional factor $\boldsymbol{F}_t$ have the same dimensions as in (2.1). The matrix-valued time series $\{\boldsymbol{F}_t\}$ captures all the temporal dynamics of FTS $\{\mathcal{F}_t(u)\}$ according to this representation.

The model in (2.1) can then be expressed as

$$\mathcal{Y}_t^{(i)}(u) = [\boldsymbol{\lambda}^{(i)}(u)]^{\top}\mathcal{F}_t(u) + \epsilon_t^{(i)}(u), \tag{2.3}$$

where $\boldsymbol{\lambda}^{(i)}(u)$ is a k-dimensional functional factor loading vector, $\boldsymbol{\lambda}^{(i)}(u) = [\lambda_1^{(i)}(u), \ldots, \lambda_k^{(i)}(u)]^{\top}$, with each entry being a function, defined previously, and $\mathcal{F}_t(u)$ is a $k \times 1$ vector of functional factors at time t. $\epsilon_t^{(i)}(u)$ is the error component with temporal dependence along the time t and cross-sectional dependence across the section i. This model differs from those of Hallin et al. (2023) and Tavakoli et al. (2023), where the factor loading is function-valued, and the factors are scalar. Front loading $\boldsymbol{\lambda}^{(i)}(u)$ depends on i, which represents the heterogeneity of each FTS, while the functional factor $\mathcal{F}_t(u)$ is common to all i, which represents the common features of HDFTS. Then $\mathcal{F}(u) = [\mathcal{F}_1(u), \ldots, \mathcal{F}_T(u)]$ is a $k \times T$ ($k \ll N$) matrix of FTS objects. In doing this, we separate the common features of HDFTS, denoted by $\mathcal{F}(u)$, from heterogeneity, denoted by $\boldsymbol{\lambda}^{(i)}(u)$, and reduce HDFTS to FTS with fewer dimensions.

Bai (2009) proposed a panel data model with interactive fixed effects, where the multiplication of the r-dimensional factor back-loading $\boldsymbol{\lambda}_i$ and common factors $\boldsymbol{F}_t$ can be viewed as the interaction of these two terms. The interactive effects model is more general than the additive effects model. It allows the individual effect of i and the time effect of t to be involved in the model interactively rather than additively, giving more flexibility. Freyberger (2018) argued that the interactive fixed effects model could handle cases where heterogeneity is not time homogeneous. Tang et al. (2022) proposed a functional panel data model with additive fixed effects. To handle time-varying heterogeneity, they further proposed a clustering algorithm to search for homogeneous subgroups before jointly modelling the FTS. Our model can be viewed as a functional extension of the interactive fixed-effects model, as stated in (2.3), where both factor loading $\boldsymbol{\lambda}^{(i)}(u)$ and common factors $\mathcal{F}_t(u)$ are function-valued, and $\boldsymbol{\lambda}^{(i)}(u)$ captures the specific effects of different cross sections and $\mathcal{F}_t(u)$ denotes common features with multiplication of $\boldsymbol{\lambda}^{(i)}(u)$ and $\mathcal{F}_t(u)$ allowing for time-varying heterogeneity.

2.2.2 *Dimension reduction on infinite-dimensional functions*

Combining (2.2) and (2.3), we can express the HDFTS $\mathcal{Y}_{\mathbf{t}}(\mathbf{u}) = [\mathcal{Y}_{\mathbf{t}}^{\mathbf{1}}(\mathbf{u}), \ldots, \mathcal{Y}_{\mathbf{t}}^{\mathbf{N}}(\mathbf{u})]^{\top}$ in a matrix format as

$$\mathcal{Y}_{\mathbf{t}}^{\top}(\mathbf{u}) = \boldsymbol{\phi}(\mathbf{u})\boldsymbol{F}_{\mathbf{t}}\boldsymbol{\Lambda}(\mathbf{u}) + \boldsymbol{\epsilon}_{\mathbf{t}}^{\top}(\mathbf{u}), \tag{2.4}$$

where the systems of back-loadings $\boldsymbol{\Lambda}(u) = [\Lambda^{(1)}(u), \ldots, \Lambda^{(N)}(u)]$ are an $k \times N$ ($k \ll N$) matrix, with the i^{th} row being $\boldsymbol{\lambda}^{(i)}(u)$. The $\boldsymbol{\phi}(u)$ is the loading function common to all cross sections, the $\boldsymbol{\Lambda}(u)$ consists of the system loading curves specific to different cross sections, and $\boldsymbol{\epsilon}_t(u)$ is a N-dimensional vector of the white noise processes. To better illustrate this model, we express the $\boldsymbol{\phi}(u)\boldsymbol{F}_t\boldsymbol{\Lambda}(u)$ component of (2.4) using matrices as

$$\boldsymbol{\phi}(u)\boldsymbol{F}_t\boldsymbol{\Lambda}(u) = \begin{bmatrix} \phi_1(u) & \phi_2(u) & \dots & \phi_r(u) \end{bmatrix} \begin{bmatrix} F_{t,11} & F_{t,12} & \dots & F_{t,1k} \\ F_{t,21} & F_{t,22} & \dots & F_{t,2k} \\ \vdots & \vdots & \ddots & \vdots \\ F_{t,r1} & F_{t,r2} & \dots & F_{t,rk} \end{bmatrix} \begin{bmatrix} \lambda_1^{(1)}(u) & \lambda_1^{(2)}(u) & \dots & \lambda_1^{(N)}(u) \\ \lambda_2^{(1)}(u) & \lambda_2^{(2)}(u) & \dots & \lambda_2^{(N)}(u) \\ \vdots & \vdots & \ddots & \vdots \\ \lambda_k^{(1)}(u) & \lambda_k^{(2)}(u) & \dots & \lambda_k^{(N)}(u) \end{bmatrix}$$
$$= \begin{bmatrix} \sum_{p=1}^{r}\sum_{q=1}^{k} \phi_p(u)F_{t,pq}\lambda_q^{(1)}(u) & \sum_{p=1}^{r}\sum_{q=1}^{k} \phi_p(u)F_{t,pq}\lambda_q^{(2)}(u) & \dots & \sum_{p=1}^{r}\sum_{q=1}^{k} \phi_p(u)F_{t,pq}\lambda_q^{(N)}(u) \end{bmatrix} \tag{2.5}$$

By reducing the dimensions of both functional continuum and cross sections, we can extract the important information of the HDFTS into an $r \times k \times T$ factor array $\boldsymbol{F}$.

2.2.3 *Model structural flexibility*

The proposed model has a flexible structure that connects the HDFTS to many popular models in the literature. We illustrate this desired property through several special cases of Equation (2.5).

Case 1: Constant back-loading matrix

When all elements in the back-loading matrix are equal to the same constant c, that is, $\lambda_q^{(i)}(u) = c$ $\forall\ q \in \{1, 2, \dots, k\}$, $i \in \{1, 2, \dots, N\}$, Equation (2.5) can be reduced to

$$\boldsymbol{\phi}(u)\boldsymbol{F}_t\boldsymbol{\Lambda}(u) = \left[c\sum_{p=1}^{r}\phi_p(u)\sum_{q=1}^{k}F_{t,pq} \quad c\sum_{p=1}^{r}\phi_p(u)\sum_{q=1}^{k}F_{t,pq} \quad \dots \quad c\sum_{p=1}^{r}\phi_p(u)\sum_{q=1}^{k}F_{t,pq} \right].$$

Let $\beta_{t,p} = c\sum_{q=1}^{k} F_{t,pq}$, then our proposed model can be expressed as

$$\mathcal{X}_t^{(i)}(u) = \mu^{(i)}(u) + \sum_{p=1}^{r}\beta_{t,p}\phi_p(u) + \epsilon_t^{(i)},\ (u),$$

where $\mu^{(i)}(u) = E[\mathcal{X}_t^{(i)}(u)]$ is the mean function of i^{th} FTS $\{\mathcal{X}_t^{(i)}(u)\}$. This format is identical to the vanilla FPCA-based time series model (see, e.g. Ramsay & Silverman, 2006).

We may also each entry in the back-loading matrix as constants depending on i, i.e. $\lambda_q^{(i)}(u) = c_i,\ \forall\ q \in \{1, 2, \dots, k\}$. Equation (2.5) can then be reduced to

$$\boldsymbol{\phi}(u)\boldsymbol{F}_t\boldsymbol{\Lambda}(u) = \left[c_1\sum_{p=1}^{r}\phi_p(u)\sum_{q=1}^{k}F_{t,pq} \quad c_2\sum_{p=1}^{r}\phi_p(u)\sum_{q=1}^{k}F_{t,pq} \quad \dots \quad c_N\sum_{p=1}^{r}\phi_p(u)\sum_{q=1}^{k}F_{t,pq} \right].$$

Denoting $\beta_{t,p}^{(i)} = c_i\sum_{q=1}^{k} F_{t,pq}$ leads to the model given by

$$\mathcal{X}_t^{(i)}(u) = \mu^{(i)}(u) + \sum_{p=1}^{r}\beta_{t,p}^{(i)}\phi_p(u) + \epsilon_t^{(i)}(u).$$

This can be viewed as a special case of Gao and Shang (2017), which employs the same set of basis functions for all populations.

Case 2: Back-loading matrix with population-specific functions

If each element of the back-loading matrix is a function specific to i, that is, $\lambda_q^{(i)}(u) = \gamma^{(i)}(u)$, $\forall\ q \in \{1, 2, \dots, k\}$, Equation (2.5) can be written as

$$\boldsymbol{\phi}(u)\boldsymbol{F}_t\boldsymbol{\Lambda}(u) = \left[\gamma^{(1)}(u)\sum_{p=1}^{r}\phi_p(u)\sum_{q=1}^{k}F_{t,pq} \quad \gamma^{(2)}(u)\sum_{p=1}^{r}\phi_p(u)\sum_{q=1}^{k}F_{t,pq} \quad \dots \quad \gamma^{(N)}(u)\sum_{p=1}^{r}\phi_p(u)\sum_{q=1}^{k}F_{t,pq} \right].$$

Let $\beta_{t,p} = \sum_{q=1}^{k} F_{t,pq}$ and $\phi_p^{(i)}(u) = \gamma^{(i)}(u)\phi_p(u)$. We then have

$$\mathcal{X}_t^{(i)}(u) = \mu^{(i)}(u) + \sum_{p=1}^{r} \beta_{t,p}\phi_p^{(i)}(u) + \epsilon_t^{(i)}(u),$$

which is a reduced version of the multivariate FPCA considered in Chiou and Müller (2014).

Case 3: Discretized back-loading matrix

The back-loading matrix in our proposed model can also take discretized values. Consider a finite collection of grid points $\{u_1, \ldots, u_m\}$ over the entire functional support $\mathcal{I}$. For any specific u_ℓ with $\ell = 1, \ldots, m$, the entries in the back-loading matrix become scalar realizations, that is, $\lambda_q^{(i)}(u_m) = c_{i,q,m}$. Hence, the proposed model turns out to be equivalent to the factor model for high-dimensional matrix-valued time series studied by Wang et al. (2019).

3 Parameter estimation

We present an estimation procedure for the parameters $\mu^{(i)}(u)$, $\boldsymbol{\phi}(u)$, $\boldsymbol{\lambda}^{i}(\boldsymbol{u})\boldsymbol{F}_t$ in (2.1). In general, the idea of the estimation procedure is to estimate the systems of back-loading first so that the HDFTS is reduced to the FTS of lower dimensions. We then proceed to estimate the common front-loading time series and the matrix-valued time series $\{\boldsymbol{F}_t\}$.

Suppose that we have T curves in each cross section, the estimation of $\mu^{(i)}(u)$ is as follows:

$$\widehat{\mu}^{(i)}(u) = \frac{1}{T}\sum_{t=1}^{T} \mathcal{X}_t^{(i)}(u).$$

Then, $\mathcal{Y}_t^{(i)}(u)$ is expressed as

$$\mathcal{Y}_t^{(i)}(u) = \mathcal{X}_t^{(i)}(u) - \widehat{\mu}^{(i)}(u).$$

3.1 Estimation of functional loadings $\boldsymbol{\Lambda(u)}$

The cross-sectional specific functional back-loadings $\boldsymbol{\lambda}^{(i)}(\boldsymbol{u})$ represent the heterogeneity of each FTS. To accommodate the serial dependence in each cross section, $\boldsymbol{\lambda}^{(i)}(\boldsymbol{u})$ is estimated by the eigen-decomposition of the long-run covariance operator corresponding to the i^{th} cross section.

Traditional FPCA is known to fall short in capturing the crucial information embedded in the serial dependence structure of functional time series. Horváth et al. (2013) and Panaretos and Tavakoli (2013b) defined smoothed periodogram-type estimates of the long-run covariance and spectral density operators for FTS. Hörmann et al. (2015) used the spectral density operator to create functional filters to construct mutually uncorrelated dynamic FPCs so that dynamic FPC scores can be analyzed component-wise. Rice and Shang (2017) proposed a bandwidth selection method for estimates of the long-run covariance function based on finite-order weight functions that aim to minimize the estimator's asymptotic mean-squared normed error. Following their work, the long-run covariance function for the i^{th} cross section $c^{(i)}(u, v)$ can be estimated by

$$\widehat{c}^{(i)}(u, v) = \sum_{s=-\infty}^{\infty} W\left(\frac{s}{h}\right)\widehat{c}_s^{(i)}(u, v),$$

where

$$\widehat{c}_s^{(i)}(u, v) = \begin{cases} \dfrac{1}{T-s}\displaystyle\sum_{t=1}^{T-s} \widehat{\mathcal{Y}}_t^{(i)}(u)\widehat{\mathcal{Y}}_{t+s}^{(i)}(v), & s \geq 0; \\ \dfrac{1}{T-s}\displaystyle\sum_{t=1-s}^{T} \widehat{\mathcal{Y}}_t^{(i)}(u)\widehat{\mathcal{Y}}_{t+s}^{(i)}(v), & s < 0. \end{cases}$$

$W(\cdot)$ is the kernel function that assigns different weights to the auto-covariance functions of different lags, and h is the bandwidth. Despite many different types of kernel functions (see Andrews, 1991; Gallant, 2009; Hansen, 1982; Newey & West, 1987; White, 1984, for types of kernel functions), it is common to assign more weight to the auto-covariance functions of small lags and less weight to the auto-covariance functions of large lags. Here, we use flat-top kernel functions as they give a reduced bias and faster rates of convergence (Politis & Romano, 1996, 1999). The choice of bandwidth in the auto-covariance estimation can greatly affect our method's finite-sample performance. Therefore, we apply the adaptive bandwidth selection procedure of Rice and Shang (2017) to gain a better estimate of the long-run covariance functions.

Remark 3.1 In the early literature of Lam et al. (2011) and Lam and Yao (2012), they estimate factor models based on the eigen-decomposition of autocovariances, by assuming the error components in factor models to be white noise. Recently, Zhang et al. (2024) extended the idiosyncratic components of white noise to a weak temporal dependence (see Assumption 4 in Zhang et al., 2024). Essentially, the auto-covariance eigen-analysis method is also feasible when temporal dependence is allowed in the error component. The intuition behind this feasibility is that, if the autocovariance of the factor model part is still the leading term in the autocovariance of the original data, factor estimation based on autocovariance can attain asymptotic consistency, although it is not efficient compared with the white noise case. Similarly, this estimation approach is also applicable when the idiosyncratic components have cross-sectional dependence to some extent. Overall, we can impose some restrictions on the cross-sectional and temporal dependence in the idiosyncratic components, to make the factor model part play the leading term in the eigen-decomposition of the covariance and autocovariance of the original data. In this way, the factor model can be identified asymptotically.

The long-run covariance operator can be estimated by

$$\widehat{C}^{(i)}(y)(u) = \sum_{s=-\infty}^{\infty} W\left(\frac{s}{h}\right)\widehat{C}_s^{(i)}(y)(u),$$

where $\widehat{C}_s^{(i)}(y)(u) = \int_{\mathcal{I}} \widehat{c}_s^{(i)}(u, v)y(v)dv$, is the sample covariance operator in lag s. With the estimated long-run covariance, we apply a principal component decomposition to extract the functional principal components and their associated scores. By Mercer's lemma, the estimator of the long-run covariance function can be expressed as

$$\widehat{C}^{(i)}(u, v) = \sum_{m=1}^{\infty} \widehat{\theta}_m^{(i)}\widehat{\lambda}_m^{(i)}(u)\widehat{\lambda}_m^{(i)}(v)$$

where $\widehat{\theta}_1^{(i)} > \widehat{\theta}_2^{(i)} > \ldots > 0$ are the estimated eigenvalues of $\widehat{C}^{(i)}(u, v)$, and $[\widehat{\lambda}_1^{(i)}(u), \widehat{\lambda}_2^{(i)}(u), \ldots]$ are the associated eigenfunctions.

The estimates of $\boldsymbol{\lambda}^{(i)}(\boldsymbol{u})$, denoted by $\widehat{\boldsymbol{\lambda}}^{(i)}(\boldsymbol{u}) = [\widehat{\lambda}_1^{(i)}(u), \ldots, \widehat{\lambda}_{k_i}^{(i)}(u)]$, are the eigenfunctions corresponding to the first k_i largest eigenvalues of $\widehat{C}^{(i)}(y)(u)$. Since we allow $\boldsymbol{\lambda}^{(i)}(\boldsymbol{u})$ to vary according to i, it represents the heterogeneity of each population.

To maintain the matrix format of the loadings of each population, k is selected to be

$$k = \max\{k_i : 1 \le i \le N\},$$

where k_i is the number of components retained for the i^{th} population.

The selection of the optimal number of functional principal components has been widely studied in the literature (for example, Chiou, 2012; Hall & Vial, 2006; Hörmann & Kidziński, 2015; Rice & Silverman, 1991; Yao et al., 2005). In this paper, we follow Ahn and Horenstein (2013), Li et al.

(2020) and Martínez-Hernández et al. (2022) and determine k_i as the integer that minimizes the ratio of two adjacent empirical eigenvalues given by

$$k_i = \underset{1\le m\le k_{\max}^{(i)}}{\operatorname{argmin}} \left\{ \frac{\widehat{\theta}_{m+1}^{(i)}}{\widehat{\theta}_m^{(i)}} \times \mathbb{1}\left(\frac{\widehat{\theta}_m^{(i)}}{\widehat{\theta}_1^{(i)}} \ge \delta \right) + \mathbb{1}\left(\frac{\widehat{\theta}_m^{(i)}}{\widehat{\theta}_1^{(i)}} < \delta \right) \right\}, \tag{3.1}$$

where $k_{\max}^{(i)}$ is a pre-specified positive integer, δ is a pre-specified small positive number, and $\mathbb{1}(\cdot)$ is the indicator function. We choose $k_{\max}^{(i)} = \#\{m \mid \widehat{\theta}_m \ge \sum_{m=1}^{T} \widehat{\theta}_m/T, m \ge 1\}$ and set the threshold constant $\delta = 1/\ln(\max\{\widehat{\theta}_1, T\})$.

3.2 Estimation of $\mathcal{F}_t(u)$

In order to estimate $\boldsymbol{\phi}(\boldsymbol{u})$ and $\boldsymbol{F}_t$, we need to estimate $\mathcal{F}_t(u)$ in (2.3). To extract the common features of the HDFTS, first rewrite (2.3) in matrix format given by

$$\mathcal{Y}_{\mathbf{t}}(\mathbf{u}) = \boldsymbol{\Lambda}^{\top}(\mathbf{u})\mathcal{F}_{\mathbf{t}}(\mathbf{u}) + \boldsymbol{\epsilon}_{\mathbf{t}}(\mathbf{u}).$$

The i^{th} entry of $\mathcal{Y}_{\mathbf{t}}(\mathbf{u})$ can be expressed as follows:

$$\mathcal{Y}_t^{(i)}(u) = \sum_{q=1}^{k} \lambda_q^{(i)}(u)\mathcal{F}_{t,q}(u) + \epsilon_t^{(i)}(u).$$

This is essentially the functional concurrent regression model in Ramsay et al. (2009), but without the intercept function, where $\lambda_q^{(i)}(u)$ is a k-dimensional functional observation and $\mathcal{F}_{t,q}(u)$ is the functional regression coefficient.

Let the N-dimensional residual function vector be

$$\boldsymbol{r}_t(\boldsymbol{u}) = \mathcal{Y}_t(u) - \boldsymbol{\Lambda}^{\top}(u)\mathcal{F}_t(u).$$

Then, $\mathcal{F}_t(u)$ can be estimated by minimizing the penalized sum of squares,[2] this is done through `fRegress` function of the `fda` package in `R`.

Concurrent functional regression is performed for each $t = 1, 2, \ldots, T$, such that $\widehat{\mathcal{F}}(u) = [\widehat{\mathcal{F}}_1(u), \widehat{\mathcal{F}}_2(u), \ldots, \widehat{\mathcal{F}}_T(u)]$ is a $k \times T$ matrix of function-valued objects. Hence, we have a panel of dimension-reduced FTS, which represents the common features.

3.3 Estimation of the front-loading function $\boldsymbol{\phi}(u)$ and factor matrix $\boldsymbol{F}_t$

Once $\widehat{\mathcal{F}}(u)$ is obtained, we can estimate $\boldsymbol{\phi}(u)$ and $\boldsymbol{F}_t$ based on (2.2). Since $\mathcal{F}(u)$ is a dimension-reduced panel of FTS, where each column may be correlated, the key is to calculate the auto-cross-covariance of $\mathcal{F}(u)$. Wang et al. (2019) used the auto-cross-covariance to capture the co-movement observations from different cross sections of different lags. We extend it to the functional setting. Let $C_{h,lj}(u, v)$ be the auto-cross-covariance function of the l^{th} and j^{th} column of $\mathcal{F}(u)$, for l and $j = 1, \ldots, k$, such that

$$C_{h,lj}(u, v) = \operatorname{Cov}\left[\mathcal{F}_{t,l}(u), \mathcal{F}_{t+h,j}(u)\right].$$

Since the operator using $C_{h,lj}(u, v)$ as a kernel is not non-negative, we define a non-negative operator (using a similar idea as Bathia et al., 2010),

$$M_{h,lj}(u, v) = \int_{\mathcal{I}} C_{h,lj}(u, z)C_{h,lj}(v, z)dz.$$

[2] See Chapter 10.2 of Ramsay et al. (2009) for details.

For a pre-determined integer h_0, the operator

$$M(u, v) = \sum_{h=1}^{h_0} \sum_{l=1}^{k} \sum_{j=1}^{k} M_{h,lj}(u, v), \tag{3.2}$$

is also nonnegative. The auto-cross-covariance function can be estimated by

$$\widehat{C}_{h,lj}(u, v) = \frac{1}{T-h} \sum_{t=1}^{T-h} \widehat{\mathcal{F}}_{t,l}(u) \widehat{\mathcal{F}}_{t+h,j}(v).$$

Hence, the operator $M(u, v)$ can be estimated by

$$\widehat{M}(u, v) = \sum_{h=1}^{h_0} \sum_{l=1}^{k} \sum_{j=1}^{k} \int_{\mathcal{I}} \widehat{C}_{h,lj}(u, z) \widehat{C}_{h,lj}(v, z) dz.$$

Estimates of $\boldsymbol{\phi}(u)$, $\widehat{\boldsymbol{\phi}}(u) = [\widehat{\phi}_1(u), \ \ldots, \widehat{\phi}_r(u)]$, are the eigenfunctions corresponding to the first r largest eigenvalues of the eigen-decomposition on $\widehat{M}(u, v)$. The optimal r can be selected using the same approach as in (3.1).

Once $\widehat{\boldsymbol{\phi}}(u)$ is obtained, the estimated $r \times k$ matrix $\widehat{\boldsymbol{F}}_t$ can be obtained through

$$\widehat{\boldsymbol{F}}_t = \int_{\mathcal{I}} \widehat{\boldsymbol{\phi}}^{\top}(u) \widehat{\mathcal{F}}_t^{\top}(u) du.$$

Therefore, all temporal information is extracted into the T sheets of $r \times k$ matrices.

We note a caveat that the front- and back-loadings of Equation (2.4) are not uniquely identifiable unless additional constraints are imposed. The lack of identifiability is not a major concern for forecasting, since point forecasts depend on the combined effect of the functional factors and their corresponding loadings. The reader is referred to (see Hörmann & Jammoul, 2022, Remark 3) for more discussions on the impacts of model identifiability on forecasting.

4 Forecasting

Based on the proposed model in (2.1), the functional panel data are extracted from T sheets of fixed-dimensional factor matrices, $\boldsymbol{F}_t$. Forecasting HDFTS is equivalent to forecasting the estimated factor matrices, $\widehat{\boldsymbol{F}}_t$. Following the standard matrix-valued time series forecasting method (see, e.g. Tsay, 2024, for more details), we initially vectorize the matrices $\{\widehat{\boldsymbol{F}}_1, \widehat{\boldsymbol{F}}_2, \ \ldots, \widehat{\boldsymbol{F}}_\kappa\}$ before applying vector autoregressive (VAR) models to produce h-step-ahead in-sample forecast $\widehat{\boldsymbol{F}}_{\kappa+h\,|\,\kappa}$, where $\kappa < T$ is the training data sample size. In selecting the order of the VAR model, we follow the method of Tsay (2013), where an information criterion is used, such as the Akaike information criterion.

For a given sample size κ, the h-step-ahead in-sample forecast of the i^{th} cross section in the HDFTS can be calculated as

$$\widehat{\mathcal{X}}^{(i)}_{\kappa+h\,|\,\kappa}(u) = \widehat{\mu}^{(i)}(u) + \widehat{\boldsymbol{\phi}}(u) \widehat{\boldsymbol{F}}_{\kappa+h\,|\,\kappa} \widehat{\boldsymbol{\lambda}}^{(i)}(u).$$

To capture the uncertainties associated with the point forecasts, point-wise prediction intervals could also be constructed. We apply the nonparametric bootstrap approach of Shang (2018) to construct point-wise prediction intervals.

For a grid point u_j on the curve, the h-step-ahead bootstrapped forecasts of the i^{th} cross section, $\widehat{\mathcal{X}}^{(i),b}_{\kappa+h\,|\,\kappa}(u_j)$, can be obtained by adding a bootstrapped residual, $\widehat{e}^{(i),b}_{\kappa+h\,|\,\kappa}(u_j)$ to the point forecast, $\widehat{\mathcal{X}}^{(i)}_{\kappa+h\,|\,\kappa}(u_j)$. The $\widehat{e}^{(i),b}_{\kappa+h\,|\,\kappa}(u_j)$ is generated using sampling with replacement from the in-sample-forecast errors,

Algorithm 1: HDFTS forecasting

1 Input: HDFTS $\{\mathbf{X}_t(u) = [\mathcal{X}_t^{(\infty)}, \mathcal{X}_t^{(\in)}, \ldots, \mathcal{X}_t^{(\mathcal{N})}], t = 1, \ldots, \kappa\}$.

2 Estimation Step:

2.1 For $i \in \{1, 2, \ldots, N\}$, compute the sample mean $\widehat{\mu}^{(i)} = \frac{1}{\kappa}\sum_{t=1}^{\kappa} \mathcal{X}_t^{(i)}(u)$ and the demeaned function $\widehat{\mathcal{Y}}_t^{(i)} = \mathcal{X}_t^{(i)}(u) - \widehat{\mu}^{(i)}$;

2.2 For $i \in \{1, 2, \ldots, N\}$, apply the FPCA based on long-run covariance function to $\widehat{\mathcal{Y}}^{(i)} = [\widehat{\mathcal{Y}}_1^{(i)}, \widehat{\mathcal{Y}}_2^{(i)}, \ldots, \widehat{\mathcal{Y}}_\kappa^{(i)}]^\top$ and obtain the eigenfunctions $\widehat{\boldsymbol{\lambda}}^{(\mathbf{i})}(\mathbf{u}) = [\widehat{\lambda}_1^{(i)}(u), \ldots, \widehat{\lambda}_{k_i}^{(i)}(u)]$ corresponding to the first k_i largest eigenvalues of the long-run covariance operator of $\widehat{\mathcal{Y}}_t^{(i)}$, where k_i is selected based on (3.1);

2.3 Choose $k = \max\{k_i : 1 \le i \le N\}$, such that the loadings of each cross section are of the same dimension such that $\boldsymbol{\Lambda}(u) = [\widehat{\boldsymbol{\lambda}}^{(1)}(u), \widehat{\boldsymbol{\lambda}}^{(2)}(u), \ldots, \widehat{\boldsymbol{\lambda}}^{(\mathbf{N})}(u)]^\top$;

2.4 For $t \in \{1, \ldots, \kappa\}$, perform the concurrent functional regression with $\{\widehat{\mathcal{Y}}_t^{(i)}, i = 1, \ldots, N\}$ as the response and $\{\widehat{\boldsymbol{\lambda}}^{(\mathbf{i})}(\mathbf{u}), i = 1, \ldots, N\}$ as covariates and obtain the coefficient functions $\widehat{\mathcal{F}}_t(u) = [\widehat{\mathcal{F}}_{t,1}(u), \ldots, \widehat{\mathcal{F}}_{t,k}(u)]$ and $\widehat{\mathcal{F}}(u) = [\widehat{\mathcal{F}}_1(u)^\top, \widehat{\mathcal{F}}_2(u)^\top, \ldots, \widehat{\mathcal{F}}_T(u)^\top]^\top$;

2.5 For $i, j \in \{1, \ldots, k\}$, compute $\widehat{C}_{h,lj}(u, v) = \frac{1}{T-h}\sum_{t=1}^{T-h} \widehat{\mathcal{F}}_{t,l}(u)\widehat{\mathcal{F}}_{t+h,j}(v)$;

2.6 For a predetermined h_0, compute $\widehat{M}(u, v) = \sum_{h=1}^{h_0}\sum_{l=1}^{k}\sum_{j=1}^{k}\int_{\mathcal{I}} \widehat{C}_{h,lj}(u, z)\widehat{C}_{h,lj}(v, z)dz$;

2.7 Conduct eigendecomposition on $\widehat{M}(u, v)$ and get $\widehat{\boldsymbol{\phi}}(u) = [\widehat{\phi}_1(u), \ldots, \widehat{\phi}_r(u)]$, the eigenfunctions corresponding to the first r largest eigenvalues of $\widehat{M}(u, v)$, where r is selected based on similar idea as in (3.1);

2.8 For $t \in \{1, 2, \ldots, \kappa\}$, compute $\widehat{\mathbf{F}}_t = \int_{\mathcal{I}} \widehat{\boldsymbol{\phi}}(u)^\top \widehat{\mathcal{F}}_t(u)^\top du$.

3. Forecasting Step:

3.1 Obtain the h-step-ahead forecast of $\{\widehat{\mathbf{F}}_t, t = 1, \ldots, \kappa\}$, $\widehat{\mathbf{F}}_{\kappa+h\,|\,\kappa}$;

3.2 Recover the h-step-ahead forecast of the i^{th} cross section of the HDFTS, $\widehat{\mathcal{X}}_{\kappa+h\,|\,\kappa}^{(i)}(u) = \widehat{\mu}^{(i)}(u) + \widehat{\boldsymbol{\Phi}}(u)\widehat{\mathbf{F}}_{\kappa+h\,|\,\kappa}\widehat{\boldsymbol{\lambda}}^{(i)}(u)$.

Result: $\widehat{\mathcal{X}}_{\kappa+h\,|\,\kappa}^{(i)}(u)$, the h-step-ahead forecast of the i^{th} cross section of the HDFTS.

$$\widehat{e}_{\delta+h\,|\,\delta}^{(i)}(u_j) = \mathcal{X}_{\delta+h}^{(i)}(u_j) - \widehat{\mathcal{X}}_{\delta+h\,|\,\delta}^{(i)}(u_j), \quad \delta = 2, \ldots, T-h.$$

For a given significance level α, the h-step-ahead bootstrapped point-wise prediction interval for the i^{th} cross section, $[\widehat{\mathcal{X}}_{\kappa+h\,|\,\kappa}^{(i),\text{lb}}(u_j), \widehat{\mathcal{X}}_{\kappa+h\,|\,\kappa}^{(i),\text{ub}}(u_j)]$, can be calculated point-wise as the $100 \times (\alpha/2)^{\text{th}}$ and $100 \times (1-\alpha/2)^{\text{th}}$ percentile of the bootstrapped forecasts at each grid point, respectively. The estimation procedure and the forecasting method are summarized in Algorithm 1.

5 Monte-Carlo simulation studies

We consider a series of Monte-Carlo simulations to evaluate the finite-sample performance of the proposed method. The data generating process (DGP) is designed to show the proposed method's ability to accurately recover the front factor loading $\boldsymbol{\phi}(u)$ and the back factor loading $\boldsymbol{\lambda}^{(i)}(u)$ of FTS.

1. The simulated $\{\mathcal{Y}_t^{(i)}(u)\}$ functions are highly positively correlated, corresponding to the strong temporal correlation. This is achieved by adopting VAR models with positive-diagonal coefficient matrices.
2. The elements of the back factor loading $\boldsymbol{\lambda}^{(i)}(u)$ are designed to slightly vary according to $i = 1, 2, \ldots, N$.
3. The elements of the front factor loading $\boldsymbol{\phi}(u)$ are selected to be independent of the individual index i. This setting is consistent with the practice that one set of homogeneous basis functions is used in modelling highly correlated data.

In total, we generate N sets of correlated FTS according to (2.1). In particular, the t^{th} curve of i^{th} FTS is generated from the following model contaminated with measurement error $\epsilon_t^{(i)}(u)$

$$\mathcal{Y}_t^{(i)}(u) = \boldsymbol{\phi}(u)\boldsymbol{F}_t\boldsymbol{\lambda}^{(i)}(u) + \epsilon_t^{(i)}(u),$$

where $\boldsymbol{F}_t$ is a 2×2 matrix with each row generated using a VAR model of order 1. Specifically, the first and second rows are generated using the VAR(1) model with coefficient matrices

$$\begin{pmatrix} 0.5 & 0.2 \\ 0.2 & 0.5 \end{pmatrix}, \quad \begin{pmatrix} 0.4 & -0.25 \\ -2 & 0.3 \end{pmatrix}.$$

The covariance matrix of the innovations for both VAR(1) models is $\begin{pmatrix} 1 & 0.5 \\ 0.5 & 1 \end{pmatrix}$. The two front-loading curves are $\phi_1(u) = \sqrt{2} * \sin(\pi u)$ and $\phi_2(u) = \sqrt{2} * \cos(\pi u)$, respectively, with $\{u = \frac{m}{100} : m = 0, 1, \ldots, 100\}$. The two back-loading curves for each i are $\lambda_2^{(i)} = \sin(2\pi u + i\pi/2)$ and $\lambda_1^{(i)} = \cos(2\pi u + i\pi/2)$, respectively. The measurement error $\epsilon_t^{(i)}(u)$ is generated from a white noise process with sigma = 0.5.

The above DGP involves different combinations of N and T. We select $T = 10$, 20, 30, 40 and $N = 20$, 30, 40. The pre-determined integer h_0 in (3.2) is taken to be one as the VAR(1) model is considered.

The in-sample model fitting can be evaluated using root mean squared error (RMSE)

$$\text{RMSE}(h) = \sqrt{\frac{1}{N \times T \times 101} \sum_{i=1}^{N} \sum_{t=1}^{T} \sum_{m=1}^{101} \left[\mathcal{Y}_t^{(i)}(u_m) - \widehat{\mathcal{Y}}_t^{(i)}(u_m)\right]^2},$$

where $\widehat{\mathcal{Y}}_t^{(i)}(u_m)$ represents the estimated function value at the grid points $u_m \in [0, 1]$. For each combination of N and T, we simulate 100 replications with pseudo-random seeds and compute the mean of the RMSE values.

Figure 1 displays the model fitting performance. For a given $N(T)$, as $T(N)$ increases, the estimation error decreases. Moreover, the decrease in RMSE is larger when increaseing T for a given N than the decrease in RMSE when increasing N for a given T. The decreasing RMSE when N increases indicates that our model can handle high-dimensional cases when $N > T$.

6 Empirical studies

We apply the proposed method to forecast sub-national mortality rates in Japan (Japanese Mortality Database, 2025). In particular, we consider the age-specific mortality rates of 47 prefectures of Japan from 1975 to 2021. We pick this time interval to ensure that the number of observation years is less than the number of prefectures, ensuring the 'curse of dimensionality' challenges are preserved in the empirical data.

We evaluated and compared the prediction accuracy of the proposed method with several competing methods. In addition, we estimate life annuity pricing to demonstrate the impact of the forecast improvement from the proposed method.

6.1 Japanese subnational age-specific mortality

In the mortality data applications, the weighted penalized regression splines method is applied to convert the observed data into continuous functions. Due to the sparsity of data for the very high age group, we have aggregated both female and male mortality for ages over 100. Hence, the number of grid points on the mortality curves is 101.

Figure 2 shows the rainbow plots of the logarithm of the smoothed female mortality rates for four prefectures of Japan, namely, Tokyo, Nagano, Kyoto, and Osaka, from 1975 to 2022. The rainbow plots of Hyndman and Shang (2010) can visualize the time order of functions. The color order of the rainbow reflects the time ordering of the functional observations. Functions from the distant past are depicted in red, whereas more recent functions are shown in purple.

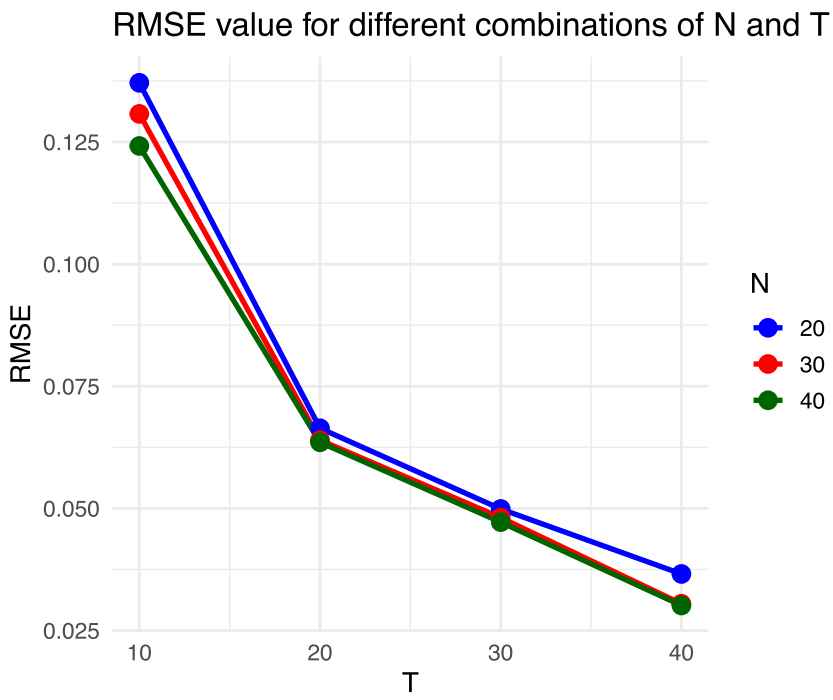

Figure 1. Fitting RMSE for different combinations of N and T.

Although the general shapes of mortality rates for these four prefectures are similar, we may observe significant differences in the troughs and peaks and the overlaps of the curves. This indicates that the mortality rates of different prefectures contain heterogeneity. To achieve improved forecasts of subnational mortality rates, we need to extract the common features and use these common features to produce forecasts. This could be achieved by separating the homogeneity and heterogeneity of mortality rates in all prefectures. From (2.4), the heterogeneity of mortality rates in each prefecture is reflected in the functional factor loading matrix $\boldsymbol{\Lambda}(u)$, as each prefecture has a different set of functional factor loadings. The term $\boldsymbol{\phi}(u)\boldsymbol{F}_t$ represents homogeneity, which is common to all FTS.

Figure 3 shows the first four functional factor loadings (the $\boldsymbol{\lambda}^{(i)}(\boldsymbol{u})$ term) of mortality rates of four chosen prefectures. All these four-factor loadings are different for these four prefectures. This represents the heterogeneity in subnational mortality rates with different troughs, peaks, and overlaps of mortality rate curves.

After separating the heterogeneity, we extract the common characteristics, $\mathcal{F}_t(u)$, in Figure 4, which are common to all the mortality rate curves. This reflects the general common shapes and color ordering of the mortality rates of all prefectures in Japan. By extracting common characteristics, we have reduced the dimension of the FTS of 47 prefectures of Japan into smaller sets of FTS. The temporal information within the original HDFTS is preserved in the fixed-dimensional FTS.

In addition, the first four common functional factor loadings, $\boldsymbol{\phi}(\boldsymbol{u})$, of both genders are depicted in Figure 5. Although the general shape is similar for men and women, the common functional loadings for males contain more troughs and peaks, indicating that male mortality rates are more volatile than those of females.

Table 1 presents the first six eigenvalues for the auto-cross-covariance function of the common FTS and the long-run-covariance function of the FTS of four selected prefectures for both genders. The eigenvalues are normalized so that the sum of all eigenvalues adds up to one to reflect the percentage of total variation explained. We observe that the first few factors would explain most of the variations in the data.

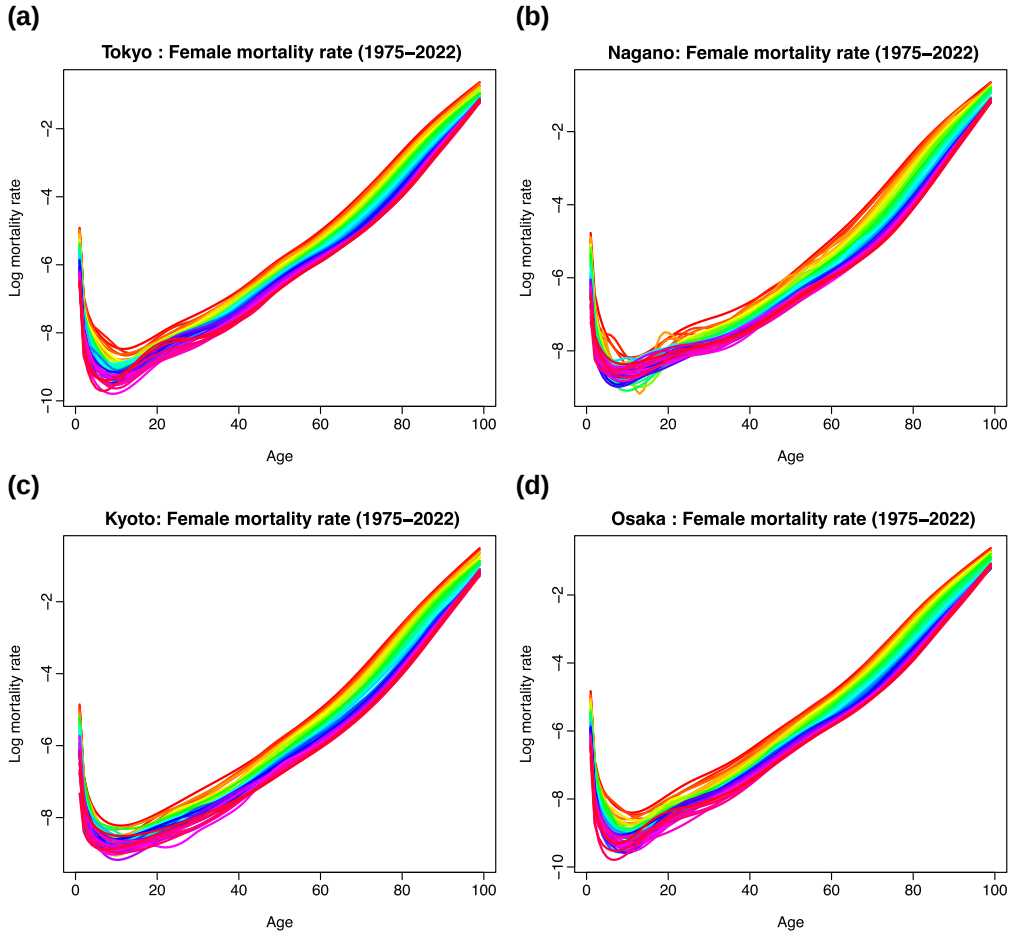

Figure 2. Functional time series displays for smoothed age-specific mortality rates of selected prefectures in Japan. (a) Female smoothed mortality rates in Tokyo. (b) Female smoothed mortality rates in Nagano. (c) Female smoothed mortality rates in Kyoto. (d) Female smoothed mortality rates in Osaka.

The number of factors that we select for common and specific FTS are four and six, respectively, for females, and seven and eight, respectively, for males. With these selections of the number of factors, we ensure that the factors explain 99% of the variation in the data. In addition, the percentages of total variation explained by the first factor for males are smaller than those for females, and this also explains that the data for males are more volatile.

6.2 Forecast evaluation

We consider the Japanese sub-national mortality data between 1975–2022 in the empirical study. To assess the stabilities of the model and the parameters over time, we adopt an expanding window analysis (see Zivot & Wang, 2006, Chapter 9 for details) and produce forecasts in iterations. Specifically, we initiate the process by fitting the proposed method to the first 38 years of observations between 1975–2012 and generating one- to 10-step-ahead point forecasts. Through an expanding window approach, we re-estimate the parameters of the considered model using the first 39 mortality curves from 1975 to 2013 and make one- to nine-step-ahead forecasts. The process is iterated with the sample size increased by one year until the end of the observation period. This process produces ten one-step-ahead forecasts, nine two-step-ahead forecasts, and so on, up to one 10-step-ahead forecast.

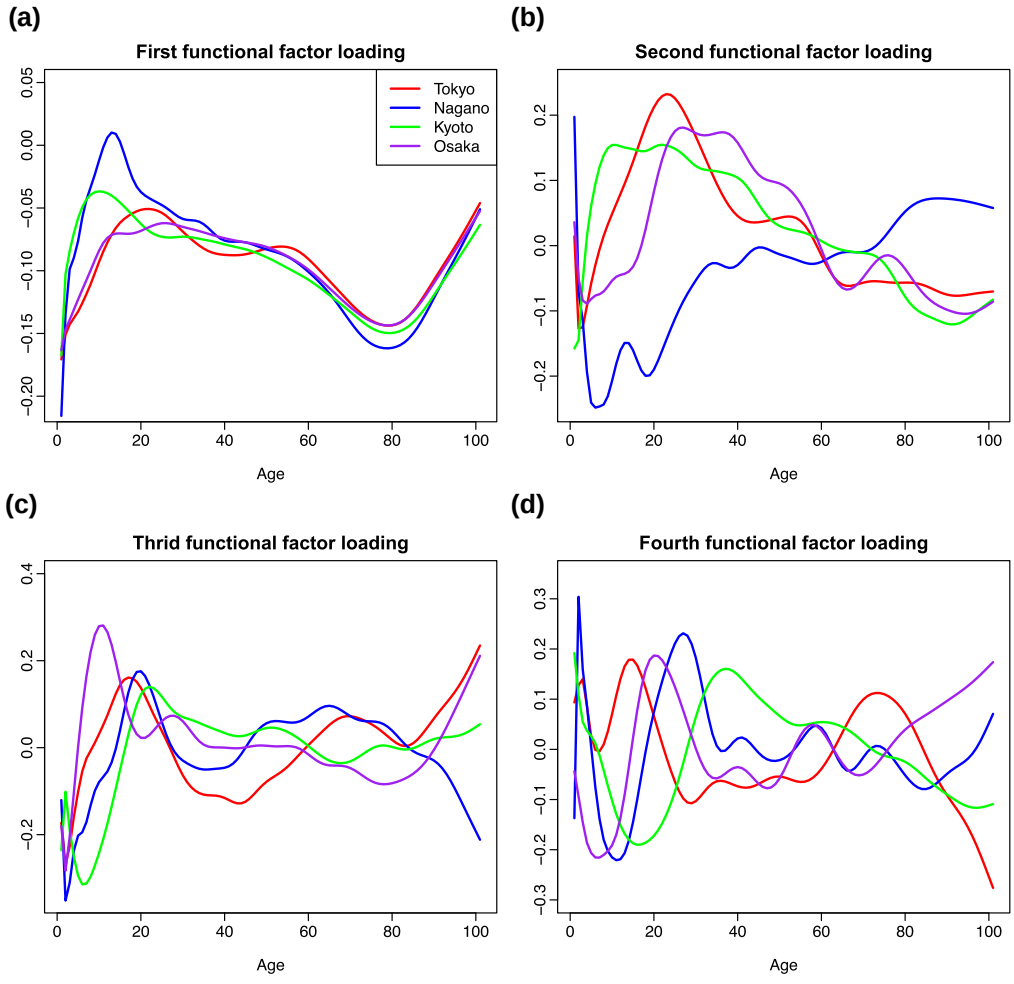

Figure 3. The first four functional factor loadings (the $\boldsymbol{\lambda}^{(i)}(\boldsymbol{u})$ terms) for four selected prefectures of Japan, namely Tokyo, Nagano, Kyoto and Osaka.(a) First functional factor loading. (b) Second functional factor loading. (c) Third functional factor loading. (d) Fourth functional factor loading.

We compare the proposed method with the HDFTS method studied in Gao et al. (2019), the high-dimensional functional factor model (HDFFM) (Hallin et al., 2023; Tavakoli et al., 2023), the two-step high-dimensional functional time series model (TSHDFTS) (Chang et al., 2024), the factor models for matrix-valued high-dimensional time series (MFM) (Wang et al., 2019), and the tensor decomposition forecast methods (Canonical Polyadic Decomposition (CPD) and Tucker of Dong et al. (2020)).

6.2.1 Point forecast accuracy

We use the root mean square forecast error (RMSFE) to evaluate the point forecast accuracy. The RMSFE measures how close the forecast results are to the actual values of the data under forecast. The RMSFE for the h-step-ahead forecasts of the i^{th} prefecture for $i \in \{1, \ldots, 47\}$ can be written as

$$\text{RMSFE}^{(i)}(h) = \sqrt{\frac{1}{J \times (T-n-h+1)} \sum_{\kappa=n}^{T-h} \sum_{m=1}^{101} \left[\mathcal{X}^{(i)}_{\kappa+h}(u_m) - \widehat{\mathcal{X}}^{(i)}_{\kappa+h\,|\,\kappa}(u_m) \right]^2},$$

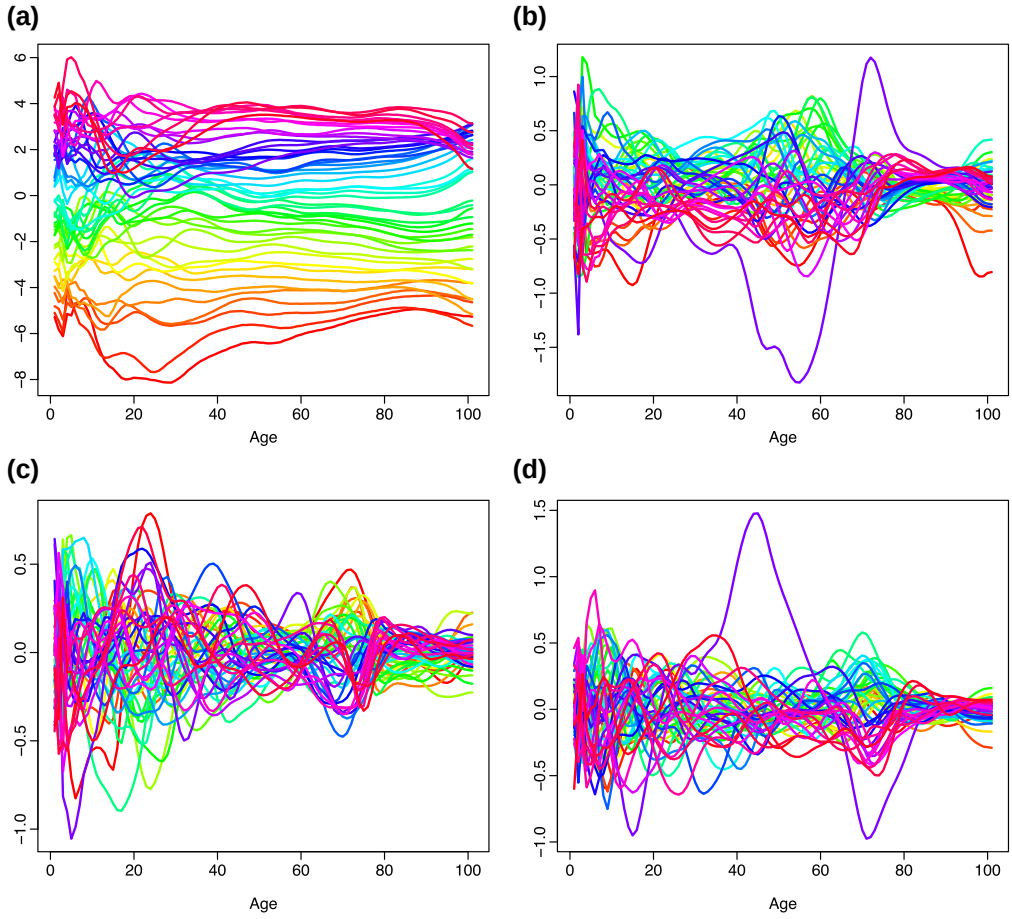

Figure 4. First four sets of common FTS ($\mathcal{F}_t(u)$) of selected prefectures (Tokyo, Nagano, Kyoto, and Osaka) in Japan. (a) First set of common FTS. (b) Second set of common FTS. (c) Third set of common FTS. (d) Fourth set of common FTS.

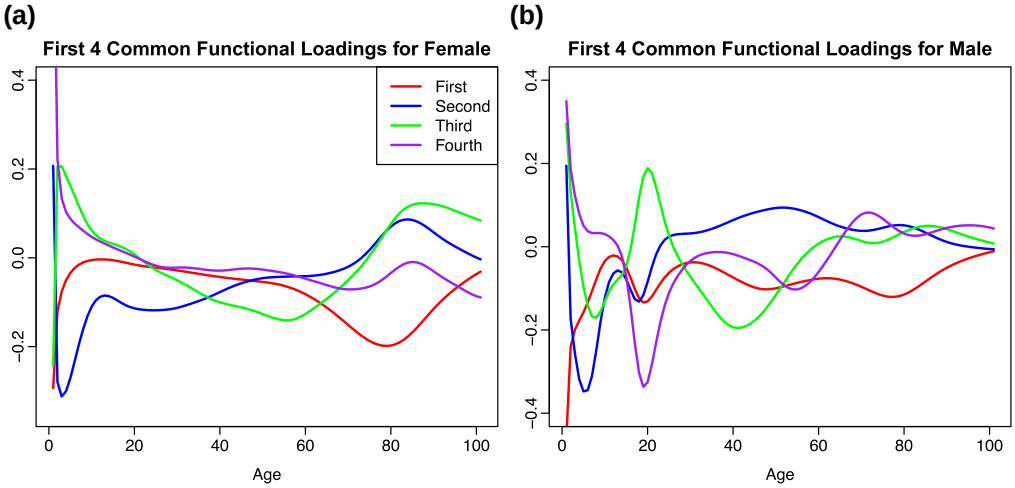

Figure 5. The first four common factor loadings ($\boldsymbol{\phi}(\boldsymbol{u})$) for female and male data. (a) Common factor loadings for female. (b) Common factor loadings for male.

Table 1. The first six normalized eigenvalues for the auto-cross-covariance function of the common FTS and the long-run-covariance function of four selected prefectures for both genders are reported

	Female					Male				
	Common	Specific				Common	Specific			
Eigenvalue		Tokyo	Nagano	Kyoto	Osaka		Tokyo	Nagano	Kyoto	Osaka
1st	0.901	0.973	0.957	0.961	0.971	0.873	0.959	0.933	0.937	0.956
2nd	0.055	0.018	0.026	0.018	0.016	0.048	0.019	0.024	0.029	0.024
3rd	0.023	0.004	0.007	0.010	0.006	0.032	0.012	0.015	0.013	0.008
4th	0.008	0.002	0.004	0.005	0.004	0.021	0.004	0.009	0.008	0.005
5th	0.007	0.001	0.002	0.003	0.002	0.009	0.002	0.005	0.005	0.003
6th	0.003	0.001	0.001	0.002	0.001	0.006	0.001	0.005	0.003	0.002
Sum	0.997	0.999	0.997	0.999	1.000	0.989	0.997	0.991	0.995	0.998

Note. The eigenvalues for the auto-cross-covariance function of the common FTS are labelled as 'Common', and the eigenvalues for the long-run-covariance function of the FTS are labelled as 'Specific'.

where κ is the number of observations used in generating the point forecasts, $\mathcal{X}_{\kappa+h}^{(i)}(u_j)$ is the actual value of the $(\kappa+h)^{\text{th}}$ curve of the i^{th} prefecture, $\widehat{\mathcal{X}}_{\kappa+h}^{(i)}(u_j)$ is the h-step-ahead point forecast based on the first κ observations, and J is the number of grid points on each curve. We use the average RMSFE for the h-step-ahead forecasts across all prefectures to reflect the accuracy of the h-step-ahead point forecasts:

$$\overline{\text{RMSFE}}(h) = \frac{1}{N}\sum_{i=1}^{N} \text{RMSFE}^{(i)}(h).$$

Table 2 tabulates the average RMSFE values across all prefectures for one-step-ahead to 10-step-ahead forecasts of different forecast methods. The bold entries highlight the method that produces the most accurate point forecast. Forecasts based on the proposed method have the smallest average RMSFE for both male and female mortality, which indicates that the proposed method outperforms other methods in terms of point forecast. This implies that the proposed model extracts the most information contained in the HDFTS.

6.2.2 Interval forecast accuracy

We use the interval scoring rule of Gneiting and Raftery (2007) and Gneiting and Katzfuss (2014) to evaluate point-wise interval forecast accuracy. The interval score for the point-wise prediction interval of the i^{th} prefecture at time point u_j is

$$\begin{aligned} S_\alpha\left[\widehat{\mathcal{X}}_{\kappa+h|\kappa}^{(i),\text{lb}}(u_m), \widehat{\mathcal{X}}_{\kappa+h|\kappa}^{(i),\text{ub}}(u_m); \mathcal{X}_{\kappa+h}^{(i)}(u_m)\right] &= \left[\widehat{\mathcal{X}}_{\kappa+h|\kappa}^{(i),\text{ub}}(u_m) - \widehat{\mathcal{X}}_{\kappa+h|\kappa}^{(i),\text{lb}}(u_m)\right] \\ &+ \frac{2}{\alpha}\left[\widehat{\mathcal{X}}_{\kappa+h|\kappa}^{(i),\text{lb}}(u_m) - \mathcal{X}_{\kappa+h}^{(i)}(u_m)\right]\mathbb{1}\left\{\mathcal{X}_{\kappa+h}^{(i)}(u_m) < \widehat{\mathcal{X}}_{\kappa+h|\kappa}^{(i),\text{lb}}(u_m)\right\} \\ &+ \frac{2}{\alpha}\left[\mathcal{X}_{\kappa+h}^{(i)}(u_m) - \widehat{\mathcal{X}}_{\kappa+h|\kappa}^{(i),\text{ub}}(u_m)\right]\mathbb{1}\left\{\mathcal{X}_{\kappa+h}^{(i)}(u_m) > \widehat{\mathcal{X}}_{\kappa+h|\kappa}^{(i),\text{ub}}(u_m)\right\}, \end{aligned}$$

where the level of significance α can be chosen conventionally as 0.2. It is not difficult to find that the smaller the interval score, the more accurate the interval forecast. An optimal (which is also minimal) interval score value can be achieved if $\mathcal{X}_{\kappa+h}^{(i)}(u_j)$ is between $\widehat{\mathcal{X}}_{\kappa+h|\kappa}^{(i)\text{lb}}(u_j)$ and $\widehat{\mathcal{X}}_{\kappa+h|\kappa}^{(i)\text{ub}}(u_j)$. Then the mean interval score for the h-step-ahead forecasts of the i^{th} prefecture can be written as

Table 2. Average RMSE values (×100) in the holdout sample based on various forecasting methods are presented

Sex	h	FDFM	TSHDFTS	HDFTS	HDFFM	MFM	CPD	Tucker
Female	1	**0.641**	0.723	2.165	2.272	1.464	0.755	0.741
	2	**0.653**	0.737	2.225	2.208	1.496	0.745	0.737
	3	**0.655**	0.755	2.268	2.261	1.518	0.730	0.703
	4	**0.715**	0.772	2.319	2.362	1.537	0.761	0.729
	5	**0.691**	0.801	2.360	2.397	1.553	0.770	0.774
	6	**0.723**	0.787	2.420	2.499	1.582	0.739	0.739
	7	0.815	0.797	2.470	2.513	1.601	0.741	**0.740**
	8	0.817	0.842	2.516	2.698	1.617	0.810	**0.809**
	9	**0.666**	0.924	2.413	2.921	1.488	0.840	0.833
	10	**0.722**	1.141	2.199	3.296	1.261	0.839	0.857
	Mean	**0.710**	0.828	2.336	2.543	1.512	0.773	0.776
Male	1	**0.998**	2.444	2.119	2.627	2.970	2.703	2.730
	2	**1.004**	2.516	2.172	2.725	2.866	2.766	2.790
	3	**1.030**	2.536	2.218	2.772	2.836	2.783	2.821
	4	**1.053**	2.642	2.258	2.637	2.911	2.795	2.810
	5	**1.086**	2.850	2.304	2.508	2.794	2.812	2.842
	6	**1.093**	3.006	2.372	2.422	2.917	2.841	2.869
	7	**1.112**	3.161	2.432	2.484	3.149	2.865	2.905
	8	**1.138**	2.634	2.473	2.496	2.937	2.913	2.954
	9	**1.290**	1.601	2.322	2.358	2.798	2.892	2.952
	10	**1.578**	2.119	2.025	2.052	2.968	2.682	2.756
	Mean	**1.138**	2.551	2.270	2.508	2.915	2.805	2.843

Note. Forecasts based on the proposed method are labelled as 'FDFM', forecasts based on high-dimensional functional time series are labelled as 'HDFTS', forecasts based on the two-step high-dimensional functional time series model are labelled as 'TSHDFTS', forecasts based on high-dimensional functional factor models are labelled as 'HDFFM', forecasts based on matrix factor models are labelled as 'MFM', forecasts based on CPD tensor decompositions are labelled as 'CPD', and forecasts based on Tucker tensor decompositions are labelled as 'Tucker'.

$$\overline{S}_{\alpha}^{(i)}(h)=\frac{1}{101\times(T-n-h+1)}\sum_{\kappa=n}^{T-h}\sum_{m=1}^{101}S_{\alpha}\left[\widehat{\mathcal{X}}_{\kappa+h\,|\,\kappa}^{(i),\mathrm{lb}}(u_m),\,\widehat{\mathcal{X}}_{\kappa+h\,|\,\kappa}^{(i),\mathrm{ub}}(u_m);\,\mathcal{X}_{\kappa+h}^{(i)}(u_m)\right].$$

We use the average mean interval score for the h-step-ahead forecasts across all prefectures to reflect the accuracy of the h-step-ahead point-wise prediction interval:

$$\overline{S}_{\alpha}(h)=\frac{1}{N}\sum_{i=1}^{N}\overline{S}_{\alpha}^{(i)}(h).$$

Figure 6 shows the side-by-side box plots of the average mean interval score values across all prefectures for one-step-ahead to 10-step-ahead forecasts of different forecast methods. From the box plots, despite HDFTS performing better for female mortality, the proposed method performs relatively the same as HDFTS, which indicates that the proposed method could generate relatively smaller average mean interval score values for some of the forecast horizons. This demonstrates that the proposed method is capable of producing more accurate interval forecasts.

After comparing the point and interval forecast results of various methods, we found that our proposed method outperforms other competitive methods in forecasting HDFTS. This indicates that the proposed model's dimension reduction process efficiently captures both serial dependence and correlation among cross sections in the HDFTS.

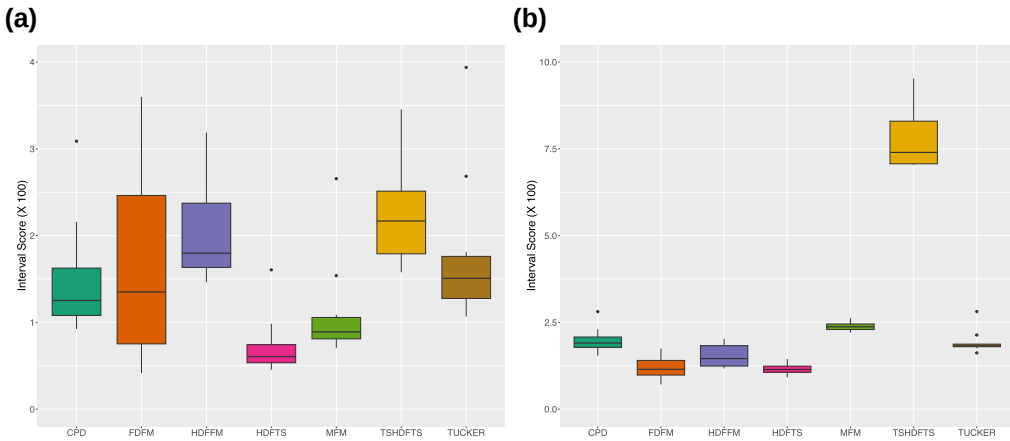

Figure 6. Average interval score values (×100) in the holdout sample of various forecasting methods. (a) Mean interval scores of various forecasting methods for female mortality. (b) Mean interval scores of various forecasting methods for male mortality.

Table 3. Computational time (in seconds) of the considered high-dimensional mortality forecasting methods

	Model	Computation Time (seconds)
Our method	FDFM	187.12
Gao et al. (2019)	HDFTS	68.95
Wang et al. (2019)	MFM	95.94
Dong et al. (2020)	CPD	0.20
Dong et al. (2020)	Tucker	0.39
Hallin et al. (2023) and Tavakoli et al. (2023)	HDFFM	108.14

6.2.3 Computation speed

We also compare the computational speed of various high-dimensional mortality forecasting methods. The methods shown in Table 3 are applied sequentially to the same training data set to obtain one-step-ahead point forecasts, with their computation time recorded in seconds. It is worth noting that CPD's fast computation speed is due to the time efficiency of Tucker's decomposition embedded in the algorithm. Our proposed method takes 50% – 90% more time to implement compared to other functional data analysis approaches shown in the table. However, the superior forecast accuracy of the proposed FDFM method justifies the additional computation time. The TSHDFTS method of Chang et al. (2024) selects the parameters by cross-validation, which requires a significant computation time in implementation. For this reason, we left the method out of the comparison presented in the table.

6.3 Life annuity pricing

Life annuities are one of the typical longevity insurance products for retirees, helping them finance their retirements. Rapid improvements in mortality have exposed life insurers to longevity risk (Ngai & Sherris, 2011). Accurate mortality forecasts could enable life insurers to manage longevity risk effectively without holding excessive levels of capital. To illustrate the impact of these forecasting improvements, we use mortality forecasts to price life annuities, that is, the amount of money an individual pays to a life insurer in return for annual payments after retirement until death. We compare the present values of the life annuities, that is, the amount of capital that the life insurer should reserve, based on mortality forecasts from different methods.

In comparing the value of a life annuity, we calculate the present value (PV) of the life annuity with \$1 annual payments. More specifically, the price of a life annuity for an individual aged x in year t is the PV of the annual payments of \$1 that the individual receives after retirement until death or a pre-agreed age (whichever occurs first). The retirement age is set to be 65, and the pre-agreed age at which the annuity terminates is assumed to be 90. Then the price of the annuity can be calculated as follows:

$$\mathrm{PV}_{x,t} = \begin{cases} \sum_{n=1}^{90-x} \frac{{}_np_{x,t}}{(1+i)^n}, & x \geq 65 \\ \sum_{n=1}^{25} \frac{{}_np_{65,t+(65-x)}}{(1+i)^{n+(65-x)}}, & x < 65, \end{cases}$$

where $\mathrm{PV}_{x,t}$ is the PV of the life annuity for an individual aged x in year t, ${}_np_{x,t}$ is the survival probability of an individual aged x at year t to survive after n years, and i is the interest rate used to discount. An individual older than 65-year-old receives payment for each year of survival, and for an individual younger than 65-year-old, the annuity is deferred with the first payment paid out in the year that they survive their 66^{th} birthday.

To demonstrate the financial impact of the forecasting improvement in mortality, we compare annuity prices based on mortality forecasting using different methods. We use the Japanese sub-national mortality data of 47 prefectures from 1975 to 2012 as the training data set and the data from years 2013 to 2022 as test data. We forecast the mortality rates for the test data based on the training data using different methods. Then we calculate the annuity prices, $\mathrm{PV}_{x,t}$, using the forecasts of mortality rates from different methods, as well as the actual mortality rates.

Table 4 exhibits the average prices of annuities with annual payment 1 and interest rate 2% in Japan for some selected ages and years. Most forecasting methods tend to underestimate annuity prices, which is a common phenomenon in actuarial science, corresponding to the underestimated longevity risk (Ngai & Sherris, 2011). Compared to others, our methods slightly overestimate

Table 4. Average annuity prices with annual payment 1 and interest rate 2% in Japan for selected ages and years

Sex	(Year, Age)	TRUE	FDFM	HDFTS	HDFFM	MFM	CPD	Tucker
Female	(1980, 55)	13.026	13.032	12.964	12.940	12.997	13.040	13.035
	(1985, 60)	14.724	14.731	14.654	14.627	14.691	14.740	14.734
	(1990, 65)	15.786	15.794	15.707	15.676	15.749	15.805	15.798
	(1995, 70)	12.980	12.989	12.888	12.852	12.937	13.001	12.994
	(2000, 75)	10.042	10.053	9.933	9.890	9.990	10.067	10.058
	(2005, 80)	6.976	6.990	6.840	6.787	6.912	7.007	6.996
	(2010, 85)	3.776	3.795	3.590	3.518	3.688	3.819	3.804
Male	(1980, 55)	10.081	10.085	10.057	10.057	10.067	10.090	10.090
	(1985, 60)	11.698	11.702	11.670	11.670	11.682	11.708	11.708
	(1990, 65)	12.863	12.869	12.830	12.830	12.844	12.876	12.875
	(1995, 70)	10.411	10.418	10.371	10.370	10.388	10.427	10.426
	(2000, 75)	7.981	7.990	7.928	7.928	7.950	8.001	8.001
	(2005, 80)	5.571	5.584	5.495	5.495	5.527	5.600	5.599
	(2010, 85)	3.161	3.182	3.031	3.029	3.085	3.210	3.209

Note. Annuity prices based on true mortality rates are labelled as 'TRUE', annuity prices based on the proposed method are labelled as 'FDFM', annuity prices based on high-dimensional functional time series are labelled as 'HDFTS', annuity prices based on high-dimensional functional factor models are labelled as 'HDFFM', annuity prices based on a matrix factor model are labelled as 'MFM', annuity prices based on CPD tensor decompositions are labelled as 'CPD' and annuity prices based on Tucker tensor decompositions are labelled as 'Tucker'.

annuity prices. Further investigation of the annuity prices reveals that the pricing errors of the proposed method are much lower than those of other methods.

To illustrate the financial impact of the mispricing on life insurers, consider the annuity pricing for females aged 75 at year 2000. The pricing error for the proposed method is \$0.011 per \$1 payment. However, the pricing errors for other competing methods range from \$0.016 to \$0.152 per \$1 payment. Suppose the annual payment for each individual is \$10,000, and 50, 000 people purchased this product, based on the proposed method, the capital that the insurer needs to reserve is reduced by at least 2.5 million $((0.016 - 0.011) \times 10,000 \times 50,000 = 2.5\text{million})$, a significant saving to the life insurer.

7 Conclusion

We propose a factor model for HDFTS. By identifying homogeneity, we can reduce the HDFTS to an FTS of low dimensions. Then, a common functional front-loading can be determined to extract the temporal information in HDFTS into a low-dimensional FTS. The proposed model provides more flexibility, with certain choices of front and back-loadings; our model coincides with some of the existing factor models for HDFTS.

Via a series of Monte-Carlo simulations, we demonstrate the performance of the proposed method for in-sample model fitting and out-of-sample forecast accuracy. An empirical study on age-specific mortality rates in 47 Japanese prefectures demonstrates the superiority of the proposed method over existing models in forecasting. This result shows that the proposed model can extract temporal information more effectively.

The proposed model can be viewed as a functional panel data model with interactive effects. We considered the multiplicity of time trend effects and population-specific effects instead of addition as in additive fixed-effects models.

In the current work, functional concurrent regression is performed for each $t \in \{1, \ldots, T\}$ to obtain the FTS of lower dimensions, which omits the temporal dependence of functional observations. A possible future study could extend our functional concurrent regression to the FTS regression of Pham and Panaretos (2018). This awaits further investigation.

Acknowledgments

The authors thank the referees for their valuable feedback, which greatly improved the article. The usual disclaimer applies.

Conflicts of interest: None declared.

Funding

This research was supported by the Australian Research Council Discovery Project (DP230102250) and Australian Research Council Future Fellowship (FT240100338).

Data availability

Code can be accessed via: https://github.com/tangchen90/Functional-Dual-Factor-Models.

References

Ahn S. C., & Horenstein A. R. (2013). Eigenvalue ratio test for the number of factors. *Econometrica: Journal of the Econometric Society*, *81*(3), 1203–1227. https://doi.org/10.3982/ECTA8968

Andrews D. (1991). Heteroskedasticity and autocorrelation consistent covariant matrix estimation. *Econometrica: Journal of the Econometric Society*, *59*(3), 817–858. https://doi.org/10.2307/2938229

Bai J. (2009). Panel data models with interactive fixed effects. *Econometrica: Journal of the Econometric Society*, *77*(4), 1229–1279. https://doi.org/10.3982/ECTA6135

Bai J., & Ng S. (2008). Large dimensional factor analysis. *Foundations and Trends® in Econometrics*, *3*(2), 89–163. https://doi.org/10.1561/0800000002

Bathia N., Yao Q., & Ziegelmann F. (2010). Identifying the finite dimensionality of curve time series. *Annals of Statistics*, *38*(6), 3352–3386. https://doi.org/10.1214/10-AOS819

Bellman R. (1966). Dynamic programming. *Science*, *153*(3731), 34–37. https://doi.org/10.1126/science.153.3731.34
Bühlmann P., & van de Geer S (2011). *Statistics for high-dimensional data: Methods, theory and applications*. Springer Science & Business Media.
Cai T., & Chen X (2010). *High-dimensional data analysis*. Higher Education Press.
Chang J., Fang Q., Qiao X., & Yao Q. (2024). On the modelling and prediction of high-dimensional functional time series. *Journal of the American Statistical Association*, 1–15. https://doi.org/10.1080/01621459.2024.2413201. in press.
Chiou J.-M. (2012). Dynamical functional prediction and classification, with application to traffic flow prediction. *The Annals of Applied Statistics*, *6*(4), 1588–1614. https://doi.org/10.1214/12-AOAS595
Chiou J.-M., & Müller H.-G. (2014). Linear manifold modelling of multivariate functional data. *Journal of the Royal Statistical Society: Series B, Statistical Methodology*, *76*(3), 605–626. https://doi.org/10.1111/rssb.12038
Di C.-Z., Crainiceanu C. M., Caffo B. S., & Punjabi N. M. (2009). Multilevel functional principal component analysis. *The Annals of Applied Statistics*, *3*(1), 458–488. https://doi.org/10.1214/08-AOAS206
Dong Y., Huang F., Yu H., & Haberman S. (2020). Multi-population mortality forecasting using tensor decomposition. *Scandinavian Actuarial Journal*, *2020*(8), 754–775. https://doi.org/10.1080/03461238.2020.1740314
Freyberger J. (2018). Non-parametric panel data models with interactive fixed effects. *The Review of Economic Studies*, *85*(3), 1824–1851. https://doi.org/10.1093/restud/rdx052
Gallant A. R (2009). *Nonlinear statistical models*. John Wiley & Sons.
Gao Y., & Shang H. L. (2017). Multivariate functional time series forecasting: Application to age-specific mortality rates. *Risks*, *5*(2), 21. https://doi.org/10.3390/risks5020021
Gao Y., Shang H. L., & Yang Y. (2019). High-dimensional functional time series forecasting: An application to age-specific mortality rates. *Journal of Multivariate Analysis*, *170*, 232–243. https://doi.org/10.1016/j.jmva.2018.10.003
Gneiting T., & Katzfuss M. (2014). Probabilistic forecasting. *Annual Review of Statistics and Its Application*, *1*(1), 125–151. https://doi.org/10.1146/statistics.2013.1.issue-1
Gneiting T., & Raftery A. E. (2007). Strictly proper scoring rules, prediction, and estimation. *Journal of the American Statistical Association: Review Article*, *102*(477), 359–378. https://doi.org/10.1198/016214506000001437
Hall P., & Vial C. (2006). Assessing the finite dimensionality of functional data. *Journal of the Royal Statistical Society: Series B, Statistical Methodology*, *68*(4), 689–705. https://doi.org/10.1111/j.1467-9868.2006.00562.x
Hallin M., Nisol G., & Tavakoli S. (2023). Factor models for high-dimensional functional time series I: Representation results. *Journal of Time Series Analysis*, *44*(5-6), 578–600. https://doi.org/10.1111/jtsa.v44.5-6
Hansen L. P. (1982). Large sample properties of generalized method of moments estimators. *Econometrica: Journal of the Econometric Society*, *50*(4), 1029–1054. https://doi.org/10.2307/1912775
Hörmann S., & Jammoul F. (2022). Preprocessing noisy functional data: A multivariate perspective. *Electronic Journal of Statistics*, *16*(2), 6232–6266. https://doi.org/10.1214/22-EJS2083
Hörmann S., & Kidziński Ł. (2015). A note on estimation in Hilbertian linear models. *Scandinavian Journal of Statistics*, *42*(1), 43–62. https://doi.org/10.1111/sjos.v42.1
Hörmann S., Kidziński Ł., & Hallin M. (2015). Dynamic functional principal components. *Journal of the Royal Statistical Society: Series B, Statistical Methodology*, *77*(2), 319–348. https://doi.org/10.1111/rssb.12076
Hörmann S., & Kokoszka P. (2010). Weakly dependent functional data. *Annals of Statistics*, *38*(3), 1845–1884. https://doi.org/10.1214/09-AOS768
Horváth L., Kokoszka P., & Reeder R. (2013). Estimation of the mean of functional time series and a two-sample problem. *Journal of the Royal Statistical Society: Series B, Statistical Methodology*, *75*(1), 103–122. https://doi.org/10.1111/j.1467-9868.2012.01032.x
Hyndman R. J., & Shang H. L. (2010). Rainbow plots, bagplots, and boxplots for functional data. *Journal of Computational and Graphical Statistics*, *19*(1), 29–45. https://doi.org/10.1198/jcgs.2009.08158
Hyndman R. J., & Ullah M. S. (2007). Robust forecasting of mortality and fertility rates: A functional data approach. *Computational Statistics & Data Analysis*, *51*(10), 4942–4956. https://doi.org/10.1016/j.csda.2006.07.028
Japanese Mortality Database (2025). *National Institute of Population and Social Security Research*. Accessed at April 12, 2024. http://www.ipss.go.jp/p-toukei/JMD/.
Jiménez-Varón C. F., Sun Y., & Shang H. L. (2024). Forecasting high-dimensional functional time series: Application to sub-national age-specific mortality. *Journal of Computational and Graphical Statistics*, *33*(4), 1160–1174. https://doi.org/10.1080/10618600.2024.2319166
Jiménez-Varón C. F., Sun Y., & Shang H. L. (2025). Forecasting density-valued functional panel data. *Australian & New Zealand Journal of Statistics*, *67*(3), 401–415. https://doi.org/10.1111/anzs.70013

Lam C., & Yao Q. (2012). Factor modeling for high-dimensional time series: Inference for the number of factors. *Annals of Statistics*, *40*(2), 694–726. https://doi.org/10.1214/12-AOS970
Lam C., Yao Q., & Bathia N. (2011). Estimation of latent factors for high-dimensional time series. *Biometrika*, *98*(4), 901–918. https://doi.org/10.1093/biomet/asr048
Lee R. D., & Carter L. R. (1992). Modeling and forecasting U.S. mortality. *Journal of the American Statistical Association*, *87*(419), 659–671. https://doi.org/10.2307/2290201
Li D., Robinson P. M., & Shang H. L. (2020). Long-range dependent curve time series. *Journal of the American Statistical Association: Theory and Methods*, *115*(530), 957–971. https://doi.org/10.1080/01621459.2019.1604362
Li J. (2013). A Poisson common factor model for projecting mortality and life expectancy jointly for females and males. *Population Studies*, *67*(1), 111–126. https://doi.org/10.1080/00324728.2012.689316
Li N., & Lee R. (2005). Coherent mortality forecasts for a group of populations: An extension of the Lee-Carter method. *Demography*, *42*(3), 575–594. https://doi.org/10.1353/dem.2005.0021
Li N., Lee R., & Gerland P. (2013). Extending the Lee-Carter method to model the rotation of age patterns of mortality decline for long-term projections. *Demography*, *50*(6), 2037–2051. https://doi.org/10.1007/s13524-013-0232-2
Martínez-Hernández I., Gonzalo J., & González-Farías G. (2022). Nonparametric estimation of functional dynamic factor model. *Journal of Nonparametric Statistics*, *34*(4), 895–916. https://doi.org/10.1080/10485252.2022.2080825
Newey W. K., & West K. D. (1987). A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica: Journal of the Econometric Society*, *55*(3), 703–708. https://doi.org/10.2307/1913610
Ngai A., & Sherris M. (2011). Longevity risk management for life and variable annuities: The effectiveness of static hedging using longevity bonds and derivatives. *Insurance: Mathematics and Economics*, *49*(1), 100–114. https://doi.org/10.1016/j.insmatheco.2011.02.009
Pampel F. (2005). Forecasting sex differences in mortality in high income nations: The contribution of smoking. *Demographic Research*, *13*(18), 455–484. https://doi.org/10.4054/DemRes.2005.13.18
Panaretos V. M., & Tavakoli S. (2013a). Cramér–Karhunen–Loève representation and harmonic principal component analysis of functional time series. *Stochastic Processes and their Applications*, *123*(7), 2779–2807. https://doi.org/10.1016/j.spa.2013.03.015
Panaretos V. M., & Tavakoli S. (2013b). Fourier analysis of stationary time series in function space. *Annals of Statistics*, *41*(2), 568–603. https://doi.org/10.1214/13-AOS1086
Pham T., & Panaretos V. M. (2018). Methodology and convergence rates for functional time series regression. *Statistica Sinica*, *28*(4), 2521–2539. https://doi.org/10.5705/ss.202016.0536
Politis D. N., & Romano J. P. (1996). On flat-top kernel spectral density estimators for homogeneous random fields. *Journal of Statistical Planning and Inference*, *51*(1), 41–53. https://doi.org/10.1016/0378-3758(95)00069-0
Politis D. N., & Romano J. P. (1999). Multivariate density estimation with general flat-top kernels of infinite order. *Journal of Multivariate Analysis*, *68*(1), 1–25. https://doi.org/10.1006/jmva.1998.1774
Ramsay J. O., Hooker G., & Graves S (2009). *Functional data analysis with R and MATLAB*. Springer.
Ramsay J. O., & Silverman B. W (2006). *Functional data analysis*. Springer Science & Business Media.
Renshaw A. E., & Haberman S. (2003). Lee-Carter mortality forecasting with age-specific enhancement. *Insurance: Mathematics and Economics*, *33*(2), 255–272. https://doi.org/10.1016/S0167-66870300138-0
Rice G., & Shang H. L. (2017). A plug-in bandwidth selection procedure for long-run covariance estimation with stationary functional time series. *Journal of Time Series Analysis*, *38*(4), 591–609. https://doi.org/10.1111/jtsa.v38.4
Rice J. A., & Silverman B. W. (1991). Estimating the mean and covariance structure nonparametrically when the data are curves. *Journal of the Royal Statistical Society: Series B, Statistical Methodology*, *53*(1), 233–243. https://doi.org/10.1111/j.2517-6161.1991.tb01821.x
Shang H. L. (2016). Mortality and life expectancy forecasting for a group of populations in developed countries: A multilevel functional data method. *The Annals of Applied Statistics*, *10*(3), 1639–1672. https://doi.org/10.1214/16-AOAS953
Shang H. L. (2018). Bootstrap methods for stationary functional time series. *Statistics and Computing*, *28*(1), 1–10. https://doi.org/10.1007/s11222-016-9712-8
Shang H. L. (2020). Dynamic principal component regression for forecasting functional time series in a group structure. *Scandinavian Actuarial Journal*, *2020*(4), 307–322. https://doi.org/10.1080/03461238.2019.1663553
Tang C., Shang H. L., & Yang Y. (2022). Clustering and forecasting multiple functional time series. *The Annals of Applied Statistics*, *16*(4), 2523–2553. https://doi.org/10.1214/22-AOAS1602

Tavakoli S., Nisol G., & Hallin M. (2023). Factor models for high-dimensional functional time series II: Estimation and forecasting. *Journal of Time Series Analysis*, *44*(5-6), 601–621. https://doi.org/10.1111/jtsa.v44.5-6

Tsay R. S (2013). *Multivariate time series analysis: With R and financial applications*. John Wiley & Sons.

Tsay R. S. (2024). Matrix-variate time series analysis: A brief review and some new developments. *International Statistical Review*, *92*(2), 246–262. https://doi.org/10.1111/insr.v92.2

Wang D., Liu X., & Chen R. (2019). Factor models for matrix-valued high-dimensional time series. *Journal of Econometrics*, *208*(1), 231–248. https://doi.org/10.1016/j.jeconom.2018.09.013

White H (1984). *Asymptotic theory for econometricians*. Academic Press.

Wiśniowski A., Smith P. W., Bijak J., Raymer J., & Forster J. J. (2015). Bayesian population forecasting: Extending the Lee-Carter method. *Demography*, *52*(3), 1035–1059. https://doi.org/10.1007/s13524-015-0389-y

Yang Y., Shang H. L., & Cohen J. E. (2022). Temporal and spatial Taylor's law: Application to Japanese subnational mortality rates. *Journal of the Royal Statistical Society Series A*, *185*(4), 1979–2006. https://doi.org/10.1111/rssa.12859

Yao F., Müller H.-G., & Wang J.-L. (2005). Functional data analysis for sparse longitudinal data. *Journal of the American Statistical Association: Theory and Methods*, *100*(470), 577–590. https://doi.org/10.1198/016214504000001745

Zhang B., Pan G., Yao Q., & Zhou W. (2024). Factor modeling for clustering high-dimensional time series. *Journal of the American Statistical Association: Theory and Methods*, *119*(546), 1252–1263. https://doi.org/10.1080/01621459.2023.2183132

Zhou Z., & Dette H. (2023). Statistical inference for high-dimensional panel functional time series. *Journal of the Royal Statistical Society: Series B, Statistical Methodology*, *85*(2), 523–549. https://doi.org/10.1093/jrsssb/qkad015

Zivot E., & Wang J (2006). *Modeling financial time series with S-PLUS* (Vol. 2). Springer.