

On the Modeling and Prediction of High-Dimensional Functional Time Series

Jinyuan Chang^{a,b,c} , Qin Fang^d , Xinghao Qiao^e, and Qiwei Yao^f 

^aJoint Laboratory of Data Science and Business Intelligence, Southwestern University of Finance and Economics, Chengdu, Sichuan, China; ^bBig Data Laboratory on Financial Security and Behavior (MOE Philosophy and Social Sciences Laboratory), Southwestern University of Finance and Economics, Chengdu, Sichuan, China; ^cAcademy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing, China; ^dBusiness School, University of Sydney, Sydney, Australia; ^eFaculty of Business and Economics, The University of Hong Kong, Pokfulam, Hong Kong, China; ^fDepartment of Statistics, London School of Economics, London, U.K.

ABSTRACT

We propose a two-step procedure to model and predict high-dimensional functional time series, where the number of function-valued time series p is large in relation to the length of time series n . Our first step performs an eigenanalysis of a positive definite matrix, which leads to a one-to-one linear transformation for the original high-dimensional functional time series, and the transformed curve series can be segmented into several groups such that any two subseries from any two different groups are uncorrelated both contemporaneously and serially. Consequently in our second step those groups are handled separately without the information loss on the overall linear dynamic structure. The second step is devoted to establishing a finite-dimensional dynamical structure for all the transformed functional time series within each group. Furthermore the finite-dimensional structure is represented by that of a vector time series. Modeling and forecasting for the original high-dimensional functional time series are realized via those for the vector time series in all the groups. We investigate the theoretical properties of our proposed methods, and illustrate the finite-sample performance through both extensive simulation and two real datasets. Supplementary materials for this article are available online, including a standardized description of the materials available for reproducing the work.

ARTICLE HISTORY

Received January 2023
Accepted October 2024

KEYWORDS

Dimension reduction;
Eigenanalysis; Functional
thresholding;
Hilbert–Schmidt norm;
Permutation; Segmentation
transformation

1. Introduction

Functional time series typically refers to continuous-time records that are naturally divided into consecutive time intervals, such as days, months or years. With recent advances in data collection technology, multivariate or even high-dimensional functional time series arise ubiquitously in many applications, including daily pollution concentration curves over different locations, annual temperature curves at different stations, annual age-specific mortality rates for different countries, and intraday energy consumption trajectories from different households. Those data can be represented as a p -dimensional functional time series $\mathbf{Y}_t(u) = \{Y_{t1}(u), \dots, Y_{tp}(u)\}^\top$ defined on a compact set $u \in \mathcal{U}$, and we observe $\mathbf{Y}_t(\cdot)$ for $t = 1, \dots, n$. In this article we tackle the high-dimensional settings when the dimension p is comparable to, or even greater than, the sample size n , which poses new challenges in modeling and forecasting $\mathbf{Y}_t(\cdot)$.



By assuming $\mathbf{Y}_t(\cdot)$ is stationary, a conventional approach is first to extract features by performing dimension reduction for each component series $Y_{tj}(\cdot)$ separately via, for example functional principal component analysis (FPCA) or dynamic FPCA (Bathia, Yao, and Ziegelmann 2010; Hörmann, Kidziński, and Hallin 2015), and then to model p vector time series by, for example, regularized vector autoregressions (Guo and Qiao 2023) or factor model (Gao, Shang, and Yang 2019). However, more effective dimension-reduction can be achieved by pulling

together the information from different component series in the first place. This is in the same spirit of multivariate FPCA (Chiou, Chen, and Yang 2014; Happ and Greven 2018) and sparse FPCA (Hu and Yao 2022), though those approaches make no use of the information on the serial dependence which is the most relevant for future prediction.

To achieve more effective dimension reduction and better predictive performance, we propose in this article a two-step approach. Our first step is a segmentation transformation step in which we seek for a linear transformation $\mathbf{Y}_t(\cdot) = \mathbf{A}\mathbf{Z}_t(\cdot)$, where \mathbf{A} is a $p \times p$ invertible constant matrix, such that the transformed series $\mathbf{Z}_t(\cdot) = \{\mathbf{Z}_t^{(1)}(\cdot)^\top, \dots, \mathbf{Z}_t^{(q)}(\cdot)^\top\}^\top$ can be segmented into q groups $\mathbf{Z}_t^{(1)}(\cdot), \dots, \mathbf{Z}_t^{(q)}(\cdot)$, and curve subseries $\mathbf{Z}_t^{(i)}(\cdot)$ and $\mathbf{Z}_t^{(j)}(\cdot)$ are uncorrelated at all time lags for any $i \neq j$, that is,

$$\text{cov}\{\mathbf{Z}_t^{(i)}(u), \mathbf{Z}_{t+k}^{(j)}(v)\} = \mathbf{0}, \quad (u, v) \in \mathcal{U}^2 \quad \text{and} \\ k = 0, \pm 1, \pm 2, \dots$$

Hence, each $\mathbf{Z}_t^{(i)}$ can be modeled and forecasted separately as far as the linear dynamics is concerned. Under the stationarity assumption, the estimation of the transformation matrix \mathbf{A} boils down to the eigenanalysis of a positive definite matrix defined by the double integral of quadratic forms in the autocovariance functions of $\mathbf{Y}_t(\cdot)$. An additional permutation on the components of $\mathbf{Z}_t(\cdot)$ will be specified in order to identify the latent group structure.

CONTACT Xinghao Qiao  xinghaoq@hku.hk  Faculty of Business and Economics, The University of Hong Kong, Pokfulam, Hong Kong, China.

 Supplementary materials for this article are available online. Please go to www.tandfonline.com/r/JASA.

© 2024 The Author(s). Published with license by Taylor & Francis Group, LLC.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. The terms on which this article has been published allow the posting of the Accepted Manuscript in a repository by the author(s) or with their consent.

Our second step is to identify a finite-dimensional dynamic structure for each transformed subseries $\mathbf{Z}_t^{(l)}(\cdot)$ separately, which is based on a latent decomposition

$$\mathbf{Z}_t^{(l)}(u) = \mathbf{X}_t^{(l)}(u) + \boldsymbol{\varepsilon}_t^{(l)}(u), \quad u \in \mathcal{U}, \quad (1)$$

where $\mathbf{X}_t^{(l)}(\cdot)$ represents the dynamics of $\mathbf{Z}_t^{(l)}(\cdot)$, $\boldsymbol{\varepsilon}_t^{(l)}(\cdot)$ is white noise with $\mathbb{E}\{\boldsymbol{\varepsilon}_t^{(l)}(u)\} = \mathbf{0}$ and $\mathbb{E}\{\boldsymbol{\varepsilon}_t^{(l)}(u)\boldsymbol{\varepsilon}_s^{(l)}(v)^\top\} = \mathbf{0}$ for any $(u, v) \in \mathcal{U}^2$ and $t \neq s$, and $\{\mathbf{X}_t^{(l)}(\cdot)\}_{t=1}^n$ are uncorrelated with $\{\boldsymbol{\varepsilon}_t^{(l)}(\cdot)\}_{t=1}^n$. Furthermore we assume that the dynamic structure of $\mathbf{X}_t^{(l)}(\cdot)$ admits a vector time series presentation via a variational multivariate FPCA. For given $\{\mathbf{Z}_t^{(l)}(\cdot)\}_{t=1}^n$, the standard multivariate FPCA performs dimension reduction based on the eigenanalysis of the sample covariance function of $\mathbf{Z}_t^{(l)}(\cdot)$, which cannot be used to identify the finite-dimensional dynamic structure of $\mathbf{X}_t^{(l)}(\cdot)$ due to the contamination of $\boldsymbol{\varepsilon}_t^{(l)}(\cdot)$. Inspired by the fact that the lag- k ($k \neq 0$) autocovariance function of $\mathbf{Z}_t^{(l)}(\cdot)$ automatically filters out the white noise, our variational multivariate FPCA is based on the eigenanalysis of a positive-definite matrix defined in terms of its nonzero lagged autocovariance functions; leading to a low-dimensional vector time series which bears all the dynamic structure of $\mathbf{X}_t^{(l)}(\cdot)$, and consequently, also that of $\mathbf{Z}_t^{(l)}(\cdot)$. This is possible as the number of components in each $\mathbf{Z}_t^{(l)}(\cdot)$ is usually small in practice. Finally, owing to the one-to-one linear transformation in the segmentation step, the good predictive performance of $\mathbf{Z}_t(\cdot)$ can be easily carried back to $\mathbf{Y}_t(\cdot)$.

Our article makes useful contributions on multiple fronts. First, the segmentation transformation in the first step transforms the serial correlations across different series into the autocorrelations within each of the identified q subseries. This not only avoids the direct modeling of the p functional time series together, but also makes each of those transformed subseries more serially correlated and, hence, more predictable. As the serial correlations across different series are valuable for future prediction, the segmentation provides an effective way to use the information. Note that the prediction directly based on a multivariate ARMA-type model with even a moderately large dimension is not recommendable, as the gain from using the autocorrelations across different component series is often canceled off by the errors in estimating too many parameters. Furthermore, even in the special case with $q = 1$, our decorrelation transformation can effectively push the cross-autocorrelations that are previously spread over p components into a block-diagonally dominate structure, where the cross-autocorrelations along the block diagonal are significantly stronger than those off the diagonal. This still leads to reasonably good segmentation by retaining the strong within-group cross-autocorrelations while ignoring the weak between-group cross-autocorrelations and, as evidenced by simulations in Section 5.3, results in more accurate future predictions than those based on models without transformation. Therefore, the proposed transformation can always be used as an initial step in modeling high-dimensional functional time series.

Second, though the segmentation transformation is motivated from the decorrelation idea of Chang, Guo, and Yao (2018) for vector time series, its adaption to the functional setting introduces additional methodological and theoretical complexities

and requires innovative advancements in both methodology and theory due to the intrinsic infinite-dimensionality of functional data. A simple extension of Chang, Guo, and Yao (2018) would be to apply their method to the p -dimensional vector $\mathbf{Y}_t(u)$ on each evaluation grid value u followed by aggregation, which fails to account for the smoothness and continuity of the functional nature of observed data. In contrast, our proposal on $\mathbf{Y}_t(\cdot)$ implements novel integral-based normalization and uses double integral over $(u, v) \in \mathcal{U}^2$ to fully leverage the autocovariance information, thus, leading to more efficient estimation. Moreover, when performing permutation on the components of the transformed series, our method relies on Hilbert-Schmidt norm to measure the magnitude of bivariate functions, which introduces extra theoretical complexities compared to the absolute value measure used in Chang, Guo, and Yao (2018). Finally, we develop a novel functional thresholding procedure, which guarantees the consistency of our estimation under high-dimensional scaling. Its theoretical analysis involves establishing novel inequalities between functional versions of matrix norms.

Third, the nonzero lagged autocovariance-based dimension reduction approach in the second step makes the good use of the serial dependence information in our estimation, which is most relevant in time series prediction. On the method side, our proposed variational multivariate FPCA extends the univariate method of Bathia, Yao, and Ziegelmann (2010) by incorporating the cross-autocovariance. This extension addresses a crucial gap in dimension-reduction techniques, enabling us to accommodate multivariate functional time series. Importantly, when p is fixed or moderately large, such method can be directly applied to the observed curve series $\mathbf{Y}_t(\cdot)$ for dimension reduction and forecasting purposes. On the theory side, we demonstrate that our proposal exhibits appealing convergence properties despite the additional transformation and estimation errors arisen from the first step, which are not involved in Bathia, Yao, and Ziegelmann (2010). By comparison, standard (multivariate) FPCA methods under (1) suffer from inconsistent estimation and less efficient dimension reduction.

Existing research on functional time series has mainly focused on adapting the univariate or low-dimensional multivariate time series methods to the functional domain. An incomplete list of the relevant references includes Bathia, Yao, and Ziegelmann (2010), Cho et al. (2013), Aue, Norinho, and Hörmann (2015), Hörmann, Kidziński, and Hallin (2015), Aue, Rice, and Sonmez (2018) and Li, Robinson, and Shang (2020). Following the recent emergence of high-dimensional functional time series data, there has been a wave of significant advancements aimed at addressing its complexities. Notable developments include functional factor models (Gao, Shang, and Yang 2019; Tavakoli, Nisol, and Hallin 2023), functional dependence analysis (Guo and Qiao 2023), functional clustering (Tang, Shang, and Yang 2022), statistical inference for mean functions (Zhou and Dette 2023), sparse vector functional autoregressions (Chang et al. 2024), and graphical PCA (Tan et al. in press).

The rest of the article is organized as follows. In Section 2, we develop the methods employed in the first step, that is the segmentation transformation, the permutation and the functional thresholding. Section 3 specifies the variational multivariate FPCA method used in the second dimension reduction step. We investigate the associated theoretical properties of the

proposed methods in Section 4. The finite-sample performance of our methods is examined through extensive simulations in Section 5. Section 6 applies our proposal to two real datasets, revealing its superior predictive performance over most frequently used competitors.

Notation. Denote by $I(\cdot)$ the indicator function. For a positive integer m , write $[m] = \{1, \dots, m\}$ and denote by \mathbf{I}_m the identity matrix of size $m \times m$. For $x, y \in \mathbb{R}$, we use $x \vee y = \max(x, y)$. For two positive sequences $\{a_n\}$ and $\{b_n\}$, we write $a_n \ll b_n$ or $b_n \gg a_n$ if $\limsup_{n \rightarrow \infty} a_n/b_n = 0$. For a $p \times q$ real matrix \mathbf{E} , denote by \mathbf{E}^\top its transpose, and write $\mathbf{E}^{\otimes 2} = \mathbf{E}\mathbf{E}^\top$ and $\|\mathbf{E}\|_2 = \lambda_{\max}^{1/2}(\mathbf{E}^\top \mathbf{E})$, where $\lambda_{\max}(\mathbf{M})$ denotes the largest eigenvalue of the matrix \mathbf{M} . Let $L_2(\mathcal{U})$ be the Hilbert space of square integrable functions defined on \mathcal{U} and equipped with the inner product $\langle f, g \rangle = \int_{\mathcal{U}} f(u)g(u) du$ for $f, g \in L_2(\mathcal{U})$ and the induced norm $\|\cdot\| = \langle \cdot, \cdot \rangle^{1/2}$. For any B in $\mathbb{S} \equiv L_2(\mathcal{U} \times \mathcal{U})$, we denote the Hilbert–Schmidt norm by $\|B\|_{\mathbb{S}} = \{\int_{\mathcal{U}} \int_{\mathcal{U}} B^2(u, v) dudv\}^{1/2}$.

2. Segmentation Transformation

2.1. Linear Decomposition of $\mathbf{Y}_t(\mathbf{u})$

We consider the following linear decomposition of $\mathbf{Y}_t(\mathbf{u})$:

$$\mathbf{Y}_t(\mathbf{u}) = \mathbf{A}\mathbf{Z}_t(\mathbf{u}) = \mathbf{A}_1\mathbf{Z}_t^{(1)}(\mathbf{u}) + \dots + \mathbf{A}_q\mathbf{Z}_t^{(q)}(\mathbf{u}), \quad \mathbf{u} \in \mathcal{U}, \quad (2)$$

where $q \in [p]$ is an unknown positive integer, $\mathbf{A} = (\mathbf{A}_1, \dots, \mathbf{A}_q)$ is a $p \times p$ unknown loading matrix, and $\mathbf{Z}_t(\mathbf{u}) = \{\mathbf{Z}_t^{(1)}(\mathbf{u})^\top, \dots, \mathbf{Z}_t^{(q)}(\mathbf{u})^\top\}^\top$ is a latent p -dimensional functional time series such that $\text{cov}\{\mathbf{Z}_t^{(l)}(\mathbf{u}), \mathbf{Z}_s^{(l')}(v)\} = \mathbf{0}$ for all $t, s \in [n]$, $l \neq l'$ and $(\mathbf{u}, v) \in \mathcal{U}^2$. Such linear decomposition possesses three key properties:

- For any p -dimensional functional time series $\mathbf{Y}_t(\mathbf{u})$, its linear decomposition (2) always exists by setting $q = 1$ and choosing $(\mathbf{A}, \mathbf{Z}_t(\mathbf{u})) = (\mathbf{H}, \mathbf{H}^{-1}\mathbf{Y}_t(\mathbf{u}))$ for some invertible matrix \mathbf{H} .
- The linear decomposition (2) is not uniquely determined. Alternative segmentations of $\mathbf{Z}_t(\mathbf{u})$ can be obtained by merging multiple uncorrelated groups into a single group.
- For a given segmentation, \mathbf{A}_l for $l \in [q]$ cannot be uniquely identified, as within-group rotations will not distort the uncorrelated group structure. In fact, only the linear spaces spanned by the columns of \mathbf{A}_l , denoted by $\mathcal{C}(\mathbf{A}_l)$, $l \in [q]$, are uniquely defined.

Our goal is then to find a linear decomposition (2) for $\mathbf{Y}_t(\cdot)$, where each group $\mathbf{Z}_t^{(l)}(\cdot)$ for $l \in [q]$ cannot be further divided into smaller uncorrelated subgroups. This allows us to model each $\mathbf{Z}_t^{(l)}(\cdot)$ separately, as there are no cross-correlations among them at all time lags. We formalize the inseparability for each $\mathbf{Z}_t^{(l)}(\cdot)$ as Condition 4 in Section 4, which in turn defines the number of groups q and the segmentation of $\mathbf{Z}_t(\cdot)$ in (2). In Section 2.2, we will present the estimation of the number of groups q , the linear spaces $\mathcal{C}(\mathbf{A}_l)$ and the associated transformed subseries $\mathbf{Z}_t^{(l)}(\cdot)$ of group size p_l for $l \in [q]$. Before that, let us first illustrate the validity and benefit of the linear decomposition (2), that is segmentation transformation, in predicting multivariate functional time series with a real-life example. As we will

demonstrate, such a decomposition (2) is commonly achieved with a relatively large q in practice. This effectively reduces the modeling burden while retaining the full linear dynamics of the original curve series $\mathbf{Y}_t(\cdot)$, thus, leading to more accurate predictions.

Example 1. We consider the global age-specific mortality dataset analyzed in Tang, Shang, and Yang (2022). To simplify the presentation, we examine only the female mortality curve series $\mathbf{Y}_t(\cdot)$ with $p = 8$ randomly selected countries (Australia, Canada, Switzerland, Denmark, Finland, Great Britain, Japan and Portugal) and the transformed curve series $\mathbf{Z}_t(\cdot)$ in (2), which are obtained by the proposed method in Section 2.2. Let

$$\hat{\sigma}_{y,k,ij}(\mathbf{u}, v) = \frac{1}{n-k} \sum_{t=1}^{n-k} \{Y_{ti}(\mathbf{u}) - \bar{Y}_i(\mathbf{u})\} \{Y_{(t+k)j}(v) - \bar{Y}_j(v)\}$$

with $\bar{Y}_i(\mathbf{u}) = n^{-1} \sum_{t=1}^n Y_{ti}(\mathbf{u})$. We use

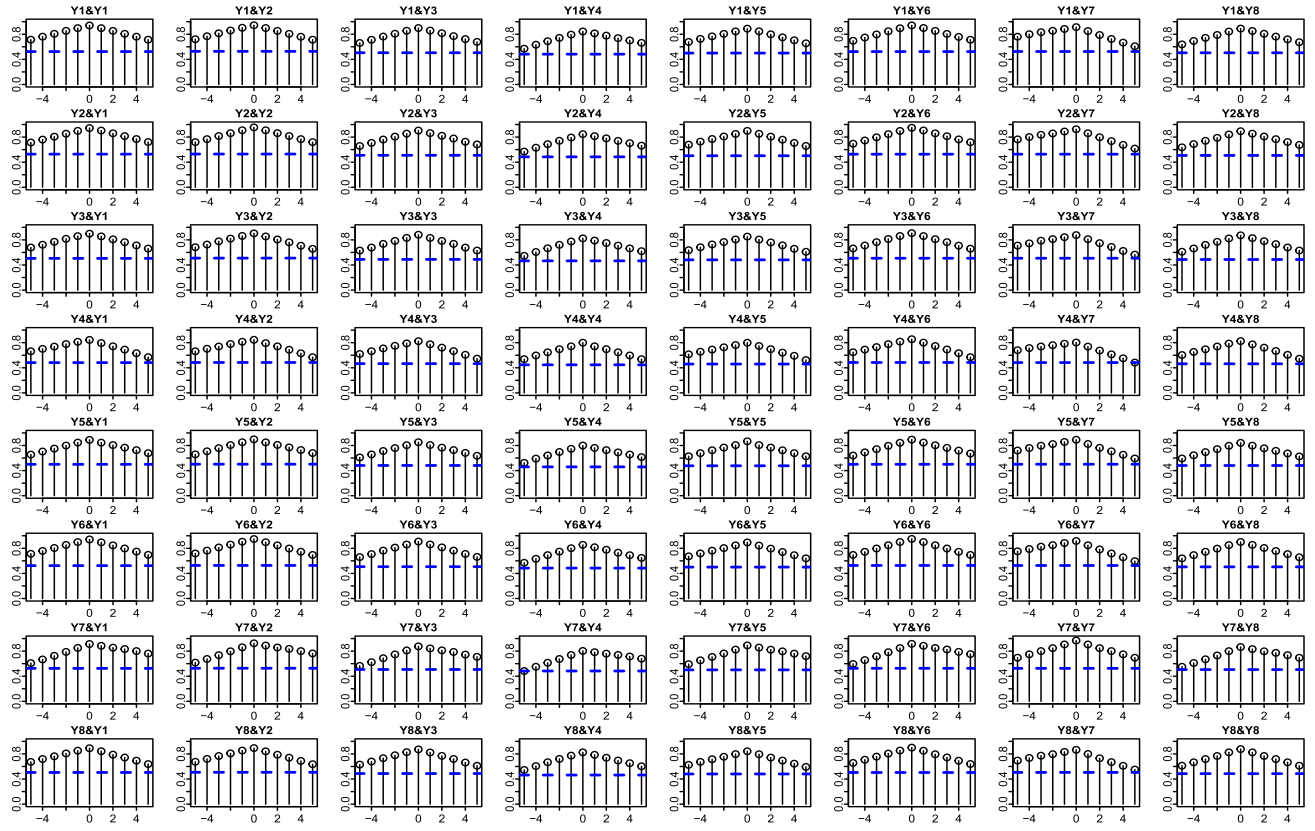
$$\hat{\omega}_{y,k,ij} = \frac{\|\hat{\sigma}_{y,k,ij}\|_{\mathbb{S}}}{\{\int \hat{\sigma}_{y,0,ii}^2(\mathbf{u}, \mathbf{u}) du \int \hat{\sigma}_{y,0,jj}^2(\mathbf{u}, \mathbf{u}) du\}^{1/2}},$$

as proposed by Rice and Shum (2019), to measure the functional cross-autocorrelation between $Y_{ti}(\cdot)$ and $Y_{tj}(\cdot)$ at lag k . Figure 1 displays $\hat{\omega}_{y,k,ij}$ and $\hat{\omega}_{z,k,ij}$ for $-5 \leq k \leq 5$, where $\hat{\omega}_{z,k,ij}$ is defined by substituting each $Y_{ti}(\cdot)$ in $\hat{\omega}_{y,k,ij}$ with $Z_{ti}(\cdot)$. It is evident that the transformation effectively channels the strong cross-autocorrelations over different time lags among all 8 countries into significant autocorrelations within each of the six groups of $\mathbf{Z}_t(\cdot)$, that is, $\{1, 2, 3\}$, $\{4\}$, $\{5\}$, $\{6\}$, $\{7\}$, and $\{8\}$, while the cross-autocorrelations among these six groups are identified as weak and statistically insignificant across all time lags at the 5% significance level.

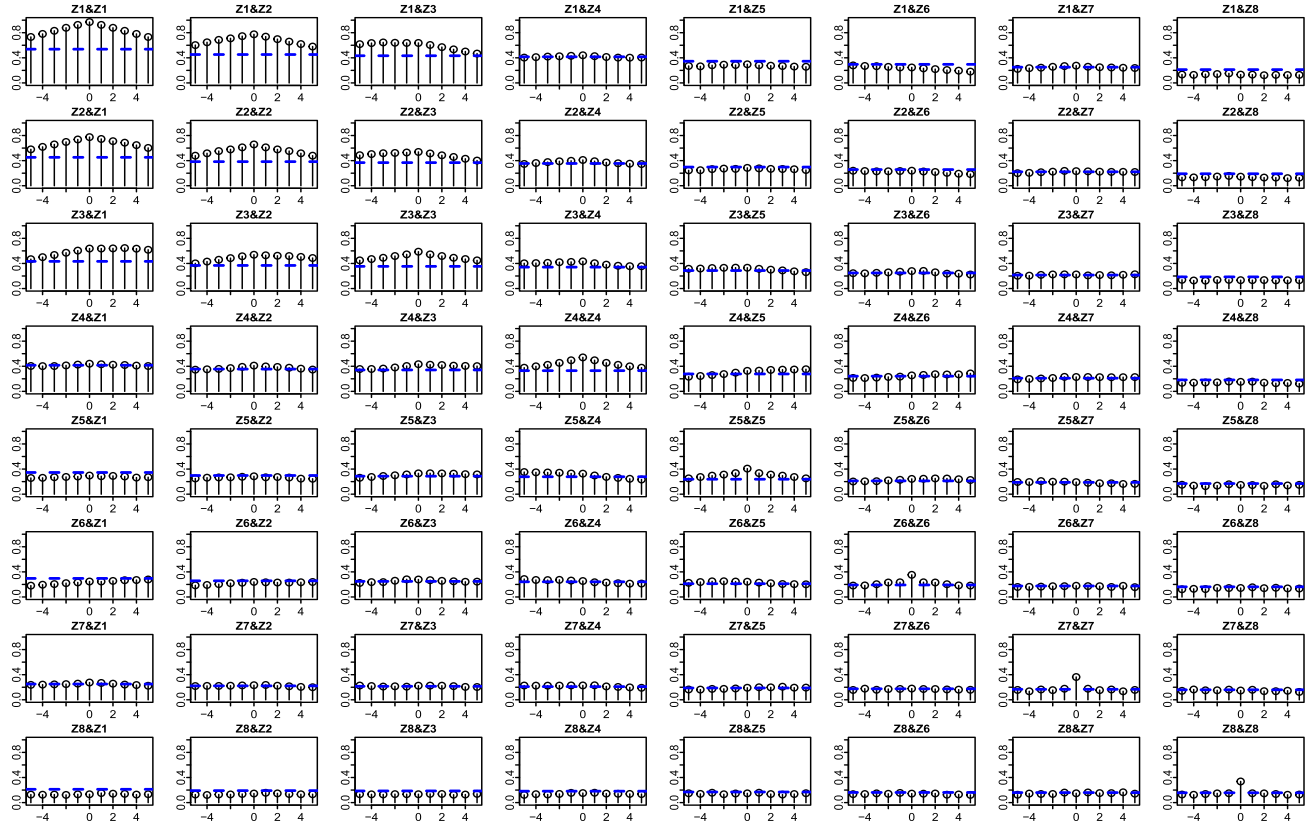
We then implement two prediction methods on $\mathbf{Y}_t(\cdot)$ and $\mathbf{Z}_t(\cdot)$, respectively, to demonstrate that forecasting $\mathbf{Y}_t(\cdot)$ through the forecasting of the transformed series $\mathbf{Z}_t(\cdot)$ can yield more accurate predictive performance than directly forecasting $\mathbf{Y}_t(\cdot)$:

- (Joint prediction) We treat the p components of $\mathbf{Y}_t(\cdot)$ as one group and perform the Variational-multivariate-FPCA-and-VAR-based procedure (VmV), that is Step (ii) of our proposed Algorithm 1 in Section 5.1, on $\mathbf{Y}_t(\cdot)$ directly. Based on the identified group structure by Figure 1, we implement SegV on $\mathbf{Z}_t(\cdot)$, which performs VmV on each of the six groups of $\mathbf{Z}_t(\cdot)$ separately.
- (Marginal prediction) We implement UniV and Uni.SegV, which respectively perform VmV on each component of $\mathbf{Y}_t(\cdot)$ and $\mathbf{Z}_t(\cdot)$ separately.

Note that the difference in each prediction method comes solely from the transformation. See details of these methods in Sections 5 and 6. Table 1 reports one-step ahead mean absolute prediction errors (MAPE) and mean squared prediction errors (MSPE) defined as (34) in Section 6, with a test size of 15. As expected, methods that employ the transformation, namely SegV and Uni.SegV, significantly outperform their counterparts VmV and UniV without any transformation. This highlights the benefit of integrating the transformation as an initial step in modeling multivariate functional time series.



(a) Functional cross-autocorrelations of the 8 original curve series.



(b) Functional cross-autocorrelations of the 8 transformed curve series.

Figure 1. Functional cross-autocorrelation of mortality (original and transformed) curve series versus size 0.95 upper confidence bound (blue dotted line).

Table 1. MAPEs and MSPEs for four competing methods on the female mortality curves.

Method	SegV	VmV	Uni.SegV	UniV
MAPE	1.205	1.890	1.454	1.765
MSPE	0.296	0.662	0.386	0.616

NOTE: All numbers are multiplied by 10. The lowest values are in bold font.

2.2. Estimation Procedure

We now consider how to find the segmentation transformation under (2). Assume that $\max_{i \in [p]} \int_{\mathcal{U}} \mathbb{E}\{Z_{ti}^2(u)\} du = O(1)$. Define $\Sigma_{y,k}(u, v) = \text{cov}\{Y_t(u), Y_{t+k}(v)\}$ and $\Sigma_{z,k}(u, v) = \text{cov}\{Z_t(u), Z_{t+k}(v)\}$. Without loss of generality, we focus on the orthogonal transformations only, that is, $\mathbf{A}^\top \mathbf{A} = \mathbf{A} \mathbf{A}^\top = \mathbf{I}_p$, as we can replace (Y_t, Z_t) in (2) by $(V_y^{-1/2} Y_t, V_z^{-1/2} Z_t)$ with $V_y = \int_{\mathcal{U}} \Sigma_{y,0}(u, u) du$ and $V_z = \int_{\mathcal{U}} \Sigma_{z,0}(u, u) du$. Then \mathbf{A} is replaced by $V_y^{-1/2} \mathbf{A} V_z^{1/2}$ which is an orthogonal matrix as

$$\mathbf{I}_p = \int_{\mathcal{U}} \text{var}\{V_y^{-1/2} Y_t(u)\} du = \int_{\mathcal{U}} \text{var}\{V_z^{-1/2} Z_t(u)\} du. \quad (3)$$

Due to the unobservability of Z_t , we can take $V_z^{-1/2} Z_t$ as Z_t since they share the same block structure. In practice, we can replace observations Y_t by $\widehat{V}_y^{-1/2} Y_t$, where \widehat{V}_y is a consistent estimator of V_y .

For a given integer $k_0 \geq 1$, let

$$\begin{aligned} \mathbf{W}_z &= \sum_{k=0}^{k_0} \int_{\mathcal{U}} \int_{\mathcal{U}} \Sigma_{z,k}(u, v)^{\otimes 2} dudv \text{ and} \\ \mathbf{W}_y &= \sum_{k=0}^{k_0} \int_{\mathcal{U}} \int_{\mathcal{U}} \Sigma_{y,k}(u, v)^{\otimes 2} dudv. \end{aligned} \quad (4)$$

Then both \mathbf{W}_y and \mathbf{W}_z are nonnegative definite. According to (2), it holds that $\Sigma_{y,k}(u, v) = \mathbf{A} \Sigma_{z,k}(u, v) \mathbf{A}^\top$, where $\Sigma_{z,k}(u, v)$ is block-diagonal with blocks on the main diagonal of sizes $p_1 \times p_1, \dots, p_q \times p_q$. Due to $\mathbf{A} \mathbf{A}^\top = \mathbf{I}_p$, by (4),

$$\mathbf{W}_z = \mathbf{A}^\top \mathbf{W}_y \mathbf{A}. \quad (5)$$

As all $\Sigma_{z,k}(u, v)$ for $k \geq 0$ and $(u, v) \in \mathcal{U}^2$ are block-diagonal matrices of the same sizes, so is \mathbf{W}_z . Perform the eigenanalysis for each of q blocks on the main diagonal of \mathbf{W}_z separately, leading to q orthogonal matrices of sizes $p_l \times p_l$ for $l \in [q]$. The columns of each of those orthogonal matrices are the p_l orthonormal eigenvectors from the corresponding eigenanalysis. We form a $p \times p$ block diagonal orthogonal matrix Γ_z with those q orthogonal matrices along the main block diagonal. Then the columns of Γ_z are the orthonormal eigenvectors of \mathbf{W}_z , that is,

$$\mathbf{W}_z \Gamma_z = \Gamma_z \mathbf{D}, \quad (6)$$

where \mathbf{D} is a diagonal matrix consisting of the p eigenvalues. Then by (5) and (6), $\mathbf{W}_y \mathbf{A} \Gamma_z = \mathbf{A} \mathbf{W}_z \Gamma_z = \mathbf{A} \Gamma_z \mathbf{D}$. Thus, the columns of $\Gamma_y \equiv \mathbf{A} \Gamma_z$ are the orthonormal eigenvectors of \mathbf{W}_y . Combining this with (2) yields that $\Gamma_y^\top Y_t(\cdot) = \Gamma_z^\top \mathbf{A}^\top Y_t(\cdot) = \Gamma_z^\top Z_t(\cdot)$. Since Γ_z is a block-diagonal orthogonal matrix with q blocks, $\Gamma_z^\top Z_t(\cdot)$ effectively applies orthogonal transformation within each of the q groups of $Z_t(\cdot)$. Thus, $\Gamma_z^\top Z_t(\cdot)$ is of the same

segmentation structure of $Z_t(\cdot)$, that is knowing $\Gamma_z^\top Z_t(\cdot)$ is as good as knowing the latent segmentation of $Z_t(\cdot)$. By (2), we have $Z_t(\cdot) = \mathbf{A}^\top Y_t(\cdot)$. Hence, Γ_y can be taken as the required transformation matrix \mathbf{A} .

Let $\widehat{\Sigma}_{y,k}(u, v)$ be some consistent estimator of $\Sigma_{y,k}(u, v)$ for $k \in \{0\} \cup [k_0]$, to be specified in Section 2.3. We define an estimator of \mathbf{W}_y as

$$\widehat{\mathbf{W}}_y = \sum_{k=0}^{k_0} \int_{\mathcal{U}} \int_{\mathcal{U}} \widehat{\Sigma}_{y,k}(u, v)^{\otimes 2} dudv, \quad (7)$$

and calculate its orthonormal eigenvectors $\widehat{\eta}_1, \dots, \widehat{\eta}_p$. Let $\widehat{\Gamma}_y = (\widehat{\eta}_1, \dots, \widehat{\eta}_p)$. Then the required transformation matrix \mathbf{A} can be estimated by a (latent) column-permutation of $\widehat{\Gamma}_y$. More specifically, put

$$\widehat{\mathbf{Z}}_t(\cdot) \equiv \{\widehat{Z}_{t1}(\cdot), \dots, \widehat{Z}_{tp}(\cdot)\}^\top = \widehat{\Gamma}_y^\top Y_t(\cdot). \quad (8)$$

We propose below a data-driven procedure to divide the p components of $\widehat{\mathbf{Z}}_t(\cdot)$ into \hat{q} uncorrelated groups.

Recall $Z_t(\cdot) = \{Z_{t1}(\cdot), \dots, Z_{tp}(\cdot)\}^\top$ with $\Sigma_{z,k}(\cdot, \cdot) = \{\Sigma_{z,k,ij}(\cdot, \cdot)\}_{i,j \in [p]}$. For two curve series $Z_{ti}(\cdot)$ and $Z_{tj}(\cdot)$ within the same group, one would expect that their lag- k cross-autocovariance function $\Sigma_{z,k,ij}(u, v)$ to be significantly different from zero for some integer k and $(u, v) \in \mathcal{U}^2$, thus, leading to at least one large $\|\Sigma_{z,k,ij}\|_S$ for some integer k . Based on $\widehat{\mathbf{Z}}_t(\cdot)$ defined as (8), we let $\widehat{\Sigma}_{z,k}(u, v) \equiv \{\widehat{\Sigma}_{z,k,ij}(u, v)\}_{i,j \in [p]} = \widehat{\Gamma}_y^\top \widehat{\Sigma}_{y,k}(u, v) \widehat{\Gamma}_y$ for any $(u, v) \in \mathcal{U}^2$. Given a fixed integer $m \geq 0$, we define the maximum cross-autocovariance over the lags between prespecified $-m$ and m as

$$\widehat{T}_{ij} = \max_{|k| \leq m} \|\widehat{\Sigma}_{z,k,ij}\|_S \quad (9)$$

for any pair $(i, j) \in [p]^2$ such that $i < j$, and regard $\widehat{Z}_{ti}(\cdot)$ and $\widehat{Z}_{tj}(\cdot)$ from the same group if \widehat{T}_{ij} takes some large value. To be specific, we rearrange $\aleph = p(p-1)/2$ values of \widehat{T}_{ij} ($1 \leq i < j \leq p$) in the descending order $\widehat{T}_{(1)} \geq \dots \geq \widehat{T}_{(\aleph)}$ and compute

$$\hat{q} = \arg \max_{j \in [\aleph]} \frac{\widehat{T}_{(j)} + \delta_n}{\widehat{T}_{(j+1)} + \delta_n} \quad (10)$$

for some $\delta_n > 0$. Corresponding to $\widehat{T}_{(1)}, \dots, \widehat{T}_{(\hat{q})}$, we identify \hat{q} pairs of cross-correlated curves. To divide the p components of $\widehat{\mathbf{Z}}_t(\cdot)$ into several uncorrelated groups, we can first start with p groups with each $\widehat{Z}_{ti}(\cdot)$ in one group and then repeatedly merge two groups if two cross-correlated curves are split over the two groups. The iteration is terminated until all the cross-correlated pairs are within one group. Hence, we obtain the estimated group structure of $\widehat{\mathbf{Z}}_t(\cdot)$ with the number of the final groups \hat{q} being the estimated value for q . Denote by $\widehat{\mathbf{Z}}_t^{(l)}(\cdot)$ the estimated l th group for $l \in [\hat{q}]$. The estimated transformation matrix $\widehat{\mathbf{A}} = (\widehat{\mathbf{A}}_1, \dots, \widehat{\mathbf{A}}_{\hat{q}})$ can then be found by reorganizing the order of $(\widehat{\eta}_1, \dots, \widehat{\eta}_p)$ such that

$$\widehat{\mathbf{Z}}_t^{(l)}(\cdot) = \widehat{\mathbf{A}}_l^\top Y_t(\cdot), \quad l \in [\hat{q}]. \quad (11)$$

Remark 1. (i) We include a small term $\delta_n > 0$ in (10) to stabilise the estimates for “0/0”. Given a suitable order of δ_n , we can establish the group recovery consistency. See Theorem 1 in

Section 4. A common practice is to set $\delta_n = 0$ and replace \aleph by $c_\theta \aleph$ in (10) for some constant $c_\theta \in (0, 1)$, see Lam and Yao (2012) and Ahn and Horenstein (2013).

(ii) All integrated terms in \mathbf{W}_y are nonnegative definite. Hence, there is no information cancellation over different lags. Therefore, the estimation is insensitive to the choice of k_0 . In practice a small k_0 (such as $k_0 \leq 5$) is often sufficient, while further enlarging k_0 tends to add more noise to \mathbf{W}_y .

2.3. Selection of $\widehat{\Sigma}_{y,k}(u, v)$

The estimate $\widehat{\Sigma}_{y,k}(u, v)$ plays a key role in Section 2.2. Let $\bar{\mathbf{Y}}(u) = n^{-1} \sum_{t=1}^n \mathbf{Y}_t(u)$. A natural candidate for $\widehat{\Sigma}_{y,k}(u, v)$ is the sample version of $\Sigma_{y,k}(u, v)$ defined as

$$\widehat{\Sigma}_{y,k}^s(u, v) = \frac{1}{n-k} \sum_{t=1}^{n-k} \{\mathbf{Y}_t(u) - \bar{\mathbf{Y}}(u)\} \{\mathbf{Y}_{t+k}(v) - \bar{\mathbf{Y}}(v)\}^\top, \quad k \in \{0\} \cup [k_0]. \quad (12)$$

When $p^2/n \rightarrow 0$, $\widehat{\Sigma}_{y,k}^s(u, v)$ is a valid estimator for $\Sigma_{y,k}(u, v)$. However, when p grows faster than $n^{1/2}$, it does not always hold that $\|\widehat{\Sigma}_{y,k}^s(u, v) - \Sigma_{y,k}(u, v)\|_2 \rightarrow 0$ in probability. Under the high-dimensional scenario, the orthogonality of \mathbf{A} naturally results in the magnitude of many of its entries being small, leading to certain sparsity on \mathbf{A} which will then pass onto the autocovariance functions $\Sigma_{y,k}(\cdot, \cdot)$, as $\Sigma_{y,k}(\cdot, \cdot) = \mathbf{A} \Sigma_{z,k}(\cdot, \cdot) \mathbf{A}^\top$.

Inspired by the spirit of threshold estimator for large covariance matrix (Bickel and Levina 2008), we apply the functional thresholding rule, which combines the functional generalizations of hard thresholding and shrinkage with the aid of the Hilbert–Schmidt norm of functions, on the entries of the sample autocovariance function $\widehat{\Sigma}_{y,k}^s(u, v) = \{\widehat{\Sigma}_{y,k,ij}^s(u, v)\}_{i,j \in [p]}$ in (12). This leads to the estimator

$$\mathcal{T}_{\omega_k}(\widehat{\Sigma}_{y,k}^s)(u, v) = [\widehat{\Sigma}_{y,k,ij}^s(u, v) I\{\|\widehat{\Sigma}_{y,k,ij}^s\|_S \geq \omega_k\}]_{i,j \in [p]}, \quad (u, v) \in \mathcal{U}^2, \quad (13)$$

where $\omega_k \geq 0$ is the thresholding parameter at lag k . Taking $\widehat{\Sigma}_{y,k}$ in (7) as $\mathcal{T}_{\omega_k}(\widehat{\Sigma}_{y,k}^s)$ yields

$$\widehat{\mathbf{W}}_y = \sum_{k=0}^{k_0} \int_{\mathcal{U}} \int_{\mathcal{U}} \mathcal{T}_{\omega_k}(\widehat{\Sigma}_{y,k}^s)(u, v) \otimes^2 du dv. \quad (14)$$

Remark 2. The thresholding parameter ω_k for each $k \in \{0\} \cup [k_0]$ can be selected using an L -fold cross-validation approach. Specifically, we sequentially divide the set $[n]$ into L validation sets V_1, \dots, V_L of approximately equal size. For each $l \in [L]$, let $\widehat{\Sigma}_{y,k}^{s,(l)}(u, v) = \{\widehat{\Sigma}_{y,k,ij}^{s,(l)}(u, v)\}_{i,j \in [p]}$ and $\widehat{\Sigma}_{y,k}^{s,(-l)}(u, v) = \{\widehat{\Sigma}_{y,k,ij}^{s,(-l)}(u, v)\}_{i,j \in [p]}$ be the sample lag- k autocovariance functions based on the l th validation set $\{\mathbf{Y}_t(\cdot) : t \in V_l\}$ and the remaining $L-1$ sets $\{\mathbf{Y}_t(\cdot) : t \in [n] \setminus V_l\}$, respectively. We select the optimal $\widehat{\omega}_k$ by minimizing

$$\text{Error}(\omega_k) = \frac{1}{L} \sum_{l=1}^L \sum_{i,j=1}^p \|\mathcal{T}_{\omega_k}(\widehat{\Sigma}_{y,k,ij}^{s,(l)}) - \widehat{\Sigma}_{y,k,ij}^{s,(-l)}\|_S^2,$$

where $\mathcal{T}_{\omega_k}(\widehat{\Sigma}_{y,k,ij}^{s,(l)})(u, v) = \widehat{\Sigma}_{y,k,ij}^{s,(l)}(u, v) I\{\|\widehat{\Sigma}_{y,k,ij}^{s,(l)}\|_S \geq \omega_k\}$.

3. Variational Multivariate FPCA

Our second step is to represent (linear) dynamic structure of each $\mathbf{Z}_t^{(l)}(\cdot)$ in terms of a vector time series via representation (1). The key idea is to identify the finite decomposition for $\mathbf{X}_t^{(l)}(\cdot)$. For $(u, v) \in \mathcal{U}^2$ and $k \geq 0$, let $\boldsymbol{\mu}^{(l)}(u) = \mathbb{E}\{\mathbf{X}_t^{(l)}(u)\}$ and

$$\mathbf{M}_k^{(l)}(u, v) = \mathbb{E}\{[\mathbf{X}_t^{(l)}(u) - \boldsymbol{\mu}^{(l)}(u)]\{[\mathbf{X}_{t+k}^{(l)}(v) - \boldsymbol{\mu}^{(l)}(v)]\}^\top\}.$$

Then the multivariate Karhunen–Loève decomposition for $\mathbf{X}_t^{(l)}(\cdot)$ serving as the foundation of multivariate FPCA (Chiou, Chen, and Yang 2014; Happ and Greven 2018) admits the form

$$\begin{aligned} \mathbf{M}_0^{(l)}(u, v) &= \sum_{j=1}^{\infty} \lambda_j^{(l)} \boldsymbol{\varphi}_j^{(l)}(u) \boldsymbol{\varphi}_j^{(l)}(v)^\top, \\ \mathbf{X}_t^{(l)}(u) - \boldsymbol{\mu}^{(l)}(u) &= \sum_{j=1}^{\infty} \xi_{tj}^{(l)} \boldsymbol{\varphi}_j^{(l)}(u), \end{aligned} \quad (15)$$

where $\lambda_1^{(l)} \geq \lambda_2^{(l)} \geq \dots \geq 0$ are the ordered eigenvalues of $\mathbf{M}_0^{(l)}(\cdot, \cdot)$, $\boldsymbol{\varphi}_1^{(l)}(\cdot), \boldsymbol{\varphi}_2^{(l)}(\cdot), \dots$ are the corresponding orthonormal eigenfunctions satisfying $\int_{\mathcal{U}} \boldsymbol{\varphi}_j^{(l)}(u)^\top \boldsymbol{\varphi}_k^{(l)}(u) du = I(j=k)$, and $\xi_{tj}^{(l)} = \int_{\mathcal{U}} \boldsymbol{\varphi}_j^{(l)}(u)^\top \{\mathbf{X}_t^{(l)}(u) - \boldsymbol{\mu}^{(l)}(u)\} du$ with $\mathbb{E}\{\xi_{tj}^{(l)}\} = 0$ and $\text{cov}\{\xi_{tj}^{(l)}, \xi_{tk}^{(l)}\} = \lambda_j^{(l)} I(j=k)$.

When $\mathbf{X}_t^{(l)}(\cdot)$ is r_l -dimensional in the sense that $\lambda_{r_l+1}^{(l)} = 0$, the dynamics of $\mathbf{X}_t^{(l)}(\cdot)$ is entirely determined by that of r_l -vector time series $\boldsymbol{\xi}_t^{(l)} = \{\xi_{t1}^{(l)}, \dots, \xi_{tr_l}^{(l)}\}^\top$. Unfortunately, under the latent decomposition (1), that is,

$$\begin{aligned} \mathbf{Z}_t^{(l)}(u) &= \mathbf{X}_t^{(l)}(u) + \boldsymbol{\varepsilon}_t^{(l)}(u) \\ &= \boldsymbol{\mu}^{(l)}(u) + \sum_{j=1}^{r_l} \xi_{tj}^{(l)} \boldsymbol{\varphi}_j^{(l)}(u) + \boldsymbol{\varepsilon}_t^{(l)}(u), \quad u \in \mathcal{U}, \end{aligned} \quad (16)$$

the standard multivariate FPCA based on (15) is inappropriate as $\mathbf{X}_t^{(l)}(\cdot)$ is unobservable and we cannot provide a consistent estimator for $\mathbf{M}_0^{(l)}(u, v)$ based on $\mathbf{Z}_t^{(l)}(\cdot)$ due to the fact $\text{cov}\{\mathbf{Z}_t^{(l)}(u), \mathbf{Z}_t^{(l)}(v)\} = \mathbf{M}_0^{(l)}(u, v) + \text{cov}\{\boldsymbol{\varepsilon}_t^{(l)}(u), \boldsymbol{\varepsilon}_t^{(l)}(v)\}$.

Now we introduce the variational multivariate FPCA based on a variational multivariate Karhunen–Loève decomposition for $\mathbf{X}_t^{(l)}(\cdot)$. Motivated from the fact $\text{cov}\{\mathbf{Z}_t^{(l)}(u), \mathbf{Z}_{t+k}^{(l)}(v)\} = \mathbf{M}_k^{(l)}(u, v)$ for any $k \geq 1$, for a prespecified small integer $k_0 \geq 1$, we define

$$\mathbf{K}^{(l)}(u, v) = \sum_{k=1}^{k_0} \int_{\mathcal{U}} \mathbf{M}_k^{(l)}(u, w) \mathbf{M}_k^{(l)}(v, w)^\top dw. \quad (17)$$

Similar to $\mathbf{M}_0^{(l)}$, $\mathbf{K}^{(l)}$ is also nonnegative definite and admits a spectral decomposition

$$\mathbf{K}^{(l)}(u, v) = \sum_{j=1}^{\infty} \theta_j^{(l)} \boldsymbol{\psi}_j^{(l)}(u) \boldsymbol{\psi}_j^{(l)}(v)^\top,$$

where $\theta_1^{(l)} \geq \theta_2^{(l)} \geq \dots \geq 0$ are the eigenvalues of $\mathbf{K}^{(l)}$, and $\boldsymbol{\psi}_1^{(l)}(\cdot), \boldsymbol{\psi}_2^{(l)}(\cdot), \dots$ are the corresponding orthonormal eigenfunctions.

Proposition 1. Let $\mathbf{\Omega}_k^{(l)} = \mathbb{E}[\xi_t^{(l)} \{\xi_{t+k}^{(l)}\}^\top]$ be a full-ranked matrix for some $k \in [k_0]$. Then it holds that (i) $\theta_{r_l}^{(l)} > 0$ and $\theta_{r_l+1}^{(l)} = 0$; (ii) $\text{span}\{\psi_1^{(l)}(\cdot), \dots, \psi_{r_l}^{(l)}(\cdot)\} = \text{span}\{\psi_1^{(l)}(\cdot), \dots, \psi_{r_l}^{(l)}(\cdot)\}$.

Proposition 1 shows that, under the expansion (16), $\mathbf{K}^{(l)}$ has exactly r_l nonzero eigenvalues, and the dynamic space spanned by $\{\psi_1^{(l)}(\cdot), \dots, \psi_{r_l}^{(l)}(\cdot)\}$ remains the same as that spanned by $\{\varphi_1^{(l)}(\cdot), \dots, \varphi_{r_l}^{(l)}(\cdot)\}$. Therefore, $\mathbf{X}_t^{(l)}(\cdot)$ can be expanded using r_l basis functions $\psi_1^{(l)}(\cdot), \dots, \psi_{r_l}^{(l)}(\cdot)$, that is,

$$\mathbf{X}_t^{(l)}(u) - \boldsymbol{\mu}^{(l)}(u) = \sum_{j=1}^{r_l} \zeta_{tj}^{(l)} \psi_j^{(l)}(u), \quad u \in \mathcal{U}, \quad (18)$$

where the basis coefficients $\zeta_{tj}^{(l)} = \int_{\mathcal{U}} \psi_j^{(l)}(u)^\top \{\mathbf{X}_t^{(l)}(u) - \boldsymbol{\mu}^{(l)}(u)\} du$. Note that we take the sum in defining $\mathbf{K}^{(l)}$ in (17) to accumulate the information from different lags, and there is no information cancellation as each term in the sum is nonnegative definite. An additional advantage for using the nonzero lagged autocovariance-based decomposition is that the identified directions $\psi_1^{(l)}(\cdot), \dots, \psi_{r_l}^{(l)}(\cdot)$ catch the most significant serial dependence, which leads to the most efficient dimension reduction and is thus advantageous for prediction.

Noting that $\mathbf{Z}_t^{(l)}(\cdot)$ is not directly observable, we can only estimate $\mathbf{M}_k^{(l)}$ and $\mathbf{K}^{(l)}$ based on \hat{p}_l -vector of estimated transformed curve subseries $\hat{\mathbf{Z}}_t^{(l)}(\cdot) = \{\hat{z}_{t1}^{(l)}(\cdot), \dots, \hat{z}_{t\hat{p}_l}^{(l)}(\cdot)\}^\top$ obtained in the segmentation transformation step. With the aid of (11), for $k \in \{0\} \cup [k_0]$, put

$$\hat{\mathbf{M}}_k^{(l)}(u, v) = \hat{\mathbf{A}}_l^\top \hat{\Sigma}_{y,k}(u, v) \hat{\mathbf{A}}_l. \quad (19)$$

It is easy to see from (1) that $\hat{\mathbf{M}}_k^{(l)}(u, v)$ is a reasonable estimator for $\mathbf{M}_k^{(l)}(u, v)$ when $k \geq 1$, as it filters out white noise $\boldsymbol{\epsilon}_t^{(l)}(\cdot)$ automatically. It is noteworthy that (19) requires the consistent estimators for $\Sigma_{y,k}(u, v)$. Its implementation under the high-dimensional setting can thus be done by setting $\hat{\Sigma}_{y,k}(u, v) = \mathcal{T}_{\omega_k}(\hat{\Sigma}_{y,k}^s)(u, v)$ defined in (13).

To estimate $\psi_j^{(l)}(\cdot)$ and $\zeta_{tj}^{(l)}$ in (18), we perform eigenanalysis of the estimator for $\mathbf{K}^{(l)}$,

$$\hat{\mathbf{K}}^{(l)}(u, v) = \sum_{k=1}^{k_0} \int_{\mathcal{U}} \hat{\mathbf{M}}_k^{(l)}(u, w) \hat{\mathbf{M}}_k^{(l)}(v, w)^\top dw, \quad (20)$$

leading to the eigenvalues $\hat{\theta}_1^{(l)} \geq \hat{\theta}_2^{(l)} \geq \dots \geq 0$, and the corresponding orthonormal eigenfunctions $\hat{\psi}_1^{(l)}(\cdot), \hat{\psi}_2^{(l)}(\cdot), \dots$. To estimate r_l (i.e., the number of nonzero eigenvalues), we take the commonly-adopted ratio-based estimator for r_l as

$$\hat{r}_l = \arg \max_{j \in [n-k_0]} \frac{\hat{\theta}_j^{(l)} + \tilde{\delta}_n}{\hat{\theta}_{j+1}^{(l)} + \tilde{\delta}_n} \quad (21)$$

for some $\tilde{\delta}_n > 0$. Under some regularity conditions, such defined \hat{r}_l is a consistent estimator for r_l ; see **Theorem 3** in **Section 4**. In practice, since $\tilde{\delta}_n$ is usually unknown, we instead adopt $\hat{r}_l = \arg \max_{j \in [c_r(n-k_0)]} \hat{\theta}_j^{(l)} / \hat{\theta}_{j+1}^{(l)}$, where $c_r \in (0, 1)$ is a prescribed

constant aiming to avoid fluctuations due to the ratios of extreme small values.

Let $\hat{\xi}_{tj}^{(l)} = \int_{\mathcal{U}} \hat{\psi}_j^{(l)}(u)^\top \{\hat{\mathbf{Z}}_t^{(l)}(u) - \bar{\mathbf{Z}}^{(l)}(u)\} du$ for $t \in [n]$, $j \in [\hat{r}_l]$ and $l \in [\hat{q}]$. We can fit a model for the \hat{r}_l -dimensional vector time series $\hat{\xi}_t^{(l)} = \{\hat{\xi}_{t1}^{(l)}, \dots, \hat{\xi}_{t\hat{r}_l}^{(l)}\}^\top$ with $t \in [n]$ to obtain its h -step ahead prediction $\hat{\xi}_{n+h}^{(l)}$ and then recover the h -step ahead functional prediction as

$$\hat{\mathbf{Z}}_{n+h}^{(l)}(u) = \bar{\mathbf{Z}}^{(l)}(u) + \sum_{j=1}^{\hat{r}_l} \hat{\xi}_{(n+h)j}^{(l)} \hat{\psi}_j^{(l)}(u), \quad h \geq 1.$$

We finally obtain the h -step ahead prediction $\hat{\mathbf{A}} \hat{\mathbf{Z}}_{n+h}^{(l)}(\cdot)$ for original functional time series, where $\hat{\mathbf{A}} = (\hat{\mathbf{A}}_1, \dots, \hat{\mathbf{A}}_{\hat{q}})$ and $\hat{\mathbf{Z}}_{n+h}^{(l)}(\cdot) = \{\hat{\mathbf{Z}}_{n+h}^{(1)}(\cdot)^\top, \dots, \hat{\mathbf{Z}}_{n+h}^{(\hat{q})}(\cdot)^\top\}^\top$.

4. Theoretical Properties

This section presents theoretical analysis of our two-step estimation procedure. To ease presentation, we focus on the high-dimensional scenario and develop the theoretical results based on the estimator $\mathcal{T}_{\omega_k}(\hat{\Sigma}_{y,k}^s)(u, v)$ in (13). To simplify notation, we use B to denote the linear operator induced from the kernel function $B \in \mathbb{S}$, that is, for any $f \in L_2(\mathcal{U})$, $B(f)(\cdot) = \int_{\mathcal{U}} B(\cdot, v) f(v) dv \in L_2(\mathcal{U})$. Denote the p -fold Cartesian product $\mathbb{H} = L_2(\mathcal{U}) \times \dots \times L_2(\mathcal{U})$. For any $\mathbf{f}, \mathbf{g} \in \mathbb{H}$, we denote the inner product by $\langle \mathbf{f}, \mathbf{g} \rangle = \int_{\mathcal{U}} \mathbf{f}(u)^\top \mathbf{g}(u) du$ with the induced norm $\|\cdot\| = \langle \cdot, \cdot \rangle^{1/2}$, and use \mathbf{B} to denote the linear operator induced from the kernel matrix function $\mathbf{B} = (B_{ij})_{m_1 \times m_2}$ with each $B_{ij} \in \mathbb{S}$, that is, for any $\mathbf{f} \in \mathbb{H}$, $\mathbf{B}(\mathbf{f})(\cdot) = \int_{\mathcal{U}} \mathbf{B}(\cdot, v) \mathbf{f}(v) dv \in \mathbb{H}$. We write $\|\mathbf{B}\|_{\mathcal{S}, \infty} = \max_{i \in [m_1]} \sum_{j=1}^{m_2} \|B_{ij}\|_{\mathcal{S}}$. Before imposing the regularity conditions, we first define the functional version of sub-Gaussianity that facilitates the development of non-asymptotic results for Hilbert space-valued random elements.

Definition 1. Let $\Upsilon_t(\cdot)$ be a mean zero random variable in $L_2(\mathcal{U})$ and $\Sigma_0 : L_2(\mathcal{U}) \rightarrow L_2(\mathcal{U})$ be a covariance operator. We call $\Upsilon_t(\cdot)$ a sub-Gaussian process if there exists a constant $c > 0$ such that $\mathbb{E}[\exp\{\langle \mathbf{f}, \Upsilon_t - \mathbb{E}(\Upsilon_t) \rangle\}] \leq \exp\{2^{-1} c^2 \langle \mathbf{f}, \Sigma_0(\mathbf{f}) \rangle\}$ for all $\mathbf{f} \in L_2(\mathcal{U})$.

Condition 1. (i) $\{\mathbf{Y}_t(\cdot)\}$ is a sequence of multivariate functional linear processes with sub-Gaussian errors, that is, $\mathbf{Y}_t(\cdot) = \sum_{l=0}^{\infty} \Psi_l(\boldsymbol{\epsilon}_{t-l})$, where $\Psi_l = (\Psi_{l,ij})_{p \times p}$ with each $\Psi_{l,ij} \in \mathbb{S}$ and $\boldsymbol{\epsilon}_t(\cdot) = \{\epsilon_{t1}(\cdot), \dots, \epsilon_{tp}(\cdot)\}^\top$ with independent components of mean-zero sub-Gaussian processes satisfying **Definition 1**; (ii) The coefficient functions satisfy $\sum_{l=0}^{\infty} \|\Psi_l\|_{\mathcal{S}, \infty} = O(1)$; (iii) $\max_{j \in [p]} \int_{\mathcal{U}} \text{cov}\{\epsilon_{tj}(u), \epsilon_{tj}(u)\} du = O(1)$.

Condition 2. For $\{\mathbf{Y}_t(\cdot)\}$, its spectral density operator $\Lambda_{y,\theta} = (2\pi)^{-1} \sum_{k \in \mathbb{Z}} \Sigma_{y,k} \exp(-k\theta\sqrt{-1})$ for $\theta \in [-\pi, \pi]$ exists and the functional stability measure

$$\mathcal{M}_y = 2\pi \text{ess sup}_{\theta \in [-\pi, \pi], \Phi \in \mathbb{H}_0} \frac{\langle \Phi, \Lambda_{y,\theta}(\Phi) \rangle}{\langle \Phi, \Sigma_{y,0}(\Phi) \rangle} < \infty, \quad (22)$$

where $\mathbb{H}_0 = \{\Phi \in \mathbb{H} : \langle \Phi, \Sigma_{y,0}(\Phi) \rangle \in (0, \infty)\}$.

Write $\Sigma_{y,k}(u, v) = \{\Sigma_{y,k,ij}(u, v)\}_{i,j \in [p]}$. **Conditions 1(ii)** and **1(iii)** guarantee the covariance-stationarity of $\{\mathbf{Y}_t(\cdot)\}$ and imply

that $\max_{j \in [p]} \int_{\mathcal{U}} \Sigma_{y,0,jj}(u, u) du = O(1)$. **Condition 2** places a finite upper bound on the functional stability measure, which characterizes the effect of small decaying eigenvalues of $\Sigma_{y,0}$ on the numerator of (22), thus, being able to handle infinite-dimensional functional objects $Y_{ij}(\cdot)$. See its detailed discussion in Guo and Qiao (2023). **Conditions 1** and **2** are essential to derive $\max_{i,j \in [p]} \|\widehat{\Sigma}_{y,k,ij}^s - \Sigma_{y,k,ij}\|_S = O_p\{\mathcal{M}_y(n^{-1} \log p)^{1/2}\}$ for $\widehat{\Sigma}_{y,k,ij}^s$ involved in (13), which plays a crucial rule in our theoretical analysis.

Condition 3. For $\mathbf{A} = (A_{ij})_{p \times p}$, $\max_{i \in [p]} \sum_{j=1}^p |A_{ij}|^\alpha \leq s_1$ and $\max_{j \in [p]} \sum_{i=1}^p |A_{ij}|^\alpha \leq s_2$ for some constant $\alpha \in [0, 1)$.

The parameters s_1 and s_2 determine the row and column sparsity levels of \mathbf{A} , respectively. The row sparsity with small s_1 entails that each component of $\mathbf{Y}_t(\cdot)$ is a linear combination of a small number of components in $\mathbf{Z}_t(\cdot)$, while the column sparsity with small s_2 corresponds to the case that each $Z_{tj}(\cdot)$ has impact on only a few components of $\mathbf{Y}_t(\cdot)$. The parameter α also controls the sparsity level of \mathbf{A} with a smaller value yielding a sparser \mathbf{A} . Write

$$p_{\dagger} = \max_{l \in [q]} p_l. \quad (23)$$

Lemma A2 in the supplementary material reveals that the functional sparsity structures in columns/rows of $\Sigma_{y,k}(\cdot, \cdot)$ are determined by $s_1 s_2 p_{\dagger}$ with smaller values of s_1 , s_2 and p_{\dagger} yielding functional sparser $\Sigma_{y,k}(\cdot, \cdot)$.

Recall that $\mathbf{W}_z = \text{diag}(\mathbf{W}_{z,1}, \dots, \mathbf{W}_{z,q})$ in (5) is a block-diagonal matrix, where $\mathbf{W}_{z,l}$ is a $p_l \times p_l$ matrix. We further define

$$\rho = \min_{j \neq l} \min_{\lambda \in \Lambda(\mathbf{W}_{z,l}), \tilde{\lambda} \in \Lambda(\mathbf{W}_{z,j})} |\lambda - \tilde{\lambda}|, \quad (24)$$

where $\Lambda(\cdot)$ denotes the set of eigenvalues of the matrix, and assume $\rho > 0$.

We first establish the group recovery consistency of the segmentation step. To do this, we reformulate the permutation in Section 2.2 in an equivalent graph representation way. Recall $\mathbf{\Gamma}_y = \mathbf{A}\mathbf{\Gamma}_z$ and $\mathbf{\Gamma}_z$ is a block-diagonal orthogonal matrix with the main block sizes p_1, \dots, p_q . Write $\mathbf{\Gamma}_z = \text{diag}(\mathbf{\Gamma}_{z,1}, \dots, \mathbf{\Gamma}_{z,q})$. Since $\mathbf{A} = (\mathbf{A}_1, \dots, \mathbf{A}_q)$, we have $\mathbf{\Gamma}_y \equiv (\boldsymbol{\eta}_1, \dots, \boldsymbol{\eta}_p) = (\mathbf{A}_1 \mathbf{\Gamma}_{z,1}, \dots, \mathbf{A}_q \mathbf{\Gamma}_{z,q})$. The columns of $\mathbf{\Gamma}_y$ are naturally partitioned in to q groups G_1, \dots, G_q , where $G_l = \{\boldsymbol{\eta}_{\sum_{l'=0}^{l-1} p_{l'}+1}, \dots, \boldsymbol{\eta}_{\sum_{l'=0}^l p_{l'}}\}$ with $p_0 = 0$. To simplify the notation, we just write

$$G_l = \left\{ \sum_{l'=0}^{l-1} p_{l'} + 1, \dots, \sum_{l'=0}^l p_{l'} \right\}, \quad l \in [q]. \quad (25)$$

Recall that the columns of such defined $\mathbf{\Gamma}_y$ are the eigenvectors of \mathbf{W}_y . For ρ defined in (24), if $\|\widehat{\mathbf{W}}_y - \mathbf{W}_y\|_2 \leq \rho/5$, by Lemma A4 in the supplementary material, there exists an orthogonal matrix $\mathbf{H} = \text{diag}(\mathbf{H}_1, \dots, \mathbf{H}_q)$ with $\mathbf{H}_l \in \mathbb{R}^{p_l \times p_l}$ for each $l \in [q]$ and a column permutation matrix \mathbf{R} for $\widehat{\mathbf{\Gamma}}_y$, such that $\widehat{\mathbf{\Gamma}}_y \mathbf{R} \equiv (\widehat{\boldsymbol{\Pi}}_1, \dots, \widehat{\boldsymbol{\Pi}}_q)$ with $\widehat{\boldsymbol{\Pi}}_l \in \mathbb{R}^{p \times p_l}$, and

$$\|\widehat{\boldsymbol{\Pi}}_l - \mathbf{A}_l \mathbf{\Gamma}_{z,l} \mathbf{H}_l\|_2 \leq 8\rho^{-1} \|\widehat{\mathbf{W}}_y - \mathbf{W}_y\|_2. \quad (26)$$

If the p eigenvalues of \mathbf{W}_y are distinct, \mathbf{H} is a diagonal matrix with elements in the diagonal being 1 or -1 . Write $\mathbf{\Gamma}_y \mathbf{H} = (\mathbf{A}_1 \mathbf{\Gamma}_{z,1} \mathbf{H}_1, \dots, \mathbf{A}_q \mathbf{\Gamma}_{z,q} \mathbf{H}_q) \equiv (\boldsymbol{\gamma}_1, \dots, \boldsymbol{\gamma}_p)$. For each $l \in [q]$, we can define a graph (G_l, E_l) such that $(i, j) \in E_l$ if and only if $\max_{|k| \leq m} \|\boldsymbol{\gamma}_i^\top \Sigma_{y,k} \boldsymbol{\gamma}_j\|_S \neq 0$.

Condition 4. There exists some $\varsigma > 0$ such that $\inf_{(i,j) \in E_l} \max_{|k| \leq m} \|\boldsymbol{\gamma}_i^\top \Sigma_{y,k} \boldsymbol{\gamma}_j\|_S \geq \varsigma$ for each $l \in [q]$, where m is specified in (9).

Condition 4 ensures that the group G_l is inseparable at the minimal signal level ς given the transformation $\mathbf{A}_l \mathbf{\Gamma}_{z,l} \mathbf{H}_l$ for each $l \in [q]$, and facilitates the specifications of the true number of groups q and the associated segmentation structure under (2). Define $T_{ij} = \max_{|k| \leq m} \|\boldsymbol{\gamma}_i^\top \Sigma_{y,k} \boldsymbol{\gamma}_j\|_S$ and $\varrho = \sum_{l=1}^q |E_l|$. Rearrange $\aleph = p(p-1)/2$ values of T_{ij} ($1 \leq i < j \leq p$) in the descending order, $T_{(1)} \geq \dots \geq T_{(\aleph)}$. We then have $T_{(i)} \geq \varsigma$ for $i \in [\varrho]$ and $T_{(i)} = 0$ for $i \geq \varrho + 1$. Denote by $E = \{(i, j) : T_{ij} \geq T_{(\varrho)}, 1 \leq i < j \leq p\}$ the edge set of $G = [p]$ under the transformation $\mathbf{\Gamma}_y \mathbf{H}$. The true segmentation $\{G_1, \dots, G_q\}$ in (25) can then be identified by splitting (G, E) into q isolated subgraphs $(G_1, E_1), \dots, (G_q, E_q)$, where q represents the true number of uncorrelated groups.

Recall that with the aid of $\widehat{\mathbf{\Gamma}}_y$, the estimated segmentation is obtained via the ratio-based estimator \hat{q} as defined in (10). To be specific, we build an estimated graph (G, \tilde{E}) with vertex set $G = [p]$ and edge set $\tilde{E} = \{(i, j) : \tilde{T}_{ij} \geq \tilde{T}_{(\hat{q})}, 1 \leq i < j \leq p\}$, and split it into \hat{q} isolated subgraphs $(\tilde{G}_1, \tilde{E}_1), \dots, (\tilde{G}_{\hat{q}}, \tilde{E}_{\hat{q}})$. Note that p columns of $\widehat{\mathbf{\Gamma}}_y = (\widehat{\boldsymbol{\eta}}_1, \dots, \widehat{\boldsymbol{\eta}}_p)$ correspond to the ordered eigenvalues $\lambda_1(\widehat{\mathbf{W}}_y) \geq \dots \geq \lambda_p(\widehat{\mathbf{W}}_y)$. Write $\widehat{\mathbf{\Gamma}}_y \mathbf{R} \equiv (\widehat{\boldsymbol{\gamma}}_1, \dots, \widehat{\boldsymbol{\gamma}}_p)$ and let $\pi : [p] \rightarrow [p]$ denote the permutation associated with \mathbf{R} , that is, $\widehat{\boldsymbol{\gamma}}_i = \widehat{\boldsymbol{\eta}}_{\pi(i)}$. Based on the permutation mapping π , we let $\widehat{G}_l = \{\pi^{-1}(i) : i \in \tilde{G}_l\}$ for $l \in [\hat{q}]$.

Theorem 1. Let **Conditions 1–4** hold. For each $|k| \leq k_0 \vee m$, select $\omega_k = c_k \mathcal{M}_y(n^{-1} \log p)^{1/2}$ in (13) for some sufficiently large constant $c_k > 0$. Assume $(\rho^{-1} s_1^2 s_2^2 p_{\dagger}^{3-\alpha})^{2/(1-\alpha)} \mathcal{M}_y^2 \log p = o(n)$ and δ_n in (10) satisfies $\rho^{-1} s_1^3 s_2^3 p_{\dagger}^{5-2\alpha} \mathcal{M}_y^{1-\alpha} (n^{-1} \log p)^{(1-\alpha)/2} \ll \delta_n \ll \varsigma^2 T_{(1)}^{-1}$, where p_{\dagger} and ρ are specified in (23) and (24), respectively. As $n \rightarrow \infty$, it holds that (i) $\mathbb{P}(\hat{q} = q) \rightarrow 1$ and (ii) there exists a permutation $\tilde{\pi} : [q] \rightarrow [q]$ such that $\mathbb{P}[\bigcap_{l=1}^q \{\widehat{G}_{\tilde{\pi}(l)} = G_l\} | \hat{q} = q] \rightarrow 1$.

Theorem 1 gives the group recovery consistency of our segmentation step. We next evaluate the errors in estimating $\mathcal{C}(\mathbf{A}_l)$ for $l \in [q]$. Based on the estimated groups $\{\widehat{G}_1, \dots, \widehat{G}_{\hat{q}}\}$, we reorganize the order of $(\widehat{\boldsymbol{\gamma}}_1, \dots, \widehat{\boldsymbol{\gamma}}_p) = (\widehat{\boldsymbol{\eta}}_{\pi(1)}, \dots, \widehat{\boldsymbol{\eta}}_{\pi(p)})$ and define $\widehat{\mathbf{A}}_l$ in (11) as $\widehat{\mathbf{A}}_l = (\widehat{\boldsymbol{\gamma}}_i)_{i \in \widehat{G}_l}$ for $l \in [\hat{q}]$. We consider a general discrepancy measure (Chang, Guo, and Yao 2015, 2018) between two linear spaces $\mathcal{C}(\mathbf{E}_1)$ and $\mathcal{C}(\mathbf{E}_2)$ spanned by the columns of $\mathbf{E}_1 \in \mathbb{R}^{p \times p_1}$ and $\mathbf{E}_2 \in \mathbb{R}^{p \times p_2}$, respectively, with $\mathbf{E}_i^\top \mathbf{E}_i = \mathbf{I}_{p_i}$ for $i \in [2]$ as

$$D\{\mathcal{C}(\mathbf{E}_1), \mathcal{C}(\mathbf{E}_2)\} = \sqrt{1 - \frac{\text{tr}(\mathbf{E}_1 \mathbf{E}_1^\top \mathbf{E}_2 \mathbf{E}_2^\top)}{\max(\tilde{p}_1, \tilde{p}_2)}} \in [0, 1]. \quad (27)$$

Then $D\{\mathcal{C}(\mathbf{E}_1), \mathcal{C}(\mathbf{E}_2)\}$ is equal to 0 if and only if $\mathcal{C}(\mathbf{E}_1) \subset \mathcal{C}(\mathbf{E}_2)$ or $\mathcal{C}(\mathbf{E}_2) \subset \mathcal{C}(\mathbf{E}_1)$, and to 1 if and only if the two spaces are orthogonal.

Theorem 2. Let conditions for [Theorem 1](#) hold. As $n \rightarrow \infty$, it holds that

$$\begin{aligned} & \max_{l \in [q]} \min_{j \in [\hat{q}]} D(\mathcal{C}(\mathbf{A}_l), \mathcal{C}(\hat{\mathbf{A}}_j)) \\ & = O_p \left\{ \rho^{-1} s_1^2 s_2^2 p_+^{3-\alpha} \mathcal{M}_y^{1-\alpha} (n^{-1} \log p)^{(1-\alpha)/2} \right\}. \end{aligned}$$

[Theorem 2](#) presents the uniform convergence rate for $\min_{j \in [\hat{q}]} D(\mathcal{C}(\mathbf{A}_l), \mathcal{C}(\hat{\mathbf{A}}_j))$ over $l \in [q]$. For given α , the rate is faster for smaller values of $\{s_1, s_2, p_+, \mathcal{M}_y\}$, while enlarging the minimum eigen-gap between different blocks (i.e., larger ρ) reduces the difficulty of estimating each $\mathcal{C}(\mathbf{A}_l)$.

Supported by [Theorems 1](#) and [2](#), our subsequent theoretical results are developed by assuming that the group structure of $\mathbf{Z}_t(\cdot)$ is correctly identified or known, that is, $\hat{q} = q$ and $\hat{G}_l = G_l$ for each l . We now turn to investigate the theoretical properties of the dimension reduction step. Inherited from the segmentation step, $\mathbf{Z}_t^{(1)}(\cdot), \dots, \mathbf{Z}_t^{(q)}(\cdot)$ rely on the specific form of $\mathbf{A} = (\mathbf{A}_1, \dots, \mathbf{A}_q)$, and thus are not uniquely defined. Yet intuitively, we only require a certain transformation matrix to make our subsequent analysis related to $\hat{\eta}_1, \dots, \hat{\eta}_p$ mathematically tractable. Based on (26), we define $\Pi_l = \mathbf{A}_l \Gamma_{z,l} \mathbf{H}_l$ and it holds that $\mathcal{C}(\Pi_l) = \mathcal{C}(\mathbf{A}_l)$ for each $l \in [q]$. Let $\mathbf{Z}_t^{(l)}(\cdot) = \Pi_l^\top \mathbf{Y}_t(\cdot)$. Recall (1) and (18). The primary goal of the second dimension reduction step is to identify each r_l and to estimate the associated dynamic space $\mathcal{C}_l = \text{span}\{\psi_1^{(l)}(\cdot), \dots, \psi_{r_l}^{(l)}(\cdot)\}$. Recall that $\{\hat{\theta}_j^{(l)}, \hat{\psi}_j^{(l)}(\cdot)\}_{j \geq 1}$ are the eigenvalue/eigenfunction pairs of $\hat{\mathbf{K}}^{(l)}(\cdot, \cdot)$ defined in (20) with $\hat{\mathbf{A}}_l = (\hat{\gamma}_i)_{i \in G_l}$ and the dimension r_l is fixed for all $l \in [q]$. Our asymptotic results are based on the following regularity condition:

Condition 5. For each $l \in [q]$, all r_l nonzero eigenvalues of $\mathbf{K}^{(l)}(\cdot, \cdot)$ are different, that is, $\theta_1^{(l)} > \dots > \theta_{r_l}^{(l)} > 0 = \theta_{r_l+1}^{(l)} = \dots$.

Theorem 3. Let [Conditions 1–3](#) and [5](#) hold. Assume $(\rho^{-1} s_1^3 s_2^3 p_+^{5-2\alpha})^{2/(1-\alpha)} \mathcal{M}_y^2 \log p = o(n)$ and $\tilde{\delta}_n$ in (21) satisfies $\rho^{-1} s_1^3 s_2^3 p_+^{7-2\alpha} \mathcal{M}_y^{1-\alpha} (n^{-1} \log p)^{(1-\alpha)/2} \ll \tilde{\delta}_n \ll \min_{l \in [q]} \{\theta_{r_l}^{(l)}\}^2 / \max_{l \in [q]} \theta_1^{(l)}$, where p_+ and ρ are specified in (23) and (24), respectively. As $n \rightarrow \infty$, it holds that $\mathbb{P}[\bigcap_{l=1}^q \{\hat{r}_l = r_l\}] \rightarrow 1$.

[Theorem 3](#) shows that r_l can be correctly identified with probability tending to one uniformly over $l \in [q]$. Let $\hat{\mathcal{C}}_l = \text{span}\{\hat{\psi}_1^{(l)}(\cdot), \dots, \hat{\psi}_{\hat{r}_l}^{(l)}(\cdot)\}$ be the dynamic space spanned by \hat{r}_l estimated eigenfunctions. To measure the discrepancy between \mathcal{C}_l and $\hat{\mathcal{C}}_l$, we introduce the following metric. For two subspaces $\mathcal{C}(\mathbf{b}_1) = \text{span}\{\mathbf{b}_{11}(\cdot), \dots, \mathbf{b}_{1\hat{r}_1}(\cdot)\}$ and $\mathcal{C}(\mathbf{b}_2) = \text{span}\{\mathbf{b}_{21}(\cdot), \dots, \mathbf{b}_{2\hat{r}_2}(\cdot)\}$ satisfying $\langle \mathbf{b}_{ij}, \mathbf{b}_{ik} \rangle = I(j=k)$ for each $i \in [2]$, the discrepancy measure between $\mathcal{C}(\mathbf{b}_1)$ and $\mathcal{C}(\mathbf{b}_2)$ is defined as

$$\begin{aligned} & \tilde{D}(\mathcal{C}(\mathbf{b}_1), \mathcal{C}(\mathbf{b}_2)) \\ & = \sqrt{1 - \frac{1}{\max(\tilde{r}_1, \tilde{r}_2)} \sum_{j=1}^{\tilde{r}_1} \sum_{k=1}^{\tilde{r}_2} \langle \mathbf{b}_{1j}, \mathbf{b}_{2k} \rangle^2} \in [0, 1], \end{aligned}$$

which equals 0 if and only if $\mathcal{C}(\mathbf{b}_1) \subset \mathcal{C}(\mathbf{b}_2)$ or $\mathcal{C}(\mathbf{b}_2) \subset \mathcal{C}(\mathbf{b}_1)$ and 1 if and only if two spaces are orthogonal.

Theorem 4. Let conditions for [Theorem 3](#) hold. Assume $(\Delta^{-1} \rho^{-1} s_1^3 s_2^3 p_+^{7-2\alpha})^{2/(1-\alpha)} \mathcal{M}_y^2 \log p = o(n)$ with $\Delta = \min_{l \in [q], j \in [r_l]} \{\theta_j^{(l)} - \theta_{j+1}^{(l)}\}$. As $n \rightarrow \infty$, it holds that

$$\begin{aligned} & \max_{l \in [q]} \tilde{D}(\hat{\mathcal{C}}_l, \mathcal{C}_l) \\ & = O_p \left\{ \Delta^{-1} \rho^{-1} s_1^3 s_2^3 p_+^{7-2\alpha} \mathcal{M}_y^{1-\alpha} (n^{-1} \log p)^{(1-\alpha)/2} \right\}. \end{aligned}$$

5. Simulation Studies

We conduct a series of simulations to illustrate the finite sample performance of the proposed methods. To simplify the data-generating process, we consider a relaxed form of (2) as

$$\begin{aligned} \tilde{\mathbf{Y}}_t(u) &= \tilde{\mathbf{A}} \tilde{\mathbf{Z}}_t(u) \\ &= \tilde{\mathbf{A}} \{\tilde{\mathbf{Z}}_t^{(1)}(u)^\top, \dots, \tilde{\mathbf{Z}}_t^{(q)}(u)^\top\}^\top, \quad u \in \mathcal{U} = [0, 1], \end{aligned} \quad (28)$$

with no orthonormality restriction on the transformation matrix $\tilde{\mathbf{A}} = (\tilde{\mathbf{A}}_1, \dots, \tilde{\mathbf{A}}_q)$. The p -dimensional functional time series $\tilde{\mathbf{Z}}_t(\cdot)$ is formed by q uncorrelated groups $\{\tilde{\mathbf{Z}}_t^{(l)}(\cdot) : l \in [q]\}$, where each $\tilde{\mathbf{Z}}_t^{(l)}(\cdot)$ arises as the sum of dynamics $\tilde{\mathbf{X}}_t^{(l)}(\cdot)$ and white noise $\tilde{\boldsymbol{\epsilon}}_t^{(l)}(\cdot)$. Based on (3) in [Section 2.2](#), (28) can then be easily reformulated as (2) by setting

$$\begin{aligned} \mathbf{Y}_t(\cdot) &= \mathbf{V}_y^{-1/2} \tilde{\mathbf{Y}}_t(\cdot), \quad \mathbf{A} = \mathbf{V}_y^{-1/2} \tilde{\mathbf{A}} \mathbf{V}_z^{1/2} \quad \text{and} \\ \mathbf{Z}_t(\cdot) &= \mathbf{V}_z^{-1/2} \tilde{\mathbf{Z}}_t(\cdot), \end{aligned} \quad (29)$$

where $\mathbf{V}_y = \int_{\mathcal{U}} \text{cov}\{\tilde{\mathbf{Y}}_t(u), \tilde{\mathbf{Y}}_t(u)\} du$ and $\mathbf{V}_z = \int_{\mathcal{U}} \text{cov}\{\tilde{\mathbf{Z}}_t(u), \tilde{\mathbf{Z}}_t(u)\} du$. Then the orthonormality of \mathbf{A} is satisfied.

Write $\tilde{\boldsymbol{\epsilon}}_t(\cdot) = \{\tilde{\boldsymbol{\epsilon}}_t^{(1)}(\cdot)^\top, \dots, \tilde{\boldsymbol{\epsilon}}_t^{(q)}(\cdot)^\top\}^\top \equiv \{\tilde{\epsilon}_{t1}(\cdot), \dots, \tilde{\epsilon}_{tp}(\cdot)\}^\top$. We generate each curve component of $\tilde{\boldsymbol{\epsilon}}_t(\cdot)$ independently by $\tilde{\epsilon}_{ij}(\cdot) = \sum_{l=1}^{10} 2^{-(l-1)} e_{ijl} \psi_l(\cdot)$ for $j \in [p]$, where e_{ijl} 's are sampled independently from $\mathcal{N}(0, 1)$ and $\{\psi_l(\cdot)\}_{l=1}^{10}$ is a 10-dimensional Fourier basis function. The finite-dimensional dynamics $\tilde{\mathbf{X}}_t(\cdot) = \{\tilde{\mathbf{X}}_t^{(1)}(\cdot)^\top, \dots, \tilde{\mathbf{X}}_t^{(q)}(\cdot)^\top\}^\top$ with prescribed group structure is generated based on some 5-dimensional curve dynamics $\vartheta_{tg}(\cdot) = \sum_{l=1}^5 \kappa_{tgl} \psi_l(\cdot)$ for $g \in [20]$. The basis coefficients $\kappa_{tg} = (\kappa_{tg1}, \dots, \kappa_{tg5})^\top$ are generated from a stationary VAR model $\kappa_{tg} = \mathbf{U}_g \kappa_{(t-1)g} + \mathbf{e}_{tg}$ for each g . To guarantee the stationarity of κ_{tg} , we generate $\mathbf{U}_g = \iota \tilde{\mathbf{U}}_g / \rho(\tilde{\mathbf{U}}_g)$ with $\iota \sim \text{Uniform}[0.5, 1]$ and $\rho(\tilde{\mathbf{U}}_g)$ being the spectral radius of $\tilde{\mathbf{U}}_g \in \mathbb{R}^{5 \times 5}$, the entries of which are sampled independently from $\text{Uniform}[-3, 3]$. The components of the innovation \mathbf{e}_{tg} are sampled independently from $\mathcal{N}(0, 1)$.

We will specify the exact forms of $\tilde{\mathbf{X}}_t(\cdot)$ under the fixed and large p scenarios in [Sections 5.1](#) and [5.2](#), respectively. The white noise sequence $\tilde{\boldsymbol{\epsilon}}_t(\cdot)$ ensures that $\tilde{\mathbf{Z}}_t(\cdot)$ as well as $\mathbf{Z}_t(\cdot)$ share the same group structure as $\tilde{\mathbf{X}}_t(\cdot)$. Unless otherwise stated, we set $k_0 = m = 5$ and $c_r = c_\rho = 0.75$ in our procedure, as our simulation results suggest that our procedure is robust to the choices of these parameters.

5.1. Cases with Fixed p

We consider the following three examples of $\tilde{\mathbf{X}}_t(\cdot) = \{\tilde{X}_{t1}(\cdot), \dots, \tilde{X}_{tp}(\cdot)\}^\top$ with different group structures for $p \in \{6, 10, 15\}$ based on independent $\vartheta_{t1}(\cdot), \dots, \vartheta_{t5}(\cdot)$.

Example 1. $\tilde{X}_{t1}(\cdot) = \vartheta_{t1}(\cdot)$, $\tilde{X}_{tj}(\cdot) = \vartheta_{(t+j-2)2}(\cdot)$ for $j \in \{2, 3\}$ and $\tilde{X}_{tj}(\cdot) = \vartheta_{(t+j-4)3}(\cdot)$ for $j \in \{4, 5, 6\}$.

Example 2. $\tilde{X}_{tj}(\cdot)$ for $j \in [6]$ are the same as those in **Example 1** and $\tilde{X}_{tj}(\cdot) = \vartheta_{(t+j-7)4}(\cdot)$ for $j \in \{7, \dots, 10\}$.

Example 3. $\tilde{X}_{tj}(\cdot)$ for $j \in [10]$ are the same as those in **Example 2** and $\tilde{X}_{tj}(\cdot) = \vartheta_{(t+j-11)5}(\cdot)$ for $j \in \{11, \dots, 15\}$.

Therefore, $\tilde{\mathbf{X}}_t(\cdot)$ consists of $q = 3, 4$, and 5 uncorrelated groups of curve subseries in **Examples 1–3**, respectively, where the number of component curves per group is $p_l = l$ for $l \in [q]$. The p -dimensional observed functional time series $\tilde{\mathbf{Y}}_t(\cdot) = \{\tilde{Y}_{t1}(\cdot), \dots, \tilde{Y}_{tp}(\cdot)\}^\top$ for $t \in [n]$ is then generated by (28) with the entries of $\tilde{\mathbf{A}}$ sampled independently from Uniform $[-3, 3]$. To obtain h -step ahead prediction of $\tilde{\mathbf{Y}}_t(\cdot)$, we integrate the segmentation and dimension reduction steps respectively in **Sections 2** and **3** into the VAR estimation as outlined in **Algorithm 1**. For each of the three examples introduced above, we select

$$\hat{\mathbf{V}}_{\tilde{\mathbf{y}}}^{(h)} = \frac{1}{n-h} \sum_{t=1}^{n-h} \int_{\mathcal{U}} \left\{ \tilde{\mathbf{Y}}_t(u) - \frac{1}{n-h} \sum_{t=1}^{n-h} \tilde{\mathbf{Y}}_t(u) \right\}^{\otimes 2} du, \quad (30)$$

$$\hat{\Sigma}_{\tilde{\mathbf{y}},k}^{(h)}(u, v) = \frac{1}{n-h-k} \sum_{t=1}^{n-h-k} \{ \tilde{\mathbf{Y}}_t(u) - \tilde{\mathbf{Y}}_*(u) \} \{ \tilde{\mathbf{Y}}_{t+k}(v) - \tilde{\mathbf{Y}}_*(v) \}^\top, \quad (31)$$

with $\tilde{\mathbf{Y}}_*(\cdot) = (n-h-k)^{-1} \sum_{t=1}^{n-h-k} \tilde{\mathbf{Y}}_t(\cdot)$, for the quantities involved in Step (i) of **Algorithm 1**. We refer to the **segmentation-(Variational-multivariate-FPCA)-and-VAR-based Algorithm 1** with selections of $\hat{\mathbf{V}}_{\tilde{\mathbf{y}}}^{(h)}$ in (30) and $\hat{\Sigma}_{\tilde{\mathbf{y}},k}^{(h)}(u, v)$ in (31) as SegV hereafter.

The performance of our two-step proposal is examined in terms of linear space estimation, group identification and post-sample prediction. For $\mathbf{A} = (\mathbf{A}_1, \dots, \mathbf{A}_q)$ specified in (29), with the aid of (27), define $f(l) = \arg \min_{j \in [\hat{q}]} D^2\{\mathcal{C}(\mathbf{A}_l), \mathcal{C}(\hat{\mathbf{A}}_j)\}$ for each $l \in [q]$. We then call $\hat{\mathbf{A}} = (\hat{\mathbf{A}}_1, \dots, \hat{\mathbf{A}}_{\hat{q}})$ an effective segmentation of \mathbf{A} if (i) $1 < \hat{q} \leq q$, and (ii) $\text{rank}(\hat{\mathbf{A}}_{l'}) = \sum_{l \in [q]: f(l)=l'} \text{rank}(\mathbf{A}_l)$ for each $l' \in [\hat{q}]$. The intuition is as follows. The effective segmentation implies that each identified group in $\hat{\mathbf{Z}}_t(\cdot)$ contains at least one, but not all, groups in $\mathbf{Z}_t(\cdot)$. Since our main target is to forecast $\tilde{\mathbf{Y}}_t(\cdot)$ based on the cross-serial dependence in $\{\mathbf{Z}_t^{(l)}(\cdot) : l \in [q]\}$, this segmentation result is effective in the sense that the linear dynamics in $\mathbf{Z}_t(\cdot)$ is well kept in $\{\hat{\mathbf{Z}}_t^{(l)}(\cdot) : l \in [\hat{q}]\}$ without any contamination or damage and a mild dimension reduction is achieved with $\hat{q} > 1$. For the special case of complete segmentation ($\hat{q} = q$), we use the maximum and averaged estimation errors for $(\hat{\mathbf{A}}_1, \dots, \hat{\mathbf{A}}_{\hat{q}})$, respectively,

Algorithm 1 General prediction procedure for multivariate functional time series

- (i) Treat the first $n - h$ observations as training data, adopt the normalization step to obtain $\tilde{\mathbf{Y}}_t(\cdot) = \{\hat{\mathbf{V}}_{\tilde{\mathbf{y}}}^{(h)}\}^{-1/2} \tilde{\mathbf{Y}}_t(\cdot)$, where $\hat{\mathbf{V}}_{\tilde{\mathbf{y}}}^{(h)}$ is the consistent estimator of $\mathbf{V}_{\tilde{\mathbf{y}}}$ in (29), and implement the procedure in **Section 2.2** on $\{\tilde{\mathbf{Y}}_t(\cdot)\}_{t=1}^{n-h}$ to obtain estimated transformation matrix $\hat{\mathbf{A}} = (\hat{\mathbf{A}}_1, \dots, \hat{\mathbf{A}}_{\hat{q}})$ and transformed curve subseries $\{\hat{\mathbf{Z}}_t^{(l)}(\cdot) : l \in [\hat{q}]\}$.
- (ii) Apply the procedure in **Section 3** on each $\{\hat{\mathbf{Z}}_t^{(l)}(\cdot)\}_{t=1}^{n-h}$ to achieve the h -step ahead prediction denoted as $\hat{\mathbf{Z}}_n^{(l)}(\cdot)$ for $l \in [\hat{q}]$. In particular, for each l , select the best VAR model that best fits each vector time series $\{\hat{\mathbf{Z}}_t^{(l)}\}_{t=1}^{n-h}$ according to the AIC criterion.
- (iii) Obtain the h -step ahead prediction $\hat{\mathbf{A}} \hat{\mathbf{Z}}_n(\cdot)$ for the normalized curves $\tilde{\mathbf{Y}}_n(\cdot)$ with $\hat{\mathbf{Z}}_n(\cdot) = \{\hat{\mathbf{Z}}_n^{(1)}(\cdot)^\top, \dots, \hat{\mathbf{Z}}_n^{(\hat{q})}(\cdot)^\top\}^\top$. Then the h -step ahead prediction for the original curves $\tilde{\mathbf{Y}}_n(\cdot)$ is given by $\hat{\mathbf{Y}}_n(\cdot) \equiv \{\hat{Y}_{n1}(\cdot), \dots, \hat{Y}_{np}(\cdot)\}^\top = \{\hat{\mathbf{V}}_{\tilde{\mathbf{y}}}^{(h)}\}^{1/2} \hat{\mathbf{A}} \hat{\mathbf{Z}}_n(\cdot)$.

defined as $\text{MaxE} = \max_{l \in [q]} D^2\{\mathcal{C}(\mathbf{A}_l), \mathcal{C}(\hat{\mathbf{A}}_{f(l)})\}$ and $\text{AvgE} = q^{-1} \sum_{l=1}^q D^2\{\mathcal{C}(\mathbf{A}_l), \mathcal{C}(\hat{\mathbf{A}}_{f(l)})\}$ to assess the ability of our method in fully recovering the spanned spaces $\mathcal{C}(\mathbf{A}_1), \dots, \mathcal{C}(\mathbf{A}_q)$. Note that \mathbf{A} in (29) can not be easily computed, as the true $\mathbf{V}_{\tilde{\mathbf{y}}}$ and $\mathbf{V}_{\tilde{\mathbf{z}}}$ are hard to find even for simulated examples. For $\tilde{\mathbf{A}}$ specified in (28), let $\tilde{\mathbf{A}} = \mathbf{V}_{\tilde{\mathbf{y}}}^{-1/2} \tilde{\mathbf{A}} \equiv (\tilde{\mathbf{A}}_1, \dots, \tilde{\mathbf{A}}_q)$ with $\tilde{\mathbf{A}}_l = \mathbf{V}_{\tilde{\mathbf{y}}}^{-1/2} \tilde{\mathbf{A}}_l$. Since $\mathbf{V}_{\tilde{\mathbf{z}}}$ is a block-diagonal matrix, then $\mathcal{C}(\tilde{\mathbf{A}}_l) = \mathcal{C}(\mathbf{A}_l)$ for $l \in [q]$. Hence, we can replace $\mathcal{C}(\mathbf{A}_l)$ by $\mathcal{C}(\{\hat{\mathbf{V}}_{\tilde{\mathbf{y}}}^{(h)}\}^{-1/2} \tilde{\mathbf{A}}_l)$ to obtain the approximations of MaxE and AvgE in our simulations.

To evaluate the post-sample predictive accuracy, we define the mean squared prediction error (MSPE) as

$$\text{MSPE} = \frac{1}{pN} \sum_{j=1}^p \sum_{i=1}^N \{\hat{Y}_{nj}(v_i) - \tilde{Y}_{nj}(v_i)\}^2 \quad (32)$$

with v_1, \dots, v_N being equally spaced points in $[0, 1]$, and compute the relative prediction error as the ratio of MSPE in (32) to that under the “oracle” case. In the oracle case, we apply the procedure in **Section 3** directly on each true $\{\tilde{\mathbf{Z}}_t^{(l)}(\cdot)\}_{t=1}^{n-h}$ to achieve the h -step ahead prediction for $\{\tilde{\mathbf{Z}}_n^{(l)}(\cdot) : l \in [q]\}$, denoted by $\{\tilde{\mathbf{Z}}_n^{(l)}(\cdot) : l \in [q]\}$, and further obtain the h -step ahead prediction $\tilde{\mathbf{A}}\{\tilde{\mathbf{Z}}_n^{(1)}(\cdot)^\top, \dots, \tilde{\mathbf{Z}}_n^{(q)}(\cdot)^\top\}^\top$ for the original curves $\tilde{\mathbf{Y}}_t(\cdot)$. By comparison, we also implement an **univariate** functional prediction method on each $\tilde{Y}_{tj}(\cdot)$ separately by performing univariate dimension reduction (Bathia, Yao, and Ziegelmann 2010), then predicting vector time series based on the best fitted VAR model and finally recovering functional prediction (denoted as UniV).

We generate $n \in \{200, 400, 800, 1600\}$ observations with $N = 30$ for each example and replicate each simulation 500 times. **Table 2** provides numerical summaries, including the relative frequencies of the effective segmentation with $\hat{q} = q$ and $\hat{q} \geq q - 1$, and the estimation errors for $\hat{\mathbf{A}} = (\hat{\mathbf{A}}_1, \dots, \hat{\mathbf{A}}_{\hat{q}})$ under the complete segmentation case. As one would expect, the proposed

Table 2. The relative frequencies of effective segmentation with respect to $\hat{q} = q$ and $\hat{q} \geq q - 1$, and the means (standard deviations) of MaxE, AvgE, and relative MSPEs (rMSPE) over 500 simulation runs.

		$n = 200$	$n = 400$	$n = 800$	$n = 1600$
Example 1 ($p = 6$)	$\hat{q} = q$	0.626	0.722	0.772	0.880
	$\hat{q} \geq q - 1$	0.930	0.988	0.998	1.000
	MaxE	0.128(0.088)	0.089(0.066)	0.053(0.048)	0.035(0.037)
	AvgE	0.079(0.052)	0.053(0.038)	0.030(0.025)	0.019(0.019)
	rMSPE - SegV	1.081(0.172)	1.048(0.105)	1.026(0.065)	1.014(0.048)
Example 2 ($p = 10$)	rMSPE - UniV	1.584(0.453)	1.598(0.423)	1.596(0.379)	1.651(0.443)
	$\hat{q} = q$	0.324	0.444	0.644	0.806
	$\hat{q} \geq q - 1$	0.490	0.688	0.874	0.972
	MaxE	0.301(0.108)	0.193(0.090)	0.117(0.064)	0.072(0.049)
	AvgE	0.183(0.059)	0.115(0.047)	0.069(0.035)	0.041(0.024)
Example 3 ($p = 15$)	rMSPE - SegV	1.291(0.271)	1.174(0.215)	1.089(0.143)	1.059(0.091)
	rMSPE - UniV	1.708(0.404)	1.836(0.410)	1.841(0.436)	1.862(0.392)
	$\hat{q} = q$	0.032	0.178	0.410	0.622
	$\hat{q} \geq q - 1$	0.086	0.344	0.616	0.832
	MaxE	0.426(0.091)	0.347(0.121)	0.241(0.113)	0.157(0.091)
	AvgE	0.273(0.054)	0.195(0.050)	0.128(0.042)	0.077(0.033)
	rMSPE - SegV	1.477(0.313)	1.363(0.277)	1.166(0.156)	1.091(0.098)
	rMSPE - UniV	1.805(0.370)	1.967(0.394)	2.033(0.394)	2.001(0.384)

method provides higher proportions of effective segmentation and lower estimation errors as n increases, and performs fairly well for reasonably large n as p increases. For $(p, n) = (6, 200)$, we observe 62.6% complete segmentation with AvgE as low as 0.079. Furthermore, the proportions of effective segmentation with $\hat{q} \geq q - 1$ are above 93% for $n \geq 200$. Similar results can be found for cases of $(p, n) = (10, 800+)$ and $(15, 1600)$, whose proportions of effective segmentation with $\hat{q} \geq q - 1$ remain higher than 87.4% and 83.2%, respectively. Table 2 also reports the relative one-step ahead prediction errors. It is evident that SegV significantly outperforms UniV in all settings, demonstrating the effectiveness of our proposed segmentation transformation and dimension reduction in predicting future values. Although the proportions of complete segmentation are not high when $p = 15$, the corresponding proportions of $\hat{q} \geq q - 1$ become satisfactorily higher, and SegV performs similarly to the oracle case with its relative prediction errors being closer to 1 as n increases.

5.2. Cases with Large p

Under a large p scenario, a natural question to ask is whether the segmentation method based on the classical estimation for autocovariance functions of $\tilde{Y}_t(\cdot)$ (denoted as NonT) as (31) in Section 5.1 still performs well, and if not, whether a satisfactory improvement is attainable via the functional-thresholding estimation (denoted as FunT) developed in Section 2.3. To this end, we generate $\tilde{Y}_t(\cdot)$ from (28) with $p \in \{30, 60\}$ and $n \in \{200, 400\}$. Specifically, we let $\tilde{X}_{t(3l-2)}(\cdot) = \vartheta_{tl}(\cdot)$, $\tilde{X}_{t(3l-1)}(\cdot) = \vartheta_{(t+1)l}(\cdot)$, $\tilde{X}_{t(3l)}(\cdot) = \vartheta_{(t+2)l}(\cdot)$ for $l \in [q]$. This setting ensures q uncorrelated groups of curve subseries in $\tilde{X}_t(\cdot)$ with $p_l = 3$ component curves per group and hence $q = 10$ and 20 correspond to $p = 30$ and 60, respectively. Let the $p \times p$ transformation matrix $\tilde{\mathbf{A}} = \mathbf{\Delta}_1 + \delta \mathbf{\Delta}_2$. Here $\mathbf{\Delta}_1 = \text{diag}\{\mathbf{\Delta}_{11}, \dots, \mathbf{\Delta}_{1(p/6)}\}$ with elements of each $\mathbf{\Delta}_{li} \in \mathbb{R}^{6 \times 6}$ being sampled independently from Uniform $[-3, 3]$ for $i \in [p/6]$, and $\mathbf{\Delta}_2$ is a matrix with two randomly selected nonzero elements from Uniform $[-1, 1]$ each row. We set $\delta \in \{0.1, 0.5\}$. It is notable that our setting results

in a very high-dimensional learning task in the sense that the intrinsic dimension $30 \times 5 = 150$ or $60 \times 5 = 300$ is large relative to the sample size $n = 200$ or 400.

We assess the performance of NonT and FunT in discovering the group structure. The optimal thresholding parameters $\hat{\omega}_k$ in FunT are selected by the 5-fold cross-validation (see Remark 2), and $\mathbf{V}_{\tilde{y}}$ in the normalization step is estimated by $\hat{\mathbf{V}}_{\tilde{y}}^{(0)}$ given in (30), as the threshold version of $\hat{\mathbf{V}}_{\tilde{y}}^{(0)}$ might not be positive definite. In practice, when p is large, FunT may lead to segmentation with a small \hat{q} , indicating that some groups of $\{\hat{\mathbf{Z}}_t^{(l)}(\cdot) : l \in [\hat{q}]\}$ contain multiple groups in $\{\mathbf{Z}_t^{(l)}(\cdot) : l \in [q]\}$. To ease the modeling burden of complex VAR process, we may consider performing further segmentation transformation on the estimated groups by repeating FunT R times. To be precise, the i th round of segmentation transformation via FunT is performed within each group discovered in the $(i - 1)$ th round with $c_{\varrho} = 1$ for $i \in [R]$, and hence $(\hat{\mathbf{A}}_1, \dots, \hat{\mathbf{A}}_{\hat{q}})$ is updated after each iteration. Table 3 reports the relative frequencies of the effective segmentation for NonT and FunT with $R \in \{1, 5, 10\}$. Finally, we apply FunT-based SegV (denoted as FTSegV) combined with the R -round segmentation transformation for $R \in \{1, 5, 10\}$ in Step (i) of Algorithm 1, and compare their one-step ahead predictive performance with UniV and SegV. Table 4 summarizes the relative prediction errors for all five comparison methods.

Several conclusions can be drawn from Tables 3 and 4. First, the performance of SegV severely deteriorates under the high-dimensional setting, as this procedure fails to detect any effective segmentation, resulting in elevated prediction errors. By comparison, FTSegV exhibits superior predictive ability over SegV and UniV. In particular, for large n , for example, $n = 400$, FTSegV does a reasonably good job in recovering the group structure of $\mathbf{Z}_t(\cdot)$ and performs comparably well to the oracle method with the relative prediction errors lower than 1.149 in all scenarios. Second, comparing the results for $n = 200$ among different R , we observe an interesting phenomenon that even though the relative frequencies of effective segmentation for FunT drop as R increases, implying that some groups in $\{\hat{\mathbf{Z}}_t^{(l)}(\cdot) : l \in [\hat{q}]\}$ are split incorrectly before forecasting,

Table 3. The relative frequencies of effective segmentation over 500 simulation runs.

(p, δ)	NonT		FunT					
	$n = 200$	$n = 400$	$R = 1$		$R = 5$		$R = 10$	
			$n = 200$	$n = 400$	$n = 200$	$n = 400$	$n = 200$	$n = 400$
(30, 0.1)	0	0	0.706	1.000	0.556	1.000	0.546	1.000
(30, 0.5)	0	0	0.588	1.000	0.436	1.000	0.420	1.000
(60, 0.1)	0	0	0.298	1.000	0.148	1.000	0.144	1.000
(60, 0.5)	0	0	0.194	0.996	0.078	0.990	0.072	0.990

Table 4. Means (standard deviations) of relative MSPEs over 500 simulation runs.

Method	(p, δ)	$n=200$	$n=400$	(p, δ)	$n=200$	$n=400$
FTSegV ($R = 1$)	(30, 0.1)	1.243(0.162)	1.095(0.105)	(60, 0.1)	1.249(0.122)	1.110(0.073)
FTSegV ($R = 5$)		1.225(0.153)	1.091(0.101)		1.250(0.123)	1.104(0.071)
FTSegV ($R = 10$)		1.222(0.151)	1.087(0.099)		1.249(0.122)	1.099(0.071)
SegV		1.814(0.376)	1.901(0.368)		1.813(0.271)	1.907(0.265)
UniV	(30, 0.5)	1.631(0.313)	1.735(0.317)	(60, 0.5)	1.599(0.214)	1.682(0.210)
FTSegV ($R = 1$)		1.268(0.176)	1.134(0.134)		1.285(0.134)	1.149(0.101)
FTSegV ($R = 5$)		1.255(0.171)	1.128(0.130)		1.282(0.136)	1.142(0.098)
FTSegV ($R = 10$)		1.250(0.168)	1.128(0.127)		1.281(0.136)	1.141(0.099)
SegV		1.815(0.377)	1.903(0.369)		1.813(0.271)	1.905(0.264)
UniV		1.635(0.315)	1.740(0.317)		1.603(0.215)	1.684(0.209)

the prediction errors stay low and slightly decrease as shown in Table 4. This is not surprising, since further segmentation based on FunT yields fewer parameters to be estimated in VAR models and thus benefits the forecasting accuracy even if a few small but significant cross-covariances of $\mathbf{Z}_t(\cdot)$ are ignored. Such finding highlights the success of FTSegV and its R -round segmentation in the sense that although FTSegV may not be able to accurately recover the group structure in $\mathbf{Z}_t(\cdot)$ for a small n , it achieves an appropriate dimension reduction to provide significant improvement in high-dimensional functional prediction.

5.3. General Data-Generating Cases

To further illustrate the advantage of our proposed segmentation transformation in predicting high-dimensional functional time series, we simulate data from a more generalized functional time series framework instead of strictly adhering to (2). Specifically, we consider the vector functional autoregressive (VFAR) model of order 1,

$$\mathbf{Y}_t(u) = \int_{\mathcal{U}} \mathbf{Q}(u, v) \mathbf{Y}_{t-1}(v) dv + \boldsymbol{\epsilon}_t(u), \quad u \in \mathcal{U}, t \in [200], \quad (33)$$

where $\boldsymbol{\epsilon}_t(\cdot) = \{\epsilon_{t1}(\cdot), \dots, \epsilon_{tp}(\cdot)\}^\top$ are independently sampled from a p -dimensional vector of mean zero Gaussian processes, independent of $\mathbf{Y}_{t-1}(\cdot)$, and $\mathbf{Q} = (Q_{ij})_{i,j \in [p]}$ is the functional transition matrix with each $Q_{ij} \in \mathbb{S}$. See Section H1 of the supplementary material for the detailed data-generating process.

We compare the predictive performance of three competing methods. The first VFAR method is developed by knowing the true data-generating process through VFAR model. We relegate the detailed prediction procedure to Section H1 of the supplementary material. We next consider two segmentation-based prediction methods:

- (Seg+Y method) For the original curve series $\{\mathbf{Y}_t(\cdot)\}_{t \in [200]}$, we compute the sample estimates $\{\widehat{\Sigma}_{y,k,ij}(u, v)\}_{i,j \in [p]}$ for $k \in \{0\} \cup [5]$ as in Section 2.3. Let $\hat{T}_{y,ij} = \max_{|k| \leq 5} \|\widehat{\Sigma}_{y,k,ij}\|_{\mathcal{S}}$

Table 5. The mean of MSPEs over 500 simulation runs.

Method	$p = 10$	$p = 15$	$p = 20$	$p = 25$	$p = 30$	$p = 35$
VFAR	6.709	7.974	10.506	13.439	20.149	40.552
Seg+Y	6.314	6.691	8.931	11.313	17.001	32.846
Seg+Z	6.324	6.682	8.267	8.998	12.605	17.275

and sort $\hat{T}_{y,ij}$'s for $1 \leq i < j \leq p$ in descending order. We recognize $Y_{ti}(\cdot)$ and $Y_{tj}(\cdot)$ as belonging to the same group if $\hat{T}_{y,ij}$ is ranked among 10% of all $p(p-1)/2$ sorted values. We then segment the p component series $Y_{tj}(\cdot)$'s into several nonoverlapping groups and apply VFAR to each identified group to obtain its one-step ahead prediction.

- (Seg+Z method) Consider the transformed curve series $\widehat{\mathbf{Z}}_t(u) = \widehat{\mathbf{A}}^\top \{\widehat{\mathbf{V}}_y^{(0)}\}^{-1/2} \mathbf{Y}_t(u)$, where $\widehat{\mathbf{A}}$ is obtained by implementing the procedure in Section 2.2 on the normalized process $\{[\widehat{\mathbf{V}}_y^{(0)}]^{-1/2} \mathbf{Y}_t(\cdot)\}_{t \in [200]}$. We perform the same segmentation procedure as in Seg+Y to $\{\widehat{\mathbf{Z}}_t(\cdot)\}_{t \in [200]}$, apply VFAR to each of the identified groups of $\{\widehat{\mathbf{Z}}_t(\cdot)\}_{t \in [200]}$ to obtain the one-step ahead prediction $\check{\mathbf{Z}}_{201}(\cdot)$, and finally obtain $\{\widehat{\mathbf{V}}_y^{(0)}\}^{1/2} \widehat{\mathbf{A}} \check{\mathbf{Z}}_{201}(\cdot)$ as the one-step ahead prediction for the original curve series.

Table 5 reports one-step ahead MSPEs for three methods with different values of p . As anticipated, the performance of VFAR deteriorates severely as p increases, demonstrating that the joint model suffers from the high-dimensionality, even when the true model is known. Meanwhile, both segmentation-based prediction methods exhibit improved predictive performance, with Seg+Z notably outperforming Seg+Y, particularly in scenarios with large p . It is crucial to emphasize that the improvement of Seg+Z over Seg+Y is attributed to the decorrelation transformation. Table S1 in the supplementary material provides further insights into the impact of transformation, where \hat{q}_y and \hat{q}_z denote the numbers of the identified groups using Seg+Y and Seg+Z, respectively. Interestingly, Seg+Z yields more groups than Seg+Y while retaining the same amount

of strongly connected pairs. This observation indicates that the decorrelation transformation effectively pushes the cross-autocorrelations that were previously spread over p components into a block-diagonally dominate structure, where the cross-autocorrelations along the block diagonal are significantly stronger than those off the diagonal. Such enhancement of within-group autocorrelations, along with the reduction of cross-autocorrelations between the groups, leads to reasonably good segmentation by only retaining the strong within-group cross-autocorrelations while ignoring the weak between-group cross-autocorrelations, and thus yields more accurate future predictions.

6. Real Data Analysis

In this section, we apply our proposed SegV and FTSegV to two real data examples arising from different fields. Our main goal is to evaluate the post-sample predictive accuracy of both methods. By comparison, we also implement componentwise univariate prediction method (UniV) and the multivariate prediction method of Gao, Shang, and Yang (2019) (denoted as GSY) to jointly predict p component series by fitting a factor model to estimated scores obtained via eigenanalysis of the long-run covariance function (Hörmann, Kidziński, and Hallin 2015). It is worth mentioning that the joint prediction model VmV (see Example 1) completely fail due to high dimensionality, so we do not report their results here. To evaluate the effectiveness of the segmentation transformation and its impact on prediction, we forge two other segmentation cases, namely **under**-segmentation and **uni**-segmentation, for both SegV and FTSegV (denoted as Under.SegV, Uni.SegV, Under.FTSegV and Uni.FTSegV, respectively). Denote by $\{\hat{G}_l : l \in [\hat{q}]\}$ the segmented groups of $\{\hat{\mathbf{Z}}_t^{(l)}(\cdot) : l \in [\hat{q}]\}$ discovered in Step (i) of Algorithm 1 (seen also as correct-segmentation). The under-segmentation updates $\{\hat{G}_l : l \in [\hat{q}]\}$ by merging two groups \hat{G}_{l_1} and $\hat{G}_{l'_1}$ together before subsequent analysis, where $\arg \max_{(i,j): i \in \hat{G}_{l_1}, j \in \hat{G}_{l'_1}, 1 \leq l \neq l' \leq \hat{q}} \hat{T}_{ij} \in \hat{G}_{l_1} \times \hat{G}_{l'_1}$ with \hat{T}_{ij} defined in (9). The uni-segmentation, on the other hand, regards each curve component of $\{\hat{\mathbf{Z}}_t^{(l)}(\cdot) : l \in [\hat{q}]\}$ as an individual group and then applies UniV componentwisely. For a fair comparison, the orders of VAR models adopted in all SegV/FTSegV-related methods and UniV are determined by the AIC criterion, while GSY is implemented using the R package `fetsa`.

To examine the predictive performance, we apply an expanding window approach to the observed data $\check{Y}_{tj}(v_i)$ for $t \in [n], j \in [p], i \in [N]$. We first split the dataset into a training set and a test set respectively consisting of the first n_1 and the remaining n_2 observations. For any positive integer h , we implement each comparison method on the training set $\{\check{Y}_{tj}(v_i) : t \in [n_1], j \in [p], i \in [N]\}$ and obtain its h -step ahead prediction, denoted as $\hat{Y}_{(n_1+h)j}^{(h)}(v_i)$, based on the fitted model. We then increase the training size by one, that is $\{\check{Y}_{tj}(v_i) : t \in [n_1 + 1], j \in [p], i \in [N]\}$, refit the model and compute the next h -step ahead prediction $\hat{Y}_{(n_1+1+h)j}^{(h)}(v_i)$ for $j \in [p], i \in [N]$. Repeat the above procedure until the last h -step ahead prediction $\hat{Y}_{nj}^{(h)}(v_i)$ is produced. Finally, we compute the h -step ahead MAPE and

MSPE as

$$\begin{aligned} \text{MAPE}(h) &= \frac{1}{(n_2 + 1 - h)pN} \sum_{t=n_1+h}^n \sum_{j=1}^p \sum_{i=1}^N |\hat{Y}_{tj}^{(h)}(v_i) - \check{Y}_{tj}(v_i)|, \\ \text{MSPE}(h) &= \frac{1}{(n_2 + 1 - h)pN} \sum_{t=n_1+h}^n \sum_{j=1}^p \sum_{i=1}^N \{\hat{Y}_{tj}^{(h)}(v_i) - \check{Y}_{tj}(v_i)\}^2. \end{aligned} \quad (34)$$

6.1. Age-Specific Mortality Data

The first dataset, analyzed in Tang, Shang, and Yang (2022), contains age-specific and gender-specific mortality rates for developed countries during 1965–2013 ($n = 49$). See Table S3 in the supplementary material for the list of $p = 29$ countries after removing certain countries with missing data. Following the proposal of Tang, Shang, and Yang (2022), we model the log transformation of the mortality rate of people aged $v_i = i - 1$ living in the j th country during year $1964 + t$ as a random curve $\check{Y}_{tj}(v_i)$ ($t \in [49], j \in [29], i \in [101]$) and perform smoothing for observed mortality curves via smoothing splines. We divide the smoothed dataset into the training set of size $n_1 = 34$ and the test set of size $n_2 = 15$. Since the smoothed curve series exhibit weak autocorrelations when lags are beyond 3 and the training size is relatively small, we use $k_0 = m = 3$ in our procedure for this example.

Table 6 reports the MAPEs and MSPEs for females and males. Several obvious patterns are observable. First, our proposed methods, SegV and FTSegV, provide the best predictive performance uniformly for both females and males, and all h . This demonstrates the effectiveness of reducing the number of parameters via the segmentation transformation in predicting high-dimensional functional time series. Second, although the cases of under- and uni-segmentation are inferior to the correct-segmentation case, they significantly outperform UniV and GSY. Note that the improvement of Uni.SegV over UniV reveals the capability of the transformation matrix $\hat{\mathbf{A}}$ to effectively decorrelate the original curves, thereby leading to more accurate predictions. One may also notice that, Uni.SegV does not perform as well as SegV and Under.SegV. In most cases, the transformed curve series exhibits $\hat{q} = 26$ groups, with 25 groups of size 1 and one large group of size 4; see Figures S2–S11 in the supplementary material. The limitation of Uni.SegV thus becomes apparent as it fails to account for the cross-serial dependence within the large group, resulting in less accurate predictions. This finding again confirms the effectiveness of our procedure, in particular, the within-group cross-autocorrelations is also valuable in forecasting future values.

6.2. Energy Consumption Data

Our second dataset contains energy consumption readings (in kWh) taken at half hourly intervals for thousands of London households, and is available at <https://data.london.gov.uk/dataset/smartmeter-energy-use-data-in-london-households>. In

Table 6. MAPEs and MSPEs for eight competing methods on the female and male mortality curves for $h \in \{1, 2, 3\}$.

Method	MAPE			MSPE			MAPE			MSPE		
	$h = 1$	$h = 2$	$h = 3$	$h = 1$	$h = 2$	$h = 3$	$h = 1$	$h = 2$	$h = 3$	$h = 1$	$h = 2$	$h = 3$
SegV	1.157	1.461	1.806	0.291	0.401	0.566	1.104	1.391	1.727	0.251	0.354	0.499
Under.SegV	1.201	1.510	1.874	0.304	0.417	0.593	1.123	1.425	1.751	0.251	0.358	0.500
Uni.SegV	1.526	1.821	2.154	0.441	0.579	0.767	1.302	1.573	1.892	0.324	0.443	0.598
FTSegV	1.175	1.458	1.801	0.301	0.405	0.569	1.101	1.391	1.732	0.251	0.353	0.502
Under.FTSegV	1.206	1.510	1.876	0.309	0.421	0.598	1.118	1.418	1.743	0.251	0.356	0.499
Uni.FTSegV	1.560	1.838	2.173	0.457	0.585	0.776	1.300	1.573	1.897	0.324	0.444	0.602
UniV	1.761	2.032	2.325	0.603	0.749	0.925	1.561	1.825	2.127	0.467	0.596	0.759
GSY	2.476	2.515	2.577	1.434	1.447	1.451	2.144	2.110	2.201	1.112	1.023	1.043

NOTE: All numbers are multiplied by 10. The lowest values for each h are in bold font.

Table 7. MAPEs and MSPEs for eight competing methods on the energy consumption curves for $h \in \{1, 2, 3\}$ and $p \in \{40, 80\}$.

Method	MAPE			MSPE			MAPE			MSPE		
	$h = 1$	$h = 2$	$h = 3$	$h = 1$	$h = 2$	$h = 3$	$h = 1$	$h = 2$	$h = 3$	$h = 1$	$h = 2$	$h = 3$
SegV	1.639	1.748	1.793	0.047	0.053	0.054	1.996	2.058	2.071	0.070	0.075	0.075
Under.SegV	1.669	1.766	1.794	0.048	0.054	0.054	2.025	2.092	2.104	0.072	0.077	0.077
Uni.SegV	1.709	1.873	1.964	0.049	0.058	0.062	2.022	2.132	2.187	0.070	0.078	0.081
FTSegV	1.637	1.747	1.791	0.047	0.053	0.054	2.012	2.055	2.070	0.071	0.074	0.074
Under.FTSegV	1.669	1.766	1.793	0.048	0.054	0.054	2.040	2.087	2.104	0.073	0.076	0.077
Uni.FTSegV	1.708	1.872	1.963	0.049	0.058	0.062	2.045	2.138	2.190	0.072	0.078	0.081
UniV	1.867	2.009	2.109	0.058	0.067	0.072	2.221	2.362	2.463	0.083	0.093	0.100
GSY	2.142	2.264	2.320	0.099	0.110	0.119	2.833	2.826	2.781	0.159	0.159	0.159

NOTE: All numbers are multiplied by 10^2 . The lowest values for each h are in bold font.

our study, we select households with flat energy prices during the period between December 2012 and May 2013 ($n = 182$) after removing samples with too many missing records, and hence construct 4000 samples of daily energy consumption curves observed at $N = 48$ equally spaced time points following the proposal of Cho et al. (2013). To alleviate the impact of randomness from individual curves, we randomly split the data into p groups of equal size, then take the sample average of curves within each group and finally smooth the averaged curves based on a 15-dimensional Fourier basis. We target to evaluate the h -day ahead predictive accuracy for the p -dimensional intraday energy consumption averaged curves in May 2013 based on the training data from December 2012 to the previous day. The eight comparison methods are built in the same manner as Section 6.1 with $k_0 = m = 5$.

Table 7 presents the mean prediction errors for $h \in \{1, 2, 3\}$ and $p \in \{40, 80\}$. A few trends are apparent. First, the prediction errors for $p = 80$ are higher than those for $p = 40$ as higher dimensionality poses more challenges in prediction. Second, likewise in previous examples, SegV and FTSegV attain the lowest prediction errors in comparison to five competing methods under all scenarios. All segmentation-based methods consistently outperform UniV and GSY by a large margin. Third, despite being developed for high-dimensional functional time series prediction, GSY provides the worst result in this example.

Supplementary Materials

The supplementary material contains all technical proofs of the main results, and additional empirical results.

Acknowledgments

We are grateful to the editor, the associate editor and two referees for their insightful comments and suggestions, which have led to significant improvement of our article.

Disclosure Statement

The authors report there are no competing interests to declare.

Funding

Chang was supported in part by the National Key R&D Program of China (No. 2023YFA1008701), and the National Natural Science Foundation of China (grant nos. 72125008 and 71991472). Qiao was partially supported by the Australian Research Council Discovery Project (grant no. DP230102250). Yao was partially supported by the U.K. Engineering and Physical Sciences Research Council (grant nos. EP/V007556/1 and EP/X002195/1).

ORCID

Jinyuan Chang  <http://orcid.org/0000-0001-7933-4449>

Qin Fang  <http://orcid.org/0009-0004-7152-1142>

Qiwei Yao  <http://orcid.org/0000-0003-2065-8486>

References

- Ahn, S. C., and Horenstein, A. R. (2013), "Eigenvalue Ratio Test for the Number of Factors," *Econometrica*, 81, 1203–1227. [6]
- Aue, A., Norinho, D. D., and Hörmann, S. (2015), "On the Prediction of Stationary Functional Time Series," *Journal of the American Statistical Association*, 110, 378–392. [2]
- Aue, A., Rice, G., and Sonmez, O. (2018), "Detecting and Dating Structural Breaks in Functional Data Without Dimension Reduction," *Journal of the Royal Statistical Society, Series B*, 80, 509–529. [2]
- Bathia, N., Yao, Q., and Ziegelmann, F. (2010), "Identifying the Finite Dimensionality of Curve Time Series," *The Annals of Statistics*, 38, 3352–3386. [1,2,10]
- Bickel, P., and Levina, E. (2008), "Covariance Regularization by Thresholding," *The Annals of Statistics*, 36, 2577–2604. [6]
- Chang, J., Chen, C., Qiao, X. and Yao, Q. (2024), "An Autocovariance-based Learning Framework for High-Dimensional Functional Time Series," *Journal of Econometrics*, 390, 105385. [2]

- Chang, J., Guo, B., and Yao, Q. (2015), “High Dimensional Stochastic Regression with Latent Factors, Endogeneity and Nonlinearity,” *Journal of Econometrics*, 189, 297–312. [8]
- Chang, J., Guo, B., and Yao, Q. (2018), “Principal Component Analysis for Second-Order Stationary Vector Time Series,” *The Annals of Statistics*, 46, 2094–2124. [2,8]
- Chiou, J.-M., Chen, Y.-T., and Yang, Y.-F. (2014), “Multivariate Functional Principal Component Analysis: A Normalization Approach,” *Statistica Sinica*, 24, 1571–1596. [1,6]
- Cho, H., Goude, Y., Brossat, X. and Yao, Q. (2013), “Modeling and Forecasting Daily Electricity Load Curves: A Hybrid Approach,” *Journal of the American Statistical Association*, 108, 7–21. [2,14]
- Gao, Y., Shang, H. L., and Yang, Y. (2019), “High-Dimensional Functional Time Series Forecasting: An Application to Age-Specific Mortality Rates,” *Journal of Multivariate Analysis*, 170, 232–243. [1,2,13]
- Guo, S., and Qiao, X. (2023), “On Consistency and Sparsity for High-Dimensional Functional Time Series with Application to Autoregressions,” *Bernoulli*, 29, 451–472. [1,2,8]
- Happ, C., and Greven, S. (2018), “Multivariate Functional Principal Component Analysis for Data Observed on Different (Dimensional) Domains,” *Journal of the American Statistical Association*, 113, 649–659. [1,6]
- Hörmann, S., Kidziński, L., and Hallin, M. (2015), “Dynamic Functional Principal Components,” *Journal of the Royal Statistical Society, Series B*, 77, 319–348. [1,2,13]
- Hu, X., and Yao, F. (2022), “Sparse Functional Principal Component Analysis in High Dimensions,” *Statistica Sinica*, 32, 1939–1960. [1]
- Lam, C., and Yao, Q. (2012), “Factor Modeling for High-Dimensional Time Series: Inference for the Number of Factors,” *The Annals of Statistics*, 40, 694–726. [6]
- Li, D., Robinson, P. M., and Shang, H. L. (2020), “Long-Range Dependent Curve Time Series,” *Journal of the American Statistical Association*, 115, 957–971. [2]
- Rice, G., and Shum, M. (2019), “Inference for the Lagged Cross-Covariance Operator between Functional Time Series,” *Journal of Time Series Analysis*, 40, 665–692. [3]
- Tan, J., Liang, D., Guan, Y. and Huang, H. (in press), “Graphical Principal Component Analysis of Multivariate Functional Time Series,” *Journal of the American Statistical Association*. [2]
- Tang, C., Shang, H.-L., and Yang, Y. (2022), “Clustering and Forecasting Multiple Functional Time Series,” *The Annals of Applied Statistics*, 16, 2523–2553. [2,3,13]
- Tavakoli, S., Nisol, G., and Hallin, M. (2023), “Factor Models for High-Dimensional Functional Time Series II: Estimation and Forecasting,” *Journal of Time Series Analysis*, 44, 601–621. [2]
- Zhou, Z., and Dette, H. (2023), “Statistical Inference for High-Dimensional Panel Functional Time Series,” *Journal of the Royal Statistical Society, Series B*, 85, 523–549. [2]