On Functional Data Analysis: Methodologies and Applications

by

Renfang Tian

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Doctor of Philosophy
in
Economics

Waterloo, Ontario, Canada, 2020

Examining Committee Membership

The following served on the Examining Committee for this thesis. The decision of the Examining Committee is by majority vote.

External Examiner: Marine Carrasco
Professor
Department of Economics
University of Montreal

Supervisor: Tao Chen
Associate Professor
Department of Economics
University of Waterloo

Internal Members: Pierre Chaussé
Associate Professor
Department of Economics
University of Waterloo

Thomas Parker
Assistant Professor
Department of Economics
University of Waterloo

Internal-External Member: Tony S. Wirjanto
Professor
Department of Statistics and Actuarial Science
University of Waterloo

Author's Declaration

This thesis consists of material all of which I authored or co-authored: see Statement of Contributions included in the thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Statement of Contributions

While Chapter 1 and 2 are not sole-authored (Chapter 1 is co-authored with my supervisor Professor Tao Chen and Professor Joseph DeJuan from the Department of Economics at the University of Waterloo; Chapter 2 is co-authored with my supervisor Professor Tao Chen and Professor Jiawen Xu from the School of Economics at Shanghai University of Finance and Economics), I have made the major contribution to the work involved in proving the results.

Abstract

In economic analyses, the variables of interest are often functions defined on continua such as time or space, though we may only have access to discrete observations – such type of variables are said to be "functional" (Ramsay, 1982). Traditional economic analyses model discrete observations using discrete methods, which can cause misspecification when the data are driven by functional underlying processes and further lead to inconsistent estimation and invalid inference. This thesis contains three chapters on functional data analysis (FDA), which concerns data that are functional in nature. As a nonparametric method accommodating functional data of different levels of smoothness, not only does FDA recover the functional underlying processes from discrete observations without misspecification, it also allows for analyses of derivatives of the functional data.

Specifically, Chapter 1 provides an application of FDA in examining the distribution equality of GDP functions across different versions of the Penn World Tables (PWT). Through our bootstrap-based hypothesis test and applying the properties of the derivatives of functional data, we find no support for the distribution equality hypothesis, indicating that GDP functions in different versions do not share a common underlying distribution. This result suggests a need to use caution in drawing conclusions from a particular PWT version, and conduct appropriate sensitivity analyses to check the robustness of results.

In Chapter 2, we utilize a FDA approach to generalize dynamic factor models. The newly proposed generalized functional dynamic factor model adopts two-dimensional loading functions to accommodate possible instability of the loadings and lag effects of the factors nonparametrically. Large sample theories and simulation results are provided. We also present an application of our model using a widely used macroeconomic data set.

In Chapter 3, I consider a functional linear regression model with a forward-in-time-only causality from functional predictors onto a functional response. In this chapter, (i) a uniform convergence rate of the estimated functional coefficients is derived depending on the degree of cross-sectional dependence; (ii) asymptotic normality of the estimated coefficients can be obtained under proper conditions, with unknown forms of cross-sectional dependence; (iii) a bootstrap method is proposed for approximating the distribution of the estimated functional coefficients. A simulation analysis is provided to illustrate the estimation and bootstrap procedures and to demonstrate the properties of the estimators.

Acknowledgements

I wish to thank Professors Tao Chen, Pierre Chaussé, Thomas Parker and Tony Wirjanto for their sound advice and kind support. Without them, this thesis would not have been possible. Special thanks to my supervisor, Tao Chen, who makes my PhD study a profound and life-changing experience. Tao has trained me with his rigorous and critical thinking, inspired me with his passion and creativity, while guided me to seek for my own motivation to keep exploring the world. I have learned from him that every milestone is a new start, and life is a process of endless exploration and discovery.

I would also like to express my gratitude to all the members of faculty, staff and my colleagues, who have helped me and offered deep insight into my PhD. They have made my experience at the University of Waterloo memorable and meaningful.

Finally, I thank my parents and my friends for their constant support and encouragement.

Dedication

To my parents I dedicate this thesis.

Table of Contents

List of Tables

List of Figures

Introduction

In economic analyses, the variables of interest are often functions defined on continua such as time or space, though we may only have access to discrete observations – such type of variables are said to be "functional" (Ramsay, 1982). For example, data for international trade in goods or services are usually by month or year; however, trade can happen at any points of time during continuous time periods, which makes the underlying processes of trade functions over time intervals. Traditional economic analyses models the discrete observations using discrete methods, which can cause misspecification when the observations are driven by such functional underlying processes and further lead to inconsistent estimation as well as invalid inference.

Functional data analysis (FDA) proposed by Ramsay (1982) and Ramsay and Dalzell (1991), as a nonparametric and continuous analysis approach concerning data that are functional in nature, started to gain attentions and become a powerful tool in various fields of studies, such as economics (e.g., Grambsch et al., 1995; Ramsay and Ramsey, 2002; Benatia et al., 2017; Chen et al., 2018, Working Paper.a), finance (e.g., Bapna et al., 2008; Laukaitis, 2008; Chen et al., Working Paper.b), environmental studies (e.g., Gao, 2007; Meiring, 2007), bioscience (e.g., Müller et al., 2009; Durá et al., 2010; Zhu et al., 2010), sports (e.g., Chen and Fan, 2018), along with others (Ullah and Finch, 2013).

This thesis contains three chapters developing methodologies and motivating applications of FDA, comprising hypothesis tests for functional data, functional factor models and functional regression. Specifically, Chapter 1, co-authored with Tao Chen and Joseph De Juan, provides an application of FDA in examining the distribution equality of GDP functions across different versions of the Penn World Tables (PWT). The idea is motivated by the fact that data in the PWT have been subject to a series of revisions since its first release in the early 1990s, and the amendments are substantial for many countries. Through our bootstrap-based hypothesis test and applying the properties of the derivatives of functional data, we find no support for the distribution equality hypothesis, indicating that GDP in different versions do not share a common underlying distribution. This result suggests a need to use caution in drawing conclusions from a particular PWT version, and conduct appropriate sensitivity analyses to check the robustness of results.

In Chapter 2, co-authored with Tao Chen and Jiawen Xu, we utilize a FDA approach

to generalize dynamic factor models. The newly proposed generalized functional dynamic factor model adopts two-dimensional loading functions to accommodate possible instability of the loadings and lag effects of the factors nonparametrically. Large sample theories and simulation results are provided. We also present an application of our model using a widely used macroeconomic data set.

In Chapter 3, I consider a functional linear regression model with a forward-in-time-only causality from functional predictors onto a functional response; such a model is also referred to as the historical functional linear model. This chapter contributes to the literature by establishing the asymptotics of B-spline-based estimated functional coefficients and developing the bootstrap inference, accommodating unknown forms of cross-sectional dependence. The main findings are (i) a uniform convergence rate of the estimated functional coefficients is derived depending on the degree of cross-sectional dependence and $\sqrt{n}$-consistency can be achieved in the absence of cross-sectional dependence; (ii) with unknown forms of cross-sectional dependence, asymptotic normality of the estimated coefficients can be obtained under proper conditions; (iii) the proposed bootstrap method has a better finite-sample performance than the asymptotics while approximating the distribution of the estimated functional coefficients. A simulation analysis is provided to illustrate the estimation and bootstrap procedures and to demonstrate the properties of the estimators.

Chapter 1

Distributions of GDP Across Versions of the Penn World Tables: A Functional Data Analysis Approach[1]

1.1 Introduction and Motivation

The Penn World Table (PWT) has become the most widely used database for empirical research aimed at explaining income differences between countries. Yet, despite its popularity, concerns have been raised regarding (i) data quality of GDP estimates in a given PWT version, and (ii) data consistency of estimates across versions. Summers and Heston (1991) note early on that GDP for about two-thirds of the countries in the database have margins of error of ten to forty percent. They summarize the severity of data inaccuracies by assigning each country a quality grade of A, B, C, or D, with A being the best and D the worst.

With data consistency, Breton (2012) and Johnson et al. (2013) report that GDP estimates in some countries for a given year are vastly different across versions despite being

[1]This chapter is co-authored with Tao Chen and Joseph De Juan.

derived from the same source and comparable data construction methodologies. Breton (2012), in particular, finds the year-by-year GDP level of the UK and the Philippines, two countries that participated in all the price benchmarking studies and hence supposed to have the most reliable data, to be consistently higher (or lower) in one version than the other. Johnson et al. (2013) also report similar data inconsistency for GDP growth across versions. Ponomareva and Katayama (2010) find considerable difference in annual mean GDP growth across versions for a given year and country.

In this chapter, we utilize FDA to examine the distribution functions of GDP from four commonly used PWT versions. We model the discrete GDP observations with FDA and construct test statistics for the hypothesis that the distribution functions of GDP are equal in any two PWT versions. The critical values of test statistics are obtained by bootstrap method.

1.2 Modelling GDP Processes Using FDA

Let $Y_{j,v,t}$ be the value of GDP at time t for country j in PWT version v. We consider four versions (namely, 6.3, 7.1, 8.0 and 8.1) over the years $1960 - 2007$ and two groups of countries (namely, 23 OECD countries and 78 non-OECD countries). As such, $j = 1$, ..., N with $N \in \{23, 78\}$, and $v = v63$, $v71$, $v80$, $v81$.

Assuming smoothness in the underlying processes of GDP, denoted $X_{j,v}(t)$, we express $Y_{j,v,t}$ as

$$Y_{j,v,t} = X_{j,v}(t) + \epsilon_{j,v}(t),$$

where $\epsilon_{j,v}(t)$ is a random error with mean zero and finite variance. Given the nature of the data, we use an order-4 B-spline basis to approximate $X_{j,v}(t)$:

$$X_{j,v}(t) \approx C_{j,v,K_v}^T B_{K_v}(t),$$

where K_v is the number of basis functions for version v, $B_{K_v}(t)$ and C_{j,v,K_v} are K_v-vectors of B-spline functions and coefficients, respectively. For a given K_v and regularization parameter λ_v, estimates of the coefficients are obtained by minimizing a penalized sum of

squared residuals,

$$m(\{C_{j,v,\lambda_v,K_v}\}_{j=1}^{N};\lambda_v,K_v) := \frac{1}{N}\sum_{j=1}^{N}\left(\frac{1}{S}\sum_{i=1}^{S}[Y_{j,v,t_i}-X_{j,v}(t_i)]^2+\lambda_v\int_0^1\left(X''_{j,v}(t)\right)^2dt\right),$$

where S denotes the number of observing time points, and $\{t_i\}_{i=1}^{S}$ is the observing time normalized to $[0,1]$ interval. The penalty is determined by the integral of squared second derivatives, and λ_v controls the trade-off between bias and variance in the curve fitting function (Ramsay, 2005).

Solving the first order condition yields estimates $\hat{C}_{j,v,\lambda_v,K_v}$:

$$\hat{C}_{j,v,\lambda_v,K_v}=\left(\frac{1}{S}\sum_{i=1}^{S}\left[B_{K_v}(t_i)B_{K_v}^T(t_i)\right]+\lambda_v\int_0^1\left(B''_{K_v}(t)\right)\left(B''_{K_v}(t)\right)^T dt\right)^{-1}\frac{1}{S}\sum_{i=1}^{S}\left[Y_{j,v,t_i}B_{K_v}(t_i)\right].$$

To this end, we utilize the standard leave-one-out cross-validation to guide our choice of the parameters λ_v and K_v. The optimal pairs, denoted (λ_v^*, K_v^*), are displayed in Table 1.1. The fitting function under $(\lambda_v^*,\ K_v^*)$ is

$$\hat{X}_{j,v}(t)=\hat{C}^T_{j,v,\lambda_v^*,K_v^*}B_{K_v^*}(t);$$

the first and second derivatives are

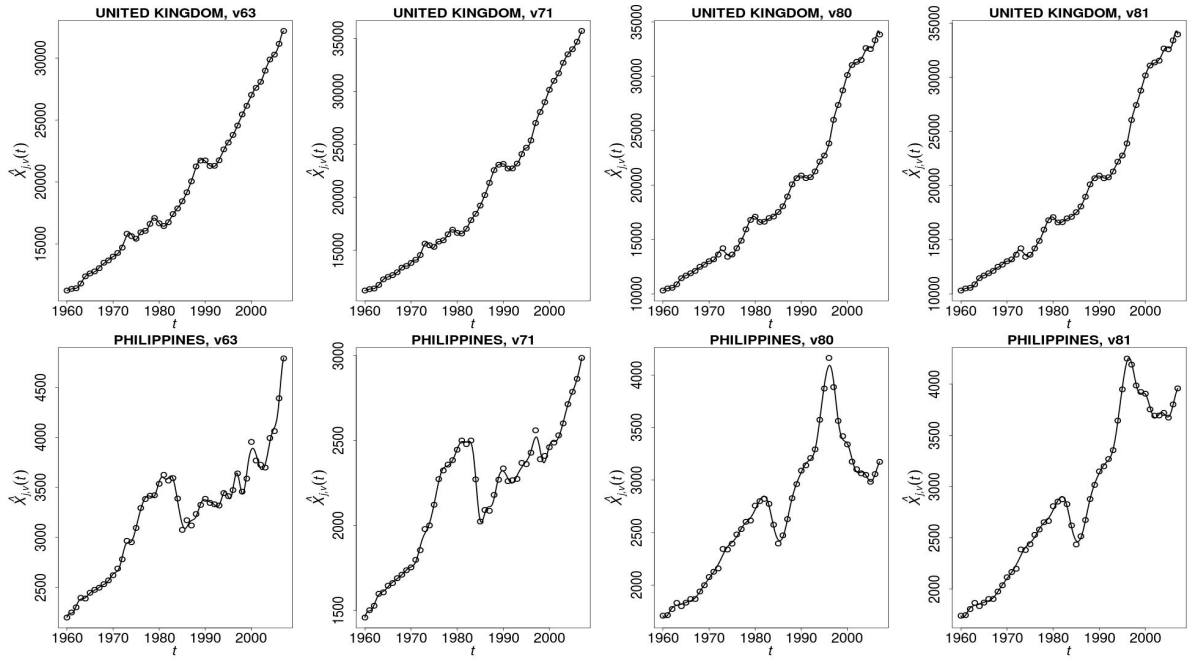
$$\hat{X}'_{j,v}(t)=\hat{C}^T_{j,v,\lambda_v^*,K_v^*}B'_{K_v^*}(t)\quad\text{and}\quad\hat{X}''_{j,v}(t)=\hat{C}^T_{j,v,\lambda_v^*,K_v^*}B''_{K_v^*}(t).$$

Table 1.1: $(\lambda_v^*,\ K_v^*)$

	$v63$	$v71$	$v80$	$v81$
OECD Countries	$(1.00\times10^{-8},\ 41)$	$(1.00\times10^{-8},\ 48)$	$(1.58\times10^{-8},\ 41)$	$(1.58\times10^{-8},\ 41)$
non-OECD Countries	$(3.98\times10^{-8},\ 40)$	$(3.98\times10^{-8},\ 37)$	$(3.98\times10^{-8},\ 30)$	$(3.98\times10^{-8},\ 30)$

To get a view of the GDP function for a country within the OECD and non-OECD groups, Figure 1.1 plots the observations and the fitting functions for the UK and the Philippines, and Figures 1.2 plots the corresponding first and second derivatives.

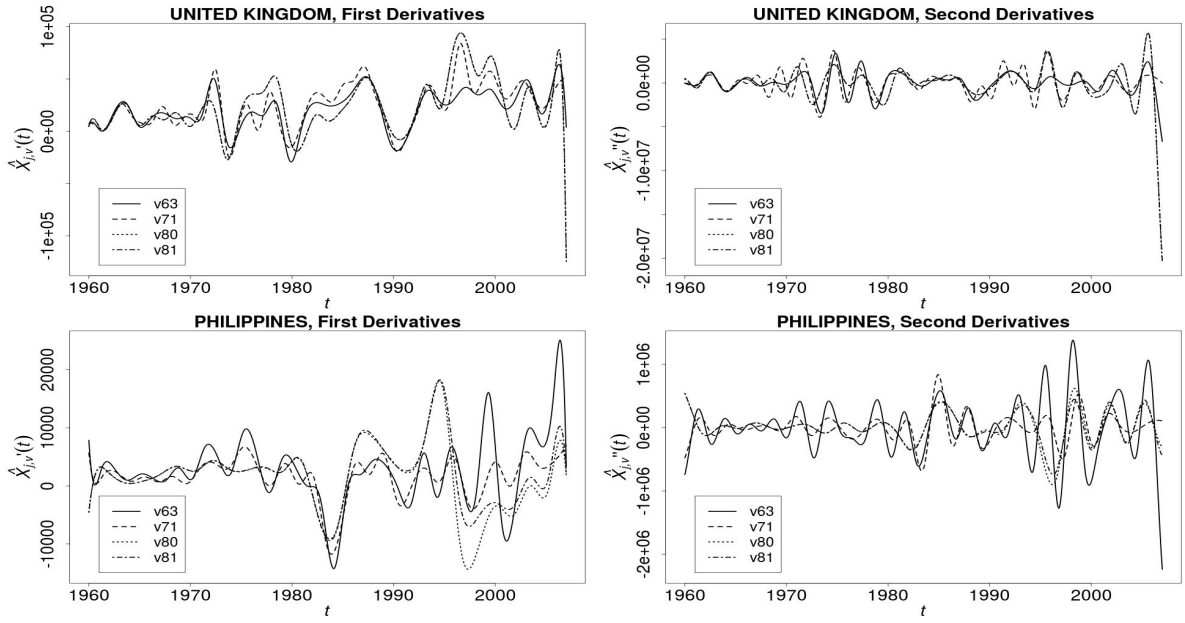
Figure 1.1: Sample countries fitting, level data



1.3 Construction of the Hypothesis Tests

Testing distributional equality of two random vectors is well known in the literature (e.g., the Kolmogorov–Smirnov test); however, to the best of our knowledge, little work has been done on testing distributional equality of random processes. Constructing such a test is one of the contributions of this chapter. The test is motivated by the notion that two random variables share a common distribution if they have the same moment generating functions, provided their existence. We generalize this notion to random processes and construct an asymptotically consistent test of comparing two random processes by verifying the equality of an enlarging set of moment functions of the sample paths at an increasing order of derivatives of the curve fitting functions. The realities of finite sample size, however, imply that we can only test a finite order of moments and derivatives. We next present the test statistics for the first two moments associated with the first three orders of derivatives (level data counts as the zero-order derivative).

Figure 1.2: Sample countries fitting, derivatives



1.3.1 Test statistics and test procedures

Let $G_{j,v}(t)$ be a general representation, such that $G_{j,v}(t) = X_{j,v}(t), X'_{j,v}(t), X''_{j,v}(t)$, and let $\hat{G}_{j,v}(t)$ denote the corresponding estimated function, for all j and v. For a given pair of PWT versions, $v_1, v_2 \in \{v63, v71, v80, v81\}$, the test statistic for testing the first moment equality is constructed as

$$W^{(1)}_{v_1,v_2} := \int_0^1 \left[\frac{1}{N} \sum_{j=1}^{N} \left(\hat{G}_{j,v_1}(t) - \hat{G}_{j,v_2}(t) \right) \right]^2 dt. \tag{1.1}$$

The idea behind this is that when time is fixed at t, the stochastic process reduces to a random variable, and the expression inside the squared bracket in Equation (1.1) represents the difference between the estimated averages of two random variables. It follows that the $W^{(1)}_{v_1,v_2}$ statistic is a functional of this collection (indexed by t) of differences, which in this case is a squared-L_2 norm[4].

[4]For a collection of functions defined on $[0,1]$, the L_2 norm of a function f is defined as $\|f\|_2 = \left(\int_0^1 |f(t)|^2 \, dt \right)^{1/2}$. Since $\int_0^1 |f(t)|^2 \, dt$ to $\left(\int_0^1 |f(t)|^2 \, dt \right)^{1/2}$ is a one-to-one transformation, we can leave out

To construct the bootstrap statistics, we use i.i.d. bootstrap resampling method to the OECD and non-OCED countries, generating B bootstrap samples with N replications, where $N \in \{23, 78\}$ and $B = 1000$. Let $\left\{G^*_{b,j,v}(t)\right\}_{j=1}^{N}$ denote the underlying processes of the b-th set of bootstrap sample and $\left\{\hat{G}^*_{b,j,v}(t)\right\}_{j=1}^{N}$ their corresponding estimated functions. The bootstrap statistic for the first moment is constructed as

$$W^{*(1)}_{b,v_1,v_2} := \int_0^1 \left[\frac{1}{N}\sum_{j=1}^{N}\left(\hat{G}^*_{b,j,v_1}(t) - \hat{G}^*_{b,j,v_2}(t)\right) - \frac{1}{N}\sum_{j=1}^{N}\left(\hat{G}_{j,v_1}(t) - \hat{G}_{j,v_2}(t)\right)\right]^2 dt. \tag{1.2}$$

The null hypothesis for the first moment equality test is

$$H_0 : \mathbb{E}\left[G_{j,v_1}(t)\right] = \mathbb{E}\left[G_{j,v_2}(t)\right],\ v_1,\ v_2 \in \{v63, v71, v80, v81\},\ \text{for all } j \text{ and almost all } t, \tag{1.3}$$

and the alternative is the complement. For a given t, the bootstrap statistic in Equation (1.2) is centered by the bootstrap population mean, whether or not the null hypothesis is true. Under the null, the original statistic is also centered and thus shares the same distribution with the bootstrap statistic, which implies that our test has the exact size. Under the alternative, the original statistic is no longer centered such that our test has power. Proofs of the exact size and power properties of the test are shown in Appendix A.

Similarly, the null hypothesis for the second moment equality test is

$$H_0 :\ \mathrm{Var}\left(G_{j,v_1}(t)\right) = \mathrm{Var}\left(G_{j,v_2}(t)\right),\ v_1,\ v_2 \in \{v63, v71, v80, v81\},\ \text{for all } j \text{ and almost all } t, \tag{1.4}$$

and the corresponding statistics are constructed as

$$W^{(2)}_{v_1,v_2} := \int_0^1 \left(\frac{1}{N}\sum_{j=1}^{N}\left[\left(\hat{G}_{j,v_1}(t) - \bar{G}_{v_1}(t)\right)^2 - \left(\hat{G}_{j,v_2}(t) - \bar{G}_{v_2}(t)\right)^2\right]\right)^2 dt, \tag{1.5}$$

$$\begin{aligned} W^{*(2)}_{b,v_1,v_2} := \int_0^1 \Bigg(&\frac{1}{N}\sum_{j=1}^{N}\left[\left(\hat{G}^*_{b,j,v_1}(t) - \bar{G}^*_{b,v_1}(t)\right)^2 - \left(\hat{G}^*_{b,j,v_2}(t) - \bar{G}^*_{b,v_2}(t)\right)^2\right] \\ &- \frac{1}{N}\sum_{j=1}^{N}\left[\left(\hat{G}_{j,v_1}(t) - \bar{G}_{v_1}(t)\right)^2 - \left(\hat{G}_{j,v_2}(t) - \bar{G}_{v_2}(t)\right)^2\right]\Bigg)^2 dt, \end{aligned} \tag{1.6}$$

the square root in the $W^{(1)}_{v_1,v_2}$ statistic without affecting the test results.

where $\bar{G}_v(t) := N^{-1}\left(\sum_{j=1}^{N}\hat{G}_{j,v}(t)\right)$ and $\bar{G}^*_{b,v}(t) := N^{-1}\left(\sum_{j=1}^{N}\hat{G}^*_{b,j,v}(t)\right)$.

Then one can implement the test as follows:
(i) Estimate $G_{j,v}(t)$ using the functional data approach introduced above, obtaining $\hat{G}_{j,v}(t)$.
(ii) Compute $W^{(m)}_{v_1,v_2}$ using Equations (1.1) and (1.5), where the superscript (m) means the mth moment, for $m = 1, 2$.
(iii) Apply i.i.d. bootstrap resampling on the OECD and non-OECD countries, generating B bootstrap samples with N replications, where $N \in \{23, 78\}$ and $B = 1000$.
(iv) Recall that $\left\{G^*_{b,j,v}(t)\right\}_{j=1}^{N}$ denote the underlying processes of the b-th set of bootstrap sample. Estimate $G^*_{b,j,v}(t)$ for all b, j, v using the functional data approach, obtaining the corresponding estimated function $\hat{G}^*_{b,j,v}(t)$.
(v) Compute the bootstrap test statistics $W^{*(m)}_{b,v_1,v_2}$ using Equations (1.2) and (1.6), where the superscript (m) indicates the mth moment, for $m = 1, 2$.
(vi) Reject the null in (1.3) or (1.4) if $W^{(m)}_{v_1,v_2}$ exceeds the bootstrap statistic $W^{*(m)}_{b,v_1,v_2}$ at a chosen confidence interval.

1.3.2 The asymptotic size and power properties of the test

The proposed test has power and the exact size under the following assumptions. Recall that S denotes the number of time observations, and K and λ are the general representations for the number of basis functions and the smoothing parameter, respectively.

Assumption 1.3.1 *$G_{j,v}(t)$ is four times continuously differentiable.*

Assumption 1.3.1 ensures the underlying function $G_{j,v}(t)$ is smooth up to a certain order.

Assumption 1.3.2 *$N, S \to \infty$, $N/S^{2\gamma} \to 0$ with $0 < \gamma < 4/9$, $K \asymp S^{1/9}$, $\lambda = \mathcal{O}\left(S^{-2/3}\right)$.*

Assumption 1.3.2 sets the divergence rates of N and S as well as the orders of parameters K and λ in terms of S. The intuition behind it is that the asymptotic bias of our functional estimators vanishes as $S \to \infty$, and the optimal convergence rate of these estimators, with the given basis functions and roughness penalty, can be achieved under properly selected K and λ (see, e.g., Claeskens et al., 2009). Since we use order-4 B-spline bases with equal-spaced knots on $[0, 1]$ time interval and the roughness penalties of order-2 derivatives,

$K \asymp S^{1/9}$ and $\lambda = \mathcal{O}\left(S^{-2/3}\right)$ imply an optimal convergence rate as $S \to \infty$ (Claeskens et al., 2009, Theorem 1), where $a \asymp b$ indicates that the ratio a/b is bounded away from zero and infinity. Also, under the assumption of cross-individual independence that will be introduced below, the central limit theorem (CLT) adopts the inflator $\sqrt{N}$, and $N/S^{2\gamma} \to 0$ with $0 < \gamma < 4/9$ is to make sure the functional estimators are converging to the true underlying processes fast enough, so that the estimation error is not inflated by $\sqrt{N}$.

Assumption 1.3.3 *For any given v and almost all t, $\{G_{j,v}(t)\}_{j=1}^{N}$ are independently distributed across j with $\mathbb{E}\left[G_{j,v}(t)\right] = \mu_{G_v}(t)$ and $\mathrm{Var}\left[G_{j,v}(t)\right] = \sigma^2_{G_{j,v}}(t)$. Let $S_{N,G_v}(t) := \sqrt{\sum_{j=1}^{N} \sigma^2_{G_{j,v}}(t)}$, $\exists\ \delta > 0$, such that*

$$\lim_{N\to\infty} \frac{1}{S^{2+\delta}_{N,G_v}(t)} \sum_{j=1}^{N} \mathbb{E}\left[\left|G_{j,v}(t) - \mu_{G_v}(t)\right|^{2+\delta}\right] = 0. \tag{1.7}$$

Assumption 1.3.3 states the conditions required for applying the CLT. Specifically, the CLT we apply requires independent but not necessarily identical distribution of the random variables $\{G_{j,v}(t)\}_{j=1}^{N}$, as long as the condition in Equation (1.7) is satisfied, which indicates that for a given v and almost all t, $\{G_{j,v}(t)\}_{j=1}^{N}$ have some moments of orders higher than 2 with limited growth rates as $N \to \infty$.

We then have the following asymptotics for the statistics:

Theorem 1.3.1 *Under the null hypothesis and Assumptions 1.3.1 – 1.3.3, $\sqrt{N}W^{(m)}_{v_1,v_2} \xrightarrow{d} \int_0^1 \left(\mathbb{G}^{(m)}(t)\right)^2 dt$ and $\sqrt{N}W^{*(m)}_{b,v_1,v_2} \xrightarrow{d} \int_0^1 \left(\mathbb{G}^{(m)}(t)\right)^2 dt$, where $\mathbb{G}^{(m)}(t)$ denotes some function of $G_{j,v}(t)$, for $m = 1, 2$.*

Theorem 1.3.2 *Under the alternatives and Assumptions 1.3.1 – 1.3.3, $\sqrt{N}W^{(m)}_{v_1,v_2} \to \infty$ and $\sqrt{N}W^{*(m)}_{b,v_1,v_2} = \mathcal{O}_p(1)$, for $m = 1, 2$.*

Theorem 1.3.1 states that under the null, $\sqrt{N}W^{*(m)}_{b,v_1,v_2}$ and $\sqrt{N}W^{(m)}_{v_1,v_2}$ have the same limiting distribution, which implies the test has the exact size; Theorem 1.3.2 implies that the power of the test against the alternatives converges to unity.

1.3.3 Do GDP data across versions of PWT follow the same distribution?

Tables 1.2 summarizes the pairwise test results for the null hypotheses of moments equality. For GDP levels, $X_{j,v}(t)$, the null hypotheses of the first and the second moments equality

Table 1.2: Bootstrap Test Results

		OECD Countries			**non-OECD Countries**		
		$v71$	$v80$	$v81$	$v71$	$v80$	$v81$
$X_{j,v}(t)$	$v63$	(**R**, **R**)	(**R**, **R**)	(**R**, **R**)	(**R**, **R**)	(**R**, **R**)	(**R**, **R**)
	$v71$		(**R**, **R**)	(**R**, **R**)		(FR, **R**)	(FR, **R**)
	$v80$			(**R**, **R**)			(**R**, **R**)
$X'_{j,v}(t)$	$v63$	(FR, **R**)	(**R**, **R**)	(**R**, **R**)	(FR, **R**)	(FR, **R**)	(FR, **R**)
	$v71$		(**R**, **R**)	(**R**, **R**)		(FR, **R**)	(FR, **R**)
	$v80$			(FR, **R**)			(FR, **R**)
$X''_{j,v}(t)$	$v63$	(FR, **R**)	(FR, **R**)	(FR, **R**)	(FR, **R**)	(FR, **R**)	(FR, **R**)
	$v71$		(FR, **R**)	(FR, **R**)		(FR, **R**)	(FR, **R**)
	$v80$			(FR, **R**)			(FR, **R**)

Note: **R**=reject H_0, FR=fail to reject H_0; 95% confidence level. The first and second elements in parenthesis indicate tests for the first and second moments.

are rejected for all pairs in both OECD and non-OECD samples, except for the first moment of the non-OECD pairs 7.1-8.0 and 7.1-8.1. For the first and second derivatives of GDP, the first moment equality is not rejected in many pairs but the second moment equality is rejected for all pairs. These results suggest that the distributions of GDP differ significantly across PWT versions. Some caveats are in order, however. First, tests based on FDA are applied not to the discrete GDP observations but rather to their continuous-time approximations (or functional objects) formed using a system of basis function expansion (order-4 B-spline function) and roughness penalty of order-2 derivatives procedure. Second, the conversion of the discrete data into functional objects depends on the number of basis functions K, as well as the smoothing parameter λ that controls the tradeoff between bias and variance in the curve fitting function. While these setups are to some extent necessary to conduct FDA, they should be kept in mind when interpreting the results.

1.4 Concluding Remarks

This chapter utilizes FDA to examine distributional equality of GDP from four PWT versions. Our principal findings provide some evidence supporting the hypothesis that the distribution functions of GDP are different across versions. In this regard, they are consistent and complement the findings of previous studies that, in many countries, the levels and growth rates of GDP for a given year vary substantially in different PWT versions.

Chapter 2

Functional Dynamic Factor Models[1]

2.1 Introduction

Modern macroeconomics data usually consists of hundreds, or even thousands of series covering an increasing time span. Due to the high dimensionality of the data, researchers face challenges not only in empirical analysis, but also in theoretical estimation and inference. Dynamic factor models (DFMs), first proposed by Geweke (1977) and Sargent et al. (1977), offer a powerful tool for the analysis of such data structure by reducing the dimensions and summarizing the co-movements of the series using a few common factors. There is a vast literature on DFMs, such as the studies introduced in survey papers by Bai and Ng (2008), Forni et al. (2000), Breitung and Eickmeier (2006), Reichlin (2003) and Stock and Watson (2006). The first two surveys mainly focus on the key theoretical results for large static factor models and DFMs respectively, while the last three emphasize the empirical applications of the estimated factors.

Specifically, in a DFM, the only observables x_{it} 's are decomposed into a K-vector of latent dynamic factors f_t, a K-vector of loadings λ_i and idiosyncratic disturbances ε_{it}, where i ($= 1, ..., n$) counts cross sections, t ($= 1, ..., T$) indicates the time index, and K is the number of common factors. The much lower K - dimensional vector f_t is assumed to

[1]This chapter is co-authored with Tao Chen and Jiawen Xu from the School of Economics at Shanghai University of Finance and Economics.

govern the co-movements of the whole data set, and its dynamic property is often modeled as a vector autoregression (VAR) process, such that

$$\begin{cases} x_{it} & = \lambda_i^T f_t + \varepsilon_{it} \\ f_t & = \Phi(L) f_{t-1} + \eta_t \end{cases},$$

where $\Phi(L)$ is a lag operator and η_t is a zero-mean random variable independent of the rest. The idiosyncratic disturbances are assumed to be uncorrelated with the factors at all leads and lags and mutually uncorrelated at all leads and lags, which is the usual assumption of the exact factor model of Sargent et al. (1977).

Many researchers, however, have raised the problems regarding parameter instabilities that are concerned with model misspecification and forecasting failure — parameters may change dramatically due to important economic events or financial crisis during the sampling period, while ignoring structural changes in factor loadings may cause misleading results in analysis such as estimating the common factors and assessing the transmission of common shocks to specific variables. In recent years, more and more researchers attempted to take model instabilities into considerations. Banerjee et al. (2008) investigated the consequences of ignoring time variations in the factor loadings for forecasting based on Monte Carlo simulations and found it to worsen the forecasts. Breitung and Eickmeier (2011), BE hereafter, proposed a sup-LM test to detect structural breaks in factor loadings and found evidence that January 1984 (which is usually associated with the beginning of the so called Great Moderation) coincided with a structural break in the factor loadings using a large US macroeconomic dataset provided by Stock and Watson (2005). Improving upon the sup-LM test of BE, Yamamoto and Tanaka (2015) proposed a modified BE test that is robust to the non-monotonic power problem. Empirical application using the U.S. Treasury yield curve data showed that three structural breaks in factor loadings occurred in the sample period from 1985 to 2011.

Apart from testing for structural breaks in factor loadings, some researchers focused on modeling time variations in factor loading parameters. The time varying parameter model has been widely applied in various model specifications to account for parameter instabilities. In these models, time varying parameters are assumed to follow certain stochastic processes. Models incorporating such features showed great potential in improving forecasting performance upon the traditional steady model setup. Xu and Perron (2014) modeled

the return volatility as a random level shift process with mean reversion and varying jump probabilities. Their model provides robust improvements in forecasting compared with many popular models, such as GARCH, ARFIMA, HAR and Regime Switching models, in various return series and multiple forecasting horizons. Xu and Perron (2017) further propose a generalized varying parameter model in which the parameters are assumed to follow a level shift process, and demonstrate that their model can help forecast out-of-sample structural breaks in parameters. This model is also applied to forecast exchange rate volatilities and forecasting gains are achieved over other competing models, see Li et al. (2017). There are still very few papers that directly model factor loadings as time varying processes. Del Negro and Otrok (2008) suggested a time-varying parameter model where the factor loadings are modeled as random walks. Mikkelsen et al. (2019) assume the factor loadings to evolve as stationary VAR and consistent estimates of the loadings parameters can be obtained by a two-step maximum likelihood estimation procedure. Motta et al. (2011) and Su and Wang (2017) consider time varying loadings as smooth evolutions that is purely deterministic, such as $\lambda_{it} = \lambda_i(t/T)$. They simultaneously estimate the factors and the time varying factor loadings via local PCA method; they also provided the limiting distributions of the estimated factors and factor loading under large T and large N framework.

As we summarized above, in the existing literature there are basically four types of models dealing with parameter instability in factor loadings: 1) abrupt structural breaks in loadings (e.g., Breitung and Eickmeier 2011; Yamamoto and Tanaka 2015), 2) smooth changes in factor loadings (e.g., Motta et al. 2011; Su and Wang 2017), 3) VAR factor loadings (e.g., Mikkelsen et al. 2019), and 4) random walk factor loadings (e.g., Del Negro and Otrok 2008). Here, we provide a new perspective to model instabilities of factor loadings using FDA methods. An essential motivation of adopting the functional data idea is to allow for continuous-time analysis — when there exist continuous-time underlying processes beyond the observables, which happens a lot with macroeconomics data, consistent estimation can be achieved from continuous-time analysis but not necessarily from a discrete-time one where the observations are treated as discrete points without taking into account the underlying continuity (e.g., Merton, 1980, 1992; Melino and Sims, 1996; Aït-Sahalia, 2002). There has been literature studying factor models with the idea of functional data. For example, Hays et al. (2012) proposes a functional DFM, where the

co-movements are specified as latent, continuous, nonrandom functions, and the individual-specific effects are constant over the continuum time dimension but follow AR(p) processes over the cross-sectional dimension that in their case is also indexed by time; Jungbacker et al. (2014) impose smoothness on the individual-specific effects applying cubic spline functions; Kokoszka et al. (2014) and Kowal et al. (2017a) model the processes of observations and the latent co-movements as continuous functions over time; Kowal et al. (2017b) model the processes of observations as a functional autoregression with Gaussian innovations, and design a nonparametric factor model for the dynamic innovation process.

In the current chapter, we propose a generalized functional DFM (GFDFM). Specifically, in the spirit of FDA, we view the observable x_{it} as the "snapshot" of the continuous-time underlying process i at time t, denoted by $x_i(t)$'s, and we motivate the GFDFM as $x_i(t) = \sum_{k=1}^{K} \int_0^t \lambda_{ik}(t,s) f_k(s) ds + \varepsilon_i(t)$. The function $f_k(t)$ represents the $k\text{-}th$ factor, and the function $\lambda_{ik}(t,s)$ represents the loading for factor k and individual i — the two time dimensions t and s in the loading function $\lambda_{ik}(t,s)$ capture the current and the past effects of the $k\text{-}th$ factor on $x_i(t)$, respectively. Such a specification generalize the conventional factor models in several aspects. First, the processes are modeled as functional data for the subsequent continuous-time analysis. Meanwhile, from the perspective of DFM, a continuous, and thus infinite-order lag effect is captured by the integration over s; from the perspective of accommodating loading instability, time-varying loading is allowed by including the concurrent time dimension t in the loading functions.

A major contribution of this chapter is that, to our best knowledge, we are the first ones who propose a functional DFM to take account for two-dimensional parameter instability in factor loadings — in previous literature, the DFMs with continuous time-varying loadings that only capture the current effect of the factors (e.g., Su and Wang, 2017) can be viewed as a "concurrent" version of the GFDFM. More specifically, when the past effects of factors are all zero, the loadings reduce to one-dimensional functions, and the GFDFM can be rewritten into a concurrent form: $x_i(t) = \sum_{k=1}^{K} \lambda_{ik}(t) f_k(t) + \varepsilon_i(t)$. Conversely, when the past effects of the factors are not all zero, the GFDFM can capture the effects of the factors on the observed processes, while the concurrent form is not able to do so. Therefore, the GFDFM possesses a time varying property in a more general form.

Furthermore, we provide derivations of the estimators as well as proofs of consistency and normality. There has been literature on generalizing time-invariant coefficients to time-

varying ones in regression analysis (e.g., Hastie and Tibshirani, 1993; Hall and Horowitz, 2007), and the effects of such generalization on the convergence rates of the estimators has also been studied (e.g., Hall and Horowitz, 2007). In the current chapter, we demonstrate that involving the two-dimensional time-varying loadings complicates the estimators and the processes of their convergence, so that the asymptotic normality of the fitted observables can no longer be approached at the standard rate of $\min\left\{\sqrt{N}, \sqrt{T}\right\}$ as shown in literature (e.g., Bai and Ng, 2002) but with a lower speed. We also propose a heuristic bootstrap test in empirical studies to justify the application of the GFDFM by testing the significance of the past-effect-dimension in loadings. Moreover, there has not been a large literature in economics applying FDA (e.g., Chen et al., 2018); hence, this chapter also contributes to the literature by motivating and developing FDA in the study of economics.

2.2 GFDFM Estimators

Recall the GFDFM defined above; we are specifically interested in the following model:

$$x_i(t) = \sum_{k=1}^{K} \int_0^t \lambda_{ik}(t,s) f_k(s) ds + \varepsilon_i(t); \quad i = 1, ..., n, \ t \in [0,1], \tag{2.1}$$

where $\lambda_{ik}(\cdot,\cdot)$'s are non-stochastic loadings, and $f_k(\cdot)$'s are stochastic common factors. Let n denote the number of cross sectional series and K the number of factors. For $i = 1, ..., n$ and $k = 1, ..., K$, $f_k(\cdot)$ represents the *k-th* factor, $\lambda_{ik}(\cdot,\cdot)$ represents the loading for replication i and factor $f_k(\cdot)$, and $x_i(\cdot)$ is the underlying process, from which the data x_{it}'s are drawn at discrete time points.

To analyze the model in Equation (2.1), if either $\lambda_{ik}(\cdot,\cdot)$'s or $f_k(\cdot)$'s have observable realizations, the others can be estimated by solving least squares problems. However, as in conventional factor models, both $\lambda_{ik}(\cdot,\cdot)$'s and $f_k(\cdot)$'s are latent; thus, extra conditions are required to make the model identifiable. In the current chapter, we first estimate the underlying processes $x_i(\cdot)$'s and denote the functional estimator as $\tilde{x}_i(\cdot)$'s, then we estimate the co-movement of $x_i(\cdot)$'s by implementing functional principal component analysis (FPCA) on $\tilde{x}_i(\cdot)$'s and estimate individual-specific time-varying effects of the co-movement on $x_i(\cdot)$'s using functional linear regression. In order to obtain the functional estimators of

$x_i(t)$'s, $f_k(t)$'s and $\lambda_{ik}(t,s)$'s, we express these underlying processes using basis expansions:

$$x_i(t) = \sum_{h=1}^{\infty} c_{i,h}\beta_h(t) \approx \sum_{h=1}^{H} c_{i,h}\beta_h(t) =: \check{x}_i(t), \tag{2.2}$$

$$f_k(t) = \sum_{h=1}^{\infty} a_{k,h}\alpha_h(t) \approx \sum_{h=1}^{H} a_{k,h}\alpha_h(t) =: \check{f}_k(t), \tag{2.3}$$

$$\lambda_{ik}(t,s) = \sum_{p=1}^{\infty}\sum_{q=1}^{\infty} b_{i,k,p,q}\theta_q(t)\psi_p(s) \approx \sum_{p=1}^{P}\sum_{q=1}^{Q} b_{i,k,p,q}\theta_q(t)\psi_p(s) =: \check{\lambda}_{ik}(t,s), \tag{2.4}$$

where $c_{i,h}$'s, $a_{k,h}$'s and $b_{i,k,p,q}$'s denote the expansion coefficients, $\beta_h(\cdot)$'s, $\alpha_h(\cdot)$'s, $\psi_p(\cdot)$'s and $\theta_q(\cdot)$'s denote the expansion bases, and $H, P, Q \in \mathbb{N}$ denote the numbers of basis functions. As H, P and Q increase to infinity, the partial sums in (2.2) - (2.4) converge to $x_i(t)$, $f_k(t)$ and $\lambda_{ik}(t,s)$, respectively, for all s, t; in other words, one can approximate $x_i(t)$, $f_k(t)$ and $\lambda_{ik}(t,s)$ arbitrarily closely by selecting proper H, P and Q.

To explain the estimation procedure, we first re-write the model into a vector form:

$$x_i(t) = \int_0^t \boldsymbol{\lambda}_i^T(t,s)\boldsymbol{f}(s)ds + \varepsilon_i(t)\ \forall i,\ \text{or}\ \boldsymbol{x}(t) = \int_0^t \boldsymbol{\Lambda}(t,s)\boldsymbol{f}(s)ds + \boldsymbol{\varepsilon}(t); \quad t \in [0,1]\,, \tag{2.5}$$

where the superscript T represents matrix transpose, and for $s, t \in [0,1]$,

$$\boldsymbol{x}(t) = \begin{bmatrix} x_1(t) \\ \vdots \\ x_n(t) \end{bmatrix}_{n\times 1}, \boldsymbol{\lambda}_i(t,s) = \begin{bmatrix} \lambda_{i1}(t,s) \\ \vdots \\ \lambda_{iK}(t,s) \end{bmatrix}_{K\times 1}, \boldsymbol{\Lambda}(t,s) = \begin{bmatrix} \boldsymbol{\lambda}_1^T(t,s) \\ \vdots \\ \boldsymbol{\lambda}_n^T(t,s) \end{bmatrix}_{n\times K},$$
$$\boldsymbol{f}(t) = \begin{bmatrix} f_1(t) \\ \vdots \\ f_K(t) \end{bmatrix}_{K\times 1}, \boldsymbol{\varepsilon}(t) = \begin{bmatrix} \varepsilon_1(t) \\ \vdots \\ \varepsilon_n(t) \end{bmatrix}_{n\times 1}.$$

According to expressions (2.2) - (2.4), we have $\boldsymbol{\lambda}_i(t,s) \approx \boldsymbol{\Psi}^T(s)\boldsymbol{\Theta}^T(t)\boldsymbol{b}_i$ and $\boldsymbol{\Lambda}(t,s) \approx \boldsymbol{B}\boldsymbol{\Theta}(t)\boldsymbol{\Psi}(s)$; the notations $\boldsymbol{b}_i$, $\boldsymbol{B}$, $\boldsymbol{\Theta}$ and $\boldsymbol{\Psi}$ are defined in Appendix B. Therefore, we can define $\boldsymbol{\lambda}_i^*(t) := \boldsymbol{\Theta}^T(t)\boldsymbol{b}_i$, $\boldsymbol{\Lambda}^*(t) := \boldsymbol{B}\boldsymbol{\Theta}(t)$ and $\boldsymbol{f}^*(t) := \int_0^t \boldsymbol{\Psi}(s)\boldsymbol{f}(s)ds$, such that Equation (2.5) can be approximated as follow:

$$x_i(t) \approx \boldsymbol{\lambda}_i^{*T}(t)\boldsymbol{f}^*(t) + \varepsilon_i(t)\ \forall i,\ \text{or}\ \boldsymbol{x}(t) \approx \boldsymbol{\Lambda}^*(t)\boldsymbol{f}^*(t) + \boldsymbol{\varepsilon}(t), \quad t \in [0,1]\,. \tag{2.6}$$

Since we first estimate functional data from the observations and then proceed to the eigenanalysis and regression using the fitted functional data, our estimation procedure and results are specifically in terms of the functional data methods we employ. In the current chapter, we use the functional estimators achieved by order-four B-spline bases defined on $[0, 1]$, and the second order derivatives of the fitting functions are adopted as the roughness penalty, which leads to the following penalized sum of squares criterion:

$$m(\{c_{i,h}\}_{i,h}; \gamma_x, H) := \frac{1}{n}\sum_{i=1}^{n}\left[\frac{1}{J}\sum_{j=1}^{J}\left\{x_{it_j} - \sum_{h=1}^{H} c_{i,h}\beta_h(t_j)\right\}^2 + \gamma_x \int_0^1 \left\{\sum_{h=1}^{H} c_{i,h}\beta_h''(t)\right\}^2 dt\right], \tag{2.7}$$

where J denotes the number of observation points, and $\{t_j\}_{j=1}^{J}$ denotes the set of time indices normalized to $[0, 1]$ interval, such that $t_1 = 0$ and $t_J = 1$. As shown in Equation (2.7), this estimation requires two parameters to be determined first — the number of the basis functions H, as well as the smoothing or tuning parameter γ_x. The smoothing parameter γ_x balances the trade-off between the minimization of bias and variance in the fitting functions. The larger the γ_x is, the more penalty is put on roughness, and the smoother the fitting functions becomes while the larger the bias is; on the other hand, the smaller the γ_x is, the less penalty is put on roughness, and the more closely the fitting functions can follow the data points while the larger the variance is. In this chapter, we use the standard leave-one-out cross-validation (CV) method to select the number of basis functions and the smoothing parameters.

The basic idea of using the CV method for parameter selection is to find the pair of parameters (H, γ_x) that jointly optimizes the out-of-sample performance of the fitting functions; i.e., the pair of (H, γ_x) that jointly minimizes a CV criterion. First, we define the estimators for the left out observations x_{it}'s as

$$\tilde{x}^{(-i)}_{i,H,\gamma_x}(t) := \left\{\sum_{h=1}^{H} \tilde{c}^{(-i)}_{i,h,H,\gamma_x}\beta_h(t)\right\}, \ \forall i, \tag{2.8}$$

where the coefficients $\left\{\tilde{c}^{(-i)}_{i,h,H,\gamma_x}\right\}_{i,h}$ are obtained based on the parameters (H, γ_x), omitting the i^{th} observation. The CV criterion can then be defined as a sum of squares

$$\mathrm{CV}(H, \gamma_x) := J^{-1}\sum_{j=1}^{J}\left[n^{-1}\sum_{i=1}^{n}\left\{x_{i,t_j} - \tilde{x}^{(-i)}_{i,H,\gamma_x}(t_j)\right\}^2\right], \tag{2.9}$$

and the optimal H and γ_x, denote (H^*, γ_x^*), can be estimated as

$$(H^*, \gamma_x^*) := \underset{(H,\gamma_x)}{\operatorname{argmin}} \operatorname{CV}(H, \gamma_x). \tag{2.10}$$

with which we can obtain the estimated coefficients $\left\{\tilde{c}_{i,h,H^*,\gamma^*_{x,H^*}}\right\}_{i,h}$ by solving the first order condition of Equation (2.7), and it follows that

$$\tilde{x}_i(t) := \left\{\sum_{h=1}^{H} \tilde{c}_{i,h,H^*,\gamma^*_{x,H^*}} \beta_h(t)\right\}. \tag{2.11}$$

Once we have the fitted functional data, we can now move on to the estimation of the functional factors and loadings.

Note that we can estimate the co-movement of $x_i(\cdot)$'s using the eigenfunction(s) of the sample covariance function $\tilde{v}_n(s,t)$, which is defined as

$$\tilde{v}_n(s,t) := n^{-1} \sum_{i=1}^{n} \tilde{x}_i(s)\tilde{x}_i(t). \tag{2.12}$$

Applying FPCA on $\tilde{v}_n(s,t)$, and let $\hat{\boldsymbol{\rho}}$ be a K-by-K diagonal matrix of the largest K eigenvalues in descending order and $\hat{\boldsymbol{f}}^*(\cdot)$ the K corresponding eigenfunctions, we then have (Ramsay, 2005):

$$\int_0^1 \hat{\boldsymbol{f}}^*(s)\tilde{v}_n(s,t)ds = \hat{\boldsymbol{\rho}}\hat{\boldsymbol{f}}^*(t), \tag{2.13}$$

where $\hat{\boldsymbol{f}}^*(\cdot)$ captures the co-movement of the processes $x_i(\cdot)$'s. However, $\hat{\boldsymbol{f}}^*(\cdot)$ does not directly correspond to the factor $\boldsymbol{f}(\cdot)$ but to $\boldsymbol{f}^*(\cdot)$. Here we define the estimator for $\boldsymbol{f}(\cdot)$, denoted $\hat{\boldsymbol{f}}(\cdot)$, as

$$\hat{\boldsymbol{f}}(t) := \frac{\partial \hat{\boldsymbol{f}}^*(t)}{\partial t}. \tag{2.14}$$

Once we have $\hat{\boldsymbol{f}}(\cdot)$, the expansion coefficients $\boldsymbol{b}_i$'s (or $\boldsymbol{B}$), and thus the loadings, can then be estimated by regressing $\tilde{x}_i(t)$ on $\boldsymbol{\Theta}(t)\int_0^t \boldsymbol{\Psi}(s)\hat{\boldsymbol{f}}(s)ds$ for each i, which leads to the penalized least squares estimators

$$\hat{\boldsymbol{b}}_i = \hat{\boldsymbol{R}}_\lambda^{-1} \int_0^1 \hat{\boldsymbol{\Omega}}_\lambda^T(t)\tilde{x}_i(t)dt,\ \hat{\boldsymbol{\lambda}}_i^*(t) = \boldsymbol{\Theta}^T(t)\hat{\boldsymbol{b}}_i, \text{ and } \hat{\boldsymbol{\lambda}}_i(t,s) = \boldsymbol{\Psi}^T(s)\boldsymbol{\Theta}^T(t)\hat{\boldsymbol{b}}_i,$$

where

$$\hat{\boldsymbol{\Omega}}_\lambda(t) := \int_0^t \hat{\boldsymbol{f}}^T(s)\boldsymbol{\Psi}^T(s)ds\boldsymbol{\Theta}^T(t) \text{ and } \hat{\boldsymbol{R}}_\lambda := \int_0^1 \left\{\hat{\boldsymbol{\Omega}}_\lambda^T(t)\hat{\boldsymbol{\Omega}}_\lambda(t)\right\} dt + \gamma_\lambda \int_0^1 \boldsymbol{\Theta}''(t)\boldsymbol{\Theta}''^T(t)dt.$$

Hence,

$$\hat{x}_i(t) = \hat{\boldsymbol{b}}_i^T\boldsymbol{\Theta}(t)\int_0^t \boldsymbol{\Psi}(s)\hat{\boldsymbol{f}}(s)ds.$$

2.3 Large Sample Theories

Now we establish the large sample properties of our functional estimators. We provide theorems to show that our estimators are consistent and asymptotically normal. However, since the true factors and loadings are not completely identifiable, their estimators can only recover some transformed underlying processes, as opposed to those underlying processes themselves. Hence, to investigate the properties of the estimators, instead of comparing the estimators with the underlying processes directly, we compare them after some transformation.

It is important to note that we have been taking the number of factors K as given in our estimation, but in practice K is unknown and needs to be estimated. The estimation of factor numbers has been studied in literature (e.g., Bai and Ng, 2002, 2007; Hallin and Liška, 2007), and one option is to utilize the idea of Bai and Ng (2002) information criteria, which can consistently estimate the number of static factors, say K_S, when K_S is finite. Since a DFM with a finite factor number K_D and a finite lag order K_l can be written as a static factor model with the factor number $K_S = K_D\,(K_l + 1)$ by treating each lag as a separate factor (e.g., Bai and Ng, 2007, 2008), the Bai and Ng (2002) information criteria can also be used to estimated the number K_S in DFMs. Our model, as explained previously, contains finitely many common factors but infinite-order lags, and we can adopt the idea of Bai and Ng (2002) information criteria with some adjustment to our setting, such as

$$\begin{aligned}\hat{K} &:= \underset{K^0>0}{\operatorname{argmin}}\, PC(K^0),\\ PC(K^0) &= \min_{\boldsymbol{\lambda}_i^{K^0}} \frac{1}{n}\sum_{i=1}^n \int_0^1 \left[\tilde{x}_i(t) - \int_0^t \boldsymbol{\lambda}_i^{K^{0^T}}(t,s)\hat{\boldsymbol{f}}^{K^0}(s)\right]^2 dt + K^0 g(N,J), \qquad (2.15)\end{aligned}$$

where $\hat{K}$ represents the estimated K, $\boldsymbol{\lambda}_i^{K^0}(t,s)$ and $\hat{\boldsymbol{f}}^{K^0}(s)$ indicate the loadings and the factors when the number of factors is K^0, and the function $g(N,J)$ needs to satisfy some proper order conditions. However, in the current paper, the large sample properties with $\hat{K}$ is not covered; instead, we will focus on those with K.

Before getting to the asymptotic theorems, we first make the following assumptions.

Assumption 2.3.1 *$n, J \to \infty$, $n \in o\left(J^{8/9}\right)$; $H, Q \asymp J^{1/9}$; $\gamma_x \in \mathcal{O}\left(J^{-2/3}\right)$, $\gamma_\lambda \to 0$.*

Assumption 2.3.1 sets the divergence rates of n and J as well as the orders of parameters H, Q, γ_x and γ_λ in terms of n and J. Essentially, under proper regularization conditions, the estimation errors of $\hat{\boldsymbol{f}}(\cdot)$, $\hat{\boldsymbol{\lambda}}_i(\cdot,\cdot)$ and $\hat{x}_i(\cdot)$ vanish as $n, J, H, Q \to \infty$. In the first step of estimation when we fit the functional data, the optimal convergence rate of $\tilde{x}_i(t)$'s with given basis functions and roughness penalty can be achieved under properly selected H and γ_x (e.g., Claeskens et al., 2009). Since we use order-four B-spline bases defined on $[0,1]$ time interval and the roughness penalties of order-two derivatives, $H \asymp J^{1/9}$ and $\gamma_x \in \mathcal{O}\left(J^{-2/3}\right)$ imply an optimal convergence rate as $J \to \infty$ (Claeskens et al., 2009, Theorem 1). Based on the fitted functional data, the estimation of the co-movements produces an estimation error of order $\mathcal{O}_p\left(n^{-1/2}\right)$ during the process of FPCA, on top of which the estimated loading has a convergence rate determined by Q and γ_λ jointly — as $\gamma_\lambda \to 0$ and $Q \to \infty$, $\hat{\boldsymbol{\lambda}}_i^*(t)$ converges to a "rotated" $\boldsymbol{\lambda}_i^*(t)$ for all t, such that $\hat{\boldsymbol{\lambda}}_i^{*T}(t) \int_0^t \boldsymbol{\Psi}(s)\hat{\boldsymbol{f}}(s)ds$ (or $\hat{x}_i(t)$) converges to $\boldsymbol{\lambda}_i^{*T}(t)\boldsymbol{f}^*(t)$. Therefore, among the conditions in Assumption 2.3.1, $n, J, H, Q \to \infty$, $H, Q \in \mathcal{O}(J)$ and $\gamma_x, \gamma_\lambda \to 0$ suffice the consistency.

Normality, however, requires stronger restrictions on the orders of parameters. We consider the case where the number of time observations grows faster than the number of replications. The leading term of the error for the estimated co-movements will then be the terms whose speeds of vanishing depend on the divergence rate of n, and for that reason, we use $\sqrt{n}$ as the inflator for the derivation of normality. More specifically, $n \in o\left(J^{8/9}\right)$ is to make sure that the functional estimators of the underlying processes are converging to the true underlying processes fast enough, under the optimal convergence rate, so that the estimation error is not inflated by $\sqrt{n}$. On the other hand, the error of $\int_0^t \hat{\boldsymbol{\lambda}}_i^T(t,s)\hat{\boldsymbol{f}}(s)ds$ is of order $\mathcal{O}_p\left(n^{-1/2}\right)$ under the conditions given in Assumption 2.3.1; however, inflating this estimation error with $\sqrt{n}$ does not guarantee normality but only a $\mathcal{O}_p(1)$, due to the interaction among all the terms of order $\mathcal{O}_p\left(n^{-1/2}\right)$. Hence, one way to obtain normality

is by replacing the integrals in the estimator $\int_0^t \hat{\boldsymbol{\lambda}}_i^T(t,s)\hat{\boldsymbol{f}}(s)ds$ with Riemann sums using a parameter of order $o(n)$, and inflating the error terms with $o(\sqrt{n})$, so that the $\mathcal{O}_p\left(n^{-1/2}\right)$'s will not be inflated. The details will be shown in the proofs.

Assumption 2.3.2 *For all i, there exists a polynomial approximation to the continuous underlying processes $x_i(t)$, say $\check{x}_i(t)$, such that $\check{x}_i(t)$ is four times continuously differentiable.*

Assumption 2.3.2 guarantees that the underlying function $x_i(t)$ has an approximation that is smooth up to a certain order, so that we can get a consistent functional estimator with a desired optimal convergence rate (Claeskens et al., 2009, Theorem 1).

Assumption 2.3.3 *$\boldsymbol{\mu}_f(t)$ and $\boldsymbol{\sigma}_f(t)$ are two absolute continuous functions, such that $\mathbb{E}\left\{\boldsymbol{f}^T(t)\right\} = \boldsymbol{\mu}_f(t)$, and $\text{Var}\left\{\boldsymbol{f}^T(t)\right\} = \boldsymbol{\sigma}_f(t)$.*

Assumption 2.3.3 is saying that the factor does not have to have constant mean or variance over time, but only need to have the mean and variance functions that are absolute continuous, so that we can obtain an upper bound of the convergence rate while performing some integration transformation. In other words, the factors do not need to be stationary.

Assumption 2.3.4
a. There exists $C \in \mathbb{R}$, such that $\int_0^t \lambda_{ik}^2(t,s)ds < C$ and $\int_0^s \lambda_{ik}^2(t,s)dt < C$, for all i and k;
b. $n^{-1}\boldsymbol{\Lambda}^T(t,s)\boldsymbol{\Lambda}(t',s') = \boldsymbol{\Sigma}_{\Lambda,1}(t,s,t',s') + \mathcal{O}\left(n^{-1/2}\right)$ and $n^{-1}\boldsymbol{\Lambda}^{*T}(t)\boldsymbol{\Lambda}^*(s) = \boldsymbol{\Sigma}_{\Lambda,2}(t,s) + \mathcal{O}\left(n^{-1/2}\right)$ for $PK \times PK$ matrix functions $\boldsymbol{\Sigma}_{\Lambda,1}(t,s,t',s')$ and $\boldsymbol{\Sigma}_{\Lambda,2}(t,s)$, for $P, K \in \mathbb{N}$.

Assumption 2.3.4 provides some boundedness constraints for the loadings.

Assumption 2.3.5
a. $\mathbb{E}\{\varepsilon_i(t)\} = 0$, for all i and t;
b. $\max_{s,t}\mathbb{E}\left[n^{-1}\left\|\boldsymbol{\varepsilon}^T(s)\boldsymbol{\varepsilon}(t)\right\|^2\right] = o(1)$;
c. $\max_{s,t}\mathbb{E}\left[n^{-1}\left\|\boldsymbol{\Lambda}^{*T}(s)\boldsymbol{\varepsilon}(t)\right\|^2\right] = \mathcal{O}(1)$;
d. $n^{-1/2}\sum_{i=1}^n \boldsymbol{\mu}_i\varepsilon_i(t) \xrightarrow{d} N(\mathbf{0}, \boldsymbol{\sigma}(t))$ for some non-stochastic real valued K-vector $\boldsymbol{\mu}_i \in \mathbb{R}^K$, where $\mathbf{0} \in \mathbb{R}^K$ and $\boldsymbol{\sigma}(t) := \lim_{n\to\infty} n^{-1}\sum_{i=1}^n\sum_{j=1}^n \boldsymbol{\mu}_i\mathbb{E}\{\varepsilon_i(t)\varepsilon_j(t)\}\boldsymbol{\mu}_j^T < \infty$ and $K \in \mathbb{N}$.

Assumption 2.3.5 states the zero-mean and the weak dependence constraints on the error term through moment conditions and weak convergence. This assumption allows for weak dependence in the error term across individuals (as in part c) and over time (in part b). Specifically, for part c, consider the case with $K = 1$ without loss of generality. Then

$$\mathbb{E}\left[n^{-1}\left\|\mathbf{\Lambda}^{*T}(s)\boldsymbol{\varepsilon}(t)\right\|^2\right] = \mathbb{E}\left[n^{-1}\left(\sum_{i=1}^{n}\lambda_i^*(s)\varepsilon_i(t)\right)^2\right] = \mathbb{E}\left[n^{-1}\sum_{i=1}^{n}\sum_{j=1}^{n}\lambda_i^*(s)\varepsilon_i(t)\lambda_j^*(s)\varepsilon_j(t)\right],$$

Since the loading is non-stochastic, it follows that

$$\mathbb{E}\left[n^{-1}\sum_{i=1}^{n}\sum_{j=1}^{n}\left(\lambda_i^*(s)\varepsilon_i(t)\right)\left(\lambda_j^*(s)\varepsilon_j(t)\right)\right] = n^{-1}\sum_{i=1}^{n}\sum_{j=1}^{n}\lambda_i^*(s)\lambda_j^*(s)\mathbb{E}\left[\varepsilon_i(t)\varepsilon_j(t)\right].$$

Therefore, when there is minimum cross sectional dependence, i.e., $\mathbb{E}\left[\varepsilon_i(t)\varepsilon_j(t)\right] = 0$ for $i \neq j$, we have $\max_{s,t} n^{-1}\sum_{i=1}^{n}\lambda_i^{*2}(s)\mathbb{E}\left[\varepsilon_i^2(t)\right]$, implied by a finite variance of $\varepsilon_i(t)$ and well-behaved loading functions. The maximum cross sectional dependence allowed will be for $\mathbb{E}\left[\varepsilon_i(t)\varepsilon_j(t)\right]$ for $i \neq j$ small enough such that $\max_{s,t} n^{-1}\sum_{i=1}^{n}\sum_{j=1}^{n}\lambda_i^*(s)\lambda_j^*(s)\mathbb{E}\left[\varepsilon_i(t)\varepsilon_j(t)\right]$ is still $\mathcal{O}(1)$ under well-behaved loading functions. Similarly for part b,

$$\max_{s,t}\mathbb{E}\left[n^{-1}\left\|\boldsymbol{\varepsilon}^T(s)\boldsymbol{\varepsilon}(t)\right\|^2\right] = n^{-1}\sum_{i}\sum_{j}\mathbb{E}\left[\varepsilon_i(s)\varepsilon_i(t)\varepsilon_j(s)\varepsilon_j(t)\right] = o(1),$$

which, together with the cross sectional dependence limited by part c, imposes the constraint on the correlation over time in $\varepsilon_i(t)$. Part d is the continuous-time-indexed versions of Assumptions A.2(i) in Su and Wang (2017) but in terms of our model setup. The term $\boldsymbol{\mu}_i$ in part d is defined in Lemma B.2 from Appendix B.

Assumption 2.3.6

a. $\boldsymbol{f}(t)$ *and* $\varepsilon_i(t)$ *are orthogonal;*
b. $\int_0^t \boldsymbol{\lambda}_i^T(t,s)\boldsymbol{f}(s)ds$ *and* $\varepsilon_i(\cdot)$ *are orthogonal;*
c. $\int_0^t \boldsymbol{\lambda}_i^T(t,s)\boldsymbol{f}(s)ds$ *is a strong mixing process over* t*, for all* i*.*

Assumption 2.3.6.a and b guarantee that the signals in $x_i(\cdot)$'s can be properly separated from the noise and that there will not be endogeneity problems when we perform the functional linear regression to estimate the loadings. Assumption 2.3.6.c constrains the serial dependence of the process $\int_0^t \boldsymbol{\lambda}_i^T(t,s)\boldsymbol{f}(s)ds$, and it is useful for the application of the CLT for strong mixing processes.

2.3.1 Consistency

Now we present the theorems for consistency.

Theorem 2.3.1 *Under Assumptions 2.3.1 to 2.3.6 and the true number of factors K, there exists an invertible operator $\boldsymbol{W}$ (specified in Appendix B), such that as $n, J \to \infty$, the followings hold:*
a. $\left\| \hat{\boldsymbol{f}}^*(t) - (\boldsymbol{W}\boldsymbol{f})(t) \right\| \xrightarrow{p} 0$, *for* $t \in [0,1]$;
b. $\left\| \int_0^t \hat{\boldsymbol{\lambda}}_i^T(t,s)\hat{\boldsymbol{f}}(s)ds - \int_0^t \boldsymbol{\lambda}_i^T(t,s)\boldsymbol{f}(s)ds \right\| \xrightarrow{p} 0, \ \forall i = 1, ..., n, \ t \in [0,1]$.

Recall that there are mainly two components in the estimation procedure — FPCA for identifying co-movements and functional linear regression for predicting the data series. Essentially, in the process of FPCA, we expect to see the co-movements can be identified, in that the estimated functional principal components converge to the transformed true factors under some invertible operator; also, in the process of functional linear regression, we expect to see the estimated factors generated from the functional principal components can contribute in the prediction of $x_i(\cdot)$ as if the true common factors were observed, and the resulting estimator $\hat{x}_i(\cdot)$ performs reasonably well.

Theorem 2.3.1.a indicates that under the necessary assumptions, the estimated functional principal components $\hat{\boldsymbol{f}}^*(t)$ converges to the transformed true factors $(\boldsymbol{W}\boldsymbol{f})(t)$ under some invertible operator $\boldsymbol{W}$. Theorem 2.3.1.b just shows that the underlying process $\int_0^t \boldsymbol{\lambda}_i^T(t,s)\boldsymbol{f}(s)ds$ can be consistently estimated by the estimators $\hat{\boldsymbol{\lambda}}_i^T(t,s)$ and $\hat{\boldsymbol{f}}(t)$, where the factors $\hat{\boldsymbol{f}}(t)$ are generated from the principal components $\hat{\boldsymbol{f}}^*(t)$ as shown in (2.14), and the loadings $\hat{\boldsymbol{\lambda}}_i^T(t,s)$ are obtained from the functional linear regression.

2.3.2 Asymptotic normality

We also have normality for the estimators.

Theorem 2.3.2 *Under Assumptions 2.3.1 to 2.3.6 and the true number of factors K, there exists an invertible operator $\boldsymbol{W}$, such that as $n, J, S \to \infty$ and $S = o\left(\min\left\{n, \gamma_\lambda^{-2}\right\}\right)$, the followings hold:*

a. $\sqrt{n}\left\{\hat{\boldsymbol{f}}^*(t) - (\boldsymbol{W}\boldsymbol{f})(t)\right\} \xrightarrow{d} N\left(0, \boldsymbol{\rho}^{-1}\Sigma_f(t)\boldsymbol{\rho}^{-1}\right)$*;*
b. $\sqrt{S}\left\{\int_0^t \hat{\boldsymbol{\lambda}}_i^T(t,s)\hat{\boldsymbol{f}}(s)ds - \int_0^t \boldsymbol{\lambda}_i^T(t,s)\boldsymbol{f}(s)ds\right\} \xrightarrow{d} N\left(0, \boldsymbol{\Omega}_\lambda(t)\boldsymbol{R}_\lambda^{-1}\Sigma_{\lambda_i,f}\boldsymbol{R}_\lambda^{-T}\boldsymbol{\Omega}_\lambda^T(t)\right)$.
($\boldsymbol{W}$*,* $\Sigma_f(t)$*,* $\Sigma_{\lambda_i,f}$*,* $\boldsymbol{\Omega}_\lambda(t)$ *and* $\boldsymbol{R}_\lambda$ *are specified in Appendix B.)*

For Theorem 2.3.2.a, recall that $\hat{\boldsymbol{f}}^*(t)$ are the functional principal components derived based on the estimates $\tilde{x}_i(t)$'s, which include both the signal $\int_0^t \boldsymbol{\lambda}_i^T(t,s)\boldsymbol{f}(s)ds$ and the idiosyncratic error $\varepsilon_i(t)$'s, while there exists some invertible operator $\boldsymbol{W}$ such that the transformed true factors $(\boldsymbol{W}\boldsymbol{f})(t)$ can be defined as the functional principal components based on the signal $\int_0^t \boldsymbol{\lambda}_i^T(t,s)\boldsymbol{f}(s)ds$ only. After subtracting $(\boldsymbol{W}\boldsymbol{f})(t)$ from $\hat{\boldsymbol{f}}^*(t)$, it is the part consisting of interaction with the errors $\varepsilon_i(t)$'s that remains, which is also the part that leads to the normality given Assumption 2.3.5.c.

As for Theorem 2.3.2.b, the statement is for each i, and the asymptotic properties are achieved by enlarging the sample size in the continuum dimension. However, the estimated factor $\hat{\boldsymbol{f}}(t)$ carries some terms from FPCA with the convergence rates $\mathcal{O}_p\left(n^{-1/2}\right)$, and as explained previously, due to the interaction among these $\mathcal{O}_p\left(n^{-1/2}\right)$ terms, inflating them by $\sqrt{n}$ does not guarantee normality. Instead, we approximate the integrals of the estimator $\int_0^t \hat{\boldsymbol{\lambda}}_i^T(t,s)\hat{\boldsymbol{f}}(s)ds$ by Riemann sums with S terms, where $S = o(n)$, and we use the inflator $\sqrt{S}$ to obtain normality. The normality is then driven by the errors $\varepsilon_i(t)$'s as well as the interaction between the true factors and the errors given their low correlation along the time dimension under the divergence of S, which is slow enough, so that other sources of randomness will vanish before being caught.

Here we briefly justify our theorems in words, and the mathematical proofs of the theorems can be found in Appendix B. There are four main statements to prove — consistency and asymptotic normality for $\hat{\boldsymbol{f}}^*(t)$ as well as $\int_0^t \hat{\boldsymbol{\lambda}}_i^T(t,s)\hat{\boldsymbol{f}}(s)ds$. The method of adding and subtracting terms is used to decompose the estimation errors $\hat{\boldsymbol{f}}^*(t) - (\boldsymbol{W}\boldsymbol{f})(t)$ and $\int_0^t \hat{\boldsymbol{\lambda}}_i^T(t,s)\hat{\boldsymbol{f}}(s)ds - \int_0^t \boldsymbol{\lambda}_i^T(t,s)\boldsymbol{f}(s)ds$. With the decompositions, we show that the main sources of errors generally lie in four types: the residuals from fitting the functional data, the errors of Riemann sum approximations to integrals, the remainders from the convergence of eigenfunctions, and the interaction involving the idiosyncratic errors. We obtain the orders of the first three sources of errors from literature, and we derive the limiting behavior of the last source of error based on our assumptions. In the proofs for consistency, we demonstrate that these sources of errors are $o_p(1)$ or $o(1)$, while in the proofs for

asymptotic normality, we further investigate their convergence rates.

2.4 Simulation Analysis

We now examine the performance of our functional estimators through simulations.

First, we generate observables x_{i,t_j}'s. Generating x_{i,t_j}'s requires the underlying processes of $\lambda_{ik}(t,s)$'s, $f_k(t)$'s and $\varepsilon_i(t)$'s for all i and k, and in the current simulation, we set $K = 1$. Recall that factor loadings, $\lambda_{ik}(t,s)$'s, are non-random functions; in the current simulation, we define $\lambda_{ik}(t,s)$'s to be local polynomials of order five on both dimensions. Specifically, we generate the coefficient matrix $\mathbf{B}$ by filling the entries with random draws from $N(1,1)$; also, we define the basis for the first dimension as an order-five B-spline containing 20 basis functions, and we define the basis for the second dimension as an order-five B-spline containing 10 basis functions. For the stochastic processes $f_k(t)$'s and $\varepsilon_i(t)$'s, we set them as continuous-time AR(1) processes, which can be written in the following differential form,

$$dz(t) = -\kappa_z z(t)dt + \sigma_z dB(t), \tag{2.16}$$

where $z(t)$ is a general representation for $f_k(t)$'s and $\varepsilon_i(t)$'s, κ_z and σ_z are the parameters for the corresponding process, and $B(t)$ is a standard Brownian motion, which follows that $dB(t)$ denotes the increments of the standard Brownian motion. In the current simulation, we set $\sigma_f, \sigma_\varepsilon = 1$ and set $\kappa_\varepsilon = 1/dt$ for the corresponding discretized model so that $\varepsilon_i(t) = dB(t)$. After generating $\int_0^t \boldsymbol{\lambda}_i^T(t,s)\boldsymbol{f}(s)ds$, we adjust the size of $\varepsilon_i(t)$ relative to $\int_0^t \boldsymbol{\lambda}_i^T(t,s)\boldsymbol{f}(s)ds$, so that comparing with the noise, the signal is not too weak to be identified. To reduce the notation load, we still use $\varepsilon_i(t)$ to represent the rescaled error, and in the current simulation, we rescale the error term such that its standard deviation is 10% as much as the standard deviation of the signal $\int_0^t \boldsymbol{\lambda}_i^T(t,s)\boldsymbol{f}(s)ds$. Finally, the observations x_{i,t_j}'s can be generated as follow:

$$x_{i,t_j} = x_i(t_j) = \int_0^{t_j} \boldsymbol{\lambda}_i^T(t_j,s)\boldsymbol{f}(s)ds + \varepsilon_i(t_j); \quad i = 1,...,n, j = 1,...,J. \tag{2.17}$$

We simulate 199 data sets with sample sizes $J \times n$. The loading functions, as well as the parameters of factor functions are fixed for all simulations.

Once the data is obtained, we derive the estimators $\hat{\boldsymbol{f}}(\cdot)$, $\hat{\boldsymbol{\lambda}}_i(\cdot,\cdot)$ and $\hat{x}_i(\cdot)$ following the procedure introduced above, and we summarize the estimation results by the following two statistics:

$$R_f^2 = \frac{\mathbb{E}\left[\int_0^1 \left\{\hat{\boldsymbol{f}}^*(t) - (\boldsymbol{W}\boldsymbol{f})(t)\right\}^T \left\{\hat{\boldsymbol{f}}^*(t) - (\boldsymbol{W}\boldsymbol{f})(t)\right\} dt\right]}{\mathbb{E}\left\{\int_0^1 \hat{\boldsymbol{f}}^{*T}(t)\hat{\boldsymbol{f}}^*(t)dt\right\}}, \tag{2.18}$$

$$R_x^2 = \frac{\mathbb{E}\left[\int_0^1 \left\{\int_0^t \hat{\boldsymbol{\Lambda}}(t,s)\hat{\boldsymbol{f}}(s)ds - \int_0^t \boldsymbol{\Lambda}(t,s)\boldsymbol{f}(s)ds\right\}^T \left\{\int_0^t \hat{\boldsymbol{\Lambda}}(t,s)\hat{\boldsymbol{f}}(s)ds - \int_0^t \boldsymbol{\Lambda}(t,s)\boldsymbol{f}(s)ds\right\} dt\right]}{\mathbb{E}\left[\int_0^1 \left\{\int_0^t \hat{\boldsymbol{\Lambda}}(t,s)\hat{\boldsymbol{f}}(s)ds\right\}^T \left\{\int_0^t \hat{\boldsymbol{\Lambda}}(t,s)\hat{\boldsymbol{f}}(s)ds\right\}^T dt\right]}. \tag{2.19}$$

The two statistics illustrate the results in Theorem 2.3.1 by measuring the relative average sizes of the estimation errors and the estimators for $\hat{\boldsymbol{f}}^*(t)$ and $\int_0^t \hat{\boldsymbol{\Lambda}}(t,s)\hat{\boldsymbol{f}}(s)ds$, respectively. For example, by Theorem 2.3.1.a, $\mathbb{E}\left[\int_0^1 \left\{\hat{\boldsymbol{f}}^*(t) - (\boldsymbol{W}\boldsymbol{f})(t)\right\}^T \left\{\hat{\boldsymbol{f}}^*(t) - (\boldsymbol{W}\boldsymbol{f})(t)\right\} dt\right]$ converges to zero, and to observe such convergence, we use the same measure for the size of $\hat{\boldsymbol{f}}^*(t)$ as a reference, i.e. $\mathbb{E}\left\{\int_0^1 \hat{\boldsymbol{f}}^{*T}(t)\hat{\boldsymbol{f}}^*(t)dt\right\}$; hence, we expect the size of the errors relative to the size of the estimators, R_f^2, vanishes as $n, J \to \infty$. The same idea applies to the construction of R_x^2.

Specifically, we generate the data using a discretized version of Equation (2.16) with $dt \approx 1/501$, and we check four different sample sizes and three different κ_f values. The results of the two statistics are shown in Table 2.1. We can see that as the sample size increases, both R_f^2 and R_x^2 are getting closer to zero in general. Another interesting result is that as κ_f gets smaller, which indicates the lag effects get larger, we can also see a decreasing trend in both R_f^2 and R_x^2. One explanation is that under the current DGP, the errors have zero lag effects, so stronger lag effects in the underlying signals can help to distinguish the signals from the errors, and thus, makes the estimation more accurate. Figures 2.1 and 2.2 show some examples of the comparison between the estimates and the transformed true processes when $\kappa_f = 400$, and we can see that the fitting is getting more accurate as the sample size increases.

To check the normality, we perform a K-S test, comparing $\hat{\boldsymbol{f}}^*(t) - (\boldsymbol{W}\boldsymbol{f})(t)$ and $\int_0^t \hat{\boldsymbol{\lambda}}_i^T(t,s)\hat{\boldsymbol{f}}(s)ds - \int_0^t \boldsymbol{\lambda}_i^T(t,s)\boldsymbol{f}(s)ds$ respectively with the normal distributions of their

Table 2.1: Consistency

		R_f^2			R_x^2		
J	n	$\kappa_f = 400$	$\kappa_f = 250$	$\kappa_f = 100$	$\kappa_f = 400$	$\kappa_f = 250$	$\kappa_f = 100$
25	20	0.0517	0.0446	0.0239	0.0348	0.0314	0.0238
51	30	0.0227	0.0163	0.0065	0.0194	0.0168	0.0127
101	55	0.0095	0.0052	0.0021	0.0138	0.0119	0.0100
201	100	0.0034	0.0015	0.0007	0.0105	0.0095	0.0082

Note: J indicates the number of time observations, n the number of individuals, κ_f the coefficient of the continuous-time AR(1) process, and R^2's the measurements of the goodness of fit for $f_k(t)$ and $x_i(t)$.

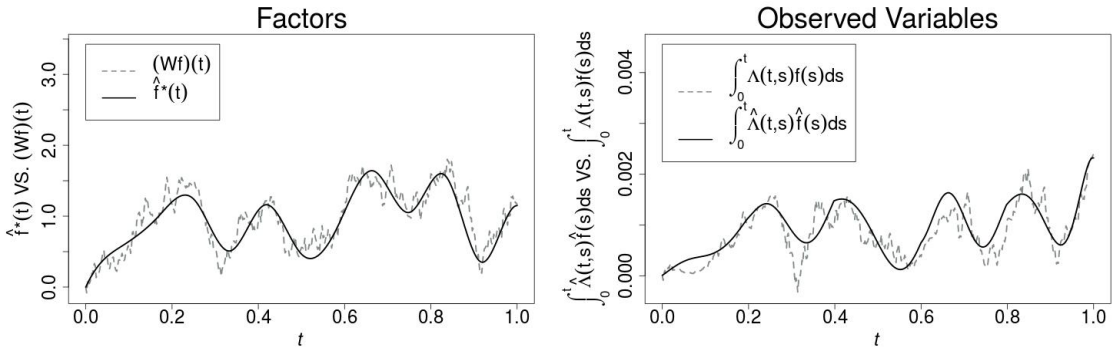


Figure 2.1 Functional Estimates VS. True Processes, $J = 25$, $n = 20$, $\kappa_f = 400$, where J indicates the number of time observations, n the number of individuals and κ_f the coefficient of the continuous-time AR(1) process.

empirical means and variances for all t. Table 2.2 presents the percentages of "fail to reject normality", for the four different sample sizes and the three different κ_f values. The results show that in all the functional estimators, for the majority of time, we fail to reject normality. Figures 2.3 and 2.4 present the p-values of the K-S test over time comparing with the 5% significance level, for some sample estimators. Again, for the majority of time, we fail to reject normality.

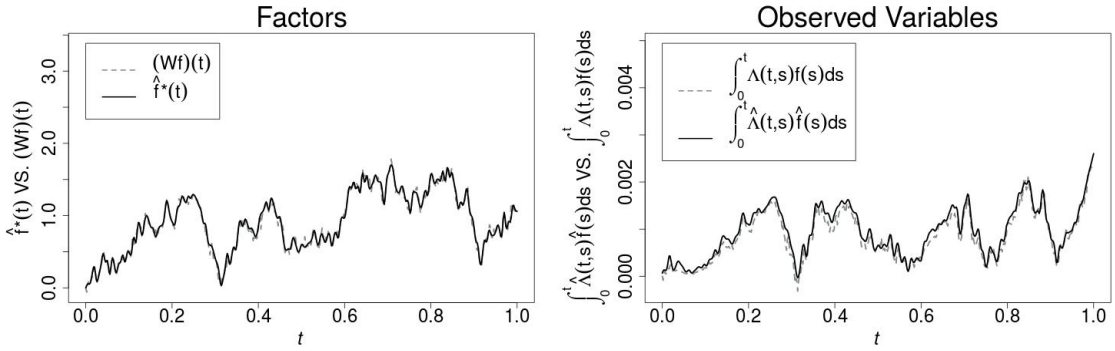


Figure 2.2 Functional Estimates VS. True Processes, $J = 201$, $n = 100$, $\kappa_f = 400$, where J indicates the number of time observations, n the number of individuals and κ_f the coefficient of the continuous-time AR(1) process.

Table 2.2: Normality

		KS_f			KS_X		
J	n	$\kappa_f = 400$	$\kappa_f = 250$	$\kappa_f = 100$	$\kappa_f = 400$	$\kappa_f = 250$	$\kappa_f = 100$
25	20	0.5768	0.5010	0.5569	0.9944	0.9937	0.9933
51	30	0.7844	0.8004	0.9321	0.9976	0.9979	0.9927
101	55	0.8503	0.9222	0.9800	0.9956	0.9948	0.9879
201	100	0.9042	0.9441	0.9780	0.9960	0.9921	0.9847

Note: J indicates the number of time observations, n the number of individuals, κ_f the coefficient of the continuous-time AR(1) process, KS_f and KS_X the average p-values of the K-S test over time for $f_k(t)$ and $x_i(t)$.

2.5 Empirical Analysis

In the current section, we apply our model to real data. Our goal is to present the application of our method through empirical analysis; meanwhile, we adopt hypothesis tests to justify the choice of the GFDFM over the factor models with constant or one-dimensional-time-varying loadings.

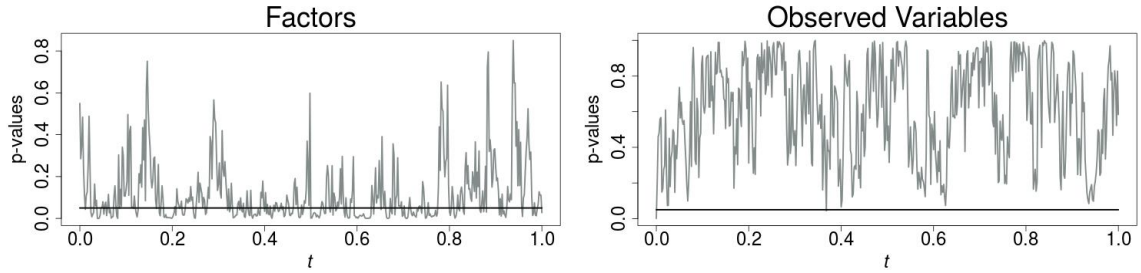


Figure 2.3 Functional Estimates VS. True Processes, $J = 25$, $n = 20$, $\kappa_f = 400$, where J indicates the number of time observations, n the number of individuals and κ_f the coefficient of the continuous-time AR(1) process.

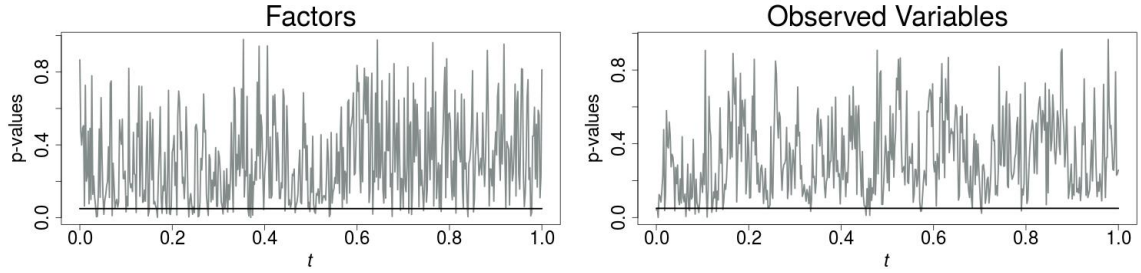


Figure 2.4 Functional Estimates VS. True Processes, $J = 201$, $n = 100$, $\kappa_f = 400$, where J indicates the number of time observations, n the number of individuals and κ_f the coefficient of the continuous-time AR(1) process.

2.5.1 Macroeconomic data (Stock and Watson, 2009)

The study by Stock and Watson (2009) investigates split-sample instability with a single break. In the current empirical analysis, we use the data set from Stock and Watson (2009), which includes 144 quarterly macroeconomics series from 1959:I to 2006:IV, total 192 observations for each series. The data are publicly available from *Publications and Replication Files* at `http://www.princeton.edu/~mwatson/publi.html`. In Stock and Watson's study, they group the series into 13 categories and transform the data by taking logarithm or differencing — in general, first differences of logarithms (growth rates) are used for real quantity variables, first differences are used for nominal interest rates, and second differences of logarithms (changes in rates of inflation) for price series. Since up to the second differences are taken, a balanced panel covering from 1959:III to 2006:IV for

144 series is used for analysis, and 109 out of 144 disaggregated series are used to compute principle components. They compared the performance of three different methods for four-step ahead in-sample forecasting with a single break in 1984:I — (a) a "full-full" model (referred to as "FF", hereafter) with full-sample estimates of the factors and full-sample forecasting regression (b) a "full-split" model ("FS") with full-sample estimates of the factors and split-sample forecasting regression, and (c) a "split-split" model ("SS") with split-sample estimates of the factors and split-sample forecasting regression. The comparison is presented (in Table 5 of their paper) by relative mean squared forecasting errors, and the results show that the "FS" method outperforms the "SS" method, and "FS" also produces improvements relative to "FF", especially for the post-84 sample.

Stock and Watson (2009) estimate the number of factors using the Bai and Ng (2002) information criteria, and eventually keep three to five factors in their analysis. Though our functional approach works with any given number of factors, in this empirical analysis, we impose $K = 1$ to obtain the estimated functional factor based on the 109 selected series by Stock and Watson (2009), and then we provide four-step ahead forecasts for all 144 series using the forecasting regression (8) from their paper.

There are two main components in our empirical analysis. First, we apply the hypothesis test on the second dimension of loadings to justify our application of our GFDFM, by showing that the null hypothesis of "the cross-individual variance function is a zero function" is rejected at 5% significance level. Second, we compute the ratios of the root mean squared forecasting errors (RMSFEs) under GFDFM with $K = 1$ to the RMSFEs obtained by (Stock and Watson, 2009) with three to five factors, to show that our functional approach has better performance in terms of RMSFEs. Note that Stock and Watson (2009) use an in-sample forecast, and in order for our results to be comparable with theirs, we follow their method for predictive assessment.

2.5.2 Hypothesis test

The general functional form allows more flexibility in our model by using a second dimension of loading functions to capture the individualized lag effects of factors as well as the dynamics of the lag effects. Thus, the utilization of the general functional form can be justified by the following two reasons: (1) for at least some of the individuals in the

sample, the second dimension of the loading functions shows non-zero pattern, and (2) the second dimension of the loading functions shows individual-specific dynamics across all the replications. Essentially, if (2) is true, then it implies (1). We justify point (2) by checking whether the second dimension of the loading functions has non-zero cross-individual variance. In the current chapter, we use bootstrap tests on the null hypothesis that the cross-individual variance function is a zero function, and we adopt an L_1-norm for the construction of the test statistics.

Specifically, we define the variance function for the second dimension of the loadings, at a given concurrent dimension time t, as $V_{n,t}(s) = n^{-1}\sum_{i=1}^{n}\left\{\lambda_i(t,s) - \bar{\lambda}_n(t,s)\right\}^2$, where $\bar{\lambda}_n(t,s) = n^{-1}\sum_{i=1}^{n}\lambda_i(t,s)$. Then we construct the test statistic as

$$W_{n,t} := \int_0^t |V_{n,t}(s)|\, ds, \tag{2.20}$$

which is the size of the variance function $V_{n,t}(s)$ on $s \in [0,t]$ for a given t and n. For the bootstrap test statistic, let $b = 1$, ..., B, where $B = 1000$, indicating the number of size-n bootstrap samples. We define $\bar{\lambda}^*_{b,n}(t,s) = n^{-1}\sum_{i=1}^{n}\lambda^*_{b,i}(t,s)$, where $\lambda^*_{b,i}(t,s)$ denotes the i^{th} replication of the b^{th} re-sample from $\{\lambda_i(t,s)\}_{i=1}^n$ with replacement. The bootstrap statistic is then constructed as

$$W^*_{b,n,t} := \int_0^t \left|V^*_{b,n,t}(s) - V_{n,t}(s)\right| ds, \tag{2.21}$$

where $V^*_{b,n,t}(s) := n^{-1}\sum_{i=1}^{n}\left\{\lambda^*_{b,i}(t,s) - \bar{\lambda}^*_{b,n}(t,s)\right\}^2$.

Intuitively, under the null, $V_{n,t}(s)$ in Equation (2.21) is a zero function; therefore, $W_{n,t}$ and $W^*_{b,n,t}$ share the same distribution, which implies that our test has correct size; under the alternative, the $W^*_{b,n,t}$'s capture the size of centered variance functions while the $W_{n,t}$'s capture the size of uncentered variance functions, which generally implies the $W_{n,t}$'s are greater than the $W^*_{b,n,t}$'s, and thus, our test has power. We defer more detailed verification for the properties of the test statistics to future research.

We select seven time points from 1959:III — 2006:IV to present the test statistics, and the results are summarized in Table 2.3. Basically, the null of zero-variance function is rejected at all seven time points under 95% confidence interval and rejected at six out of the seven time points under 99% confidence interval, implying that the second dimension of the loadings does pick up individual-specific lag effects of factors, and for at least some individuals the lag effect is non-zero, which justify the application of the GFDFM.

Table 2.3: Hypothesis Test, Macroeconomic Data

	61:IV	69:II	76:IV	84:II	91:IV	99:II	06:IV
$W_{n,t}$	0.48	3.68	80.47	392.99	2322.22	6977.38	15334.05
95% $W^*_{n,t}$	0.22	2.18	70.17	169.41	1127.81	3848.30	8710.40
99% $W^*_{n,t}$	0.28	2.90	105.44	195.95	1403.75	4858.23	10957.59

Note: the table presents the sample statistics, $W_{n,t}$, as well as the 95% and 99% bootstrap criteria (95% $W^*_{n,t}$ and 99% $W^*_{n,t}$).

2.5.3 Comparison of the forecasting results

For the comparisons of the four-step ahead forecasting, we split our estimated functions into "pre-84" and "post-84" at the time of "1984:I", and we take the ratios of the RMSFEs from GFDFM to the three RMSFEs from Stock and Watson's estimates for each of the 144 series. The complete list of ratios for all 144 series are shown in Table 2.4, and Table 2.5 provides a summary in terms of the quantiles, the means as well as the percentages of RMSFEs smaller than one. For example, for the comparison between our GFDFM and FF with the pre-84 sample, the ratios of the RMSFEs range from 0.3742 to 1.9885 over all 144 series with the mean 0.7355 and the median 0.6971, and for about 83.33% of the series, GFDFM produces a better forecast than FF in terms of RMSFE. From Table 2.5 we can see that on average, in terms of RMSFE, GFDFM outperforms the three methods from Stock and Watson (2009) for both the pre-84 and the post-84 periods. The corresponding histograms of the ratios over all 144 series in Figures 2.5 - Figure 2.7 also illustrate the point.

Table 2.4: Relative Root Mean Square Forecasting Errors, Macroeconomic Data

	Pre-84 Sample			**Post-84 Sample**		
Series	GFDFM-FF	GFDFM-FS	GFDFM-SS	GFDFM-FF	GFDFM-FS	GFDFM-SS
RGDP	0.65	0.67	0.68	0.54	0.65	0.60
Cons	0.55	0.57	0.57	0.45	0.54	0.50
Cons-Dur	0.74	0.76	0.76	0.64	0.70	0.69
Cons-NonDur	0.63	0.67	0.63	0.48	0.55	0.52
Cons-Serv	0.43	0.45	0.45	0.54	0.73	0.64
GPDInv	0.90	0.95	0.94	0.77	0.87	0.84
FixedInv	0.62	0.66	0.66	0.36	0.43	0.42
NonResInv	0.53	0.56	0.56	0.41	0.47	0.47
NonResInv-Struct	0.50	0.53	0.53	0.65	0.73	0.72

NonResInv-Bequip	0.59	0.64	0.63	0.36	0.42	0.42
Res.Inv	0.69	0.71	0.71	0.52	0.66	0.60
Exports	0.93	0.97	0.98	0.68	0.72	0.72
Imports	0.82	0.83	0.82	0.63	0.67	0.66
Gov	0.41	0.41	0.41	0.50	0.52	0.52
Gov Fed	0.40	0.40	0.40	0.54	0.57	0.58
Gov State/Loc	0.47	0.47	0.47	0.55	0.61	0.60
IP: total	0.71	0.74	0.74	0.48	0.55	0.54
IP: products	0.61	0.64	0.64	0.47	0.55	0.53
IP: final prod	0.60	0.62	0.62	0.49	0.57	0.55
IP: cons gds	0.76	0.78	0.77	0.54	0.72	0.67
IP: cons dble	0.84	0.85	0.85	0.69	0.74	0.72
iIP:cons nondble	0.59	0.63	0.60	0.56	0.78	0.71
IP:bus eqpt	0.58	0.62	0.61	0.45	0.49	0.49
IP: matls	0.82	0.85	0.86	0.54	0.62	0.62
IP: dble mats	0.83	0.87	0.86	0.50	0.59	0.58
IP:nondble mats	0.80	0.87	0.86	0.53	0.68	0.64
IP: mfg	0.74	0.76	0.77	0.48	0.54	0.53
IP: fuels	0.67	0.68	0.68	0.65	0.72	0.70
NAPM prodn	0.71	0.72	0.73	0.66	0.74	0.67
Capacity Util	0.48	0.50	0.50	0.39	0.45	0.43
Emp: total	0.58	0.61	0.63	0.34	0.43	0.40
Emp: gds prod	0.64	0.67	0.68	0.36	0.48	0.45
Emp: mining	0.47	0.49	0.49	0.50	0.55	0.54
Emp: const	0.52	0.54	0.55	0.36	0.43	0.41
Emp: mfg	0.62	0.67	0.68	0.37	0.52	0.50
Emp: dble gds	0.60	0.64	0.64	0.39	0.52	0.50
Emp: nondbles	0.67	0.78	0.76	0.35	0.48	0.46
Emp: services	0.42	0.45	0.46	0.34	0.41	0.38
Emp: TTU	0.48	0.54	0.54	0.33	0.42	0.38
Emp: wholesale	0.48	0.57	0.56	0.38	0.45	0.43
Emp: retail	0.48	0.54	0.54	0.30	0.40	0.37
Emp: FIRE	0.38	0.41	0.41	0.46	0.53	0.50
Emp: Govt	0.37	0.39	0.38	0.46	0.57	0.56
Help wanted indx	0.54	0.59	0.59	0.43	0.48	0.46
Help wanted/emp	0.58	0.59	0.59	0.38	0.45	0.44
Emp CPS total	0.58	0.63	0.63	0.47	0.59	0.50
Emp CPS nonag	0.57	0.63	0.63	0.51	0.63	0.54
Emp. Hours	0.64	0.69	0.70	0.33	0.40	0.38
Avg hrs	0.71	0.72	0.72	0.62	0.66	0.66
Overtime: mfg	0.81	0.84	0.84	0.71	0.75	0.73
U: all	0.70	0.71	0.71	0.43	0.51	0.46
U: mean duration	0.79	0.83	0.82	0.70	0.85	0.78
U $<$ 5 wks	0.73	0.76	0.75	0.66	0.71	0.68
U 5-14 wks	0.75	0.76	0.76	0.53	0.61	0.55
U 15+ wks	0.61	0.63	0.63	0.46	0.57	0.52
U 15-26 wks	0.72	0.74	0.74	0.56	0.68	0.61
U 27+ wks	0.61	0.63	0.62	0.50	0.61	0.55
HStarts: Total	0.45	0.46	0.46	0.36	0.40	0.41
BuildPermits	0.44	0.44	0.45	0.30	0.34	0.34

HStarts: NE	0.54	0.55	0.56	0.50	0.57	0.54
HStarts: MW	0.50	0.51	0.51	0.64	0.65	0.62
HStarts: South	0.47	0.48	0.49	0.46	0.53	0.52
HStarts: West	0.44	0.45	0.45	0.38	0.41	0.40
PMI	0.69	0.72	0.74	0.59	0.68	0.62
NAPM new ordrs	0.73	0.73	0.74	0.69	0.77	0.70
NAPM vendor del	0.64	0.65	0.67	0.48	0.65	0.59
NAPM Invent	0.72	0.78	0.83	0.39	0.60	0.55
Orders (ConsGoods)	0.79	0.84	0.86	0.56	0.68	0.66
Orders (NDCapGoods)	0.77	0.81	0.81	0.65	0.74	0.73
PGDP	1.03	1.05	1.06	1.08	1.35	1.27
PCED	0.88	0.89	0.90	1.02	1.23	1.17
CPI-ALL	0.97	0.99	0.99	1.02	1.21	1.20
PCED-Core	0.89	0.90	0.89	1.09	1.41	1.31
CPI-Core	0.80	0.81	0.80	0.98	1.33	1.28
PCED-DUR	1.04	1.06	1.04	0.91	1.14	1.05
PCED-DUR-MOTORVEH	1.20	1.21	1.20	1.11	1.22	1.19
PCED-DUR-HHEQUIP	0.96	1.00	0.96	1.11	1.44	1.34
PCED-DUR-OTH	0.92	0.93	0.93	1.38	1.63	1.45
PCED-NDUR	0.93	0.95	0.98	1.30	1.37	1.35
PCED-NDUR-FOOD	1.08	1.08	1.09	1.31	1.52	1.41
PCED-NDUR-CLTH	1.17	1.20	1.16	1.59	1.71	1.64
PCED-NDUR-ENERGY	1.05	1.15	1.15	1.31	1.31	1.36
PCED-NDUR-OTH	0.96	1.00	0.99	1.16	1.34	1.28
PCED-SERV	1.03	1.03	1.05	1.40	1.61	1.61
PCED-SERV-HOUS	1.05	1.07	1.08	1.17	1.24	1.22
PCED-SERV-HOUSOP	1.08	1.12	1.14	1.24	1.28	1.24
PCED-SERV-H0-ELGAS	1.04	1.25	1.24	0.92	0.96	0.95
PCED-SERV-HO-OTH	1.08	1.09	1.09	1.64	1.86	1.73
PCED-SERV-TRAN	0.95	1.26	1.22	0.54	0.64	0.64
PCED-SERV-MED	1.16	1.20	1.20	0.87	1.03	1.03
PCED-SERV-REC	1.20	1.19	1.21	1.26	1.35	1.30
PCED-SERV-OTH	0.97	0.99	0.99	1.25	1.43	1.57
PGPDI	1.15	1.18	1.14	0.83	1.13	1.07
PFI	1.16	1.20	1.16	0.84	1.13	1.07
PFI-NRES	0.89	0.93	0.90	0.84	1.10	1.06
PFI-NRES-STR Price Index	0.91	0.93	0.93	0.85	0.99	0.95
PFI-NRES-EQP	0.97	1.02	0.97	0.88	1.08	1.05
PFI-RES	0.80	0.80	0.81	0.95	1.45	1.44
PEXP	1.02	1.04	1.06	0.89	1.08	1.02
PIMP	0.81	0.83	0.85	1.19	1.31	1.30
PGOV	1.51	1.60	1.60	1.46	1.73	1.71
PGOV-FED	1.99	2.04	2.03	2.31	2.48	2.49
PGOV-SL	1.08	1.14	1.16	1.00	1.22	1.18
Com: spot price (real)	0.55	0.59	0.57	0.42	0.48	0.46
OilPrice (Real)	0.42	0.49	0.49	0.64	0.70	0.70
NAPM com price	0.58	0.62	0.64	0.50	0.62	0.58
Real AHE: goods	0.47	0.50	0.48	0.45	0.53	0.51
Real AHE: const	0.41	0.42	0.41	0.36	0.41	0.41
Real AHE: mfg	0.51	0.55	0.54	0.45	0.52	0.50

Labor Prod	0.70	0.72	0.71	0.67	0.72	0.72
Real Comp/Hour	0.44	0.46	0.45	0.55	0.57	0.56
Unit Labor Cost	0.54	0.53	0.55	0.50	0.65	0.63
FedFunds	0.66	0.70	0.71	0.53	0.65	0.64
3 mo T-bill	0.68	0.73	0.73	0.49	0.58	0.57
6 mo T-bill	0.66	0.71	0.69	0.46	0.54	0.53
1 yr T-bond	0.63	0.67	0.64	0.45	0.51	0.49
5 yr T-bond	0.51	0.53	0.52	0.51	0.54	0.56
10 yr T-bond	0.47	0.49	0.49	0.53	0.58	0.60
Aaabond	0.44	0.46	0.45	0.51	0.55	0.58
Baa bond	0.40	0.41	0.41	0.52	0.56	0.58
fygm6-fygm3	0.73	0.75	0.75	0.58	0.68	0.64
fygt1-fygm3	0.62	0.68	0.65	0.52	0.61	0.59
fygt10-fygm3	0.58	0.59	0.59	0.50	0.60	0.60
FYAAAC-Fygt10	0.50	0.56	0.54	0.49	0.52	0.51
FYBAAC-Fygt10	0.52	0.55	0.56	0.43	0.47	0.47
M1	1.46	1.57	1.56	0.98	1.01	1.10
MZM	0.85	0.87	0.88	0.89	0.99	1.10
M2	1.10	1.18	1.16	1.03	1.22	1.33
MB	1.22	1.35	1.36	0.79	0.81	0.82
Reserves tot	1.17	1.51	1.53	0.85	0.92	0.93
Reserves nonbor	0.71	0.81	0.78	0.83	0.95	0.93
BUSLOANS	1.10	1.15	1.13	0.98	1.10	1.06
Cons credit	0.88	0.95	0.93	0.76	0.82	0.82
Ex rate: avg	0.44	0.48	0.48	0.51	0.54	0.50
Ex rate: Switz	0.52	0.55	0.53	0.53	0.57	0.55
Ex rate: Japan	0.55	0.59	0.56	0.51	0.53	0.52
Ex rate: UK	0.46	0.52	0.51	0.61	0.69	0.63
EX rate: Canada	0.43	0.50	0.49	0.48	0.50	0.51
S&P 500	0.72	0.81	0.79	0.44	0.52	0.52
S&P: indust	0.71	0.80	0.78	0.45	0.51	0.51
S&P div yield	0.83	0.88	0.83	0.45	0.57	0.57
S&P PE ratio	0.78	0.94	0.89	0.59	0.67	0.66
DJIA	0.75	0.85	0.84	0.51	0.62	0.62
Consumer expect	0.70	0.77	0.76	0.68	0.82	0.82

As mentioned previously, according to Stock and Watson (2009), the FS model outperforms both the FF and the SS models in general. Hence, we now present a comparison between the FS model and GFDFM. In Figures 2.8 to 2.13, we show the four-step ahead forecast, using FS and GFDFM respectively, for six variables — real GDP, exports, imports, unemployment rates, CPI, as well as effective interest rates, and Table 2.6 shows the corresponding comparison of RMSFEs.

Table 2.5: Relative Root Mean Square Forecasting Errors, Macroeconomic Data

	Pre-84 Sample			**Post-84 Sample**		
	GFDFM-FF	GFDFM-FS	GFDFM-SS	GFDFM-FF	GFDFM-FS	GFDFM-SS
Min.	0.3742	0.3861	0.3828	0.2991	0.3412	0.3448
1st Qu.	0.5366	0.5656	0.5562	0.4620	0.5354	0.5171
Median	0.6971	0.7223	0.7301	0.5381	0.6468	0.6193
Mean	0.7355	0.7760	0.7721	0.6728	0.7814	0.7581
3rd Qu.	0.8883	0.9305	0.9062	0.8402	0.9635	0.9503
Max.	1.9885	2.0394	2.0255	2.3123	2.4805	2.4946
¡ 1	0.8333	0.8056	0.8264	0.8403	0.7639	0.7569

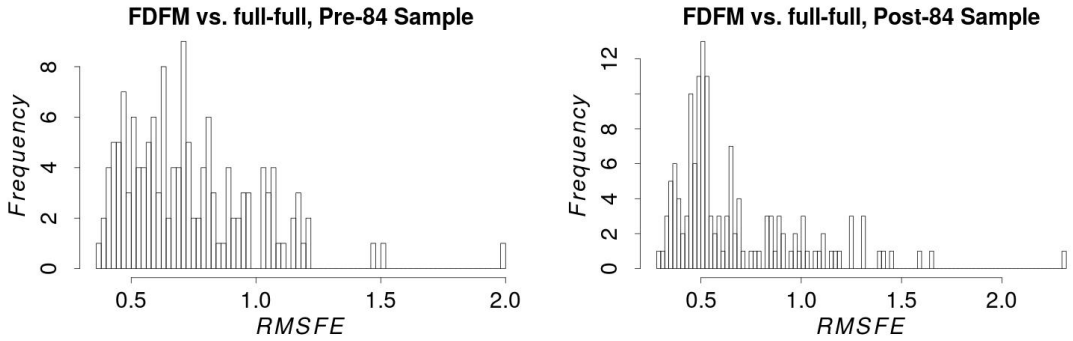


Figure 2.5 RMSFE, GFDFM vs. Full-Full

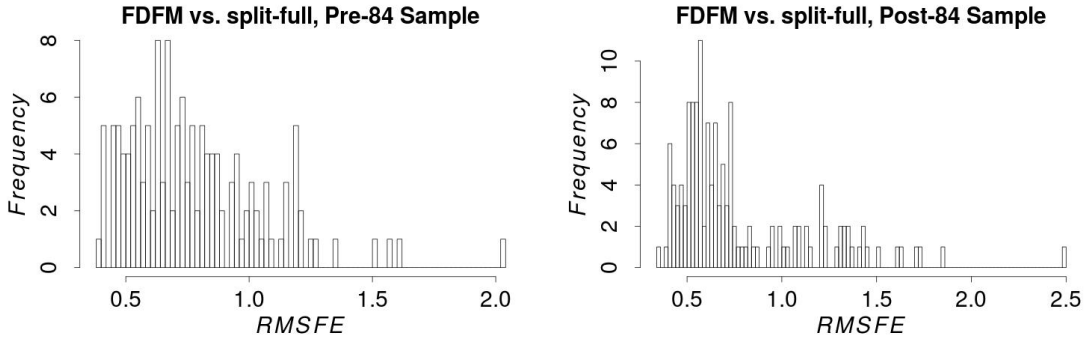


Figure 2.6 RMSFE, GFDFM vs. Split-Full

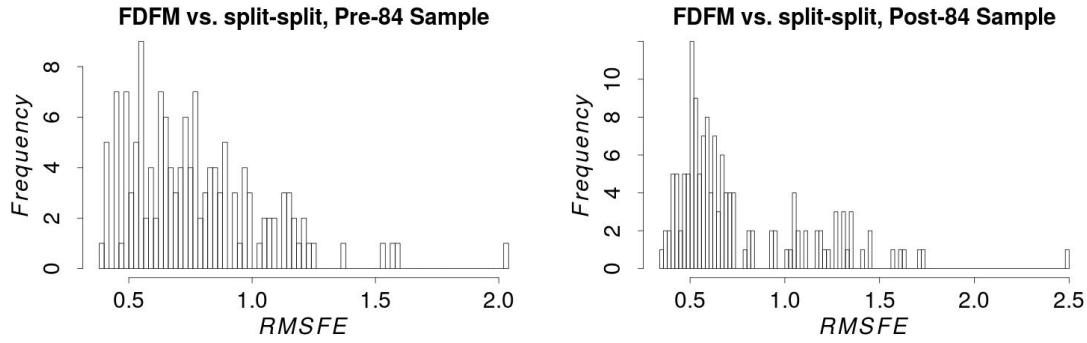


Figure 2.7 RMSFE, GFDFM vs. Split-Split

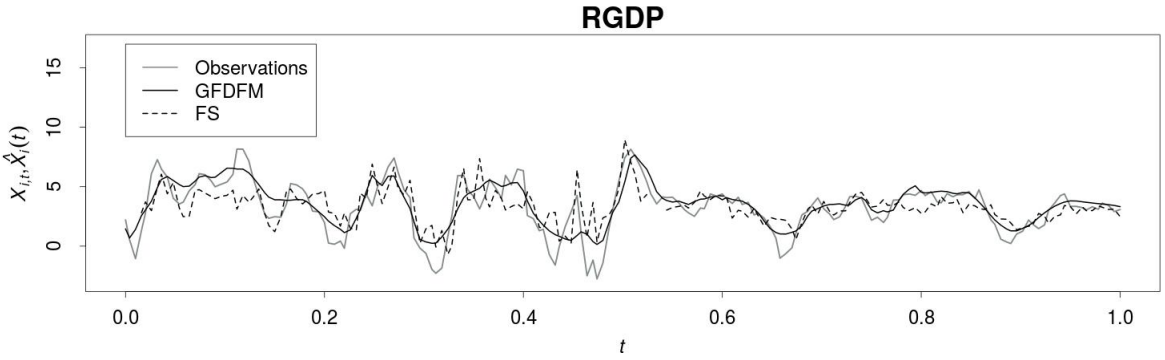


Figure 2.8 Real GDP, quantity index (2000=100)

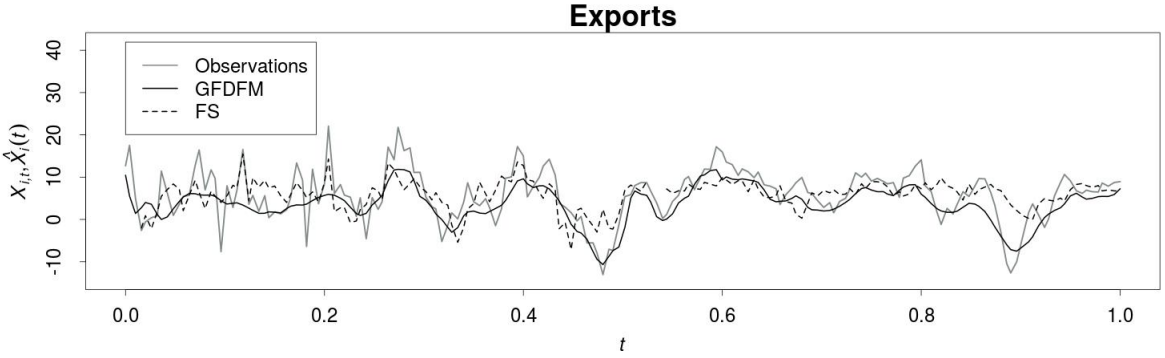


Figure 2.9 Real GDP and Unemployment rate four-step ahead forecasting

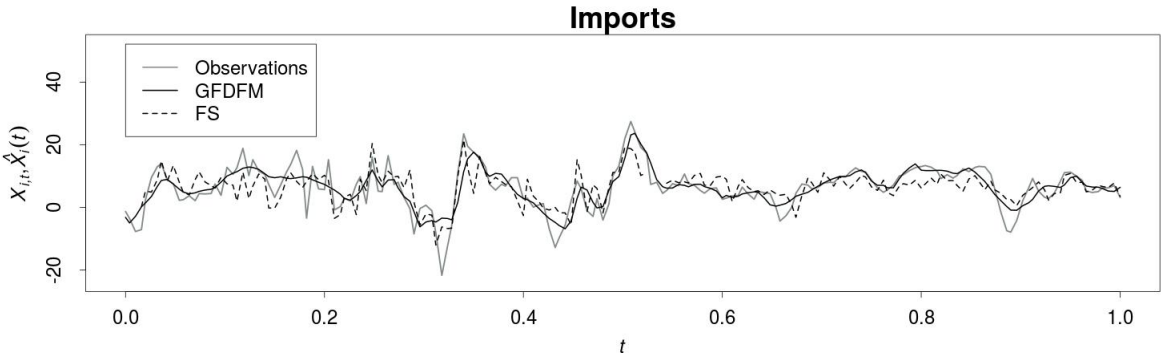


Figure 2.10 Real GDP and Unemployment rate four-step ahead forecasting

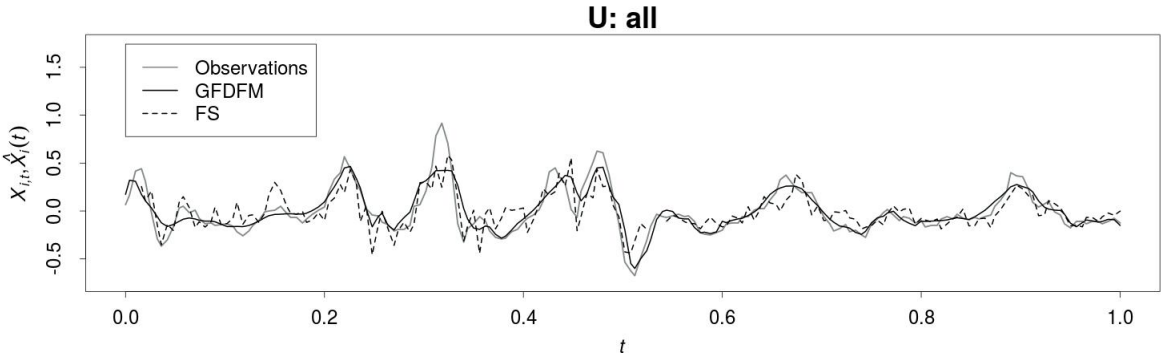


Figure 2.11 Unemployment rate, all workers, 16 years & over (%)

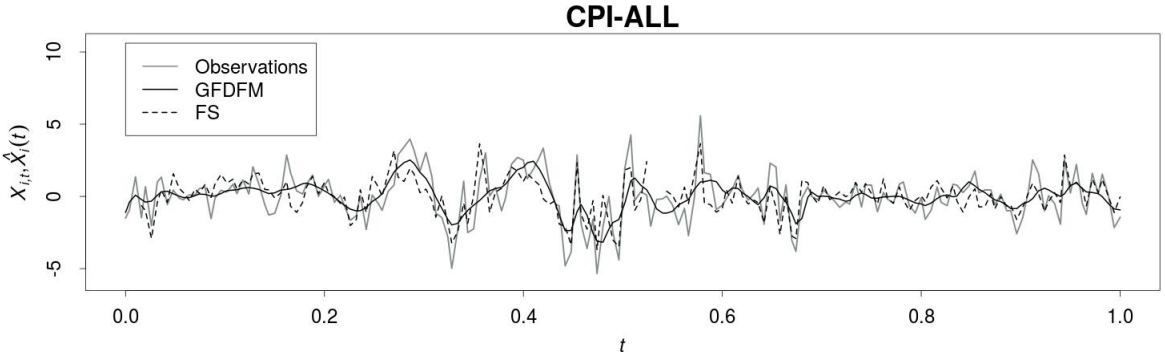


Figure 2.12 CPI, all items

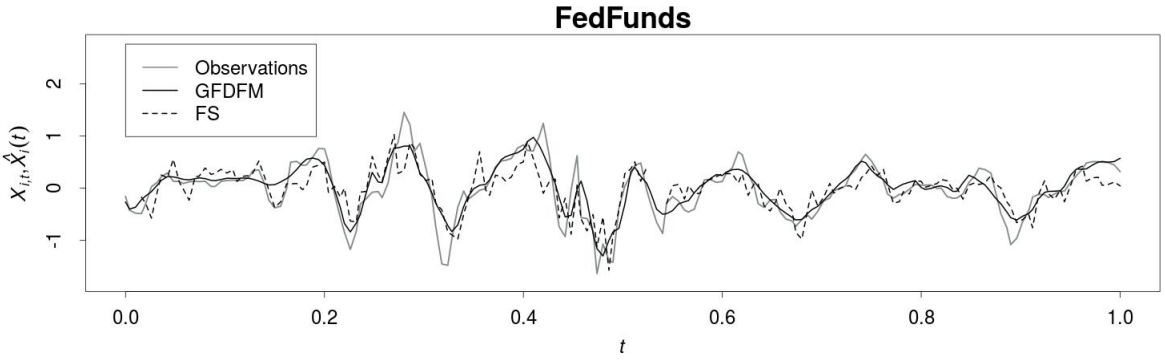


Figure 2.13 Interest rate, federal funds (effective) (% per annum)

Table 2.6: Relative Root Mean Square Forecasting Errors, Macroeconomic Data

	Pre-84 Sample			**Post-84 Sample**		
	GFDFM-FF	GFDFM-FS	GFDFM-SS	GFDFM-FF	GFDFM-FS	GFDFM-SS
RGDP	0.65	0.67	0.68	0.54	0.65	0.60
Exports	0.93	0.97	0.98	0.68	0.72	0.72
Imports	0.82	0.83	0.82	0.63	0.67	0.66
U: all	0.70	0.71	0.71	0.43	0.51	0.46
CPI-ALL	0.97	0.99	0.99	1.02	1.21	1.20
FedFunds	0.66	0.70	0.71	0.53	0.65	0.64

2.6 Conclusion

The DFMs constructed in previous literature either assume constant loadings over time or concurrent time-varying loadings. This chapter suggests a more general specification for dynamic factor models in terms of functional data. There are two main remarks of our model. First, given the fact that the data of interest for DFMs usually has continuous underlying processes, e.g., macroeconomic data, we model the processes as continuous functions and adopt a continuous-time analysis to achieve estimation consistency and asymptotic normality. On the other hand, our general model setup allows the loadings to capture both the lag and the concurrent time-varying individual-specific effects of the factors on the data series.

We provide theorems for the consistency and normality of the estimators as well as the

proofs of the asymptotics; then we use a simulation study to illustrate the theorems. We also present an empirical study using macroeconomic data from Stock and Watson (2009) and propose a heuristic bootstrap test to justify the application of our GFDFM by testing the significance of the second dimension of loadings — the test results basically show that the second dimension of the loadings is statistically significant and thus should be included in the model, and by comparing the RMSFEs of GFDFM with that in Stock and Watson (2009), we show that the estimators under GFDFM perform better.

Chapter 3

Historical-time Functional Linear Model and its Inference with Cross-sectional Dependence

3.1 Introduction

In the current chapter, I study another major branch of FDA — functional linear regression and its inference with longitudinal functional data. In general, functional linear regression can be modeled with a mixture of either scalar or functional response, predictors and coefficients. For example, models with scalar responses, say Y_i, in the form of $Y_i = \sum_{k_z=1}^{K_Z} Z_{k_z i}\beta_{Z,k_z} + \sum_{k=1}^{K} \int_{\mathcal{S}} X_{ki}(t)\beta_k(t)dt + U_i$ have been studied either only with the functional predictors $X_{ki}(t)$'s and coefficients $\beta_k(t)$'s on the domain $\mathcal{S}$ (see, e.g., Cardot et al., 1999, 2003; Hall and Horowitz, 2007) or including both the functional terms and the scalar ones $Z_{k_z i}$'s and β_{Z,k_z}'s (see, e.g., Hu et al., 2004). Moreover, Fan and Zhang (1999) and Zhu et al. (2014), among others, develop models with functional responses, functional predictors and univariate functional coefficients as such $Y_i(t) = \sum_{k=1}^{K} X_{ki}(t)\beta_k(t) + U_i(t)$, while similar models with bivariate functional coefficients defined as

$$Y_i(t) = \sum_{k=1}^{K} \left[\int_{\mathcal{S}} X_{ki}(s)\beta_k(s,t)ds \right] + U_i(t) \tag{3.1}$$

have also been investigated (see, e.g., Ramsay and Dalzell, 1991; Benatia et al., 2017; Lin et al., 2019). In this chapter, I focus on a generalized version of the models with bivariate functional coefficients, allowing the domain $\mathcal{S}$ of the integral variable s to vary along t, such that

$$Y_i(t) = \sum_{k=1}^{K} \left[\int_0^t X_{ki}(s)\beta_k(s,t)ds \right] + U_i(t), \tag{3.2}$$

which is one type of the "historical-time functional linear model" (hFLM) (see, e.g., Malfait and Ramsay, 2003; Ramsay, 2005; Harezlak et al., 2007; Kim et al., 2011; Wang et al., 2016).

As stated by Malfait and Ramsay (2003) and Ramsay (2005), a more general form of the hFLM is specified as $Y_i(t) = \sum_{k=1}^{K} \left[\int_{\mathcal{S}_t} X_{ki}(s)\beta_k(s,t)ds \right] + U_i(t)$ with $\mathcal{S}_t$ being an arbitrary function of t, and one common situation is where $\mathcal{S}_t = [0,t]$ imposing a forward-in-time-only causality by treating $\beta_k(s,t) = 0$ for $s > t$. There has been literature developing estimation procedures and estimator properties for the hFLM. Namely, Malfait and Ramsay (2003) construct a bivariate "tent-like" piecewise linear basis system, use it to expand the coefficient surface $\beta_k(s,t)$ on a triangular domain, and fit the model assuming zero cross-sectional correlation in the error terms. Harezlak et al. (2007) suggest to compare the performance of a variety of regularization methods for linear B-spline basis functions, including basis truncation, roughness penalties, and sparsity penalties using an extension of the Akaike Information Criterion. Kim et al. (2011) propose an estimation procedure for the recent hFLM that is oriented towards sparse longitudinal data, where the observation times across subjects are irregular and the total number of measurements per subject is small. Assuming i.i.d. samples, they obtain the basis-approximated functional response and predictors by utilizing FPCA built upon the kernel smoothed auto-covariance functions, and then estimate the bivariate functional coefficients. They establish uniform consistency of the proposed estimators with a convergence rate depending on the sample size and the kernel smoothing parameter; they also derive the asymptotic distribution of the fitted response trajectories, which can be used to construct asymptotic point-wise confidence intervals. Apart from these previous studies, the inference on the functional coefficients is also an essential aspect of functional regression analysis yet has not been developed much for hFLMs.

In this chapter, I first propose an estimator for the functional coefficients in the hFLM as defined in (3.2) by using B-spline-based expansions. Upon the estimation, I study the

asymptotic properties of the estimated coefficients, providing a uniform convergence rates as well as an asymptotic normality result under certain conditions. Lastly, since the asymptotic distribution of the estimator does not necessarily provide a good approximation with finite samples, I develop a bootstrap method that can better approximate the distribution of the estimated coefficients in finite sample situations.

Bootstrap methods involving functional data have been developed under variety of conditions. For example, Cuevas and Fraiman (2004) provides a general discussion on the asymptotic validity of the bootstrap methodology for functional data. Sharipov et al. (2016) and Paparoditis (2018) propose bootstrap functional central limit theorems; specifically, Sharipov et al. (2016) develop a functional central limit theorem for the block bootstrap in a Hilbert space to capture the dependence structure among the random functions, while Paparoditis (2018) illustrate a bootstrap central limit theorem for functional finite Fourier transforms and demonstrate the validity of their bootstrap method. Particularly for functional linear regression, Herwartz and Xu (2009) suggest a factor based bootstrap method for the inference of functional coefficient models, allowing for heteroskedastic error terms; González-Manteiga and Martínez-Calvo (2011) propose a bootstrap procedure to obtain the pointwise confidence intervals in a functional linear model with scalar responses and demonstrate the asymptotic validity of their bootstrap method.

However, it is worth noting that for data with dependence, bootstrap can easily fail by ignoring the order of observations. In previous studies with functional settings, constraints on dependence have been imposed across functions. For instance, bootstrap with longitudinal functional data mostly assumes cross-sectional independence (see, e.g., Hall et al., 1989; Cuevas et al., 2006; González-Manteiga and Martínez-Calvo, 2011; Shang, 2015); while bootstrap with functional time series has assumed stationarity and weak dependence across functions (see, e.g., Dehling et al., 2015; Sharipov et al., 2016; Paparoditis, 2018). There is also literature on bootstrapping functional data with dependence across replications in various model specifications, such as linear autoregressive Hilbertian model, functional linear models with scalar responses, functional linear models with functional time series data or functional autoregressive models (see, e.g., De Castro et al., 2005; Rana et al., 2016; Shang, 2018; Franke and Nyarige, 2019).

In the current chapter, I adopt the idea of the moving blocks bootstrap (MBB) developed by Kunsch (1989) and Liu and Singh (1992). I represent the regressor and residual

functions using local polynomial basis expansions. Then under some stationarity and ergodicity assumptions, I generate the bootstrap functions for the regressors and the residuals via a moving block bootstrap on their basis coefficients. Such a bootstrap accommodates unknown forms of cross-sectional dependence that can be either weak or strong.

This chapter contributes to the literature by establishing the asymptotics of the B-spline-based estimated functional coefficients and developing the bootstrap inference, accommodating unknown forms of cross-sectional dependence. The main findings are (i) a uniform convergence rate of the estimated functional coefficients is derived depending on the degree of cross-sectional dependence and $\sqrt{n}$-consistency can be achieved in the absence of cross-sectional dependence, which is a faster convergence than that of the estimators proposed in Kim et al. (2011); (ii) under proper stationarity and ergodicity conditions on the functional variables, asymptotic normality of the estimated coefficients can be obtained with unknown forms of cross-sectional dependence; (iii) the proposed bootstrap method has a better finite-sample performance than the asymptotics while approximating the distribution of the estimated functional coefficients, and it can be used to construct percentile confidence intervals and perform hypothesis tests for the functional coefficients. To my best knowledge, this study is the first to discuss the asymptotic and bootstrap inferences for the functional coefficients in hFLMs with cross-sectional dependence.

3.2 The Model and its Estimation

Linear regression is one of the most powerful and widely used approach to analyzing the relationship between the response and the predictors. In the case with functional data, where either the response or the predictors (or both) are functions defined on continua, such a relationship may vary along time, and the response at time t may be affected by the predictors from different time points. In order to capture such effects, Ramsay and Dalzell (1991), who devised the term "functional data analysis", introduced an infinite dimensional regression with bivariate functional coefficients as in Equation (3.1). While Malfait and Ramsay (2003) and Ramsay (2005) pointed out that to properly specify such a dynamic relationship, different versions of functional linear regression models shall be adopted, given different scenarios. In this section, I first introduce the model of interest, the hFLM (see, e.g., Malfait and Ramsay, 2003; Ramsay, 2005; Harezlak et al., 2007; Kim

et al., 2011; Wang et al., 2016); then I present an estimation procedure for the regression coefficients based on the B-Spline basis system.

3.2.1 The historical functional linear model

As mentioned above, previous studies has considered the general functional linear model $Y_i(t) = \sum_{k=1}^{K} \left[\int_{\mathcal{S}} X_{ki}(s)\beta_k(s,t)ds\right] + U_i(t)$, for all $t \in \mathcal{S}$, indicating that the response $Y_i(t)$ at any point of time $t \in \mathcal{S}$ depends on the predictors $X_{ki}(s)$ at all points of time $s \in \mathcal{S}$. Practically, it is however a common circumstance where the behavior of the response is only affected by the movements of the predictors from the past but not from the future; in other words, $\beta_k(s,t)$ is not defined for $s > t$, and the set $\mathcal{S}$ becomes a function of t, denoted $\mathcal{S}_t$, specifying when the effects of the predictors start to appear and vanish given the current time t. In this chapter, I specifically consider the model with $\mathcal{S}_t := [0,t]$, such that

$$Y_i(t) = \int_0^t \boldsymbol{X}_i{}^T(s)\boldsymbol{\beta}(s,t)ds + U_i(t),\ t \in [0,\bar{t}]\,,\ \boldsymbol{X}_i(s) = \begin{bmatrix} X_{1i}(s) \\ \vdots \\ X_{Ki}(s) \end{bmatrix},\ \boldsymbol{\beta}(s,t) = \begin{bmatrix} \beta_1(s,t) \\ \vdots \\ \beta_K(s,t) \end{bmatrix}, \quad (3.3)$$

where the response $Y_i(t)$, predictors $X_{ki}(s)$ and the error $U_i(t)$ are defined on $\mathcal{T} := [0,\bar{t}]$, and the coefficient surface $\beta_k(s,t)$ is defined on $\mathcal{T}_2 := \{(s,t) : 0 \le s \le t \le \bar{t}\}$. For all $k = 1,...,K$ and $i = 1,...,n$, $Y_i, X_{ki}, U_i \in \mathcal{C}^D(\mathcal{T})$ and $\beta_k(s,t) \in \mathcal{C}^D(\mathcal{T}_2)$ with $D \in \mathbb{N}$, where $\mathcal{C}^D(\mathcal{T})$ represents the space of all real-valued functions defined on $\mathcal{T}$ with continuous D-th order derivatives and $\mathcal{C}^D(\mathcal{T}_2)$ represents the space of all real-valued functions defined on $\mathcal{T}_2$ with continuous D-th order partial derivatives[1]; meanwhile, $\beta_k(s,t)$'s are square integrable on $\mathcal{T}_2$ such that $\int_0^{\bar{t}}\int_0^t \beta_k^2(s,t)dsdt < \infty$.

3.2.2 The functional estimators

By the Weierstrass approximation theorem, the functional quantities in Equation (3.3) can be uniformly approximated arbitrarily closely by polynomial functions, and the represen-

[1] A more general setup is to let β_k have different orders of continuous differentiability; to keep the notations simple, here I assume both dimensions have the same order, without loss of generality.

tations using basis expansions can be written as follow (Ramsay, 2005):

$$Y_i(t) = \sum_{h=1}^{\infty} a_{i,h}\eta_h(t) \approx \sum_{h=1}^{H} a_{i,h}\eta_h(t) =: \check{Y}_i(t), \tag{3.4}$$

$$X_{ki}(s) = \sum_{h=1}^{\infty} c_{k,i,l}\phi_l(s) \approx \sum_{l=1}^{L} c_{k,i,l}\phi_l(s) =: \check{X}_{ki}(s), \tag{3.5}$$

$$\beta_k(s,t) = \sum_{p=1}^{\infty}\sum_{q=1}^{\infty} b_{k,p,q}\psi_p(s)\theta_q(t) \approx \sum_{p=1}^{P}\sum_{q=1}^{Q} b_{k,p,q}\psi_p(s)\theta_q(t) =: \check{\beta}_k(s,t), \tag{3.6}$$

where $a_{i,h}$'s, $c_{i,l}$'s and $b_{k,p,q}$'s denote the expansion coefficients, $\eta_h(t)$'s, $\phi_l(s)$'s, $\psi_p(s)$'s and $\theta_q(t)$'s the expansion bases, and $H, L, P, Q \in \mathbb{N}$ the numbers of basis functions. As H, L, P and Q increase to infinity, the partial sums in (3.4) - (3.6) converge to $Y_i(t)$, $X_{ki}(s)$ and $\beta_k(s,t)$, respectively, for all s, t, and one can approximate these functional quantities arbitrarily closely by selecting proper numbers of basis functions. With the basis expansions of $\boldsymbol{\beta}(s,t)$, model (3.3) can be re-written as

$$Y_i(t) \approx \int_0^t \boldsymbol{X}_i{}^T(s)\boldsymbol{\Psi}(s)\boldsymbol{\Theta}(t)\boldsymbol{b}ds + U_i(t), \ \forall i = 1, ..., n, \tag{3.7}$$

where I define the following vectors and matrices for the basis expansions of the functional data.

$$\boldsymbol{b} = \begin{bmatrix} \boldsymbol{b}_1 \\ \vdots \\ \boldsymbol{b}_K \end{bmatrix}_{QPK\times 1}, \ \boldsymbol{b}_k = \begin{bmatrix} \boldsymbol{b}_{k,1} \\ \vdots \\ \boldsymbol{b}_{k,P} \end{bmatrix}_{QP\times 1}, \ \boldsymbol{b}_{k,p} = \begin{bmatrix} b_{k,p,1} \\ \vdots \\ b_{k,p,Q} \end{bmatrix}_{Q\times 1},$$

$$\boldsymbol{\Theta}(\cdot) = \begin{bmatrix} \boldsymbol{\theta}_P(\cdot) & & \\ & \ddots & \\ & & \boldsymbol{\theta}_P(\cdot) \end{bmatrix}_{PK\times QPK}, \ \boldsymbol{\theta}_P(\cdot) = \begin{bmatrix} \boldsymbol{\theta}^T(\cdot) & & \\ & \ddots & \\ & & \boldsymbol{\theta}^T(\cdot) \end{bmatrix}_{P\times QP}, \ \boldsymbol{\theta}(\cdot) = \begin{bmatrix} \theta_1(\cdot) \\ \vdots \\ \theta_Q(\cdot) \end{bmatrix}_{Q\times 1},$$

$$\boldsymbol{\Psi}(\cdot) = \begin{bmatrix} \boldsymbol{\psi}^T(\cdot) & & \\ & \ddots & \\ & & \boldsymbol{\psi}^T(\cdot) \end{bmatrix}_{K\times PK}, \boldsymbol{\psi}(\cdot) = \begin{bmatrix} \psi_1(\cdot) \\ \vdots \\ \psi_P(\cdot) \end{bmatrix}_{P\times 1}.$$

It is worth noting that while selecting the basis for $\boldsymbol{\beta}(s,t)$ in such a regression model, one needs to take into account the fact that $\boldsymbol{\beta}(s,t)$ is only defined for $s \leq t$. It is natural to adopt a basis that does not force $\boldsymbol{\beta}(s,t)$ to go off to the area where $s > t$. In the current chapter, I suggest to use B-spline bases for both dimensions of $\boldsymbol{\beta}(s,t)$, relying on the "local" property of the B-spline system. Without loss of generality, I define $\boldsymbol{\psi}$ and $\boldsymbol{\theta}$ each to be a B-spline basis of order D and has non-overlapping equally-spaced knots, denoted $\kappa_\psi := [0 = \kappa_0, ..., \kappa_{P-D} = \bar{t}]$ and $\kappa_\theta := [0 = \kappa_0, ..., \kappa_{Q-D} = \bar{t}]$, respectively. Then one can construct a set of polynomials $\Gamma(D, \kappa_\psi, \kappa_\theta)$ of order D, which is a two-dimensional analogue of the set $S(m, \underline{t})$ defined in Zhou et al. (1998), such that

$$\begin{aligned}\Gamma(D, \kappa_\psi, \kappa_\theta) = \big\{ & \gamma \in \mathcal{C}^{D-2}(\mathcal{T}_2) : D \geq 2, \\ & \gamma \text{ is a bivariate polynomial of order } D \text{ on } [\kappa_\iota, \kappa_{\iota+1}] \times [\kappa_{\iota'}, \kappa_{\iota'+1}] \\ & \text{for all } \iota = 0, ..., P-D-1 \text{ and } \iota' = 0, ..., Q-D-1 \text{ with } \kappa_\iota \leq \kappa_{\iota'+1} \big\},\end{aligned}$$

and an estimator $\hat{\boldsymbol{\beta}} \in \Gamma(D, \kappa_\psi, \kappa_\theta)$ for $\boldsymbol{\beta}$ can be defined as a penalized least squares minimizer

$$\hat{\boldsymbol{\beta}} := \min_{\substack{\gamma_k \in \Gamma(D, \kappa_\psi, \kappa_\theta) \\ \forall k = 1, ..., K}} \left\{ \frac{1}{n} \sum_{i=1}^{n} \int_0^{\bar{t}} \left[Y_i(t) - \sum_{k=1}^{K} \int_0^t X_{ki}(s) \gamma_k(s,t) ds \right]^2 dt + \lambda_\Psi \boldsymbol{\Lambda}_\Psi + \lambda_\Theta \boldsymbol{\Lambda}_\Theta \right\}, \quad (3.8)$$

where $\boldsymbol{\Lambda}_\Psi$ and $\boldsymbol{\Lambda}_\Theta$ denote some penalties for the two dimensions of $\hat{\boldsymbol{\beta}}$, associated with the tuning parameters λ_Ψ and λ_Θ. Here, I adopt a commonly used roughness penalty such that

$$\begin{aligned}\boldsymbol{\Lambda}_\Psi &:= \int_0^{\bar{t}} \int_0^t \left[\boldsymbol{\Psi}^{(2)}(s) \boldsymbol{\Theta}(t) \right] \left[\boldsymbol{\Psi}^{(2)}(s) \boldsymbol{\Theta}(t) \right]^T ds dt, \text{and} \\ \boldsymbol{\Lambda}_\Theta &:= \int_0^{\bar{t}} \int_0^t \left[\boldsymbol{\Psi}(s) \boldsymbol{\Theta}^{(2)}(t) \right] \left[\boldsymbol{\Psi}(s) \boldsymbol{\Theta}^{(2)}(t) \right]^T ds dt.\end{aligned}$$

Under the B-spline basis expansion specified in (3.6), the estimation of $\boldsymbol{\beta}$ as in (3.8) reduces to the estimation of $\boldsymbol{b}$, and the estimator, denoted $\hat{\boldsymbol{b}}$, can be obtained by minimizing the criterion

$$m\left(\tilde{\boldsymbol{b}}; P, Q, \{Y_i, \boldsymbol{X}_i\}_{i=1}^n\right) := \frac{1}{n} \sum_{i=1}^{n} \int_0^{\bar{t}} \left[Y_i(t) - \int_0^t \boldsymbol{X}_i{}^T(s) \boldsymbol{\Psi}(s) \boldsymbol{\Theta}(t) \tilde{\boldsymbol{b}} ds \right]^2 dt + \lambda_\Psi \boldsymbol{\Lambda}_\Psi + \lambda_\Theta \boldsymbol{\Lambda}_\Theta. \quad (3.9)$$

However, we do not directly observe Y_i's and X_{ki}'s but their "snapshots", denoted Y_{it_j}'s and X_{kit_j}'s for $j = 1, ..., J_Y$ or $j = 1, ..., J_X$, where J_Y and J_X denote the number of observation points for the response and the predictor variables, respectively. To simplify the notations, let $\{t_j\}_j$ denotes the set of time indices for either the response or the predictors, such that $t_1 = 0$ and $t_{J_Y} = t_{J_X} = \bar{t}$. Hence, to obtain $\hat{\boldsymbol{\beta}}$ (i.e., $\hat{\boldsymbol{b}}$), Y_i's and X_{ki}'s in (3.9) need to be replaced by their estimates. The coefficients in the basis expansions for Y_i's and X_{ki}'s can be estimated by solving the following least squares problems, based on their observations Y_{it_j}'s and X_{kit_j}'s:

$$(\tilde{a}_{i,1}, ..., \tilde{a}_{i,H}) := \underset{(r_{i,1},...,r_{i,H})}{\operatorname{argmin}} \frac{1}{J_Y} \sum_{j=1}^{J_Y} \left[Y_{it_j} - \sum_{h=1}^{H} r_{i,h} \eta_h(t_j) \right]^2, \tag{3.10}$$

$$(\tilde{c}_{k,i,1}, ..., \tilde{c}_{k,i,L}) := \underset{(r_{k,i,1},...,r_{k,i,L})}{\operatorname{argmin}} \frac{1}{J_X} \sum_{j=1}^{J_X} \left[X_{kit_j} - \sum_{l=1}^{L} r_{k,i,l} \phi_l(t_j) \right]^2. \tag{3.11}$$

The functional estimators can then be obtained as

$$\tilde{Y}_i(t) := \sum_{h=1}^{H} \tilde{a}_{i,h} \eta_h(t) = \boldsymbol{\eta}^T(t) \left[\int_0^{\bar{t}} \boldsymbol{\eta}(\tau) \boldsymbol{\eta}^T(\tau) d\tau \right]^{-1} \left[\frac{1}{J_Y} \sum_{j=1}^{J_Y} \boldsymbol{\eta}(t_j) Y_{it_j} \right],$$

$$\tilde{X}_{ki}(t) := \sum_{l=1}^{L} \tilde{c}_{k,i,l} \phi_l(t) = \boldsymbol{\phi}^T(t) \left[\int_0^{\bar{t}} \boldsymbol{\phi}(\tau) \boldsymbol{\phi}^T(\tau) d\tau \right]^{-1} \left[\frac{1}{J_X} \sum_{j=1}^{J_X} \boldsymbol{\phi}(t_j) X_{kit_j} \right],$$

where $\boldsymbol{\eta} := [\eta_1, ..., \eta_H]^T$ and $\boldsymbol{\phi} := [\phi_1, ..., \phi_L]^T$. Note that the choices of the bases as well as the functional fitting methods can affect the convergence behavior of the functional estimators. The asymptotic properties of the functional estimators have been established in literature utilizing variety of bases and fitting methods (see, e.g., Cox, 1983; Schwetlick and Kunert, 1993; Zhou et al., 1998; Speckman and Sun, 2003). In the current chapter, $\boldsymbol{\eta}$ and $\boldsymbol{\phi}$ are set to be B-spline bases of order D defined on $[0, \bar{t}]$ with non-overlapping equally-spaced knots. I adopt the regression spline method to obtain the fitted functional data as shown in (3.10) and (3.11). With proper numbers of basis functions, the functional estimators $\tilde{Y}_i$ and $\tilde{X}_{ki}$ are uniformly consistent and do not suffer from boundary effects asymptotically (e.g., Gasser and Müller, 1984; Zhou et al., 1998).

Under the fitted functions $\tilde{Y}_i(t)$'s and $\tilde{X}_{ki}$'s, the estimator $\hat{\boldsymbol{b}}$ shall write

$$\left[\int_0^{\bar{t}} \frac{1}{n}\sum_{i=1}^{n} \underline{\tilde{\boldsymbol{X}}}_i(t)\underline{\tilde{\boldsymbol{X}}}_i^{\,T}(t)dt + \lambda_\Psi \boldsymbol{\Lambda}_\Psi + \lambda_\Theta \boldsymbol{\Lambda}_\Theta\right]^{-1}\left[\int_0^{\bar{t}} \frac{1}{n}\sum_{i=1}^{n} \underline{\tilde{\boldsymbol{X}}}_i(t)\tilde{Y}_i(t)dt\right], \tag{3.12}$$

where $\underline{\tilde{\boldsymbol{X}}}_i^{\,T}(t) := \int_0^t \tilde{\boldsymbol{X}}_i^T(s)\boldsymbol{\Psi}(s)\boldsymbol{\Theta}(t)ds$. However, with the B-spline bases $\boldsymbol{\psi}$ and $\boldsymbol{\theta}$ of order D, the value of $\boldsymbol{\beta}$ at each given time point is approximated by a linear combination of D basis functions only, while the rest $P-D$ basis functions in $\boldsymbol{\psi}$ and the $Q-D$ in $\boldsymbol{\theta}$ are all irrelevant. As a result, in both matrices $\boldsymbol{\Psi}$ and $\boldsymbol{\Theta}$, the entries corresponding to the irrelevant basis functions will be zeros. Hence, one can drop all the basis functions that are irrelevant to the area of $s < t$, which reduces the number of the to-be-estimated $b_{k,p,q}$'s and change the expression in Equation (3.12) to the following

$$\hat{\boldsymbol{b}}^{(-)} := \left[\int_0^{\bar{t}} \frac{1}{n}\sum_{i=1}^{n} \underline{\tilde{\boldsymbol{X}}}_i^{\,(-)}(t)\underline{\tilde{\boldsymbol{X}}}_i^{\,(-)T}(t)dt + \lambda_\Psi \boldsymbol{\Lambda}_\Psi^{(-)} + \lambda_\Theta \boldsymbol{\Lambda}_\Theta^{(-)}\right]^{-1}\left[\int_0^{\bar{t}} \frac{1}{n}\sum_{i=1}^{n} \underline{\tilde{\boldsymbol{X}}}_i^{\,(-)}(t)\tilde{Y}_i(t)dt\right], \tag{3.13}$$

where $\underline{\tilde{\boldsymbol{X}}}_i^{\,(-)}$, $\boldsymbol{\Lambda}_\Psi^{(-)}$ and $\boldsymbol{\Lambda}_\Theta^{(-)}$ denote the $\underline{\tilde{\boldsymbol{X}}}_i$, $\boldsymbol{\Lambda}_\Psi$ and $\boldsymbol{\Lambda}_\Theta$ while replacing the matrices $\boldsymbol{\Psi}$ and $\boldsymbol{\Theta}$ with the ones after removing the columns and rows corresponding to the irrelevant basis functions. To reduce the notation load, I omit the superscript "$(-)$" hereafter while indicating the reduced matrices unless otherwise stated. The estimated functional coefficient can then be written as

$$\hat{\boldsymbol{\beta}}(s,t) = \boldsymbol{\Psi}(s)\boldsymbol{\Theta}(t)\hat{\boldsymbol{b}}. \tag{3.14}$$

It should be empathized that when I apply the B-spline bases for $\boldsymbol{\beta}$, this process of dropping basis functions (and thus, the number of estimates) is crucial, since otherwise, the matrix $\int_0^{\bar{t}} n^{-1}\sum_{i=1}^n \underline{\tilde{\boldsymbol{X}}}_i^{\,T}(t)\underline{\tilde{\boldsymbol{X}}}_i(t)dt + \lambda_s\boldsymbol{\Lambda}_s + \lambda_t\boldsymbol{\Lambda}_t$ in (3.12) will be singular and thus non-invertible even with the penalties. There are two reasons jointly contributing to this fact: first, the B-spline basis functions are only "locally" non-zero; second, $\mathcal{S}_t \subset \mathcal{S}_{t'} \subset \mathcal{T}$ for all $t < t' < \bar{t}$ — the two together will leave entire columns/rows of the matrix $\int_0^{\bar{t}} n^{-1}\sum_{i=1}^n \underline{\tilde{\boldsymbol{X}}}_i^{\,T}(t)\underline{\tilde{\boldsymbol{X}}}_i(t)dt + \lambda_s\boldsymbol{\Lambda}_s + \lambda_t\boldsymbol{\Lambda}_t$ zero.

3.3 Large Sample Theorems

In this section, I establish the asymptotics of the estimator $\hat{\boldsymbol{\beta}}(s,t)$ with cross-sectional dependence, denoted ρ. Essentially, ρ is a parameter in $[1/2, 1]$ indicating the level of cross-sectional dependence, such that $\left\|\sum_{i=1}^{n}\sum_{\iota=1}^{n}\mathbb{E}\left[\boldsymbol{X}_i(s)U_i(t)U_\iota(t')\boldsymbol{X}_\iota{}^T(s')\right]\right\|_F \in \mathcal{O}\left(n^{2\rho}\right)$ for all $(s,t),(s',t') \in \mathcal{T}_2$. When $\rho = 1/2$, the cross sectional dependence vanishes as $n \to \infty$, and the maximum level of cross sectional dependence allowed is for $\rho = 1$, whence we have $\left\|\sum_{i=1}^{n}\sum_{\iota=1}^{n}\mathbb{E}\left[\boldsymbol{X}_i(s)U_i(t)U_\iota(t')\boldsymbol{X}_\iota{}^T(s')\right]\right\|_F \in \mathcal{O}\left(n^2\right)$. First, let $\|\cdot\|_F$ represent a Frobenius norm for any M_1-by-M_2 matrix $\boldsymbol{M}$ (which in this case reduces to a vector L_2-norm) such that $\|\boldsymbol{M}\|_F = \sqrt{\sum_{m_1=1}^{M_1}\sum_{m_2=1}^{M_2}|\boldsymbol{M}_{m_1,m_2}|^2}$. In the current chapter, I use the asymptotic notations $\mathcal{O}$, o, $\mathcal{O}_p$ and o_p for the cases with matrices in any dimensions (i.e., scalars, vectors or higher dimensional matrices) indicating element-wise bounds or convergence, and I let the notations adjust to the conformable dimensions without specifying repeatedly. Then I state the following assumptions.

Assumption 3.3.1
Equation (3.3) holds for all $k = 1, ..., K$ *and* $i = 1, ..., n$, *where* $Y_i, X_{ki}, U_i \in \mathcal{C}^D(\mathcal{T})$, $\beta_k(s,t) \in \mathcal{C}^D(\mathcal{T}_2)$ *with* $K, D \in \mathbb{N}$ *and* $D \geq 2$; *also,* $\mathbb{E}[Y_i(t)] = \mathbb{E}[X_{ki}(s)] = 0$ *for all* $t, s \in [0, \bar{t}]$.

Assumption 3.3.1 states the model specification on the relationship between the functional response and the functional predictors, imposing smoothness over the functional terms as well as finitely many predictors. I also assume that processes Y_i's and X_{ki}'s are centered around zero, without loss of generality.

Assumption 3.3.2
$n, J_Y, J_X, H, L, P, Q \to \infty$; $H \in o(J_Y)$, $L \in o(J_X)$, $PQ \in o(n)$ *and* $[\min\{P,Q\}]^{-D} \in \mathcal{O}\left(n^{\rho-1}\right)$.

Assumption 3.3.2 states the divergence conditions for the sample size as well as the number of basis functions, some of which depend on the level of cross sectional dependence ρ. First of all, large samples over time and increasing numbers of basis functions H and L

lead to a consistent estimator for the functional data Y_i's and X_{ki}'s, then with such estimated functional data, consistency of the estimated functional coefficients can be achieved under sufficiently large numbers of basis functions P and Q. The condition $PQ \in o(n)$ suffices the asymptotic invertibility of matrices when needed. Note that with finite sample, the invertibility is satisfied by $PQK \leq n$. Moreover, as demonstrated in literature on asymptotics of functional data estimation with local polynomials, one component of the estimation error is associated with the step length between adjacent knots as well as the order of the polynomials. The constraint $[\min\{P, Q\}]^{-D} \in \mathcal{O}(n^{\rho-1})$ is to control the behavior of this error component.

Assumption 3.3.3
Given any $\bar{t} < \infty$, there exist estimators $\tilde{Y}_i, \tilde{X}_{ki} \in \mathcal{C}^D(\mathcal{T})$, such that as $n, J_Y, J_X \to \infty$,

$$n^{1-\rho} \sup_{t\in\mathcal{T}} \left\| \tilde{Y}_i(t) - Y_i(t) \right\|_F \in o_p(1) \text{ and } n^{1-\rho} \sup_{t\in\mathcal{T}} \left\| \tilde{X}_{ki}(t) - X_{ki}(t) \right\|_F \in o_p(1).$$

Assumption 3.3.3 imposes the existence of the uniformly consistent functional estimators for the response and the predictors under the given convergence conditions. These controls on the convergence rates of the estimated response and predictors guarantee that the asymptotics of the estimated functional coefficients $\boldsymbol{\beta}$ are not sensitive to the estimation errors from the procedure of functional fitting. One can justify this assumption by selecting the proper diverging parameters. For instance, under a commonly used B-spline basis setting with $D = 4$, one can obtain $\sup_{t\in\mathcal{T}} \left\| \tilde{Y}_i(t) - Y_i(t) \right\|_F \in \mathcal{O}_p\left(J_Y^{-4/9}\right)$ and $\sup_{t\in\mathcal{T}} \left\| \tilde{X}_{ki}(t) - X_{ki}(t) \right\|_F \in \mathcal{O}p\left(J_X^{-4/9}\right)$ by having $H \in \mathcal{O}\left(J_Y^{1/9}\right)$ and $L \in \mathcal{O}\left(J_X^{1/9}\right)$ (see, e.g., Zhou et al. (1998)). Then $n^{1-\rho} \in o\left(J^{4/9}\right)$ suffices the convergence conditions stated above.

Assumption 3.3.4
There exists a K-by-K full-rank positive-definite matrix of real-valued bivariate functions $\boldsymbol{\Sigma}_X(\cdot,\cdot) \in \mathcal{O}(1)$ such that for all $t, \tau \in \mathcal{T}$, $\left\| n^{-1} \sum_{i=1}^n \boldsymbol{X}_i(t) \boldsymbol{X}_i{}^T(\tau) - \boldsymbol{\Sigma}_X(t,\tau) \right\|_F \in o_p(1)$.

Assumption 3.3.4 corresponds to the conventional asymptotic full rank assumption for linear regression model. Note that this assumption can be satisfied by having the condition $PQ \in o(n)$ from Assumption 3.3.2.

Assumption 3.3.5
For all $i = 1, ..., n$ *and* $t \in \mathcal{T}$, $\mathbb{E}[U_i(t)] = 0$ *and* $\mathbb{E}\left[{U_i}^2(t)\right] < \infty$*; for* $(s,t) \in \mathcal{T}_2$, $n^{-1}\sum_{i=1}^{n}\mathbb{E}[\boldsymbol{X}_i(s)U_i(t)] = o(n^{\rho-1})$.

Assumption 3.3.5 states that for each individual i, the error term has zero mean and finite variance over time, and its correlation with the predictors both contemporaneously and across different time points is centered close to zero so that the average of such correlation across all individuals tends to zero at a given rate, i.e., $n^{-1}\sum_{i=1}^{n}\mathbb{E}[\boldsymbol{X}_i(s)U_i(t)] = o(n^{\rho-1})$.

Assumption 3.3.6
There exist a $K \times K$ *matrix of real-valued multivariate functions* $\boldsymbol{\Sigma}_{XU}(\cdot,\cdot,\cdot,\cdot) \in \mathcal{O}(1)$ *and a parameter* $\rho \in [1/2, 1]$ *indicating the level of cross-sectional dependence, such that for all* $(s,t),(s',t') \in \mathcal{T}_2$, $\left\|n^{-2\rho}\sum_{i=1}^{n}\sum_{\iota=1}^{n}\mathbb{E}\left[\boldsymbol{X}_i(s)U_i(t)U_\iota(t'){\boldsymbol{X}_\iota}^T(s')\right] - \boldsymbol{\Sigma}_{XU}(s,t,t',s')\right\|_F \in o(1)$.

While Assumption 3.3.5 indicates that $n^{-1}\sum_{i=1}^{n}\boldsymbol{X}_i(s)U_i(t)$ is centered close to zero, Assumption 3.3.6 further controls the order of $n^{-1}\sum_{i=1}^{n}\boldsymbol{X}_i(s)U_i(t)$ through its second moment. Such an order condition is essential in the achievement of the asymptotic results for $\hat{\boldsymbol{\beta}}$, especially with the appearance of the cross-sectional dependence.

Theorem 3.3.1 *Suppose Assumptions 3.3.1 to 3.3.6 hold, then for any* $0 < \bar{t} \in \mathbb{R}$*, the estimator* $\hat{\boldsymbol{\beta}}(s,t)$ *obtained from (3.13) - (3.14) over the domain* $\mathcal{T}_2$ *is uniformly consistent with the convergence rate* $\sup_{(s,t)\in\mathcal{T}_2}\left\|\hat{\boldsymbol{\beta}}(s,t) - \boldsymbol{\beta}(s,t)\right\|_F \in \mathcal{O}_p(n^{\rho-1})$.

Theorem 3.3.1 indicates the uniform convergence rate of the estimated functional coefficients $\hat{\boldsymbol{\beta}}$. Under the functional linear specification given by Assumption 3.3.1, $\hat{\boldsymbol{\beta}}$ has the expression as shown in (3.13) and (3.14). Meanwhile, for $\beta_k(s,t) \in \mathcal{C}^D(\mathcal{T}_2)$, there exists some $\gamma(s,t) \in \Gamma(D,P,Q)$ such that the estimation error can be decomposed into two components, $\hat{\boldsymbol{\beta}}(s,t) - \gamma(s,t)$ and $\gamma(s,t) - \boldsymbol{\beta}(s,t)$. As demonstrated in the proof, $\hat{\boldsymbol{\beta}}(s,t) - \gamma(s,t)$ tends to zero at the rate of $n^{\rho-1}$ under the boundedness and convergence conditions stated in Assumptions 3.3.2 to 3.3.6, while $\gamma(s,t) - \boldsymbol{\beta}(s,t)$ vanishes faster by the condition $[\min\{P,Q\}]^{-D} \in \mathcal{O}(n^{\rho-1})$ from Assumption 3.3.2. Then with the smoothness

conditions for $\boldsymbol{\beta}$ as well as the properties of the order-D B-spline bases for $\hat{\boldsymbol{\beta}}$, the uniform convergence result can be justified using Chebyshev's inequality. Note that the convergence rate of $\hat{\boldsymbol{\beta}}$ given in Theorem 3.3.1 is obtained base on the order conditions of parameters and sample size determined in the corresponding assumptions. These order conditions select the leading term in the estimation error by specifying the relative divergence rate among parameters. A different set of order conditions can result in a different leading term of the estimation error, which in turn may generate a different convergence rate for $\hat{\boldsymbol{\beta}}$.

As mentioned above, a uniform convergence result has been derived in Kim et al. (2011) with i.i.d. samples such that the estimation error is of order $\mathcal{O}_p\left(n^{-1/2}\left(h_X^{-2}+h_1^{-1}h_2^{-1}\right)\right)$, where h_X, h_1 and h_2 are the bandwidths used in obtaining the kernel smoothed covariance surfaces. On the other hand, one special case for the convergence result is when there is no cross-sectional dependence, i.e., when $\rho = 1/2$. Then a $\sqrt{n}$ convergence of $\hat{\boldsymbol{\beta}}$ can be obtained, which is a faster convergence than $\mathcal{O}_p\left(n^{-1/2}\left(h_X^{-2}+h_1^{-1}h_2^{-1}\right)\right)$. Intuitively, the expansion used by Kim et al. (2011) to represent their $\hat{\boldsymbol{\beta}}$ is based on the kernel smoothed covariance surfaces, and the error produced in this smoothing process is carried through the estimation of $\hat{\boldsymbol{\beta}}$. Corresponding to the estimator $\hat{\boldsymbol{\beta}}$, there is also such a component in the estimation error of $\hat{\boldsymbol{\beta}}$ coming from the process of functional fitting. However, as explained above, Assumptions 3.3.2 and 3.3.3 control the order of this component of error, so that it does not lead the estimation error of $\hat{\boldsymbol{\beta}}$, and hence, allows for this $\sqrt{n}$ convergence rate in the absence of cross-sectional dependence.

Assumption 3.3.7
$n, J_Y, J_X, H, L, P, Q\bar{t} \to \infty$, $H \in o(J_Y)$, $L \in o(J_X)$, $\bar{t}^2 \in o(\min\{J_Y, J_X\})$, $PQ \in o(n)$, $\bar{t}\left[\min\{P,Q\}\right]^{-D} \in \mathcal{O}(n^{\rho-1})$, such that there exist estimators $\tilde{Y}_i, \tilde{X}_{ki} \in \mathcal{C}^D(\mathcal{T})$, where

$$n^{1-\rho}\sqrt{\bar{J}}\sup_{t\in\mathcal{T}}\left\|\tilde{Y}_i(t)-Y_i(t)\right\|_F \in o_p(1) \text{ and } n^{1-\rho}\sqrt{\bar{J}}\sup_{t\in\mathcal{T}}\left\|\tilde{X}_{ki}(t)-X_{ki}(t)\right\|_F \in o_p(1),$$

for some $\bar{J}$ such that $\bar{t}^2/\bar{J} \in \mathcal{O}(1)$.

Assumption 3.3.7 upgrades the order conditions in Assumption 3.3.2 and the convergence rate conditions in Assumption 3.3.3 under large $\bar{t}$. This large $\bar{t}$ set up allows for further assumptions on the underlying functions, which will be explained below. Note that Assumption 3.3.3 was justified using the results from Zhou et al. (1998), which define the

functions over a fixed interval. Now with the interval $[0, \bar{t}]$ for $\bar{t} \to \infty$, I include the condition $\bar{t} \in o(\min\{J_Y, J_X\})$, so that as the range of the time period becomes wider and the number of observations increases, the observation number within any fixed interval of time increases as well. In this way, the convergence of $\tilde{Y}_i(t)$ and $\tilde{X}_{ki}(t)$ stated in Assumption 3.3.7 can still be justified by Zhou et al. (1998).

Assumption 3.3.8
There exist some $K \times K$ matrix of functions $\boldsymbol{C}_{X_iU_i}(\tau) \in \mathcal{O}(1)$, such that for some Lebesgue measure λ and any given $\tau \geq 0$ with $\tau \in o(\bar{t})$, there is

$$\int_{\mathcal{T}} \left\| \boldsymbol{C}_{X_iU_i}(\tau) - n^{-2\rho} \sum_{i=1}^{n} \sum_{\iota=1}^{n} \mathbb{E}\left[\boldsymbol{X}_i(t) U_i(t) U_\iota(t+\tau) \boldsymbol{X}_\iota{}^T(t+\tau) \right] \right\|_F d\lambda \in o(\bar{t}).$$

Assumption 3.3.8 together with Assumption 3.3.5 indicate that the term $X_i(t)U_i(t)$ becomes stationary as $\bar{t} \to \infty$ in that it has a near-zero first moment across the entire domain and an autocovariance function that is asymptotically almost surely coincide with $\boldsymbol{C}_{X_iU_i}(\tau)$ which does not depend on the point in time t. Assumption 3.3.8 also controls the level of the cross-sectional dependence, such that

$$\sum_{i=1}^{n} \sum_{\iota=1}^{n} \mathbb{E}\left[\boldsymbol{X}_i(t) U_i(t) U_\iota(t+\tau) \boldsymbol{X}_\iota{}^T(t+\tau) \right] \in \mathcal{O}\left(n^{2\rho}\right),$$

almost always over $\mathcal{T}$, as $\bar{t} \to \infty$. This assumption serves in the derivation of the asymptotic normality of $\hat{\boldsymbol{\beta}}$.

Assumption 3.3.9
For any two bounded functions $f : \mathbb{R} \to \mathbb{R}$ and $g : \mathbb{R} \to \mathbb{R}$, we have

$$\begin{aligned} &\left| \mathbb{E}\left[f\left(X_i(t)U_i(t)\right) g^T\left(X_i(t+\tau)U_i(t+\tau)\right) \right] \right| - \\ &\quad \left| \mathbb{E}\left[f\left(X_i(t)U_i(t)\right) \right] \right| \left| \mathbb{E}\left[g^T\left(X_i(t+\tau)U_i(t+\tau)\right) \right] \right| = \mathcal{O}\left(\tau^{-2\xi}\right), \end{aligned}$$

with $-\xi$ being the order of the mixing coefficient of the process $X_i(t)U_i(t)$, $\tau \in o(\bar{t})$ and $\tau, \bar{t} \to \infty$.

Assumption 3.3.9 states that the term $X_i(t)U_i(t)$ is ergodic in that the dependence of $X_i(t)U_i(t)$ at any two points in time is vanishing as the two points become further apart. This assumption also implies that with $\bar{m}_{U_i} := \bar{t}^{-1} \int_0^{\bar{t}} U_i(t)dt$, one has $\lim_{\bar{t}\to\infty} \mathrm{Var}\left(\bar{m}_{U_i}\right) = 0$, and there exists some $m_{U_i} \in \mathbb{R}$, such that $\lim_{\bar{t}\to\infty} \bar{m}_{U_i} = m_{U_i}$.

Theorem 3.3.2 *Suppose Assumptions 3.3.1 and 3.3.4 to 3.3.9 hold, then as* $\bar{t} \to \infty$, $\hat{\boldsymbol{\beta}}(s,t)$ *obtained from (3.13) - (3.14) is asymptotically normal in that*

$$\boldsymbol{V}_{\beta,\rho}{}^{-1/2}(s,t) n^{1-\rho}\sqrt{\bar{J}}\left[\hat{\boldsymbol{\beta}}(s,t) - \boldsymbol{\beta}(s,t)\right] \xrightarrow{d} N\left(\mathbf{0}, \boldsymbol{I}_K\right), \ \forall (s,t) \in \mathcal{T}_2,$$

where $\boldsymbol{V}_{\beta,\rho}(s,t) := \mathrm{Var}\left(n^{1-\rho}\sqrt{\bar{J}}\left[\hat{\boldsymbol{\beta}}(s,t) - \boldsymbol{\beta}(s,t)\right]\right) \in \mathcal{O}(1)$, *for some* $\bar{J}$ *such that* $\bar{t}^2/\bar{J} \in \mathcal{O}(1)$.

Recall that in Theorem 3.3.1, upon the consistency of the functional estimators for Y_i's and X_{ki}'s achieved through large J_Y and J_X, the $n^{1-\rho}$ consistency of the estimator $\hat{\boldsymbol{\beta}}$ is obtained on the fixed domain $[0,\bar{t}]$ through the enlargement of n. However, with the unknown form of cross-sectional dependence and interactions among different components of the estimation errors, normality cannot be achieved simply by increasing n. Theorem 3.3.2 states that if the time interval $[0,\bar{t}]$ extends to infinity as the number of observation time points increases, then under the stationarity and ergodicity conditions given in Assumptions 3.3.8 and 3.3.9, the asymptotic normality of $\hat{\boldsymbol{\beta}}$ can be achieved through the time dimension.

One important implication of Theorem 3.3.2 is that even with an enlarging sample size along the time dimension, the in-filling asymptotics itself does not lead to normality at the limit; rather that the time period needs to be long enough to reveal a repetitive and low-correlated pattern of the functions over time, then averaging over the time dimension can result in asymptotic normality by using some CLT for dependent processes. This result delivers a message that in order to apply the asymptotic normality of the estimated coefficients, instead of only increasing the observation frequency, one needs to extend the length of the time domain as well.

3.4 Bootstrap Methodology

As explained above, the asymptotic normality is achieved under large $\bar{t}$. However, with a finite time domain, or with a small sample size in general, the asymptotic theorem does not necessarily provide a good approximation to the distribution of $\hat{\boldsymbol{\beta}}$. In this section, I will develop a bootstrap method that outperforms the asymptotic theorem in approximating

the distribution of $\hat{\boldsymbol{\beta}}$, especially with finite time domains or small samples. The bootstrap method accommodates unknown forms of cross-sectional dependence that is either weak or strong, and it can be used to construct functional confidence intervals or perform hypothesis test for the estimated coefficient $\hat{\boldsymbol{\beta}}$.

3.4.1 Bootstrap procedure

The idea of the bootstrap is briefly summarized as follow. First, I obtain the consistent estimates of the error functions, denoted $\hat{U}_i(t)$; then I represent them using B-spline basis expansions. Under certain stationarity and ergodicity conditions, I adopt the idea of the MBB (see, e.g., Gonçalves, 2011; Kunsch, 1989; Liu and Singh, 1992) on the basis coefficients of the functional predictors and residuals and generate the bootstrap predictors and residuals, denoted $\boldsymbol{X}_i^*$'s and U_i^*'s respectively. From this process, I can obtain the corresponding bootstrap responses $Y_{it_j}^*$'s based on $\boldsymbol{X}_i^*$'s, U_i^*'s and the estimate $\hat{\boldsymbol{\beta}}$. With the pairs $\left\{Y_{it_j}^*, \boldsymbol{X}_{it_j}^*\right\}$'s, I can then obtain the bootstrap estimated functional coefficients $\hat{\boldsymbol{\beta}}^*(s,t)$. Such a bootstrap method captures both the time-wise smoothness and the cross-sectional dependence — while resampling blocks of the basis coefficients, the smoothness over time is imposed by the basis functions, and the cross-sectional dependence is preserved within the blocks.

Specifically, the bootstrap can be implemented as follow.
(B.i) Compute the residuals $\hat{U}_i(t) = \tilde{Y}_i(t) - \int_{\mathcal{S}_t} \tilde{\boldsymbol{X}}_i(s)\hat{\boldsymbol{\beta}}^T(s,t)ds$ for $i = 1, ..., n$.
(B.ii) Represent the residuals $\hat{U}_i(t)$ using the same basis as for $\tilde{Y}_i(t)$ and the residual values at the observing points $\hat{U}_i(t_j)$. Denoted these fitted residual functions by $\tilde{U}_i(t)$, such that $\tilde{U}_i(t) := \sum_{h=1}^{H} \tilde{w}_{i,h}\eta_h(t)$.
(B.iii) Let $\Delta_U \in \mathbb{N}$ denote the length of the blocks and $b_{\Delta_U,d}$ the dth size-Δ_U block of the basis coefficients, such that $b_{\Delta_U,d} := \{\tilde{\boldsymbol{w}}_{d+D-1}, ..., \tilde{\boldsymbol{w}}_{d+\Delta_U+D-2}\}$, where $\tilde{\boldsymbol{w}}_d = [\tilde{w}_{1,d}, ..., \tilde{w}_{n,d}]^T$. Then resample $\lceil (H - 2 * D + 2)/\Delta_{X,k} \rceil$ blocks with replacement from the set of overlapping blocks $b_{\Delta_U,1}, ..., b_{\Delta_U,H-\Delta_U-5}$. Truncate the resampled blocks in order to form the bootstrap basis coefficients of the original length $[w^*{}_{i,1}, ..., w^*{}_{i,H}]^T$. Then the bootstrap residuals can be expressed as $U_i^*(t) := \sum_{h=1}^{H} w^*{}_{i,h}\eta_h(t)$.
(B.iv) Let $\Delta_{X,k} \in \mathbb{N}$ denote the length of the blocks and $b_{\Delta_{X,k},d}$ the dth size-$\Delta_{X,k}$ block of the basis coefficients, such that $b_{\Delta_{X,k},d} := \left\{\tilde{\boldsymbol{c}}_{k,d+D-1}, ..., \tilde{\boldsymbol{c}}_{k,d+\Delta_{X,k}+D-2}\right\}$, where $\tilde{\boldsymbol{c}}_{k,d} = [\tilde{c}_{k,1,d}, ..., \tilde{c}_{k,n,d}]^T$. Resample $\lceil (L - 2 * D + 2)/\Delta_{X,k} \rceil$ blocks with replacement from the set

of overlapping blocks $b_{\Delta_{X,k},1}, ..., b_{\Delta_{X,k},L-\Delta_{X,k}-5}$, and truncate the resampled blocks to form the bootstrap basis coefficients $[{c^*}_{k,i,1}, ..., {c^*}_{k,i,L}]^T$. Then the bootstrap predictors write $X^*_{ki}(t) := \sum_{l=1}^{L} {c^*}_{k,i,l}\phi_l(t)$.
(B.v) For $i = 1, ..., n$, generate the bootstrap functional response, denoted $Y_i^*(t)$, such that

$$Y_i^*(t) = \int_0^t {\boldsymbol{X}_i^*}^T(s)\hat{\boldsymbol{\beta}}(s,t)ds + U_i^*(t),\ t \in [0, \bar{t}]\,, \tag{3.15}$$

where $\boldsymbol{X}_i^*(s) := [X^*_{1i}(s), ..., X^*_{Ki}(s)]^T$.
(B.vi) For $i = 1, ..., n$ and $j = 1, ..., J_Y$, generate the observation errors of the response $\varepsilon^*_{it_j} \sim i.i.d.\ \mathrm{N}(0, \sigma^2_{\varepsilon i})$ B times with $\sigma^2_{\varepsilon i} := J_Y^{-1}\sum_{j=1}^{J_Y}\left[Y_{it_j} - \tilde{Y}_i(t_j)\right]^2$, and generate B sets of observations for the response $Y^*_{it_j}$:

$$Y^*_{it_j} = Y_i^*(t_j) + \varepsilon^*_{it_j}. \tag{3.16}$$

(B,vii) Repeat the same procedure to obtain B sets of observations for the predictors $X^*_{kit_j}$.
(B.viii) Fit the functional response and predictors using B-spline basis expansion, denoted $\tilde{Y}_i^*(t)$ and $\tilde{\boldsymbol{X}}_i^*(t)$, and obtain the B bootstrap estimated coefficients $\hat{\boldsymbol{\beta}}^*(s,t)$, such that

$$\hat{\boldsymbol{b}}^* := \left[\int_0^{\bar{t}} \frac{1}{n}\sum_{i=1}^{n} \underline{\tilde{\boldsymbol{X}}}_i^*(t)\underline{\tilde{\boldsymbol{X}}}_i^{*T}(t)dt + \lambda_\Psi\boldsymbol{\Lambda}_\Psi + \lambda_\Theta\boldsymbol{\Lambda}_\Theta\right]^{-1}\left[\int_0^{\bar{t}} \frac{1}{n}\sum_{i=1}^{n} \underline{\tilde{\boldsymbol{X}}}_i^{*T}(t)\tilde{Y}_i^*(t)dt\right], \tag{3.17}$$

where $\underline{\tilde{\boldsymbol{X}}}_i^{*T}(t) := \int_0^t \tilde{\boldsymbol{X}}_i^{*T}(s)\boldsymbol{\Psi}(s)\boldsymbol{\Theta}(t)ds$, and thus, $\hat{\boldsymbol{\beta}}^*(s,t) = \boldsymbol{\Psi}(s)\boldsymbol{\Theta}(t)\hat{\boldsymbol{b}}^*$.

It is worth noting that for a B-spline basis of order D, the first and the last $D-1$ basis functions are not identical to the rest; therefore, while resampling the basis coefficients, I do not involve those coefficients that correspond to the basis functions at the two ends. For this reason, in steps (B.iii) and (B.iv), I start the blocks from the D-th basis functions and end at the D-th last one. Also, I again drop the basis functions over the domain of $\hat{\boldsymbol{\beta}}^*(s,t)$ where $s > t$, like I did in previous sections. Hence, the notations $\hat{\boldsymbol{b}}^*$ and $\underline{\tilde{\boldsymbol{X}}}_i^*(t)$ used in (3.17) denote the corresponding vector or matrix while replacing $\boldsymbol{\Psi}$ and $\boldsymbol{\Theta}$ with the ones after removing the columns and rows corresponding to the irrelevant basis functions.

3.4.2 Bootstrap validity

The bootstrap method I introduced above generalizes the MBB for longitudinal data that satisfies certain stationarity and ergodicity conditions to functional data. Intuitively, the

cross-sectional structure can be preserved by resampling the approximately independent moving blocks over time. However, with functional data, the difficulty is that such a rearrangement does not preserve the smoothness within functions. By using the "local" property of the B-spline representation for the functions, I discretize the smooth functions of stationary and ergodic properties onto vectors of basis coefficients, on which I can perform a longitudinal MBB.

Theorem 3.4.1 *Suppose Assumptions Assumptions 3.3.1 to 3.3.9 hold. With $\hat{\boldsymbol{\beta}}(s,t)$ obtained from (3.13) - (3.14), $\hat{\boldsymbol{\beta}}^*(s,t)$ from the bootstrap procedure (B.i) to (B.viii) and $l_J \in o\left(J^{1/2}\right)$, I have for $(s,t) \in \mathcal{T}_2$ and some $\bar{J}$ such that $\bar{t}^2/\bar{J} \in \mathcal{O}(1)$,*

$$\sup_{r\in\mathbb{R}^K} \left| \mathbb{P}^* \left(n^{1-\rho}\sqrt{\bar{J}} \left[\hat{\boldsymbol{\beta}}^*(s,t) - \hat{\boldsymbol{\beta}}(s,t) \right] \leq r \right) - \mathbb{P} \left(n^{1-\rho}\sqrt{\bar{J}} \left[\hat{\boldsymbol{\beta}}(s,t) - \boldsymbol{\beta}(s,t) \right] \leq r \right) \right| = o(1),$$

where $\mathbb{P}^*$ is induced by the bootstrap, conditional on the data.

Theorem 3.4.1 states that under the conditions sufficing a consistent estimator $\hat{\boldsymbol{\beta}}$ as well as stationarity and ergodicity in the predictors and errors, the bootstrap provides an asymptotically valid approximation to the distribution of $\hat{\boldsymbol{\beta}}$. The bootstrap statistic $n^{1-\rho}\sqrt{\bar{t}}\left[\hat{\boldsymbol{\beta}}^*(s,t) - \hat{\boldsymbol{\beta}}(s,t)\right]$ can also be used to construct percentile confidence intervals and perform hypothesis tests for the functional coefficients.

3.5 Simulation Analysis

In this section, I illustrate the estimation and bootstrap methods using a simulation study, and I demonstrate the consistency and asymptotic normality of the estimated coefficients $\hat{\boldsymbol{\beta}}$ as well as the validity of the bootstrap coefficients $\hat{\boldsymbol{\beta}}^*$ under different degrees of cross-sectional dependence.

3.5.1 Data generating process

Without loss of generality, I set $D = 4$, $\boldsymbol{\beta}(s,t) = 0$ and $K = 1$. The data for the simulation study is generated as follow.

(D.i) I construct a pseudo-continuous interval of $\mathcal{T} := [0, \bar{t}]$ consisting of 1001 equally-spaced points, denoted $\mathcal{T}^p$, and then I take $\mathcal{S}_t{}^p := [0, t] \cap \mathcal{T}^p$ for all $t \in \mathcal{T}$.
(D.ii) I generate n functional predictors X_i's using basis expansions with a degree of cross-sectional dependence controlled by a parameter ϱ and imposed through the basis coefficients. While ρ captures a more general sense of cross-sectional dependence, which can be in any unknown form, in the current simulation, a fixed correlation between curves is considered as indicated by ϱ, and this is just one type of the cross-sectional dependence. Specifically, I use the basis expansion $\sum_{l=1}^{L} c_{1,i,l}\phi_l(s)$, letting $\phi_l(s)$'s be order-four B-spline basis functions and $\boldsymbol{C}$ be a matrix of basis coefficients, such that $\boldsymbol{C} = \boldsymbol{\Sigma}_{\boldsymbol{C},1}\boldsymbol{\Sigma}_{\boldsymbol{C},2}$ and

$$\boldsymbol{C} = \begin{bmatrix} c_{1,1,1} & \cdots & c_{1,n,1} \\ \vdots & \cdots & \vdots \\ c_{1,1,L} & \cdots & c_{1,n,L} \end{bmatrix}.$$

I define $\boldsymbol{\Sigma}_{\boldsymbol{C},1}$ to be an L-by-L matrix of i.i.d. t-distribution with 2 degrees of freedom, and the rows of the L-by-n matrix $\boldsymbol{\Sigma}_{\boldsymbol{C},2}$ to be random vectors drawn from $N(\mathbf{0}, \boldsymbol{\Sigma}_{\varrho,X})$ with

$$\boldsymbol{\Sigma}_{\varrho,X} := \begin{bmatrix} r_0\sigma_1^2 & r_1\sigma_1\sigma_2 \cdots & r_{n-1}\sigma_1\sigma_n \\ r_1\sigma_2\sigma_1 & r_0\sigma_2^2 \cdots & r_{n-2}\sigma_2\sigma_n \\ \vdots & \ddots & \vdots \\ r_{n-1}\sigma_n\sigma_1 & r_{n-2}\sigma_n\sigma_2 \cdots & r_0\sigma_n^2 \end{bmatrix}.$$

Let $1 = r_0 > r_1 > ... > r_{\lfloor n^\varrho \rfloor} = 0.1$ denote an array of descending equally-spaced values on $[1, 0.1]$ followed by $r_{\lfloor n^\varrho \rfloor} + 1 = ... = r_{n-1} = 0$, indicating the correlation coefficients[2]. Meanwhile, I obtain $\sigma_i^2 \sim \chi^2(1)$ and $\sigma_i := \sqrt{\sigma_i^2}$ for $i = 1, ..., n$, indicating the variances and the standard deviations respectively.
(D.iii) I generate U_i's in the same way as for X_i's, with the basis expansion $\sum_{h=1}^{H} w_{i,h}\eta_h(s)$, where $H = L$, $\eta_h(s)$'s are order-four B-spline basis functions and $\boldsymbol{W} = \boldsymbol{\Sigma}_{\boldsymbol{W},1}\boldsymbol{\Sigma}_{\boldsymbol{W},2}$ where

$$\boldsymbol{W} = \begin{bmatrix} w_{1,1} & \cdots & w_{n,1} \\ \vdots & \cdots & \vdots \\ w_{1,H} & \cdots & w_{n,H} \end{bmatrix}.$$

The matrices $\boldsymbol{\Sigma}_{\boldsymbol{W},1}$, $\boldsymbol{\Sigma}_{\boldsymbol{W},1}$ and $\boldsymbol{\Sigma}_{\varrho,U}$ are defined in the same way as $\boldsymbol{\Sigma}_{\boldsymbol{C},1}$, $\boldsymbol{\Sigma}_{\boldsymbol{C},1}$ and $\boldsymbol{\Sigma}_{\varrho,X}$, while I use different notations to indicate that the ones for U_i's are independently generated.

[2] $\lfloor n^\varrho \rfloor$ represents the largest integer that is smaller than n^ϱ.

(D.iv) With the functional predictors, coefficient and error terms generated from previous steps, I can obtain n functional responses according to the specification in (3.2).
(D.v) I draw equally-spaced discrete observations of the response and the predictor on $\{t_j\}_{j=1}^{J_Y}$ and $\{t_j\}_{j=1}^{J_X}$ with observational errors $\varepsilon_{it_j} \sim i.i.d.\ \mathrm{N}(0, \sigma_\varepsilon^2)$ for the response and $\epsilon_{it_j} \sim i.i.d.\ \mathrm{N}(0, \sigma_\epsilon^2)$ for the predictor, obtaining Y_{it_j}'s and X_{it_j}'s, where σ_ε^2 is set to be 1% of the variance of Y_i and σ_ϵ^2 is set to be 1% of the variance of X_i.

I can obtain B sets of observations for Y_i's and X_i's by repeating steps (D.1) to (D.v) B times. In this simulation study, I set $B = 199$, and I look at the cases where $\varrho \in \{0.1, 0.5, 0.9\}$ under three different sample sizes $(J, n) \in \{(51, 50), (101, 80), (251, 130)\}$ respectively, with $J_Y = J_X = J$. The values for $\bar{t}$, H and L will be determined in the following discussion.

3.5.2 Simulation results

With the B sets of simulated data, one can obtain B estimated functional coefficients $\hat{\boldsymbol{\beta}}$'s. In the current simulation, I set the parameters for the functional estimators according to the order conditions stated in the assumptions. Also, since for B-spline basis expansion of order D, every point in the functional data is spanned by D consecutive basis functions, I let the block size be $D = 4$ and the blocks overlap with a 1 step of jump, so that I keep together the basis functions that span every single point of the functional data. Then I first demonstrate the consistency of $\hat{\boldsymbol{\beta}}$ using the following statistic:

$$R^2_{\beta,\varrho} = \mathbb{E}\left[\sup_{(s,t)\in\mathcal{T}_2} \left| \left[\hat{\boldsymbol{\beta}}_\varrho(s,t) - \boldsymbol{\beta}(s,t)\right]^T \left[\hat{\boldsymbol{\beta}}_\varrho(s,t) - \boldsymbol{\beta}(s,t)\right] \right|\right], \tag{3.18}$$

where $\hat{\boldsymbol{\beta}}_\varrho(s,t)$ denotes the estimated functional coefficients $\hat{\boldsymbol{\beta}}(s,t)$ obtained under the cross-sectional dependence of degree ϱ specifically. The statistic $R^2_{\beta,\varrho}$ measures the size of the maximal error of estimated functional coefficients over the domain, where the expectation is approximated by averaging across all 199 sets of simulated samples. When the estimator $\hat{\boldsymbol{\beta}}_\varrho(s,t)$ is consistent, I expect the statistic $R^2_{\beta,\varrho}$ to vanish as $n, J \to \infty$. The statistics under different sample sizes and different degrees of cross-sectional dependence are shown in Table 3.1. The results indicate that in general, the estimation error is decreasing as the sample size increases, and the estimation error is generally smaller when the cross-sectional dependence is weaker.

Table 3.1: Consistency

		Fixed $\bar{t}$			$\bar{t} \to \infty$				
J	n	$\varrho = 0.1$	$\varrho = 0.5$	$\varrho = 0.9$	$\bar{t}$	H (or L)	$\varrho = 0.1$	$\varrho = 0.5$	$\varrho = 0.9$
51	50	0.1808	0.0980	0.1455	100	43	0.0938	0.1126	0.1949
101	80	0.1087	0.2603	0.2176	200	83	0.0489	0.0456	0.0461
251	130	0.0965	0.1586	0.1723	400	163	0.0046	0.0141	0.0097

Note: J indicates the number of time observations, n the number of individuals, ϱ the cross sectional dependence, H and L the numbers of basis functions of $Y_i(t)$ and $X_{ki}(s)$ respectively, and $\bar{t}$ the upper bound of the considered time domain.

Figures 3.1 to 3.4 present the comparison between the underlying and estimated functional coefficients with $(\varrho = 0.9, J = 51, n = 50)$ and $(\varrho = 0.9, J = 251, n = 130)$ with infilling asymptotics as examples. Specifically, Figures 3.1 and 3.3 plot the true coefficient surface (left) versus the estimated surface (right) from a set of randomly selected data out of the 199 simulated data sets. Figures 3.2 and 3.4 then chooses four time points on $[0, \bar{t}]$, namely $t^\star = \{t_{25}, t_{50}, t_{75}, t_{100}\}$, and plots the true coefficient $\boldsymbol{\beta}(s, t^\star)$ (the black dashed lines) versus the estimated coefficients $\hat{\boldsymbol{\beta}}(s, t^\star)$'s from all 199 simulated data sets (the grey solid lines).

Second, I show the asymptotic normality of the estimated functional coefficients using the Kolmogorov-Smirnov (K-S) Test. Specifically, in both settings with a fixed or an enlarging $\bar{t}$, I compare the statistic $\hat{\boldsymbol{V}}_{\beta,\rho}^{-1/2}(s,t) n^{1-\rho}\sqrt{\bar{t}} \left[\hat{\boldsymbol{\beta}}(s,t) - \boldsymbol{\beta}(s,t)\right]$ with a standard normal distribution through the K-S test over the domain $\mathcal{T}_2$ pointwisely and report the probability of rejection, where $\hat{\boldsymbol{V}}_{\beta,\rho}^{-1/2}(s,t)$ denotes the empirical variance of $\hat{\boldsymbol{\beta}}_\rho(s,t)$ obtained from the DGP. If the estimator $\hat{\boldsymbol{\beta}}_\rho(s,t)$ is asymptotically normally distributed, the rejection rates of the K-S test are expected to decline as the sampel size increases.

As shown in Table 3.2, with and enlarging $\bar{t}$, the rejection rates of the K-S test generally decreases as the sample size increases. However, with fixed $\bar{t}$, the rejection rates stay relatively high despite of the increase in sample size. This results demonstrate the point I made earlier that with a domain of a fix length, the stationarity or the ergodicity of the functions do not present in the data; hence, even with large sample, averaging over the time dimension does not necessarily result in asymptotic normality.

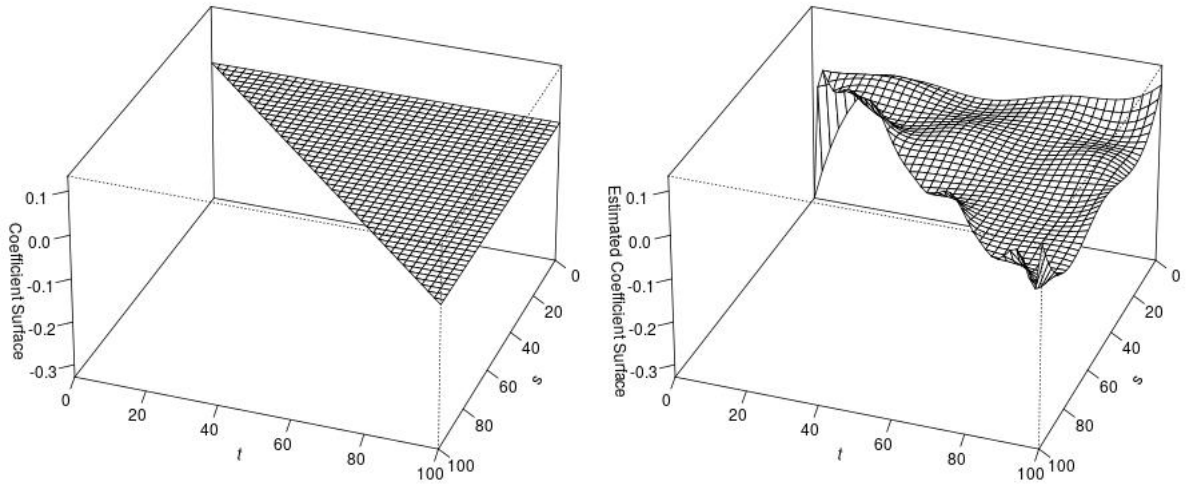


Figure 3.1: $\beta(s,t)$ (left) vs. $\hat{\beta}(s,t)$ (right); $\varrho = 0.9$, $J = 51$, $n = 50$

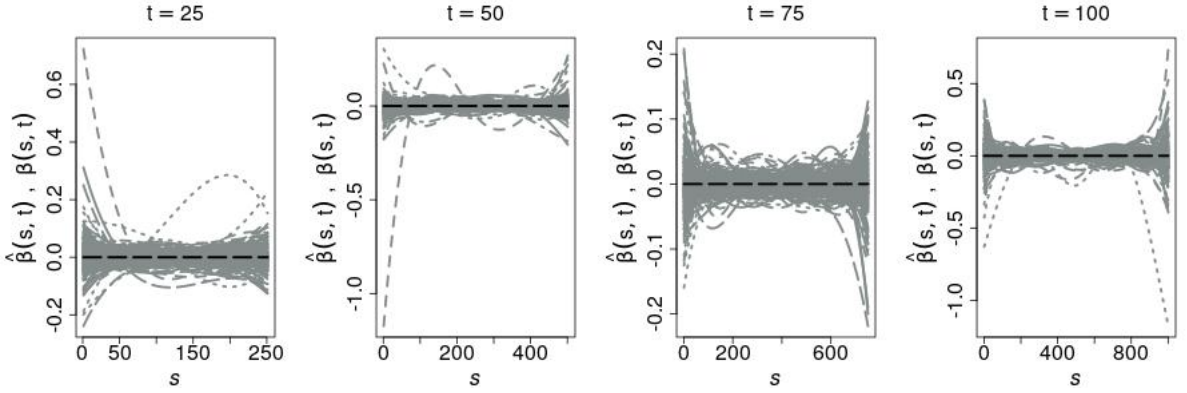


Figure 3.2: $\beta(s,t)$ (black dashed lines) vs. $\hat{\beta}(s,t)$ (grey solid lines) at fixed t; $\varrho = 0.9$, $J = 51$, $n = 50$

Then I present the bootstrap for the functional coefficients. I again randomly select one set of simulated data and the corresponding estimate $\hat{\beta}(s,t)$. Then I compare between the settings of $(\varrho = 0.9, J = 51, n = 50)$ and $(\varrho = 0.9, J = 251, n = 130)$ with in-filling asymptotics in Figures 3.5 to 3.8 as examples. Furthermore, similarly to the tests on asymptotic normality, the distributions of the bootstrapped coefficients and the estimated coefficients

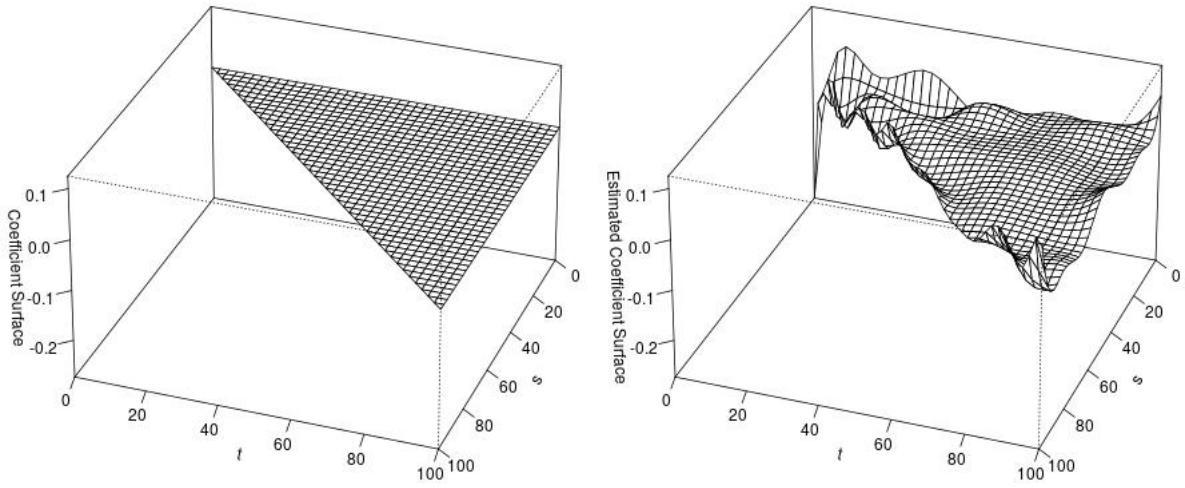


Figure 3.3: $\beta(s,t)$ (left) vs. $\hat{\beta}(s,t)$ (right); $\varrho = 0.9$, $J = 251$, $n = 130$

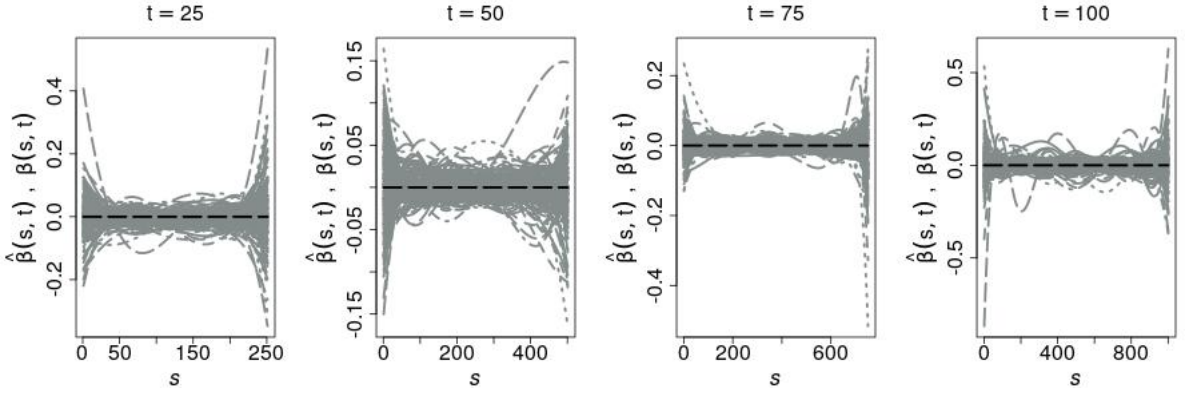


Figure 3.4: $\beta(s,t)$ (black dashed lines) vs. $\hat{\beta}(s,t)$ (grey solid lines) at fixed t; $\varrho = 0.9$, $J = 251$, $n = 130$

from the simulated data are compared through K-S tests over the domain $\mathcal{T}_2$ pointwisely; the probability of rejection rates are shown in Table 3.3 – a lower (higher) rejection rate indicates the bootstrap distribution provides a better (worse) approximation to the distribution of the estimated coefficients. Comparing the rejection rates of the K-S tests on asymptotic normality and on bootstrap validity, I can see that the bootstrap approx-

Table 3.2: Asymptotic Normality

		Fixed $\bar{t}$			$\bar{t} \to \infty$				
J	n	$\varrho = 0.1$	$\varrho = 0.5$	$\varrho = 0.9$	$\bar{t}$	H (or L)	$\varrho = 0.1$	$\varrho = 0.5$	$\varrho = 0.9$
51	50	0.6806	0.6843	0.4030	100	43	0.5407	0.6614	0.7964
101	80	0.5275	0.4947	0.6711	200	83	0.3757	0.4521	0.2852
251	130	0.5832	0.6304	0.4267	400	163	0.3085	0.4290	0.2368

Note: J indicates the number of time observations, n the number of individuals, ϱ the cross sectional dependence, H and L the numbers of basis functions of $Y_i(t)$ and $X_{ki}(s)$ respectively, and $\bar{t}$ the upper bound of the considered time domain.

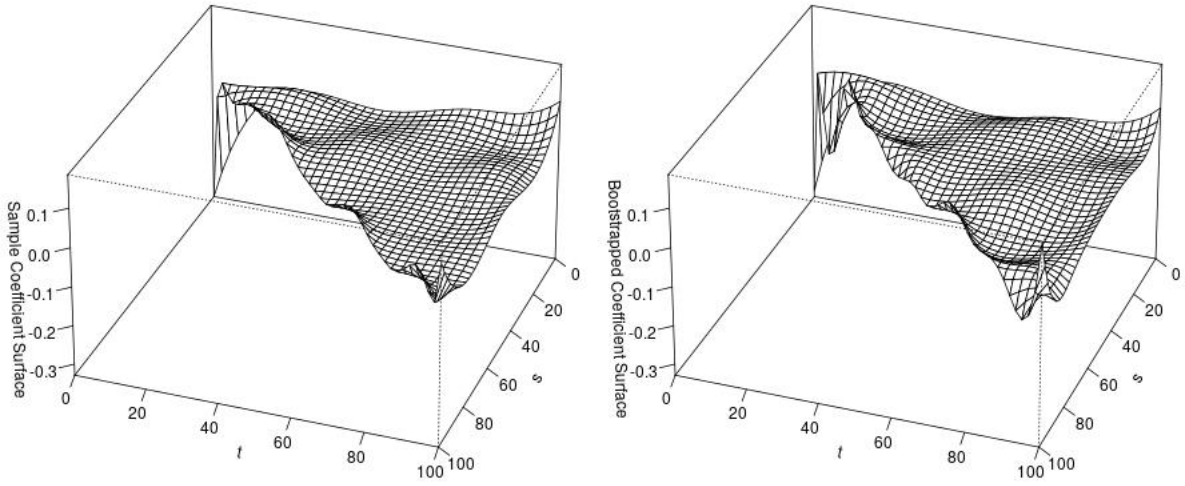


Figure 3.5: $\hat{\boldsymbol{\beta}}^*(s,t)$ (left) vs. $\hat{\beta}(s,t)$ (right); $\varrho = 0.9$, $J = 51$, $n = 50$

imation outperforms the normal approximation to the true distribution of the estimated functional coefficients, under both fixed $\bar{t}$ or large $\bar{t}$. Also, the performance of the bootstrap approximation is relatively stable with either finite or large $\bar{t}$.

Finally, in Figures 3.9 and 3.10, I present the p-values versus the 5% significance level over time from the K-S tests on asymptotic normality (the first plots in both figures) and on bootstrap validity (the second plot in both figures). I again choose four time points $t = \{t_{25}, t_{50}, t_{75}, t_{100}\}$ on $[0, \bar{t}]$, and plots the p-values over the corresponding time intervals

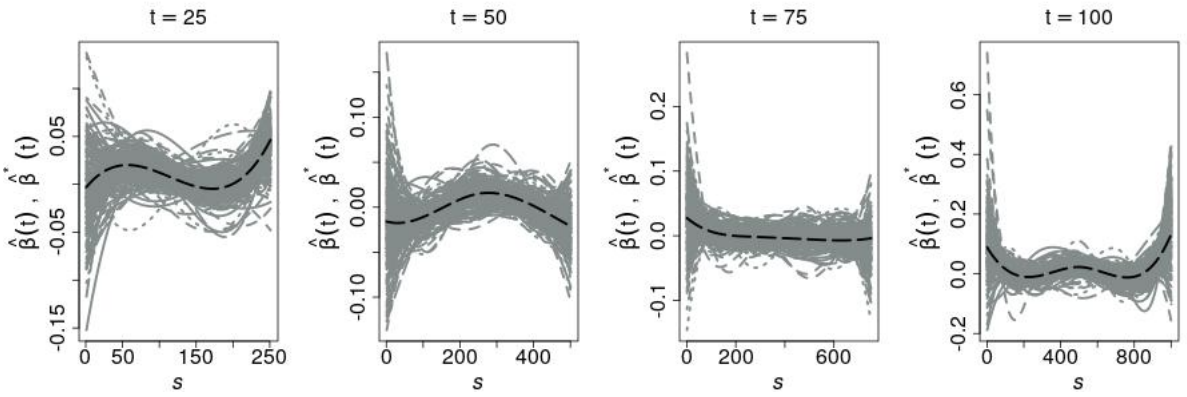


Figure 3.6: $\hat{\boldsymbol{\beta}}^*(s,t)$ (black dashed lines) vs. $\hat{\beta}(s,t)$ (grey solid lines) at fixed t; $\varrho = 0.9$, $J = 51$, $n = 50$

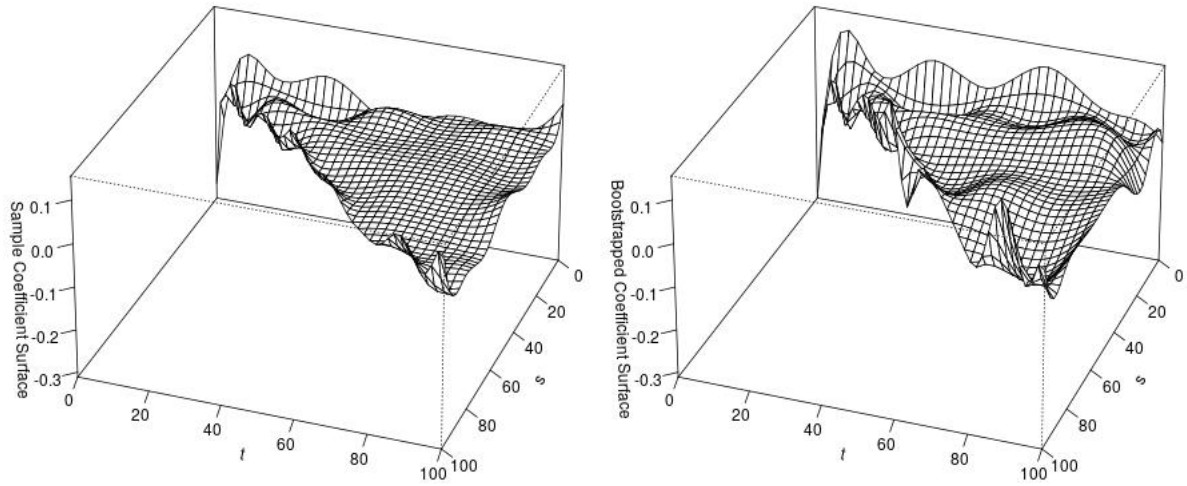


Figure 3.7: $\hat{\boldsymbol{\beta}}^*(s,t)$ (left) vs. $\hat{\beta}(s,t)$ (right); $\varrho = 0.9$, $J = 251$, $n = 130$

(s,t) for all $s \in [0,t]$. Figure 3.9 shows that when the sample size is relatively small, the K-S test on asymptotic normality is mostly rejected, indicating that the standard normal distribution cannot provide a good approximation to the distribution of the estimated coefficients, while the K-S test for the bootstrap approximation is mostly not rejected, indicating that the bootstrap distribution formed by $\hat{\boldsymbol{\beta}}^*(s,t)$'s provides a much better

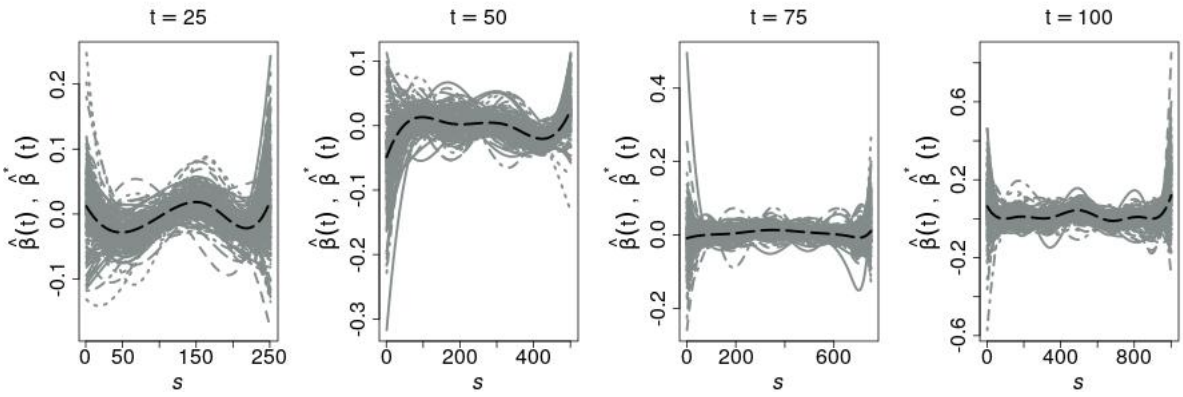


Figure 3.8: $\hat{\boldsymbol{\beta}}^*(s,t)$ (black dashed lines) vs. $\hat{\beta}(s,t)$ (grey solid lines) at fixed t; $\varrho = 0.9$, $J = 251$, $n = 130$

Table 3.3: Bootstrap Validity

		Fixed $\bar{t}$			$\bar{t} \to \infty$				
J	n	$\varrho = 0.1$	$\varrho = 0.5$	$\varrho = 0.9$	$\bar{t}$	H (or L)	$\varrho = 0.1$	$\varrho = 0.5$	$\varrho = 0.9$
51	50	0.1835	0.4102	0.1883	100	43	0.1256	0.2741	0.4384
101	80	0.1837	0.1642	0.3797	200	83	0.2194	0.2570	0.0914
251	130	0.1275	0.0992	0.0944	400	163	0.1468	0.2378	0.0914

Note: J indicates the number of time observations, n the number of individuals, ϱ the cross sectional dependence, H and L the numbers of basis functions of $Y_i(t)$ and $X_{ki}(s)$ respectively, and $\bar{t}$ the upper bound of the considered time domain.

approximation. While Figure 3.10 shows the test results with a large sample — both the asymptotic normality and the bootstrap have better performance in terms of the p-values from the K-S tests. Another interesting phenomenon is that the approximation by the asymptotic normality seems to be better in the later stage (when $t = \{t_{100}\}$) than the earlier one (when $t = \{t_{25}, t_{50}\}$). One explanation is that, in the hFLM, estimation with B-spline-based estimated functional coefficients uses fewer data points while t is small than when t is large; hence in general, the asymptotic normality forms faster in the later stage, i.e., when t is large.

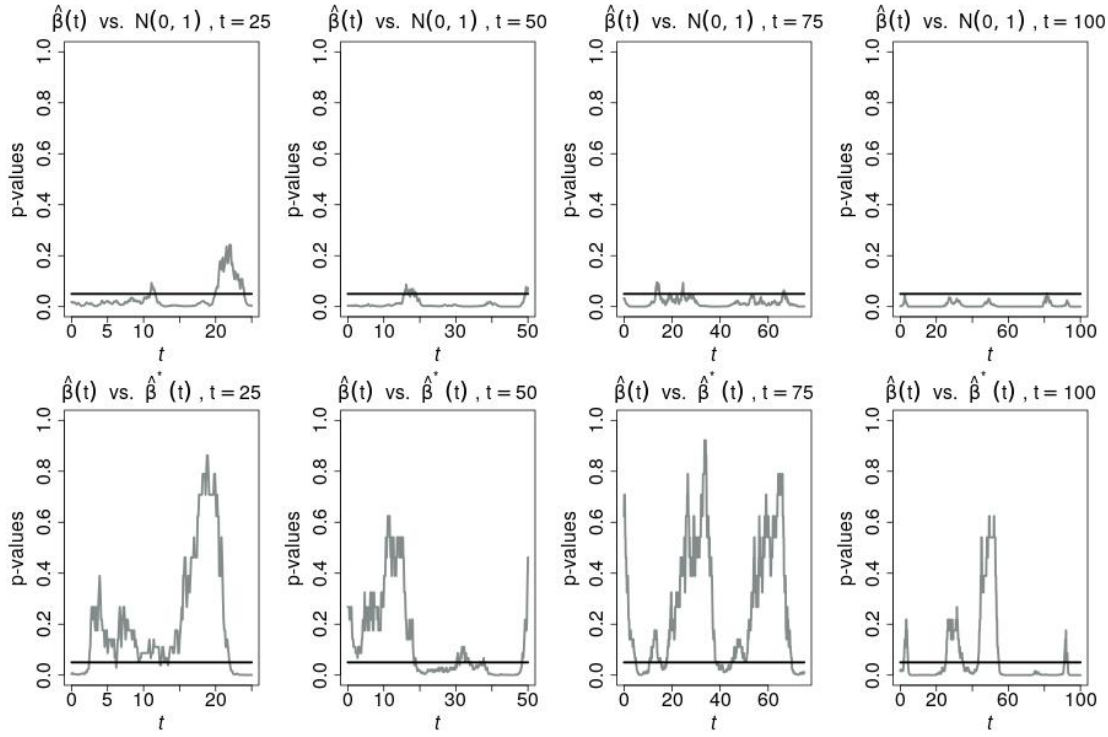


Figure 3.9: p-values of the K-S tests; $\varrho = 0.9$, $J = 51$, $n = 50$

3.6 Conclusion

In this chapter, I consider the historical functional linear model for longitudinal data with unknown cross-sectional dependence. This chapter contributes to the literature by establishing the asymptotics of the B-spline-based estimated functional coefficients and developing the bootstrap inference, accommodating unknown forms of cross-sectional dependence. The main findings are (i) a uniform convergence rate of the estimated functional coefficients is derived depending on the degree of cross-sectional dependence and $\sqrt{n}$-consistency can be achieved in the absence of cross-sectional dependence, which is a faster convergence than the estimators proposed in Kim et al. (2011); (ii) under proper stationary and ergodicity conditions on the functional variables, asymptotic normality of the estimated coefficients can be obtained with unknown forms of cross-sectional dependence; (iii) the proposed bootstrap method has a better finite-sample performance than the asymptotics

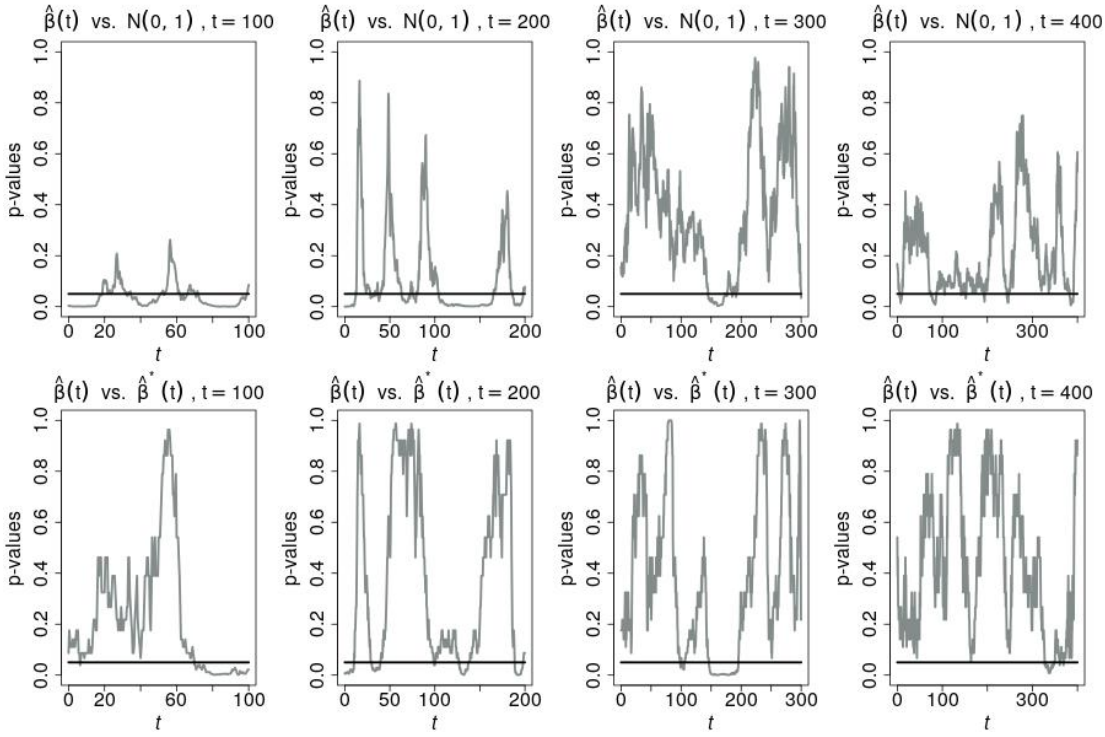


Figure 3.10: p-values of the K-S tests; $\varrho = 0.9$, $J = 251$, $n = 130$

while approximating the distribution of the estimated functional coefficients, and it can be used to construct percentile confidence intervals and perform hypothesis tests for the functional coefficients. I also provide simulation analysis to illustrate the estimation and bootstrap procedures and to demonstrate the properties of the estimators. Moreover, I also demonstrate through the theorem as well as the simulation that even with an enlarging sample size along the time dimension, the in-filling asymptotics does not necessarily lead to normality at the limit. The time period needs to be long enough to reveal the stationarity and ergodicity of the functions, then averaging over the time dimension can result in asymptotic normality. To my best knowledge, this study is the first to discuss the asymptotic and bootstrap inferences for the functional coefficients in hFLMs with cross-sectional dependence.

There are several aspects in the current study that motivate future research. For example, a bootstrap approach provides a practical alternative for estimating the distribution

of statistics when the the analytical results become unavailable or too complicated to achieve; in some circumstances, a transformation operator can also provide an alternative by transforming the statistics of interest into ones that are asymptotically distribution-free. Then the question becomes, for the historical functional linear model with unknown cross-sectional dependence, does there exist such a transformation scheme that can turn the statistic $\boldsymbol{V}_{\beta,\rho}{}^{-1/2}(s,t)n^{1-\rho}\sqrt{\bar{t}}\left[\hat{\boldsymbol{\beta}}(s,t)-\boldsymbol{\beta}(s,t)\right]$ into one following a known distribution, which can also be used to construct percentile confidence intervals or perform hypothesis tests; if the answer is yes, then how would the properties of such transformed statistics compare to the properties of bootstrap statistics. Moreover, endogeneity is a very common issue in regression analysis. Benatia et al. (2017) address the situation with endogeneity in a functional linear regression model with bivariate functional coefficients, providing asymptotics of the estimators. A further study along this line is to develop the estimators and their properties in hFLMs when endogeneity is involved, or furthermore, to construct specification test to detect endogenous predictors under functional setups.

References

Agarwal, G. G. and W. Studden (1980): "Asymptotic integrated mean square error using least squares and bias minimizing splines," *The Annals of Statistics*, 1307–1325.

Aït-Sahalia, Y. (2002), "Maximum Likelihood Estimation of Discretely Sampled Diffusions: A Closed-Form Approximation Approach," *Econometrica*, 70, 223–262.

Bai, J. and S. Ng (2002), "Determining the Number of Factors in Approximate Factor Models," *Econometrica*, 70, 191–221.

——— (2007), "Determining the Number of Primitive Shocks in Factor Models," *Journal of Business & Economic Statistics*, 25, 52–60.

——— (2008), "Large Dimensional Factor Analysis," *Foundations and Trends® in Econometrics*, 3, 89–163.

Banerjee, A., M. Marcellino, and I. Masten (2008), "Chapter 4 Forecasting Macroeconomic Variables Using Diffusion Indexes in Short Samples with Structural Change," in *Forecasting in the Presence of Structural Breaks and Model Uncertainty*, Emerald Group Publishing Limited, 149–194.

Bapna, R., W. Jank, and G. Shmueli (2008): "Price formation and its dynamics in online auctions," *Decision Support Systems*, 44, 641–656.

Benatia, D., M. Carrasco, and J.-P. Florens (2017): "Functional linear regression with functional response," *Journal of econometrics*, 201, 269–291.

Breitung, J. and S. Eickmeier (2006), "Dynamic Factor Models," *Allgemeines Statistisches Archiv*, 90, 27–42.

——— (2011), "Testing for Structural Breaks in Dynamic Factor Models," *Journal of Econometrics*, 163, 71–84.

Breton, T. R. (2012): "Penn World Table 7.0: Are the data flawed?" *Economics Letters*, 117, 208–210.

Cardot, H., F. Ferraty, and P. Sarda (1999): "Functional linear model," *Statistics & Probability Letters*, 45, 11–22.

——— (2003): "Spline estimators for the functional linear model," *Statistica Sinica*, 571–591.

Chen, T., J. DeJuan, and R. Tian (2018): "Distributions of GDP across versions of the Penn World Tables: A functional data analysis approach," *Economics Letters*, 170, 179–184.

Chen, T. and Q. Fan (2018): "A functional data approach to model score difference process in professional basketball games," *Journal of Applied Statistics*, 45, 112–127.

Chen, T., Y. Li, and R. Tian (Working Paper.b): "A Tale of Two Continuous Limits,".

Chen, T., R. Tian, and J. Xu (Working Paper.a): "Functional dynamic factor models,".

Chui, C. K. (1971), "Concerning Rates of Convergence of Riemann Sums," *Journal of Approximation Theory*, 4, 279–287.

Claeskens, G., T. Krivobokova, and J. D. Opsomer (2009): "Asymptotic properties of penalized spline estimators," *Biometrika*, 96, 529–544.

Cox, D. D. (1983): "Asymptotics for M-type smoothing splines," *The Annals of Statistics*, 530–551.

Cuevas, A., M. Febrero, and R. Fraiman (2006): "On the use of the bootstrap for estimating functions with functional data," *Computational statistics & data analysis*, 51, 1063–1074.

Cuevas, A. and R. Fraiman (2004): "On the bootstrap methodology for functional data," in *COMPSTAT 2004—Proceedings in Computational Statistics*, Springer, 127–135.

De Castro, B. F., S. Guillas, and W. G. Manteiga (2005): "Functional samples and bootstrap for predicting sulfur dioxide levels," *Technometrics*, 47, 212–222.

Dehling, H., O. S. Sharipov, and M. Wendler (2015): "Bootstrap for dependent Hilbert space-valued random variables with application to von Mises statistics," *Journal of Multivariate Analysis*, 133, 200–215.

Del Negro, M. and C. Otrok (2008), "Dynamic Factor Models with Time-Varying Parameters: Measuring Changes in International Business Cycles," *Federal Reserve Bank of New York Staff Reports, No. 326.*

Durá, J. V., J. M. Belda, R. Poveda, Á. Page, J. Laparra, J. Das, J. Prat, and A. C. García (2010): "Comparison of functional regression and nonfunctional regression approaches to the study of the walking velocity effect in force platform measures," *Journal of applied biomechanics*, 26, 234–239.

Fan, J. and W. Zhang (1999): "Statistical estimation in varying coefficient models," *The annals of Statistics*, 27, 1491–1518.

Forni, M., M. Hallin, M. Lippi, and L. Reichlin (2000), "The Generalized Dynamic-Factor Model: Identification and Estimation," *Review of Economics and Statistics*, 82, 540–554.

Franke, J. and E. G. Nyarige (2019): "A residual-based bootstrap for functional autoregressions," *arXiv preprint arXiv:1905.07635.*

Gao, H. O. (2007): "Day of week effects on diurnal ozone/NOx cycles and transportation emissions in Southern California," *Transportation Research Part D: Transport and Environment*, 12, 292–305.

Gasser, T. and H.-G. Müller (1984): "Estimating regression functions and their derivatives by the kernel method," *Scandinavian Journal of Statistics*, 171–185.

Geweke, J. (1977), "The Dynamic Factor Analysis of Economic Time Series," in *Latent Variables in Socio-Economic Models*, eds. D. J. Aigner and A. S. Goldberger, Amsterdam: North-Holland.

Gonçalves, S. (2011): "The moving blocks bootstrap for panel linear regression models with individual fixed effects," *Econometric Theory*, 27, 1048–1082.

González-Manteiga, W. and A. Martínez-Calvo (2011): "Bootstrap in functional linear regression," *Journal of Statistical Planning and Inference*, 141, 453–461.

Grambsch, P. M., B. L. Randall, R. M. Bostick, J. D. Potter, and T. A. Louis (1995): "Modeling the labeling index distribution: An application of functional data analysis," *Journal of the American Statistical Association*, 90, 813–821.

Hall, P., T. J. DiCiccio, and J. P. Romano (1989): "On smoothing and the bootstrap," *The Annals of Statistics*, 17, 692–704.

Hall, P. and Horowitz, J.L. (2007), "Methodology and Convergence Rates for Functional Linear Regression," *The Annals of Statistics*, 35, 70–91.

Hall, P., Müller, H. G. and J. L. Wang (2006), "Properties of Principal Component Methods for Functional and Longitudinal Data Analysis," *The Annals of Statistics*, 1493–1517.

Hallin, M. and R. Liška (2007), "Determining the Number of Factors in the General Dynamic Factor Model," *Journal of the American Statistical Association*, 102, 603–617.

Harezlak, J., B. A. Coull, N. M. Laird, S. R. Magari, and D. C. Christiani (2007): "Penalized solutions to functional regression problems," *Computational statistics & data analysis*, 51, 4911–4925.

Hastie, T. and R. Tibshirani (1993), "Varying-coefficient Models," *Journal of the Royal Statistical Society: Series B (Methodological)*, 55, 757–779.

Hays, S., H. Shen, and J. Z. Huang (2012), "Functional Dynamic Factor Models with Application to Yield Curve Forecasting," *The Annals of Applied Statistics*, 6, 870–894.

Herwartz, H. and F. Xu (2009): "A new approach to bootstrap inference in functional coefficient models," *Computational Statistics & Data Analysis*, 53, 2155–2167.

Hu, Z., N. Wang, and R. J. Carroll (2004): "Profile-kernel versus backfitting in the partially linear models for longitudinal/clustered data," *Biometrika*, 91, 251–262.

Hörmann, S. and P. Kokoszka (2012): "Functional Time Series," *Handb. Stat.*, 30.

Johnson, S., W. Larson, C. Papageorgiou, and A. Subramanian (2013): "Is newer better? Penn World Table revisions and their impact on growth estimates," *Journal of Monetary Economics*, 60, 255–274.

Jungbacker, B., Koopman, S. J., and Van der Wel, M. (2014), "Smooth Dynamic Factor Analysis with Application to the US Term Structure of Interest Rates," *Journal of Applied Econometrics*, 29, 65–90.

Kim, K., D. Şentürk, and R. Li (2011): "Recent history functional linear models for sparse longitudinal data," *Journal of statistical planning and inference*, 141, 1554–1566.

Kokoszka, P., Miao, H., and Zhang, X. (2014), "Functional Dynamic Factor Model for Intraday Price Curves," *Journal of Financial Econometrics*, 13, 456–477.

Kowal, D. R., Matteson, D. S., and Ruppert, D. (2017a), "A Bayesian Multivariate Functional Dynamic Linear Model," *Journal of the American Statistical Association*, 112, 733–744.

——— (2017b), "Functional Autoregression for Sparsely Sampled Data," *Journal of Business & Economic Statistics*, 1–13.

Kunsch, H. R. (1989): "The jackknife and the bootstrap for general stationary observations," *The annals of Statistics*, 1217–1241.

Laukaitis, A. (2008): "Functional data analysis for cash flow and transactions intensity continuous-time prediction using Hilbert-valued autoregressive processes," *European Journal of Operational Research*, 185, 1607–1614.

Li, Y., P. Perron, and J. Xu (2017), "Modelling Exchange Rate Volatility with Random Level Shifts," *Applied Economics*, 49, 2579–2589.

Lin, H., X. Jiang, H. Lian, and W. Zhang (2019): "Reduced rank modeling for functional regression with functional responses," *Journal of Multivariate Analysis*, 169, 205–217.

Liu, R. Y. and K. Singh (1992): "Moving blocks jackknife and bootstrap capture weak dependence," *Exploring the limits of bootstrap*, 225, 248.

Malfait, N. and J. O. Ramsay (2003): "The historical functional linear model," *Canadian Journal of Statistics*, 31, 115–128.

Meiring, W. (2007): "Oscillations and time trends in stratospheric ozone levels: a functional data analysis approach," *Journal of the American Statistical Association*, 102, 788–802.

Melino, A. and C. Sims (1996), "Estimation of Continuous-Time Models in Finance," in *Advances in Econometrics Sixth World Congress, ed. Sims, CA, Cambridge: Cambridge University Press*, 2, 313–351.

Merton, R. C. (1980), "On Estimating the Expected Return on the Market: An Exploratory Investigation," *Journal of Financial Economics*, 8, 323–361.

——— (1992), *Continuous-Time Finance*, Revised Edition, Oxford: Basil Blackwell.

Mikkelsen, J. G., E. Hillebrand, and G. Urga (2019), "Consistent Estimation of Time-Varying Loadings in High-Dimensional Factor Models," *Journal of Econometrics*, 208, 535–562.

Motta, G., C. M. Hafner, and R. von Sachs (2011), "Locally Stationary Factor Models: Identification and Nonparametric Estimation," *Econometric Theory*, 27, 1279–1319.

Müller, H.-G., S. Wu, A. D. Diamantidis, N. T. Papadopoulos, and J. R. Carey (2009): "Reproduction is adapted to survival characteristics across geographically isolated medfly populations," *Proceedings of the Royal Society B: Biological Sciences*, 276, 4409–4416.

Paparoditis, E. (2018): "Sieve bootstrap for functional time series," *The Annals of Statistics*, 46, 3510–3538.

Park, S. Y. and A.-M. Staicu (2015): "Longitudinal functional data analysis," *Stat*, 4, 212–226.

Ponomareva, N. and H. Katayama (2010): "Does the Version of the Penn World Tables Matter? An Analysis of the Relationship Between Growth and Volatility," *Canadian Journal of Economics*, 43, 152–179.

Ramsay, J. O. (1982): "When the data are functions," *Psychometrika*, 47, 379–396.

Ramsay, J. O. (2005), *Functional Data Analysis*, Wiley Online Library.

Ramsay, J. O. and C. Dalzell (1991): "Some tools for functional data analysis," *Journal of the Royal Statistical Society: Series B (Methodological)*, 53, 539–561.

Ramsay, J. O. and J. B. Ramsey (2002): "Functional data analysis of the dynamics of the monthly index of nondurable goods production," *Journal of Econometrics*, 107, 327–344.

Rana, P., G. Aneiros, J. Vilar, and P. Vieu (2016): "Bootstrap confidence intervals in functional nonparametric regression under dependence," *Electronic Journal of Statistics*, 10, 1973–1999.

Reichlin, L. (2003), "Factor Models in Large Cross Sections of Time Series," *Econometric Society Monographs*, 37, 47–86.

Sargent, T. J. and C. A. Sims (1977), "Business Cycle Modeling without Pretending to Have Too Much A Priori Economic Theory," *New Methods in Business Cycle Research*, 1, 145–168.

Schwetlick, H. and V. Kunert (1993): "Spline smoothing under constraints on derivatives," *BIT Numerical Mathematics*, 33, 512–528.

Shang, H. L. (2015): "Resampling techniques for estimating the distribution of descriptive statistics of functional data," *Communications in Statistics-Simulation and Computation*, 44, 614–635.

——— (2018): "Bootstrap methods for stationary functional time series," *Statistics and Computing*, 28, 1–10.

Sharipov, O., J. Tewes, and M. Wendler (2016): "Sequential block bootstrap in a Hilbert space with application to change point analysis," *Canadian Journal of Statistics*, 44, 300–322.

Speckman, P. L. and D. Sun (2003): "Fully Bayesian spline smoothing and intrinsic autoregressive priors," *Biometrika*, 90, 289–302.

Stock, J. H. and M. W. Watson (2005), "Implications of Dynamic Factor Models for VAR Analysis," Tech. rep., National Bureau of Economic Research.

——— (2006), "Forecasting with Many Predictors," *Handbook of Economic Forecasting*, 1, 515–554.

——— (2009), "Forecasting in Dynamic Factor Models Subject to Structural Instability," *The Methodology and Practice of Econometrics. A Festschrift in Honour of David F. Hendry*, 173, 205.

Su, L. and X. Wang (2017), "On Time-Varying Factor Models: Estimation and Testing," *Journal of Econometrics*, 198, 84–101.

Summers, R. and A. Heston (1991): "The Penn World Table (Mark 5): an expanded set of international comparisons, 1950–1988," *The Quarterly Journal of Economics*, 106, 327–368.

Ullah, S. and C. F. Finch (2013): "Applications of functional data analysis: A systematic review," *BMC medical research methodology*, 13, 43.

Wang, J.-L., J.-M. Chiou, and H.-G. Müller (2016): "Functional data analysis," *Annual Review of Statistics and Its Application*, 3, 257–295.

Xu, J. and P. Perron (2014), "Forecasting Return Volatility: Level Shifts with Varying Jump Probability and Mean Reversion," *International Journal of Forecasting*, 30, 449–463.

——— (2017), "Forecasting in the Presence of in and out of Sample Breaks," Boston University - Department of Economics - Working Papers Series WP2018-014, Boston University - Department of Economics, revised Nov 2018.

Yamamoto, Y. and S. Tanaka (2015), "Testing for Factor Loading Structural Change Under Common Breaks," *Journal of Econometrics*, 189, 187–206.

Zhou, S., X. Shen, and D. Wolfe (1998): "Local asymptotics for regression splines and confidence regions," *The Annals of Statistics*, 26, 1760–1782.

Zhu, H., J. Fan, and L. Kong (2014): "Spatially varying coefficient model for neuroimaging data with jump discontinuities," *Journal of the American Statistical Association*, 109, 1084–1098.

Zhu, H., M. Styner, N. Tang, Z. Liu, W. Lin, and J. H. Gilmore (2010): "Frats: Functional regression analysis of dti tract statistics," *IEEE transactions on medical imaging*, 29, 1039–1049.

APPENDICES

A Appendices of Chapter 1

We now prove Theorems 1.3.1 and 1.3.2. Since the proofs for $m = 1$ and $m = 2$ follow the same idea, we will show the proofs for $m = 1$ only.

Recall that $G_{j,v}(t) = X_{j,v}(t), X'_{j,v}(t), X''_{j,v}(t)$ and $\hat{G}_{j,v}(t)$ denotes the corresponding estimated function, for all j and v.

A.1 Proof of Theorem 1.3.1

We can expand $\frac{1}{N}\sum_{j=1}^{N}\left(\hat{G}_{j,v_1}(t) - \hat{G}_{j,v_2}(t)\right)$ as

$$\begin{aligned}
&\frac{1}{N}\sum_{j=1}^{N}\left(\hat{G}_{j,v_1}(t) - \hat{G}_{j,v_2}(t)\right) \\
=&\frac{1}{N}\sum_{j=1}^{N}\left(G_{j,v_1}(t) - \mu_{G_{v_1}}(t)\right) - \frac{1}{N}\sum_{j=1}^{N}\left(G_{j,v_2}(t) - \mu_{G_{v_2}}(t)\right) + \left(\mu_{G_{v_1}}(t) - \mu_{G_{v_2}}(t)\right) + \\
&\frac{1}{N}\sum_{j=1}^{N}\left(\hat{G}_{j,v_1}(t) - G_{j,v_1}(t)\right) - \frac{1}{N}\sum_{j=1}^{N}\left(\hat{G}_{j,v_2}(t) - G_{j,v_2}(t)\right). \qquad \text{(A.1)}
\end{aligned}$$

Under Assumptions 1.3.1 and 1.3.2, applying Chebyshev's inequality and Theorem 2 from Claeskens et al. (2009), we have $\hat{G}_{j,v}(t) - G_{j,v}(t) = \mathcal{O}_p\left(S^{-\gamma}\right)$ for given j, v and almost all

t (Claeskens et al., 2009, Theorem 2) [3], which implies that

$$\begin{aligned}
&\sqrt{N}\left[\frac{1}{N}\sum_{j=1}^{N}\left(\hat{G}_{j,v_1}(t)-\hat{G}_{j,v_2}(t)\right)\right] \\
=&\sqrt{N}\left[\frac{1}{N}\sum_{j=1}^{N}\left(G_{j,v_1}(t)-\mu_{G_{v_1}}(t)\right)\right]-\sqrt{N}\left[\frac{1}{N}\sum_{j=1}^{N}\left(G_{j,v_2}(t)-\mu_{G_{v_2}}(t)\right)\right]+ \\
&\sqrt{N}\left(\mu_{G_{v_1}}(t)-\mu_{G_{v_2}}(t)\right)+\mathcal{O}_p\left(N^{1/2}S^{-\gamma}\right). \qquad \text{(A.2)}
\end{aligned}$$

By Assumption 1.3.3 and Lyapunov CLT,

$$\sqrt{N}\left[\frac{1}{N}\sum_{j=1}^{N}\left(G_{j,v}(t)-\mu_{G_v}(t)\right)\right]\xrightarrow{d}N\left(0,N^{-1}S^2_{N,G_v}(t)\right), \qquad \text{(A.3)}$$

and under the null hypothesis, $\mu_{G_{v_1}}(t)-\mu_{G_{v_2}}(t)=0$; then symmetry of normal distributions implies that

$$\sqrt{N}\left[\frac{1}{N}\sum_{j=1}^{N}\left(\hat{G}_{j,v_1}(t)-\hat{G}_{j,v_2}(t)\right)\right]\xrightarrow{d}\mathbb{G}^{(1)}(t),$$

where $\mathbb{G}^{(1)}(t)\sim N\left(0,N^{-1}S^2_{N,G_{v_1}}(t)\right)+N\left(0,N^{-1}S^2_{N,G_{v_2}}(t)\right)$. Hence, applying the continuous mapping theorem, we have

$$\sqrt{N}W^{(1)}_{v_1,v_2}=\sqrt{N}\int_0^1\left[\frac{1}{N}\sum_{j=1}^{N}\left(\hat{G}_{j,v_1}(t)-\hat{G}_{j,v_2}(t)\right)\right]^2dt\xrightarrow{d}\int_0^1\left(\mathbb{G}^{(1)}(t)\right)^2dt.$$

Recall that $\left\{G^*_{b,j,v}(t)\right\}_{j=1}^N$ denotes the b-th set of bootstrap sample from $\{G_{j,v}(t)\}_{j=1}^N$, where we apply an i.i.d. bootstrap, and $\hat{G}^*_{b,j,v}(t)$ denotes the corresponding estimated functions. We again apply Chebyshev's inequality and Theorem 2 from Claeskens et al. (2009), so that $\hat{G}^*_{b,j,v}(t)-G^*_{b,j,v}(t)=\mathcal{O}_p\left(S^{-\gamma}\right)$ for given j,v,b and almost all t. We can

[3] Under the optimal K and λ that satisfy Assumption 1.3.2, the pointwise asymptotic bias and the square root of the pointwise asymptotic variance are both $\mathcal{O}\left(S^{-4/9}\right)$ (Claeskens et al., 2009, Theorem 2). For the verification, see Claeskens et al. (2009).

then obtain

$$\begin{aligned}
&\frac{1}{N}\sum_{j=1}^{N}\left(\hat{G}^{*}_{b,j,v_1}(t)-\hat{G}^{*}_{b,j,v_2}(t)\right)-\frac{1}{N}\sum_{j=1}^{N}\left(\hat{G}_{j,v_1}(t)-\hat{G}_{j,v_2}(t)\right)\\
=&\frac{1}{N}\sum_{j=1}^{N}\left(G^{*}_{b,j,v_1}(t)-G^{*}_{b,j,v_2}(t)\right)-\frac{1}{N}\sum_{j=1}^{N}\left(G_{j,v_1}(t)-G_{j,v_2}(t)\right)+\mathcal{O}_p\left(S^{-\gamma}\right).
\end{aligned}$$

Noting that $\{G_{j,v}(t)\}_{j=1}^{N}$ is the population of $\left\{G^{*}_{b,j,v}(t)\right\}_{j=1}^{N}$, we can define $\mu_{G^{*}_{v,b}}(t) := N^{-1}\sum_{j=1}^{N}G_{j,v}(t)$. Hence, we have the following:

$$\begin{aligned}
&\frac{1}{N}\sum_{j=1}^{N}\left(\hat{G}^{*}_{b,j,v_1}(t)-\hat{G}^{*}_{b,j,v_2}(t)\right)-\frac{1}{N}\sum_{j=1}^{N}\left(\hat{G}_{j,v_1}(t)-\hat{G}_{j,v_2}(t)\right)\\
=&\frac{1}{N}\sum_{j=1}^{N}\left(G^{*}_{b,j,v_1}(t)-\mu_{G^{*}_{v_1,b}}(t)\right)-\frac{1}{N}\sum_{j=1}^{N}\left(G^{*}_{b,j,v_2}(t)-\mu_{G^{*}_{v_2,b}}(t)\right)+\mathcal{O}_p\left(S^{-\gamma}\right).
\end{aligned}$$

Since $\left\{G^{*}_{b,j,v}(t)\right\}_{j=1}^{N}$ is obtained from i.i.d. resampling, it is implied that

$$G^{*}_{b,j,v}(t)-\mu_{G^{*}_{v,b}}(t)\overset{d}{=}G_{j,v}(t)-\mu_{G_v}(t),$$

which therefore implies that

$$\begin{aligned}
\sqrt{N}W^{*(m)}_{b,v_1,v_2}=&\sqrt{N}\int_0^1\left[\frac{1}{N}\sum_{j=1}^{N}\left(\hat{G}^{*}_{b,j,v_1}(t)-\hat{G}^{*}_{b,j,v_2}(t)\right)-\frac{1}{N}\sum_{j=1}^{N}\left(\hat{G}_{j,v_1}(t)-\hat{G}_{j,v_2}(t)\right)\right]^2dt\\
\xrightarrow{d}&\int_0^1\left(\mathbb{G}^{(1)}(t)\right)^2dt. \qquad \text{(A.4)}
\end{aligned}$$

□

A.2 Proof of Theorem 1.3.2

According to Equations (A.1) to (A.3), we have

$$\sqrt{N}\left[\frac{1}{N}\sum_{j=1}^{N}\left(\hat{G}_{j,v_1}(t)-\hat{G}_{j,v_2}(t)\right)\right]=\mathbb{G}^{(1)}(t)+\sqrt{N}\left(\mu_{G_{v_1}}(t)-\mu_{G_{v_2}}(t)\right)+\mathcal{O}_p\left(N^{1/2}S^{-\gamma}\right).$$

Under the alternatives, $\mu_{G_{v_1}}(t) - \mu_{G_{v_2}}(t) = \mathcal{O}(1)$, and $\sqrt{N}\left(\mu_{G_{v_1}}(t) - \mu_{G_{v_2}}(t)\right) = \mathcal{O}(N^{1/2})$; hence,

$$\sqrt{N}\left[\frac{1}{N}\sum_{j=1}^{N}\left(\hat{G}_{j,v_1}(t) - \hat{G}_{j,v_2}(t)\right)\right] = \mathcal{O}_p(N^{1/2}).$$

According to the proofs of Theorem 1.3.1, the result in Equation (A.4) does not depend on the null hypothesis. Therefore, we can conclude that under the alternatives, $\sqrt{N}W^{*(m)}_{b,v_1,v_2} = \mathcal{O}_p(1)$.

□

B Appendices of Chapter 2

This appendix contains definitions for the notations, derivations of the estimators, as well as the proofs to the theorems and the to-be-stated Lemmas.

B.1 Notations

Recall that n denotes the number of replications, and J is the number of observations for each replication, which forms an index set $\{t_j\}_{j=1}^{J}$. We can then define the following vectors:

$$\boldsymbol{x}_i = \begin{bmatrix} x_i(t_1) \\ \vdots \\ x_i(t_J) \end{bmatrix}_{J\times 1}, \boldsymbol{\varepsilon}_i = \begin{bmatrix} \varepsilon_i(t_1) \\ \vdots \\ \varepsilon_i(t_J) \end{bmatrix}_{J\times 1};$$

$\tilde{\boldsymbol{x}}_i$ is constructed in the same way. The matrices of the basis functions are defined as follows:

$$\boldsymbol{\beta}(\cdot) = \begin{bmatrix} \beta_1(\cdot) \\ \vdots \\ \beta_M(\cdot) \end{bmatrix}_{M\times 1}, \boldsymbol{\alpha}(\cdot) = \begin{bmatrix} \alpha_1(\cdot) \\ \vdots \\ \alpha_H(\cdot) \end{bmatrix}_{H\times 1}, \boldsymbol{A}(\cdot) = \begin{bmatrix} \boldsymbol{\alpha}(\cdot) & & \\ & \ddots & \\ & & \boldsymbol{\alpha}(\cdot) \end{bmatrix}_{HK\times K},$$

$$\boldsymbol{\theta}(\cdot) = \begin{bmatrix} \theta_1(\cdot) \\ \vdots \\ \theta_Q(\cdot) \end{bmatrix}_{Q\times 1}, \boldsymbol{\Theta}_P(\cdot) = \begin{bmatrix} \boldsymbol{\theta}(\cdot) & & \\ & \ddots & \\ & & \boldsymbol{\theta}(\cdot) \end{bmatrix}_{QP\times P}, \boldsymbol{\Theta}(\cdot) = \begin{bmatrix} \boldsymbol{\Theta}_P(\cdot) & & \\ & \ddots & \\ & & \boldsymbol{\Theta}_P(\cdot) \end{bmatrix}_{QPK\times PK},$$

$$\boldsymbol{\psi}(\cdot) = \begin{bmatrix} \psi_1(\cdot) \\ \vdots \\ \psi_P(\cdot) \end{bmatrix}_{P\times 1}, \boldsymbol{\Psi}(\cdot) = \begin{bmatrix} \boldsymbol{\psi}(\cdot) & & \\ & \ddots & \\ & & \boldsymbol{\psi}(\cdot) \end{bmatrix}_{PK\times K};$$

the matrices of the basis coefficients are

$$\boldsymbol{c}_i = \begin{bmatrix} c_{i,1} \\ \vdots \\ c_{i,M} \end{bmatrix}_{M\times 1}, \boldsymbol{C} = \begin{bmatrix} \boldsymbol{c}_1' \\ \vdots \\ \boldsymbol{c}_n' \end{bmatrix}_{n\times M}, \boldsymbol{d}_k = \begin{bmatrix} d_{k,1} \\ \vdots \\ d_{k,M} \end{bmatrix}_{M\times 1}, \boldsymbol{D} = \begin{bmatrix} \boldsymbol{d}_1' \\ \vdots \\ \boldsymbol{d}_K' \end{bmatrix}_{K\times M},$$

$$\boldsymbol{a}_k = \begin{bmatrix} a_{k,1} \\ \vdots \\ a_{k,H} \end{bmatrix}_{H\times 1}, \boldsymbol{a} = \begin{bmatrix} \boldsymbol{a}_1 \\ \vdots \\ \boldsymbol{a}_K \end{bmatrix}_{HK\times 1},$$

$$\boldsymbol{b}_{i,k,p} = \begin{bmatrix} b_{i,k,p,1} \\ \vdots \\ b_{i,k,p,Q} \end{bmatrix}_{Q\times 1}, \boldsymbol{b}_{i,k} = \begin{bmatrix} \boldsymbol{b}_{i,k,1} \\ \vdots \\ \boldsymbol{b}_{i,k,P} \end{bmatrix}_{QP\times 1}, \boldsymbol{b}_i = \begin{bmatrix} \boldsymbol{b}_{i,1} \\ \vdots \\ \boldsymbol{b}_{i,K} \end{bmatrix}_{QPK\times 1}, \boldsymbol{B} = \begin{bmatrix} \boldsymbol{b}_1' \\ \vdots \\ \boldsymbol{b}_n' \end{bmatrix}_{n\times QPK},$$

where the corresponding estimators are constructed in the same way. Specifically, $\boldsymbol{\beta}(\cdot)$, $\boldsymbol{c}_i$ and $\boldsymbol{C}$ are for the estimation of the functional data from the observations x_{it}; $\boldsymbol{\beta}(\cdot)$, $\boldsymbol{d}_k$ and $\boldsymbol{D}$ are for the estimation of the functional principal components; $\boldsymbol{\alpha}(\cdot)$, $\boldsymbol{A}(\cdot)$, $\boldsymbol{a}_k$ and $\boldsymbol{a}$ are for the estimation of the functional factors; the rest of the basis functions and the coefficients are for the estimation of the bi-variate functional loadings. For simplicity of expression, we introduce the following notations:

$\hat{\boldsymbol{\Omega}}_\lambda(t) := \int_0^t \hat{\boldsymbol{f}}^T(s)\boldsymbol{\Psi}^T(s)ds\boldsymbol{\Theta}^T(t)$;
$\hat{\boldsymbol{R}}_\lambda := \int_0^1 \left\{\hat{\boldsymbol{\Omega}}_\lambda^T(t)\hat{\boldsymbol{\Omega}}_\lambda(t)\right\} dt + \gamma_\lambda \int_0^1 \boldsymbol{\Theta}''(t)\boldsymbol{\Theta}''^T(t)dt.$

Also, let f_k^0 and ρ_k denote the limits of the estimated eigenfunctions $\hat{f}_k^*$'s and eigenvalues $\hat{\rho}_k$'s as $n, J \to \infty$, then Assumption 2.3.1 implies that $\hat{f}_k^* - f_k^0 = \mathcal{O}_p\left(n^{-1/2}\right)$ and $\hat{\rho}_k - \rho_k = \mathcal{O}_p\left(n^{-1/2}\right)$ for all k (e.g., Hall et al., 2006)[3], and correspondingly, $\hat{\boldsymbol{f}}^* - \boldsymbol{f}^0 = \mathcal{O}_p\left(n^{-1/2}\right)$ and $\hat{\boldsymbol{\rho}} - \boldsymbol{\rho} = \mathcal{O}_p\left(n^{-1/2}\right)$[4].

B.2 Derivation

Recall that order-four B-spline bases with equal-spaced knots on $[0, 1]$ time interval are used to estimate the functional data $x_i(\cdot)$ and the functional principal components $\hat{\boldsymbol{f}}^*(\cdot)$, order-four B-spline bases are used to estimate the functional factors $\hat{\boldsymbol{f}}(\cdot)$ as well as the first dimension of the loading functions $\boldsymbol{\lambda}_i^*(\cdot)$, and order-one B-spline bases are used for the second dimension of the loading functions. Meanwhile, we use the roughness penalties of order-two derivatives for all the functional estimators but that for the second dimension of the functional loadings.

The estimates of $\{\boldsymbol{c}_i\}_i$'s, denoted $\{\tilde{\boldsymbol{c}}_i\}_i$'s, can be obtained by solving the first order conditions of Equation (2.7), such that

$$\tilde{\boldsymbol{c}}_i = \left\{ \frac{1}{J} \sum_{j=1}^{J} \boldsymbol{\beta}(t_j) \boldsymbol{\beta}^T(t_j) + \gamma_x \int_0^1 \boldsymbol{\beta}''(t) \boldsymbol{\beta}''^T(t) dt \right\}^{-1} \frac{1}{J} \sum_{j=1}^{J} \boldsymbol{\beta}(t_j) x_{it_j}, \ \forall i, \tag{B.1}$$

then the fitted functional data can be expressed as

$$\tilde{\boldsymbol{x}}(t) = \tilde{\boldsymbol{C}} \boldsymbol{\beta}(t). \tag{B.2}$$

Approximating $\boldsymbol{f}^*(t)$ with the basis expansion $\boldsymbol{D}\boldsymbol{\beta}(t)$ and defining the estimator $\hat{\boldsymbol{f}}^*(t) := \hat{\boldsymbol{D}}\boldsymbol{\beta}(t)$ correspondingly, where $\hat{\boldsymbol{D}}$ denotes the estimator of $\boldsymbol{D}$, we can then re-write Equa-

[3]Hall et al. (2006) states that if the process $x_i(t)$ is fully observed without noise, the eigenfunctions $\hat{f}_k^*$'s and the eigenvalues $\hat{\rho}_k$'s are both $\mathcal{O}_p\left(n^{-1/2}\right)$, but if the observations come with noise, the convergence of the eigenvalues will still be at the rate of $\mathcal{O}_p\left(n^{-1/2}\right)$, while that of the eigenfunctions will drop to a lower speed; however, as the number of observations J goes to infinity, one can treat the process $x_i(t)$ as fully observed in a continuum.

[4]In our proofs, we use the $\mathcal{O}$, $\mathcal{O}_p$ and o_p notations for the cases with matrices in any dimensions (i.e., scalars, vectors or higher dimensional matrices), and we let the notations adjust to the comfortable dimensions without specifying repeatedly.

tion (2.13) as

$$\int_0^1 n^{-1}\hat{\boldsymbol{D}}\boldsymbol{\beta}(s)\boldsymbol{\beta}^T(s)\tilde{\boldsymbol{C}}^T\tilde{\boldsymbol{C}}\boldsymbol{\beta}(t)ds = \hat{\boldsymbol{\rho}}\hat{\boldsymbol{D}}\boldsymbol{\beta}(t), \tag{B.3}$$

where

$$\hat{\boldsymbol{\rho}} = n^{-1}\int_0^1 \hat{\boldsymbol{D}}\boldsymbol{\beta}(t)\boldsymbol{\beta}^T(t)\tilde{\boldsymbol{C}}^T dt \int_0^1 \tilde{\boldsymbol{C}}\boldsymbol{\beta}(t)\boldsymbol{\beta}^T(t)\hat{\boldsymbol{D}}^T dt.$$

Since Equation (B.3) holds for $\boldsymbol{\beta}(t)$ at all t, it can be reduced to

$$\hat{\boldsymbol{D}}n^{-1}\int_0^1 \boldsymbol{\beta}(s)\boldsymbol{\beta}^T(s)ds\tilde{\boldsymbol{C}}^T\tilde{\boldsymbol{C}} = \hat{\boldsymbol{\rho}}\hat{\boldsymbol{D}},$$

which follows

$$\begin{aligned}&\hat{\boldsymbol{D}}\left\{\int_0^1 \boldsymbol{\beta}(s)\boldsymbol{\beta}^T(s)ds\right\}^{1/2} n^{-1}\left\{\int_0^1 \boldsymbol{\beta}(s)\boldsymbol{\beta}^T(s)ds\right\}^{1/2}\tilde{\boldsymbol{C}}^T\tilde{\boldsymbol{C}}\left\{\int_0^1 \boldsymbol{\beta}(s)\boldsymbol{\beta}^T(s)ds\right\}^{1/2}\\ =&\hat{\boldsymbol{\rho}}\hat{\boldsymbol{D}}\left\{\int_0^1 \boldsymbol{\beta}(s)\boldsymbol{\beta}^T(s)ds\right\}^{1/2},\end{aligned} \tag{B.4}$$

with the identification constraint

$$\int_0^1 \hat{\boldsymbol{f}}^*(s)\hat{\boldsymbol{f}}^{*T}(s)ds = \hat{\boldsymbol{D}}\int_0^1 \boldsymbol{\beta}(s)\boldsymbol{\beta}^T(s)ds\hat{\boldsymbol{D}}^T = \boldsymbol{I}_K, \tag{B.5}$$

where $\boldsymbol{I}_K$ is a $K\times K$ identity matrix. The $K\times M$ matrix $\hat{\boldsymbol{D}}\left\{\int_0^1 \boldsymbol{\beta}(s)\boldsymbol{\beta}^T(s)ds\right\}^{1/2}$ can be computed by filling up the K rows with the eigenvectors corresponding to the largest K eigenvalues of the $M\times M$ matrix

$$n^{-1}\left\{\int_0^1 \boldsymbol{\beta}(s)\boldsymbol{\beta}^T(s)ds\right\}^{1/2}\tilde{\boldsymbol{C}}^T\tilde{\boldsymbol{C}}\left\{\int_0^1 \boldsymbol{\beta}(s)\boldsymbol{\beta}^T(s)ds\right\}^{1/2},$$

and $\hat{\boldsymbol{D}}$ can be computed as $\left[\hat{\boldsymbol{D}}\left\{\int_0^1 \boldsymbol{\beta}(s)\boldsymbol{\beta}^T(s)ds\right\}^{1/2}\right]\left\{\int_0^1 \boldsymbol{\beta}(s)\boldsymbol{\beta}^T(s)ds\right\}^{-1/2}$. Once $\hat{\boldsymbol{D}}$ is obtained, we can get $\hat{\boldsymbol{f}}^*(t)$ as

$$\hat{\boldsymbol{f}}^*(t) = \hat{\boldsymbol{D}}\boldsymbol{\beta}(t), \tag{B.6}$$

and the estimated factors

$$\hat{\boldsymbol{f}}(t) := \frac{\partial\hat{\boldsymbol{f}}^*(t)}{\partial t}.$$

For the estimation of the loadings, the coefficients $\boldsymbol{b}_i$'s, and thus $\boldsymbol{\lambda}_i^*(\cdot)$, can be estimated through the following penalized least squares criteria, where $\tilde{\boldsymbol{b}}_i$ represents any estimator of $\boldsymbol{b}_i$:

$$m(\left\{\tilde{\boldsymbol{b}}_i\right\}; \gamma_\lambda, Q) := \int_0^1 \left\{\tilde{x}_i(t) - \tilde{\boldsymbol{b}}_i^T \hat{\boldsymbol{\Omega}}_\lambda^T(t)\right\}^2 dt + \gamma_\lambda \int_0^1 \tilde{\boldsymbol{b}}_i^T \boldsymbol{\Theta}''(t) \boldsymbol{\Theta}''^T(t) \tilde{\boldsymbol{b}}_i dt, \tag{B.7}$$

and the estimator $\hat{\boldsymbol{b}}_i$ (and thus $\hat{\boldsymbol{B}}$), can be obtained by solving the first order condition, such that

$$\hat{\boldsymbol{b}}_i = \hat{\boldsymbol{R}}_\lambda^{-1} \int_0^1 \hat{\boldsymbol{\Omega}}_\lambda^T(t) \tilde{x}_i(t) dt, \tag{B.8}$$

and the corresponding estimator for $\tilde{\boldsymbol{\lambda}}_i^*(\cdot)$ as well as the estimated loading function can then be expressed as

$$\hat{\boldsymbol{\lambda}}_i^*(t) = \boldsymbol{\Theta}^T(t) \hat{\boldsymbol{b}}_i, \ \forall i; \quad \hat{\boldsymbol{\Lambda}}^*(t) = \hat{\boldsymbol{B}} \boldsymbol{\Theta}(t),$$

and

$$\hat{\boldsymbol{\lambda}}_i(t,s) = \boldsymbol{\Psi}^T(s) \boldsymbol{\Theta}^T(t) \hat{\boldsymbol{b}}_i, \ \forall i; \quad \hat{\boldsymbol{\Lambda}}(t,s) = \hat{\boldsymbol{B}} \boldsymbol{\Theta}(t) \boldsymbol{\Psi}(s). \tag{B.9}$$

B.3 Proofs of theorems

In this section, we present the proofs of the theorems. We first introduce another index set, $\{\tau_j\}_{j=1}^T$, consisting of points that are equally spaced on $[0,1]$, where $T \to \infty$[5], and we define the integral transformation $\boldsymbol{W}$, such that

$$(\boldsymbol{W}\boldsymbol{f})(t) := \hat{\boldsymbol{\rho}}^{-1} n^{-1} T^{-1} \sum_{j=1}^T \hat{\boldsymbol{f}}^*(\tau_j) \boldsymbol{f}^{*T}(\tau_j) \boldsymbol{\Lambda}^{*T}(\tau_j) \boldsymbol{\Lambda}^*(t) \int_0^t \boldsymbol{\Psi}(s) \boldsymbol{f}(s) ds.$$

Also, for the latent factors $\boldsymbol{f}(t)$ and the estimator $\hat{\boldsymbol{f}}(t)$, there exists some PK-by-PK continuous matrix function $\boldsymbol{W}^*(t)$, such that $\int_0^t \boldsymbol{\Psi}(s) \boldsymbol{f}(s) ds = \boldsymbol{W}^*(t) \int_0^t \boldsymbol{\Psi}(s) \hat{\boldsymbol{f}}(s) ds$, for all $t \in [0,1]$.

We now begin the proofs with the statement of the following lemmas.

[5]The use of $\{\tau_j\}_{j=1}^T$ and T is for the proof, in order to separate from the index for observations when necessary.

Lemma B.1 *Under Assumption 2.3.1, we have that* $\hat{\mathbf{\Omega}}_\lambda(s)\left\{\int_0^1 \hat{\mathbf{\Omega}}_\lambda^T(\tau)\hat{\mathbf{\Omega}}_\lambda(\tau)d\tau\right\}^{-1}\hat{\mathbf{\Omega}}_\lambda^T(t) - \hat{\mathbf{\Omega}}_\lambda(s)\hat{\boldsymbol{R}}_\lambda^{-T}\hat{\mathbf{\Omega}}_\lambda^T(t) = \mathcal{O}(\gamma_\lambda)$.

Lemma B.2 *Under Assumptions 2.3.3, 2.3.5 and 2.3.6.c, there exists a vector* $\boldsymbol{\mu}_i$*, where* $\boldsymbol{\mu}_i = \int_0^1 \boldsymbol{f}^0(t)\int_0^t \boldsymbol{\mu}_f(s)\mathbf{\Psi}^T(s)ds\boldsymbol{\lambda}_i^*(t)dt$ *for every* i *so that* $T^{-1}\sum_{j=1}^T \boldsymbol{f}^0(\tau_j)\boldsymbol{f}^{*T}(\tau_j)\boldsymbol{\lambda}_i^*(\tau_j) = \boldsymbol{\mu}_i + \mathcal{O}_p\left(T^{-1/2}\right)$.

The proof of Theorem 2.3.1.a

By subtracting and adding terms, we have

$$\begin{aligned}\hat{\boldsymbol{f}}^*(t) - (\boldsymbol{W}\boldsymbol{f})(t) = &\left\{\hat{\boldsymbol{f}}^*(t) - \hat{\boldsymbol{\rho}}^{-1}n^{-1}T^{-1}\sum_{j=1}^T \hat{\boldsymbol{f}}^*(\tau_j)\tilde{\boldsymbol{x}}^T(\tau_j)\tilde{\boldsymbol{x}}(t)\right\} + \\ &\left\{\hat{\boldsymbol{\rho}}^{-1}n^{-1}T^{-1}\sum_{j=1}^T \hat{\boldsymbol{f}}^*(\tau_j)\tilde{\boldsymbol{x}}^T(\tau_j)\tilde{\boldsymbol{x}}(t) - \hat{\boldsymbol{\rho}}^{-1}n^{-1}T^{-1}\sum_{j=1}^T \hat{\boldsymbol{f}}^*(\tau_j)\boldsymbol{x}^T(\tau_j)\boldsymbol{x}(t)\right\} + \\ &\left\{\hat{\boldsymbol{\rho}}^{-1}n^{-1}T^{-1}\sum_{j=1}^T \hat{\boldsymbol{f}}^*(\tau_j)\boldsymbol{x}^T(\tau_j)\boldsymbol{x}(t) - (\boldsymbol{W}\boldsymbol{f})(t)\right\} = (I) + (II) + (III).\end{aligned} \tag{B.10}$$

Recall that $\hat{\hat{\boldsymbol{f}}}^*(t) = \hat{\boldsymbol{\rho}}^{-1}n^{-1}\int_0^1 \hat{\boldsymbol{f}}^*(s)\tilde{\boldsymbol{x}}^T(s)ds\tilde{\boldsymbol{x}}(t)$, and applying Theorem 1(c) from Chui (1971), we have $\int_0^1 \hat{\boldsymbol{f}}^*(s)\tilde{x}_i(s)ds - T^{-1}\sum_{j=1}^T \hat{\boldsymbol{f}}^*(\tau_j)\tilde{x}_i(\tau_j) = \mathcal{O}(T^{-1})$[6]; therefore,

$$\begin{aligned}(I) &= \hat{\boldsymbol{f}}^*(t) - \hat{\boldsymbol{\rho}}^{-1}n^{-1}T^{-1}\sum_{j=1}^T \hat{\boldsymbol{f}}^*(\tau_j)\tilde{\boldsymbol{x}}^T(\tau_j)\tilde{\boldsymbol{x}}(t) \\ &= \hat{\boldsymbol{\rho}}^{-1}n^{-1}\sum_{i=1}^n \left\{\int_0^1 \hat{\boldsymbol{f}}^*(s)\tilde{x}_i(s)ds - T^{-1}\sum_{j=1}^T \hat{\boldsymbol{f}}^*(\tau_j)\tilde{x}_i(\tau_j)\right\}\tilde{x}_i(t) = \mathcal{O}\left(T^{-1}\right).\end{aligned} \tag{B.11}$$

For (II), first note that $\check{x}_i(t)$ is arbitrarily close to $x_i(t)$ for all i, and applying Assumption 2.3.1 and Theorem 2 from Claeskens et al. (2009), we have $\tilde{x}_i(t) - \check{x}_i(t) = \mathcal{O}_p\left(J^{-4/9}\right)$;

[6]Since $\hat{\boldsymbol{f}}^*(t)$ and $\tilde{x}_i(t)$ both consist of order-four local polynomials, they are both absolutely continuous, which suffices the result in Theorem 1(c) in Chui (1971).

hence, it follows that

$$\tilde{x}_i(t) - x_i(t) = \mathcal{O}_p\left(J^{-4/9}\right).$$

Therefore,

$$\begin{aligned}
(II) &= \hat{\boldsymbol{\rho}}^{-1} n^{-1} T^{-1} \sum_{j=1}^{T} \hat{\boldsymbol{f}}^*(\tau_j) \tilde{\boldsymbol{x}}^T(\tau_j) \tilde{\boldsymbol{x}}(t) - \hat{\boldsymbol{\rho}}^{-1} n^{-1} T^{-1} \sum_{j=1}^{T} \hat{\boldsymbol{f}}^*(\tau_j) \boldsymbol{x}^T(\tau_j) \boldsymbol{x}(t) \\
&= \hat{\boldsymbol{\rho}}^{-1} n^{-1} T^{-1} \sum_{j=1}^{T} \hat{\boldsymbol{f}}^*(\tau_j) \left\{\tilde{\boldsymbol{x}}(\tau_j) - \boldsymbol{x}(\tau_j)\right\}^T \left\{\tilde{\boldsymbol{x}}(t) - \boldsymbol{x}(t)\right\} + \\
&\quad \hat{\boldsymbol{\rho}}^{-1} n^{-1} T^{-1} \sum_{j=1}^{T} \hat{\boldsymbol{f}}^*(\tau_j) \left\{\tilde{\boldsymbol{x}}(\tau_j) - \boldsymbol{x}(\tau_j)\right\}^T \boldsymbol{x}(t) + \\
&\quad \hat{\boldsymbol{\rho}}^{-1} n^{-1} T^{-1} \sum_{j=1}^{T} \hat{\boldsymbol{f}}^*(\tau_j) \boldsymbol{x}^T(\tau_j) \left\{\tilde{\boldsymbol{x}}(t) - \boldsymbol{x}(t)\right\} \\
&= \mathcal{O}_p\left(J^{-8/9}\right) + \mathcal{O}_p\left(J^{-4/9}\right) + \mathcal{O}_p\left(J^{-4/9}\right) = \mathcal{O}_p\left(J^{-4/9}\right). \qquad \text{(B.12)}
\end{aligned}$$

For (III), recall $(\boldsymbol{W}\boldsymbol{f})(t) := \hat{\boldsymbol{\rho}}^{-1} n^{-1} T^{-1} \sum_{j=1}^{T} \hat{\boldsymbol{f}}^*(\tau_j) \boldsymbol{f}^{*T}(\tau_j) \boldsymbol{\Lambda}^{*T}(\tau_j) \boldsymbol{\Lambda}^*(t) \int_0^t \boldsymbol{\Psi}(s) \boldsymbol{f}(s) ds$, then we have

$$\begin{aligned}
(III) &= \hat{\boldsymbol{\rho}}^{-1} n^{-1} T^{-1} \sum_{j=1}^{T} \hat{\boldsymbol{f}}^*(\tau_j) \boldsymbol{x}^T(\tau_j) \boldsymbol{x}(t) - (\boldsymbol{W}\boldsymbol{f})(t) \\
&= \hat{\boldsymbol{\rho}}^{-1} n^{-1} T^{-1} \sum_{j=1}^{T} \hat{\boldsymbol{f}}^*(\tau_j) \boldsymbol{x}^T(\tau_j) \boldsymbol{x}(t) - \\
&\quad \hat{\boldsymbol{\rho}}^{-1} n^{-1} T^{-1} \sum_{j=1}^{T} \hat{\boldsymbol{f}}^*(\tau_j) \boldsymbol{f}^{*T}(\tau_j) \boldsymbol{\Lambda}^{*T}(\tau_j) \boldsymbol{\Lambda}^*(t) \int_0^t \boldsymbol{\Psi}(s) \boldsymbol{f}(s) ds \\
&= \hat{\boldsymbol{\rho}}^{-1} n^{-1} T^{-1} \sum_{j=1}^{T} \hat{\boldsymbol{f}}^*(\tau_j) \boldsymbol{f}^{*T}(\tau_j) \boldsymbol{\Lambda}^{*T}(\tau_j) \boldsymbol{\varepsilon}(t) + \hat{\boldsymbol{\rho}}^{-1} n^{-1} T^{-1} \sum_{j=1}^{T} \hat{\boldsymbol{f}}^*(\tau_j) \boldsymbol{\varepsilon}^T(\tau_j) \boldsymbol{\Lambda}^*(t) \boldsymbol{f}^*(t) + \\
&\quad \hat{\boldsymbol{\rho}}^{-1} n^{-1} T^{-1} \sum_{j=1}^{T} \hat{\boldsymbol{f}}^*(\tau_j) \boldsymbol{\varepsilon}^T(\tau_j) \boldsymbol{\varepsilon}(t) = (III.1) + (III.2) + (III.3). \qquad \text{(B.13)}
\end{aligned}$$

By Assumption 2.3.5.c, we have the followings:

$$\begin{aligned}(III.1) &= \hat{\boldsymbol{\rho}}^{-1}n^{-1}T^{-1}\sum_{j=1}^{T}\hat{\boldsymbol{f}}^{*}(\tau_j)\boldsymbol{f}^{*T}(\tau_j)\boldsymbol{\Lambda}^{*T}(\tau_j)\boldsymbol{\varepsilon}(t)\\ &= \hat{\boldsymbol{\rho}}^{-1}T^{-1}\sum_{j=1}^{T}\hat{\boldsymbol{f}}^{*}(\tau_j)\boldsymbol{f}^{*T}(\tau_j)\left\{n^{-1}\sum_{i=1}^{n}\boldsymbol{\lambda}_i^{*}(\tau_j)\varepsilon_i(t)\right\}\\ &= \hat{\boldsymbol{\rho}}^{-1}T^{-1}\sum_{j=1}^{T}\hat{\boldsymbol{f}}^{*}(\tau_j)\boldsymbol{f}^{*T}(\tau_j)\mathcal{O}_p\left(n^{-1/2}\right)=\mathcal{O}_p\left(n^{-1/2}\right), \end{aligned} \tag{B.14}$$

and similarly,

$$\begin{aligned}(III.2) &= \hat{\boldsymbol{\rho}}^{-1}n^{-1}T^{-1}\sum_{j=1}^{T}\hat{\boldsymbol{f}}^{*}(\tau_j)\boldsymbol{\varepsilon}^{T}(\tau_j)\boldsymbol{\Lambda}^{*}(t)\boldsymbol{f}^{*}(t)\\ &= \hat{\boldsymbol{\rho}}^{-1}T^{-1}\sum_{j=1}^{T}\hat{\boldsymbol{f}}^{*}(\tau_j)\left\{n^{-1}\sum_{i=1}^{n}\varepsilon_i(\tau_j)\boldsymbol{\lambda}_i^{*T}(t)\right\}\boldsymbol{f}^{*}(t)=\mathcal{O}_p\left(n^{-1/2}\right); \end{aligned} \tag{B.15}$$

by Assumption 2.3.5.b,

$$\begin{aligned}(III.3) &= \hat{\boldsymbol{\rho}}^{-1}n^{-1}T^{-1}\sum_{j=1}^{T}\hat{\boldsymbol{f}}^{*}(\tau_j)\boldsymbol{\varepsilon}^{T}(\tau_j)\boldsymbol{\varepsilon}(t)\\ &= \hat{\boldsymbol{\rho}}^{-1}T^{-1}\sum_{j=1}^{T}\hat{\boldsymbol{f}}^{*}(\tau_j)\left\{n^{-1}\sum_{i=1}^{n}\varepsilon_i(\tau_j)\varepsilon_i(t)\right\}=o_p\left(n^{-1/2}\right). \end{aligned} \tag{B.16}$$

Therefore, $(III)=\mathcal{O}_p\left(n^{-1/2}\right)$. Applying triangle inequality, we then have

$$\left\|\hat{\boldsymbol{f}}^{*}(t)-(\boldsymbol{W}\boldsymbol{f})(t)\right\|\leq\|(I)\|+\|(II)\|+\|(III)\|=\mathcal{O}_p\left(\max\left\{T^{-1},J^{-4/9},n^{-1/2}\right\}\right)=o_p(1). \tag{B.17}$$

□

The proof of Theorem 2.3.1.b

To separate the convergence rates of different estimators, we now introduce another index set, $\{s_j\}_{j=1}^{S}$, consisting of points that are equally spaced on $[0,1]$, where $S\to\infty$. Recall the

definition of $\boldsymbol{W}^*(t)$, we have $x_i(t) = \boldsymbol{\lambda}_i^{*T}(t)\boldsymbol{f}^*(t) = \boldsymbol{\lambda}_i^{*T}(t)\boldsymbol{W}^*(t)\int_0^t \boldsymbol{\Psi}(s)\hat{\boldsymbol{f}}(s)ds$. Note that the continuous function $\boldsymbol{W}^{*T}(t)\boldsymbol{\lambda}_i^*(T)$ has an arbitrarily close polynomial approximation, then again, by Assumption 2.3.1 and Theorem 2 from Claeskens et al. (2009), we can find some coefficients, say ${\boldsymbol{b}^0}_i$, for the base $\boldsymbol{\Theta}(\cdot)$, such that

$$\boldsymbol{W}^{*T}(t)\boldsymbol{\lambda}_i^*(T) - \boldsymbol{\Theta}^T(t){\boldsymbol{b}^0}_i = \mathcal{O}_p\left(J^{-4/9}\right),$$

and therefore,

$$\begin{aligned} x_i(t) &= \boldsymbol{\lambda}_i^{*T}(t)\boldsymbol{W}^*(t)\int_0^t \boldsymbol{\Psi}(s)\hat{\boldsymbol{f}}(s)ds \\ &= {\boldsymbol{b}^{0T}}_i\boldsymbol{\Theta}(t)\int_0^t \boldsymbol{\Psi}(s)\hat{\boldsymbol{f}}(s)ds + \mathcal{O}_p\left(J^{-4/9}\right) = {\boldsymbol{b}^{0T}}_i\boldsymbol{\Omega}_\lambda^T(t) + \mathcal{O}_p\left(J^{-4/9}\right); \end{aligned}$$

a penalized least squares estimator of ${\boldsymbol{b}^{0T}}_i$ can then be expressed as

$$\hat{\boldsymbol{b}}_i = \left\{\int_0^1 \hat{\boldsymbol{\Omega}}_\lambda^T(t)\hat{\boldsymbol{\Omega}}_\lambda(t)dt + \gamma_\lambda\int_0^1 \boldsymbol{\Theta}''(t)\boldsymbol{\Theta}''^T(t)dt\right\}^{-1}\left\{\int_0^1 \hat{\boldsymbol{\Omega}}_\lambda^T(t)\tilde{x}_i(t)dt\right\}.$$

Hence, we have the following:

$$\begin{aligned} &\int_0^t \hat{\boldsymbol{\lambda}}_i^T(t,s)\hat{\boldsymbol{f}}(s)ds - \int_0^t \boldsymbol{\lambda}_i^T(t,s)\boldsymbol{f}(s)ds = \hat{\boldsymbol{b}}_i^T\boldsymbol{\Omega}_\lambda^T(t) - \int_0^t \boldsymbol{\lambda}_i^T(t,s)\boldsymbol{f}(s)ds \\ =\ &\left\{\hat{\boldsymbol{b}}_i^T\hat{\boldsymbol{\Omega}}_\lambda^T(t) - S^{-1}\sum_{j=1}^S x_i(s_j)\hat{\boldsymbol{\Omega}}_\lambda(s_j)\hat{\boldsymbol{R}}_\lambda^{-T}\hat{\boldsymbol{\Omega}}_\lambda^T(t)\right\} + \\ &\left\{S^{-1}\sum_{j=1}^S x_i(s_j)\hat{\boldsymbol{\Omega}}_\lambda(s_j)\hat{\boldsymbol{R}}_\lambda^{-T}\hat{\boldsymbol{\Omega}}_\lambda^T(t) - \int_0^t \boldsymbol{\lambda}_i^T(t,s)\boldsymbol{f}(s)ds\right\} \\ =\ &(IV) + (V). \end{aligned} \tag{B.18}$$

Substituting in previous results, it follows that

$$\begin{aligned} (IV) &= \hat{\boldsymbol{b}}_i^T\hat{\boldsymbol{\Omega}}_\lambda^T(t) - S^{-1}\sum_{j=1}^S x_i(s_j)\hat{\boldsymbol{\Omega}}_\lambda(s_j)\hat{\boldsymbol{R}}_\lambda^{-T}\hat{\boldsymbol{\Omega}}_\lambda^T(t) \\ &= \left\{\int_0^1 \tilde{x}_i(s)\hat{\boldsymbol{\Omega}}_\lambda(s)ds - S^{-1}\sum_{j=1}^S x_i(s_j)\hat{\boldsymbol{\Omega}}_\lambda(s_j)\right\}\hat{\boldsymbol{R}}_\lambda^{-T}\hat{\boldsymbol{\Omega}}_\lambda^T(t). \end{aligned} \tag{B.19}$$

We again apply Theorem 1 (c) from Chui (1971) and a previous result $\tilde{x}_i(t) = x_i(t) + \mathcal{O}_p\left(J^{-4/9}\right)$, so that

$$\int_0^1 \tilde{x}_i(s)\hat{\boldsymbol{\Omega}}_\lambda(s)ds - S^{-1}\sum_{j=1}^{S} x_i(s_j)\hat{\boldsymbol{\Omega}}_\lambda(s_j) = \mathcal{O}\left(S^{-1}\right) + \mathcal{O}_p\left(J^{-4/9}\right) = \mathcal{O}_p\left(\max\left\{S^{-1}, J^{-4/9}\right\}\right).$$

Hence, under Assumption 2.3.1,

$$(IV) = \mathcal{O}_p\left(\max\left\{S^{-1}, J^{-4/9}\right\}\right)\hat{\boldsymbol{R}}_\lambda^{-T}\hat{\boldsymbol{\Omega}}_\lambda^T(t) = \mathcal{O}_p\left(\max\left\{S^{-1}, J^{-4/9}\right\}\right). \tag{B.20}$$

For (V), by subtracting and adding terms, we have the followings:

$$\begin{aligned}
(V) &= S^{-1}\sum_{j=1}^{S} x_i(s_j)\hat{\boldsymbol{\Omega}}_\lambda(s_j)\hat{\boldsymbol{R}}_\lambda^{-T}\hat{\boldsymbol{\Omega}}_\lambda^T(t) - \int_0^t \boldsymbol{\lambda}_i^T(t,s)\boldsymbol{f}(s)ds \\
&= S^{-1}\sum_{j=1}^{S}\left\{\boldsymbol{\lambda}_i^{*T}(s_j)\boldsymbol{W}^*(s_j) - \boldsymbol{b}^{0T}_i\boldsymbol{\Theta}(s_j)\right\}\int_0^{s_j}\boldsymbol{\Psi}(s)\hat{\boldsymbol{f}}(s)ds\hat{\boldsymbol{\Omega}}_\lambda(s_j)\hat{\boldsymbol{R}}_\lambda^{-T}\hat{\boldsymbol{\Omega}}_\lambda^T(t) + \\
&\quad S^{-1}\sum_{j=1}^{S}\boldsymbol{b}^{0T}_i\boldsymbol{\Theta}(s_j)\int_0^{s_j}\boldsymbol{\Psi}(s)\hat{\boldsymbol{f}}(s)ds\hat{\boldsymbol{\Omega}}_\lambda(s_j)\hat{\boldsymbol{R}}_\lambda^{-T}\hat{\boldsymbol{\Omega}}_\lambda^T(t) - \int_0^t \boldsymbol{\lambda}_i^T(t,s)\boldsymbol{f}(s)ds + \\
&\quad S^{-1}\sum_{j=1}^{S}\varepsilon_i(s_j)\hat{\boldsymbol{\Omega}}_\lambda(s_j)\hat{\boldsymbol{R}}_\lambda^{-T}\hat{\boldsymbol{\Omega}}_\lambda^T(t) = (V.1) + (V.2) + (V.3).
\end{aligned} \tag{B.21}$$

According to Assumption 2.3.1 and previous results, it is implied that

$$\begin{aligned}
(V.1) &= S^{-1}\sum_{j=1}^{S}\left\{\boldsymbol{\lambda}_i^{*T}(s_j)\boldsymbol{W}^*(s_j) - \boldsymbol{b}^{0T}_i\boldsymbol{\Theta}(s_j)\right\}\int_0^{s_j}\boldsymbol{\Psi}(s)\hat{\boldsymbol{f}}(s)ds\hat{\boldsymbol{\Omega}}_\lambda(s_j)\hat{\boldsymbol{R}}_\lambda^{-T}\hat{\boldsymbol{\Omega}}_\lambda^T(t) \\
&= S^{-1}\sum_{j=1}^{S}\mathcal{O}_p\left(J^{-4/9}\right)\int_0^{s_j}\boldsymbol{\Psi}(s)\hat{\boldsymbol{f}}(s)ds\hat{\boldsymbol{\Omega}}_\lambda(s_j)\hat{\boldsymbol{R}}_\lambda^{-T}\hat{\boldsymbol{\Omega}}_\lambda^T(t) = \mathcal{O}_p\left(J^{-4/9}\right),
\end{aligned} \tag{B.22}$$

For (V.2), applying Theorem 1 (c) from Chui (1971) and Lemma B.1, we have

$$
\begin{aligned}
(V.2) &= S^{-1}\sum_{j=1}^{S} \boldsymbol{b}_i^{0T}\boldsymbol{\Theta}(s_j)\int_0^{s_j}\boldsymbol{\Psi}(s)\hat{\boldsymbol{f}}(s)ds\hat{\boldsymbol{\Omega}}_\lambda(s_j)\hat{\boldsymbol{R}}_\lambda^{-T}\hat{\boldsymbol{\Omega}}_\lambda^T(t) - \int_0^t \boldsymbol{\lambda}_i^T(t,s)\boldsymbol{f}(s)ds \\
&= \boldsymbol{b}_i^{0T} S^{-1}\sum_{j=1}^{S}\hat{\boldsymbol{\Omega}}_\lambda^T(s_j)\hat{\boldsymbol{\Omega}}_\lambda(s_j)\left\{\int_0^1 \hat{\boldsymbol{\Omega}}_\lambda^T(s)\hat{\boldsymbol{\Omega}}_\lambda(s)ds\right\}^{-1}\hat{\boldsymbol{\Omega}}_\lambda^T(t) - \int_0^t \boldsymbol{\lambda}_i^T(t,s)\boldsymbol{f}(s)ds + \mathcal{O}\left(\gamma_\lambda\right) \\
&= \boldsymbol{b}_i^{0T}\hat{\boldsymbol{\Omega}}_\lambda^T(t) - \int_0^t \boldsymbol{\lambda}_i^T(t,s)\boldsymbol{f}(s)ds + \mathcal{O}\left(\gamma_\lambda\right) + \mathcal{O}\left(S^{-1}\right) \\
&= \left\{\boldsymbol{b}_i^{0T}\boldsymbol{\Theta}(t) - \boldsymbol{\lambda}_i^{*T}(t)\boldsymbol{W}^*(t)\right\}\int_0^t \boldsymbol{\Psi}(s)\hat{\boldsymbol{f}}(s)ds + \boldsymbol{\lambda}_i^{*T}(t)\boldsymbol{W}^*(t)\int_0^t\boldsymbol{\Psi}(s)\hat{\boldsymbol{f}}(s)ds - \\
&\quad \int_0^t \boldsymbol{\lambda}_i^T(t,s)\boldsymbol{f}(s)ds + \mathcal{O}\left(\gamma_\lambda\right) + \mathcal{O}\left(S^{-1}\right) \\
&= \mathcal{O}_p\left(J^{-4/9}\right) + \boldsymbol{\lambda}_i^{*T}(t)\boldsymbol{f}^*(t) - \int_0^t \boldsymbol{\lambda}_i^T(t,s)\boldsymbol{f}(s)ds + \mathcal{O}\left(\gamma_\lambda\right) + \mathcal{O}\left(S^{-1}\right) \\
&= \mathcal{O}_p\left(\max\left\{J^{-4/9}, \gamma_\lambda, S^{-1}\right\}\right). \qquad \text{(B.23)}
\end{aligned}
$$

By adding and subtracting terms,

$$
\begin{aligned}
(V.3) &= S^{-1}\sum_{j=1}^{S}\varepsilon_i(s_j)\hat{\boldsymbol{\Omega}}_\lambda(s_j)\hat{\boldsymbol{R}}_\lambda^{-T}\hat{\boldsymbol{\Omega}}_\lambda^T(t) \\
&= S^{-1}\sum_{j=1}^{S}\varepsilon_i(s_j)\left\{\int_0^{s_j}\hat{\boldsymbol{f}}^T(s)\boldsymbol{\Psi}^T(s)ds + \hat{\boldsymbol{\mathcal{C}}}_0^T(s_j) - (\boldsymbol{W}\boldsymbol{f})^T(s_j)\boldsymbol{\Psi}^T(s_j)\right\}\boldsymbol{\Theta}^T(s_j)\hat{\boldsymbol{R}}_\lambda^{-T}\hat{\boldsymbol{\Omega}}_\lambda^T(t) + \\
&\quad S^{-1}\sum_{j=1}^{S}\varepsilon_i(s_j)\left\{(\boldsymbol{W}\boldsymbol{f})^T(s_j)\boldsymbol{\Psi}^T(s_j) - \hat{\boldsymbol{\mathcal{C}}}_0^T(s_j)\right\}\boldsymbol{\Theta}^T(s_j)\hat{\boldsymbol{R}}_\lambda^{-T}\hat{\boldsymbol{\Omega}}_\lambda^T(t) = (V.3.a) + (V.3.b), \qquad \text{(B.24)}
\end{aligned}
$$

where $\hat{\boldsymbol{\mathcal{C}}}_0$ is a PK-1 vector of step functions breaking at the knots of $\boldsymbol{\Psi}$ and $\boldsymbol{\Psi}(t)\hat{\boldsymbol{f}}^*(t) = \int_0^t \boldsymbol{\Psi}(s)\hat{\boldsymbol{f}}(s)ds + \hat{\boldsymbol{\mathcal{C}}}_0(t)$, since $\boldsymbol{\Psi}$ is a matrix of order-one B-spline basis functions. Hence, Theorem 2.3.1.a implies that

$$
\begin{aligned}
(V.3.a) &= \frac{1}{S}\sum_{j=1}^{S}\varepsilon_i(s_j)\left\{\int_0^{s_j}\hat{\boldsymbol{f}}^T(s)\boldsymbol{\Psi}^T(s)ds + \hat{\boldsymbol{\mathcal{C}}}_0^T(s_j) - (\boldsymbol{W}\boldsymbol{f})^T(s_j)\boldsymbol{\Psi}^T(s_j)\right\}\boldsymbol{\Theta}^T(s_j)\hat{\boldsymbol{R}}_\lambda^{-T}\hat{\boldsymbol{\Omega}}_\lambda^T(t) \\
&= \frac{1}{S}\sum_{j=1}^{S}\varepsilon_i(s_j)\left\{\hat{\boldsymbol{f}}^{*T}(s_j)\boldsymbol{\Psi}^T(s_j) - (\boldsymbol{W}\boldsymbol{f})^T(s_j)\boldsymbol{\Psi}^T(s_j)\right\}\boldsymbol{\Theta}^T(s_j)\hat{\boldsymbol{R}}_\lambda^{-T}\hat{\boldsymbol{\Omega}}_\lambda^T(t) = \mathcal{O}_p\left(n^{-1/2}\right). \qquad \text{(B.25)}
\end{aligned}
$$

For (V.3.b), it follows from Lemma B.2 and previous results that there exists some PK-1 vector of step functions, say $\boldsymbol{\mathcal{C}}_0(t)$, such that $\hat{\boldsymbol{\mathcal{C}}}_0(t) - \boldsymbol{\mathcal{C}}_0(t) = \mathcal{O}_p\left(n^{-1/2}\right)$ for all t; hence, applying Assumption 2.3.6,

$$\begin{aligned}(V.3.b) &= S^{-1}\sum_{j=1}^{S}\varepsilon_i(s_j)\left\{(\boldsymbol{W}\boldsymbol{f})^T(s_j)\boldsymbol{\Psi}^T(s_j) - \hat{\boldsymbol{\mathcal{C}}}_0^T(s_j)\right\}\boldsymbol{\Theta}^T(s_j)\hat{\boldsymbol{R}}_\lambda^{-T}\hat{\boldsymbol{\Omega}}_\lambda^T(t) \\ &= S^{-1}\sum_{j=1}^{S}\varepsilon_i(s_j)\left\{(\boldsymbol{W}\boldsymbol{f})^T(s_j)\boldsymbol{\Psi}^T(s_j) - \boldsymbol{\mathcal{C}}_0^T(s_j)\right\}\boldsymbol{\Theta}^T(s_j)\hat{\boldsymbol{R}}_\lambda^{-T}\hat{\boldsymbol{\Omega}}_\lambda^T(t) + \mathcal{O}_p\left(n^{-1/2}\right) \\ &= \mathcal{O}_p\left(\max\left\{T^{-1/2}, \frac{1}{\sqrt{S}}\right\}\right) + \mathcal{O}_p\left(n^{-1/2}\right). \end{aligned} \tag{B.26}$$

Therefore, combining the above results, it follows that for all i,

$$\begin{aligned}\left\|\int_0^t \hat{\boldsymbol{\lambda}}_i^T(t,s)\hat{\boldsymbol{f}}(s)ds - \int_0^t \boldsymbol{\lambda}_i^T(t,s)\boldsymbol{f}(s)ds\right\| \leq\ & \|(IV)\| + \|(V)\| \\ &= \mathcal{O}_p\left(\max\left\{J^{-4/9}, \gamma_\lambda, T^{-1/2}, \frac{1}{\sqrt{S}}, n^{-1/2}\right\}\right) = o_p(1).\end{aligned}$$

The proof of Theorem 2.3.2.a

Let $n = o(T)$, then according to the proofs of Theorem 2.3.1.a,

$$\sqrt{n}\left\{\hat{\boldsymbol{f}}^*(t) - (\boldsymbol{W}\boldsymbol{f})(t)\right\} = \sqrt{n}(III) + o_p(1). \tag{B.27}$$

We now check $\sqrt{n}(III)$ through $(III.1)$, $(III.2)$ and $(III.3)$.

Recall that $\int_0^1\left\{\hat{\boldsymbol{f}}^*(t) - \boldsymbol{f}^0(t)\right\}^2 dt = \mathcal{O}_p(n^{-1})$. Under Assumption 2.3.5.c and Slutzky's Lemma, we have

$$\begin{aligned}\sqrt{n}(III.1) &= \hat{\boldsymbol{\rho}}^{-1}T^{-1}\sum_{j=1}^{T}\left\{\hat{\boldsymbol{f}}^*(\tau_j) - \boldsymbol{f}^0(\tau_j) + \boldsymbol{f}^0(\tau_j)\right\}\boldsymbol{f}^{*T}(\tau_j)\left\{n^{-1/2}\sum_{i=1}^{n}\boldsymbol{\lambda}_i^*(\tau_j)\varepsilon_i(t)\right\} \\ &= \hat{\boldsymbol{\rho}}^{-1}T^{-1}\sum_{j=1}^{T}\boldsymbol{f}^0(\tau_j)\boldsymbol{f}^{*T}(\tau_j)\left\{n^{-1/2}\sum_{i=1}^{n}\boldsymbol{\lambda}_i^*(\tau_j)\varepsilon_i(t)\right\} + \\ &\quad \hat{\boldsymbol{\rho}}^{-1}T^{-1}\sum_{j=1}^{T}\left\{\hat{\boldsymbol{f}}^*(\tau_j) - \boldsymbol{f}^0(\tau_j)\right\}\boldsymbol{f}^{*T}(\tau_j)\left\{n^{-1/2}\sum_{i=1}^{n}\boldsymbol{\lambda}_i^*(\tau_j)\varepsilon_i(t)\right\} \\ &= (III.1.a) + (III.1.b),\end{aligned} \tag{B.28}$$

where

$$
\begin{aligned}
(III.1.a) &= \hat{\boldsymbol{\rho}}^{-1}T^{-1}\sum_{j=1}^{T}\boldsymbol{f}^{0}(\tau_j)\boldsymbol{f}^{*T}(\tau_j)\left\{n^{-1/2}\sum_{i=1}^{n}\boldsymbol{\lambda}_i^{*}(\tau_j)\varepsilon_i(t)\right\} \\
&= \hat{\boldsymbol{\rho}}^{-1}n^{-1/2}\sum_{i=1}^{n}\left\{T^{-1}\sum_{j=1}^{T}\boldsymbol{f}^{0}(\tau_j)\boldsymbol{f}^{*T}(\tau_j)\boldsymbol{\lambda}_i^{*}(\tau_j)\right\}\varepsilon_i(t) \xrightarrow{d} N\left(0,\boldsymbol{\rho}^{-1}\Sigma_f(t)\boldsymbol{\rho}^{-1}\right), \\
(III.1.b) &= \hat{\boldsymbol{\rho}}^{-1}T^{-1}\sum_{j=1}^{T}\left\{\hat{\boldsymbol{f}}^{*}(\tau_j)-\boldsymbol{f}^{0}(\tau_j)\right\}\boldsymbol{f}^{*T}(\tau_j)\left\{n^{-1/2}\sum_{i=1}^{n}\boldsymbol{\lambda}_i^{*}(\tau_j)\varepsilon_i(t)\right\} = \mathcal{O}_p\left(n^{-1/2}\right);
\end{aligned}
$$

hence, we have $\sqrt{n}(III.1) \xrightarrow{d} N(0, \boldsymbol{\Psi}(t)\boldsymbol{\rho}^{-1}\Sigma_f(t)\boldsymbol{\rho}^{-1}\boldsymbol{\Psi}^{T}(t))$. Similarly, for $\sqrt{n}(III.2)$,

$$
\begin{aligned}
\sqrt{n}(III.2) &= \sqrt{n}\hat{\boldsymbol{\rho}}^{-1}T^{-1}\sum_{j=1}^{T}\left\{\hat{\boldsymbol{f}}^{*}(\tau_j)-\boldsymbol{f}^{0}(\tau_j)+\boldsymbol{f}^{0}(\tau_j)\right\}\left\{n^{-1}\sum_{i=1}^{n}\varepsilon_i(\tau_j)\boldsymbol{\lambda}_i^{*T}(t)\right\}\boldsymbol{f}^{*}(t) \\
&= \sqrt{n}\hat{\boldsymbol{\rho}}^{-1}T^{-1}\sum_{j=1}^{T}\left\{\hat{\boldsymbol{f}}^{*}(\tau_j)-\boldsymbol{f}^{0}(\tau_j)\right\}\left\{n^{-1}\sum_{i=1}^{n}\varepsilon_i(\tau_j)\boldsymbol{\lambda}_i^{*T}(t)\right\}\boldsymbol{f}^{*}(t)+ \\
&\quad \sqrt{n}\hat{\boldsymbol{\rho}}^{-1}T^{-1}\sum_{j=1}^{T}\boldsymbol{f}^{0}(\tau_j)\left\{n^{-1}\sum_{i=1}^{n}\varepsilon_i(\tau_j)\boldsymbol{\lambda}_i^{*T}(t)\right\}\boldsymbol{f}^{*}(t) = (III.2.a)+(III.2b),
\end{aligned}
\tag{B.29}
$$

where

$$
\begin{aligned}
(III.2.a) &= \sqrt{n}\hat{\boldsymbol{\rho}}^{-1}T^{-1}\sum_{j=1}^{T}\left\{\hat{\boldsymbol{f}}^{*}(\tau_j)-\boldsymbol{f}^{0}(\tau_j)\right\}\left\{n^{-1}\sum_{i=1}^{n}\varepsilon_i(\tau_j)\boldsymbol{\lambda}_i^{*T}(t)\right\}\boldsymbol{f}^{*}(t) = \mathcal{O}_p\left(n^{-1/2}\right), \\
(III.2.b) &= n^{-1}\sum_{i=1}^{n}\hat{\boldsymbol{\rho}}^{-1}\sqrt{n}\left\{T^{-1}\sum_{j=1}^{T}\boldsymbol{f}^{0}(\tau_j)\varepsilon_i(\tau_j)\right\}\boldsymbol{\lambda}_i^{*T}(t)\boldsymbol{f}^{*}(t) = \mathcal{O}_p\left(T^{-1/2}n^{1/2}\right) = o_p(1).
\end{aligned}
$$

For $\sqrt{n}(III.3)$, applying Assumption 2.3.5.b yields

$$
\sqrt{n}(III.3) = \sqrt{n}\hat{\boldsymbol{\rho}}^{-1}T^{-1}\sum_{j=1}^{T}\hat{\boldsymbol{f}}^{*}(\tau_j)\left\{n^{-1}\sum_{i=1}^{n}\varepsilon_i(\tau_j)\varepsilon_i(t)\right\} = o_p(1). \tag{B.30}
$$

Therefore,

$$
\sqrt{n}\left\{\hat{\boldsymbol{f}}^{*}(t)-(\boldsymbol{W}\boldsymbol{f})(t)\right\} \xrightarrow{d} N\left(0,\boldsymbol{\rho}^{-1}\Sigma_f(t)\boldsymbol{\rho}^{-1}\right).
$$

□

The proof of Theorem 2.3.2.b

Let $S = o\left(\min\left\{J^{8/9}, n, T, \gamma_\lambda^{-2}\right\}\right)$, then according to the proofs of Theorems 2.3.1.a, b and 2.3.2.a,

$$\sqrt{S}\left\{\int_0^t \hat{\boldsymbol{\lambda}}_i^T(t,s)\hat{\boldsymbol{f}}(s)ds - \int_0^t \boldsymbol{\lambda}_i^T(t,s)\boldsymbol{f}(s)ds\right\} = \sqrt{S}(V.3) + o_p(1). \tag{B.31}$$

Here, we show that $\sqrt{S}$ inflates $(V.3.b)$ to a distribution, while $\sqrt{S}(V.3.a)$ remains $o_p(1)$.

First, according to previous results,

$$\begin{aligned}\sqrt{S}(V.3) = &\frac{1}{\sqrt{S}}\sum_{j=1}^{S}\varepsilon_i(s_j)\left\{\hat{\boldsymbol{f}}^{*T}(s)\boldsymbol{\Psi}^T(s) - (\boldsymbol{W}\boldsymbol{f})^T(s_j)\boldsymbol{\Psi}^T(s_j)\right\}\boldsymbol{\Theta}^T(s_j)\hat{\boldsymbol{R}}_\lambda^{-T}\hat{\boldsymbol{\Omega}}_\lambda^T(t) + \\ &\frac{1}{\sqrt{S}}\sum_{j=1}^{S}\varepsilon_i(s_j)\left\{(\boldsymbol{W}\boldsymbol{f})^T(s_j)\boldsymbol{\Psi}^T(s_j) - \hat{\boldsymbol{\mathcal{C}}}_0^T(s_j)\right\}\boldsymbol{\Theta}^T(s_j)\hat{\boldsymbol{R}}_\lambda^{-T}\hat{\boldsymbol{\Omega}}_\lambda^T(t) \\ = &\sqrt{S}(V.3.a) + \sqrt{S}(V.3.b),\end{aligned} \tag{B.32}$$

where $\sqrt{S}(V.3.a) = \mathcal{O}_p\left(S^{1/2}n^{-1/2}\right) = o_p(1)$. For $\sqrt{S}(V.3.b)$, recall that $\hat{\boldsymbol{\mathcal{C}}}_0(t) - \boldsymbol{\mathcal{C}}_0(t) = \mathcal{O}_p\left(n^{-1/2}\right)$, and Lemma B.2 implies $(\boldsymbol{W}\boldsymbol{f})(t) - \rho^{-1}n^{-1}\sum_{\iota=1}^n \boldsymbol{\mu}_\iota\boldsymbol{\lambda}_\iota^{*T}(t)\boldsymbol{f}^*(t) = \mathcal{O}_p\left(T^{-1/2}\right)$; hence,

$$\begin{aligned}&\sqrt{S}(V.3.b) \\ = &\frac{1}{\sqrt{S}}\sum_{j=1}^{S}\varepsilon_i(s_j)\left\{\boldsymbol{f}^{*T}(s_j)\frac{1}{n}\sum_{\iota=1}^{n}\boldsymbol{\lambda}_\iota^*(s_j)\boldsymbol{\mu}_\iota^T\rho^{-T}\boldsymbol{\Psi}^T(s_j) - \boldsymbol{\mathcal{C}}_0^T(s_j)\right\}\boldsymbol{\Theta}^T(s_j)\hat{\boldsymbol{R}}_\lambda^{-T}\hat{\boldsymbol{\Omega}}_\lambda^T(t) + o_p(1).\end{aligned} \tag{B.33}$$

Applying Assumption 2.3.6 and the CLT for strong mixing processes,

$$\frac{1}{\sqrt{S}}\sum_{j=1}^{S}\varepsilon_i(s_j)\left\{\boldsymbol{f}^{*T}(s_j)n^{-1}\sum_{\iota=1}^{n}\boldsymbol{\lambda}_\iota^*(s_j)\boldsymbol{\mu}_\iota^T\rho^{-T}\boldsymbol{\Psi}^T(s_j) - \boldsymbol{\mathcal{C}}_0^T(s_j)\right\}\boldsymbol{\Theta}^T(s_j) \xrightarrow{d} N\left(\boldsymbol{\mu}_{\lambda_i,f}, \Sigma_{\lambda_i,f}\right),$$

where

$$\boldsymbol{\mu}_{\lambda_i,f} := \mathbb{E}\left[\frac{1}{\sqrt{S}}\sum_{j=1}^{S}\varepsilon_i(s_j)\left\{\boldsymbol{f}^{*T}(s_j)\frac{1}{n}\sum_{\iota=1}^{n}\boldsymbol{\lambda}_\iota^*(s_j)\boldsymbol{\mu}_\iota^T\rho^{-T}\boldsymbol{\Psi}^T(s_j) - \boldsymbol{\mathcal{C}}_0^T(s_j)\right\}\boldsymbol{\Theta}^T(s_j)\right] = 0,$$

$$\Sigma_{\lambda_i,f} := \lim_{S\to\infty} \mathrm{Var}\left[\frac{1}{\sqrt{S}}\sum_{j=1}^{S}\varepsilon_i(s_j)\left\{\boldsymbol{f}^{*T}(s_j)\frac{1}{n}\sum_{\iota=1}^{n}\boldsymbol{\lambda}_\iota^*(s_j)\boldsymbol{\mu}_\iota^T\rho^{-T}\boldsymbol{\Psi}^T(s_j)-\boldsymbol{\mathcal{C}}_0^T(s_j)\right\}\boldsymbol{\Theta}^T(s_j)\right] < \infty.$$

For $\hat{\boldsymbol{R}}_\lambda^{-T}$ and $\hat{\boldsymbol{\Omega}}_\lambda^T(t)$, according to previous results, we have

$$\begin{aligned}\hat{\boldsymbol{\Omega}}_\lambda(t) &= \int_0^t \hat{\boldsymbol{f}}^T(s)\boldsymbol{\Psi}^T(s)ds\boldsymbol{\Theta}^T(t) = \left\{\hat{\boldsymbol{f}}^{*T}(t)\boldsymbol{\Psi}^T(t)-\hat{\boldsymbol{\mathcal{C}}}_0(t)\right\}\boldsymbol{\Theta}^T(t)\\ &= \left\{\boldsymbol{f}^{0T}(t)\boldsymbol{\Psi}^T(t)-\boldsymbol{\mathcal{C}}_0(t)\right\}\boldsymbol{\Theta}^T(t)+\mathcal{O}_p\left(n^{-1/2}\right).\end{aligned}$$

Then $\hat{\boldsymbol{\Omega}}_\lambda(t)\to\boldsymbol{\Omega}_\lambda(t)$ for all t with $\boldsymbol{\Omega}_\lambda(t) := \left\{\boldsymbol{f}^{0T}(t)\boldsymbol{\Psi}^T(t)-\boldsymbol{\mathcal{C}}_0(t)\right\}\boldsymbol{\Theta}^T(t)$, and by the continuous mapping theorem, we have $\hat{\boldsymbol{R}}_\lambda \to \boldsymbol{R}_\lambda$ with $\boldsymbol{R}_\lambda := \int_0^1\left\{\boldsymbol{\Omega}_\lambda^T(t)\boldsymbol{\Omega}_\lambda(t)\right\}dt + \gamma_\lambda\int_0^1\boldsymbol{\Theta}''(t)\boldsymbol{\Theta}''^T(t)dt$. Therefore, applying Slutzky's Lemma,

$$\sqrt{S}(V.3.b)\xrightarrow{d} N\left(0,\boldsymbol{\Omega}_\lambda(t)\boldsymbol{R}_\lambda^{-1}\Sigma_{\lambda_i,f}\boldsymbol{R}_\lambda^{-T}\boldsymbol{\Omega}_\lambda^T(t)\right),$$

and it is implied that for all t,

$$\sqrt{S}\left\{\int_0^t\hat{\boldsymbol{\lambda}}_i^T(t,s)\hat{\boldsymbol{f}}(s)ds-\int_0^t\boldsymbol{\lambda}_i^T(t,s)\boldsymbol{f}(s)ds\right\}\xrightarrow{d}N\left(0,\boldsymbol{\Omega}_\lambda(t)\boldsymbol{R}_\lambda^{-1}\Sigma_{\lambda_i,f}\boldsymbol{R}_\lambda^{-T}\boldsymbol{\Omega}_\lambda^T(t)\right).$$

B.4 Proofs of lemmas

The proof of Lemma B.1

For any given $\hat{\boldsymbol{\Omega}}_\lambda(t)$, let M_l be a squared matrix, such that

$$M_l^{-1}M_l^{-T} = \int_0^1\hat{\boldsymbol{\Omega}}_\lambda^T(t)\hat{\boldsymbol{\Omega}}_\lambda(t)dt,$$

and let U_l be orthonormal and L_l be diagonal, such that

$$M_l\int_0^1\boldsymbol{\Theta}''(t)\boldsymbol{\Theta}''^T(t)dtM_l^T = U_lL_lU_l^T,$$

which implies that $\int_0^1\boldsymbol{\Theta}''(t)\boldsymbol{\Theta}''^T(t)dt = M_l^{-1}U_lL_lU_l^TM_l^{-T}$. Then we have

$$\begin{aligned}\hat{\boldsymbol{\Omega}}_\lambda(s)\hat{\boldsymbol{R}}_\lambda^{-T}\hat{\boldsymbol{\Omega}}_\lambda^T(t) &= \hat{\boldsymbol{\Omega}}_\lambda(s)\left\{\int_0^1\hat{\boldsymbol{\Omega}}_\lambda^T(\tau)\hat{\boldsymbol{\Omega}}_\lambda(\tau)d\tau+\gamma_\lambda\int_0^1\boldsymbol{\Theta}''(\tau)\boldsymbol{\Theta}''^T(\tau)d\tau\right\}^{-1}\hat{\boldsymbol{\Omega}}_\lambda^T(t)\\ &= \hat{\boldsymbol{\Omega}}_\lambda(s)\left(M_l^{-1}U_lU_l^TM_l^{-T}+\gamma_\lambda M_l^{-1}U_lL_lU_l^TM_l^{-T}\right)^{-1}\hat{\boldsymbol{\Omega}}_\lambda^T(t)\\ &= \hat{\boldsymbol{\Omega}}_\lambda(s)M_l^TU_l^{-T}\left(\mathbf{I}+\gamma_\lambda L_l\right)^{-1}U_l^{-1}M_l\hat{\boldsymbol{\Omega}}_\lambda^T(t).\end{aligned}$$

Since L_l is diagonal, let $l_{l,r}$ be the rth diagonal element, then we have that $(\mathbf{I}+\gamma_\lambda L_l)^{-1}$ is also diagonal, with the rth diagonal element $1-\gamma_\lambda l_{l,r}/(1+\gamma_\lambda l_{l,r})$, which is $1+\mathcal{O}(\gamma_\lambda)$. Therefore, $(\mathbf{I}+\gamma_f L_l)^{-1}=\mathbf{I}+\mathcal{O}(\gamma_\lambda)$, and it follows that

$$\begin{aligned}\hat{\mathbf{\Omega}}_\lambda(s)\hat{\boldsymbol{R}}_\lambda^{-T}\hat{\mathbf{\Omega}}_\lambda^T(t) &= \hat{\mathbf{\Omega}}_\lambda(s)M_l^T U_l^{-T}\left(\mathbf{I}+\gamma_\lambda L_f\right)^{-1}U_l^{-1}M_l\hat{\mathbf{\Omega}}_\lambda^T(t)\\ &= \hat{\mathbf{\Omega}}_\lambda(s)\left\{\int_0^1\hat{\mathbf{\Omega}}_\lambda^T(\tau)\hat{\mathbf{\Omega}}_\lambda(\tau)d\tau\right\}^{-1}\hat{\mathbf{\Omega}}_\lambda^T(t)+\mathcal{O}(\gamma_\lambda).\end{aligned}$$

□

The proof of Lemma B.2

Under Assumptions 2.3.3 and 2.3.6.c, we have

$$\begin{aligned}&\mathbb{E}\left\{T^{-1}\sum_{j=1}^T\boldsymbol{f}^0(\tau_j)\boldsymbol{f}^{*T}(\tau_j)\boldsymbol{\lambda}_i^*(\tau_j)\right\}\\ &=T^{-1}\sum_{j=1}^T\boldsymbol{f}^0(\tau_j)\mathbb{E}\left\{\boldsymbol{f}^{*T}(\tau_j)\right\}\boldsymbol{\lambda}_i^*(\tau_j)=T^{-1}\sum_{j=1}^T\boldsymbol{f}^0(\tau_j)\int_0^{\tau_j}\mathbb{E}\left\{\boldsymbol{f}^T(s)\right\}\mathbf{\Psi}^T(s)ds\boldsymbol{\lambda}_i^*(\tau_j)\\ &=T^{-1}\sum_{j=1}^T\boldsymbol{f}^0(\tau_j)\int_0^{\tau_j}\boldsymbol{\mu}_f(s)\mathbf{\Psi}^T(s)ds\boldsymbol{\lambda}_i^*(\tau_j)\\ &=\int_0^1\boldsymbol{f}^0(t)\int_0^t\boldsymbol{\mu}_f(s)\mathbf{\Psi}^T(s)ds\boldsymbol{\lambda}_i^*(t)dt+\mathcal{O}\left(T^{-1}\right)=\boldsymbol{\mu}_i+\mathcal{O}\left(T^{-1}\right);\end{aligned}$$

$$\operatorname{Var}\left\{T^{-1}\sum_{j=1}^T\boldsymbol{f}^0(\tau_j)\boldsymbol{f}^{*T}(\tau_j)\boldsymbol{\lambda}_i^*(\tau_j)\right\}=T^{-2}\operatorname{Var}\left\{\sum_{j=1}^T\boldsymbol{f}^0(\tau_j)\boldsymbol{f}^{*T}(\tau_j)\boldsymbol{\lambda}_i^*(\tau_j)\right\}=\mathcal{O}\left(T^{-1}\right).$$

□

C Appendices of Chapter 3

This appendix contains the proofs of the theorems and the to-be-stated Lemmas for Chapter 3.

C.1 Proofs of Theorems

To simplify the notations, we define $\mathbf{\Lambda} := \lambda_\Psi \mathbf{\Lambda}_\Psi + \lambda_\Theta \mathbf{\Lambda}_\Theta$. Then we begin the proofs by stating the following lemmas.

Lemma C.1 *Let f be a function that maps a squared matrix to a real value; then for full rank squared matrices X, Y and Z, say with dimension J-by-J, there is*

$$f(\boldsymbol{M}_1) = f(\boldsymbol{M}_2) + \operatorname{tr}\left\{ \left[\frac{\partial f\left(\bar{\boldsymbol{M}}\right)}{\partial \bar{\boldsymbol{M}}} \right]^T (\boldsymbol{M}_1 - \boldsymbol{M}_2) \right\},$$

where $\min\{\boldsymbol{M}_{1,ij}, \boldsymbol{M}_{2,ij}\} < \bar{\boldsymbol{M}}_{ij} < \max\{\boldsymbol{M}_{1,ij}, \boldsymbol{M}_{2,ij}\}$ *for all elements* $\boldsymbol{M}_{1,ij}$, $\boldsymbol{M}_{2,ij}$ *and* $\bar{\boldsymbol{M}}_{ij}$ *of the matrices* $\boldsymbol{M}_1$, $\boldsymbol{M}_2$ *and* $\bar{\boldsymbol{M}}$, *respectively, with* $i, j = 1, ..., J$.

Lemma C.2 *Under Assumptions 3.3.2 and 3.3.3, we have*
a. $n^{-1}\sum_{i=1}^n \tilde{\boldsymbol{X}}_i(\xi)\left[\tilde{Y}_i(\tau) - Y_i(\tau)\right] \in o_p\left(n^{\rho-1}\right)$ *for all* $\xi, \tau \in \mathcal{T}$,
b. $n^{-1}\sum_{i=1}^n \left[\tilde{\boldsymbol{X}}_i(\xi) - \boldsymbol{X}_i(\xi)\right] Y_i(\tau) \in o_p\left(n^{\rho-1}\right)$ *for all* $\xi, \tau \in \mathcal{T}$.

Lemma C.3 *Suppose Assumptions 3.3.2 and 3.3.3 hold. Given the bases* $\mathbf{\Psi}$ *and* $\boldsymbol{\theta}$, *we have the following:*
a. $\sup_{t\in\mathcal{T}} \left\| \underline{\tilde{\boldsymbol{X}}}_i(t) - \underline{\boldsymbol{X}}_i(t) \right\|_F \in o_p\left(n^{\rho-1}\right)$;
b. $\bar{t}^{-1}\int_0^{\bar{t}} n^{-1}\sum_{i=1}^n \underline{\tilde{\boldsymbol{X}}}_i(\tau)\underline{\tilde{\boldsymbol{X}}}_i{}^T(\tau)d\tau - \bar{t}^{-1}\int_0^{\bar{t}} n^{-1}\sum_{i=1}^n \underline{\boldsymbol{X}}_i(\tau)\underline{\boldsymbol{X}}_i{}^T(\tau)d\tau \in o_p\left(n^{\rho-1}\right)$;
c. for given $(s,t), (\tau,\xi) \in \mathcal{T}_2$,

$$\begin{aligned}\mathbf{\Psi}(s)\mathbf{\Theta}(t)\left[\bar{t}^{-1}\int_0^{\bar{t}} n^{-1}\sum_{i=1}^n \underline{\tilde{\boldsymbol{X}}}_i(t')\underline{\tilde{\boldsymbol{X}}}_i{}^T(t')dt' + \mathbf{\Lambda}\right]^{-1}\mathbf{\Theta}^T(\tau)\mathbf{\Psi}^T(\xi) - \\ \mathbf{\Psi}(s)\mathbf{\Theta}(t)\left[\bar{t}^{-1}\int_0^{\bar{t}} n^{-1}\sum_{i=1}^n \underline{\boldsymbol{X}}_i(\tau)\underline{\boldsymbol{X}}_i{}^T(\tau)d\tau\right]^{-1}\mathbf{\Theta}^T(\tau)\mathbf{\Psi}^T(\xi) \in o_p\left(n^{\rho-1}\right).\end{aligned}$$

Lemma C.4 *Under Assumption 3.3.7, we have*
a. $n^{-1}\sum_{i=1}^n \tilde{\boldsymbol{X}}_i(\xi)\left[\tilde{Y}_i(\tau) - Y_i(\tau)\right] \in o_p\left(n^{\rho-1}\bar{t}^{-1/2}\right)$ *for all* $\xi, \tau \in \mathcal{T}$,
b. $n^{-1}\sum_{i=1}^n \left[\tilde{\boldsymbol{X}}_i(\xi) - \boldsymbol{X}_i(\xi)\right] Y_i(\tau) \in o_p\left(n^{\rho-1}\bar{t}^{-1/2}\right)$ *for all* $\xi, \tau \in \mathcal{T}$.

Lemma C.5 *Suppose Assumption 3.3.7 holds. Given the bases* $\boldsymbol{\Psi}$ *and* $\boldsymbol{\theta}$*, we have the following:*
a. $\sup_{t\in\mathcal{T}} \left\| \underline{\tilde{\boldsymbol{X}}}_i(t) - \underline{\boldsymbol{X}}_i(t) \right\|_F \in o_p\left(n^{\rho-1}\bar{t}^{-1/2}\right)$*;*
b. $\bar{t}^{-1}\int_0^{\bar{t}} n^{-1}\sum_{i=1}^n \underline{\tilde{\boldsymbol{X}}}_i(\tau)\underline{\tilde{\boldsymbol{X}}}_i{}^T(\tau)d\tau - \bar{t}^{-1}\int_0^{\bar{t}} n^{-1}\sum_{i=1}^n \underline{\boldsymbol{X}}_i(\tau)\underline{\boldsymbol{X}}_i{}^T(\tau)d\tau \in o_p\left(n^{\rho-1}\bar{t}^{-1/2}\right)$*;*
c. for given $(s,t),(\tau,\xi)\in\mathcal{T}_2$*,*

$$\boldsymbol{\Psi}(s)\boldsymbol{\Theta}(t)\left[\bar{t}^{-1}\int_0^{\bar{t}} n^{-1}\sum_{i=1}^n \underline{\tilde{\boldsymbol{X}}}_i(t')\underline{\tilde{\boldsymbol{X}}}_i{}^T(t')dt' + \boldsymbol{\Lambda}\right]^{-1}\boldsymbol{\Theta}^T(\tau)\boldsymbol{\Psi}^T(\xi) -$$
$$\boldsymbol{\Psi}(s)\boldsymbol{\Theta}(t)\left[\bar{t}^{-1}\int_0^{\bar{t}} n^{-1}\sum_{i=1}^n \underline{\boldsymbol{X}}_i(\tau)\underline{\boldsymbol{X}}_i{}^T(\tau)d\tau\right]^{-1}\boldsymbol{\Theta}^T(\tau)\boldsymbol{\Psi}^T(\xi) \in o_p\left(n^{\rho-1}\bar{t}^{-1/2}\right).$$

Lemma C.6 *Suppose Assumptions 3.3.2, and 3.3.4 to 3.3.9 hold, then as* $\bar{t}\to\infty$*, we have the following:*
a. $\bar{t}^{-1}\int_0^{\bar{t}} n^{-1}\sum_{i=1}^n \underline{\tilde{\boldsymbol{X}}}_i^*(\tau)\underline{\tilde{\boldsymbol{X}}}_i^{*T}(\tau)d\tau - \bar{t}^{-1}\int_0^{\bar{t}} n^{-1}\sum_{i=1}^n \underline{\boldsymbol{X}}_i(\tau)\underline{\boldsymbol{X}}_i{}^T(\tau)d\tau \in o_p\left(n^{\rho-1}\bar{t}^{-1/2}\right)$*;*
b. for given $(s,t),(\tau,\xi)\in\mathcal{T}_2$*,*

$$\boldsymbol{\Psi}(s)\boldsymbol{\Theta}(t)\left[\bar{t}^{-1}\int_0^{\bar{t}} n^{-1}\sum_{i=1}^n \underline{\tilde{\boldsymbol{X}}}_i^*(t')\underline{\tilde{\boldsymbol{X}}}_i^{*T}(t')dt' + \boldsymbol{\Lambda}\right]^{-1}\boldsymbol{\Theta}^T(\tau)\boldsymbol{\Psi}^T(\xi) -$$
$$\boldsymbol{\Psi}(s)\boldsymbol{\Theta}(t)\left[\bar{t}^{-1}\int_0^{\bar{t}} n^{-1}\sum_{i=1}^n \underline{\boldsymbol{X}}_i(t')\underline{\boldsymbol{X}}_i{}^T(t')dt'\right]^{-1}\boldsymbol{\Theta}^T(\tau)\boldsymbol{\Psi}^T(\xi) \in o_p\left(n^{\rho-1}\bar{t}^{-1/2}\right).$$

Proof of Theorem 3.3.1

Extending the arguments in the proof of Theorem 2.1 by Zhou et al. (1998) and the proof of Lemma 6.10 by Agarwal and Studden (1980), it is implied that there exists some $\gamma\in\Gamma(D,\kappa_\psi,\kappa_\theta)$ where $\gamma(s,t)=\boldsymbol{\Psi}(s)\boldsymbol{\Theta}(t)\boldsymbol{b}$, so that $\|\gamma(s,t)-\boldsymbol{\beta}(s,t)\|_F \in o\left(\left[\min\{P,Q\}\right]^{-D}\right)$ for all $(s,t)\in\mathcal{T}_2$; then by adding and subtracting terms as well as the triangle inequality, we can write

$$\begin{aligned}\left\|\hat{\boldsymbol{\beta}}(s,t)-\boldsymbol{\beta}(s,t)\right\|_F &\le \left\|\hat{\boldsymbol{\beta}}(s,t)-\gamma(s,t)\right\|_F + \|\gamma(s,t)-\boldsymbol{\beta}(s,t)\|_F \\ &= \left\|\hat{\boldsymbol{\beta}}(s,t)-\gamma(s,t)\right\|_F + o\left(\left[\min\{P,Q\}\right]^{-D}\right).\end{aligned}$$

Meanwhile,

$$
\begin{aligned}
&\hat{\boldsymbol{\beta}}(s,t)-\gamma(s,t)\\
&=\boldsymbol{\Psi}(s)\boldsymbol{\Theta}(t)\left[\int_0^{\bar{t}} n^{-1}\sum_{i=1}^{n}\underline{\tilde{\boldsymbol{X}}}_i(\tau)\underline{\tilde{\boldsymbol{X}}}_i{}^T(\tau)d\tau+\boldsymbol{\Lambda}\right]^{-1}\int_0^{\bar{t}} n^{-1}\sum_{i=1}^{n}\underline{\tilde{\boldsymbol{X}}}_i(\tau)\tilde{Y}_i(\tau)d\tau-\boldsymbol{\Psi}(s)\boldsymbol{\Theta}(t)\boldsymbol{b}\\
&=\left\{\boldsymbol{\Psi}(s)\boldsymbol{\Theta}(t)\left[\frac{1}{\bar{t}}\int_0^{\bar{t}}\frac{1}{n}\sum_{i=1}^{n}\underline{\tilde{\boldsymbol{X}}}_i(\tau)\underline{\tilde{\boldsymbol{X}}}_i{}^T(\tau)d\tau+\boldsymbol{\Lambda}\right]^{-1}\frac{1}{\bar{t}}\int_0^{\bar{t}}\frac{1}{n}\sum_{i=1}^{n}\underline{\tilde{\boldsymbol{X}}}_i(\tau)\left[\tilde{Y}_i(\tau)-Y_i(\tau)\right]d\tau\right\}+\\
&\quad\left\{\boldsymbol{\Psi}(s)\boldsymbol{\Theta}(t)\left[\frac{1}{\bar{t}}\int_0^{\bar{t}}\frac{1}{n}\sum_{i=1}^{n}\underline{\tilde{\boldsymbol{X}}}_i(\tau)\underline{\tilde{\boldsymbol{X}}}_i{}^T(\tau)d\tau+\boldsymbol{\Lambda}\right]^{-1}\frac{1}{\bar{t}}\int_0^{\bar{t}}\frac{1}{n}\sum_{i=1}^{n}\underline{\tilde{\boldsymbol{X}}}_i(\tau)Y_i(\tau)d\tau-\boldsymbol{\Psi}(s)\boldsymbol{\Theta}(t)\boldsymbol{b}\right\}\\
&=(I)+(II).
\end{aligned}
$$

By Assumptions 3.3.2 to 3.3.4 and Lemma C.3.b, $\bar{t}^{-1}\int_0^{\bar{t}} n^{-1}\sum_{i=1}^{n}\underline{\tilde{\boldsymbol{X}}}_i(\tau)\underline{\tilde{\boldsymbol{X}}}_i{}^T(\tau)d\tau+\boldsymbol{\Lambda}$ is a positive definite symmetric matrix. Note that for any rank R positive definite symmetric matrix $\boldsymbol{M}$ with singular values $\{\zeta_r(\boldsymbol{M})\}_{r=1}^{R}$, applying the facts $\|\boldsymbol{M}\|_{max}\leq\|\boldsymbol{M}\|_2$[7] and $\|\boldsymbol{M}^{-1}\|_2=[\min\{\zeta_r(\boldsymbol{M})\}_r]^{-1}$ yields $\|\boldsymbol{M}^{-1}\|_{max}\leq\|\boldsymbol{M}^{-1}\|_2=[\min\{\zeta_r(\boldsymbol{M})\}_r]^{-1}\in\mathcal{O}(1)$; hence, $\left[\bar{t}^{-1}\int_0^{\bar{t}} n^{-1}\sum_{i=1}^{n}\underline{\tilde{\boldsymbol{X}}}_i(\tau)\underline{\tilde{\boldsymbol{X}}}_i{}^T(\tau)d\tau+\boldsymbol{\Lambda}\right]^{-1}\in\mathcal{O}_p(1)$ given $\boldsymbol{\Psi}(s)$ and $\boldsymbol{\Theta}(t)$. Meantime, due to the "locally nonzero" property of $\boldsymbol{\psi}(s)$ and $\boldsymbol{\theta}(t)$ of finite orders, the product $\boldsymbol{\Psi}(s)\boldsymbol{\Theta}(t)$ at any point $(s,t)\in\mathcal{T}_2$ is a $K\times QPK$ matrix with only finitely many non-zero elements of order $\mathcal{O}(1)$. Hence, it is implied that for $\boldsymbol{\Psi}(s)\boldsymbol{\Theta}(t)$ and $\boldsymbol{\Psi}(\xi)\boldsymbol{\Theta}(\tau)$ at fixed s, t and τ where $(s,t),(\xi,\tau)\in\mathcal{T}_2$, $\boldsymbol{\Psi}(s)\boldsymbol{\Theta}(t)\left[\bar{t}^{-1}\int_0^{\bar{t}} n^{-1}\sum_{i=1}^{n}\underline{\tilde{\boldsymbol{X}}}_i(\tau)\underline{\tilde{\boldsymbol{X}}}_i{}^T(\tau)d\tau+\boldsymbol{\Lambda}\right]^{-1}\boldsymbol{\Theta}^T(\tau)\boldsymbol{\Psi}^T(\xi)$ as a function of ξ is $\mathcal{O}_p(1)$ only on a finite interval of ξ and zero elsewhere. Then applying

[7] $\|\cdot\|_2$ and $\|\cdot\|_{max}$ are two matrix norms, such that $\|\boldsymbol{M}\|_2=\max\{\zeta_r(\boldsymbol{M})\}_r$ and $\|\boldsymbol{M}\|_{max}=\max_{ij}|m_{ij}|$, where $\max\{\zeta_r(\boldsymbol{M})\}_r$ denotes the largest singular value of $\boldsymbol{M}$ and m_{ij} the element on the ith row and jth column of $\boldsymbol{M}$.

Assumptions 3.3.2, 3.3.3 and Lemma C.2.a, we have the following:

$$
\begin{aligned}
(I) &= \boldsymbol{\Psi}(s)\boldsymbol{\Theta}(t)\left[\frac{1}{\bar{t}}\int_0^{\bar{t}}\frac{1}{n}\sum_{i=1}^{n}\tilde{\underline{\boldsymbol{X}}}_i(\tau)\tilde{\underline{\boldsymbol{X}}}_i{}^T(\tau)d\tau+\boldsymbol{\Lambda}\right]^{-1}\frac{1}{\bar{t}}\int_0^{\bar{t}}\frac{1}{n}\sum_{i=1}^{n}\tilde{\underline{\boldsymbol{X}}}_i(\tau)\left[\tilde{Y}_i(\tau)-Y_i(\tau)\right]d\tau\\
&= \frac{1}{\bar{t}}\int_0^{\bar{t}}\int_0^{\tau}\boldsymbol{\Psi}(s)\boldsymbol{\Theta}(t)\left[\frac{1}{\bar{t}}\int_0^{\bar{t}}\frac{1}{n}\sum_{i=1}^{n}\tilde{\underline{\boldsymbol{X}}}_i(\tau)\tilde{\underline{\boldsymbol{X}}}_i{}^T(\tau)d\tau+\boldsymbol{\Lambda}\right]^{-1}\\
&\qquad\qquad \boldsymbol{\Theta}^T(\tau)\boldsymbol{\Psi}^T(\xi)\frac{1}{n}\sum_{i=1}^{n}\tilde{\boldsymbol{X}}_i(\xi)\left[\tilde{Y}_i(\tau)-Y_i(\tau)\right]d\xi d\tau\\
&= \frac{1}{\bar{t}}\int_0^{\bar{t}}\mathcal{O}_p(1)o_p\left(n^{\rho-1}\right)d\tau = o_p\left(n^{\rho-1}\right).
\end{aligned}
$$

For (II), under Assumptions 3.3.1 to 3.3.3, Lemmas C.2.b and C.3.c, substituting in (3.7) yields

$$
\begin{aligned}
(II) &= \boldsymbol{\Psi}(s)\boldsymbol{\Theta}(t)\left[\frac{1}{\bar{t}}\int_0^{\bar{t}}\frac{1}{n}\sum_{i=1}^{n}\tilde{\underline{\boldsymbol{X}}}_i(\tau)\tilde{\underline{\boldsymbol{X}}}_i{}^T(\tau)d\tau+\boldsymbol{\Lambda}\right]^{-1}\frac{1}{\bar{t}}\int_0^{\bar{t}}\frac{1}{n}\sum_{i=1}^{n}\tilde{\underline{\boldsymbol{X}}}_i(\tau)Y_i(\tau)d\tau-\boldsymbol{\Psi}(s)\boldsymbol{\Theta}(t)\boldsymbol{b}\\
&= \boldsymbol{\Psi}(s)\boldsymbol{\Theta}(t)\left[\frac{1}{\bar{t}}\int_0^{\bar{t}}\frac{1}{n}\sum_{i=1}^{n}\underline{\boldsymbol{X}}_i(\tau)\underline{\boldsymbol{X}}_i{}^T(\tau)d\tau\right]^{-1}\frac{1}{\bar{t}}\int_0^{\bar{t}}\frac{1}{n}\sum_{i=1}^{n}\underline{\boldsymbol{X}}_i(\tau)Y_i(\tau)d\tau-\\
&\quad \boldsymbol{\Psi}(s)\boldsymbol{\Theta}(t)\boldsymbol{b}+o_p\left(n^{\rho-1}\right)\\
&= \boldsymbol{\Psi}(s)\boldsymbol{\Theta}(t)\left[\frac{1}{\bar{t}}\int_0^{\bar{t}}\frac{1}{n}\sum_{i=1}^{n}\underline{\boldsymbol{X}}_i(\tau)\underline{\boldsymbol{X}}_i{}^T(\tau)d\tau\right]^{-1}\frac{1}{\bar{t}}\int_0^{\bar{t}}\frac{1}{n}\sum_{i=1}^{n}\underline{\boldsymbol{X}}_i(\tau)U_i(\tau)d\tau+o_p\left(n^{\rho-1}\right).
\end{aligned}
$$

Under Assumptions 3.3.2, 3.3.5 and 3.3.6, we have $n^{-1}\sum_{i=1}^{n}\boldsymbol{X}_i(\xi)U_i(\tau)d\tau\in\mathcal{O}_p\left(n^{\rho-1}\right)$, and similarly to the previous, $\boldsymbol{\Psi}(s)\boldsymbol{\Theta}(t)\left[\bar{t}^{-1}\int_0^{\bar{t}}n^{-1}\sum_{i=1}^{n}\underline{\boldsymbol{X}}_i(\tau)\underline{\boldsymbol{X}}_i{}^T(\tau)d\tau\right]^{-1}\boldsymbol{\Theta}^T(\tau)\boldsymbol{\Psi}^T(\xi)$ as a function of ξ is $\mathcal{O}_p(1)$ only on a finite interval of ξ and zero elsewhere. Thus, we have

$$
\begin{aligned}
(II) &= \boldsymbol{\Psi}(s)\boldsymbol{\Theta}(t)\left[\frac{1}{\bar{t}}\int_0^{\bar{t}}\frac{1}{n}\sum_{i=1}^{n}\underline{\boldsymbol{X}}_i(\tau)\underline{\boldsymbol{X}}_i{}^T(\tau)d\tau\right]^{-1}\frac{1}{\bar{t}}\int_0^{\bar{t}}\frac{1}{n}\sum_{i=1}^{n}\underline{\boldsymbol{X}}_i(\tau)U_i(\tau)d\tau+o_p\left(n^{\rho-1}\right)\\
&= \frac{1}{\bar{t}}\int_0^{\bar{t}}\int_0^{\tau}\boldsymbol{\Psi}(s)\boldsymbol{\Theta}(t)\left[\frac{1}{\bar{t}}\int_0^{\bar{t}}\frac{1}{n}\sum_{i=1}^{n}\underline{\boldsymbol{X}}_i(\tau)\underline{\boldsymbol{X}}_i{}^T(\tau)d\tau\right]^{-1}\boldsymbol{\Theta}^T(\tau)\boldsymbol{\Psi}^T(\xi)\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{X}_i(\xi)U_i(\tau)d\xi d\tau\\
&\quad +o_p\left(n^{\rho-1}\right)\\
&= \frac{1}{\bar{t}}\int_0^{\bar{t}}\mathcal{O}_p(1)\mathcal{O}_p\left(n^{\rho-1}\right)d\tau = \mathcal{O}_p\left(n^{\rho-1}\right).
\end{aligned}
$$

Therefore, $\left\|\hat{\boldsymbol{\beta}}(s,t)-\gamma(s,t)\right\|_F \leq \|(I)\|_F + \|(II)\|_F = \mathcal{O}_p\left(n^{\rho-1}\right)$ for all $(s,t) \in \mathcal{T}_2$, and under Assumption 3.3.2, $\left\|\hat{\boldsymbol{\beta}}(s,t)-\boldsymbol{\beta}(s,t)\right\|_F = \mathcal{O}_p\left(n^{\rho-1}\right)+o\left(\left[\min\{P,Q\}\right]^{-D}\right) = \mathcal{O}_p\left(n^{\rho-1}\right)$. Since $\hat{\boldsymbol{\beta}} \in \Gamma\left(D,\kappa_\psi,\kappa_\theta\right)$ with $D \in \mathbb{N}$, the estimators $\hat{\boldsymbol{\beta}}$'s are asymptotically stochastically equicontinuous on $\mathcal{T}_2$; hence, the uniform convergence follows, such that $\sup_{(s,t)\in\mathcal{T}_2}\left\|\hat{\boldsymbol{\beta}}(s,t)-\boldsymbol{\beta}(s,t)\right\|_F \in \mathcal{O}_p\left(n^{\rho-1}\right)$.

□

Proof of Theorem 3.3.2

First of all, according to previous results, we have

$$\begin{aligned} n^{1-\rho}\sqrt{\bar{t}}\left[\hat{\boldsymbol{\beta}}(s,t)-\boldsymbol{\beta}(s,t)\right] &= n^{1-\rho}\sqrt{\bar{t}}\left[\hat{\boldsymbol{\beta}}(s,t)-\gamma(s,t)\right] + n^{1-\rho}\sqrt{\bar{t}}\left[\gamma(s,t)-\boldsymbol{\beta}(s,t)\right] \\ &= n^{1-\rho}\sqrt{\bar{t}}(I) + n^{1-\rho}\sqrt{\bar{t}}(II) + o(1); \end{aligned}$$

by Assumptions 3.3.4 to 3.3.7, Lemmas C.3.b and C.4.a,

$$\begin{aligned} n^{1-\rho}\sqrt{\bar{t}}(I) &= n^{1-\rho}\sqrt{\bar{t}}\boldsymbol{\Psi}(s)\boldsymbol{\Theta}(t)\left[\frac{1}{\bar{t}}\int_0^{\bar{t}}\frac{1}{n}\sum_{i=1}^{n}\underline{\tilde{\boldsymbol{X}}}_i(\tau)\underline{\tilde{\boldsymbol{X}}}_i{}^T(\tau)d\tau + \boldsymbol{\Lambda}\right]^{-1} \\ &\qquad \frac{1}{\bar{t}}\int_0^{\bar{t}}\frac{1}{n}\sum_{i=1}^{n}\underline{\tilde{\boldsymbol{X}}}_i(\tau)\left[\tilde{Y}_i(\tau)-Y_i(\tau)\right]d\tau \\ &= \frac{1}{\bar{t}}\int_0^{\bar{t}}\int_0^{\tau}\boldsymbol{\Psi}(s)\boldsymbol{\Theta}(t)\left[\frac{1}{\bar{t}}\int_0^{\bar{t}}\frac{1}{n}\sum_{i=1}^{n}\underline{\tilde{\boldsymbol{X}}}_i(\tau)\underline{\tilde{\boldsymbol{X}}}_i{}^T(\tau)d\tau + \boldsymbol{\Lambda}\right]^{-1} \\ &\qquad \boldsymbol{\Theta}^T(\tau)\boldsymbol{\Psi}^T(\xi)n^{1-\rho}\sqrt{\bar{t}}\frac{1}{n}\sum_{i=1}^{n}\tilde{\boldsymbol{X}}_i(\xi)\left[\tilde{Y}_i(\tau)-Y_i(\tau)\right]d\xi d\tau \\ &= \frac{1}{\bar{t}}\int_0^{\bar{t}}\mathcal{O}_p(1)o_p(1)d\tau = o_p(1), \end{aligned}$$

and by Assumptions 3.3.1, 3.3.7, C.4.b and Lemmas C.5.c, substituting in (3.7) yields

$$
\begin{aligned}
n^{1-\rho}\sqrt{\bar{t}}(II) = n^{1-\rho}\sqrt{\bar{t}}\boldsymbol{\Psi}(s)\boldsymbol{\Theta}(t)\left[\frac{1}{\bar{t}}\int_0^{\bar{t}}\frac{1}{n}\sum_{i=1}^{n}\underline{\boldsymbol{X}}_i(\tau)\underline{\boldsymbol{X}}_i{}^T(\tau)d\tau\right]^{-1} \\
\frac{1}{\bar{t}}\int_0^{\bar{t}}\frac{1}{n}\sum_{i=1}^{n}\underline{\boldsymbol{X}}_i(\tau)U_i(\tau)d\tau + o_p(1) \\
= \sqrt{\bar{t}}\frac{1}{\bar{t}}\int_0^{\bar{t}}\int_0^{\tau}\boldsymbol{\Psi}(s)\boldsymbol{\Theta}(t)\left[\frac{1}{\bar{t}}\int_0^{\bar{t}}\frac{1}{n}\sum_{i=1}^{n}\underline{\boldsymbol{X}}_i(\tau')\underline{\boldsymbol{X}}_i{}^T(\tau')d\tau'\right]^{-1} \\
\boldsymbol{\Theta}^T(\tau)\boldsymbol{\Psi}^T(\xi)\left[n^{1-\rho}\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{X}_i(\xi)U_i(\tau)\right]d\xi d\tau + o_p(1).
\end{aligned}
$$

Since for given s, t and τ, $\boldsymbol{\Psi}(s)\boldsymbol{\Theta}(t)\left[\bar{t}^{-1}\int_0^{\bar{t}}n^{-1}\sum_{i=1}^{n}\underline{\boldsymbol{X}}_i(\tau')\underline{\boldsymbol{X}}_i{}^T(\tau')d\tau'\right]^{-1}\boldsymbol{\Theta}^T(\tau)\boldsymbol{\Psi}^T(\xi)$ as a function of ξ is $\mathcal{O}_p(1)$ only on a finite interval of ξ and zero elsewhere; also, $n^{1-\rho}n^{-1}\sum_{i=1}^{n}\boldsymbol{X}_i(\xi)U_i(\tau) \in \mathcal{O}_p(1)$. Hence, the above equation can be re-written as

$$
n^{1-\rho}\sqrt{\bar{t}}\left[\hat{\boldsymbol{\beta}}(s,t) - \boldsymbol{\beta}(s,t)\right] = \sqrt{\bar{t}}\frac{1}{\bar{t}}\int_0^{\bar{t}}n^{-\rho}\sum_{i=1}^{n}\boldsymbol{\Omega}_i(s,t,\tau)U_i(\tau)d\tau + o_p(1),
$$

where $\boldsymbol{\Omega}_i(s,t,\tau) := \boldsymbol{\Psi}(s)\boldsymbol{\Theta}(t)\left[\bar{t}^{-1}\int_0^{\bar{t}}n^{-1}\sum_{\iota=1}^{n}\underline{\boldsymbol{X}}_\iota(\tau')\underline{\boldsymbol{X}}_\iota{}^T(\tau')d\tau'\right]^{-1}\underline{\boldsymbol{X}}_i(\tau)$. Then by Assumptions 3.3.8 and 3.3.9, $n^{-\rho}\sum_{i=1}^{n}\boldsymbol{\Omega}_i(s,t,\tau_j)U_i(\tau_j) \in \mathcal{O}_p(1)$ is a stationary and ergodic process over τ_j, given any $(s,t) \in \mathcal{T}_2$, and thus, by the CLT for strong mixing processes, for all $(s,t) \in \mathcal{T}_2$, we have the asymptotic normality as such
$\boldsymbol{V}_{\beta,\rho}{}^{-1/2}(s,t)\bar{t}^{-1/2}\int_0^{\bar{t}}n^{-\rho}\sum_{i=1}^{n}\boldsymbol{\Omega}_i(s,t,\tau)U_i(\tau)d\tau \xrightarrow{d} N(\mathbf{0},\boldsymbol{I}_K)$,
where $\boldsymbol{V}_{\beta,\rho}(s,t) := \text{Var}\left(\bar{t}^{-1/2}\int_0^{\bar{t}}n^{-\rho}\sum_{i=1}^{n}\boldsymbol{\Omega}_i(s,t,\tau)U_i(\tau)d\tau\right) \in \mathcal{O}(1)$.

□

Proof of Theorem 3.4.1

Recall that the regression residual is obtained as such $\hat{U}_i(t) = \tilde{Y}_i(t) - \int_{\mathcal{S}_t} \tilde{\boldsymbol{X}}_i(s)\hat{\boldsymbol{\beta}}^T(s,t)ds$ and has the B-spline representation $\tilde{U}_i(t) := \sum_{h=1}^{H} \tilde{w}_{i,h}\eta_h(t)$, where

$$\begin{aligned}\begin{bmatrix}\tilde{w}_{i,1}\\ \vdots \\ \tilde{w}_{i,H}\end{bmatrix} &:= \underset{(r_{i,1},\ldots,r_{i,H})}{\operatorname{argmin}} \frac{1}{J_Y}\sum_{j=1}^{J_Y}\left[\hat{U}_i(t_j) - \sum_{h=1}^{H} r_{i,h}\eta_h(t_j)\right]^2 \\ &= \left[\int_0^{\bar{t}} \boldsymbol{\eta}(\tau)\boldsymbol{\eta}^T(\tau)d\tau\right]^{-1}\left[\frac{1}{J_Y}\sum_{j=1}^{J_Y}\boldsymbol{\eta}(t_j)\hat{U}_i(t_j)\right].\end{aligned}$$

Since $\eta_h(t)$'s are local polynomials, $\int_{\mathcal{T}} \boldsymbol{\eta}(\tau)\boldsymbol{\eta}^T(\tau)d\tau$, and thus, it's inverse $\left[\int_{\mathcal{T}} \boldsymbol{\eta}(\tau)\boldsymbol{\eta}^T(\tau)d\tau\right]^{-1}$ is a block-diagonal matrix; also, each element in the vector $\left[J_Y^{-1}\sum_{j=1}^{J_Y}\boldsymbol{\eta}(t_j)\hat{U}_i(t_j)\right]$ corresponds to a basis function η_h, which is locally non-zero. Hence, intuitively, each estimated basis coefficients in the vector $[\tilde{w}_{i,1},...,\tilde{w}_{i,H}]^T$ summarizes the local information of $\hat{U}_i(t)$ through the corresponding basis function, so the vector $[\tilde{w}_{i,1},...,\tilde{w}_{i,H}]^T$ copies the behaviour of $\hat{U}_i(t)$ over time. A similar argument follows for the estimated basis coefficients $[\tilde{c}_{k,i,1},...,\tilde{c}_{k,i,L}]^T$ for $\tilde{X}_i$'s, where

$$\begin{aligned}\begin{bmatrix}\tilde{c}_{k,i,1}\\ \vdots \\ \tilde{c}_{k,i,L}\end{bmatrix} &:= \underset{(r_{k,i,1},\ldots,r_{k,i,L})}{\operatorname{argmin}} \frac{1}{J_X}\sum_{j=1}^{J_X}\left[X_{kit_j} - \sum_{l=1}^{L} r_{k,i,l}\phi_l(t_j)\right]^2 \\ &= \left[\int_0^{\bar{t}} \boldsymbol{\phi}(\tau)\boldsymbol{\phi}^T(\tau)d\tau\right]^{-1}\left[\frac{1}{J_X}\sum_{j=1}^{J_X}\boldsymbol{\phi}(t_j)X_{kit_j}\right].\end{aligned}$$

Following the bootstrap steps, the "mean-preserving" property of MBB is satisfied for the bootstrap coefficients $\tilde{\boldsymbol{w}}_i^*$ and $\tilde{\boldsymbol{c}}_{ki}^*$, so that Lemma C.6.b holds; furthermore, together with the result

in Theorem 3.3.1, Lemma C.6.c holds. Hence, applying results from above, we have

$$\bar{t}^{1/2}n^{1-\rho}\left[\hat{\boldsymbol{\beta}}^*(s,t)-\hat{\boldsymbol{\beta}}(s,t)\right]=\boldsymbol{\Psi}(s)\boldsymbol{\Theta}(t)\boldsymbol{\Omega}^{-1}_{\underline{\tilde{\boldsymbol{X}}}^*_i,\boldsymbol{\Lambda}}\frac{1}{\sqrt{\bar{t}}}\int_0^{\bar{t}}\frac{1}{n^\rho}\sum_{i=1}^n\underline{\tilde{\boldsymbol{X}}}^*_i(\tau)\tilde{Y}^*_i(\tau)d\tau-\hat{\boldsymbol{\beta}}(s,t)$$

$$=\boldsymbol{\Psi}(s)\boldsymbol{\Theta}(t)\boldsymbol{\Omega}^{-1}_{\underline{\tilde{\boldsymbol{X}}}^*_i,\boldsymbol{\Lambda}}\frac{1}{\sqrt{\bar{t}}}\int_0^{\bar{t}}\frac{1}{n^\rho}\sum_{i=1}^n\underline{\tilde{\boldsymbol{X}}}^*_i(\tau)\left[\tilde{Y}^*_i(\tau)-Y^*_i(\tau)\right]d\tau+$$

$$\boldsymbol{\Psi}(s)\boldsymbol{\Theta}(t)\boldsymbol{\Omega}^{-1}_{\underline{\tilde{\boldsymbol{X}}}^*_i,\boldsymbol{\Lambda}}\frac{1}{\sqrt{\bar{t}}}\int_0^{\bar{t}}\frac{1}{n^\rho}\sum_{i=1}^n\underline{\tilde{\boldsymbol{X}}}^*_i(\tau)\underline{\tilde{\boldsymbol{X}}}^{*T}_i(\tau)d\tau\hat{\boldsymbol{b}}-\bar{t}^{1/2}n^{1-\rho}\hat{\boldsymbol{\beta}}(s,t)+$$

$$\boldsymbol{\Psi}(s)\boldsymbol{\Theta}(t)\boldsymbol{\Omega}^{-1}_{\underline{\tilde{\boldsymbol{X}}}^*_i,\boldsymbol{\Lambda}}\frac{1}{\sqrt{\bar{t}}}\int_0^{\bar{t}}\frac{1}{n^\rho}\sum_{i=1}^n\underline{\tilde{\boldsymbol{X}}}^*_i(\tau)\tilde{U}^*_i(\tau)d\tau=(III)+(IV)+(V).$$

where $\boldsymbol{\Omega}_{\underline{\tilde{\boldsymbol{X}}}^*_i,\boldsymbol{\Lambda}}:=\bar{t}^{-1}\int_0^{\bar{t}}n^{-1}\sum_{i=1}^n\underline{\tilde{\boldsymbol{X}}}^*_i(\tau)\underline{\tilde{\boldsymbol{X}}}^{*T}_i(\tau)d\tau+\boldsymbol{\Lambda}$. Assumption 3.3.7 implies that $(III)\in o_p(1)$, Lemma C.6 implies $(IV)\in o_p(1)$ and that

$$\begin{aligned}\bar{t}^{1/2}n^{1-\rho}\left[\hat{\boldsymbol{\beta}}^*(s,t)-\hat{\boldsymbol{\beta}}(s,t)\right]=&\boldsymbol{\Psi}(s)\boldsymbol{\Theta}(t)\boldsymbol{\Omega}^{-1}_{\underline{\tilde{\boldsymbol{X}}}^*_i,\boldsymbol{\Lambda}}\frac{1}{\sqrt{\bar{t}}}\int_0^{\bar{t}}\frac{1}{n^\rho}\sum_{i=1}^n\underline{\tilde{\boldsymbol{X}}}^*_i(\tau)\tilde{U}^*_i(\tau)d\tau+o_p(1)\\&\xrightarrow{d}N\left(\mathbf{0},\boldsymbol{V}_{\beta,\rho}(s,t)\right),\end{aligned}$$

where $\boldsymbol{V}_{\beta,\rho}(s,t):=\mathrm{Var}\left(n^{1-\rho}\sqrt{\bar{t}}\left[\hat{\boldsymbol{\beta}}(s,t)-\boldsymbol{\beta}(s,t)\right]\right)\in\mathcal{O}(1)$.

□

C.2 Proofs of Lemmas

Proof of Lemma C.1

First, let $\lambda(q):=f(\boldsymbol{M}_2+q(\boldsymbol{M}_1-\boldsymbol{M}_2))$ for $q\in[0,1]$. Then taking the first order derivative of $\lambda(q)$ with respect to q through the matrix argument of the function f yields

$$\begin{aligned}\lambda^{(1)}(q)&=\mathrm{tr}\left\{\left[\frac{\partial f\left(\boldsymbol{M}_2+q(\boldsymbol{M}_1-\boldsymbol{M}_2)\right)}{\partial\left(\boldsymbol{M}_2+q(\boldsymbol{M}_1-\boldsymbol{M}_2)\right)}\right]^T\left[\frac{\partial\left(\boldsymbol{M}_2+q(\boldsymbol{M}_1-\boldsymbol{M}_2)\right)}{\partial q}\right]\right\}\\&=\mathrm{tr}\left\{\left[\frac{\partial f\left(\boldsymbol{M}_2+q(\boldsymbol{M}_1-\boldsymbol{M}_2)\right)}{\partial\left(\boldsymbol{M}_2+q(\boldsymbol{M}_1-\boldsymbol{M}_2)\right)}\right]^T(\boldsymbol{M}_1-\boldsymbol{M}_2)\right\}.\end{aligned}$$

By the mean-value theorem, there exists some $q \in [0,1]$, such that $\lambda(1) - \lambda(0) = \lambda^{(1)}(q)$, which is equivalent to

$$f(\boldsymbol{M}_1) = f(\boldsymbol{M}_2) + \operatorname{tr}\left\{ \left[\frac{\partial f\left(\bar{\boldsymbol{M}}\right)}{\partial \bar{\boldsymbol{M}}} \right]^T \left(\boldsymbol{M}_1 - \boldsymbol{M}_2\right) \right\},$$

where $\min\left\{\boldsymbol{M}_{1,ij}, \boldsymbol{M}_{2,ij}\right\} < \bar{\boldsymbol{M}}_{ij} < \max\left\{\boldsymbol{M}_{1,ij}, \boldsymbol{M}_{2,ij}\right\}$ for all elements $\boldsymbol{M}_{1,ij}$, $\boldsymbol{M}_{2,ij}$ and $\bar{\boldsymbol{M}}_{ij}$ of the matrices $\boldsymbol{M}_1$, $\boldsymbol{M}_2$ and $\bar{\boldsymbol{M}}$, respectively, with $i, j = 1, ..., J$.

□

Proof of Lemma C.2

Proof of part a. By the sub-additivity and the sub-multiplicity of the Frobenius norm, we have

$$\begin{aligned}
\sup_{t\in\mathcal{T}} \left\| \frac{1}{n}\sum_{i=1}^{n} \tilde{\boldsymbol{X}}_i(\xi) \left[\tilde{Y}_i(\tau) - Y_i(\tau)\right] \right\|_F &\leq \frac{1}{n}\sum_{i=1}^{n} \sup_{t\in\mathcal{T}} \left\| \tilde{\boldsymbol{X}}_i(\xi) \left[\tilde{Y}_i(\tau) - Y_i(\tau)\right] \right\|_F \\
&\leq \frac{1}{n}\sum_{i=1}^{n} \sup_{t\in\mathcal{T}} \left\| \tilde{\boldsymbol{X}}_i(\xi) \right\|_F \sup_{t\in\mathcal{T}} \left\| \tilde{Y}_i(\tau) - Y_i(\tau) \right\|_F .
\end{aligned}$$

Applying the fact that $\sup_{t\in\mathcal{T}} \left\| \tilde{\boldsymbol{X}}_i(\xi) \right\|_F \in \mathcal{O}_p(1)$ as well as Assumption 3.3.3, the result in part a follows.

□

Proof of part b. The verification for part b follows the same idea as that for part a, using the convergence rate of $\tilde{\boldsymbol{X}}_i$ from Assumption 3.3.3.

□

Proof of Lemma C.3

Proof of part a. First, recall that $\underline{\tilde{\boldsymbol{X}}}_i^{\,T}(t) := \int_0^t \tilde{\boldsymbol{X}}_i^T(s)\boldsymbol{\Psi}(s)\boldsymbol{\Theta}(t)ds$. By the "local" property of the B-spline basis, for any given pair $(s,t) \in \mathcal{T}_2$, we have $\|\boldsymbol{\Psi}(s)\boldsymbol{\Theta}(t)\|_F \in \mathcal{O}(1)$. Hence, under Assumptions 3.3.2 and 3.3.3, there is $\left\|\boldsymbol{\Theta}^T(t)\boldsymbol{\Psi}^T(s)\right\|_F \sup_{t\in\mathcal{T}} \left\| \tilde{X}_{ki}(t) - X_{ki}(t) \right\|_F \in o_p\left(n^{\rho-1}\right)$, and by the triangle inequality as well as the sub-multiplicity of the Frobenius norm, it follows

that for $t \in \mathcal{S}_t$,

$$\begin{aligned}\sup_{t\in\mathcal{T}} \left\| \underline{\tilde{\boldsymbol{X}}}_i(t) - \underline{\boldsymbol{X}}_i(t) \right\|_F &\leq \int_0^t \left\| \boldsymbol{\Theta}^T(t)\boldsymbol{\Psi}^T(s) \right\|_F \left\| \tilde{X}_{ki}(s) - X_{ki}(s) \right\|_F ds \\ &\leq \int_0^t \left\| \boldsymbol{\Theta}^T(t)\boldsymbol{\Psi}^T(s) \right\|_F \sup_{\tau\in\mathcal{T}} \left\| \tilde{X}_{ki}(\tau) - X_{ki}(\tau) \right\|_F ds = o_p\left(n^{\rho-1}\right),\end{aligned}$$

which justifies the result. □

Proof of part b. Applying the triangle inequality and the sub-multiplicity of the Frobenius norm, the "locally non-zero" property of B-spline as explained in the proof for part a, as well as Assumption 3.3.3, we have

$$\begin{aligned}&\left\| \frac{1}{\bar{t}} \int_0^{\bar{t}} \frac{1}{n} \sum_{i=1}^n \left[\underline{\tilde{\boldsymbol{X}}}_i(s) \underline{\tilde{\boldsymbol{X}}}_i{}^T(s) - \underline{\boldsymbol{X}}_i(s) \underline{\boldsymbol{X}}_i{}^T(s) \right] ds \right\|_F \\ \leq\ & \frac{1}{\bar{t}} \int_0^{\bar{t}} \frac{1}{n} \sum_{i=1}^n \left\| \underline{\tilde{\boldsymbol{X}}}_i(s) \underline{\tilde{\boldsymbol{X}}}_i{}^T(s) - \underline{\boldsymbol{X}}_i(s) \underline{\boldsymbol{X}}_i{}^T(s) \right\|_F ds \\ \leq\ & \frac{1}{\bar{t}} \int_0^{\bar{t}} \frac{1}{n} \sum_{i=1}^n \left\| \underline{\tilde{\boldsymbol{X}}}_i(s) \underline{\tilde{\boldsymbol{X}}}_i{}^T(s) - \underline{\boldsymbol{X}}_i(s) \underline{\tilde{\boldsymbol{X}}}_i{}^T(s) \right\|_F + \left\| \underline{\boldsymbol{X}}_i(s) \underline{\tilde{\boldsymbol{X}}}_i{}^T(s) - \underline{\boldsymbol{X}}_i(s) \underline{\boldsymbol{X}}_i{}^T(s) \right\|_F ds \\ \leq\ & \frac{1}{\bar{t}} \int_0^{\bar{t}} \frac{1}{n} \sum_{i=1}^n \left\| \underline{\tilde{\boldsymbol{X}}}_i(s) - \underline{\boldsymbol{X}}_i(s) \right\|_F \left\| \underline{\tilde{\boldsymbol{X}}}_i{}^T(s) \right\|_F + \left\| \underline{\boldsymbol{X}}_i(s) \right\|_F \left\| \underline{\tilde{\boldsymbol{X}}}_i{}^T(s) - \underline{\boldsymbol{X}}_i{}^T(s) \right\|_F ds \\ \leq\ & \frac{1}{\bar{t}} \int_0^{\bar{t}} \frac{1}{n} \sum_{i=1}^n \sup_{\tau\in\mathcal{T}} \left\| \underline{\tilde{\boldsymbol{X}}}_i(\tau) - \underline{\boldsymbol{X}}_i(\tau) \right\|_F \left\| \underline{\tilde{\boldsymbol{X}}}_i{}^T(s) \right\|_F + \left\| \underline{\boldsymbol{X}}_i(s) \right\|_F \sup_{\tau\in\mathcal{T}} \left\| \underline{\tilde{\boldsymbol{X}}}_i{}^T(\tau) - \underline{\boldsymbol{X}}_i{}^T(\tau) \right\|_F d\tau \\ =\ & o_p\left(n^{\rho-1}\right),\end{aligned}$$

which justifies the result.

Proof of part c. Applying Lemma C.1 with

$$f(\boldsymbol{M}) := \boldsymbol{\Psi}(s)\boldsymbol{\Theta}(t)\boldsymbol{M}^{-1}\boldsymbol{\Theta}^T(\tau)\boldsymbol{\Psi}^T(\xi) \quad \text{for some matrix } \boldsymbol{M},$$

$$\boldsymbol{M}_1 := \frac{1}{\bar{t}} \int_0^{\bar{t}} \frac{1}{n} \sum_{i=1}^n \underline{\tilde{\boldsymbol{X}}}_i(t') \underline{\tilde{\boldsymbol{X}}}_i{}^T(t') dt' + \boldsymbol{\Lambda} \text{ and } \boldsymbol{M}_2 := \frac{1}{\bar{t}} \int_0^{\bar{t}} \frac{1}{n} \sum_{i=1}^n \underline{\boldsymbol{X}}_i(\tau) \underline{\boldsymbol{X}}_i{}^T(\tau) d\tau,$$

and Lemma C.3.b, the result follows.

□

Proof of Lemma C.4

The results in Lemma C.4 follows directly by applying Assumption 3.3.7 and the same proof as for Lemma C.2.

□

Proof of Lemma C.5

The results in Lemma C.5 follows directly by applying Assumption 3.3.7 and the same proof as for Lemma C.3.

□

Proof of Lemma C.6

Recall that for the estimated basis coefficients $[\tilde{c}_{k,i,1}, ..., \tilde{c}_{k,i,L}]^T$ of $\tilde{X}_i$'s, we have

$$\begin{aligned}\begin{bmatrix}\tilde{c}_{k,i,1}\\ \vdots \\ \tilde{c}_{k,i,L}\end{bmatrix} &:= \underset{(r_{k,i,1},...,r_{k,i,L})}{\operatorname{argmin}} \frac{1}{J_X}\sum_{j=1}^{J_X}\left[X_{kit_j} - \sum_{l=1}^{L} r_{k,i,l}\phi_l(t_j)\right]^2 \\ &= \left[\int_0^{\bar{t}} \boldsymbol{\phi}(\tau)\boldsymbol{\phi}^T(\tau)d\tau\right]^{-1}\left[\frac{1}{J_X}\sum_{j=1}^{J_X}\boldsymbol{\phi}(t_j)X_{kit_j}\right].\end{aligned}$$

$\int_0^{\bar{t}} \boldsymbol{\phi}(\tau)\boldsymbol{\phi}^T(\tau)d\tau$ and thus $\left[\int_0^{\bar{t}} \boldsymbol{\phi}(\tau)\boldsymbol{\phi}^T(\tau)d\tau\right]^{-1}$ are block diagonal matrices. $\frac{1}{J_X}\sum_{j=1}^{J_X}\boldsymbol{\phi}(t_j)X_{kit_j}$ aggregates the values in X_{kit_j} on to the local non-zero area of each basis coefficient ϕ_l. Hence, the vector $[\tilde{c}_{k,i,1}, ..., \tilde{c}_{k,i,L}]^T$ can be viewed as a discretization of the process X_{kit_j} over time, and therefore, the stationarity and ergodicity conditions also hold in $[\tilde{c}_{k,i,1}, ..., \tilde{c}_{k,i,L}]^T$. Following the bootstrap steps and proofs of Lemma C.3, the results hold.

□