Instrumental Factor Models for High-Dimensional Functional Data*

Jihyun Kim[†] and Young Kim[‡]

May 20, 2023

Abstract

This paper introduces the instrumental factor model, which extends conventional factor models in two ways. First, we develop a factor model for high-dimensional data, moving from scalar-valued data to functional data, which has gained fast-growing popularity. Second, while the standard estimation approach using Principal Component Analysis (PCA) requires both a large cross-sectional dimension and a long time horizon of data, our proposed method, which incorporates additional characteristic variables as instruments, ensures estimator consistency as long as the cross-sectional dimension is sufficiently large. We then introduce the eigenvalue ratio method to consistently estimate the unknown number of factors. Our numerical experiments suggest that our estimation approach outperforms the conventional PCA-based method, especially for short panel data. We conclude by conducting an empirical study to examine the long-term relationship between climate change and the European cereal market.

*We thank Nour Meddahi, Eric Gautier, Thierry Magnac and participants at Toulouse School of Economics, Sungkyunkwan University, and Korean Econometric Society Conference for helpful discussions and comments.

[†]Sungkyunkwan University and Toulouse School of Economics

[‡]Toulouse School of Economics

1 Introduction

In recent years, factor models have become a popular framework to analyze high-dimensional data, which are commonly available in economics, finance, and many other areas due to rapid advancements in data technology. In such contexts, high-dimensional data is typically characterized as a large collection of N time series observed over T period of time. The mechanism of factor models is built upon the idea that a dataset can be decomposed into two components; a common component that can be explained by a few number of factors, and an idiosyncratic component that is weakly correlated. The pioneering work of *approximated factor model* by Chamberlain (1983), and Chamberlain and Rothschild (1983) greatly broadens the scope of factor models, especially when the factors are unobserved, and it has been extensively studied in economics and finance. Early empirical studies of latent factor model include Stock and Watson (2002a), Stock and Watson (2002b), Bernanke et al. (2005), and Bai and Ng (2010) focusing on dimensionality reduction of high-dimensional data. Further theoretical developments are done by Bai and Ng (2002), Bai (2003), Onatski (2010), and Ahn and Horenstein (2013), providing asymptotic properties of model estimators. The factor model is also applied in the realm of correlated panel data studies by Pesaran (2006), Bai (2009), Pesaran and Tosetti (2011), and Beyhum and Gautier (2021), known as the interactive fixed-effect model. Xiong and Pelger (2019), and Bai and Ng (2021) use the factor model for the matrix completion and causal inference.

The principal component analysis (hereafter, PCA) is the most commonly used method to estimate the factors, and the factor loadings by recovering the first few principal components that capture the variation in the data. A major limitation of the PCA method is that the existing asymptotic properties are built upon not only the large cross-sectional dimension (N) but also the large time dimension (T). However, such a requirement may be infeasible in many applications where only a short panel data is available. Even if a long panel is available, there may be structural instability over a long time span, and thus the standard factor framework may not be directly applicable. Hence, this is an important constraint we wish to handle in our methodology.

The other aspect of our interest is functional data analysis (hereafter, FDA). Classical functional data consist of a random sample of square-integrable functions $y(r)$ on an interval $[a, b]$. The well-known monographs of FDA include Bosq (2000), Ramsay and Silverman (2005), Horváth and Kokoszka (2012), Hsing and Eubank (2015), and Kokoszka and Reimherr (2017), providing theoretical overview and FDA applications in many fields. With various regularization approaches, FDA is also actively used in regression analysis. See, for example, Ramsay and Dalzell (1991), Hastie and Mallows (1993), Hall and Horowitz (2007), and Benatia et al. (2017). Combining the elements of high-dimensionality and FDA, Tavakoli et al. (2021) introduced a factor model for high-dimensional functional time series data, allowing each observation is defined in a general

Hilbert space. Specifically, they consider datasets that consist of a large N collection of function-valued time series over a T period of time. Their model and estimation strategy can be viewed as the functional counterparts of the conventional factor models for scalar-valued data by Stock and Watson (2002a), Stock and Watson (2002b), Bai and Ng (2002). In the analogy of Bai and Ng (2002), Tavakoli et al. (2021) demonstrate their PCA estimators are consistent when $\min\{N, T\} \to \infty$, and thus, the very same limitation we have discussed still exists for functional data.

To overcome the earlier mentioned limitation, we propose the instrumental factor model for high-dimensional functional data and provide the identification and estimation methodology of our model. Under some regularity conditions analogous to Bai (2003), we show our estimators are consistent as long as $N \to \infty$ regardless of T being fixed or diverging. The existing asymptotic theory of factor analysis cannot be applied directly as the random elements are now extended from the real number to functional spaces. However, we illustrate that the underline intuition and mechanism remain to be similar to scalar data. We then provide the eigenvalue-ratio estimator, analogous to Ahn and Horenstein (2013), for the number of factors which is unknown in practice.

Our estimation strategy is motivated by the Projected-PCA (hereafter, PPCA) approach of Fan et al. (2016) that achieves the consistency of the factors and the loadings estimators as long as N is large, and labels our method as Functional-PPCA (hereafter, FPPCA). The important feature of the PPCA is, in addition to the main panel, there exist time-invariant characteristics relevant to the main panel through the factor loadings, the framework of Connor and Linton (2007) and Connor et al. (2012) that model the loading as a function of characteristics. Assuming the idiosyncratic noise is independent of the characteristics, it can be removed by projecting the data matrix onto a space spanned by the characteristics even when the time dimension T is finite. The PPCA then recovers the loadings and factors by applying the PCA to the projected data. This allows us to consistently estimate the model when the cross-sectional size is large while the conventional method requires the size of both dimensions to be large. This so-called 'characteristics-based' factor model has been widely recognized especially in Finance literature, see for instance Fan et al. (2017), Kokoszka and Reimherr (2017), Lettau and Pelger (2020), Kelly et al. (2020), Fan et al. (2021), and Kim et al. (2021). In particular, Kim et al. (2021) extended the PPCA method by introducing a non-zero mispricing term that is a function of covariates. Their estimators neither require large T as in the PPCA method.[1] Similarly, Kelly et al. (2020) proposed Instrumented-PCA for dynamic factor models, but their asymptotic properties require both N and T to be large, which is unsuited for our study.

Our finite-sample experiments confirm that the factors and loadings are accurately recovered by the FPPCA method as long as N is sufficiently large. In particular, our estimators

[1] The model by Fan et al. (2015) is a special case of zero mispricing term.

significantly outperform the PCA method by Tavakoli et al. (2021) for short panel data even if not all characteristic variables are observed. We then apply our methodology to the economic analysis of climate change. Using a continually collected globally gridded air temperature dataset, we conducted a factor-augmented VAR approach to analyze the long-run relationship between the global temperature and world GDP.[2]

The remainder of the paper is organized as the following. Section 2 introduces the instrumental factor model. Section 2.2 describes the estimators for the factor and the loadings. In Section 3, we provide the asymptotic properties of the proposed estimators and address the identification of the number of factors. Section 4 assesses the finite-sample properties of the proposed estimators using simulated data, and the empirical application is presented in Section 5. Section 6 concludes the paper. All technical derivations are provided in Appendices.

Notation and Preliminaries: Throughout the paper, we use the following notations. A panel of our interest is denoted by

$$\mathcal{Y}_{N,T} = \{y_{it} : \; i = 1, \ldots, N, \; t = 1, \ldots, T\}, \tag{1}$$

where $\{y_{it} \; : \; t \geq 1\}$ takes values in a real separable Hilbert space $\mathcal{H}_i$ with the inner product $\langle \cdot , \cdot \rangle_{\mathcal{H}_i}$ and the norm $\| \cdot \|_{\mathcal{H}_i} = \langle \cdot , \cdot \rangle_{\mathcal{H}_i}^{1/2}$. Here, N is the number of cross-sections, and T is the time horizon of data. We call $\mathcal{Y}_{N,T}$ high-dimensional functional data when N is large enough.

Let $\mathcal{S}_N = \mathcal{H}_1 \oplus \mathcal{H}_2 \oplus \cdots \oplus \mathcal{H}_N$, where $\oplus$ denotes the orthogonal direct sum of spaces. Then any element $v \in \mathcal{S}_N$ can be expressed as $v = (v_1, \ldots, v_N)'$, where $v_i \in \mathcal{H}_i$ for $i = 1, \ldots, N$. The inner product of the space $\mathcal{S}_N$ is defined by

$$\langle v, w \rangle_{\mathcal{S}_N} = \sum_{i=1}^{N} \langle v_i, w_i \rangle_{\mathcal{H}_i}, \quad \forall v, w \in \mathcal{S}_N,$$

and the norm $\| \cdot \|_{\mathcal{S}_N} = \langle \cdot , \cdot \rangle_{\mathcal{S}_N}^{1/2}$. The simplest example is when $\mathcal{H}_i = \mathbb{R}$ for all i, then $\mathcal{S}_N = \mathbb{R}^N$. Hence, $\langle v, w \rangle_{\mathcal{S}_N} = \sum_{i=1}^{N} \langle v_i, w_i \rangle_{\mathbb{R}} = \sum_{i=1}^{N} v_i \, w_i$, which is the usual inner product of N-dimensional real vectors. A case of our interest is $\mathcal{H}_i = L^2([a, b], \mathbb{R})$ for all i, the space of square-integrable functions from a closed interval $[a, b]$ to $\mathbb{R}$, equipped with the inner product $\langle x, y \rangle_{L^2} = \int_a^b x(r)y(r)dr$ for $x, y \in L^2([a, b], \mathbb{R})$. Without loss of generality, we set the interval to $[0, 1]$. Then we have

$$\langle v, w \rangle_{\mathcal{S}_N} = \sum_{i=1}^{N} \langle v_i, w_i \rangle_{L^2} = \sum_{i=1}^{N} \int_0^1 v_i(r) w_i(r) dr,$$

where the inner product in $\mathcal{S}_N$ involves integral for each v_i, w_i for $i \leq N$. In this paper, we assume $\mathcal{H}_i = \mathcal{H} = L^2([0, 1], \mathbb{R})$ for all $i \leq N$, hence $\mathcal{S}_N$ is the N direct sum of square

[2]The air temperature dataset is collected from the National Centers for Environmental Prediction (NCEP) and the National Center for Atmospheric Research (NCAR).

integrable functions.

Let $\mathcal{D}_1$ and $\mathcal{D}_2$ be any separable Hilbert spaces equipped with inner products $\langle \cdot\, , \cdot \rangle_{\mathcal{D}_1}$ and $\langle \cdot\, , \cdot \rangle_{\mathcal{D}_2}$, respectively, and define $\mathbb{L}(\mathcal{D}_1, \mathcal{D}_2)$ to be the space of bounded linear operators from $\mathcal{D}_1$ to $\mathcal{D}_2$. An operator $V \in \mathbb{L}(\mathcal{D}_1, \mathcal{D}_2)$ is said to be a Hilbert-Schmidt operator if for an (complete) orthonormal basis $\{e_j : j \geq 1\}$ of $\mathcal{D}_1$

$$\|V\|_{HS} := \left(\sum_{j\geq 1} \|Ve_j\|^2_{\mathcal{D}_2} \right)^{1/2} < \infty,$$

where $\|\cdot\|_{HS}$ is called the Hilbert-Schmidt norm. Then for any operator $V \in \mathbb{L}(\mathcal{D}_1, \mathcal{D}_2)$, the corresponding adjoint operator $V^* \in \mathbb{L}(\mathcal{D}_2, \mathcal{D}_1)$ satisfy $\langle Vx_1\, , x_2 \rangle_{\mathcal{D}_2} = \langle x_1\, , V^* x_2 \rangle_{\mathcal{D}_1}$ for all $x_1 \in \mathcal{D}_1,\ x_2 \in \mathcal{D}_2$. For the brevity of notations, we write $\|\cdot\| = \|\cdot\|_{HS}$. In a special case $\mathcal{D}_1 = \mathbb{R}^N$ and $\mathcal{D}_2 = \mathbb{R}^T$, $\|\cdot\|$ is the Frobenius norm, and $V^* = V'$.

For an illustration, we present an operator example relevant to our factor model. Define an operator $\Lambda \in \mathbb{L}(\mathbb{R}^K, \mathcal{H}_N)$ such that for any $f \in \mathbb{R}^K$,

$$\Lambda f = \begin{pmatrix} \sum_{k=1}^K \lambda_{1k} f_k \\ \vdots \\ \sum_{k=1}^K \lambda_{Nk} f_k \end{pmatrix} \in \mathcal{H}_N,$$

where $\lambda_{ik} \in \mathcal{H} = L^2([0,1], \mathbb{R})$ for all $i \leq N$. Note that Λf is similar to the usual matrix multiplication of $N \times K$ matrix Λ on a $K \times 1$ vector f, except now λ_{ik} is a functional element. The matrix Λ will be the loading matrix and f will be the latent factors at a given time in our model. If we fix $r \in [0,1]$, Λf corresponds to the usual real-valued matrix multiplication of $\Lambda(r)f$. The adjoint operator Λ^* must satisfy $\langle \Lambda f, y \rangle_{\mathcal{H}_N} = \langle f, \Lambda^* y \rangle_{\mathbb{R}^K}$ for all $f \in \mathbb{R}^K,\ y \in \mathcal{H}_N$. Then we define the adjoint operator as

$$\Lambda^* y = \begin{pmatrix} \sum_{i=1}^N \int_0^1 \lambda_{i1}(r) y_i(r)\, dr \\ \vdots \\ \sum_{i=1}^N \int_0^1 \lambda_{iK}(r) y_i(r)\, dr \end{pmatrix} \in \mathbb{R}^K.$$

With some abuse of notations, $\Lambda^* y = \int_0^1 \Lambda(r)' y(r)\, dr$ where we treated $\Lambda(r), y(r)$ as a real-valued matrix and vector for a fixed $r \in [0,1]$. By the above definitions, we have

$$\langle \Lambda f, y \rangle_{\mathcal{H}_N} = \langle f, \Lambda^* y \rangle_{\mathbb{R}^K} = \sum_{k=1}^K \sum_{i=1}^N f_k \int_0^1 \lambda_{ik}(r) y_i(r)\, dr.$$

Let us introduce the operator of the inner product matrix. The operator $\Lambda^*\Lambda \in \mathbb{L}(\mathbb{R}^K, \mathbb{R}^K)$ can be written as $K \times K$ real-valued matrix:

$$\Lambda^*\Lambda = \begin{pmatrix} \sum_{i=1}^N \int_0^1 \lambda_{i1}(r)\lambda_{i1}(r)\, dr & \cdots & \sum_{i=1}^N \int_0^1 \lambda_{i1}(r)\lambda_{iK}(r)\, dr \\ \vdots & \ddots & \vdots \\ \sum_{i=1}^N \int_0^1 \lambda_{iK}(r)\lambda_{i1}(r)\, dr & \cdots & \sum_{i=1}^N \int_0^1 \lambda_{iK}(r)\lambda_{iK}(r)\, dr \end{pmatrix} \in \mathbb{R}^{K\times K},$$

where $\mathbb{R}^{K\times K}$ denotes $K \times K$ real-valued matrix. We can treat this matrix as the integral of $\Lambda(r)'\Lambda(r)$ over $r \in [0,1]$, that is, $\Lambda^*\Lambda = \int_0^1 \Lambda(r)'\Lambda(r)\,dr$. If we take a special case that $\lambda_{ik} \in \mathbb{R}$, the matrix $\Lambda^*\Lambda$ is the inner product matrix, $\Lambda'\Lambda$.

2 Instrumental Factor Model

2.1 Model

Our interest lies in estimating the instrumental factor model (hereafter, IFM)

$$y_{it}(r) = \sum_{k=1}^{K} \lambda_{ik}(r) f_{tk} + \varepsilon_{it}(r), \quad r \in [0,1], \quad i \leq N, \quad t \leq T, \tag{2}$$

where $y_{it}(r), \lambda_{ik}(r), \varepsilon_{it}(r) \in \mathcal{H}$ and $f_{tk} \in \mathbb{R}$. The scalar variables f_{tk} are called latent factors; $\lambda_{ik}(r)$ are the factor loadings; $\varepsilon_{it}(r)$ are called idiosyncratic errors. The model (2) follows a typical factor structure in the literature that is decomposed into two unobserved components; the first summation term is called the common component with K factors, and the error component is orthogonal and additively separable to the common component. In addition, we give a structure to the loading coefficient as follows:

$$\lambda_{ik}(r) = g_k(X_i, r) + \gamma_{ik}(r),$$

where $X_i = (X_{i1}, \ldots, X_{iH})' \in \mathbb{R}^H$ is a vector of time-invariant characteristics, and $\gamma_{ik}(r)$ is the remaining component unexplained by X_i. For example, if y_{it} is the air temperature in region i, X_i can be regional-specific characteristics such as longitude, latitude, and amplitude. Let $f_t = (f_{t1}, \ldots, f_{tK})'$ be a vector of K factors, and the error component $\varepsilon_t(r) = (\varepsilon_{1t}(r), \ldots, \varepsilon_{Nt}(r))'$. We assume $\{X_i, f_t\}_{i\leq N, t\leq T}$ is independent of $\{\varepsilon_t(r)\}_{t\leq T}$ for all $r \in [0,1]$, and $\{X_i\}_{i\leq N}$ is independent of $\{\gamma_{ik}(r)\}_{i\leq N}$ for all $r \in [0,1]$, $k \leq K$. Moreover, we assume $g_k(X_i, r)$ has an additively separable structure such that

$$g_k(X_i, r) = g_k(X_{i1}, \ldots, X_{iH}, r) = \sum_{h=1}^{H} g_{kh}(X_{ih}, r),$$

for all $i \leq N, k \leq K$. Then for each $h \leq H$, g_{kh} is approximated by the J sieve functions $\{\phi_1(x), \phi_2(x), \ldots, \phi_J(x)\}$, i.e.,

$$g_{kh}(X_{ih}, r) = \sum_{j=1}^{J} b_{j,kh}(r)\phi_j(X_{ih}) + R_{kh}(X_{ih}, r),$$

where R_{kh} is the sieve approximation error that tends to zero as $J \to \infty$. For an illustration, suppose $H = 2$ and $J = 2$. Then we can write

$$\begin{aligned} g_k(X_i, r) &= g_{k1}(X_{i1}, r) + g_{k2}(X_{i2}, r) \\ &= \sum_{j=1}^{2}\sum_{h=1}^{2} b_{j,kh}(r)\phi_j(X_{ih}) + \sum_{h=1}^{2} R_{kh}(X_{ih}, r). \end{aligned}$$

Thus, for each k, we have $J \times H = 4$ number of sieve coefficients.

The IFM can be viewed as a general framework that encompasses various forms of factor models for high-dimensional data in the literature.

Factor Model for Time Series: Consider the factor model for scalar data studied by Bai and Ng (2002), and Bai (2003) that

$$y_{it} = \sum_{k}^{K} \lambda_{ik} f_{tk} + \varepsilon_{it}. \tag{3}$$

The model (3) is a special case of IFM with $y_{it}, \lambda_{ik}, \varepsilon_{it} \in \mathbb{R}$, and the characteristics X has no explanation power, that is, $g_k(X) = 0$. Hence, the remaining component γ_{ik} itself becomes the loading.

Semiparametric Factor Model for Time Series: If we maintain the above assumption but $g(X) \neq 0$, the IFM represents the semiparametric factor model

$$y_{it} = \sum_{k=1}^{K} \{g_k(X_i) + \gamma_{ik}\} f_{tk} + \varepsilon_{it},$$

which is proposed by Fan et al. (2016). If we further assume $\gamma_{ik} = 0$, our model coincides with Connor and Linton (2007), and Connor et al. (2012).

Factor Model for Functional Time Series: The factor model representation for high-dimensional functional time series by Tavakoli et al. (2021) is

$$y_{it}(r) = \sum_{k}^{K} \lambda_{ik}(r) f_{tk} + \varepsilon_{it}(r), \tag{4}$$

where $y_{it}(r), \lambda_{ik}(r), \varepsilon_{it}(r) \in \mathcal{H}$. Similar to the conventional factor model for scalar data, our model reduces to (4) if we assume that $g_k(X, r) = 0$, hence $\gamma_{ik}(r)$ becomes the loading component.

In the following, we describe the matrix representation of the model. Hereafter, whenever possible, we omit r in functional elements for the brevity of notations. In matrix form, the model is expressed as

$$\begin{aligned} Y &= \Lambda F' + \varepsilon \\ &= [G(X) + \Gamma] F' + \varepsilon, \end{aligned} \tag{5}$$

where Y is the $N \times T$ matrix of y_{it}; Λ is the $N \times K$ matrix of λ_{ik}; F is the $T \times K$ matrix of f_{tk}; ε is the $N \times T$ matrix of ε_{it}; $G(X)$ is the $N \times K$ matrix of $g_k(X_i)$; X is the $N \times H$ matrix of X_{ih}; and Γ is the $N \times K$ matrix of γ_{ik}. Let us elaborate on $G(X)$ in terms of the basis functions. Define JH-dimensional vector $b_k' = (b_{1,k1}, \ldots, b_{J,k1}, \ldots, b_{1,kH}, \ldots, b_{J,kH})$, and $\phi(X_i)' = (\phi_1(X_{i1}), \ldots, \phi_J(X_{i1}), \ldots, \phi_1(X_{iH}), \ldots, \phi_J(X_{iH}))$. Then for each $k \leq K$,

$$g_k(X_i) = \phi(X_i)' b_k + \sum_{h=1}^{H} R_{kh}(X_{ih}).$$

Let $B = (b_1, \dots, b_K)$ be the $(JH) \times K$ matrix of $b_{j,kh}$; $\Phi(X) = (\phi(X_1), \cdots, \phi(X_N))'$ be the $N \times (JH)$ matrix of $\phi_j(X_{ih})$; and $R(X)$ be $N \times K$ matrix of $\sum_{h=1}^{H} R_{kh}(X_{ih})$. Then, we can write $G(X) = \Phi(X)B + R(X)$, and thus we have

$$Y = [\Phi(X)B + \Gamma]F' + R(X)F' + \varepsilon.$$

Lastly, we define P as the projection matrix onto the space spanned by basis functions of X, written as

$$P = \Phi(X) \left[\Phi(X)'\Phi(X)\right]^{-1} \Phi(X)' \in \mathbb{R}^{N \times N},$$

and denote $\widehat{Y} = PY$, the projected data onto the space spanned by X. Assuming the sieve error is negligible and the orthogonal condition of (X, ε, Γ),

$$\begin{aligned}\widehat{Y} &= P[\Phi(X)B + \Gamma]F' + PR(X)F' + P\varepsilon \\ &\approx \Phi(X)BF',\end{aligned}$$

as $N \to \infty$. Hence, $\widehat{Y}$ is approximately noiseless when N is large enough and T can be either fixed or divergent.

2.2 Estimation

Our estimation strategy is inspired by the PPCA method proposed by Fan et al. (2016), which applies the PCA method to the projected data $\widehat{Y}$ rather than the original data Y. In the conventional factor model, it is assumed that the main panel data of interest is the only observed information to econometricians. On the other hand, the key assumption of the PPCA method is that, in addition to the main panel, there exist characteristics variables that are relevant to the main panel through the loading components and independent of the idiosyncratic errors. By the independence condition, $\widehat{Y}$ is asymptotically noiseless as long as $N \to \infty$, whether T is fixed or not. Hence, the factors and the loadings can be consistently estimated by applying the PCA method on $\widehat{Y}$. The above asymptotic properties can also be heuristically explained in terms of the eigenvalues of the covariance matrix. Consider the conventional factor model

$$Y = \Lambda F' + \varepsilon,$$

where $(Y, \Lambda, \varepsilon)$ are matrices of real numbers, and we impose the normalization conditions

$$\frac{F'F}{T} = I_K, \qquad \frac{\Lambda'\Lambda}{N} = \text{diagonal matrix with non-zero elements},$$

where I_K denotes the $K \times K$ identity matrix. For simplicity, we assume that $P\Lambda = \Lambda$, that is, the observed characteristics fully explain the loading. Let $\psi_k(A)$ denote the k-th largest eigenvalue of a matrix A. In the case of PCA estimation, for $k \leq K$

$$\psi_k\left(\frac{1}{NT}Y'Y\right) = \underbrace{\psi_k\left(\frac{1}{NT}F\Lambda'\Lambda F'\right)}_{O_p(1)} + \underbrace{\psi_k\left(\frac{1}{NT}\varepsilon'\varepsilon\right)}_{O_p(1/\min\{N,T\})} + \underbrace{\psi_k\left(\frac{1}{NT}(F\Lambda'\varepsilon + \varepsilon'\Lambda F')\right)}_{O_p(1/\sqrt{NT})},$$

that is, the eigenvalues of the noise terms of the sample covariance matrix vanish when both the cross-sectional dimension (N) and the time horizon (T) are large. On the other hand, the sample covariance matrix of the projected data has

$$\psi_k\left(\frac{1}{NT}Y'PY\right) = \underbrace{\psi_k\left(\frac{1}{NT}F\Lambda'P\Lambda F'\right)}_{O_p(1)} + \underbrace{\psi_k\left(\frac{1}{NT}\varepsilon'P\varepsilon\right)}_{O_p(1/N)} + \underbrace{\psi_k\left(\frac{1}{NT}(F\Lambda'P\varepsilon + \varepsilon P'\Lambda F')\right)}_{O_p(1/(\sqrt{NT}))}.$$

Therefore, the eigenvalues of the noise terms vanish as long as $N \to \infty$, and thus F and Λ are consistently estimated without large T.

We now elaborate on our FPPCA estimation procedure of the instrumental factor model for high-dimensional functional data. Suppose we have a factor model for high-dimensional functional data

$$Y = [\Phi(X)B + \Gamma]F' + R(X)F' + \varepsilon,$$

and the number of factors K is assumed to be known for the moment, and we will treat the identification issue of unknown K in Section 3. For scalar data, the PCA/PPCA method is simply applying the eigendecomposition of the sample covariance matrix which can be seen as the inner product of the data matrix. Likewise, we first compute the inner product matrix of the projected functional data

$$\frac{\widehat{Y}^*\widehat{Y}}{NT} = \frac{1}{NT}\begin{pmatrix} \langle \hat{y}_1\,, \hat{y}_1\rangle & \dots & \langle \hat{y}_1\,, \hat{y}_T\rangle \\ \vdots & \ddots & \vdots \\ \langle \hat{y}_T\,, \hat{y}_1\rangle & \dots & \langle \hat{y}_T\,, \hat{y}_T\rangle \end{pmatrix} \in \mathbb{R}^{T\times T}, \tag{6}$$

where $\hat{y}_t = (\hat{y}_{1t}, \dots, \hat{y}_{Nt})'$ is the t-th column vector of $\widehat{Y}$. We name the matrix (6) as the **integrated autocovariance matrix** [3] since each element is integrated, $\langle \hat{y}_s\,, \hat{y}_t\rangle = \sum_{i=1}^N \int_0^1 \hat{y}_{is}(r)\hat{y}_{it}(r)dr$. The integrated covariance matrix can be decomposed as

$$\frac{\widehat{Y}^*\widehat{Y}}{NT} = UVU' \tag{7}$$

where $V \in \mathbb{R}^{K\times K}$ is a diagonal matrix where the diagonal elements are the K leading eigenvalues of (6) in descending order, and $U \in \mathbb{R}^{T\times K}$ is a matrix whose k-th column is the eigenvector corresponding to k-th largest eigenvalue. In analogy to the PPCA method by Fan et al. (2016), the estimator for the factors is $\widehat{F} = \sqrt{T}U$, and for $r \in [0,1]$, the estimator for $G(X,r)$ and $\Lambda(r)$ are

$$\widehat{G}(X,r) = \frac{1}{T}\widehat{Y}(r)\widehat{F}, \quad \widehat{\Lambda}(r) = \frac{1}{T}Y(r)\widehat{F}.$$

Given that $\Gamma(r) = \Lambda(r) - G(X,r)$, the estimator for the remaining component Γ is

$$\widehat{\Gamma}(r) = \widehat{\Lambda}(r) - \widehat{G}(X,r).$$

[3] Our definition is distinctive from the notion of the integrated covariance matrix in finance literature. Following our notations, suppose $T = 1$ and $y_{i1}(r)$ denotes stock i's price in a continuous time interval $r \in [a,b]$. Then, the integrated covariance matrix refers to $E\left[\int_a^b y_1(r)y_1'(r)dr\right] \in \mathbb{R}^{N\times N}$, and if $N = 1$, it is known as integrated volatility.

Lastly, the estimator for the sieve parameter matrix is

$$\widehat{B}(r) = \frac{1}{T}[\Phi(X)'\Phi(X)]^{-1}\Phi(X)'Y(r)\widehat{F},$$

where it comes from regressing $PY(r)\widehat{F}$ on $\Phi(X)$.

A covariance matrix for scalar data can be understood as a special case of (6) if we assume that $\hat{y}_t \in \mathbb{R}^N$, and our estimation procedure coincides with the PCA/PPCA method for scalar data. In addition, the estimation of the parameters is also possible by applying the eigendecomposition of $\widehat{Y}\widehat{Y}^*/(NT)$, and its eigenfunctions are now the estimators for the loading components. However, the computation would be much heavier than the eigendecomposition of the matrix (6) as the matrix $\widehat{Y}\widehat{Y}^*$ is a functional operator in $\mathbb{L}(\mathcal{H}_N, \mathcal{H}_N)$. This is an analogy to the scalar data case that when $N > T$, it is more costly to handle a larger matrix $YY' \in \mathbb{R}^{N\times N}$ than $Y'Y \in \mathbb{R}^{T\times T}$.

3 Asymptotic Theory

In this section, we present the asymptotic properties of the estimators introduced in Section 2.2. For now, we maintain the assumption that K is known, and the estimator of the number of factors and its asymptotic results are provided in Section 3.2.

3.1 Factors, Loadings, and Common Component

We first formally state the regularity conditions to demonstrate the consistency of the estimators described in Section 2.2.

Assumption 3.1 (Random elements).
For all $i \leq N, t \leq T, k \leq K$, y_{it}, λ_{ik} and ε_{it} belong to $L^2([0,1], \mathbb{R})$, and $f_{tk} \in \mathbb{R}$.

Assumption 3.1 is to incorporate functional elements to the latent factor model that allows us to build a framework for high-dimensional functional data.

Assumption 3.2 (Identification).

(i) Almost surely,

$$\frac{F'F}{T} = I_K, \qquad \frac{G(X)^*G(X)}{N} = D,$$

where $D \in \mathbb{R}^{K\times K}$ is a diagonal matrix with distinctive elements.

(ii) There exists two positive constants c_{min} and c_{max} such that with probability approaching one, as $N \to \infty$

$$c_{min} \leq \psi_{min}\left(\frac{G(X)^*G(X)}{N}\right) \leq \psi_{max}\left(\frac{G(X)^*G(X)}{N}\right) \leq c_{max}.$$

Condition *(i)* is a typical identification restriction that allows us to separately identify factors and loading. Condition *(ii)* is known as the pervasiveness that all K factors have a nontrivial contribution to the integrated covariance matrix.

Assumption 3.3 (Basis functions)**.**

(i) As $N \to \infty$, with probability approaching one

$$d_{min} \leq \psi_{min}\left(\frac{\Phi(X)'\Phi(X)}{N}\right) \leq \psi_{max}\left(\frac{\Phi(X)'\Phi(X)}{N}\right) \leq d_{max},$$

where d_{min} and d_{max} denote two positive constants.

(ii) $\max_{j\leq J, i\leq N, h\leq H} E[\phi_j(X_{ih})^2] < \infty.$

For simplicity, if we assume that $\Phi(X) = X$, Assumption 3.3 implies the covariance matrix $E[X_i X_i']$ is well-defined and $\{X_i\}_{i\leq N}$ can be weakly dependent as long as the law of large number holds that $N^{-1}\sum_{i=1}^N X_i X_i' \xrightarrow{p} E[X_i X_i']$.

Assumption 3.4 (Data generating process)**.**

(i) A mean zero functional process $\{\varepsilon_t\}_{t\leq T}$ is independent of $\{X_i, f_t\}_{i\leq N, t\leq T}$.

(ii) $\{f_t, \varepsilon_t\}_{t\leq T}$ is strictly t-stationary.[4]

(iii) Weak dependence: Let M_1 be a positive constant. Then

$$\max_{i\leq N}\sum_{q=1}^N \int_0^1 |E[\varepsilon_{it}(r_1)\varepsilon_{qt}(r_2)]| dr_1 dr_2 < M_1$$

$$\frac{1}{NT}\sum_{i=1}^N\sum_{q=1}^N\sum_{t=1}^T\sum_{s=1}^T \int_0^1 |E[\varepsilon_{it}(r_1)\varepsilon_{qs}(r_2)]| dr_1 dr_2 < M_1$$

$$\max_{i\leq N}\frac{1}{NT}\sum_{q=1}^N\sum_{m=1}^N\sum_{t=1}^T\sum_{s=1}^T \int_0^1 |cov[\varepsilon_{it}(r_1)\varepsilon_{qt}(r_2), \varepsilon_{is}(r_1)\varepsilon_{ms}(r_2)]| dr_1 dr_2 < M_1$$

(iv) Mixing condition: There exist A, $a_1 > 0$ such that for all $T > 0$,

$$\alpha(T) < \exp(-A\, T^{a_1}),$$

where the α-mixing coefficient is defined as

$$\alpha(T) = \sup_{\Theta_1\in\mathcal{F}_{-\infty}^0, \Theta_2\in\mathcal{F}_T^\infty} |P(\Theta_1)P(\Theta_2) - P(\Theta_1\cap\Theta_2)|$$

for σ-algebras $\mathcal{F}_{-\infty}^0$ and $\mathcal{F}_T^\infty$ generated by $\{f_t, \varepsilon_t\}_{t\leq 0}$ and $\{f_t, \varepsilon_t\}_{t\geq T}$, respectively.

(v) Light-tailed distribution: There are positive constants a_2, a_3 and c_1, c_2 such that $a_1^{-1} + a_2^{-1} + a_3^{-1} > 1$, and for any $\delta > 0$,

$$\sup_{r\in[0,1]} P(|\varepsilon_{it}(r)| > \delta) \leq \exp\{-(\delta/c_1)^{a_2}\}, \quad P(|f_{tk}| > \delta) \leq \exp\{-(\delta/c_2)^{a_3}\},$$

for all $i \leq N$, $t \leq T$, and $k \leq K$.

[4] If $\{\varepsilon_{it}(r)\}_{t\leq T}$ is a strictly t-stationary process, $E[\varepsilon_{it}(r)]$ does not depend on t, and $E[\varepsilon_{it}(r_1)\varepsilon_{js}(r_2)]$ only depends on $|t-s|$ for all $i, j \leq N, r_1, r_2 \in [a, b]$.

Assumption 3.4 is analogous to the standard weak cross-sectional and time dependency conditions in the approximate factor model, see for instance Bai and Ng (2002), and Fan et al. (2016). Our assumption essentially implies that the standard weak dependency conditions are satisfied for each $r \in [0,1]$.

Assumption 3.5 (Loading components)**.**

(i) $\{\gamma_{ik}\}_{i\le N, k\le K}$ *is independent of* $\{X_i\}_{i\le N}$*, and* $E[\gamma_{ik}(r)] = 0$ *for all* $r \in [0,1]$*.*

(ii) Let $\rho_N = \sup_{r\in[0,1],\, k\le K} \frac{1}{N}\sum_{i=N}^{N} E[\gamma_{ik}^2(r)]$*, and* $\rho_N < \infty$*. Then we have*

$$\sup_{r_1,r_2\in[0,1], i\le N, k\le K} \sum_{q=1}^{N} |E[\gamma_{ik}(r_1)\gamma_{qk}(r_2)]| = O(\rho_N).$$

(iii) $\sup_{r\in[0,1], i\le N, k\le K} E[g_k^2(X_i, r)] < \infty.$

Here, we assume that the unexplained component is independent of the characteristic variables and allow weak cross-sectional dependency.

Assumption 3.6 (Sieve approximation)**.**

(i) For all $h \le H$, $k \le K$*, the loading component* $g_{kh}(\cdot)$ *belongs to a Hölder space* $\mathcal{G}(\omega, \beta, L)$ *defined as*

$$\mathcal{G}(\omega,\beta,L) = \{g : |D^\omega g(v_1) - D^\omega g(v_2)| \le L\|v_1 - v_2\|^\beta\}$$

for some $L > 0$, $v_1, v_2 \in \mathbb{R} \times [0,1]$*.*

(ii) Suppose $\kappa = (\omega + \beta) \ge 2$*. As* $J \to \infty$*, the sieve coefficients* $\{b_{k,jh}\}_{j\le J}$ *satisfy*

$$\sup_{r\in[0,1],\, x\in\mathcal{X}_h} |g_{kh}(x,r) - \sum_{j=1}^{J} b_{k,jh}(r)\phi_j(x)|^2 = O(J^{-\kappa}), \quad \text{for all } h \le H,\, k \le K,$$

where $\mathcal{X}_h$ *denotes the support of* X_{ih}*.*

(iii) $\sup_{r\in[0,1],\, k\le K, j\le J, h\le H} b_{k,jh}^2(r) < \infty.$

Assumption 3.6 is a technical condition to control the smoothness of the function $g_{kh}(,)$ to ensure the approximation error decays at the rate of $O(J^{-\kappa})$ for $\kappa \ge 2$.

Theorem 3.1.
Suppose $J = o(\sqrt{N})$*, and Assumptions 3.1-3.6 are satisfied. As* $N, J \to \infty$ *(*T *may stay constant or simultaneously grow with* N *and* J*),*

$$\frac{1}{T}\|\widehat{F} - F\|^2 = O_p\Big(\frac{1}{N} + \frac{1}{J^\kappa}\Big),$$
$$\frac{1}{N}\|\widehat{G}(X) - G(X)\|^2 = O_p\Big(\frac{J}{N^2} + \frac{J}{NT} + \frac{J}{J^\kappa} + \frac{J\rho_N}{N}\Big),$$
$$\|\widehat{B} - B\|^2 = O_p\Big(\frac{J}{N^2} + \frac{J}{NT} + \frac{J}{J^\kappa} + \frac{J\rho_N}{N}\Big).$$

Theorem 3.1 implies the consistency of the estimators without a condition $T \to \infty$, which is an important feature of the FPPCA method in the case of a short-time horizon. In practice, there is no guarantee that a time series model is stable over a long period of time, hence we may need to restrict ourselves to a short period even if large sample data is available. Therefore, our FPPCA method would be more appealing than the PCA method by Tavakoli et al. (2021) in such context.

Corollary 3.1.
Under the assumptions of Theorem 3.1, as $T \to \infty$ simultaneously with N and J,

$$\frac{1}{N}\|\widehat{\Gamma} - \Gamma\|^2 = O_p\Big(\frac{J}{N^2} + \frac{1}{T} + \frac{1}{J^\kappa} + \frac{J\rho_N}{N}\Big). \tag{8}$$

Given that $\widehat{\Gamma}(r) = \widehat{\Lambda}(r) - \widehat{G}(X, r)$, and $\widehat{\Lambda}(r)$ comes from the regression of the original data Y on $\widehat{F}$, the consistency requires $T \to \infty$.

3.2 Number of Factors

We have so far assumed that the number of factors is known, however, such an assumption is unrealistic in practice. For scalar data, Bai and Ng (2002), Hallin and Liška (2007), Bai and Ng (2007), and Alessi et al. (2010) proposed an information criteria-based method to estimate the number of factors. Other popular approaches based on eigenvalues criteria include Onatski (2010), Lam and Yao (2012), Ahn and Horenstein (2013), and Fan et al. (2016). In the functional context, Tavakoli et al. (2021) proposed a consistent estimator based on an information criterion extending the idea of Bai and Ng (2002). An attractive feature of the eigenvalue ratio estimator is that a penalty function and the maximum number of factors do not need to be pre-specified. Moreover, it is well-known that the finite-sample accuracy of the information criteria approach could heavily depend on the choice of a penalty function and the maximum number. Hence, we adopt the eigenvalue ratio method by Ahn and Horenstein (2013) and Fan et al. (2016) in the context of projected functional data. The criterion employed is the ratio of two adjacent eigenvalues of $\widehat{Y}^*\widehat{Y}$ defined as

$$ER(\ell) = \frac{\psi_\ell\left(\widehat{Y}^*\widehat{Y}\right)}{\psi_{\ell+1}\left(\widehat{Y}^*\widehat{Y}\right)}, \quad \ell = 1, \ldots, \ell_{max}.$$

Here, we heuristically describe the underline mechanism of the eigenvalue ratio criterion to identify the number of factors. As we have discussed in Section 2.2, $\psi_\ell(\cdot)$ is persistent for $\ell \leq K$ uniformly over N and T. For $\ell > K$, $\psi_\ell(\cdot)$ is solely determined by the noise terms that vanish asymptotically. Therefore, $ER(\ell)$ is bounded for $\ell \neq K$ and diverges when ℓ is correctly chosen as K. Following the above intuition, we now define the number of factors estimator as

$$\widehat{K} = \underset{1 \leq \ell < \ell_{max}}{\operatorname{argmax}} \frac{\psi_\ell\left(\widehat{Y}^*\widehat{Y}\right)}{\psi_{\ell+1}\left(\widehat{Y}^*\widehat{Y}\right)}.$$

Assumption 3.7 (Number of factors).
The error matrix $\varepsilon(r)$ *can be decomposed as*

$$\varepsilon(r) = A_N^{1/2} U(r) Z_T^{1/2},$$

(i) $A_N \in \mathbb{R}^{N\times N}$ *and* $Z_T \in \mathbb{R}^{T\times T}$ *are non-stochastic positive definite matrices where eigenvalues are bounded away from zero and infinity.*

(ii) $U(r)$ *is the* $N \times T$ *matrix of* $u_{it}(r)$*, where the functional process* $\{u_{it}\}_{i\le N, t\le T}$ *is mean-zero and independent over* i *and* t*. In addition,* $u_t = (u_{1t}, \dots, u_{Nt})'$ *is iid sub-Gaussian, that is, there exists* $M_2 > 0$ *such that*

$$E\left[\exp\{\tau\langle u_t, v\rangle\}\right] \le \exp\{\tau^2 M_2 \|v\|^2\},$$

for all $\tau > 0$*,* $v \in \mathcal{H}_N$*.*

(iii) Almost surely,

$$d_{min} \le \psi_{min}\left(\frac{\Phi(X)'\Phi(X)}{N}\right) \le \psi_{max}\left(\frac{\Phi(X)'\Phi(X)}{N}\right) \le d_{max},$$

where d_{min} *and* d_{max} *denote two positive constants.*

Condition *(i)* is analogous to the standard assumption proposed by Ahn and Horenstein (2013). A_N captures the cross-sectional correlation of ε, while R_T controls the structure of time dependency. In a special case that A_N and R_T are diagonal matrices, then ε_{it} is independent over i and t. The sub-Gaussianity condition in *(ii)* is to apply the asymptotic theory of high-dimensional random matrix by Vershynin (2010) and Vershynin (2018).

Theorem 3.2.
Suppose Assumptions 3.1-3.7 holds, and $1 \le K < JH/2$*. Then as* $N, T \to \infty$*, and* $J = o(\min\{N, T\})$*, we have*

$$P(\widehat{K} = K) \to 1.$$

Theorem 3.2 shows that the estimator $\widehat{K}$ is consistent when both $N, T \to \infty$. Large T is a necessary condition to ensure that the eigenvalues of $U(r)$ are bounded away from zero and infinity, hence denominator of $ER(\ell)$ is well-defined for $\ell \ge K$. If we impose an alternative condition that the eigenvalues of $U(r)$ are bounded away from zero and infinity almost surely, Theorem 3.2 still holds without $T \to \infty$.

4 Numerical Experiments

This section reports numerical experiment results to assess the finite-sample performance of the FPPCA estimators for factors, loadings, and common components. The purpose of the numerical experiment is to compare the finite-sample performance of our estimators in Section 2.2 to the PCA proposed by Tavakoli et al. (2021) in various settings. We first introduce the data generating process for the panel $\{y_{it} : i \leq N, t \leq T\}$. Consider a model with K factors

$$y_{it}(r) = \sum_{k=1}^{K} g_k(X_i, r) f_{tk} + \varepsilon_{it}(r),$$

where all functional elements are generated using the Fourier series $\{\phi_1(r), \ldots, \phi_L(r)\}$, and our results are invariant with the choice of basis functions. The factor loadings are generated as

$$g_k(X_i, r) = \sum_{h=1}^{H} X_{ih} b_{kh}(r) = \sum_{h=1}^{H} X_{ih} \beta_{kh} \phi_k(r), \quad k \leq K,$$

where $X_{ih} \sim N(0, 1)$. The coefficient β_{kh} is initially drawn from $U(0, 1)$ and treated as a deterministic parameter for all data generations. In this design, each k-th loading component is generated by using one basis function $\phi_k(r)$, and thus we ensure to set $L \geq K$. The factors f_{tk} are drawn from the standard normal distribution. The idiosyncratic errors are generated by $\{\phi_1(r), \ldots, \phi_L(r)\}$ that

$$\varepsilon_{it}(r) = \sum_{\ell=1}^{L} a_{\ell,it} \phi_\ell(r),$$

where $a_{\ell,it} \sim N(0, 1)$. By using the above model, we simulate the data with two different sets of panel sizes; the fixed N samples, and the fixed T samples. The fixed N samples allow us to demonstrate the relative performance of the FPPCA to the PCA, especially when the T is small. The fixed N samples visually illustrate our theoretical result that the estimators are consistent as long as N is large. It is well-known that the factors and the loadings are not separately identified; for any invertible $H \in \mathbb{R}^{K \times K}$, we have $\Lambda_0 F_0' = \Lambda_0 (H^{-1})' H' F_0' = \Lambda F'$, where $F = F_0 H$ and $\Lambda = \Lambda_0 (H^{-1})'$. Hence, we choose H such that the identification conditions, $\frac{1}{T} H F_0' F_0 H = I_K$ and $\frac{1}{N} H^{-1} \Lambda_0^* \Lambda_0 (H^{-1})' = D$ are satisfied. Then the proposed estimators target the rotated factors and loadings F and Λ.

4.1 Estimation of Factors, Loadings, and Common Component

We set $(K, H, L) = (2, 3, 5)$, and the fixed N samples and the fixed T samples are $\{N = 100 \ \& \ T = 10, 20, \ldots, 150\}$ and $\{T = 100 \ \& \ N = 10, 20, \ldots, 150\}$, respectively. For each choice of N and T, we estimate the factors, loadings and common components using the method proposed in Section 2.2 and the PCA estimators by Tavakoli et al. (2021). In analogy to the conventional PCA estimators for scalar data, Tavakoli et al. (2021) applies the eigendecomposition on the real $T \times T$ matrix $\frac{Y^* Y}{NT}$. Their method coincides with the conventional PCA method if $Y \in \mathbb{R}^{N \times T}$. The finite-sample performance is measured by $\frac{1}{\sqrt{T}} \|\widehat{F} - F\|$, $\frac{1}{\sqrt{N}} \|\widehat{\Lambda} - \Lambda\|$, and $\frac{1}{\sqrt{NT}} \|\widehat{C} - C\|$, where C and $\widehat{C}$ denotes the common component matrix and its estimator, respectively. The three values are computed as the average over 1000 replications for each size of N and T. The simulation results are depicted in Figure 1, the left column shows the fixed N sample results and the right column shows the fixed T samples.

Figure 1 demonstrates that the FPPCA estimators outperform the PCA estimators in all cases. For fixed N samples in the left column, the performance gap is much wider when T is small which is in line with our theory. As N is fixed, the improvements on the FPPCA are marginal as T increases, hence the performance gap becomes smaller. Nevertheless, our estimators still achieve lower errors than the PCA estimators for larger T. In the case of fixed T samples, our estimators start to perform even better than the PCA estimators when N grows larger. This matches with the theory that the convergence rate of the PCA estimator is $O_p(1/\min\{\sqrt{N}, \sqrt{T}\})$.

4.2 Estimation with Weak Instrument

In this simulation, we modify the loading component that

$$y_{it}(r) = \left[X_{i1}\beta_1\phi_1(r) + X_{i2}\frac{\beta_2\phi_2(r)}{D} \right] f_t + \varepsilon_{it}(r), \tag{9}$$

where $D \to \infty$ when $\max\{N, T\} \to \infty$. The model (9) is similar to the weak instrument model studied by Staiger and Stock (1994) that the parameter for X_2 is modeled as local-to-zero. Hence, the strength of X_2 on the loading component decays as the panel size increases. The objective of this experiment is to examine the performance our estimators when we partially observe the characteristics that are either strong or weak. Even if not all characteristics are observed, the experiment result shows that our estimator for factors to work well as long as the characteristics are strong.

We fix $(K, H, L) = (1, 2, 3)$, and the fixed N samples and the fixed T samples are $\{N = 100,\ T = 20, 40, \ldots, 140,\ D = T\}$ and $\{T = 100,\ N = 20, 40, \ldots, 140,\ D = N\}$. The estimation errors are averaged over the 500 replications for each N and T. X_1 and X_2 follow the standard normal distribution with $cov(X_1, X_2) = 0.5$. The results are presented in three different cases; *Case 1*: (X_2, X_2) are observed, *Case 2*: (X_1) is observed, and *Case 3*: (X_2) is observed. We choose $\beta_1, \beta_2 \in \{0, 0.5, 1.0\}$, hence there are nine possible values of (β_1, β_2). The size of βs may be important when the characteristics are partially observed. For instance, in *Case 2*, the best scenario is when $(\beta_1, \beta_2) = (1, 0)$, and the worst scenario is when $(\beta_1, \beta_2) = (0, 1)$. If $\beta_1 = 0$, we expect all estimators for factors to fail in all cases. Followed by the eigenvalue argument in Section 2.2, not only the eigenvalues of the error terms but also the estimation of the eigenvalues of the common component is not persistent, hence the estimation is not possible.

Figure 2 and 3 report the estimation results of *Case 1* for the fixed N samples and the fixed T samples, respectively. The results are similar to the ones in Section 4.1 that the FPPCA estimators clearly outperform the PCA when $\beta_1 > 0$. As we anticipated, both the FPPCA and the PCA do not work well if $\beta_1 = 0$. The estimation results for *Case 2* are presented in Figure 4 and 5. As long as $\beta_1 > 0$, our estimators still outperform the PCA, and thus the results demonstrate that our method works well even if we partially observe characteristics, as long as they are strong characteristics.

In contrast to the previous results, our estimator does not work well when only the weak characteristic X_2 is observed. In Figure 6 and 7, the factor is better estimated by the PCA method in all cases. This confirms our asymptotic theory and intuition that as X_2 loses its relevant power on the dependent variable y, the projection not only eliminates the idiosyncratic noise but also the common component, hence our estimator fails to perform well.

5 European Cereal Markets and Temperature

Recent climate changes in Europe are characterized by record-breaking high temperatures and severe weather events, which pose a great risk to crop yields and are expected to intensify in the near future [Campbell (2022), and Pörtner et al. (2022)]. Identifying the climate threat to agricultural production is essential not only to sustain the current food security but in the coming decades. To examine the effect of climate change on food security, we quantify the effect of temperature rise on the European cereal markets; barley, maize, and wheat which are the three major cereal crops compromise up to 70 percent of the world's production of cereals in a year. Cereal grains have been the prime component of the human diet, accounting for more than half of daily caloric intake and 47 percent of daily protein intake worldwide [Awika (2011)]. The importance of cereals and cereal products is also supported by the annual figure that global food security depends largely on cereal production, which yearly amounts to approximately 2,700 million tons. Production shares of cereal production by continent are summarized in Figure 8. Overall, Europe is an influential player in cereals production. For barley, it is the largest producer in the world with a 65 percent share of the production per year. Moreover, Europe accounts for 37 percent of the global wheat production which is only one percentage point below the largest producer, Asia, and 13 percent of maize is cultivated in Europe. Another crucial cereal, rice, is not considered in this analysis since Europe only accounts for less than 1 percent of global rice production.

To quantify the effect of temperature rise on the barley, maize, and wheat markets, we employ the VAR approach to estimate the impulse response of short-run temperature shock on the annual growth rate of cereal prices and productions for the three cereal markets. It is well-known that simply inspecting the global average temperature does not fully capture the temperature trend; there is significant cross-sectional heterogeneity in the trends [IPCC (2014), and Rivas and Gonzalo (2020)]. Therefore, we consider the NCEP/NCAR[5] Reanalysis dataset that recorded daily air temperature across globally gridded 2629 stations from 1948 to 2020. Given a large number of stations, the standard VAR model falls into the curse of dimensionality problem, and we thus use the factor-

[5]The National Centers for Environmental Prediction, and the National Center for Atmospheric Research.

augmented VAR approach in the spirit of Bernanke et al. (2005) by applying the FPPCA method to recover factors from the NCEP/NCAR dataset.

Following the notations in Section 2, we denote $y_{it}(r)$ as the station i's air temperature (in Kelvin scale) of r-th day in a given month t from January 1948 to December 2020. We use three characteristics variables for each station i: latitude, longitude, and the Köppen–Geiger climate classification system. Latitude and longitude are the primary characteristics that affect the temperature by determining the amount and the angle of solar radiation to a location. The Köppen-Geiger system classifies the climate of a location into thirty different types based on the geographical and ecological elements that affect the air temperature.

In addition to temperature, precipitation has also been regarded as a vital weather factor for crop cultivation [Deschênes and Greenstone (2007), Auffhammer (2018), and Xie et al. (2019)]. Similar to temperature, regional heterogeneity in precipitation needs to be considered for the analysis. Hence, we acquire information on precipitation from NOAA[6] Precipitation Reconstruction over Land (PREC/L) dataset. It records a monthly precipitation rate (mm per day) over the global grid of 144×72 stations from 1948 to 2021. Using the three characteristics explained above, we employ the IFM method to recover factors from the dataset.

Figure 9 shows the IFM estimation result of the NCEP/NCAR Reanalysis and NOAA PREC/L data. For our analysis, stations in European territories are selected only; NCEP/NCAR Reanalysis contains 447 stations and NOAA PREC/L has 593 stations. The number of factors is chosen as one according to the eigenvalue ratio estimator proposed in Section 3.2. The first factor explains over 95% of the data in terms of the relative size of its eigenvalue for both temperate and precipitation datasets, and we find that all the empirical results in this section are robust to the choice of the number of factors. For ease of interpretation, we plot the annual values averaged over all stations. In comparison to the actual trends, our estimation method performs well in capturing the overall trajectories of temperature and precipitation in European territories.

To estimate the impulse response of short-run temperature shock for each cereal market, we employ a VAR model with four key variables; temperature, precipitation rate, cereal production growth rate, and cereal price growth rate. We collected annual figures on cereal production and cereal price from the FAO[7] statistical database and the World Bank Commodity Price dataset for the period of 1962 - 2020. Given the limitation of the average temperature and precipitation rate, we make use of the estimated factors to capture the climate information in Europe. The model of our interest is the following. For $t = 1962, \ldots, 2020$, define $Z_t = (F_t^{tem}, F_t^{pre}, Q_t, P_t)'$ where (F_t^{tem}, F_t^{pre}) indicates factors

[6]National Oceanic and Atmospheric Administration.

[7]Food and Agriculture Organization of the United Nations.

for temperature and precipitation rate in Europe respectively, and (Q_t, P_t) denotes the growth rates of cereal production and price. Then, the VAR(1) model[8] we consider is

$$Z_t = AZ_{t-1} + Du_t, \tag{10}$$

where we assume $E[u_t u_t'] = I$, and D is a lower triangular impact matrix, that is, we assume that a temperature shock has a contemporaneous effect on cereal markets.

Figure 10 outlines the response of the barley market to a temperature shock (by one standard deviation). The contemporaneous effect occurs in the opposite directions for price and production; the growth rate of price is expected to rise by 2.8 percentage points whereas the rate for production decreases by 2.6 percentage points. We provided weak evidence that the contemporaneous responses to the shock are statistically significant at 68 percent confidence level in Figure 11. Overall, the short-run temperature shock dissipates after period six while there is an oscillation in the growth rate of production. The temperature shock causes the price (resp. production) level to increase (resp. decrease) permanently by 3.0 percent (resp. 1.2). As evidenced by Figure 12 and 13, the effect of temperature shock on maize market analogous to barley market. Our results are in line with the existing evidence that the temperature rise negatively affects agricultural outputs [Dell et al. (2012), Chen et al. (2016), Burke and Emerick (2016), and Pörtner et al. (2022)].

6 Conclusion

This paper proposes a new factor model approach, labeled the instrumental factor model, to analyze high-dimensional data with function-valued observations. The approach follows the characteristics-based factor model by Connor and Linton (2007) and Connor et al. (2012) that assume the factor loading as a function of observed characteristic variables. Our model also provides a general framework that encompasses the conventional factor model for scalar data, yet the existing intuition and mechanism of the conventional model still apply.

We then introduce the FPPCA estimators for the factors and the loadings, motivated by Fan et al. (2016), that first, we project functional data onto the space spanned by the characteristic variables and then apply the PCA on the projected data to recover the factors and the loadings. Since idiosyncratic errors are removed after projecting data onto the characteristics space, our theory demonstrates that the FPPCA estimators for the factors and the loadings are asymptotically valid even if T is fixed, an appealing feature when only a short panel data is available. In addition, we introduce an eigenvalue ratio estimator for the number of factors that can be used for high-dimensional functional data. The finite-sample experiment supports our expectation that the estimators

[8] We found that the estimation results are invariant to the VAR models with larger lags.

are more attractive for short-panel data. We show the FPPCA estimators outperform the PCA estimators by Tavakoli et al. (2021), especially when T is small and/or not all characteristics are observed. However, the FPPCA does not perform well if we observe weak instruments only. We then apply the FPPCA method to quantify the relationship between global warming and the cereal markets by conducting a factor-augmented VAR approach. The results support the existing evidence that temperature rise leads to a decline in agricultural output.

We conclude by highlighting some key areas for future research. The first issue is to develop a statistical procedure to test whether the observed characteristic variables have explaining power on the loading, and if so, whether the characteristics fully/partially explain the loading. In addition, our asymptotic theory is limited to the consistency of the estimators only, and therefore, another important area for the future is to provide the asymptotic distribution of the estimators comparable to Bai (2003).

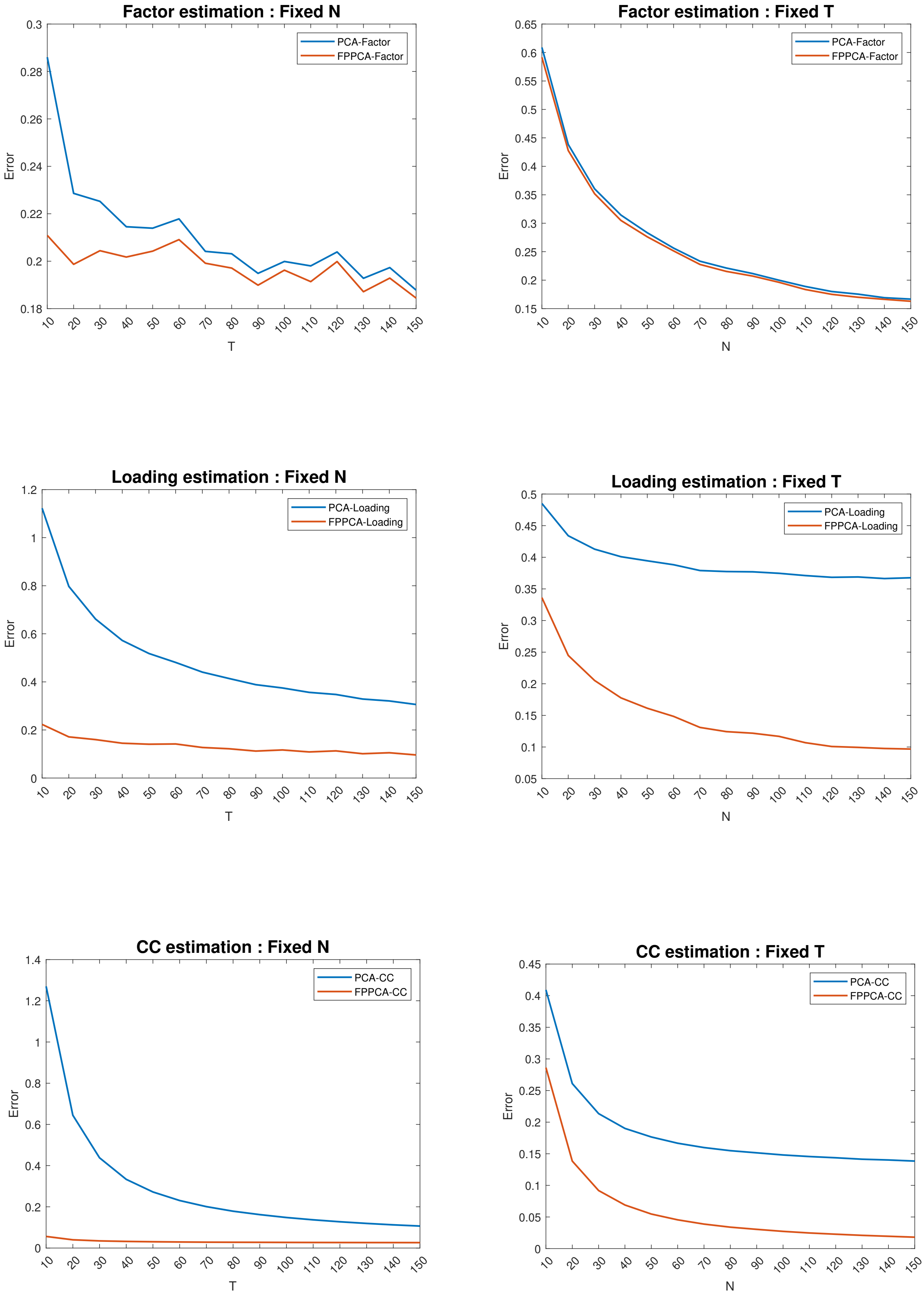

Figure 1: *Averaged errors of estimators for factors, loadings and common components in Section 4.1. Left column shows the fixed N samples errors and the right column is for the fixed T samples.*

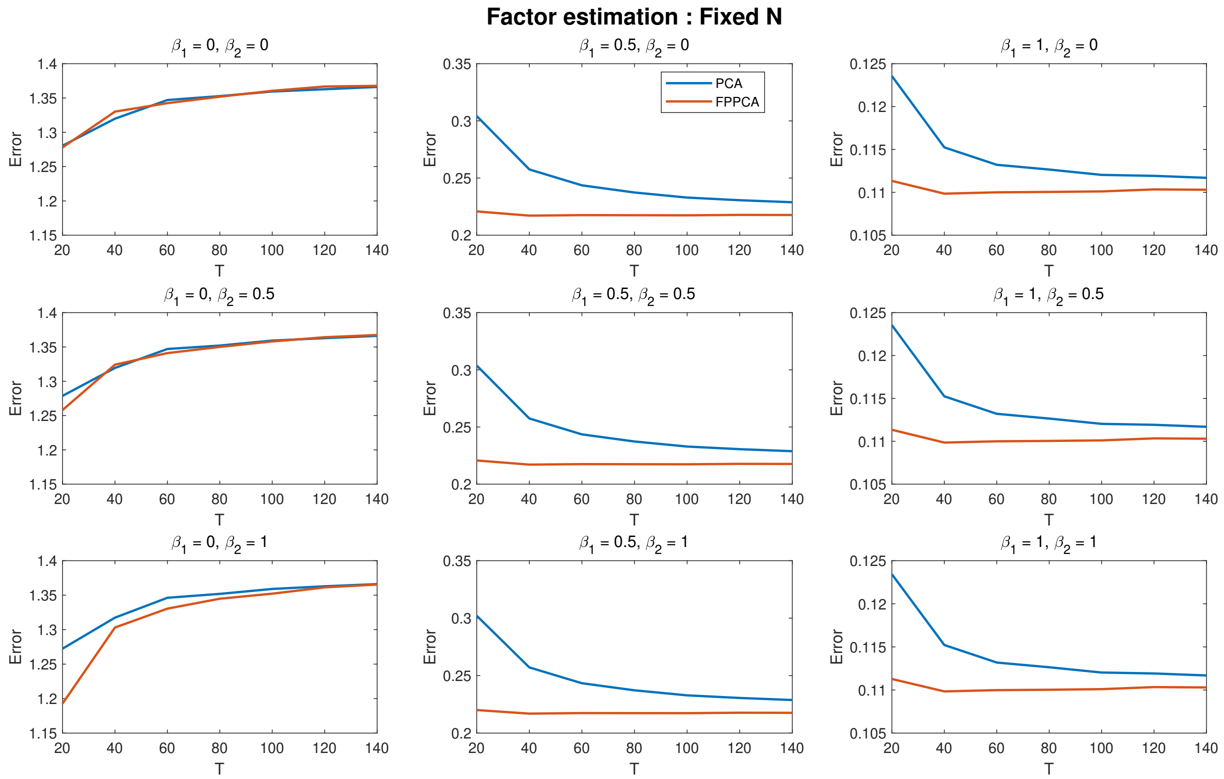

Figure 2: *Averaged errors of estimator for factors with fixed N samples, where both X_1, X_2 are observed.*

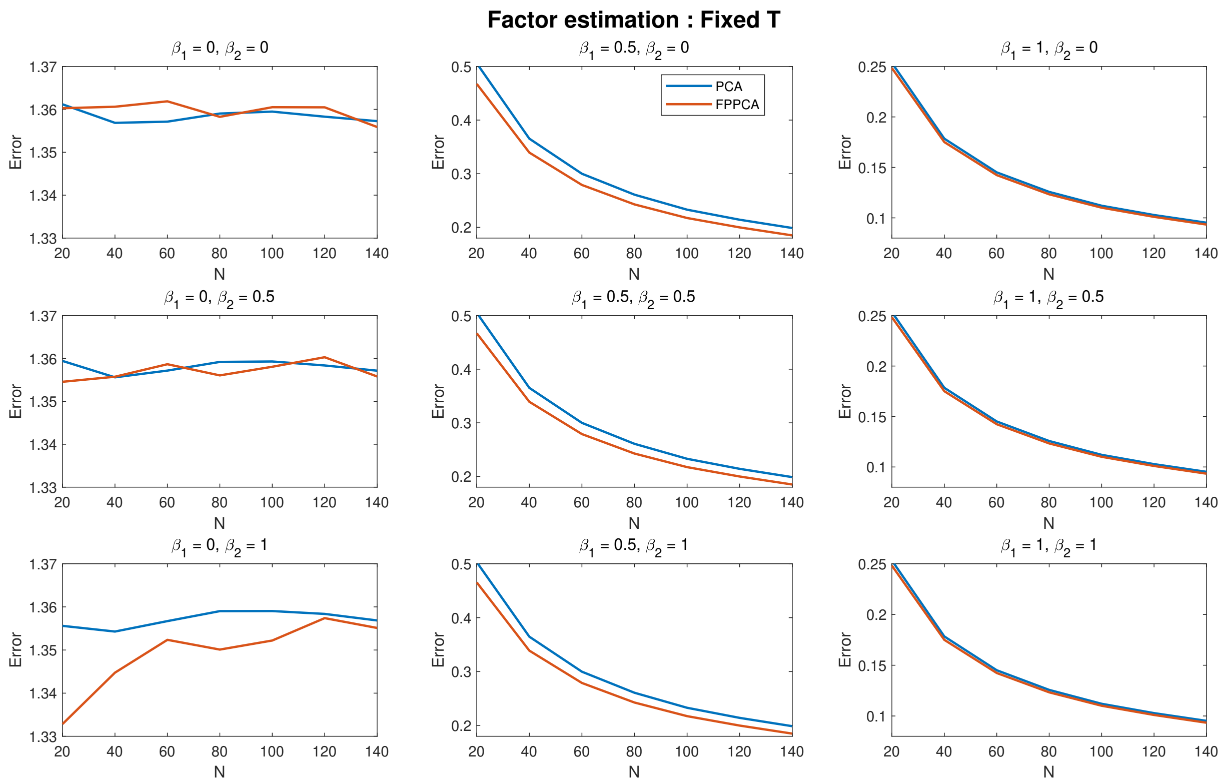

Figure 3: *Averaged errors of estimator for factors with fixed T samples, where both X_1, X_2 are observed.*

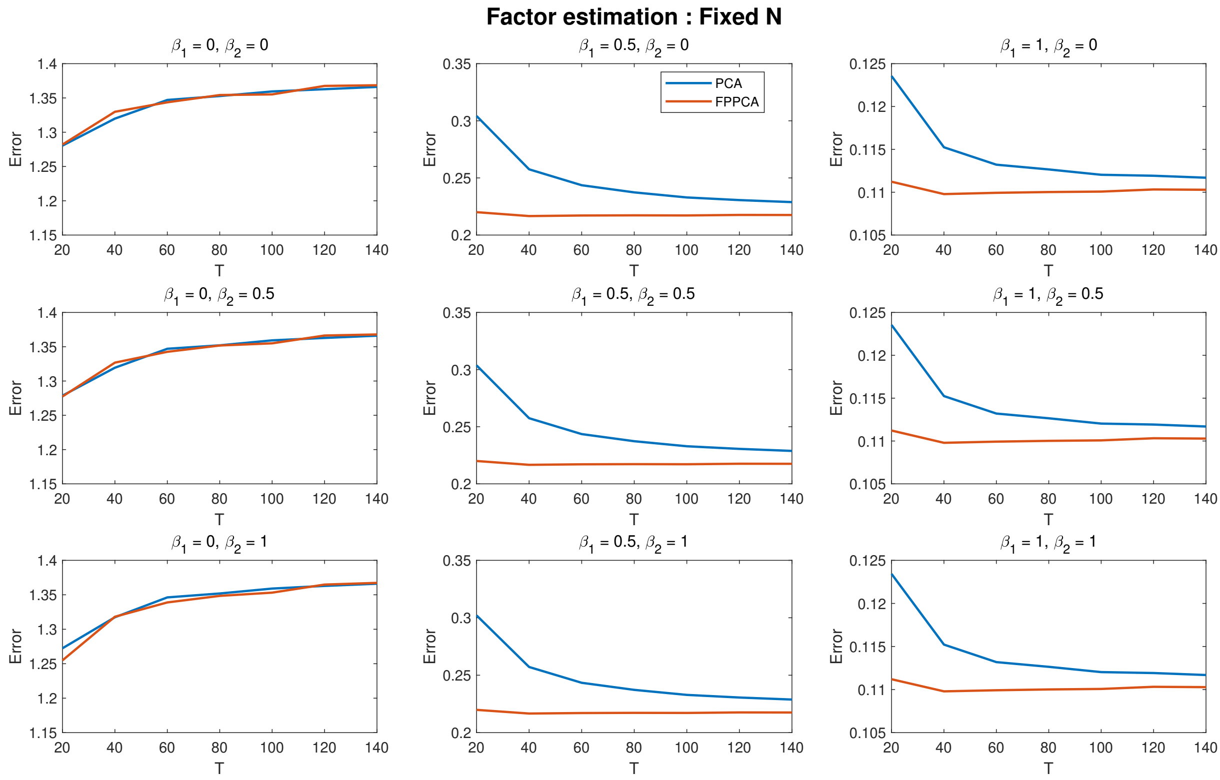

Figure 4: *Averaged errors of estimator for factors with fixed N samples, where only X_1 is observed.*

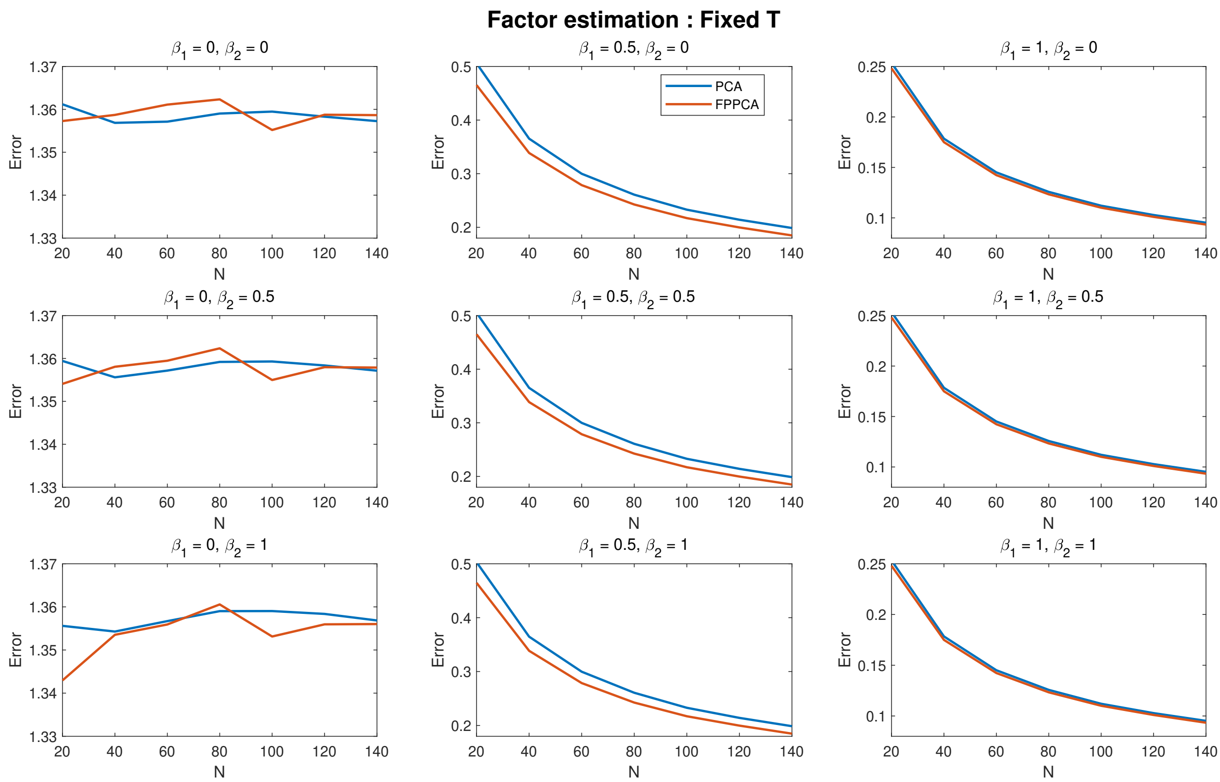

Figure 5: *Averaged errors of estimator for factors with fixed T samples, where only X_1 is observed.*

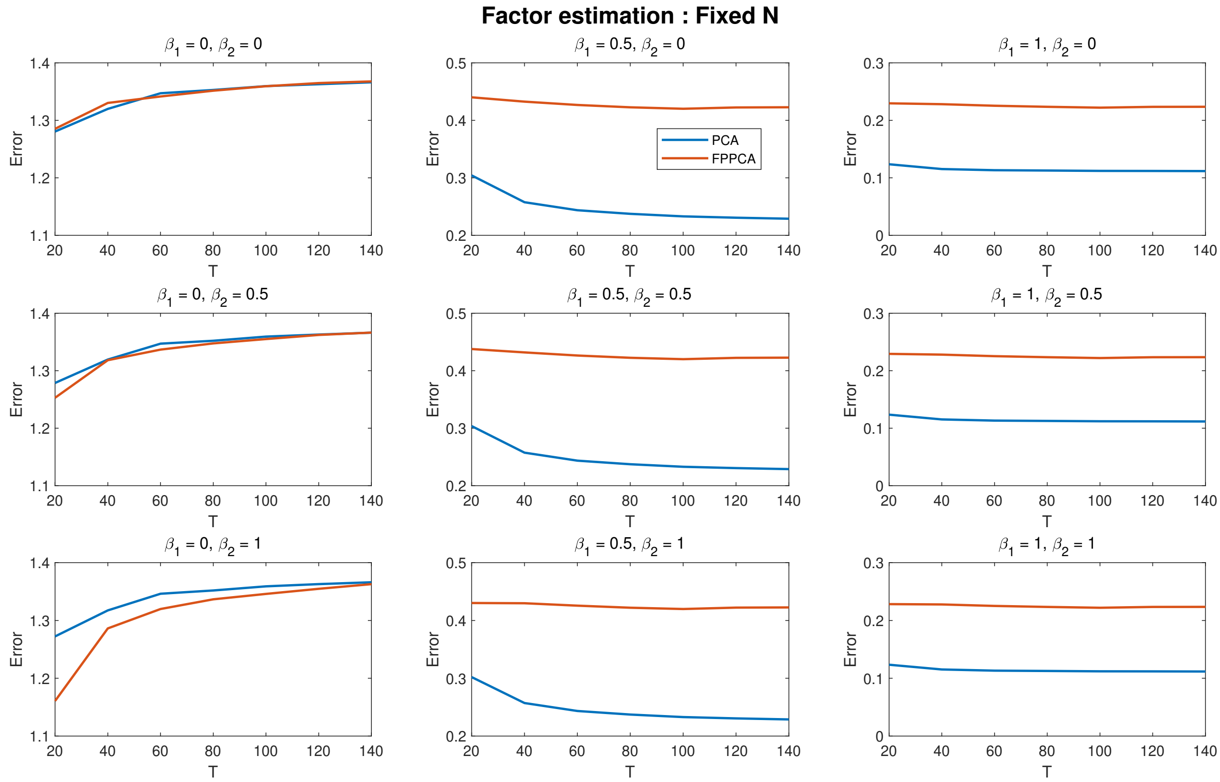

Figure 6: *Averaged errors of estimator for factors with fixed N samples, where only X_2 is observed.*

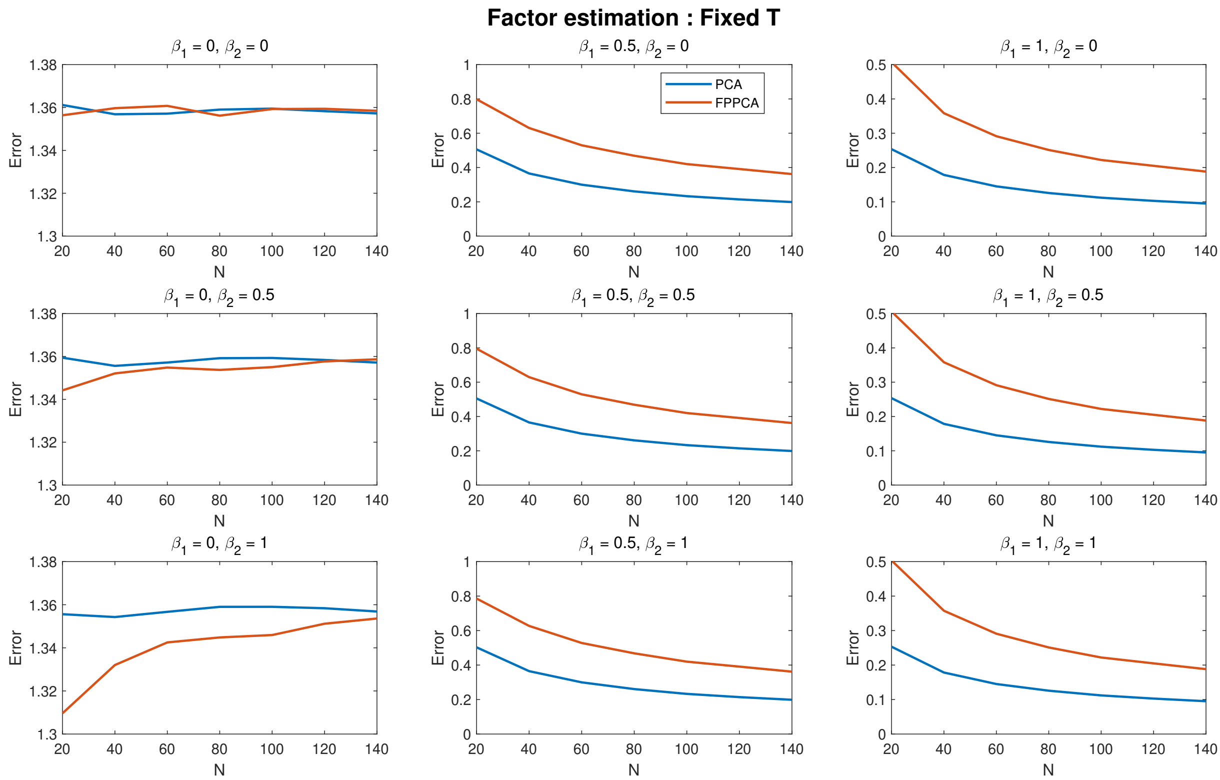

Figure 7: *Averaged errors of estimator for factors with fixed T samples, where only X_2 is observed.*

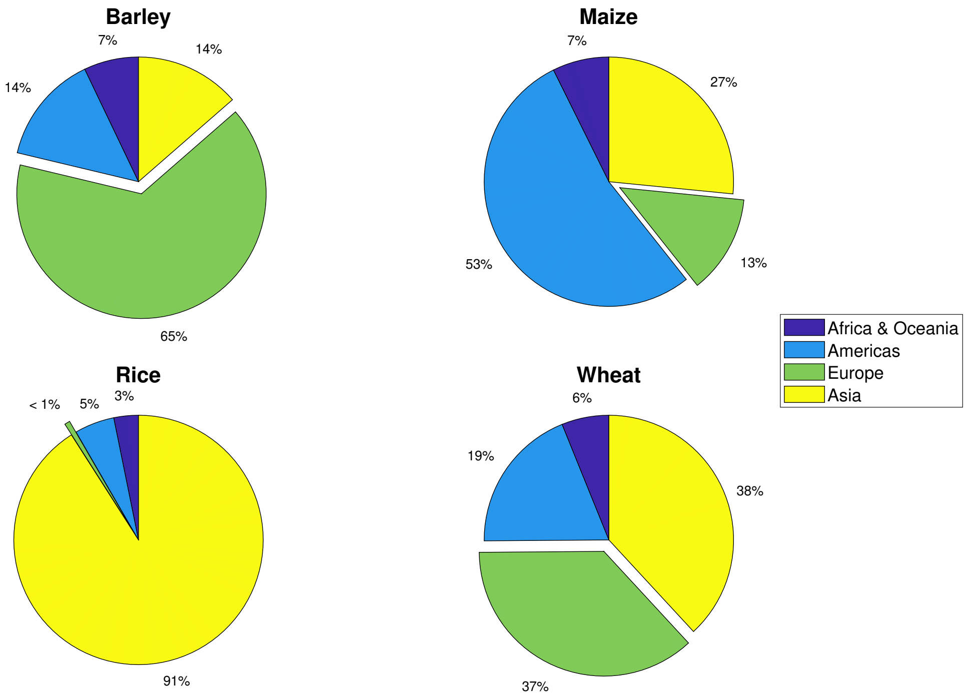

Figure 8: *Annual average production share by region 1961 - 2020*

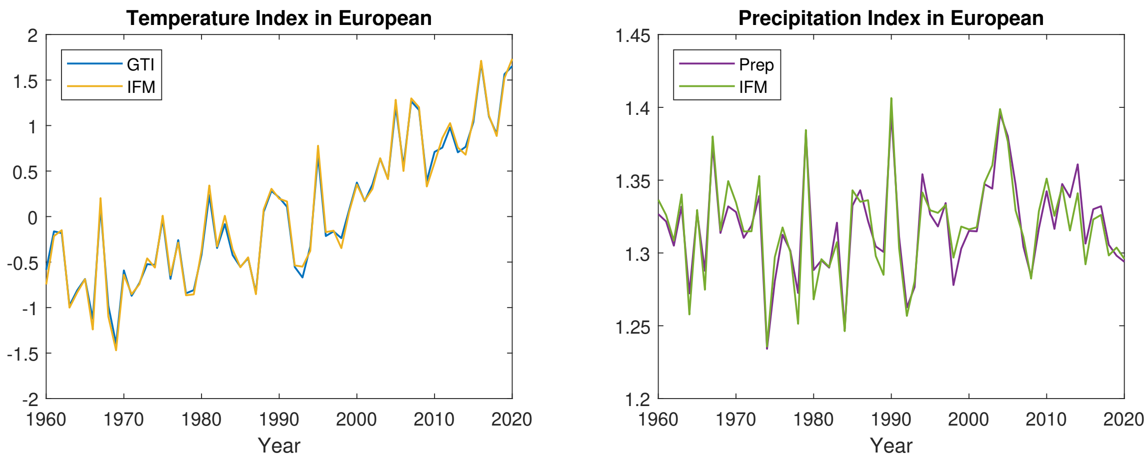

Figure 9: *IFM estimation results of the NCEP/NCAR Reanalysis dataset and NOAA PREC/L from 1960 to 2020 in European territory. The left panel shows the temperature index and the estimated yearly average temperature, and the right panel shows the actual and estimated annual average precipitation rate.*

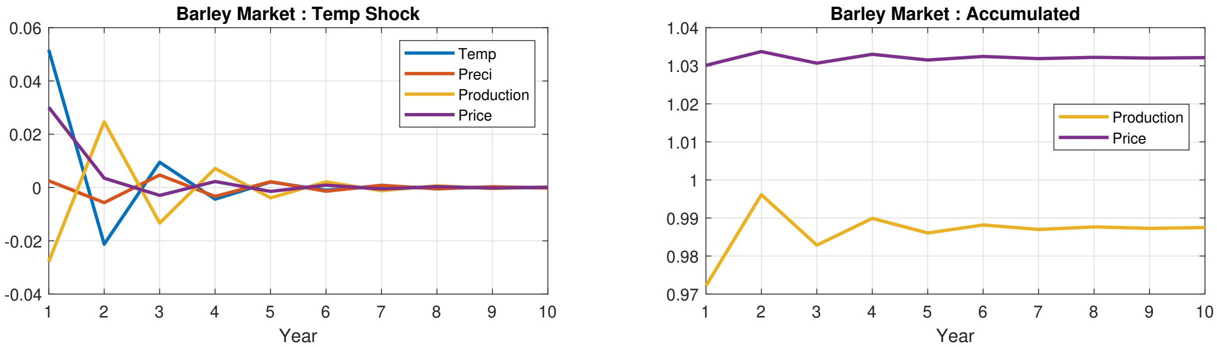

Figure 10: *Impulse response functions temperature shock on barely market*

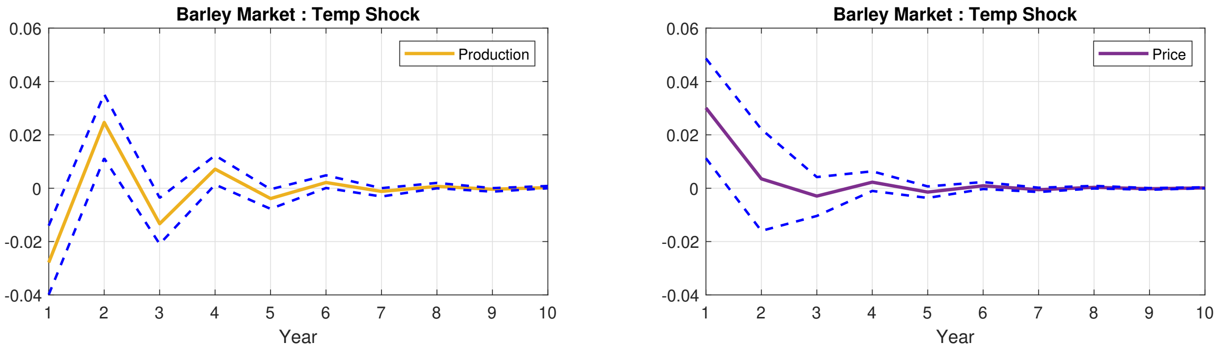

Figure 11: *68% confidence intervals of IRF on barely market*

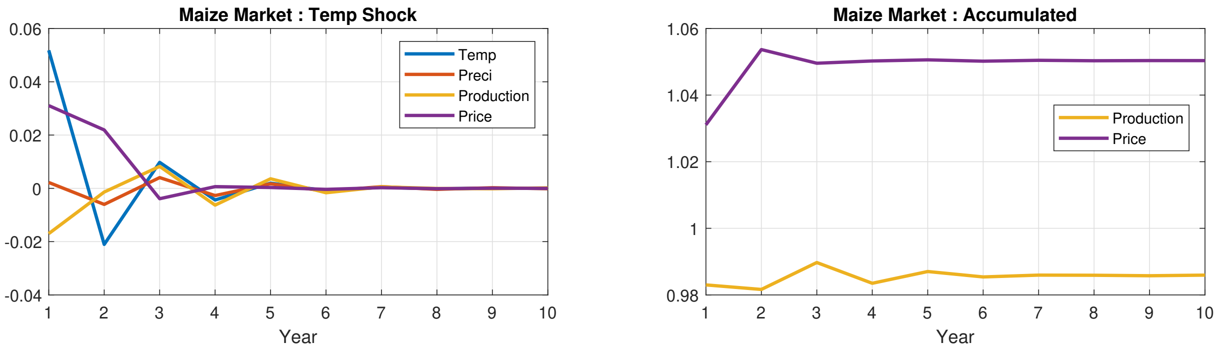

Figure 12: *Impulse response functions temperature shock on maize market*

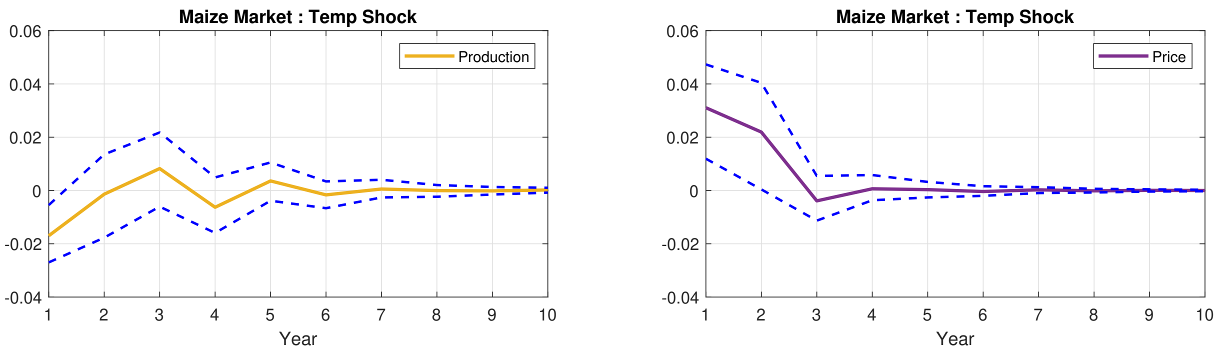

Figure 13: *68% confidence intervals of IRF on maize market*

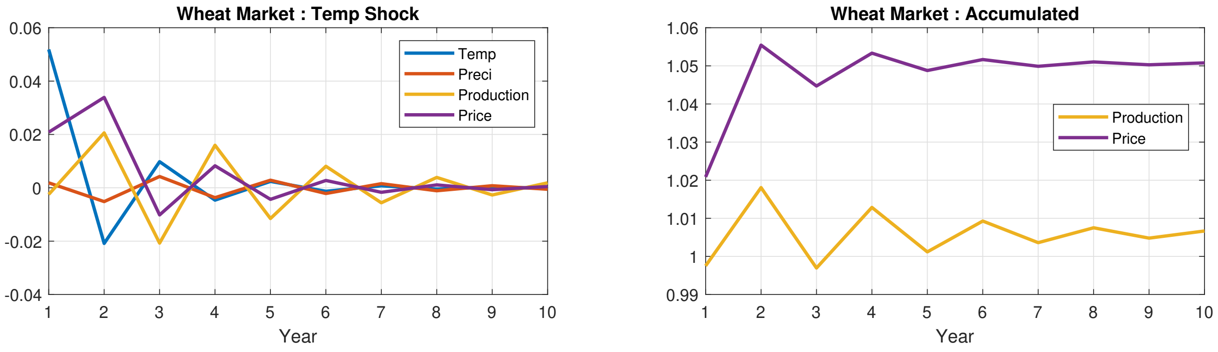

Figure 14: *Impulse response functions temperature shock on wheat market*

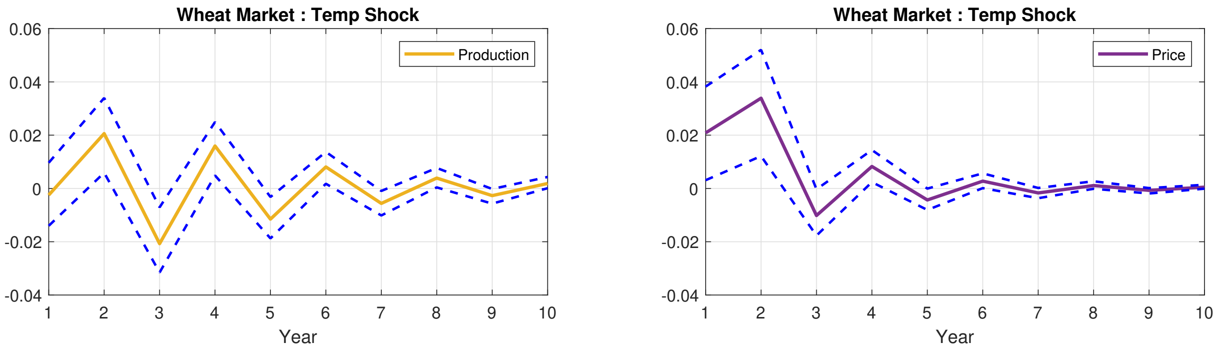

Figure 15: *68% confidence intervals of IRF on wheat market*

Appendix

A Bounded Linear Operators

In this section, we briefly define bounded linear operators used for the theoretical proofs. Let $\mathcal{D}_1$, $\mathcal{D}_2$ be any separable Hilbert spaces equipped with inner products $\langle \cdot\,,\cdot\rangle_{\mathcal{D}_1}$, $\langle \cdot\,,\cdot\rangle_{\mathcal{D}_2}$ respectively, and define $\mathbb{L}(\mathcal{D}_1, \mathcal{D}_2)$ to be the space of bounded linear operators from $\mathcal{D}_1$ to $\mathcal{D}_2$. For any $V \in \mathbb{L}(\mathcal{D}_1, \mathcal{D}_2)$, the operator norm is

$$\|V\|_\infty := \sup_{x\in\mathcal{D}_1, x\neq 0} \frac{\|Vx\|_{\mathcal{D}_2}}{\|x\|_{\mathcal{D}_1}} < \infty.$$

An operator V is called trace class if for an (complete) orthonormal basis $\{e_j : j \geq 1\}$ of $\mathcal{D}_1$

$$\|V\|_{TR} := \sum_{j\geq 1} \langle (V^*V)^{1/2} e_j, e_j\rangle_{\mathcal{D}_1} < \infty.$$

Let $s_j[\cdot]$ be the j-th largest singular value of an operator. By the singular value decomposition, $\|V\|_{HS} = \left(\sum_{j\geq 1} s_j^2[V]\right)^{1/2}$ and $\|V\|_{TR} = \sum_{j\geq 1} s_j[V]$. Then, a trace-class operator V is also Hilbert–Schmidt since

$$\|V\|_{HS} \leq \left(s_1[A] \sum_{j\geq 1} s_j[V]\right)^{1/2} = (s_1[V]\|V\|_{TR})^{1/2} < \infty.$$

In addition, for any $W \in \mathbb{L}(\mathcal{D}_2, \mathcal{D}_3)$ and Hilbert-Schmidt operator $V \in \mathbb{L}(\mathcal{D}_1, \mathcal{D}_2)$, we have $\|WV\|_{HS} \leq \|W\|_\infty \|V\|_{HS}$. For more details of bounded linear operators, see Hsing and Eubank (2015).

B Proofs for Section 3

B.1 Proof of Theorem 3.1

Proof. We begin our proof by showing the convergence rate of $\widehat{F}$ in **Part 1**, and the convergence rates for $\widehat{G}(X)$ and $\widehat{B}$ are proved in **Part 2** and **Part 3** respectively.

Part 1: $\frac{1}{T}\|\widehat{F} - F\|^2$

Define V to be a $K \times K$ diagonal matrix of the K largest eigenvalues of $\frac{1}{NT}(PY)^*PY$ such that the following holds

$$\frac{1}{NT}(PY)^*PY\widehat{F} = \widehat{F}V, \tag{11}$$

and since V is invertible $\widehat{F} = \frac{1}{NT}(PY)^*PY\widehat{F}V^{-1}$. In addition, we define $H = \frac{1}{NT}(\Phi(X)B)^*(\Phi(X)B)F'\widehat{F}V^{-1}$. Given that

$$\widehat{F} - F = \widehat{F} - FH + F(H - I_K),$$

we require to bound $\frac{1}{T}\|\widehat{F}-FH\|^2$ and $\frac{1}{T}\|F(H-I_K)\|^2$. First, we prove $\frac{1}{T}\|\widehat{F}-FH\|^2 = O_p(\frac{1}{N}+\frac{1}{J^\kappa})$. Recall the model in matrix representation,

$$\begin{aligned} Y &= [G(X)+\Gamma]F'+\varepsilon \\ &= [\Phi(X)B+R(X)+\Gamma]F'+\varepsilon. \end{aligned}$$

Then,

$$\widehat{F}-FH = \left(\sum_{m=1}^{15} W_m\right)V^{-1},$$

where

$$\begin{aligned}
&W_1 = \frac{1}{NT}FB^*\Phi(X)'\varepsilon\widehat{F}, && W_2 = \frac{1}{NT}\varepsilon^*\Phi(X)BF'\widehat{F}, && W_3 = \frac{1}{NT}\varepsilon^*P\varepsilon\widehat{F},\\
&W_4 = \frac{1}{NT}FB^*\Phi(X)'R(X)F'\widehat{F}, && W_5 = \frac{1}{NT}FR(X)^*\Phi(X)BF'\widehat{F}, && W_6 = \frac{1}{NT}FR(X)^*PR(X)F'\widehat{F},\\
&W_7 = \frac{1}{NT}FR(X)^*P\varepsilon\widehat{F}, && W_8 = \frac{1}{NT}\varepsilon^*PR(X)F'\widehat{F}, && W_9 = \frac{1}{NT}FB^*\Phi(X)'\Gamma F'\widehat{F},\\
&W_{10} = \frac{1}{NT}F\Gamma^*\Phi(X)BF'\widehat{F}, && W_{11} = \frac{1}{NT}F\Gamma^*P\varepsilon\widehat{F}, && W_{12} = \frac{1}{NT}\varepsilon^*P\Gamma F'\widehat{F},\\
&W_{13} = \frac{1}{NT}FR(X)^*P\Gamma F'\widehat{F}, && W_{14} = \frac{1}{NT}F\Gamma^*PR(X)F'\widehat{F}, && W_{15} = \frac{1}{NT}F\Gamma^*P\Gamma F'\widehat{F}.
\end{aligned}$$

By Lemma B.1,

$$\begin{aligned} \frac{1}{T}\|\widehat{F}-FH\|^2 &= \left[O_p\left(\frac{1}{N}\right)+O_p\left(\frac{1}{J^\kappa}\right)\right]O_p(1) \\ &= O_p\left(\frac{1}{N}+\frac{1}{J^\kappa}\right). \end{aligned} \tag{12}$$

ρ_N is at most a constant. Hence, the speed of $W_9 = O_p(\rho_N/N)$ at least as fast as $W_1 = O_p(1/N)$.

Second, to prove $\frac{1}{T}\|F(H-I_K)\|^2$ is bounded, it is equivalent to bound $\|H-I_K\|^2$ since $\frac{1}{T}\|F\|^2 = O_p(1)$. By Lemma B.6, we have

$$\|H-I_K\|^2 = O_p\left(\frac{1}{N^2}+\frac{1}{NT}+\frac{1}{J^\kappa}+\frac{\rho_N}{N}\right). \tag{13}$$

Hence, by (12) and (13),

$$\begin{aligned} \frac{1}{T}\|\widehat{F}-F\|^2 &= \frac{1}{T}\|\widehat{F}-FH+HF-F\|^2 \\ &\le \frac{1}{T}\|\widehat{F}-FH\|^2+\|H-I_K\|^2+\frac{2}{T}\|\widehat{F}-FH\|\|H-I_K\| \\ &\le \frac{2}{T}\|\widehat{F}-FH\|^2+2\|H-I_K\|^2 \\ &= O_p\left(\frac{1}{N}+\frac{1}{J^\kappa}\right), \end{aligned}$$

where the third line is implied by the inequality $\left(\|\widehat{F}-FH\|-\|H-I_K\|\right)^2 \geq 0$.

Part 2: $\frac{1}{N}\|\widehat{G}(X)-G(X)\|^2$

By Lemma B.7, $\frac{1}{N}\|\widehat{G}(X)-G(X)H\|^2 = O_p\left(\frac{J}{N^2}+\frac{J}{NT}+\frac{J}{J^\kappa}+\frac{J\rho_N}{N}\right)$. Hence,

$$\begin{aligned}\frac{1}{N}\|\widehat{G}(X)-G(X)\|^2 &= \frac{1}{N}\|\widehat{G}(X)-G(X)H+G(X)H-G(X)\|^2\\ &\leq \frac{2}{N}\|\widehat{G}(X)-G(X)H\|^2+\frac{2}{N}\|G(X)(H-I_K)\|^2\\ &\leq \frac{2}{N}\|\widehat{G}(X)-G(X)H\|^2+\frac{2}{N}\|G(X)\|_\infty^2\|H-I_K\|^2\\ &= O_p\left(\frac{J}{N^2}+\frac{J}{NT}+\frac{J}{J^\kappa}+\frac{J\rho_N}{N}\right),\end{aligned}$$

where the third inequality holds because $\frac{1}{N}\|G(X)\|_2^2 = O_p(1)$ by the assumption 3.2, and $\|AB\| \leq \|A\|_\infty\|B\|$.

Part 3: $\|\widehat{B}-B\|^2$

By Lemma B.6 and B.7, we have $\|H-I_K\|^2 = O_p\left(\frac{1}{N^2}+\frac{1}{NT}+\frac{1}{J^\kappa}+\frac{\rho_N}{N}\right)$, and $\|\widehat{B}-BH\|^2 = O_p\left(\frac{J}{N^2}+\frac{J}{NT}+\frac{J}{J^\kappa}+\frac{J\rho_N}{N}\right)$. Then

$$\begin{aligned}\|\widehat{B}-B\|^2 &= \|\widehat{B}-BH+BH-B\|^2\\ &\leq 2\|\widehat{B}-BH\|^2+2\|B(H-I_K)\|^2\\ &\leq 2\|\widehat{B}-BH\|^2+2\|B\|^2\|H-I_K\|^2\\ &= O_p\left(\frac{J}{N^2}+\frac{J}{NT}+\frac{J}{J^\kappa}+\frac{J\rho_N}{N}\right),\end{aligned}$$

since $\|B\|^2 = O_p(J)$. □

B.2 Proof of Corollary 3.1

By Lemma B.6 and B.7, we have $\|H-I_K\|^2 = O_p\left(\frac{1}{N^2}+\frac{1}{NT}+\frac{1}{J^\kappa}+\frac{\rho_N}{N}\right)$, and $\frac{1}{N}\|\widehat{\Gamma}-\Gamma H\|^2 = O_p(\frac{1}{J^\kappa}+\frac{J\rho_N}{N}+\frac{J}{N^2}+\frac{1}{T})$. Then

$$\begin{aligned}\frac{1}{N}\|\widehat{\Gamma}-\Gamma\|^2 &= \frac{1}{N}\|\widehat{\Gamma}-\Gamma H+\Gamma H-\Gamma\|^2\\ &\leq \frac{2}{N}\|\widehat{\Gamma}-\Gamma H\|^2+\frac{2}{N}\|\Gamma(H-I_K)\|^2\\ &\leq \frac{2}{N}\|\widehat{\Gamma}-\Gamma H\|^2+\frac{2}{N}\|\Gamma\|_\infty^2\|H-I_K\|^2\\ &= O_p\left(\frac{1}{J^\kappa}+\frac{J\rho_N}{N}+\frac{J}{N^2}+\frac{1}{T}\right)+O_p(\rho_N)\,O_p\left(\frac{1}{N^2}+\frac{1}{NT}+\frac{1}{J^\kappa}+\frac{1}{N}\right)\\ &= O_p\Big(\frac{1}{J^\kappa}+\frac{J\rho_N}{N}+\frac{J}{N^2}+\frac{1}{T}\Big).\end{aligned}$$

B.3 Lemmas for Theorem 3.1

Lemma B.1.

1. $\|V\|_\infty = O_p(1), \quad \|V^{-1}\|_\infty = O_p(1), \quad \|H\|_\infty = O_p(1).$
2. $\frac{1}{T}\|W_1\|^2 = O_p\left(\frac{1}{N}\right), \quad \frac{1}{T}\|W_2\|^2 = O_p\left(\frac{1}{N}\right).$
3. $\frac{1}{T}\|W_3\|^2 = O_p\left(\frac{J^2}{N^2}\right), \quad \frac{1}{T}\|W_4\|^2 = O_p\left(J^{-\kappa}\right), \quad \frac{1}{T}\|W_5\|^2 = O_p\left(J^{-\kappa}\right).$
4. $\frac{1}{T}\|W_6\|^2 = O_p\left(J^{-2\kappa}\right), \quad \frac{1}{T}\|W_7\|^2 = O_p\left(\frac{1}{NJ^{\kappa-1}}\right), \quad \frac{1}{T}\|W_8\|^2 = O_p\left(\frac{1}{NJ^{\kappa-1}}\right).$
5. $\frac{1}{T}\|W_9\|^2 = O_p\left(\frac{\rho_N}{N}\right), \quad \frac{1}{T}\|W_{10}\|^2 = O_p\left(\frac{\rho_N}{N}\right), \quad \frac{1}{T}\|W_{11}\|^2 = O_p\left(\frac{J^2\rho_N}{N^2}\right), \quad \frac{1}{T}\|W_{12}\|^2 = O_p\left(\frac{J^2\rho_N}{N^2}\right).$
6. $\frac{1}{T}\|W_{13}\|^2 = O_p\left(\frac{\rho_N}{NJ^{\kappa-1}}\right), \quad \frac{1}{T}\|W_{14}\|^2 = O_p\left(\frac{\rho_N}{NJ^{\kappa-1}}\right), \quad \frac{1}{T}\|W_{15}\|^2 = O_p\left(\frac{J\rho_N}{N^3}\right).$

Proof of Lemma B.1.

1. Recall that V is a $K \times K$ diagonal matrix of the first K largest eigenvalues of $\frac{1}{NT}(PY)^*PY$ such that the following holds

$$\frac{1}{NT}(PY)^*(PY)\widehat{F} = \widehat{F}V.$$

Since the square matrix $\Phi(X)'\Phi(X)/N$ has the full rank in probability, we can write

$$\frac{1}{NT}(PY)^*PY = \frac{1}{NT}Y^*\Phi(X)(\Phi(X)'\Phi(X))^{-1/2}(\Phi(X)'\Phi(X))^{-1/2}\Phi(X)'Y.$$

Define A to be the outer product of $\frac{1}{\sqrt{NT}}(\Phi(X)'\Phi(X))^{-1/2}\Phi(X)'Y$,

$$A = \frac{1}{NT}(\Phi(X)'\Phi(X))^{-1/2}\Phi(X)'YY^*\Phi(X)(\Phi(X)'\Phi(X))^{-1/2},$$

then, A and V have the same K largest eigenvalues. Given $Y = \Lambda F' + \varepsilon$, we can write

$$A = \sum_i^4 A_i,$$

where

$$\begin{aligned}
A_1 &= \frac{1}{N}(\Phi(X)'\Phi(X))^{-1/2}\Phi(X)'\Lambda\Lambda^*\Phi(X)(\Phi(X)'\Phi(X))^{-1/2},\\
A_2 &= \frac{1}{NT}(\Phi(X)'\Phi(X))^{-1/2}\Phi(X)'\Lambda F'\varepsilon^*\Phi(X)(\Phi(X)'\Phi(X))^{-1/2},\\
A_3 &= \frac{1}{NT}(\Phi(X)'\Phi(X))^{-1/2}\Phi(X)'\varepsilon F\Lambda^*\Phi(X)(\Phi(X)'\Phi(X))^{-1/2},\\
A_4 &= \frac{1}{NT}(\Phi(X)'\Phi(X))^{-1/2}\Phi(X)'\varepsilon\varepsilon^*\Phi(X)(\Phi(X)'\Phi(X))^{-1/2}.
\end{aligned}$$

We show that the K largest eigenvalues of A are determined by A_1 (in probability), and the rest three terms are negligible when $N \to \infty$ and $J = o(N)$. Similar to the construction of A, we know that the eigenvalues of A_1 are the same as those of $\frac{1}{NT}(P\Lambda)^*P\Lambda$, therefore the K largest eigenvalues of A_1 are bounded away from zero and infinity under the assumption 3.2. Hence, $\|V\|_\infty$ and $\|V^{-1}\|_\infty$ are both $O_p(1)$. Let us now prove that the eigenvalues of A and A_1 coincides in probability. Since $A_2^* = A_3$, we only prove the convergence rate of $\|A_2\|_\infty$.

$$
\begin{aligned}
\|A_2\|_\infty =& \frac{1}{NT}\| \left(\Phi(X)'\Phi(X)\right)^{-1/2} \Phi(X)' \left(\Lambda F'\varepsilon^*\right) \Phi(X) \left(\Phi(X)'\Phi(X)\right)^{-1/2} \|_\infty \\
\leq & \frac{1}{NT}\| \left(\Phi(X)'\Phi(X)\right)^{-1} \|_\infty \|\Phi(X)\|_\infty \|\Lambda\| \|F'\varepsilon^*\Phi(X)\| \\
= & \frac{1}{NT} O_p\left(\frac{1}{N}\right) O_P(\sqrt{N}) O_P(\sqrt{N}) O_p(\sqrt{NTJ}) \\
= & O_p\left(\sqrt{\frac{J}{NT}}\right),
\end{aligned}
$$

as $\|F'\varepsilon^*\Phi(X)\| = O_p(\sqrt{NTJ})$ by Lemma C.1.

By Lemma C.1, we have

$$
\begin{aligned}
\|A_4\|_\infty =& \frac{1}{NT}\| \left(\Phi(X)'\Phi(X)\right)^{-1/2} \Phi(X)'\varepsilon\varepsilon^*\Phi(X) \left(\Phi(X)'\Phi(X)\right)^{-1/2} \|_\infty \\
\leq & \frac{1}{NT}\| \left(\Phi(X)'\Phi(X)\right)^{-1} \|_\infty \|\Phi(X)'\varepsilon\|^2 \\
= & \frac{1}{NT} O_p\left(\frac{1}{N}\right) O_p(NTJ) \\
= & O_p\left(\frac{J}{N}\right).
\end{aligned}
$$

Hence, as $N \to \infty$, $J = o(N)$,

$$
\begin{aligned}
|\psi_k(A) - \psi_k(A_1)| &\leq \|A - A_1\|_\infty \\
&= \|A_2 - A_3 - A_4\|_\infty \\
&= O_p\left(\sqrt{\frac{J}{NT}} + \frac{J}{N}\right),
\end{aligned}
$$

for $k \geq 1$.

Recall $H = \frac{1}{NT} B^*\Phi(X)'\Phi(X)BF'\widehat{F}V^{-1}$. By the assumptions 3.2 and 3.6, we have $\|G(X)\|_\infty^2 = O_p(N)$, $\|R(X)\|^2 = O_p(NJ^{-\kappa})$. Then

$$
\begin{aligned}
\|H\|_\infty &= \|\frac{1}{NT} B^*\Phi(X)'\Phi(X)BF'\widehat{F}V^{-1}\|_\infty \\
&= \frac{1}{N}\|(G(X) - R(X))^*(G(X) - R(X))\frac{F'\widehat{F}}{T}V^{-1}\|_\infty \\
&\leq \frac{1}{N}\|G(X)^*G(X) + R(X)^*R(X) - 2R(X)^*G(X)\|_\infty \|\frac{F'\widehat{F}}{T}\| \|V^{-1}\|_\infty \\
&\leq \frac{1}{N}\left(\|G(X)^*G(X)\|_\infty + \|R(X)^*R(X)\|_\infty + 2\|R(X)^*G(X)\|_\infty\right) \|\frac{F'\widehat{F}}{T}\| \|V^{-1}\|_\infty \\
&\leq \frac{2}{N}\left(\|G(X)\|_2^2 + \|R(X)\|^2\right) \|\frac{F'\widehat{F}}{T}\| \|V^{-1}\|_\infty \\
&= \frac{2}{N}\left(O_p(N) + O_p(J^{-\kappa})\right) O_p(1) O_p(1) = O_p(1).
\end{aligned}
$$

2. By Lemma C.2, $\|B^*\Phi(X)'\varepsilon\|^2 = O_p(NT)$. Then
$$\begin{aligned}\frac{1}{T}\|W_1\|^2 &= \frac{1}{T}\|\frac{1}{NT}FB^*\Phi(X)'\varepsilon\widehat{F}\|^2\\ &\le \frac{1}{T^3N^2}\|F\|^2\|B^*\Phi(X)'\varepsilon\|^2\|\widehat{F}\|^2\\ &= O_p\left(\frac{1}{N}\right).\end{aligned}$$
Similarly, $\|W_2\| = O_p\left(\frac{1}{N}\right)$.

3. By Lemma C.1, $\|\Phi(X)'\varepsilon\| = O_p(NTJ)$, and the assumption 3.3 implies that $\|(\Phi(X)'\Phi(X))^{-1}\|_\infty^2 = O_p\left(\frac{1}{N^2}\right)$. Then
$$\begin{aligned}\frac{1}{T}\|W_3\|^2 &= \frac{1}{T}\|\frac{1}{NT}\varepsilon^*P\varepsilon\widehat{F}\|^2\\ &\le \frac{1}{T^3N^2}\|\Phi(X)'\varepsilon\|^4\|(\Phi(X)'\Phi(X))^{-1}\|_\infty^2\|\widehat{F}\|^2\\ &= O_p\left(\frac{J^2}{N^2}\right).\end{aligned}$$
By the assumption 3.5 and 3.6, $\|R(X)\|^2 = O_p(NJ^{-\kappa})$, and $\|\Phi(X)B\|_\infty \le \|G(X)\|_\infty + \|R(X)\| = O_p(\sqrt{N})$, dominated by $\|G(X)\|_\infty$. Hence
$$\begin{aligned}\frac{1}{T}\|W_4\|^2 &= \frac{1}{T}\|\frac{1}{NT}FB^*\Phi(X)'R(X)F'\widehat{F}\|^2\\ &\le \frac{1}{T^3N^2}\|F\|^4\|\widehat{F}\|^2\|\Phi(X)B\|_\infty^2\|R(X)\|^2\\ &= \frac{1}{T^3N^2}O_p(T^2)O_p(T)O_p(N)O_p(NJ^{-\kappa})\\ &= O_p(J^{-\kappa}).\end{aligned}$$
The same method applies to $\frac{1}{T}\|W_5\|^2$.

4. As $\|P\|_\infty = \|\Phi(X)(\Phi(X)'\Phi(X))^{-1}\Phi(X)'\|_\infty = O_p(1)$,
$$\begin{aligned}\frac{1}{T}\|W_6\|^2 &= \frac{1}{T}\|\frac{1}{NT}FR(X)^*PR(X)F'\widehat{F}\|^2\\ &\le \frac{1}{T^3N^2}\|F\|^4\|\widehat{F}\|^2\|R(X)\|^4\|P\|_\infty^4\\ &= \frac{1}{T^3N^2}O_p(T^2)O_p(T)O_p(N^2J^{-2\kappa})O_p(1)\\ &= O_p(J^{-2\kappa}).\end{aligned}$$
By Lemma C.1,
$$\begin{aligned}\frac{1}{T}\|W_7\|^2 &= \frac{1}{T}\|\frac{1}{NT}FR(X)^*P\varepsilon\widehat{F}\|^2\\ &\le \frac{1}{T^3N^2}\|F\|^2\|\widehat{F}\|^2\|R(X)\|^2\|P\varepsilon\|^2\\ &= \frac{1}{T^3N^2}O_p(T)O_p(T)O_p(NJ^{-\kappa})O_p(TJ)\\ &= O_p\left(\frac{1}{NJ^{\kappa-1}}\right).\end{aligned}$$
Similarly, $\frac{1}{T}\|W_8\|^2 = O_p(\frac{1}{NJ^{\kappa-1}})$.

5. By Lemma C.2, $\|B^*\Phi(X)'\Gamma\| = O_p(N\rho_N)$.

$$\begin{aligned}\frac{1}{T}\|W_9\|^2 &= \frac{1}{T}\|\frac{1}{NT}FB^*\Phi(X)'\Gamma F'\widehat{F}\|^2\\ &\leq \frac{1}{T^3N^2}\|F\|^4\|\widehat{F}\|^2\|B^*\Phi(X)'\Gamma\|^2\\ &= O\left(\frac{\rho_N}{N}\right).\end{aligned}$$

The same method applies to $\frac{1}{T}\|W_{10}\|^2$.

By Lemma C.2, $\|\Phi(X)'\Gamma\|^2 = O_p(JN\rho_p)$.

$$\begin{aligned}\frac{1}{T}\|W_{11}\|^2 &= \frac{1}{T}\|\frac{1}{NT}F\Gamma^*P\varepsilon\widehat{F}\|^2\\ &\leq \frac{1}{T^3N^2}\|F\|^2\|\widehat{F}\|^2\|\Gamma^*\Phi(X)\|^2\|(\Phi(X)'\Phi(X))^{-1})\|_\infty^2\|\Phi(X)'\varepsilon\|^2\\ &= O_p\left(\frac{J^2\rho_N}{N^2}\right).\end{aligned}$$

Similarly, $\frac{1}{T}\|W_{12}\|^2 = O_p\left(\frac{J^2\rho_N}{N^2}\right)$.

6. By Lemma C.2,

$$\begin{aligned}\frac{1}{T}\|W_{13}\|^2 &= \frac{1}{T}\|\frac{1}{NT}FR(X)^*P\Gamma F'\widehat{F}\|^2\\ &\leq \frac{1}{T^3N^2}\|F\|^4\|\widehat{F}\|^2\|R(X)\|^2\|\Phi(X)(\Phi(X)'\Phi(X))^{-1}\|_\infty^2\|\Phi(X)'\Gamma\|^2\\ &= O_p\left(\frac{\rho_N}{NJ^{\kappa-1}}\right)\end{aligned}$$

Similarly, $\frac{1}{T}\|W_{14}\|^2 = O_p\left(\frac{\rho_N}{NJ^{\kappa-1}}\right)$.

7. Again, by Lemma C.2

$$\begin{aligned}\frac{1}{T}\|W_{15}\|^2 &= \frac{1}{T}\|\frac{1}{NT}F\Gamma^*P\Gamma F'\widehat{F}\|^2\\ &\leq \frac{1}{T^3N^2}\|F\|^4\|\widehat{F}\|^2\|\Gamma^*\Phi(X)\|^4\|(\Phi(X)'\Phi(X))^{-1}\|_\infty^2\\ &= O_p\left(\frac{J^2\rho_N^2}{N^2}\right).\end{aligned}$$

□

Lemma B.2 (Improved convergence rate)**.**

1. $\frac{1}{T}\|W_1\|^2 = O_p\left(\frac{1}{N^2} + \frac{1}{NJ^\kappa} + \frac{1}{NT}\right)$.

2. $\frac{1}{T}\|W_3\|^2 = O_p\left(\frac{J^2}{N^3} + \frac{J^4}{N^4} + \frac{J^{2-\kappa}}{N^2} + \frac{J^2}{N^2T}\right)$.

3. $\frac{1}{T}\|W_7\|^2 = O_p\left(\frac{J^{1-\kappa}}{N^2} + \frac{J^{3-\kappa}}{N^3} + \frac{J^{1-2\kappa}}{N} + \frac{J^{1-\kappa}}{NT}\right)$.

4. $\frac{1}{T}\|W_{11}\|^2 = O_p\left(\frac{J^2\rho_N}{N^2}[\frac{1}{N} + \frac{1}{T} + \frac{1}{J^\kappa}]\right)$.

Proof of Lemma B.2.

The improved convergence rates are achieved by using the by product result of Theorem 1, $\frac{1}{T}\|\widehat{F}-FH\|^2 = O_p\left(\frac{1}{N}+\frac{1}{J^\kappa}\right)$.

1. Using Lemma C.1, and C.2

$$\begin{aligned}\frac{1}{T}\|W_1\|^2 &= \frac{1}{T}\|\frac{1}{NT}FB^*\Phi(X)'\varepsilon\widehat{F}\|^2 \\ &\leq \frac{1}{T^3N^2}\|F\|^2\|B^*\Phi(X)'\varepsilon(\widehat{F}-FH+FH)\|^2 \\ &\leq \frac{1}{T^3N^2}\|F\|^2\left(2\|B^*\Phi(X)'\varepsilon\|^2\|\widehat{F}-FH\|^2+2\|B^*\Phi(X)'\varepsilon F\|^2\|H\|^2\right) \\ &= \frac{1}{T^3N^2}O_p(T)\left[O_p(NT)O_p\left(\frac{T}{N}+\frac{T}{J^\kappa}\right)+O_p(NT)O_p(1)\right] \\ &= O_p\left(\frac{1}{N^2}+\frac{1}{NJ^\kappa}+\frac{1}{NT}\right).\end{aligned}$$

2.

$$\begin{aligned}\frac{1}{T}\|W_3\|^2 &= \frac{1}{T}\|\frac{1}{NT}\varepsilon^*P\varepsilon\widehat{F}\|^2 \\ &\leq \frac{1}{T^3N^2}\|\varepsilon^*\Phi(X)\|^2\|(\Phi(X)'\Phi(X))^{-1}\|_\infty^2\left(\|\Phi(X)'\varepsilon(\widehat{F}-FH+FH)\|^2\right) \\ &\leq \frac{1}{T^3N^2}\|\varepsilon^*\Phi(X)\|^2\|(\Phi(X)'\Phi(X))^{-1}\|_\infty^2\left(\|\Phi(X)'\varepsilon\|^2\|\widehat{F}-FH\|^2+\|\Phi(X)'\varepsilon F\|^2\|H\|^2\right) \\ &= \frac{1}{T^3N^2}O_p(NTJ)O_p\left(\frac{1}{N^2}\right)\left[O_p(NTJ)O_p\left(\frac{T}{N}+\frac{T}{J^\kappa}\right)+O_p(NTJ)O_p(1)\right] \\ &= O_p\left(\frac{J^2}{N^3}+\frac{J^4}{N^4}+\frac{J^{2-\kappa}}{N^2}+\frac{J^2}{N^2T}\right).\end{aligned}$$

3.

$$\begin{aligned}\frac{1}{T}\|W_7\|^2 &= \frac{1}{T}\|\frac{1}{NT}FR(X)^*P\varepsilon\widehat{F}\|^2 \\ &\leq \frac{1}{T^3N^2}\|F\|^2\|R(X)\|^2\|\Phi(X)(\Phi(X)'\Phi(X))^{-1}\|_\infty^2\left(\|\Phi(X)'\varepsilon(\widehat{F}-FH+FH)\|^2\right) \\ &\leq \frac{1}{T^3N^2}\|F\|^2\|R(X)\|^2\|\Phi(X)(\Phi(X)'\Phi(X))^{-1}\|_\infty^2\left(\|\Phi(X)'\varepsilon\|^2\|\widehat{F}-FH\|^2+\|\Phi(X)'\varepsilon F\|^2\|H\|^2\right) \\ &= \frac{1}{T^3N^2}O_p(T)O_p(NJ^{-\kappa})O_p\left(\frac{1}{N}\right)\left[O_p(NTJ)O_p\left(\frac{T}{N}+\frac{T}{J^\kappa}\right)+O_p(NTJ)O_p(1)\right] \\ &= O_p\left(\frac{J^{1-\kappa}}{N^2}+\frac{J^{3-\kappa}}{N^3}+\frac{J^{1-2\kappa}}{N}+\frac{J^{1-\kappa}}{NT}\right).\end{aligned}$$

4.

$$\begin{aligned}\frac{1}{T}\|W_{11}\|^2 &= \frac{1}{T}\|\frac{1}{NT}F\Gamma^*P\varepsilon\widehat{F}\|^2 \\ &\leq \frac{1}{T^3N^2}\|F\|^2\|\Gamma^*\Phi(X)\|^2\|(\Phi(X)'\Phi(X))^{-1}\|_\infty^2\left(\|\Phi(X)'\varepsilon(\widehat{F}-FH+FH)\|^2\right) \\ &\leq \frac{1}{T^3N^2}\|F\|^2\|\Gamma^*\Phi(X)\|^2\|(\Phi(X)'\Phi(X))^{-1}\|_\infty^2\left(\|\Phi(X)'\varepsilon\|^2\|\widehat{F}-FH\|^2+\|\Phi(X)'\varepsilon F\|^2\|H\|^2\right) \\ &= \frac{1}{T^3N^2}O_p(T)O_p(JN\rho_N)O_p\left(\frac{1}{N^2}\right)\left[O_p(NTJ)O_p\left(\frac{T}{N}+\frac{T}{J^\kappa}\right)+O_p(NTJ)O_p(1)\right] \\ &= O_p\left(\frac{J^2\rho_N}{N^2}\left[\frac{1}{N}+\frac{1}{T}+\frac{1}{J^\kappa}\right]\right).\end{aligned}$$

□

Lemma B.3.

1. $\frac{1}{T^2}\|F'W_2\|^2 = O_p\left(\frac{1}{NT}\right)$.
2. $\frac{1}{T^2}\|F'W_3\|^2 = O_p\left(\frac{J^2}{N^3T} + \frac{J^2}{N^2TJ^\kappa} + \frac{J^2}{N^2T^2}\right)$.
3. $\frac{1}{T^2}\|F'W_8\|^2 = O_p\left(\frac{J}{NTJ^\kappa}\right)$.
4. $\frac{1}{T^2}\|F'W_{12}\|^2 = O_p\left(\frac{J^2}{N^3T}N\rho_N\right)$.

Proof of Lemma B.3.

1. By Lemma C.2, $\|B^*\Phi(X)'\varepsilon F\|^2 = O_p(NT)$. Hence,

$$\begin{aligned}\frac{1}{T^2}\|F'W_2\|^2 &= \frac{1}{T^4N^2}\|F'\varepsilon^*\Phi(X)BF'\widehat{F}\|^2 \\ &= \frac{1}{T^4N^2}\|F'\varepsilon^*\Phi(X)B\|^2\|F\|^2\|\widehat{F}\|^2 \\ &= O_p\left(\frac{1}{NT}\right).\end{aligned}$$

2. By Lemma C.1,

$$\begin{aligned}\frac{1}{T^2}\|F'W_3\|^2 &= \frac{1}{T^4N^2}\|F'\varepsilon^*P\varepsilon\widehat{F}\|^2 \\ &\le \frac{1}{T^4N^2}\|F'\varepsilon^*\Phi(X)\|^2\|(\Phi(X)'\Phi(X))^{-1}\|_\infty^2\|\Phi(X)'\varepsilon\widehat{F}\|^2 \\ &\le \frac{1}{T^4N^2}\|F'\varepsilon^*\Phi(X)\|^2\|(\Phi(X)'\Phi(X))^{-1}\|_\infty^2\left[\|\Phi(X)'\varepsilon(\widehat{F}-FH)\|^2 + \|\Phi(X)'\varepsilon FH\|^2\right] \\ &= \frac{1}{T^4N^2}O_p(NTJ)O_p\left(\frac{1}{N^2}\right)\left[O_p(NTJ)O_p\left(\frac{T}{N}+\frac{T}{J^\kappa}\right) + O_p(NTJ)O_p(1)\right] \\ &= O_p\left(\frac{J^2}{N^3T} + \frac{J^{2-\kappa}}{NT} + \frac{J^2}{N^2T^2}\right).\end{aligned}$$

3. By Lemma C.1

$$\begin{aligned}\frac{1}{T^2}\|F'W_8\|^2 &= \frac{1}{T^4N^2}\|F'\varepsilon^*PR(X)F'\widehat{F}\|^2 \\ &\le \frac{1}{T^4N^2}\|F'\varepsilon^*\Phi(X)\|^2\|(\Phi(X)'\Phi(X))^{-1}\Phi(X)\|_\infty^2\|R(X)\|^2\|F\|^2\|\widehat{F}\|^2 \\ &= \frac{1}{T^4N^2}O_p(NTJ)O_p\left(\frac{1}{N}\right)O_p(NJ^{-\kappa})O_p(T^2) \\ &= O_p\left(\frac{J^{1-\kappa}}{NT}\right).\end{aligned}$$

4. By Lemma C.1 and C.2,

$$\begin{aligned}\frac{1}{T^2}\|F'W_{12}\|^2 &= \frac{1}{T^4N^2}\|F'\varepsilon^*P\Gamma F'\widehat{F}\|^2 \\ &\le \frac{1}{T^4N^2}\|F'\varepsilon^*\Phi(X)\|^2\|(\Phi(X)'\Phi(X))^{-1}\|_\infty^2\|\Phi(X)'\Gamma\|^2\|F\|^2\|\widehat{F}\|^2 \\ &= \frac{1}{T^4N^2}O_p(NTJ)O_p\left(\frac{1}{N^2}\right)O_p(JN\rho_N)O_p(T^2) \\ &= O_p\left(\frac{J^2}{N^3T}N\rho_N\right).\end{aligned}$$

□

Lemma B.4.

1. $\frac{1}{T}\|F'(\widehat{F}-FH)\| = O_p\left(\frac{1}{N}+\frac{1}{\sqrt{NT}}+\frac{1}{J^{\kappa/2}}+\sqrt{\frac{\rho_N}{N}}\right)$.

2. $\frac{1}{T}\|\widehat{F}'(\widehat{F}-FH)\| = O_p\left(\frac{1}{N}+\frac{1}{\sqrt{NT}}+\frac{1}{J^{\kappa/2}}+\sqrt{\frac{\rho_N}{N}}\right)$.

Proof of Lemma B.4.

1. Note that $F'(\widehat{F}-FH) = \sum_{n=1}^{15} F'W_nV^{-1}$, where W_n are defined in Section B.1. Then, it follows by Lemma B.3,

$$\frac{1}{T^2}\left(\|F'W_2\|^2+\|F'W_3\|^2+\|F'W_8\|^2+\|F'W_{12}\|^2\right) = O_p\left(\frac{1}{NT}\right).$$

Let us consider the rest of the terms. By Lemma B.1-B.2,

$$\frac{1}{T^2}\|F'W_n\|^2 \le \frac{1}{T}\|W_n\|^2,$$

for all $n = 1$, 4-7, 9-11, 13-15 (Those terms all start with F, and thus we use the assumption that $F'F/T = I_K$). Then,

$$\begin{aligned}\frac{1}{T^2}\left(\|F'W_1\|^2+\|F'W_4\|^2+\cdots+\|F'W_{15}\|^2\right) &\le \frac{1}{T}\left(\|W_1\|^2+\|W_4\|^2+\cdots+\|W_{15}\|^2\right) \\ &= O_p\left(\frac{1}{N^2}+\frac{1}{NT}+\frac{1}{J^\kappa}+\frac{\rho_N}{N}\right).\end{aligned}$$

Using the result of Lemma B.3 of the rest of terms (W_2, W_3, W_8, W_{12}),

$$\begin{aligned}\frac{1}{T}\|F'(\widehat{F}-FH)\| &\le \frac{1}{T}\sum_{n=1}^{15}\|F'W_n\|\|V^{-1}\|_\infty \\ &= O_p\left(\frac{1}{N}+\frac{1}{\sqrt{NT}}+\frac{1}{J^{\kappa/2}}+\sqrt{\frac{\rho_N}{N}}\right).\end{aligned}$$

2. By the result above,

$$
\begin{aligned}
\frac{1}{T}\|\widehat{F}'(\widehat{F}-FH)\| &= \frac{1}{T}\|(\widehat{F}-FH+FH)'(\widehat{F}-FH)\| \\
&= \frac{1}{T}\|(\widehat{F}-FH)'(\widehat{F}-FH)+H'F'(\widehat{F}-FH)\| \\
&\leq \frac{1}{T}\|\widehat{F}-FH\|^2 + \frac{1}{T}\|F'(\widehat{F}-FH)\|\|H\|_\infty \\
&= O_p\left(\frac{1}{N}+\frac{1}{J^\kappa}\right) + \frac{1}{T}O_p(1)O_p\left(\frac{1}{N}+\frac{1}{\sqrt{NT}}+\frac{1}{J^{\kappa/2}}+\sqrt{\frac{\rho_N}{N}}\right) \\
&= O_p\left(\frac{1}{N}+\frac{1}{\sqrt{NT}}+\frac{1}{J^{\kappa/2}}+\sqrt{\frac{\rho_N}{N}}\right).
\end{aligned}
$$

□

Lemma B.5.

1. $\|H'H - I_K\| = O_p\left(\frac{1}{N}+\frac{1}{\sqrt{NT}}+\frac{1}{J^{\kappa/2}}+\sqrt{\frac{\rho_N}{N}}\right)$.

2. $\|H^{-1}\|_\infty = O_p(1)$.

Proof of Lemma B.5.

1. By the assumption 3.2, $(FH)'(FH)/T = H'H$ almost surely. Then,

$$
\begin{aligned}
H'H &= \frac{1}{T}(FH-\widehat{F}+\widehat{F})'FH \\
&= \frac{1}{T}(FH-\widehat{F})'FH + \frac{1}{T}\widehat{F}'FH \\
&= \frac{1}{T}(FH-\widehat{F})'FH + \frac{1}{T}\widehat{F}'(FH-\widehat{F}+\widehat{F}) \\
&= \frac{1}{T}(FH-\widehat{F})'FH + \frac{1}{T}\widehat{F}'(FH-\widehat{F}) + I_K,
\end{aligned}
$$

where $\widehat{F}'\widehat{F}/T = I_K$ by the definition of the estimator. Hence,

$$
\begin{aligned}
\|H'H - I_K\| &\leq \frac{1}{T}\|(FH-\widehat{F})'F\|\|H\|_\infty + \frac{1}{T}\|\widehat{F}'(FH-\widehat{F})\| \\
&= O_p\left(\frac{1}{N}+\frac{1}{\sqrt{NT}}+\frac{1}{J^{\kappa/2}}+\sqrt{\frac{\rho_N}{N}}\right),
\end{aligned}
$$

where it is follows by Lemma B.4.

2. Given the previous result, we have $\psi_{min}(H'H) \geq 1 + o_p(1)$. Hence,

$$
\|H^{-1}\|_\infty^2 = \psi_{max}(H^{-1}(H^{-1})') = \psi_{min}^{-1}(H'H) = O_p(1).
$$

□

Lemma B.6.

$$
\|H - I_K\| = O_p\left(\frac{1}{N}+\frac{1}{\sqrt{NT}}+\frac{1}{J^{\kappa/2}}+\sqrt{\frac{\rho_N}{N}}\right).
$$

Proof of Lemma B.6.
Let $\delta_{NTJ} = \frac{1}{N} + \frac{1}{\sqrt{NT}} + \frac{1}{J^{\kappa/2}} + \sqrt{\frac{\rho_N}{N}}$. Then by Lemma B.4,

$$\begin{aligned}\frac{1}{T}\widehat{F}'F &= \frac{1}{T}(\widehat{F} - FH + FH)'F \\ &= \frac{1}{T}(\widehat{F} - FH)'F + H' \\ &= H' + O_p(\delta_{NTJ}).\end{aligned}$$

By the definition of $H = \frac{1}{NT}B^*\Phi(X)'\Phi(X)BF'\widehat{F}V^{-1}$ and the above, we have

$$\begin{aligned}HV &= \frac{1}{NT}B^*\Phi(X)'\Phi(X)BF'\widehat{F} \\ &= \frac{1}{N}B^*\Phi(X)'\Phi(X)B(F'\widehat{F}/T) \\ &= \frac{1}{N}B^*\Phi(X)'\Phi(X)BH + O_p(\delta_{NYJ}).\end{aligned}$$

By the assumption 3.6, and since $G(X) = \Phi(X)B + R(X)$, we have $\|G(X)^*G(X)/N - B^*\Phi(X)'\Phi(X)B/N\| = O_p\left(\frac{1}{J^{\kappa/2}}\right)$. Hence,

$$\frac{1}{N}G(X)^*G(X)H = HV + O_p(\delta_{NTJ}). \tag{14}$$

In addition, Lemma B.5 implies that

$$I_K = H'H + O_p(\delta_{NTJ}). \tag{15}$$

We will show that H satisfies (14) and (15) if and only if $\|H - I_K\| = O_p(\delta_{NTJ})$.
Let h_{nm} be the (n,m)-th element of H, g_n be the n-th diagonal element of $G(X)^*G(X)/N$, and v_n be the n-th diagonal element of V. Note that $G(X)^*G(X)/N$ and V are $K \times K$ diagonal matrices. Our task here is to prove that for $n, m = 1, \ldots, K$,

$$h_{nm} = \begin{cases} 1 + O_p(\delta_{NTJ}), & \text{if } n = m \\ O_p(\delta_{NTJ}), & \text{if } n \neq m \end{cases}$$

In element-wise representation, (14) implies

$$g_n h_{nm} = h_{nm} v_m + O_p(\delta_{NTJ}), \quad n, m \leq K. \tag{16}$$

By Lemma B.5, $\|H'H - I_K\| = o_p(1)$, and thus h_{nn} must be non-zero elements in probability for all $n = 1, \ldots, K$. Hence,

$$g_n = v_n + O_p(\delta_{NTJ}), \quad n \leq K. \tag{17}$$

Therefore, by (16) and (17),

$$(g_n - g_m)h_{nm} = O_p(\delta_{NTJ}), \quad n, m \leq K. \tag{18}$$

The assumption 3.2 ensures g_n is non-zero and distinctive for all $n = 1, \ldots, K$, almost surely. So we have $h_{nm} = O_p(\delta_{NTJ})$ for all $n \neq m$ by the equality (18). Combined with this result, (15) implies that $h_{nn}^2 - 1 = O_p(\delta_{NTJ})$ for all $n = 1, \ldots, K$, thus, $h_{nn} = \pm 1 + O_p(\sqrt{\delta_{NTJ}})$. Without loss of generality, we assume that $h_{nn} = 1 + O_p(\sqrt{\delta_{NTJ}})$ since the sign of h_{nn} can be adjusted by simply multiplying -1 to both $\widehat{F}$ and $\widehat{G}(X)$. Note that our task is to show $h_{nn} = 1 + O_p(\delta_{NTJ})$, and it can be achieved since

$$h_{nn} - 1 = \frac{1}{2}\left[h_{nn}^2 - 1 - (h_{nn} - 1)^2\right] = O_p(\delta_{NTJ}).$$

As a result,

$$\|H - I_K\|^2 = \sum_{n \neq m} h_{nm}^2 + \sum_{n=1}^{K} (h_{nn} - 1)^2 = O_p(\delta_{NTJ}^2). \tag{19}$$

□

Lemma B.7.

1. $\|\widehat{B} - BH\|^2 = O_p\left(\frac{J}{N^2} + \frac{J}{NT} + \frac{J}{J^\kappa} + \frac{J\rho_N}{N}\right)$
2. $\frac{1}{N}\|\widehat{G}(X) - G(X)H\|^2 = O_p\left(\frac{J}{N^2} + \frac{J}{NT} + \frac{J}{J^\kappa} + \frac{J\rho_N}{N}\right)$
3. $\frac{1}{N}\|\widehat{\Gamma}(X) - \Gamma H\|^2 = O_p\left(\frac{1}{J^\kappa} + \frac{J\rho_N}{N} + \frac{J}{N^2} + \frac{1}{T}\right)$

Proof of Lemma B.7.

1. Recall that $Y = \Lambda F' + \varepsilon = [\Phi(X)B + \Gamma + R(X)]F' + \varepsilon$.

$$\begin{aligned}\widehat{B} &= \frac{1}{T}(\Phi(X)'\Phi(X))^{-1}\Phi(X)'Y\widehat{F} \\ &= BH + \sum_{n=1}^{5} D_n\end{aligned}$$

where

$$\begin{aligned}D_1 &= \frac{1}{T}(\Phi(X)'\Phi(X))^{-1}\Phi(X)'R(X)F'\widehat{F}, \quad D_2 = \frac{1}{T}(\Phi(X)'\Phi(X))^{-1}\Phi(X)'\varepsilon FH, \\ D_3 &= \frac{1}{T}(\Phi(X)'\Phi(X))^{-1}\Phi(X)'\varepsilon(\widehat{F} - FH), \quad D_4 = \frac{1}{T}BF'(\widehat{F} - FH), \\ D_5 &= \frac{1}{T}(\Phi(X)'\Phi(X))^{-1}\Phi(X)'\Gamma F'\widehat{F}.\end{aligned}$$

Let us now begine with $\|D_1\|$. By the assumption 3.2, 3.3 and 3.6

$$\begin{aligned}\|D_1\|^2 &= \frac{1}{T^2}\|(\Phi(X)'\Phi(X))^{-1}\Phi(X)'R(X)F'\widehat{F}\|^2 \\ &\leq \frac{1}{T^2}\|(\Phi(X)'\Phi(X))^{-1}\Phi(X)'\|_\infty^2\|R(X)\|^2\|F\|^2\|\widehat{F}\|^2 \\ &= \frac{1}{T^2}O_p\left(\frac{1}{N}\right)O_p(NJ^{-\kappa})O_p(T)O_p(T) \\ &= O_p(J^{-\kappa}).\end{aligned}$$

By Lemma B.1 and Lemma C.1,

$$
\begin{aligned}
\|D_2\|^2 &= \|\frac{1}{T}(\Phi(X)'\Phi(X))^{-1}\Phi(X)'\varepsilon FH\|^2 \\
&\leq \frac{1}{T^2}\|(\Phi(X)'\Phi(X))^{-1}\|_\infty^2\|\Phi(X)'\varepsilon F\|^2\|H\|_\infty^2 \\
&= \frac{1}{T^2}O_p\left(\frac{1}{N^2}\right)O_p(NTJ)O_p(1) \\
&= O_p\left(\frac{J}{NT}\right)
\end{aligned}
$$

By Lemma C.1, and the by-product result of Section B.1,

$$
\begin{aligned}
\|D_3\|^2 &= \frac{1}{T^2}\|(\Phi(X)'\Phi(X))^{-1}\Phi(X)'\varepsilon(\widehat{F} - FH)\|^2 \\
&\leq \frac{1}{T}\|(\Phi(X)'\Phi(X))^{-1}\|_\infty^2\|\Phi(X)'\varepsilon\|^2\frac{1}{T}\|\widehat{F} - FH)\|^2 \\
&= \frac{1}{T}O_p\left(\frac{1}{N^2}\right)O_p(NTJ)O_p\left(\frac{1}{N} + \frac{1}{J^{-\kappa}}\right) \\
&= O_p\left(\frac{J}{N^2} + \frac{J}{NJ^\kappa}\right).
\end{aligned}
$$

By Lemma B.4,

$$
\begin{aligned}
\|D_4\|^2 &= \frac{1}{T^2}\|BF'(\widehat{F} - FH)\|^2 \\
&\leq \|B\|^2\frac{1}{T^2}\|F'(\widehat{F} - FH)\|^2 \\
&= O_p(J)O_p\left(\frac{1}{N^2} + \frac{1}{NT} + \frac{1}{J^\kappa} + \frac{\rho_N}{N}\right) \\
&= O_p\left(\frac{J}{N^2} + \frac{J}{NT} + \frac{J}{J^\kappa} + \frac{J\rho_N}{N}\right).
\end{aligned}
$$

By Lemma C.2,

$$
\begin{aligned}
\|D_5\|^2 &= \frac{1}{T^2}\|(\Phi(X)'\Phi(X))^{-1}\Phi(X)'\Gamma F'\widehat{F}\|^2 \\
&\leq \frac{1}{T^2}\|(\Phi(X)'\Phi(X))^{-1}\|_\infty^2\|\Phi(X)'\Gamma\|^2\|F\|^2\|\widehat{F}\|^2 \\
&= O_p\left(\frac{1}{N^2}\right)O_p(JN\rho_N) \\
&= O_p\left(\frac{J\rho_N}{N}\right).
\end{aligned}
$$

Finally,

$$
\begin{aligned}
\|\widehat{B} - BH\|^2 &\leq \sum_{n=1}^{5}\|D_n\|^2 \\
&= O_p\left(\frac{J}{N^2} + \frac{J}{NT} + \frac{J}{J^\kappa} + \frac{J\rho_N}{N}\right).
\end{aligned}
$$

2. By definition, $G(X) = \Phi(X)B + R(X)$, and $\widehat{G}(X) = \Phi(X)\widehat{B}$. Then

$$\begin{aligned}
\frac{1}{N}\|\widehat{G}(X) - G(X)H\|^2 &= \frac{1}{N}\|\Phi(X)\widehat{B} - \Phi(X)BH + R(X)H\|^2 \\
&\leq \frac{2}{N}\|\Phi(X)\|_\infty^2\|\widehat{B} - BH\|^2 + \frac{2}{N}\|R(X)\|^2\|H\|_\infty^2 \\
&= \frac{1}{N}O_p(N)O_p\left(\frac{J}{N^2} + \frac{J}{NT} + \frac{J}{J^\kappa} + \frac{J\rho_N}{N}\right) + \frac{1}{N}O_p(NJ^{-\kappa})O_p(1) \\
&= O_p\left(\frac{J}{N^2} + \frac{J}{NT} + \frac{J}{J^\kappa} + \frac{J\rho_N}{N}\right).
\end{aligned}$$

3. Recall that $\widehat{\Gamma} = \widehat{\Lambda} - \widehat{G}(X) = \frac{1}{T}(I_K - P)Y\widehat{F}$. Then

$$\widehat{\Gamma} = \Gamma H + \sum_{n=1}^{6} E_n,$$

where

$$\begin{aligned}
E_1 &= \frac{1}{T}(I_K - P)\Gamma F'(\widehat{F} - FH), & E_2 &= \frac{1}{T}\varepsilon(\widehat{F} - FH), \\
E_3 &= -P\Gamma H, & E_4 &= (I_K - P)R(X)\left[\frac{1}{T}F'(\widehat{F} - FH) + H\right], \\
E_5 &= -\frac{1}{T}P\varepsilon(\widehat{F} - FH), & E_6 &= \frac{1}{T}(I_K - P)\varepsilon FH.
\end{aligned}$$

Finally, by Lemma B.8,

$$\begin{aligned}
\frac{1}{N}\|\widehat{\Gamma} - \Gamma H\|^2 &\leq \frac{1}{N}\sum_{n=1}^{6}\|E_n\|^2 \\
&= O_p\left(\frac{1}{J^\kappa} + \frac{J\rho_N}{N} + \frac{J}{N^2} + \frac{1}{T}\right),
\end{aligned}$$

where

$$\begin{aligned}
E_1 &= \frac{1}{T}(I_K - P)\Gamma F'(\widehat{F} - FH), & E_2 &= \frac{1}{T}\varepsilon(\widehat{F} - FH), \\
E_3 &= -P\Gamma H, & E_4 &= (I_K - P)R(X)\left[\frac{1}{T}F'(\widehat{F} - FH) + H\right], \\
E_5 &= -\frac{1}{T}P\varepsilon(\widehat{F} - FH), & E_6 &= \frac{1}{T}(I_K - P)\varepsilon FH.
\end{aligned}$$

□

Lemma B.8.

This lemma is for the third part of Lemma B.7 only.

1. $\|\frac{1}{T}(I_K - P)\Gamma F'(\widehat{F} - FH)\|^2 = O_p\left(\rho_N\left[\frac{1}{N} + \frac{1}{T} + \frac{N}{J^\kappa} + \rho_N\right]\right)$.

2. $\|\frac{1}{T}\varepsilon(\widehat{F} - FH)\|^2 = O_p\left(\frac{N}{J^\kappa T} + \frac{1}{T} + \frac{1}{N}\right)$. *3.* $\|P\Gamma H\|^2 = O_p(J\rho_N)$.

4. $\|(I_K - P)R(X)[\frac{1}{T}F'(\widehat{F} - FH) + H]\|^2 = O_p\left(\frac{N}{J^\kappa}\right)$. *5.* $\|\frac{1}{T}P\varepsilon(\widehat{F} - FH)\|^2 = O_p\left(\frac{J}{N} + \frac{J}{J^\kappa}\right)$.

6. $\|\frac{1}{T}(I_K - P)\varepsilon FH\|^2 = O_p\left(\frac{N}{T}\right)$.

Proof of Lemma B.8.

1. By Lemma B.4,

$$\begin{aligned}\|\frac{1}{T}(I_K - P)\Gamma F'(\widehat{F} - FH)\|^2 &\le \|I_K - P\|_\infty^2 \|\Gamma\| \|\frac{1}{T}F'(\widehat{F} - FH)\|^2 \\ &= O_p(1)O_p(N\rho_N)O_p\left(\frac{1}{N^2} + \frac{1}{NT} + \frac{1}{J^\kappa} + \frac{\rho_N}{N}\right) \\ &= O_p\left(\rho_N\left[\frac{1}{N} + \frac{1}{T} + \frac{N}{J^\kappa} + \rho_N\right]\right).\end{aligned}$$

2. Recall that $\widehat{F} - FH = \sum_{m=1}^{15} W_m V^{-1}$. Then

$$\begin{aligned}\|\frac{1}{T}\varepsilon(\widehat{F} - FH)\|^2 &\le \sum_{m=1}^{15} \|\frac{1}{T}\varepsilon W_m\|^2 \|V^{-1}\|_\infty^2 \\ &= O_p\left(\frac{N}{J^\kappa T} + \frac{1}{T} + \frac{1}{N}\right).\end{aligned}$$

This part is imcomplete.

(a) Assumption 3.4 *(iii)* : third one is need here.

(b) Fan et al. (2015) lemma C7-9 must be added here.

(c) lemma C7 and 9 = convergence for all εW_m for all m.

(d) lemma C8 is needed for lemma c9. It can go to tech lemma.

3. By Lemma B.1 and C.2,

$$\begin{aligned}\|P\Gamma H\|^2 &\le \|\Phi(X)(\Phi(X)'\Phi(X))^{-1}\|_\infty^2 \|\Phi(X)'\Gamma\|^2 \|H\|_\infty^2 \\ &= O_p\left(\frac{1}{N}\right)O_p(JN\rho_N)O_p(1) \\ &= O_p(J\rho_N).\end{aligned}$$

4. By Lemma B.4,

$$\begin{aligned}\|(I_K - P)R(X)[\frac{1}{T}F'(\widehat{F} - FH) + H]\|^2 &\le \|(I_K - P)\|_\infty^2 \|R(X)\|^2 \left(\frac{1}{T^2}\|F'(\widehat{F} - FH)\|^2 + \|H\|^2\right) \\ &= O_p(1)O_p(NJ^{-\kappa})\left(O_p\left(\frac{1}{N^2} + \frac{1}{NT} + \frac{1}{J^\kappa} + \frac{\rho_N}{N} +\right) + O_p(1)\right) \\ &= O_p\left(\frac{N}{J^\kappa}\right).\end{aligned}$$

5. Using the result in Section B.1 and Lemma C.1,

$$\begin{aligned}\|\frac{1}{T}P\varepsilon(\widehat{F} - FH)\|^2 &\le \|\Phi(X)(\Phi(X)'\Phi(X))^{-1}\|_\infty^2 \|\Phi(X)'\varepsilon\|^2 \frac{1}{T^2}\|\widehat{F} - FH\|^2 \\ &= O_p\left(\frac{1}{N}\right)O_p(NTJ)\frac{1}{T}O_p\left(\frac{1}{N} + \frac{1}{J^\kappa}\right) \\ &= O_p\left(\frac{J}{N} + \frac{J}{J^\kappa}\right).\end{aligned}$$

6. Following Lemma B.1,

$$
\begin{aligned}
\|\frac{1}{T}(I_K - P)\varepsilon F H\|^2 &\leq \frac{1}{T^2}\|I_K - P\|_\infty \|\varepsilon F\|^2 \|H\|_\infty^2 \\
&= \frac{1}{T^2} O_p(1) O_p(NT) O_p(1) \\
&= O_p\left(\frac{N}{T}\right).
\end{aligned}
$$

□

B.4 Proof of Theorem 3.2

Proof. Assume that $K < \ell_{max}$ where $\ell_{max} = JH/2$. Let us reiterate the eigenvalue ratio:

$$ER(\ell) = \frac{\psi_\ell\left(\frac{\widehat{Y}^*\widehat{Y}}{NT}\right)}{\psi_{\ell+1}\left(\frac{\widehat{Y}^*\widehat{Y}}{NT}\right)}, \quad \ell = 1, \ldots, \ell_{max},$$

and the estimator for the number of factors is

$$\widehat{K} = \underset{1 \le \ell < \ell_{max}}{\text{argmax}} \frac{\psi_\ell\left(\frac{\widehat{Y}^*\widehat{Y}}{NT}\right)}{\psi_{\ell+1}\left(\frac{\widehat{Y}^*\widehat{Y}}{NT}\right)},$$

where the choice of ℓ_{max} is explained in Lemma B.11. Roughly speaking, the choice of ℓ_{max} is to ensure that the lower bound of $\psi_\ell\left(\widehat{Y}^*\widehat{Y}\right)$ is strictly greater than zero uniformly in ℓ when $\ell < \ell_{max}$, thus, $ER(\ell)$ is well defined.

We first show that $ER(\ell) = O_p(1)$ when $\ell \neq K$. If $\ell \le K - 1$, it follows from Lemma B.9 that $ER(\ell) = O_p(1)$. Now consider the case $K + 1 \le \ell \le JH/2 - 1$. From Lemma B.11, we have

$$0 < \delta_1/N + o_p(1/N) \le \psi_{K+m}\left(\frac{\widehat{Y}^*\widehat{Y}}{NT}\right) \le \delta_2/N + o_p(1/N) < \infty, \tag{20}$$

uniformly for $m = 1, \ldots, JH - 2K$, thus, $ER(\ell) = O_p(1)$.

We now show the case for $\ell = K$. It follows from Lemma B.9 that $\psi_K\left(\frac{\widehat{Y}^*\widehat{Y}}{NT}\right) = O_p(1)$ and the relation (20), there exist some $\bar{\delta} > 0$,

$$ER(K) \ge \bar{\delta}N + o_p(1).$$

Define $\mathcal{L} = \{\ell : \ell \neq K \text{ and } \ell < JH/2\}$. Then for $\eta > 0$, there exist $C_\eta > 0$ such that

$$P\left(\max_{\ell \in \mathcal{L}} ER(\ell) > C_\eta\right) < \eta.$$

This is because $ER(\ell) = O_p(1)$ for $\ell \neq K$. Hence, for all $\eta > 0$,

$$\begin{aligned}
P(\widehat{K} \neq K) &\le P\left(\max_{\ell \in \mathcal{L}} ER(\ell) \ge ER(K)\right) \\
&\le P\left(\max_{\ell \in \mathcal{L}} ER(\ell) > C_\eta\right) + P\left(ER(K) \le \max_{\ell \in \mathcal{L}} ER(\ell) \le C_\eta\right) \\
&= P\left(\max_{\ell \in \mathcal{L}} ER(\ell) > C_\eta\right) + P\Big(ER(K) \le C_\eta\Big) \\
&< \eta,
\end{aligned}$$

as $ER(K)$ is unbounded. Hence, $P(\widehat{K} = K) \to 1$. □

B.5 Lemmas for Theorem 3.2

Lemma B.9.

Suppose $J = o(N)$. For $\ell = 1, \ldots, K-1$, as $N \to \infty$,

$$ER(\ell) = \frac{\psi_\ell\left(\frac{\widehat{Y}^*\widehat{Y}}{NT}\right)}{\psi_{\ell+1}\left(\frac{\widehat{Y}^*\widehat{Y}}{NT}\right)} = O_p(1). \tag{21}$$

Proof of Lemma B.9.

By Lemma B.1, $\|V\|_\infty$ and $\|V^{-1}\|_\infty$ are both $O_p(1)$. Hence, (21) is satisfied for all $\ell \leq K-1$. $\square$

Lemma B.10.

Define

$$\Theta = \frac{1}{NT}\Phi(X)' A_N^{1/2} U U^* A_N^{1/2}\Phi(X) - \frac{1}{N}\Phi(X)' A_N^{1/2}\Sigma_u A_N^{1/2}\Phi(X),$$

where $\Sigma_u = E[u_t \otimes u_t]$. Suppose $J = o(T)$. Under the assumptions 3.4 and 3.7, as $N, T \to \infty$,

$$\|\Theta\|_\infty = o_p(1).$$

Note that we have

$$\frac{1}{T}UU^* = \frac{1}{T}\sum_t^T (u_t \otimes u_t), \quad \frac{1}{T}E[UU^*] = E[u_t \otimes u_t],$$

since u_t is iid over t.

Proof of Lemma B.10.

By the assumption 3.7, for $M_2 > 0$,

$$E\left[\exp\{\tau\langle v, u_t\rangle\}\right] \leq \exp\{\tau^2 M_2 \|v\|^2\},$$

for all $\tau > 0$, $v \in \mathcal{H}_N$. Then conditional on X, for any $w \in \mathcal{H}_{JH}$, $\|w\| = 1$

$$E\left[\exp\left\{\tau\langle A_N^{1/2}\Phi(X)w/\sqrt{N}, u_t\rangle\right\}\big|X\right] \leq \exp\left\{\tau^2 M_2 \|A_N^{1/2}\Phi(X)/\sqrt{N}\|^2\right\} \tag{22}$$

since $A_N^{1/2}\Phi(X)w \in \mathcal{H}_N$. The right hand side of (22) is bounded because

$$\|A_N^{1/2}\Phi(X)/\sqrt{N}\|^2 = \|A_N^{1/2}\Phi(X)w/\sqrt{N}\|^2 \leq \|A_N^{1/2}\Phi(X)/\sqrt{N}\|_\infty^2 \|w\| = a_{max}\, d_{max},$$

where a_{max} denotes the largest eigenvalue of A_N. Therefore, $q_t := \Phi(X)' A_N^{1/2} u_t/\sqrt{N}$ is also a sub-Gaussian vector conditional on X, that is, for all w

$$E\left[\exp\{\tau\langle w, q_t\rangle\}\right] \leq \exp\left\{\tau^2 M_2 \|A_N^{1/2}\Phi(X)/\sqrt{N}\|^2\right\}.$$

Then by Lemma B.12, there exists $C_1, C_2 > 0$ such that

$$P\left(\|\Theta\|_\infty \leq C_1\sqrt{\frac{JH}{T}} + \sqrt{\frac{J}{T}}\bigg|X\right) \geq 1 - 2exp(-C_2 J),$$

where C_1, C_2 does not depend on X. Specifically, C_1, C_2 depends on the sub-Gaussian norm of q_t :

$$\|q_t\|_{\psi_2} = \sup_{\|w\|=1} \sup_{m\geq 1} \frac{E[|\langle w, q_t\rangle|^m]^{1/m}}{\sqrt{m}} < (a_{max} d_{max} M_2)^{1/2}.$$

Hence, unconditional on X,

$$\|\Theta\|_\infty = O_p\left(\sqrt{\frac{J}{T}}\right) = o_p(1),$$

given that $J = o(T)$. □

Lemma B.11.

Suppose $J = o(T)$, and $K \leq JH/2$. Under the assumptions of 3.1-3.6, and 3.7, for $m = 1, 2, \ldots, JH - 2K$,

$$\delta_1 + o_p(1) \leq \psi_{K+m}\left(\widehat{Y}^*\widehat{Y}/T\right) \leq \delta_2 + o_p(1) \tag{23}$$

where δ_1, δ_2 are positive constants (uniformly for $m \leq JH - 2K$).

Proof of Lemma B.11.
By assumption 3.7, there exist two positive constants π_1, π_2 such that uniformly in N, T,

$$\pi_1 \leq \psi_N(A_N), \psi_T(Z_T) \quad \text{and} \quad \psi_1(A_N), \psi_1(Z_T) \leq \pi_2.$$

From Lemma B.14, we have

$$\begin{aligned}
\psi_1(\varepsilon^* P\varepsilon/T) &= \psi_1((A_N^{1/2} U Z_T^{1/2})^* P A_N^{1/2} U Z_T^{1/2}/T) = \psi_1(Z_T^{1/2} U^* A_N^{1/2} P A_N^{1/2} U Z_T^{1/2}/T) \\
&= \psi_1(Z_T U^* A_N^{1/2} P A_N^{1/2} U/T) \\
&\leq \psi_1(Z_T)\psi_1(U^* A_N^{1/2} P A_N^{1/2} U/T) \\
&= \psi_1(Z_T)\psi_1(U^* A_N^{1/2} \Phi(X)(\Phi(X)'\Phi(X))^{-1}\Phi(X)' A_N^{1/2} U/T) \\
&= \psi_1(Z_T)\psi_1(\Phi(X)' A_N^{1/2} U U^* A_N^{1/2} \Phi(X)(\Phi(X)'\Phi(X))^{-1}/T) \\
&= \psi_1(Z_T)\psi_1(\Phi(X)' A_N^{1/2} U U^* A_N^{1/2} \Phi(X)/(NT)\,(\Phi(X)'\Phi(X)/N)^{-1}) \\
&\leq \psi_1(Z_T)\psi_1(\Phi(X)' A_N^{1/2} U U^* A_N^{1/2} \Phi(X)/(NT))\psi_1(\Phi(X)'\Phi(X)/N)^{-1}) \\
&\leq \pi_2 \psi_1(\Phi(X)' A_N^{1/2} U U^* A_N^{1/2} \Phi(X)/(NT)) d_{min}^{-1} \\
&= \pi_2 d_{min}^{-1} \psi_1(\Phi(X)' A_N^{1/2} U U^* A_N^{1/2} \Phi(X)/(NT)).
\end{aligned}$$

From Lemma B.14, we obtain

$$\begin{aligned}
\psi_{JH}(\varepsilon^* P\varepsilon/T) &= \psi_{JH}(Z_T U^* A_N^{1/2} P A_N^{1/2} U/T) \\
&\geq \psi_T(Z_T)\psi_{JH}(U^* A_N^{1/2} P A_N^{1/2} U/T) \\
&= \psi_T(Z_T)\psi_{JH}(\Phi(X)' A_N^{1/2} U U^* A_N^{1/2} \Phi(X)/(NT)\,(\Phi(X)'\Phi(X)/N)^{-1}) \\
&\geq \psi_T(Z_T)\psi_{JH}(\Phi(X)' A_N^{1/2} U U^* A_N^{1/2} \Phi(X)/(NT))\psi_{JH}(\Phi(X)'\Phi(X)/N)^{-1}) \\
&\geq \pi_1 \psi_{JH}(\Phi(X)' A_N^{1/2} U U^* A_N^{1/2} \Phi(X)/(NT)) d_{max}^{-1} \\
&= \pi_1 d_{max}^{-1} \psi_{JH}(\Phi(X)' A_N^{1/2} U U^* A_N^{1/2} \Phi(X)/(NT)).
\end{aligned}$$

Therefore, the eigenvalues of $\varepsilon^* P\varepsilon/T$ are bounded by those of $\Phi(X)' A_N^{1/2} UU^* A_N^{1/2}\Phi(X)/(NT)$.

Let $\widetilde{F} = F + \varepsilon^* P\Lambda(\Lambda^* P\Lambda)^{-1}$. Then

$$\begin{aligned} Y = \Lambda F' + \varepsilon &= \Lambda(\widetilde{F} - \varepsilon^* P\Lambda(\Lambda^* P\Lambda)^{-1})' + \varepsilon \\ &= \Lambda\widetilde{F}' - \Lambda(\Lambda^* P\Lambda)^{-1}\Lambda^* P\varepsilon + \varepsilon \\ &= \Lambda\widetilde{F}' + [I_N - \Lambda(\Lambda^* P\Lambda)^{-1}\Lambda^* P]\varepsilon \end{aligned}$$

We denote $\widetilde{M} = I_N - \Lambda(\Lambda^* P\Lambda)^{-1}\Lambda^* P$, and $\widetilde{M} \in \mathbb{L}(\mathcal{H}_N, \mathcal{H}_N)$. Then $\widetilde{M}^* = I_N - P\Lambda(\Lambda^* P\Lambda)^{-1}\Lambda^*$. Note that $\widetilde{M} P\Lambda = \Lambda^* P\widetilde{M} = 0$. Moreover,

$$P\widetilde{M} = P - P\Lambda(\Lambda^* P\Lambda)^{-1}\Lambda^* P = (I_N - P\Lambda(\Lambda^* P\Lambda)^{-1}\Lambda^*)P = \widetilde{M}^* P$$

Hence, we have

$$\begin{aligned} Y^* PY &= (\Lambda\widetilde{F}' + \widetilde{M}\varepsilon)^* P(\Lambda\widetilde{F}' + \widetilde{M}\varepsilon) \\ &= \widetilde{F}\Lambda^* P\Lambda\widetilde{F}' + \varepsilon^*\widetilde{M}^* P\Lambda\widetilde{F}' + \widetilde{F}\Lambda^* P\widetilde{M}\varepsilon + \varepsilon^*\widetilde{M}^* P\widetilde{M}\varepsilon \\ &= \widetilde{F}\Lambda^* P\Lambda\widetilde{F}' + \varepsilon^*\widetilde{M}^* P\widetilde{M}\varepsilon \\ &= \widetilde{F}\Lambda^* P\Lambda\widetilde{F}' + \varepsilon^* P\widetilde{M}\widetilde{M}^* P\varepsilon \\ &= \widetilde{F}\Lambda^* P\Lambda\widetilde{F}' + \varepsilon^* P(P\widetilde{M}\widetilde{M}^* P)P\varepsilon \\ &= \widetilde{F}\Lambda^* P\Lambda\widetilde{F}' + \varepsilon^* P(PP - P\Lambda(\Lambda^* P\Lambda)^{-1}\Lambda^* P)P\varepsilon \\ &= \widetilde{F}\Lambda^* P\Lambda\widetilde{F}' + \varepsilon^* P(I_N - \Lambda(\Lambda^* P\Lambda)^{-1}\Lambda^*)P\varepsilon. \end{aligned}$$

Then, by Lemma B.13

$$\begin{aligned} \psi_{K+m}(Y^* PY) &= \psi_{K+m}(\widetilde{F}\Lambda^* P\Lambda\widetilde{F}' + \varepsilon^* P(I_N - \Lambda(\Lambda^* P\Lambda)^{-1}\Lambda^*)P\varepsilon) \\ &\leq \psi_{K+1}(\widetilde{F}\Lambda^* P\Lambda\widetilde{F}') + \psi_m(\varepsilon^* P(I_N - \Lambda(\Lambda^* P\Lambda)^{-1}\Lambda^*)P\varepsilon) \\ &= \psi_m(\varepsilon^* P(I_N - \Lambda(\Lambda^* P\Lambda)^{-1}\Lambda^*)P\varepsilon) \\ &= \psi_m(\varepsilon^* P\varepsilon - \varepsilon^* P\Lambda(\Lambda^* P\Lambda)^{-1}\Lambda^* P\varepsilon) \\ &\leq \psi_m(\varepsilon^* P\varepsilon), \end{aligned}$$

where the third equality used the fact that $rank(\Lambda^* P\Lambda) = K$; the last inequality is by Lemma B.14. To find the lower bound,

$$\begin{aligned} \psi_{K+m}(Y^* PY) &= \psi_{K+m}(\widetilde{F}\Lambda^* P\Lambda\widetilde{F}' + \varepsilon^* P(I_N - \Lambda(\Lambda^* P\Lambda)^{-1}\Lambda^*)P\varepsilon) \\ &\geq \psi_{K+m}(\varepsilon^* P(I_N - \Lambda(\Lambda^* P\Lambda)^{-1}\Lambda^*)P\varepsilon) \\ &= \psi_{K+m}(\varepsilon^* P\varepsilon - \varepsilon^* P\Lambda(\Lambda^* P\Lambda)^{-1}\Lambda^* P\varepsilon) + \psi_{K+1}(\varepsilon^* P\Lambda(\Lambda^* P\Lambda)^{-1}\Lambda^* P\varepsilon)) \\ &\geq \psi_{2K+m}(\varepsilon^* P\varepsilon), \end{aligned}$$

where the second and the last lines follow by Lemma B.14, and the third line holds by $\psi_{K+1}(\varepsilon^* P\Lambda(\Lambda^* P\Lambda)^{-1}\Lambda^* P\varepsilon)) = 0$.

Hence, for $1 \geq m \leq JH - 2K$, the bounds for $\psi_{K+m}(Y^*PY/T)$ are

$$\pi_1 d_{max}^{-1} \psi_{JH}(\Phi(X)' A_N^{1/2} UU^* A_N^{1/2} \Phi(X)/(NT)) \leq \psi_{JH}(\varepsilon^* P \varepsilon/T) \\ \leq \psi_{K+m}(Y^*PY/T) \leq \psi_1(\varepsilon^* P\varepsilon/T) \leq \pi_2 d_{min}^{-1} \psi_1(\Phi(X)' A_N^{1/2} UU^* A_N^{1/2} \Phi(X)/(NT)),$$

which can be simplied to

$$\pi_1 d_{max}^{-1} \psi_{JH}(\Phi(X)' A_N^{1/2} UU^* A_N^{1/2} \Phi(X)/(NT)) \leq \psi_{K+m}(Y^*PY/T) \\ \leq \pi_2 d_{min}^{-1} \psi_1(\Phi(X)' A_N^{1/2} UU^* A_N^{1/2} \Phi(X)/(NT)).$$

By Lemma B.10,

$$\| \frac{1}{NT} \Phi(X)' A_N^{1/2} UU^* A_N^{1/2} \Phi(X) - \frac{1}{N} \Phi(X)' A_N^{1/2} \Sigma_u A_N^{1/2} \Phi(X) \|_\infty = o_p(1),$$

where the eigenvalues of second term is bounded away from zero and infinity under the assumption 3.7. Hence, there exist positive constants δ_1, δ_2 such that

$$\delta_1 + o_p(1) \leq \pi_1 d_{max}^{-1} \psi_{JH}(\Phi(X)' A_N^{1/2} UU^* A_N^{1/2} \Phi(X)/(NT)) \leq \psi_{K+m}(Y^*PY/T) \\ \leq \pi_2 d_{min}^{-1} \psi_1(\Phi(X)' A_N^{1/2} UU^* A_N^{1/2} \Phi(X)/(NT)) \leq \delta_2 + o_p(1).$$

As a result, we prove that

$$\delta_1 + o_p(1) \leq \psi_{K+m}\left(\widehat{Y}^* \widehat{Y}/T\right) \leq \delta_2 + o_p(1).$$

□

For the remaining, we state a few technical Lemmas necessary for this section.

Lemma B.12.

Let A be a $N \times T$ random matrix with column vectors a_t that are independent and sub-Gaussian, and denote $\Sigma = E(a_t \otimes a_t)$. Then for $\tau \geq 0$, with probability at least $1 - 2exp(-C_2\tau^2)$ one has

$$\| \frac{1}{T} AA^* - \Sigma \|_\infty \leq \max\{\delta, \delta^2\}, \quad \delta = C_1 \sqrt{\frac{N}{T}} + \frac{\tau}{\sqrt{T}},$$

where $C_1, C_2 > 0$ depend only on the sub-Gaussian norm $\max_{t \leq T} \|a_t\|_{\psi_2}$ of the column vectors. (See Theorem 5.39 of Vershynin 2010).

Lemma B.13.

Let A and B be $T \times T$ symmetric matrices. For $m, n \geq 1$ and $m + n - 1 \leq T$,

$$\psi_{m+n-1}(A + B) \leq \psi_m(A) + \psi_n(B).$$

This is Weyl inequality.

Lemma B.14.

Suppose A and B are $T \times T$ positive definite and positive semidefinite matrices, then

$$\psi_m(AB) \leq \psi_n(A)\psi_\ell(B), \quad \text{for } n + \ell \leq m + 1$$

$$\psi_m(AB) \geq \psi_n(A)\psi_\ell(B), \quad \text{for } n + \ell \leq m + T.$$

In addition, if both A and B are positive definite,

$$\psi_m(A) \leq \psi_m(A + B), \quad \text{for } m1, 2, \ldots, T.$$

C Technical Lemmas

Lemma C.1.

1. $\sup_{r\in[0,1],k\leq K,i\leq N}\max_{t\leq T}\sum_{s=1}^{T}|cov(f_{tk}u_{it}(r),f_{sk}u_{is}(r))|=O(1)$.

2. $\|\varepsilon F\|^2=O_p(NT)$. *3.* $\|\Phi(X)'\varepsilon\|^2=O_p(NTJ)$. *4.* $\|\Phi(X)'\varepsilon F\|^2=O_p(NTJ)$.

5. $\|P\varepsilon\|^2=O_p(TJ)$.

Proof of Lemma C.1.

1. By Davydov's inequality, for any $t,s\leq T$, there exists $\eta>0$

$$|cov(f_{tk}u_{it}(r),f_{sk}u_{is}(r))|\leq\eta[\alpha(|t-s|)]^{1/2},$$

for all $r\in[0,1],k\leq K,i\leq N,t\leq T$, where $\alpha(t)$ is the α-mixing coefficient of $\{f_t,u_t\}$. By the assumption 3.4, $\sum_{t=1}^{T}[\alpha(t)]^{1/2}<\infty$. Hence,

$$\sup_{r\in[0,1],k\leq K,i\leq N}\max_{t\leq T}\sum_{s=1}^{T}|cov(f_{tk}u_{it}(r),f_{sk}u_{is}(r))|\leq\max_{t\leq T}\sum_{s=1}^{T}\eta[\alpha(|t-s|)]^{1/2}<\infty.$$

2. Note that $\|F\|^2,\|\widehat{F}\|^2=O_p(T)$ by the assumption 3.2. Then

$$\begin{aligned}
E\|\varepsilon F\|^2&=E\sum_{k=1}^{K}\sum_{i=1}^{N}\langle\sum_{t=1}^{T}\varepsilon_{it}f_{tk},\sum_{t=1}^{T}\varepsilon_{it}f_{tk}\rangle\\
&=\sum_{k=1}^{K}\sum_{i=1}^{N}E\left[\left(\int_0^1\sum_{t=1}^{T}\varepsilon_{it}(r)f_{tk}dr\right)^2\right]\\
&=\sum_{k=1}^{K}\sum_{i=1}^{N}\int\sum_{t=1}^{T}\sum_{s=1}^{T}cov(f_{tk}u_{it}(r),f_{sk}u_{is}(r))dr\\
&\leq\sum_{k=1}^{K}\sum_{i=1}^{N}\int\sum_{t=1}^{T}\left(\sup_{r\in[0,1],k\leq K,i\leq N}\max_{t\leq T}\sum_{s=1}^{T}|cov(f_{tk}u_{it}(r),f_{sk}u_{is})(r)|\right)dr\\
&=O(NT).
\end{aligned}$$

3.

$$\begin{aligned}
E\|\Phi(X)'\varepsilon\|^2&=E\sum_{j=1}^{J}\sum_{h=1}^{H}\sum_{i=1}^{N}\sum_{q=1}^{N}\sum_{t=1}^{T}\left[\phi_j(X_{ih})\phi_j(X_{qh})\int\varepsilon_{it}(r)\varepsilon_{qt}(r)\,dr\right]\\
&=\sum_{j=1}^{J}\sum_{h=1}^{H}\sum_{i=1}^{N}\left[\sum_{q=1}^{N}E[\phi_j(X_{ih})\phi_j(X_{qh})]\sum_{t=1}^{T}\int E[\varepsilon_{it}(r)\varepsilon_{qt}(r)]dr\right]\\
&\leq JHN\times\left(\max_{j\leq J,h\leq H,i\leq N,q\leq N,}E[\phi_j(X_{ih})\phi_j(X_{qh})]\right)\left(\sum_{t=1}^{T}\int\sup_{r\in[0,1],i\leq N}\sum_{q=1}^{N}|E[\varepsilon_{it}(r)\varepsilon_{qt}(r)]|dr\right)\\
&\leq JHN\times\left(\max_{j\leq J,h\leq H,i\leq N}E[\phi_j^2(X_{ih})]\right)\left(\sum_{t=1}^{T}\int\sup_{r\in[0,1],i\leq N}\sum_{q=1}^{N}|E[\varepsilon_{it}(r)\varepsilon_{qt}(r)]|dr\right)\\
&=O(NTJ),
\end{aligned}$$

where the fourth equality holds by the Cauchy-Schwarz inequality, and the assumptions 3.3 and 3.4 implies that $\int\max_{i\leq N}\sum_{q=1}^{N}|E[\varepsilon_{it}(r)\varepsilon_{qt}(r)]|dr<M_2<\infty$.

4.

$$E\|\Phi(X)'\varepsilon F\|^2 = \sum_{k=1}^{K}\sum_{j=1}^{J}\sum_{h=1}^{H}\sum_{i=1}^{N}\sum_{q=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T} E[\phi_j(X_{ih})\phi_j(X_{qh})]E[f_{tk}f_{sk}] \int E[\varepsilon_{it}(r)\varepsilon_{qs}(r)]\, dr$$

$$\leq O(NTJ)\frac{1}{NT}\int \sum_{i=1}^{N}\sum_{q=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T} |E[\varepsilon_{it}(r)\varepsilon_{qs}(r)]|\, dr$$

$$= O(NTJ),$$

where the last equality is due to the assumptions 3.4.

5. By the assumption 3.3,

$$\|P\varepsilon\|^2 = \|\Phi(X)\left(\Phi(X)'\Phi(X)\right)^{-1}\Phi(X)'\varepsilon\|^2$$
$$\leq \|\Phi(X)\|_2^2\|\left(\Phi(X)'\Phi(X)\right)^{-1}\|_2^2\|\Phi(X)'\varepsilon\|^2.$$
$$= O_p(N)O_p\left(\frac{1}{N^2}\right)O_p(NTJ)$$
$$= O_p(TJ).$$

□

Lemma C.2.
1. $\|\varepsilon^*\Phi(X)B\|^2 = O_p(NT)$. *2.* $\|B^*\Phi(X)'\varepsilon F\|^2 = O_p(NT)$.
3. $\|\Phi(X)'\Gamma\|^2 = O_p(JN\rho_N)$. *4.* $\|B^*\Phi(X)'\Gamma\|^2 = O_p(N\rho_N)$. *5.* $\|P\Gamma\|^2 = O_p(J\rho_N)$.
6. $\|\Gamma\|^2 = O_p(N\rho_N)$.

Proof of Lemma C.2.

1. By the definition of $G(X)$,

$$E\|\varepsilon^*\Phi(X)B\|^2 = E\|\varepsilon^*G(X) - \varepsilon^*R(X)\|^2$$
$$\leq E\|\varepsilon^*G(X)\|^2 + E\|\varepsilon^*R(X)\|^2.$$

Hence, our task is to bound the two terms on the right hand side of the inequality.

$$E\|\varepsilon^*G(X)\|^2 = \sum_{k=1}^{K}\sum_{i=1}^{N}\sum_{q=1}^{N}\sum_{t=1}^{T} E\left[\int \varepsilon_{it}(r)g_k(X_i,r)\,dr \int \varepsilon_{qt}(w)g_k(X_q,w)\,dw\right]$$
$$= \sum_{k=1}^{K}\sum_{i=1}^{N}\sum_{q=1}^{N}\sum_{t=1}^{T}\int E\left[\varepsilon_{it}(r)\varepsilon_{qt}(w)\right]E\left[g_k(X_i,r)g_k(X_q,w)\right]drdw$$
$$\leq NK \times \sup_{r\in[0,1],k\leq K,i\leq N} E[g_k^2(X_i,r)]\sum_{t=1}^{T}\max_{i\leq N}\sum_{q=1}^{N}\int_0^1 |E[\varepsilon_{it}(r)\varepsilon_{qs}(w)]|drdw,$$
$$= O(NT).$$

where the last equality holds by the assumption 3.4 and 3.5, and Cauchy-Schwarz inequality.

Define $R_{ik} = \sum_{h=1}^{H} R_{kh}(X_{ih})$, which is the (i,k) element of $R(X)$.

$$
\begin{aligned}
E\|\varepsilon^* R(X)\|^2 &= \sum_{k=1}^{K}\sum_{t=1}^{T}\sum_{i=1}^{N}\sum_{q=1}^{N} E\left[\int \varepsilon_{it}(r)R_{ik}(r)\,dr \int \varepsilon_{qt}(w)R_{qk}(w)\,dw\right] \\
&= \int \sum_{k=1}^{K}\sum_{t=1}^{T}\sum_{i=1}^{N}\sum_{q=1}^{N} E\left[\varepsilon_{it}(r)\varepsilon_{qt}(w)\right] E\left[R_{ik}(r)R_{qk}(w)\right] drdw \\
&\leq NK \times \sup_{r\in[0,1],k\leq K,i\leq N} E[R_{ik}^2(r)] \sum_{t=1}^{T} \max_{i\leq N} \sum_{q=1}^{N} \int_0^1 |E[\varepsilon_{it}(r)\varepsilon_{qt}(w)]|drdw, \\
&= O(NTJ^{-\kappa}).
\end{aligned}
$$

where the inequality is due to the assumption 3.4 and 3.6, and Cauchy Schwarz inequality. Given the two results,

$$E\|\varepsilon^*\Phi(X)B\|^2 = O(NT).$$

2. The same procedure applies in this section as before. By the definition of $G(X)$,

$$
\begin{aligned}
E\|B^*\Phi(X)'\varepsilon F\|^2 &= E\|G(X)^*\varepsilon F - R(X)^*\varepsilon F\|^2 \\
&\leq E\|G(X)^*\varepsilon F\|^2 + E\|R(X)^*\varepsilon F\|^2.
\end{aligned}
$$

$$
\begin{aligned}
E\|G(X)^*\varepsilon F\|^2 &= \sum_{k=1}^{K}\sum_{\ell=1}^{K}\sum_{i=1}^{N}\sum_{q=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T} E\left[\int g_\ell(X_i,r)\varepsilon_{it}(r)f_{tk}\,dr \int g_\ell(X_q,w)\varepsilon_{qs}(w)f_{sk}\,dw\right] \\
&= \int \sum_{k=1}^{K}\sum_{\ell=1}^{K}\sum_{i=1}^{N}\sum_{q=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T} E[g_\ell(X_i,r)g_\ell(X_q,w)]E[f_{tk}f_{sk}]E[\varepsilon_{it}(r)\varepsilon_{qs}(w)]drdw \\
&\leq K^2 \times \sup_{r\in[0,1],k,\ell\leq K,i\leq N,t\leq T} |E[g_\ell^2(X_i,r)]E[f_{tk}^2]| \sum_{i=1}^{N}\sum_{q=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T} \int_0^1 |E[\varepsilon_{it}(r)\varepsilon_{qs}(w)]|drdw \\
&= O(NT),
\end{aligned}
$$

where the third inequality holds by the assumption 3.4, and Cauchy-Schwarz inequality.

$$
\begin{aligned}
E\|R(X)^*\varepsilon F\|^2 &= \sum_{k=1}^{K}\sum_{\ell=1}^{K}\sum_{i=1}^{N}\sum_{q=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T} E\left[\int R_{i\ell}(r)\varepsilon_{it}(r)f_{tk}\,dr \int R_{q\ell}(w)\varepsilon_{qs}(w)f_{sk}\,dw\right] \\
&= \sum_{k=1}^{K}\sum_{\ell=1}^{K}\sum_{i=1}^{N}\sum_{q=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T} \int E[R_{i\ell}(r)R_{q\ell}(w)]E[f_{tk}f_{sk}]E[\varepsilon_{it}(r)\varepsilon_{qs}(w)]drdw \\
&\leq K^2 \times \sup_{r\in[0,1],k,\ell\leq K,i\leq N,t\leq T} |E[R_{i\ell}^2(r)]E[f_{tk}^2]| \sum_{i=1}^{N}\sum_{q=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T} \int_0^1 |E[\varepsilon_{it}(r)\varepsilon_{qs}(w)]|drdw \\
&= O(NTJ^{-\kappa}).
\end{aligned}
$$

Hence,

$$E\|B^*\Phi(X)'\varepsilon F\|^2 = O\left(NT\right).$$

3. By the assumption 3.5,

$$
\begin{aligned}
E\|\Phi(X)'\Gamma\|^2 &= \sum_{k=1}^{K}\sum_{h=1}^{H}\sum_{j=1}^{J}\sum_{i=1}^{N}\sum_{q=1}^{N} E\left[\int \phi_j(X_{ih})\gamma_{ik}(r)\,dr \int \phi_j(X_{qh})\gamma_{qk}(w)\,dw\right] \\
&= \sum_{k=1}^{K}\sum_{h=1}^{H}\sum_{j=1}^{J}\sum_{i=1}^{N}\sum_{q=1}^{N} E[\phi_j(X_{ih})\phi_j(X_{qh})] \int E[\gamma_{ik}(r)\gamma_{qk}(w)]\,drdw \\
&\leq KHJN \times \left(\max_{j\leq J, h\leq H, i\leq N} E[\phi_j^2(X_{ih})]\right)\left(\max_{k\leq K, i\leq N}\sum_{q=1}^{N}\int_0^1 |E[\gamma_{ik}(r)\gamma_{qk}(w)]|drdw\right) \\
&= O(JN\rho_N).
\end{aligned}
$$

4. By the definition of $G(X)$,

$$
E\|B^*\Phi(X)'\Gamma\|^2 \leq E\|G(X)^*\Gamma\|^2 + E\|R(X)^*\Gamma\|^2.
$$

   First, we bound $E\|G(X)^*\Gamma\|^2$.

$$
\begin{aligned}
E\|G(X)^*\Gamma\|^2 &= \sum_{k=1}^{K}\sum_{\ell=1}^{K}\sum_{i=1}^{N}\sum_{q=1}^{N} E\left[\int g_\ell(X_i, r)\gamma_{ik}(r)\,dr \int g_\ell(X_q, w)\gamma_{qk}(w)\,dw\right] \\
&= \int \sum_{k=1}^{K}\sum_{\ell=1}^{K}\sum_{i=1}^{N}\sum_{q=1}^{N} E\left[g_\ell(X_i, r)g_\ell(X_q, w)\right] E\left[\gamma_{ik}(r)\gamma_{qk}(w)\right]\,drdw \\
&\leq KHN \times \left(\sup_{r\in[0,1], k\leq K, i\leq N} E[g_k^2(X_i, r)]\right)\left(\max_{k\leq K, i\leq N}\sum_{q=1}^{N}\int |E[\gamma_{ik}(r)\gamma_{qk}(w)]|drdw\right) \\
&= O(N\rho_N).
\end{aligned}
$$

   Second, we bound $E\|R(X)^*\Gamma\|^2$.

$$
\begin{aligned}
E\|R(X)^*\Gamma\|^2 &= \sum_{k=1}^{K}\sum_{\ell=1}^{K}\sum_{i=1}^{N}\sum_{q=1}^{N} E\left[\int R_{i\ell}(r)\gamma_{ik}(r)\,dr \int R_{q\ell}(w)\gamma_{qk}(w)\,dw\right] \\
&= \int \sum_{k=1}^{K}\sum_{\ell=1}^{K}\sum_{i=1}^{N}\sum_{q=1}^{N} E\left[R_{i\ell}(r)R_{q\ell}(w)\right] E\left[\gamma_{ik}(r)\gamma_{qk}(w)\right]\,drdw \\
&\leq KHN \times \int \left(\sup_{r\in[0,1], k\leq K, i\leq N} E[R_{ik}^2(r)]\right)\left(\sup_{r,w\in[0,1], k\leq K, i\leq N}\sum_{q=1}^{N} |E[\gamma_{ik}(r)\gamma_{qk}(w)]|\right)drdw \\
&= O(J^{-\kappa}N\rho_N).
\end{aligned}
$$

   Hence, we have $E\|B^*\Phi(X)'\Gamma\|^2 = O(N\rho_N)$.

5.

$$
\begin{aligned}
\|P\Gamma\|^2 &= \|\Phi(X)\left(\Phi(X)'\Phi(X)\right)^{-1}\Phi(X)'\Gamma\|^2 \\
&\leq \|\Phi(X)\left(\Phi(X)'\Phi(X)\right)^{-1}\|_2^2\|\Phi(X)'\Gamma\|^2 \\
&= O_p\left(\frac{1}{N}\right)O_p(JN\rho_N) \\
&= O_p(J\rho_N).
\end{aligned}
$$

   where $\|\Phi(X)'\Gamma\|^2 = O_p(JN\rho_N)$ follows by the previous result.

6. We have

$$
\begin{aligned}
E\|\Gamma\|^2 &= \int \sum_{k=1}^{K} \sum_{i=1}^{N} E\left[\gamma_{ik}^2(r)\right] dr \\
&\leq \int K N \rho_N \, dr \\
&= O(N\rho_N),
\end{aligned}
$$

where the second inequality is due to the assumption 3.5. □

References

Ahn, S. C. and Horenstein, A. R. (2013). Eigenvalue ratio test for the number of factors, *Econometrica* **81**(3): 1203–1227.

Alessi, L., Barigozzi, M. and Capasso, M. (2010). Improved penalization for determining the number of factors in approximate factor models, *Statistics & Probability Letters* **80**(23-24): 1806–1813.

Auffhammer, M. (2018). Quantifying economic damages from climate change, *Journal of Economic Perspectives* **32**(4): 33–52.

Awika, J. M. (2011). Major cereal grains production and use around the world, *Advances in cereal science: implications to food processing and health promotion*, ACS Publications, pp. 1–13.

Bai, J. (2003). Inferential theory for factor models of large dimensions, *Econometrica* **71**(1): 135–171.

Bai, J. (2009). Panel data models with interactive fixed effects, *Econometrica* **77**(4): 1229–1279.

Bai, J. and Ng, S. (2002). Determining the number of factors in approximate factor models, *Econometrica* **70**(1): 191–221.

Bai, J. and Ng, S. (2007). Determining the number of primitive shocks in factor models, *Journal of Business & Economic Statistics* **25**(1): 52–60.

Bai, J. and Ng, S. (2010). Instrumental variable estimation in a data rich environment, *Econometric Theory* **26**(6): 1577–1606.

Bai, J. and Ng, S. (2021). Matrix completion, counterfactuals, and factor analysis of missing data, *Journal of the American Statistical Association* **116**(536): 1746–1763.

Benatia, D., Carrasco, M. and Florens, J.-P. (2017). Functional linear regression with functional response, *Journal of econometrics* **201**(2): 269–291.

Bernanke, B. S., Boivin, J. and Eliasz, P. (2005). Measuring the effects of monetary policy: a factor-augmented vector autoregressive approach, *The Quarterly journal of economics* **120**(1): 387–422.

Beyhum, J. and Gautier, E. (2021). Factor and factor loading augmented estimators for panel regression with possibly non-strong factors, *Journal of Business & Economic Statistics* (just-accepted): 1–36.

Bosq, D. (2000). *Linear processes in function spaces: theory and applications*, Vol. 149, Springer Science & Business Media.

Burke, M. and Emerick, K. (2016). Adaptation to climate change: Evidence from us agriculture, *American Economic Journal: Economic Policy* **8**(3): 106–40.

Campbell, B. (2022). *Climate change impacts and adaptation options in the agrifood system – A summary of recent IPCC Sixth Assessment Report findings*, Food and Agriculture Organization of the United Nations.

Chamberlain, G. (1983). Funds, factors, and diversification in arbitrage pricing models, *Econometrica* .

Chamberlain, G. and Rothschild, M. (1983). Arbitrage, factor structure, and mean-variance analysis on large asset markets, *Econometrica* .

Chen, S., Chen, X. and Xu, J. (2016). Impacts of climate change on agriculture: Evidence from china, *Journal of Environmental Economics and Management* **76**: 105–124.

Connor, G., Hagmann, M. and Linton, O. (2012). Efficient semiparametric estimation of the fama–french model and extensions, *Econometrica* **80**(2): 713–754.

Connor, G. and Linton, O. (2007). Semiparametric estimation of a characteristic-based factor model of common stock returns, *Journal of Empirical Finance* **14**(5): 694–717.

Cook, J., Oreskes, N., Doran, P. T., Anderegg, W. R., Verheggen, B., Maibach, E. W., Carlton, J. S., Lewandowsky, S., Skuce, A. G., Green, S. A. et al. (2016). Consensus on consensus: a synthesis of consensus estimates on human-caused global warming, *Environmental Research Letters* **11**(4): 048002.

Dell, M., Jones, B. F. and Olken, B. A. (2012). Temperature shocks and economic growth: Evidence from the last half century, *American Economic Journal: Macroeconomics* **4**(3): 66–95.

Deschênes, O. and Greenstone, M. (2007). The economic impacts of climate change: evidence from agricultural output and random fluctuations in weather, *American economic review* **97**(1): 354–385.

Deschênes, O. and Greenstone, M. (2011). Climate change, mortality, and adaptation: Evidence from annual fluctuations in weather in the us, *American Economic Journal: Applied Economics* **3**(4): 152–85.

Fan, J., Ke, Y. and Liao, Y. (2021). Augmented factor models with applications to validating market risk factors and forecasting bond risk premia, *Journal of Econometrics* **222**(1): 269–294.

Fan, J., Liao, Y. and Wang, W. (2015). Supplementary appendix to the paper projected principal component analysis in factor models, *Annals of statistics* .

Fan, J., Liao, Y. and Wang, W. (2016). Projected principal component analysis in factor models, *Annals of statistics* **44**(1): 219.

Fan, J., Xue, L. and Yao, J. (2017). Sufficient forecasting using factor models, *Journal of econometrics* **201**(2): 292–306.

Hall, P. and Horowitz, J. L. (2007). Methodology and convergence rates for functional linear regression, *The Annals of Statistics* **35**(1): 70–91.

Hallin, M. and Liška, R. (2007). Determining the number of factors in the general dynamic factor model, *Journal of the American Statistical Association* **102**(478): 603–617.

Hastie, T. and Mallows, C. (1993). [a statistical view of some chemometrics regression tools]: Discussion, *Technometrics* **35**(2): 140–143.

Horváth, L. and Kokoszka, P. (2012). *Inference for functional data with applications*, Vol. 200, Springer Science & Business Media.

Hsiang, S. and Kopp, R. E. (2018). An economist's guide to climate change science, *Journal of Economic Perspectives* **32**(4): 3–32.

Hsing, T. and Eubank, R. (2015). *Theoretical foundations of functional data analysis, with an introduction to linear operators*, Vol. 997, John Wiley & Sons.

IPCC (2014). The synthesis report: Climate change 2014, *IPCC Geneva, Switzerland* .

Kelly, B. T., Pruitt, S. and Su, Y. (2020). Instrumented principal component analysis, *Available at SSRN 2983919* .

Kim, S., Korajczyk, R. A. and Neuhierl, A. (2021). Arbitrage portfolios, *The Review of Financial Studies* **34**(6): 2813–2856.

Kokoszka, P. and Reimherr, M. (2017). *Introduction to functional data analysis*, Chapman and Hall/CRC.

Kozak, S., Nagel, S. and Santosh, S. (2020). Shrinking the cross-section, *Journal of Financial Economics* **135**(2): 271–292.

Lam, C. and Yao, Q. (2012). Factor modeling for high-dimensional time series: inference for the number of factors, *The Annals of Statistics* pp. 694–726.

Lettau, M. and Pelger, M. (2020). Estimating latent asset-pricing factors, *Journal of Econometrics* **218**(1): 1–31.

Li, D., Qian, J. and Su, L. (2016). Panel data models with interactive fixed effects and multiple structural breaks, *Journal of the American Statistical Association* **111**(516): 1804–1819.

Myers, K. F., Doran, P. T., Cook, J., Kotcher, J. E. and Myers, T. A. (2021). Consensus revisited: quantifying scientific agreement on climate change and climate expertise among earth scientists 10 years later, *Environmental Research Letters* **16**(10): 104030.

Onatski, A. (2010). Determining the number of factors from empirical distribution of eigenvalues, *The Review of Economics and Statistics* **92**(4): 1004–1016.

Pesaran, M. H. (2006). Estimation and inference in large heterogeneous panels with a multifactor error structure, *Econometrica* **74**(4): 967–1012.

Pesaran, M. H. and Tosetti, E. (2011). Large panels with common factors and spatial correlation, *Journal of Econometrics* **161**(2): 182–202.

Pörtner, H.-O., Roberts, D. C., Adams, H., Adler, C., Aldunce, P., Ali, E., Begum, R. A., Betts, R., Kerr, R. B., Biesbroek, R. et al. (2022). Climate change 2022: Impacts, adaptation and vulnerability, *IPCC Sixth Assessment Report* .

Ramsay, J. O. and Dalzell, C. (1991). Some tools for functional data analysis, *Journal of the Royal Statistical Society: Series B (Methodological)* **53**(3): 539–561.

Ramsay, J. O. and Silverman, B. W. (2005). *Functional data analysis*, Springer Series in Statistics, Springer, New York.

Rivas, M. D. G. and Gonzalo, J. (2020). Trends in distributional characteristics: Existence of global warming, *Journal of Econometrics* **214**(1): 153–174.

Staiger, D. O. and Stock, J. H. (1994). Instrumental variables regression with weak instruments.

Stock, J. H. and Watson, M. W. (2002a). Forecasting using principal components from a large number of predictors, *Journal of the American statistical association* **97**(460): 1167–1179.

Stock, J. H. and Watson, M. W. (2002b). Macroeconomic forecasting using diffusion indexes, *Journal of Business & Economic Statistics* **20**(2): 147–162.

Tavakoli, S., Nisol, G. and Hallin, M. (2021). Factor models for high-dimensional functional time series, *arXiv preprint arXiv:1905.10325* .

Tol, R. S. (2009). The economic effects of climate change, *Journal of economic perspectives* **23**(2): 29–51.

Vershynin, R. (2010). Introduction to the non-asymptotic analysis of random matrices, *arXiv preprint arXiv:1011.3027* .

Vershynin, R. (2018). *High-dimensional probability: An introduction with applications in data science*, Vol. 47, Cambridge university press.

Xie, L., Lewis, S. M., Auffhammer, M. and Berck, P. (2019). Heat in the heartland: crop yield and coverage response to climate change along the mississippi river, *Environmental and resource economics* **73**(2): 485–513.

Xiong, R. and Pelger, M. (2019). Large dimensional latent factor modeling with missing observations and applications to causal inference, *arXiv preprint arXiv:1910.08273* .

Zheng, X. and Li, Y. (2011). On the estimation of integrated covariance matrices of high dimensional diffusion processes, *The Annals of Statistics* **39**(6): 3121–3151.