Semiparametric panel data models with observable and latent factors

Jean-David Fermanian[a,*], Léonard Thelot[b]

[a] *Crest-ENSAE, 5, avenue Henry Le Chatelier, 91120 Palaiseau cedex, France, jean-david.fermanian@ensae.fr*
[b] *Crest & HSBC Global Asset Management, 5, avenue Henry Le Chatelier, 91120 Palaiseau cedex, France, leonard.thelot@ensae.fr*

Abstract

We extend static linear factor models within a semiparametric framework by allowing both latent and observable factors to explain the endogenous variable through some loadings that are modeled as unknown individual functions of a vector of covariates. These covariates are exogenous, time-varying, and differ across individuals. By explicitly combining observable and latent factors alongside explanatory variables, our model provides a more flexible representation than traditional factor models. We establish the consistency of the estimated loading functions and the latent factors under "large N and large T" asymptotics. The practical relevance of the proposed methodology is assessed through Monte Carlo simulations and an empirical application to stock returns.

Keywords: latent factors; observable factors; principal components; identifiability; kernel regression
JEL classification: C14, C33, C38, C55, G12

1. Introduction

In statistics, economics, finance, and many other fields, latent factor models have become a workhorse for the study of panel data and discrete-time multivariate dynamic processes. They offer an intuitive and flexible framework for modeling and understanding time-dependent and cross-sectional dependencies between endogenous variables. Under a "static" approach, neither lagged factors nor factor dynamics are assumed in the underlying DGP. In this setting, a simple spectral decomposition of the covariance matrix (Principal Component Analysis or PCA) of panel data yield estimated factors and loadings, partly explaining the popularity of this framework. See some seminal papers as Stock and Watson (2002), Bai (2003) and the surveys of Bai and Wang (2016), Gagliardini et al. (2020), Fan et al. (2021), for instance. An impressive number of papers have refined and extended the original approach, in particular by considering dynamic factors and more sophisticated inference methods (Fourier analysis, iterative methods as Kalman

*Corresponding author

filter or EM-algorithms, sieves, etc.). See Stock and Watson (2011, 2016) and a historical perspective in Barigozzi and Hallin (2024).

Incorporating additional information contained in explanatory variables $(Z_{it})_{i\geqslant 1,t\geqslant 1}$ appears to be key to enriching latent factor models, especially for forecasting. Moreover, assuming time-independent loadings may lead to poor or even misspecified model dynamics (Gagliardini and Ma, 2019). One method consists of linearly adding Z_{it}-based indices in model dynamics, yielding the family of interactive fixed-effect models (or time-varying individual effects). This approach was initiated by Pesaran (2006), who introduced the so-called "Common Correlated Effect" estimator. Alternative PCA-based inference strategies for the same model have been developed by Moon and Weidner (2015, 2017), among others. In particular, Bai (2009) established the interactive fixed effects framework, where outcomes are driven by observed regressors and unobserved common factors with index dependent loadings. Bai and Li (2014) extended this framework by providing a general inferential theory for panel models with interactive fixed effects. Notably, Chen et al. (2021) proposed a transformation (nonlinear) factor models for panel data with similar fixed effects.

A different and increasingly influential approach treats loadings not as individual-specific constant vectors, but as (parametric or unknown) functions of exogenous covariates. These covariates are observable, may be dynamic, systematic or idiosyncratic [1]. We will adopt this point of view. This stream of the literature has been fueled by the influential contributions of Connor and Linton (2007), Connor et al. (2012) and Fan et al. (2016). More recently, Zhang et al. (2021) proposed a semiparametric latent factor model in which loadings are modeled as nonparametric functions of observed covariates, as an extension of Bai (2009), Bai and Li (2014). Their projection based estimation strategy recovers both the latent factors and the covariate dependent loadings, while allowing for index specific heteroscedasticity and cross-sectional and temporal dependence. In the same vein, Kelly et al. (2019, 2020) assume that loadings are linear functions of asset characteristics and estimate the model using Instrumented Principal Component Analysis (IPCA). In Gagliardini and Ma (2019), time and stock-dependent loading maps are obtained by sieve non-parametric estimation through Artificial Neural Networks (ANN). In Pelger and Xiong (2022), loadings are nonlinear functions of a systematic univariate state process only, and are estimated trough localized PCA. Alternatively but still using PCA, Fermanian and Thélot (2025) propose to specify loadings as map of single-indices of covariates. Under the assumption of additive loading functions, Chen et al. (2023) apply the method of sieves, extending Fan et al. (2016) to the case of nonlinear loading functions of static individual characteristics. This is still the approach of Zhang et al. (2021), Keilbar et al. (2025), among others. Under a machine learning perspective, Gu et al. (2021) apply the deep learning machinery (autoencoders) to estimate nonlinear loadings functions of time-varying covariates. Their empirical results seem to be good,

[1] A notable exception is Su and Wang (2017) who assume individual loadings are deterministic smooth functions of time only (without any covariate), and they estimate their model by a time-localized version of PCA. See Motta et al. (2011) too.

but without clear-cut theoretical guarantees. Note that similar ideas have been recently extended in quantile regression settings (Ma et al., 2021) and for Generalized Method of Moments methods (Cheng et al., 2024).

Most of the latter papers suffer from some limitations. First, the covariate processes $(Z_{it})_{i\geqslant 1,t\geqslant 1}$ are often constant over time (any Z_{it} does not depend on t) as in Connor et al. (2012) or in the "Projected PCA" method of Fan et al. (2016); less frequently, covariates are the same between individuals (Z_{it} does not depend on i) as in Pelger and Xiong (2022). Second, when covariates are included in factor models, most authors assume there is a single loading map and heterogeneity across individuals is only transmitted by such covariates: Connor et al. (2007), Zhang et al. (2021), Fermanian and Thélot (2025), Chen et al. (2023), etc. Third, the latter models have difficulty incorporating high dimensional vectors Z_{it} due to the curse of dimensionality, except Fermanian and Thélot (2025). In such a case, most if not all authors have until now assumed that loadings are additive functions of the Z_{it}-components. This restriction is mot often associated with expansions of loading maps on some arbitrarily chosen functional basis, allowing to recover the usual machinery of linear models. Nonetheless, aside from the questionable nature of the choice of an additive model, the simplicity and elegance of the PCA method is sacrificed. Moreover, in sieve-based techniques, the number of parameters to be estimated by optimization very rapidly increases with the dimension of Z_{it}.

In this paper, we propose a new methodology for estimating panel factor models in which both observable and latent factors operate through smooth, covariate-dependent loadings. Unlike standard approaches that rely exclusively on one or the other, our framework explicitly integrates both types of factors, allowing them to operate simultaneously in shaping the behavior of the outcome variable. Traditional factor models typically assume constant factor loadings across units and time, which can mask important heterogeneities driven by observable characteristics. Our approach addresses this limitation by allowing the impact of any factor, latent and/or observable, to flexibly vary depending on observed covariates. In asset pricing, such features are typically firm characteristics or some macroeconomic indicators, thereby capturing richer cross-sectional dynamics and more realistic economic relationships.

Identification is consistently a challenge in latent factor models. Following Bai and Ng (2013), our model will be identifiable under some convenient identification restrictions (up to a sign, as usual in this literature). To estimate our model, we develop a two-step procedure that combines nonparametric regression and localized PCA. As a first stage and for each individual unit in the panel, we use kernel-based smoothing techniques to estimate the multivariate loadings function of covariates related to the observable factors. This produces a flexible, data-driven estimate of how observable characteristics shape the exposure to known risk drivers or economic variables. After isolating the component explained by the observable factors and as a second step, we estimate the latent factor structure that remains in the estimated residuals. To this aim, we apply a localized version of PCA, that gathers units with similar covariate profiles, effectively uncovering latent factors whose influence also varies across the covariate space. Our method estimates a localized covariance matrix of the estimated residuals, inspired by the recent paper of Xiong and Pelger (2023).

The resulting framework is a hybrid of two powerful ideas, the explanatory strength of observable factor models and the flexibility and dimensionality reduction of latent factor models. By allowing both types of factors to interact with covariates in a smooth, nonparametric fashion, we provide a unified and comprehensive approach to modeling complex panel data. This methodology is not only theoretically grounded, but also practically implementable and well-suited to a wide range of empirical applications where heterogeneity and nonlinearity are central features of the data. We state the consistency of our estimated factors, loadings and common components on average and point-wise, when the number of names N and the time length T both tend to the infinity ("large N, large T" asymptotic results).

The rest of the paper is organized as follows: Section 2 provides a description of our semiparametric latent factor model, in addition to sufficient conditions for its identification; Section 3 specifies the estimation method of factors, loading maps and common components; in Section 4, we establish the asymptotic properties of the latter estimators; Section 5 illustrates our method by simulation and on a real dataset of stock returns; finally, Section 6 concludes. Some appendices gather all proofs, technicalities and figures.

2. Model specification

Assume we observe a panel dataset of univariate random elements X_{it}, for some so-called "names" $i \in \{1, \dots, N\}$ and some dates $t \in \{1, \dots, T\}$. We will use the term "name" to cover usual situations in macroeconomics, microeconomics, finance, etc. Depending on the case, an index i may represent an individual, a firm, a security, a plant, a household, etc. In the case of financial econometrics, for instance, X_{it} could be the daily return at time t of the stock price associated with the firm i. We will assume N and T are simultaneously "large", i.e. our theoretical results will be stated under the assumption that N and T both tend to the infinity in a "convenient way" (see below).

Factor models are commonly used in many fields, in particular because they easily induce some reduction of dimensions. Factors are often observable (as in Fama and French (1993) or Fama and French (1995) in finance, e.g.), or unobservable/latent. In the latter (static) case, they may typically be estimated by PCA, as in Bai (2003). It is recognized that finding the right number and identifying the "best" underlying factors is a difficult task. Therefore, since the risk of forgetting some relevant factors is high, it makes sense to simultaneously consider both types of factors in the same model specification.

Beside, a possibly time-varying vector of covariates $Z_{it} \in \mathbb{R}^q$ is observable for every name and at every date. For instance, in the case of stock returns, Z_{it} may concatenate some micro-economic or financial information about firm i at time t: industry, country, size, financial ratios, realized volatility, technical indicators related to its past stock variations, etc. As explained in the introduction, it makes sense to assume our underlying factor loadings are not constant but may depend on such covariates.

Formally, we will consider the so-called Generalized static Factor Model with Functional Loadings (called GFLFM hereafter), defined as

$$X_{it} = B_i(Z_{it})^\top F_t + C_i(Z_{it})^\top G_t + \varepsilon_{it},\ \mathbb{E}[\varepsilon_{it}|F_t, G_t, Z_{it}] = 0, \tag{GFLFM}$$

for every i and every t. The unknown latent factors $F_1, \dots, F_T$ belong to $\mathbb{R}^r$ and are considered as realizations of a r-dimensional hidden random process, for some known integer r. There possibly exist additional observable factors $G_1, \dots, G_T$ that belong to $\mathbb{R}^p$ for some integer p. The factor loadings are defined as unknown maps $B_i : \mathbb{R}^q \to \mathbb{R}^r$ and $C_i : \mathbb{R}^q \to \mathbb{R}^p$, $i \in \{1, \dots, N\}$. It should be noted that our model extends most of the (static) factor models proposed in the literature. When $B_i(\cdot)$ does not depend on covariates, we recover the usual latent factor model (if $G_t = 0$) or the panel data model with unobservable interactive effects (if $G_t = 1$ and C_i is a linear map [2]). When $p = 0$ (no observable factor), we get static latent factor models where loadings depend on covariates, and we include most of the models mentioned above as special cases. In particular, we encompass the model of Fan et al. (2016) (Projected-PCA) because our loadings $B_i(\cdot)$ may depend on i. Still when $p = 0$ and if $B_i(\cdot)$ are linear, our model is similar to that in Kelly et al. (2020) (IPCA). When $r = 0$ (no latent factor), GFLFM boils down to functional-coefficient regression model for time series (Cai et al., 2000).

As a particular case of GFLFM and when $p \geqslant 2$, we could impose $G_t = [1, G_{2t}, \dots, G_{pt}]$. The model is then rewritten

$$X_{it} = B_i(Z_{it})^\top F_t + C_{i1}(Z_{it}) + \sum_{k=2}^{p} C_{ik}(Z_{it}) G_{kt} + \varepsilon_{it},\ \mathbb{E}[\varepsilon_{it}|F_t, G_t, Z_{it}] = 0. \tag{2.1}$$

The latter model explicitly introduces a functional individual effect that depends on covariates. In (2.1), we have to impose $\mathbb{E}[F_t] = 0$ (or another constant) for the sake of identifiability, but such a constraint is not required for general GFLFM with non constant processes (G_t).

3. Estimation strategy

We propose a two-step estimation strategy for our GFLFM. In a first step, the loading maps $C_i(\cdot)$ are nonparametrically estimated by $\widehat{C}_i(\cdot)$, $i \in \{1, \dots, N\}$. In a second step, we estimate the factors F_t, $t \in \{1, \dots, T\}$, and the loading maps $B_i(\cdot)$, $i \in \{1, \dots, N\}$, using PCA through localization techniques. Herafter and in notational terms, every unkown random or deterministic element, say E, covered by a wide hat, say $\widehat{E}$, will denote an estimator of E. Moreover, for any vector $\boldsymbol{x}$, $\|\boldsymbol{x}\|$ denotes an arbitrarily chosen (unless explicitly stated) norm; for any matrix A, $\|A\|$ denotes the multiplicative matrix norm of A induced by the latter norm for vectors.

[2] If $G_t = 1$ and $C_i(\cdot)$ is single-index, we get the framework of Feng et al. (2019)

3.1. First stage estimators: the loadings $C_i(\cdot)$

When there are no observed factors G_t, this section can be skipped and go directly to Section 3.2. Otherwise, for a fixed i and any value z in the support $\mathcal{S}_i$ of Z_{it}, deduce from (GFLFM)

$$X_{it} - \mathbb{E}[X_{it}|Z_{it} = z] = B_i(z)^\top \big(F_t - \mathbb{E}[F_t|Z_{it} = z]\big) \\ + \quad C_i(z)^\top \big(G_t - \mathbb{E}[G_t|Z_{it} = z]\big) + \varepsilon_{it}. \tag{3.1}$$

This yields

$$\mathrm{Cov}(X_{it}, G_t|Z_{it} = z) = B_i(z)^\top \mathrm{Cov}(F_t, G_t|Z_{it} = z) + C_i(z)^\top \mathrm{Var}(G_t|Z_{it} = z) \\ = \quad C_i(z)^\top \mathrm{Var}(G_t|Z_{it} = z), \tag{3.2}$$

under the orthogonality condition

$$\mathrm{Cov}(F_t, G_t|Z_{it} = z) = 0. \tag{3.3}$$

The conditional expectations $\mathbb{E}[X_{it}|Z_{it} = z]$, $\mathbb{E}[G_t|Z_{it} = z]$, $\mathbb{E}[X_{it}G_t|Z_{it} = z]$ and $\mathbb{E}[G_t G_t^\top|Z_{it} = z]$ can be consistently estimated since G_t is observable. Many methods are available to this aim: kernel smoothing, local polynomials, wavelets, etc. Finally, the vector $C_i(z)$ will be estimated by inverting the estimated matrix of $\mathrm{Var}(G_t|Z_{it} = z)$ in (3.2). This yields the estimators $\widehat{C}_i(z)$ of $C_i(z)$ for any $i \in \{1, \dots, N\}$ and convenient vectors z. With obvious notations, this means

$$\widehat{C}_i(z) := \widehat{\mathrm{Var}}(G_t|Z_{it} = z)^{-1} \widehat{\mathrm{Cov}}(X_{it}, G_t|Z_{it} = z). \tag{3.4}$$

Note that an alternative estimator based on local linear regression techniques has been proposed in Cai et al. (2000). For the sake of generality, we assume we have conveniently chosen an estimator $\widehat{C}_i(\cdot)$ so that the maps $\widehat{C}_i(z)$ converge in probability towards $C_i(z)$ uniformly w.r.t. (i, z).

Assumption 1. *For any $z \in \bigcap_{i\geqslant 1} \mathcal{S}_i$, there exists a deterministic sequence $(\theta_{N,T})$ s.t. $\theta_{N,T} \underset{N,T\to\infty}{\longrightarrow} 0$ and*

$$\sup_{i\leqslant N, z'\in\mathcal{C}_z} \|\widehat{C}_i(z') - C_i(z')\| = O_P(\theta_{N,T}), \tag{3.5}$$

for some compact subset $\mathcal{C}_z$ in $\mathbb{R}^q$ whose interior contains z.

The simplest estimators $\widehat{C}_i(z)$ are surely obtained by kernel smoothing w.r.t. the values of Z_{it}, using a bandwidth h_0 sequence that tends with zero with T: see (C.4). In Appendix C, we prove that Assumption 1 is then satisfied for the latter estimator, under some conditions of regularity, with

$$\theta_{N,T} = \Big(\frac{\ln T}{T h_0^q}\Big)^{1/2} + h_0^2. \tag{3.6}$$

Note that the latter uniform rate of convergence does not depend on N because we will assume $N = O(T^\xi)$ for some $\xi > 0$. To summarize, Assumption 1 is more the definition of $\theta_{N,T}$ than an assumption, strictly speaking.

To establish (3.6), we will invoke and state some results on kernel regression under temporal dependence. In particular, under some conditions of regularity, Nadaraya–Watson estimators admit uniform in z expansions whose biases are controlled uniformly over any compact set (Robinson, 1983, Bosq, 1998). If $(Z_{it}, \varepsilon_{it})_{t\in\mathbb{Z}}$ is alpha-mixing, with mixing coefficients decaying at a polynomial or exponential rate, we obtain the weak convergence of $\widehat{C}_i(z')$ uniformly w.r.t. (z', i), following Masry (1996), Hansen (2008).

Remark 1. *When the observed covariates include an intercept, as in (2.1), and if we impose $\mathbb{E}[F_t] = 0$, the estimator of $C_i(\cdot)$ is simpler because centering is no longer required. With obvious notations, the relevant estimator of $C_i(z)$ is then*

$$\widetilde{C}_i(z) := \widehat{\mathbb{E}}[G_t G_t^\top | Z_{it} = z]^{-1} \widehat{\mathbb{E}}[X_{it} G_t | Z_{it} = z], \tag{3.7}$$

and it will be assumed that $\mathbb{E}[F_t G_t^\top | Z_{it} = z] = 0$ instead of (3.3). In this case, it can be proved that (3.5) is satisfied with (3.6) exactly in the same way as for the general case.

Remark 2. *Centering the observed factors G_t is usually done in practice. This procedure means rewriting (GFLFM) as*

$$X_{it} = B_i(Z_{it})^\top F_t + C_i(Z_{it})^\top \mathbb{E}[G_t] + C_i(Z_{it})^\top (G_t - \mathbb{E}[G_t]) + \varepsilon_{it}. \tag{3.8}$$

Note that the latter model is not the same as (2.1). Indeed, its intercepts now depend on the maps $C_i(\cdot)$ and may not be arbitrarily chosen (contrary to (2.1)). Nonetheless, the loadings $C_i(\cdot)$ in (3.8) can be similarly estimated by (3.4), replacing G_t with $G_t - \widehat{\mathbb{E}}[G_t]$. The estimation of factors and loadings $B_i(z)$ is then unchanged.

3.2. Second stage estimators: the loadings $B_i(\cdot)$ and the factors F_t

Once the $C_i(\cdot)$ are estimated, we define pseudo-observations[3] by $\widehat{u}_{it} := X_{it} - \widehat{C}_i(Z_{it})^\top G_t$, which approximate the unobservable quantities $u_{it} := X_{it} - C_i(Z_{it})^\top G_t$. The latter ones satisfy the latent factor structure

$$u_{it} = B_i(Z_{it})^\top F_t + \varepsilon_{it}, \; \mathbb{E}[\varepsilon_{it} | F_t, G_t, Z_{it}] = 0. \tag{3.9}$$

Now, we will evaluate our loadings and factors by localized PCA, using the pseudo-observations $\widehat{u}_{it}$ instead of the unobservable quantities u_{it}.

Hereafter, set $\boldsymbol{F}^\top := \big[F_1, \dots, F_T\big]$ in $\mathbb{R}^{r\times T}$, and $\boldsymbol{B}(z)^\top := \big[B_1(z), \dots, B_N(z)\big]$ in $\mathbb{R}^{r\times N}$. In our second step, we estimate the sequence of latent factors $F_1, \dots, F_T$ and the z-values $B_1(z), \dots, B_N(z)$ of the loading functions. To this aim, we identify the couples (i, t) for which their covariates Z_{it} are close to the fixed chosen value z. After

[3] also called estimated residuals when they are centered, as in (2.1)

weighting and averaging by kernel smoothing, we diagonalize a convenient matrix built from cross-products between the pseudo-observations $\widehat{u}_{it}$.

Under the conventional latent factor estimator with constant loadings, we would apply PCA to the matrix $\hat{\Sigma} = N^{-1}\hat{u}\hat{u}^\top$, where $\hat{u} = [\hat{u}_{it}]_{1\leqslant t\leqslant T, 1\leqslant i\leqslant N}$, that could be called "sample covariance matrix" by a misuse of language [4]. Since we are focusing on observations for which the covariates Z_{it} are close to z, we need a new way of defining a convenient sample covariance matrix that accounts for localization before applying PCA. This will be done by kernel smoothing again.

Therefore, denote by K a multivariate kernel function of dimension q, i.e. a map $K : \mathbb{R}^q \mapsto \mathbb{R}$ s.t. $\int K = 1$. Consider a bandwidth sequence $h = h(T)$, i.e. a tuning parameter that will tend to zero when T tends to the infinity. To simplify, we will define our kernel weights using the same univariate bandwidth for all components even if extensions to matrices of bandwidths are feasible. For any $z \in \mathbb{R}^q$ and any couple (i,t), denote $w_{it}(z) := h^{-q}K\big((Z_{it} - z)/h\big)$ and $N_{ts}(z) = \sum_{i=1}^N w_{it}(z)w_{is}(z)$. Even if the kernel K and the bandwidths $h(T)$ may depend on i, we omit this refinement to lighten our notations [5]. Our localized "sample covariance" matrix $\tilde{\Sigma}(z) := [\tilde{\Sigma}_{ts}(z)]_{1\leqslant t,s\leqslant T}$ is then defined by

$$\tilde{\Sigma}_{ts}(z) := \frac{1}{N_{ts}(z)}\sum_{i=1}^N w_{it}(z)w_{is}(z)\widehat{u}_{it}\widehat{u}_{is},\ 1 \leqslant t,s \leqslant T. \tag{3.10}$$

The definitions of $\tilde{\Sigma}(z)$ and $\tilde{\Sigma}_{ts}(z)$ can be understood by drawing a parallel with factor models with missing data, i.e. unbalanced panels. Since we focus on the observations indexed by (i,t) for which $Z_{it} \simeq z$, all the others will be considered as uninformative and could be removed. Then, they would be "missing" in the panel dataset. Our weights can be seen as smoothed versions of binary variables indicating whether (i,t) belongs to the panel. Thus, our sample covariance $\tilde{\Sigma}(z)$ is similar to the covariance matrix defined in Pelger and Xiong (2023, Equation (1)), on which a PCA is then applied, in the framework of usual static latent factor models.

This leads us to apply a PCA to the normalized covariance matrix $T^{-1}\tilde{\Sigma}(z)$ to estimate the factors F_t. The latter estimates are stacked in the $T \times r$ matrix $\widehat{\boldsymbol{F}} := [\widehat{F}_1, \dots, \widehat{F}_T]^\top$. This matrix is obtained as $\sqrt{T}$ times the eigenvectors of the r largest eigenvalues of $\tilde{\Sigma}(z)$, that is

$$\frac{1}{T}\tilde{\Sigma}(z)\widehat{\boldsymbol{F}} = \widehat{\boldsymbol{F}}V_z, \tag{3.11}$$

where V_z is a diagonal matrix containing the r largest eigenvalues of $\tilde{\Sigma}(z)$. Then, for every individual i, the loading $B_i(z)$ is estimated by regressing $\widehat{u}_{it}$ on $\widehat{F}_t$ given $Z_{it} = z$, i.e.

$$\widehat{B}_i(z) := \Big(\sum_{t=1}^T w_{it}(z)\widehat{F}_t\widehat{F}_t^\top\Big)^{-1}\Big(\sum_{t=1}^T w_{it}(z)\widehat{F}_t\widehat{u}_{it}\Big). \tag{3.12}$$

[4] Indeed, the quantities u_{it} and $\hat{u}_{it}$ are not centered in general.

[5] In particular, it is easy to check that our theoretical results still apply if there exist two positive constants $\underline{\ell}$ and $\overline{\ell}$ s.t. $\underline{\ell} \leqslant h_i(T)/h_1(T) \leqslant \overline{\ell}$ for every (i,T).

Remark 3. *Note that our model specification and Lemma 4 below imply*

$$\tilde{\Sigma}_{ts}(z) \simeq \frac{F_t^\top}{N_{ts}(z)}\Big(\sum_{i=1}^N w_{it}(z)w_{is}(z)B_i(z)B_i(z)^\top\Big)F_s \simeq F_t^\top\Big(\frac{1}{N}\sum_{i=1}^N B_i(z)B_i(z)^\top\Big)F_s =: \bar{\Sigma}_{ts}(z).$$

Setting the matrix $\bar{\Sigma}(z) := \big[\bar{\Sigma}_{ts}(z)\big]_{1\leqslant s,t\leqslant T}$, *check that* $\boldsymbol{x}^\top\bar{\Sigma}(z)\boldsymbol{x} \geqslant 0$ *for any* $\boldsymbol{x} \in \mathbb{R}^T$. *Therefore,* $\bar{\Sigma}(z)$ *is nonnegative and* $\tilde{\Sigma}(z)$ *is "approximately" nonnegative definite.*

To ensure the identifiability of the factors and the loadings in (3.9), we refer to the usual problem of identification in latent factor models with fixed factor loadings. In our framework, it is well-known that the factors and the loadings are only estimated up to an invertible matrix R because $\boldsymbol{F}\boldsymbol{B}(z)^\top = \boldsymbol{F}RR^{-1}\boldsymbol{B}(z)^\top$. Among the different solutions proposed by Bai and Ng (2013), we selected their set of identifying restrictions called PC1: the two matrices $\boldsymbol{F}$ and $\boldsymbol{B}(z)$ are uniquely defined [6] under the following Assumption.

Definition 1. *The couple* $\big(\boldsymbol{F}, \boldsymbol{B}(z)\big)$ *satisfies PC1 if* $T^{-1}\sum_{t=1}^T F_tF_t^\top = I_r$ *and*

$$\boldsymbol{B}(z)^\top\boldsymbol{B}(z) \text{ is diagonal with } r \text{ distinct entries.} \tag{3.13}$$

We will impose the latter PC1 condition on the matrix of estimated factors $\widehat{\boldsymbol{F}}$, i.e. $T^{-1}\widehat{\boldsymbol{F}}^\top\widehat{\boldsymbol{F}} = I_r$ herafter. Concerning the actual underlying factors, it will be sufficient (and more realistic) to only impose the weaker condition

$$\frac{1}{T}\sum_{t=1}^T F_tF_t^\top \longrightarrow I_r, \tag{3.14}$$

in probability, when $T \to \infty$.

In the next section, we show that the previous estimators of factors and loadings are consistent. The last step is to estimate the localized common components $Com_{it}(z) := B_i(z)^\top F_t$ by using the plug-in estimator, $\widehat{Com}_{it}(z) = \widehat{B}_i(z)^\top\widehat{F}_t$.

Remark 4. *Our latter estimators can straightforwardly be extended to unbalanced panels strictly speaking as in Pelger and Xiong (2023), at least under a "missing at random" attrition process: just replace the previously defined weights* w_{it} *by* $w_{it}\pi_{it}$, *where* $\pi_{it} = 1$ *when* (X_{it}, Z_{it}) *is observed, and* $\pi_{it} = 0$ *otherwise.*

3.3. Extension with multiple z values

The previous analysis allows the inference of the factors F_t based on the statistical behavior of our data when $Z_{it} = z$ approximately, for a single value z. Therefore, different z values will provide different estimates of F_t. If the model is well specified, i.e. if the

[6] up to column sign changes, i.e. the sign of all $F_{k,t}$, $t \in \{1,\dots,T\}$, for some $k \in \{1,\dots,r\}$ may be reversed if the sign of the univariate factor loading map $x \mapsto B_k(x)$ is reversed too.

actual underlying GDP is really the factor model (GFLFM), all the estimated factors F_t (for the same t but different z) should not be very different from each other. Nonetheless, in the case of misspecification, this source of variability may become problematic. To mitigate the uncertainty around the true value of F_t induced by the arbitrariness of the choice of the covariate reference value, it is easy to extend our methodology by considering a grid of values $g := \{z_1, \dots, z_m\}$ where every z_k belongs to $\mathbb{R}^q$, instead of a single $z \in \mathbb{R}^q$. Then, we estimate the factors (F_t) by averaging the localized losses on all points z_k of the latter grid. To be specific and with the same notations as above, define a new matrix $\tilde{\Sigma}_g := [\tilde{\Sigma}_{g,ts}]_{1 \leqslant s,t \leqslant T}$, with

$$\tilde{\Sigma}_{g,ts} := \sum_{k=1}^{m} \tilde{\Sigma}_{ts}(z_k) = \sum_{k=1}^{m} \frac{1}{N_{ts}(z_k)} \sum_{i=1}^{N} w_{it}(z_k) w_{is}(z_k) \widehat{u}_{it} \widehat{u}_{is}, \; 1 \leqslant t, s \leqslant T.$$

Then, apply a PCA to the normalized covariance matrix $T^{-1}\tilde{\Sigma}_g$ to estimate the factors. Such estimators are defined as $\sqrt{T}$ times the eigenvectors of the r largest eigenvalues of $\tilde{\Sigma}_g$ as usual, i.e.

$$\frac{1}{T} \tilde{\Sigma}_g \widehat{\boldsymbol{F}} = \widehat{\boldsymbol{F}} V_g,$$

where V_g is a diagonal matrix containing the r largest eigenvalues of $\tilde{\Sigma}_g$. Then, for every individual $i \in \{1, \dots, N\}$ and any $k \in \{1, \dots, m\}$, the loading $B_i(z_k)$ is estimated by regressing $\widehat{u}_{it}$ on $\widehat{F}_t$ given $Z_{it} = z_k$, i.e.

$$\widehat{B}_i(z_k) := \big(\sum_{t=1}^{T} w_{it}(z_k) \widehat{F}_t \widehat{F}_t^\top \big)^{-1} \big(\sum_{t=1}^{T} w_{it}(z_k) \widehat{F}_t \widehat{u}_{it} \big).$$

Moreover, to ensure identifiability, we just need to replace condition (3.13) in PC1 by

$$\sum_{k=1}^{m} \boldsymbol{B}(z_k)^\top \boldsymbol{B}(z_k) \text{ is diagonal with } r \text{ distinct entries.}$$

Remark 5. *If we assume that the loadings $B_i(\cdot)$ are constant maps, then their estimated values are simply*

$$\widetilde{B}_i := \frac{1}{T} \sum_{t=1}^{T} \widehat{F}_t \widehat{u}_{it}, \; i \in \{1, \dots, N\}. \tag{3.15}$$

4. Asymptotic results

In this section, we prove the consistency of our estimated factors, loadings and common components. To simplify the presentation, we state the results for a single covariate value z. It can be straightforwardly checked that all our theoretical results are valid for the estimates built from a grid g (as in Section 3.3) instead from a single z only. It is just sufficient to check that our regularity assumptions hold for every $z = z_k$, $k \in \{1, \dots, m\}$. Since we establish our results under the double asymptotics "large N, large T", we assume hereafter there are some positive numbers κ_k, $k \in \{1,\dots,4\}$, s.t. $N^{\kappa_1} \leqslant T \leqslant N^{\kappa_2}$ and $N^{-\kappa_3} \leqslant h \leqslant N^{-\kappa_4}$, $\kappa_3 < 1$.

4.1. Technical assumptions

Before stating our main theoretical results, we need to impose some regularity assumptions.

Assumption 2. *The multivariate kernel K is nonnegative, symmetrical and its support is included in a compact subset of $\mathbb{R}^q$. Moreover, K is Lipschitz continuous: for every $(x,y) \in \mathbb{R}^{2q}$, we have $|K(x) - K(y)| \leqslant L_K \|x - y\|$ for some constant L_K.*

The latter condition on the kernel K is rather weak. Without a lack of generality, we assume hereafter that $\text{supp}(K) = (-1,1)^q$. The product of q univariate Epanechnikov kernels $K(z) = \prod_{k=1}^q K_0(z_k)$, with $K_0(t) := 3(1-t^2)\mathbf{1}(|t| \leqslant 1)/4$, satisfies Assumption 2 in particular.

Assumption 3. *The loading maps $B_{il}(\cdot)$, $i \in \{1,\dots,N\}$ and $l \in \{1,\dots,r\}$, are twice continuously differentiable on $\mathcal{V}(z)$, a neighborhood of z, and $\sup_i \sup_{z' \in \mathcal{V}(z)} \|\nabla^k B_i(z')\| < \infty$, $k \in \{1,2\}$. Moreover, $\sup_i \|B_i(z)\| =: M_B < \infty$. As $N \to \infty$, we have $\boldsymbol{B}(z)^T \boldsymbol{B}(z)/N \xrightarrow{P} \Sigma_B$ for some $r \times r$ positive definite matrix Σ_B.*

Assumption 4. *As $T \to \infty$, we have $\boldsymbol{F}^T \boldsymbol{F}/T \xrightarrow{P} \Sigma_F$ for some $r \times r$ positive definite matrix Σ_F. The r largest eigenvalues of $\tilde{\Sigma}(z)$, denoted $\lambda_1,\dots,\lambda_r$, are bounded below almost surely: there exists a positive constant c_v s.t. $\lambda_j > c_v$ for every $j \in \{1,\dots,r\}$ a.s. Moreover, the eigenvalues of the $r \times r$ matrix $\Sigma_B^{1/2} \Sigma_F \Sigma_B^{1/2}$ are distinct.*

The latter assumption is standard in the literature of latent factor models. Set the $r \times r$ diagonal matrix $V_z := Diag(\lambda_1,\dots,\lambda_r)$. Note that $\tilde{\Sigma}(z)$ and V_z are random matrices that implicitly depend on N, T and h. Assumption 4 implies that the elements of $V_z^{-1} = Diag(\lambda_1^{-1},\dots,\lambda_r^{-1})$ are bounded above a.s. by a positive number that does not depend on (N,T,h).

Define $W_{it} := (\varepsilon_{it}, Z_{it})$ and the infinite dimensional random vector $W_t := (W_{it})_{i \geqslant 1}$.

Assumption 5. *For all i,k in $\{1,\dots,N\}$ and t,s in $\{1,\dots,T\}$,*

$$\mathbb{E}\big[\varepsilon_{it}\varepsilon_{kt} | Z_{it}, Z_{kt}, Z_{is}, Z_{ks}, F_s\big] = 0 \text{ a.s. if } i \neq k. \tag{4.1}$$

The process $(W_t, F_t)_{t \geqslant 1}$ is strongly stationary and α-mixing. Its sequence of mixing coefficients (α_t) satisfies $\alpha_t \leqslant \exp(-C_\alpha t)$, for some constant $C_\alpha > 0$.

Hereafter, the stationarity of $(W_t, F_t)_{t \geqslant 1}$ is supposed to apply. As a consequence of Assumption 5, $(Z_{it})_{t \geqslant 1}$ (resp. $(F_t)_{t \geqslant 1}$) is a strongly stationary process, for any i. Thus, the density of Z_{it} does not depend on t and is denoted f_i. Moreover, the density of (Z_{it}, Z_{is}) depends on $t-s$ and is denoted $f_{i,t-s}$. Note that $f_{i,t-s} = f_{i,s-t}$ by the stationarity of our covariates.

Assumption 6. *For every couple (t,s) (possibly with $t=s$), $(Z_{it},Z_{is})_{i\geqslant 1}$ is a sequence of mutually independent random vectors. Any f_i (resp. $f_{i,t-s}$) is twice continuously differentiable on a neighborhood of z (resp. (z,z)). Moreover, assume* $\inf_i f_i(z) > 0$,

$$\liminf_N \inf_{1\leqslant s\leqslant T} \frac{1}{N}\sum_{i=1}^N \inf_{\{(u,v);\|u\|<h,\|v\|<h\}} f_{i,s}(z+u,z+v) =: \underline{\ell}_\infty(z) > 0, \tag{4.2}$$

$$\liminf_N \frac{1}{N}\sum_{i=1}^N \inf_{\{u;\|u\|<h\}} f_i(z+u) =: \tilde{\underline{\ell}}_\infty(z) > 0, \tag{4.3}$$

$$\limsup_N \sup_{1\leqslant s\leqslant T} \frac{1}{N}\sum_{i=1}^N \sup_{\{(u,v);\|u\|<h,\|v\|<h\}} f_{i,s}(z+u,z+v) < \infty, \tag{4.4}$$

$$\limsup_N \sup_{1\leqslant s\leqslant T} \frac{1}{N}\sum_{i=1}^N \sup_{\{(u,v);\|u\|<h,\|v\|<h\}} \|\nabla^k f_{i,s}(z+u,z+v)\| < \infty,\ k\in\{1,2\},\ \textit{and} \tag{4.5}$$

$$\limsup_N \frac{1}{N}\sum_{i=1}^N \sup_{\{u;\|u\|<h\}} f_i(z+u) < \infty. \tag{4.6}$$

Assumption 7. $\mathbb{E}[\|F_t\|^6] < \infty$, $T^{-1}\sum_{t=1}^T \|G_t\|^2 = O_P(1)$ *and* $\sup_N N^{-1}\sum_{i=1}^N \mathbb{E}[\varepsilon_{it}^4] < \infty$.

Technically speaking, our demonstrations require controlling certain conditional expectations for some values of the covariates, but on average across all names. This leads us to introduce a convenient concept of regularity. For any positive integer m, denote $I_{N,m}$ the set of m-tuplets $(i_1,\dots,i_m)$ in $\{1,\dots,N\}^m$ that are all distinct in pairs. Consider arbitrary measurable maps $h^{(\ell)}_{i_1,\dots,i_m}:\mathbb{R}^{mq}\to\mathbb{R}$, whose arguments will be m random vectors Z_{i_k,t_k}, $t_k\in\{1,\dots,T\}$ and $k\in\{1,\dots,m\}$. The dependence w.r.t. the indices t_k is summarized into the additional index $\ell=(t_1,\dots,t_m)$ that belongs to a subset $\mathcal{T}$ of $\{1,\dots,T\}^m$. Typically, $\ell=(t,s)$ with $s\neq t$, or even $\ell=(t,t)$, when $m=2$. For any $(i_1,\dots,i_m)\in I_{N,m}$ and any integer N, assume that the density of $(Z_{i_1,t_1},\dots,Z_{i_m,t_m})$ exists and is denoted $f^{(\ell)}_{i_1,\dots,i_m}$ with $\ell=(t_1,\dots,t_m)$.

Definition 2. *A family of maps $h^{(\ell)}_{i_1,\dots,i_m}:\mathbb{R}^{qm}\to\mathbb{R}$ indexed by $(i_1,\dots,i_m)\in I_{N,m}$, $N\geqslant m$, and $\ell\in\mathcal{T}\subset\{1,\dots,T\}^m$ is called index-regular of order s_0 at z, $s_0\geqslant 0$, if it is s_0-times differentiable in an open neighborhood of $(z,\dots,z)$ and if there exists some $\epsilon>0$ s.t.*

$$\sup_N \sup_{\ell\in\mathcal{T}} \frac{1}{N^m} \sum_{(i_1,\dots,i_m)\in I_{N,m}} \sup_{\{(u_1,\dots,u_m);\|u_k\|<\epsilon,\forall k\}} \|\nabla^{s_0}(h^{(\ell)}_{i_1,\dots,i_m} f^{(\ell)}_{i_1,\dots,i_m})(z+u_1,\dots,z+u_m)\| < \infty.$$

Hereafter and to be short, we will omit specifying the local nature of the latter concept. Note that, if a family of maps is index-regular of order $s_0>0$, this does not imply that it is regular of order $s_0'<s_0$.

Assumption 8. *The families of maps* $(x, y) \mapsto \mathbb{E}\big[\varepsilon_{it}^2 \varepsilon_{is}^2 | Z_{it} = x, Z_{is} = y\big]$, $(x, y) \mapsto \mathbb{E}[\varepsilon_{is}^2 \|F_t\|^2 | Z_{is} = x, Z_{it} = y]$, *and* $(x, y) \mapsto \mathbb{E}[\varepsilon_{is}^2 \|G_t\|^2 | Z_{is} = x, Z_{it} = y]$, *indexed by* (s, t) *with* $s \neq t$, *are index-regular of order zero. The families of maps* $x \mapsto \mathbb{E}[\varepsilon_{it}^2 \|F_t\|^2 | Z_{it} = x]$ *and* $x \mapsto \mathbb{E}[\varepsilon_{it}^2 \|G_t\|^2 | Z_{it} = x]$ *are index-regular of order zero. Moreover, for some* $\nu > 2$,

$$\sup_{i \neq k} \sup_{\{(u,v); \|u\| < h, \|v\| < h\}} \mathbb{E}\big[|\varepsilon_{it} \varepsilon_{kt}|^\nu | Z_{it} = z + u, Z_{kt} = z + v\big] < \infty. \tag{4.7}$$

Note that the family of maps $x \mapsto \mathbb{E}[\varepsilon_{it}^2 \|G_t\|^2 | Z_{it} = x]$ is indexed by t since we have not assumed the stationarity of (G_t).

A technical assumption is required to approximate $\widehat{\pi}_t$ by another matrix $\widehat{\pi}$ that does not depend on t (c.f. Lemma 4).

Assumption 9. *For any* (t, s), $s \neq t$, *the series* $N^{-1} \sum_{i=1}^N f_{i,t-s}(z, z)$ *tends to a finite limit denoted* $\ell_\infty(z)$ *when* $N \to \infty$. *Moreover,* $\gamma_{N,T} = O\big(h^2 + \sqrt{N}^{-1} h^{-q} \ln(N)\big)$, *where*

$$\gamma_{N,T} := \sup_{1 \leqslant s \leqslant T} \Big\| \frac{1}{N} \sum_{i=1}^N f_{i,s}(z, z) B_i(z) B_i(z)^\top - \Big(\frac{1}{N} \sum_{i=1}^N f_{i,s}(z, z) \Big) \Big(\frac{1}{N} \sum_{i=1}^N B_i(z) B_i(z)^\top \Big) \Big\|.$$

Intuitively, Assumption 9 is satisfied when the values $f_{i,t-s}(z, z)$ and the loadings $B_i(z)$ are approximately "not correlated", once the latter quantities are considered as randomly selected [7]. In other words, for any name i, knowing the density values $f_{i,t-s}(z, z)$ does not provide any information on its loading values and vice versa, which seems reasonable. The required magnitude of $\gamma_{N,T}$, an "empirical covariance", is realistic since its distance to the "true covariance" zero is of order $1/\sqrt{N}$, typically. Note that Assumption 9 is satisfied when $f_{i,t-s}$ does not depend on i (in particular if the DGP of covariates is the same for every name), since $\gamma_{N,T} = 0$ in this case.

To state the consistency of the estimated loadings $\widehat{B}_i(z)$, we need additional conditions of regularity, particularly on the tail behavior of the law of F_t.

Assumption 10. *For some* $\gamma_0 > 0$ *and some* $b > 0$,

$$\mathbb{P}(\|F_t\|_\infty > x) < \exp\big(1 - (x/b)^{\gamma_0}\big), \; x > 0, \; \textit{and}$$

$$\inf_i \min \Big(\textit{Spectrum of } \mathbb{E}[F_t F_t^\top | Z_{it} = z] \Big) > 0.$$

For any $(l, l') \in \{1, \ldots, r\}^2$, denote $G_{ll',i}(x) := \mathbb{E}[F_{tl} F_{tl'} | Z_{it} = x]$, $G_{F,i}(x) := \mathbb{E}[\|F_t\|^2 | Z_{it} = x]$ and $G_{G,i}^{(t)}(x) := \mathbb{E}[\|G_t\|^2 | Z_{it} = x]$, that may depend on t because we have not assumed the stationarity of (G_t), contrary to (F_t). For any $(k, k', l) \in \{1, \ldots, r\}^3$, define the family of maps $(x, y) \mapsto \bar{G}_{kk'l,i}^{(s,t)}(x, y) := \mathbb{E}[F_{tk} F_{sk'} F_{tl} F_{sl} | Z_{it} = x, Z_{is} = y]$. For any $l \in \{1, \ldots, r\}$ and any $\mu > 0$, set the maps $H_{i,\mu,l} : x \mapsto \mathbb{E}\big[|\varepsilon_{it} F_{tl}|^\mu | Z_{it} = x\big]$.

[7] in particular through the random choice of the index i, since the order of names is arbitrary

Assumption 11. *For every* $(l,l') \in \{1,\dots,r\}^2$,

$$\sup_i \sup_{\{u;\|u\|<h\}} \|\nabla^2(f_i G_{ll',i})(z+u)\| < \infty, \text{ and } \sup_i \sup_{\{u;\|u\|<h\}} \|(f_i G_{F,i})(z+u)\| < +\infty. \quad (4.8)$$

For any i, $\sup_t \sup_{\{u;\|u\|<h\}} \left(f_i G^{(t)}_{G,i}\right)(z+u) < +\infty$, *and the maps* $(G^{(t)}_{G,i})_{i\geqslant 1}$ *are index-regular of order zero. For any* $l \in \{1,\dots,r\}$ *and some* $\bar{\nu} > 2$, *the maps* $(H_{i,\bar{\nu},l})_{i\geqslant 1}$ *and* $(H_{i,2,l})_{i\geqslant 1}$ *are index-regular of order zero. For any* $(k,k',l) \in \{1,\dots,r\}^3$, *the family of maps* $(\bar{G}^{(s,t)}_{kk'l,i})_{i\geqslant 1}$, *indexed by* (t,s) *with* $s \neq t$, *is index-regular of orders zero and one.*

4.2. Consistency

To demonstrate that our estimators are consistent, we substitute the values of $\widehat{u}_{it}$ for all (i,t) into (3.11), which leads to the following decomposition (see details in Appendix A):

$$\widehat{F}_t - \widehat{\pi}_t^\top F_t = o_P(1), \text{ with } \widehat{\pi}_t := \frac{1}{T}\sum_{i=1}^N \sum_{s=1}^T \frac{w_{it}(z) w_{is}(z)}{N_{ts}(z)} B_i(Z_{it}) B_i(Z_{is})^\top F_s \widehat{F}_s^\top V_z^{-1}. \quad (4.9)$$

Similarly to Bai (2003), the latter equation expresses the estimated factors in terms of the actual factors multiplied by a matrix $\widehat{\pi}_t$, along with some residual terms which will be negligible (see the proof of Theorem 1). The key difference from traditional factor analysis stems from the fact that the latter matrix depends on t, preventing an overall assessment of the estimated factors. Fortunately, it can be proved (see Lemma 4) that $\widehat{\pi}_t$ can be replaced by another matrix $\widehat{\pi}$ that no longer depends on t. Moreover, when the model is identifiable, $\widehat{\pi}_t$ can even be replaced by the identity matrix (Corollary 4), providing consistency in the classical sense.

The next theorems establish our main results. They are related to the weak consistency of the estimated factors (for any t and in average), of the loadings (for any i and in average) and the common components evaluated at z. All proofs are postponed in Appendix A. To this aim, define our convenient rate of convergence

$$v_{N,T} = \frac{\ln^2 N}{N h^{2q}} + \frac{1}{T h^{4q(1-1/\nu)}} + h^4 + \theta^2_{N,T},$$

where $\theta_{N,T}$ (resp. ν) has been introduced in (3.5) (resp. Assumption 8).

Theorem 1 (Consistency of the estimated factors)**.** *If Assumption 1 to Assumption 9 hold and* $v_{N,T} \to 0$ *when* N *and* T *tend to the infinity, then* $\|\widehat{F}_t - \widehat{\pi}^\top F_t\|^2 = O_P(v_{N,T})$ *for any* $t \in \{1,\dots,T\}$ *and*

$$\frac{1}{T}\sum_{t=1}^T \|\widehat{F}_t - \widehat{\pi}^\top F_t\|^2 = O_P(v_{N,T}),$$

where $\widehat{\pi} := \mathbf{B}(z)^\top \mathbf{B}(z) \boldsymbol{F}^\top \widehat{\boldsymbol{F}} V_z^{-1}/(NT)$.

Note that $\widehat{\pi}$ is the usual transformation matrix in latent factor models (Bai, 2003, Theorem 1) when the vector of covariates is degenerate and equal to z a.s. If Assumption 9 is not satisfied, Theorem 1 still applies, replacing $\widehat{F}_t - \widehat{\pi}^\top F_t$ with $\widehat{F}_t - \widehat{\pi}_t^\top F_t$ and $v_{N,T}$ with $1/(Nh^{2q}) + 1/(Th^{4q(1-1/\nu)}) + \theta^2_{N,T}$ in its statement (see (A.6) in the proof).

Theorem 2 (Consistency of the estimated loadings)**.** *If Assumption 1 to Assumption 11 hold and $v_{N,T} = o(h^q)$, then $\|\widehat{B}_i(z) - \widehat{\pi}^{-1}B_i(z)\|^2 = O_P\big(v_{N,T}/h^q\big)$ for every $i \in \{1,\dots,N\}$ and $N^{-1}\sum_{i=1}^N \|\widehat{B}_i(z) - \widehat{\pi}^{-1}B_i(z)\|^2 = O_P(v_{N,T}/h^q)$. If, in addition $\sup_{i,t}|\varepsilon_{it}| < \infty$ a.s., $T^{-1}\sum_{t=1}^T \|G_t\|^2\|F_t\|^2 = O_P(1)$ and $\theta^2_{N,T} = O(h^q)$, then*

$$\frac{1}{N}\sum_{i=1}^N \|\widehat{B}_i(z) - \widehat{\pi}^{-1}B_i(z)\|^2 = O_P(v_{N,T}).$$

It is easy to check that the latter rates of convergence for $\widehat{B}_i(z)$ applied to $\widetilde{B}_i$, as defined in (3.15).

Remark 6. *The uniform boundedness of the innovations (ε_{it}) in Theorem 2 has been assumed only for convenience. It could be replaced by some painful conditions of moments, as in Bosq (1998, Theorem 1.4) for instance, so that we could apply an exponential inequality for alpha-mixing sequences of unbounded random variables.*

Corollary 3 (Consistency of Estimated Common Components)**.** *Under the assumptions of Theorem 1 and Theorem 2, $\widehat{Com}_{it}(z) - Com_{it}(z) = O_P(\sqrt{v_{N,T}}/h^{q/2})$ for any $t \in \{1,\dots,T\}$ and any $i \in \{1,\dots,N\}$. Moreover,*

$$\frac{1}{NT}\sum_{t=1}^T\sum_{i=1}^N \big(\widehat{Com}_{it}(z) - Com_{it}(z)\big)^2 = O_P\big(v_{N,T}\big).$$

We have proven that the estimated factors $\widehat{F}_t$ approximately belong to the same linear subspace as the actual factors F_t (Theorem 1). There remains an indeterminacy concerning F_t due to the presence of the random matrix $\widehat{\pi}$. By adding some identifiability constraints, the latter matrix will be asymptotically the identity matrix. Therefore, every estimated factors/loadings will converge in probability to the true underlying factors/loadings. This is the purpose of the following corollary.

Corollary 4 (Convergence of the transformation-type matrix $\widehat{\pi}$)**.** *Under the assumptions of Theorem 1, $\widehat{\pi}$ converges in probability to a deterministic matrix that depends on z only. If, in addition, we assume Conditions (3.13) and (3.14), then Model (3.9) is identifiable and $\widehat{\pi} \xrightarrow{P} I_r$.*

As a consequence of Corollary 4, $\widehat{F}_t - F_t = o_P(1)$, $\widehat{B}_i(z) - B_i(z) = o_P(1)$, and $\widehat{Com}_{it}(z) - Com_{it}(z) = o_P(1)$ for every fixed couple (i,t).

Remark 7. *A more relevant common component of name i at any time t is surely $Com_{it}(Z_{it}) = B(Z_{it})^\top F_t$ in practice. It should be estimated by $\widehat{Com}_{it}(Z_{it}) := \widehat{B}(Z_{it})^\top \widehat{F}_t$.*

The latter quantity is consistently estimated when the particular Z_{it} belongs to a grid of z-values, as explained in Section 3.3. In practice, this cannot be made for every Z_{it}, and $\widehat{B}(Z_{it})$ is most often obtained by interpolation between two estimated values $\widehat{B}(z_k)$ and $\widehat{B}(z_{k+1})$ on a fixed grid, as in Section 3.3. Thus, $\widehat{B}(Z_{it})$ will provide a (slightly) biased estimation of $B(Z_{it})$, and this bias effect will propagate across $\widehat{Com}_{it}(Z_{it})$. See Section 5.2 for a discussion.

4.3. Asymptotic normality

Under the assumptions of Theorem 1 and additional conditions of regularity, it can be proven that $\sqrt{Nh^q}\big(\widehat{F}_t - \widehat{\pi}_t^\top F_t\big)$ is asymptotically normal for any t, where $\widehat{\pi}_t$ has been defined in (4.9). Moreover, under the assumptions of Theorem 2 and other conditions of regularity, $\sqrt{Th^q}\big(\widehat{B}_i(z) - \widehat{\pi}^{-1}B_i(z)\big)$ is asymptotically normal for any i, at least when $q = 1$. These two results are valid when T and N tend toward infinity in an appropriate manner, particularly $T = o(N)$, and for a suitable choice of h. We provide a sketch of proofs of these two announced results in Appendix B, without specifying their particularly tedious and lengthy technicalities.

Under identifiability, $\widehat{\pi}_t$ and $\widehat{\pi}^{-1}$ tend to the identity matrix I_r (Corollary 4); since $\widehat{F}_t - F_t = \widehat{F}_t - \widehat{\pi}_t^\top F_t + (\widehat{\pi}_t - I_r)^\top F_t$, constructing a confidence interval around F_t requires specifying the limiting behavior of $\widehat{\pi}_t - I_r$; similarly, a confidence interval around $B_i(z)$ requires studying $\widehat{\pi} - I_r$. Both are rather difficult tasks. Indeed, PCA-based estimators are not really well suited for establishing such results, in contrast to usual M-estimators: see the arguments in the proof of Corollary 4. Thus, asymptotic normality results remain of primarily theoretical interest in our setting. Since formally deriving them would significantly lengthen our paper, we leave this objective for future work.

Remark 8. *It would be significantly more difficult to properly state the limiting law of $\widehat{F}_t - \widehat{\pi}^\top F_t$, with a convenient rate of convergence. Indeed, such a result requires to know the limiting distribution of $\widehat{\pi}_t - \widehat{\pi}$, that depends on the rate of convergence of $\gamma_{N,T}$ in Assumption 9. Nonetheless, if we assume that $Nh^q\gamma_{N,T}^2 = o(1)$, then $\widehat{\pi}_t - \widehat{\pi}$ becomes negligible and the same limiting law as above applies to $\widehat{F}_t - \widehat{\pi}^\top F_t$.*

4.4. Bandwidth choice

Since the convergence rates obtained above depend on the bandwidth h, it is natural to consider using our theoretical results to guide the choice of h. When $\theta_{N,T}$ does not depend on h (i.e. when h and h_0 are independently selected), a so-called "asymptotically optimal" bandwidth h^* could be defined by minimizing $v_{N,T}$, seen as a function of h for some fixed (N,T). By simple calculations, we obtain the order of magnitude of h^*:

- if $N/\ln^2 N = O(T^\iota)$ with $\iota := (q+2)/(2+q/\nu)$, then $h^* \sim (\ln N)^{1/(q+2)}N^{-1/(2q+4)}$;
- otherwise, $h^* \sim T^{-1/\{4+4q(1-1/\nu)\}}$.

Since the latter analysis is purely asymptotic and only provides an order of magnitude, a more pragmatic approach is desirable. Here, we can propose a data-driven bandwidth selector, by applying a k-fold cross-validation technique ($k = 5$, typically) to determine an "optimal" bandwidth h: split the dataset into k subsamples [8]; for each value of h on a given grid, estimate the factors F_t and the loadings $B_i(z)$ on $k-1$ subsets (the training set), and finally compute a performance criterion $\mathcal{L}(h,z)$ (a "loss") on the last subsample $\mathcal{S}$ (the test set). An intuitive choice would be to set $\mathcal{L}\big(h,z,\mathcal{S}\big) := \sum_{(i,t)\in\mathcal{S}} w_{it}\big\{\widehat{u}_{it} - \widehat{B}_i(z)^\top \widehat{F}_t\big\}^2$. We repeat this procedure k times so that each subsample is a test subset only once. Then, calculate the mean of the k losses we have obtained, after a plug-in of the estimated factors and loadings. After redoing this task for each value of h on the grid, we select the "optimal" bandwidth h for which the latter average loss is minimal.

Remark 9. *The latter k-fold cross validation can still be applied even if we impose $h = h_0$, evaluating the $C_i(\cdot)$ maps by kernel smoothing (see Appendix C). The methodology is exactly the same as before except that the loss becomes $\tilde{\mathcal{L}}\big(h,z,\mathcal{S}\big) := \sum_{(i,t)\in\mathcal{S}} w_{it}\big\{X_{it} - \widehat{B}_i(z)^\top \widehat{F}_t - \widehat{C}_i(z)^\top G_t\big\}^2$.*

4.5. Cross-sectionally dependent covariates

In Assumption 6, we assumed the cross-sectional independence between our covariates, but mainly for the sake of simplicity. Actually, the latter feature is only required to apply usual exponential inequalities in Lemma 3 and Lemma 4. It could be weakened in several ways. Note that assuming serial dependence between the random vectors $(Z_{it})_{i\geqslant 1}$, for any fixed date t [9], is surely not relevant here since the order between the indices i is arbitrary. Thus, many usual types of dependencies such as martingales, Markov processes, mixing processes, etc., are not really convenient in our case.

A more relevant approach could be to apply the clustering idea of Connor et al. (2012): there exist "clusters of names" (homogeneous buckets of individuals i) within which the covariates Z_{it} are dependent, possibly equal, for any fixed t. Conversely, covariate vectors present in different clusters are independent. The cluster sizes tend to the infinity with N when the number of clusters is fixed. Our results easily apply under this slightly extended framework, at the price of heavier notations and technicalities.

A second meaningful approach is to assume that cross-sectional covariates are conditionally independent. The latter approach respects the absence of order among individuals and allows cross-sectional dependencies between the vectors Z_{it}, $i \in \{1,\dots,N\}$. This extended framework is specified in the following assumption.

Assumption 12. *For any couple (t,s), there exists a random element $V_{t,s}$ s.t. $(Z_{it}, Z_{is})_{i\geqslant 1}$ is a sequence of mutually independent random vectors given $V_{t,s}$. The density of Z_{it} given $V_{t,t} = v$ does not depend on t, is denoted $f_{i|v}$ and $\inf_i f_{i|v}(z) > 0$ for almost every value v of $V_{t,t}$. Moreover, the density of (Z_{it}, Z_{is}) given $V_{t,t} = \tilde{v}$ depends on $t-s$ and is denoted $f_{i,t-s|\tilde{v}}$. Any $f_{i|v}$ (resp. $f_{i,t-s|\tilde{v}}$) is twice continuously differentiable on a neighborhood*

[8] There are several ways of splitting the NT observations indexed by (i,t). A naive and simple solution is to randomly draw NT/k couples (i,t) without replacement among the initial dataset.

[9] for instance asuming that $Z_{i,t}$ and $Z_{i+1,t}$ are more strongly correlated than $Z_{i,t}$ and $Z_{i+100,t}$

$\mathcal{V}(z)$ of z (resp. (z,z)). There exist positive constants ℓ_k, $k \in \{1,\dots,5\}$, s.t., for almost every value of v (resp. $\tilde{v}$) of $V_{t,t}$ (resp. $V_{t,s}$),

$$\liminf_N \inf_{1\leqslant s\leqslant T} \frac{1}{N}\sum_{i=1}^N \inf_{\{(u,u');\|u\|<h,\|u'\|<h\}} f_{i,s|\tilde{v}}(z+u, z+u') > \ell_1,$$

$$\liminf_N \frac{1}{N}\sum_{i=1}^N \inf_{\{u;\|u\|<h\}} f_{i|v}(z+u) > \ell_2,$$

$$\limsup_N \sup_{1\leqslant s\leqslant T} \frac{1}{N}\sum_{i=1}^N \sup_{\{(u,u');\|u\|<h,\|u'\|<h\}} f_{i,s|\tilde{v}}(z+u, z+u') < \ell_3,$$

$$\limsup_N \sup_{1\leqslant s\leqslant T} \frac{1}{N}\sum_{i=1}^N \sup_{\{(u,u');\|u\|<h,\|u'\|<h\}} \|\nabla^k f_{i,s|\tilde{v}}(z+u, z+u')\| < \ell_4,\ k \in \{1,2\},\ \textit{and}$$

$$\limsup_N \frac{1}{N}\sum_{i=1}^N \sup_{\{u;\|u\|<h\}} f_{i|v}(z+u) < \ell_5.$$

By invoking Hoeffding and Bernstein's exponential inequalities conditionally on the hidden factors $V_{t,s}$, and integrating the latter inequalities, Lemma 3 and Lemma 4 can straightforwardly be proven under Assumption 12. This justifies the next result.

Corollary 5. *Theorem 1, Theorem 2, Corollary 3 and Corollary 4 apply, replacing Assumption 6 with Assumption 12.*

To illustrate the scope of this extension, assume the underlying DGP is as follows: for some measurable maps ψ_G and ψ_i, $i \geqslant 1$, and for some independent processes (v_t), (ν_t) and (e_t) with $e_t := (e_{it})_{i\geqslant 1}$, set

$$\begin{cases} (a) & Z_{it} = \psi_i(v_t, e_{it}), \\ (b) & G_t = \psi_G(v_t, \nu_t), \\ (c) & F_t \text{ is independent of } (v_t, \nu_t, e_t). \end{cases}$$

Then, (3.3) is satisfied and the latter model specification, in addition to (GFLFM), can be estimated with our methodology. Assumption 12 is satisfied with $V_{t,s} := (v_t, v_s)$. Obviously, the roles of F_t and G_t can be interchanged: the same conclusions applies if F_t is defined as a map of v_t and a noise $\bar{\nu}_t$, when G_t is independent of $(v_t, e_t, \bar{\nu}_t)$.

Alternatively, the latter condition *(c)* could be replaced with the existence of a process $(\tilde{F}_t)$ s.t. $F_t = \psi_F(v_t) \odot \tilde{F}_t$ for some map ψ_F and componentwise vector multiplication, with $\mathbb{E}[\tilde{F}_t] = 0$ and $\tilde{F}_t$ is independent of (v_t, e_t, ν_t).

4.6. Choosing the number of latent and observable factors

We have so far assumed that the numbers of latent factors r and observable factors p are known. However, in the factor-model literature, determining the correct number of latent factors remains a classical yet challenging problem. Likewise, selecting both the number and identity of observable factors is a key practical issue, particularly in empirical asset pricing. Traditional approaches rely on model-comparison criteria such as the Akaike or Bayesian Information Criteria (AIC/BIC), which assess the trade-off between goodness of fit and model parsimony (Lewellen et al., 2010, Barillas and Shanken, 2017, Harvey et al., 2016). Cross-validation and out-of-sample forecast evaluation offer complementary data-driven tools for identifying the relevant observable factors (Welch and Goyal, 2008, Feng et al., 2020). In the asset-pricing literature, nested-model and spanning tests are also widely used to determine whether additional factors provide incremental explanatory power beyond existing ones Kan and Zhang (1999). More recently, penalization and regularization techniques such as LASSO and elastic net have been employed to handle high-dimensional sets of candidate observable factors while mitigating overfitting concerns (Feng et al., 2020, Bryzgalova et al., 2025). These approaches could be combined to form a coherent framework for determining the optimal number of observable factors prior to introducing additional latent components.

Following Bai and Ng (2002), the number of latent factors is typically selected using penalized information criteria. In the context of Generalized Dynamic Factor Models Forni et al. (2000), Hallin and Liška (2007) extended this methodology, and Alessi et al. (2010) further improved its robustness and small-sample performance. More recently, Gagliardini et al. (2019) proposed a diagnostic criterion to assess whether the error terms are merely weakly cross-sectionally correlated or share at least one common factor.

A rigorous analysis of the theoretical properties of analogous methods adapted to our framework would require substantial additional developments that lie beyond the scope of this paper. We therefore leave this investigation for future work. Nevertheless, in practice, one could conceive a two-step procedure: first, determine the number and identity of the observable factors using the techniques discussed above; then, in a second step, estimate the number of latent factors based on the estimated residuals $\widehat{u}_{it}$, conditional on this first-stage selection.

5. Empirical Analysis

5.1. Monte-Carlo experiments

In this section, we generate a dataset from the DGP given by (GFLFM). The vectors of true factors $(F_t)_{1\leqslant t\leqslant T}$ are drawn using of a r-dimensional VAR(1) process: we postulate $F_t = AF_{t-1} + \varepsilon_t$, with $\varepsilon_t \sim \mathcal{N}(0,\Sigma)$, for some (known) $r\times r$ matrices A and Σ. The observable factors $(G_t)_{t\geqslant 1}$ are drawn using p independent AR(1) processes. We draw F_0 and G_0 according to the stationary laws of $(F_t)_{t\geqslant 1}$ and $(G_t)_{t\geqslant 1}$. We want to compare our simulated factors and also our loadings with the true ones, which requires to analyze an identifiable model. In order to satisfy the identifiability condition PC1, we impose (3.14) concerning the latent factors $(F_t)_{t\geqslant 1}$. To this aim, it is sufficient to satisfy $\mathbb{E}[F_tF_t^\top] = I_r$. For example, if $r = 2$, this means $\mathbb{E}[F_{t,1}F_{t,2}] = 0$ and $\mathbb{E}[F_{t,1}^2] = \mathbb{E}[F_{t,2}^2] = 1$. When A and Σ are diagonal matrices, $A = \mathrm{Diag}(a_{11}, a_{22})$ and $\Sigma = \mathrm{Diag}(\sigma_{11}, \sigma_{22})$, this is equivalent to $a_{11}^2 + \sigma_{11}^2 = a_{22}^2 + \sigma_{22}^2 = 1$. This will be our setup in this section and we will set $a_{11} = 0.5$

and $a_{22} = 0.7$ in our experiments. Obviously, no identifiability condition is required for the observable factors $(G_t)_{t\geqslant 1}$.

Concerning covariates, for each $i \in \{1, \ldots, N\}$, $(Z_{it})_{t\geqslant 0}$ is simulated according to a VAR(1) process of dimension q. We impose that, for any i, the process of covariates $(Z_{it})_{t\geqslant 0}$ follows the stationary autoregressive dynamics $Z_{it} = 0.9Z_{it-1} + 0.05e_{it}$. The innovations $(e_{it})_{t\geqslant 1,i\geqslant 1}$ are mutually independent and follow strong Gaussian white noises, $e_{it} \sim \mathcal{N}(0, I_q)$, that are independent of the process $(\varepsilon_{it})_{t\geqslant 1,i\geqslant 1}$. Any initial value Z_{i0} is drawn according to the stationary law of the process $(Z_{it})_{t\geqslant 1}$.

The loadings are defined as second-order polynomial functions of Z_{it}. To be specific, for all $i \leqslant N$, set $B_i(Z_{it}) = b_{i0} + b_{i1}Z_{it} + b_{i2}(Z_{it} \odot Z_{it})$, where the vector b_{i0} will be drawn according to a Gaussian law $\mathcal{N}(0, I_r)$ and the $r \times q$ matrices b_{i1}, b_{i2} to a $\mathcal{N}(0, I_{r\times q})$. The same specification applies to the loadings $C_i(\cdot)$. By shifting the initial loadings functions, we can make the matrix $\boldsymbol{B}(z)^\top \boldsymbol{B}(z)$ diagonal. Thus, it satisfies (3.13) and we ensure PC1.

As a first experiment, we estimate the factors/loadings/common components from the previously defined model, with a given value z randomly drawn componentwise from the distribution of (Z_{it}), restricted to its 10%-90% quantile range. To get started, we are interested in studying a model with two latent factors ($r = 2$) and two observables factors ($p = 2$). Set $N = 1000$ and $T = 500$. Such large but nonetheless realistic sample sizes avoid prohibitive computational time. Finally, we have selected three covariates ($q = 3$).

Choose the product of q univariate Epanechnikov kernels $K(z) = \prod_{k=1}^{q} K_0(z_k)$, with $K_0(t) := 3(1-t^2)\mathbf{1}(|t| \leqslant 1)/4$, which satisfies Assumption 2. We generate mutually independent idiosyncratic noises $\varepsilon_{it} \sim \mathcal{N}(0, \sigma_i^2)$, for any (i, t), after drawing the parameters $\sigma_i^2 \sim \mathcal{U}(0, 0.5)$ independently. Obviously, such innovations satisfy Assumption 5.

The first stage $\widehat{C}_i(Z_{it})$ is obtained in a first step as detailed in Section 3.1, using the estimator studied in Appendix C. Choosing the couple of bandwidths (h_0, h^*) as in Remark 9 in Section 4.4, we illustrate Theorem 1, Theorem 2 and Corollary 3 on the consistency of the estimated factors, loadings and common components respectively. Figure D.2 shows the actual and the estimated latent factors at two hundred dates. They have been obtained on a particular simulated dataset. We can check that the estimated factors estimate the actual underlying factors very accurately [10]. Concerning loading functions (Figure D.3), the estimated values $B_j(z)$, $j \in \{1, 2\}$, are pretty close to the true underlying loadings. To improve the clarity and readability of this figure, we plotted the sorted values of each $B_{i1}(z)$ and $B_{i2}(z)$ w.r.t i, compared to the corresponding estimated values. Finally, from the same simulated data, Figure D.4 plots the two common components for two randomly chosen values of i, for two hundred dates and given $Z_{it} = z$. Even if the estimator of the common components is the product of the two latter estimators of factors and loadings, we observe that the accuracy of $\widehat{Com}_{it}(z)$ is slightly better than that of $\widehat{F}_t$ or $\widehat{B}_i(z)$.

[10] We will see later a more formal comparison between actual and estimated factors in terms of mean squared errors, for different values of r.

In a second step, by performing the same Monte-Carlo experiment as above n_{mc} times for $r \in \{1, 2, 3\}$, we evaluate the robustness of our previous findings. For the same data generating process (as defined in the beginning of this section), we generate several new paths by simulation. To be specific, for each simulation, the values of the parameters of the model like the matrix A in the definition of the factors and the vectors b_0, b_1, b_2 remain unchanged but the factors and loadings themselves are redrawn. The errors ε_{it}, the covariates Z_{it} and the value z will also be drawn at each step as explained previously. Then, we evaluate some accuracy measures between the estimated and actual factors/loadings: for every $k \in \{1, \ldots, r\}$ and $l \in \{1, \ldots, p\}$, define $\hat{M}_{F_k} := T^{-1} \sum_{t=1}^{T} (\widehat{F}_{t,k} - F_{t,k})^2$, $\hat{M}_{B_k} := N^{-1} \sum_{i=1}^{N} \big(\widehat{B}_{i,k}(z) - B_{i,k}(z)\big)^2$, $\hat{M}_{C_l} := N^{-1} \sum_{i=1}^{N} \big(\widehat{C}_{i,l}(z) - C_{i,l}(z)\big)^2$, and $\hat{C}om := (NT)^{-1} \sum_{i=1}^{N} \sum_{t=1}^{T} \big(\widehat{Com}_{it}(z) - Com_{it}(z)\big)^2$. Such MSE-type accuracy measures will be averaged over $n_{mc} = 100$ simulated paths. These means will be denoted M_Y for $Y \in \{F_k, B_k, C_l\}$, $k \in \{1, \ldots, r\}$, $l \in \{1, \ldots, p\}$, and Com respectively. They will be compared with the average second-order moments of the corresponding quantities $V_Y^2 := n_{mc}^{-1} \sum_{j=1}^{n_{mc}} \hat{\mathbb{E}}[Y_j^2]$ for $Y \in \{F_k, B_k, C_l\}$, $k \in \{1, \ldots, r\}$ and $l \in \{1, \ldots, p\}$, with obvious notations, in addition to V_X^2, the empirical second order moment of $(X_{it})_{i,t \geqslant 1}$. See the results in Table 1 for two observable factors. Table D.4 and Table D.5 in Appendix D show similar results for $p = 1$ and $p = 3$.

	$V_{F_1}^2$	$V_{F_2}^2$	$V_{F_3}^2$	M_{F_1}	M_{F_2}	M_{F_3}
r=1	1.009	-	-	0.112	-	-
r=2	1.165	1.038	-	0.165	0.105	-
r=3	0.999	0.968	0.935	0.187	0.129	0.111
	$V_{B_1}^2$	$V_{B_2}^2$	$V_{B_3}^2$	M_{B_1}	M_{B_2}	M_{B_3}
r=1	0.357	-	-	0.060	-	-
r=2	0.351	0.324	-	0.047	0.052	-
r=3	0.362	0.327	0.341	0.053	0.061	0.079
	$V_{C_1}^2$	$V_{C_2}^2$	$V_{C_3}^2$	M_{C_1}	M_{C_2}	M_{C_3}
r=1	0.346	-	-	0.076	-	-
r=2	0.340	0.341	-	0.079	0.083	-
r=3	0.354	0.352	-	0.125	0.083	-
		V_X^2			Com	
r=1		1.647			0.011	
r=2		1.502			0.013	
r=3		2.321			0.077	

Table 1: Empirical MSE and variances of factors, loadings and common components. Results of Monte-Carlo experiment for $r \in \{1, 2, 3\}, p = 2$ and 100 simulated paths.

Several consistent patterns emerge across all considered specifications. First, the MSE associated with our estimators are systematically a lot smaller than the second-order moments of their target variables. This illustrates the relevance of our inference strategy for realistic sample sizes. Second, the MSEs of the estimated factors increase moderately with the number of latent factors r. For instance, when $p = 2$, the factor MSE rises from 0.11 for $r = 1$ to 0.19 for $r = 3$, and a similar pattern holds for $p = 1$ and $p = 3$. At the opposite, the MSEs of the loadings remain stable across both p and r. This does not mean that loadings are more precisely estimated than factors. Indeed, the orders of magnitudes of the two quantities differ. Third, it seems that the $C_i(\cdot)$ coefficients are estimated with slightly less precision than the loadings $B_i(\cdot)$. Finally, common components are those recovered most accurately, by comparing Com and V_X^2. Since common components are key for prediction purpose, this result is very promising. The latter results are remarkably robust across p: Overall, the Monte Carlo experiment shows that the proposed estimation method consistently recovers both latent and observable components with relatively small errors in average, and that its performance is robust to the number of covariates p as well as to the number of latent factors r.

5.2. Empirical Application to Stock Returns

Now, we estimate the latent factor model GFLFM on individual monthly excess stock returns in the CRSP universe, and using some characteristics/covariates described in Kozak et al. (2020) (Internet Appendix D). We selected 6 530 firms during 678 months of observations, from July, 1963 to December, 2019. The three observable factors considered in our analysis correspond to the well-known Fama-French factors (Fama and French, 1993), i.e. the market excess return, the "size" factor (also called SMB), and the "value" factor (called HML). To avoid the curse of dimensionality, we will lead our empirical applications with no more than two covariables. We selected *Momentum 6 months (mom)* which is the cumulated past performance of a given stock in the previous 6 months after skipping the most recent month [11], and *Idiosyncratic Volatility (ivol)* which represents the standard deviation of the residuals from a firm-level regression of daily stock returns on the three Fama-French factors, using an estimation window of three months.

To evaluate the relevance of the estimated common components $\widehat{Com}_{it}(Z_{it}) = \widehat{B}_i(Z_{it})^\top \widehat{F}_t$, we compare them with the observed returns, for three latent factors. To this aim, we need to evaluate "the whole" maps $B_i(\cdot)$ and $C_i(\cdot)$, not only $B_i(z)$ and $C_i(z)$ at a first point z. Thus, we introduce a grid of values $\{z_1, \dots, z_m\}$ in $\mathbb{R}^q$. The loadings $B_i(\cdot)$ and $C_i(\cdot)$ are estimated on this grid, as suggested in Section 3.3. Elsewhere, every map $\widehat{B}_i$ is linearly interpolated and, possibly, extrapolated. For example when $q = 1$, the values z_k correspond to the quantiles of the distribution of the covariates Z_{it}, for the levels $\{k/(m+1); k = 1, \dots, m\}$ with $m = 10$. Thus, for each z, define $\widehat{B}_i(z) = \widehat{B}_i(z_s) + (z - z_s)(\widehat{B}_i(z_{s+1}) - \widehat{B}_i(z_s))/(z_{s+1} - z_s)$, for z_{s+1} and z_s if $z_s \leqslant z \leqslant z_{s+1}$, and assuming that the z-grid is sorted. If $z < z_1$, we simply set $\widehat{B}_i(z) = \widehat{B}_i(z_1)$ and if $z > z_m$, set $\widehat{B}_i(z) = \widehat{B}_i(z_m)$. In dimension two, we apply the interpolation method called *Delaunay triangulation*. Obviously, many alternative interpolation techniques could have been applied: radial basis functions, kriging, splines, etc. See Wendland (2004), e.g.

When the covariate is set to *mom*, approximately 20% of the estimated loading functions associated with the first factor are increasing, while 21% are decreasing. In addition,

[11] This means *mom* of i at time t is the past return of i between $t - 1$ and $t - 6$.

51% of the loading functions may be considered as "flat", and the remaining 8% exhibit no clear monotonic pattern. For the covariate *ivol*, the corresponding proportions are 21% (increasing), 24% (decreasing), 46% (flat) and 9% (non-monotonic). These values remain relatively stable across factors. Figure 1 displays the estimated loading functions, related to the first latent factor, when there is only one covariate, *mom* (left panel) or *ivol* (right panel) respectively, for ten typical stocks i. Thus, some assets with higher values of *mom* (or *ivol*) exhibit stronger exposure to the common factor, but this is the opposite for others. In other words, the marginal effect of each covariate on factor exposure is not uniform across firms. Instead, such characteristics may be seen as sorting variables that split the universe of assets: for some assets, increasing the chosen covariate amplifies its exposure to systematic risk, for others it attenuates it, while no clear pattern appear for still others. Hence, *mom* and *ivol* capture heterogeneous sensitivities to the first common factor. Moreover, Figure D.5 and Figure D.6 display typical estimated loading functions, related to the second and third latent factor and similar findings can be made.

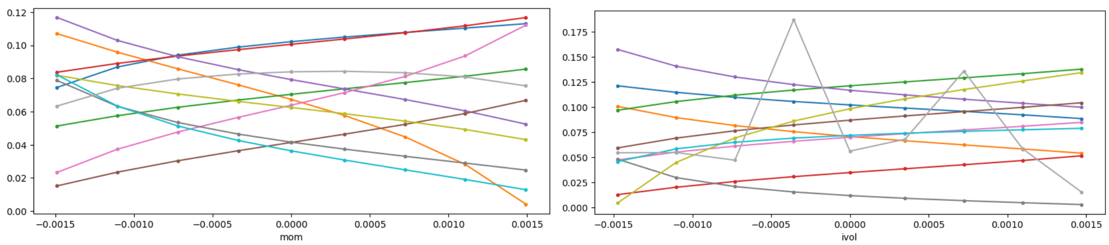

Figure 1: Estimated loadings related to the first latent factor, on a grid of *mom* and *ivol* values, for some indexes i ($q = 1$ and $r = 3$)

Similarly, Figure D.7, Figure D.8 and Figure D.9 show the estimated loading functions related to the three observable factors proposed by Fama and French. When the covariate is set to *mom*, approximately 26% of the estimated loading functions associated with the first observable factor (*market excess return*) are increasing, while 26% are decreasing. In addition, 37% of the loading functions appear to be flat, and the remaining 11% exhibit no clear monotonic pattern. For the covariate *ivol*, the corresponding proportions are 26% (increasing), 25% (decreasing), 44% (flat) and 5% (non-monotonic). These values remain relatively stable across factors. Globally, building covariate-dependent loadings in an (observable) factor model as in Fama and French (1993) seems to be relevant statistically speaking, even if a certain amount of heterogeneity across firms complicates their economic interpretation.

The MSE of the estimated common components (calculated on all asset returns and dates) is $(NT)^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(X_{it} - \widehat{Com}_{it}(Z_{it})\right)^2 = 1.2 \times 10^{-3}$, that is a lot smaller than the second moment of our returns $(NT)^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T} X_{it}^2 = 9.8 \times 10^{-3}$.

For the sake of comparison, we consider the classical factor model (Bai, 2003) defined as

$$X_{it} = \lambda_i^\top F_t + e_{it},\ t \in \{1, \dots, T\},\ i \in \{1, \dots, N\},$$

and also the IPCA model (Kelly et al., 2020) given by the equations

$$X_{it} = Z_{it}\Gamma F_t + e_{it},\ t \in \{1,\ldots,T\},\ i \in \{1,\ldots,N\},$$

where X_{it} is the observed return of i at time t, F_t is a vector $r \times 1$ of latent common factors, λ_i is a constant vector $r \times 1$ of factor loadings, Z_{it} are still the covariables of asset i at date t, Γ is an unknown matrix and e_{it} is the idiosyncratic risk of X_{it}. The variables F_t and e_{it} are obviously not observable. The estimated factors and loadings are obtained by PCA in Bai (2003) and by Alternative Least Squares in Kelly et al. (2020). The quantities of interest are their common components, denoted $Com_{it,B} = \lambda_i^\top F_t$ and $Com_{it,IPCA} = Z_{it}\Gamma F_t$ respectively. Their corresponding estimators are $\widehat{Com}_{it,B} = \hat{\lambda}_i^\top \hat{F}_t$ and $\widehat{Com}_{it,IPCA} = Z_{it}\hat{\Gamma}\hat{F}_t$. Then, we compute the MSE between the observed returns and the estimated common components for these two alternative models as we did before. We get $(NT)^{-1}\sum_{i=1}^N\sum_{t=1}^T(X_{it} - \widehat{Com}_{it,B})^2 = 1.5 \times 10^{-3}$ and $(NT)^{-1}\sum_{i=1}^N\sum_{t=1}^T(X_{it} - \widehat{Com}_{it,IPCA})^2 = 1.6 \times 10^{-3}$, that are slightly larger than what we obtained with the GFLFM. Thus, using the information contained in covariates in a flexible non-linear way apparently allows to outperform the classical factor models, at least in-sample.

So far, we studied only in-sample performances. For the purpose of return prediction, we need to perform also out-of-sample evaluation. Thus, the last five years of our dataset will be kept as an out-of-sample period of time. We apply an expanding window procedure to perform the estimation method described in Section 3 between the origin of time (July 1963, here) and $T-k$, for any $k \in \{1,\ldots,60\}$. For each value of k, we get estimated factors and loadings on all living stocks at each date $t \in \{1,\ldots,T-k\}$. Then, still for each value of k, we fit a VAR(1) model on the estimated factors to obtain their out-of-sample predictors for the next month. We also get the out-of-sample predictor of the matrix of covariates by selecting its last observable value. In other words, for every k, we estimate a $r \times r$ matrix A_k and a $r \times 1$ constant vector a_k s.t. the "out-of-sample" predicted factor at $T-k+1$ is $\hat{F}_{T-k+1,oos} = a_k + A_k\hat{F}_{T-k}$, and $\hat{Z}_{i,T-k+1,oos} = Z_{i,T-k}$, $i \in \{1,\ldots,N\}$.

Note that $\hat{F}_{T-k}$ has been obtained in-sample and $\hat{F}_{T-k+1,oos}$ is a predicted value. Moreover, $Z_{i,T-k}$ is observed when $\hat{Z}_{i,T-k+1,oos}$ is required. Hence, for all $k \in \{1,\ldots,60\}$, we have built an out-of-sample value of the estimated factor $\hat{F}_{T-k+1,oos}$ and the out-of-sample covariates $\hat{Z}_{i,T-k+1,oos}$. The latter one allows to compute the out-of-sample loadings $\hat{B}_{i,T-k}(\hat{Z}_{i,T-k+1,oos})$ [12] by linear interpolation as explained above. We perform also the same procedure for the standard constant loading model of Bai (2003).

To assess model performance in economic terms, we evaluate how return predictions deduced from each model translate into Sharpe ratios, for some portfolio strategies. To this aim and for any model, we sort our universe of stocks into deciles based on out-of-sample return forecasts. We construct a zero net-investment long-short portfolio that

[12] Obviously, $\hat{B}_{i,T-k}(\cdot)$ denotes the loading map estimated in sample from the observations until $T-k$.

buys the highest expected return stocks (decile 10) and sells the lowest (decile 1). We rebalance portfolios each month, and consider equally weighted portfolios.

Remark 10. *Note that, for theoretical reasons, the IPCA model requires the number of factors to be less than or equal to the number of characteristics.*

	Number of latent factors	1	2	3
PCA	Sharpe Ratio	0.49	0.57	0.55
	Portfolio returns	0.013	0.022	0.024
	Standard deviation	0.027	0.039	0.043
	Maximum Drawdown (%)	-3.95	-4.70	-5.29
	Omega Ratio	1.45	1.54	1.52
IPCA (*mom*)	Sharpe Ratio	0.71		
	Portfolio returns	0.031		
	Standard deviation	0.044		
	Maximum Drawdown (%)	-4.67		
	Omega Ratio	1.93		
IPCA (*mom* and *ivol*)	Sharpe Ratio	0.66	0.77	
	Portfolio returns	0.030	0.027	
	Standard deviation	0.046	0.036	
	Maximum Drawdown (%)	-5.07	-5.58	
	Omega Ratio	1.80	1.79	
GFLFM (*mom*)	Sharpe Ratio	1.56	1.46	1.29
	Portfolio returns	0.043	0.024	0.014
	Standard deviation	0.028	0.017	0.011
	Maximum Drawdown (%)	-1.72	-1.69	-2.58
	Omega Ratio	4.90	3.57	2.98
GFLFM (*mom* and *ivol*)	Sharpe Ratio	1.86	1.50	1.74
	Portfolio returns	0.034	0.028	0.046
	Standard deviation	0.018	0.019	0.026
	Maximum Drawdown (%)	-0.83	-1.28	-1.49
	Omega Ratio	6.76	3.64	5.87

Table 2: Out-of-sample Sharpe ratios, returns and standard deviations of long–short portfolios according to the number of latent factors for the three competing models, with a single covariate (*mom* : momentum six months) or two covariates (*mom* and *ivol* : idiosyncratic volatility) : constant loadings as in Bai (2003) (PCA), loadings defined as linear function of covariates as in Kelly et al. (2020) (IPCA) and our own extended semiparametric loading maps model (GFLFM).

Table 2 reports the out-of-sample performance of long-short portfolios constructed using different models and sets of characteristics, with $r \in \{1, 2, 3\}$ latent factors. The table displays Sharpe ratios, portfolio returns, standard deviations, maximum drawdowns and Omega ratios with a threshold set at zero.

Our main insights are as follows. First, the PCA benchmark provides only modest economic value: its Sharpe ratios range from 0.49 to 0.57, its portfolio returns are around 1-2% per month, and its Omega ratios remain close to 1.5. This indicates that constant-loading models are able to capture only limited predictive content from asset returns only (no covariates).

Second, IPCA based on momentum (*mom*) slightly improves upon PCA, with a Sharpe ratio of 0.71, a monthly return of 3.1%, and an Omega ratio of 1.93, though at the cost of relatively high volatility and drawdowns. When both momentum and idiosyncratic volatility (*mom* and *ivol*) are included, IPCA achieves Sharpe ratios of 0.66-0.77 and Omega ratios around 1.8, but again suffers from large drawdowns in excess of 5%.

Third, the proposed GFLFM methodology delivers substantially stronger performance. With momentum only, its Sharpe ratios range from 1.29 to 1.56, its portfolio returns between 1.4% and 4.3% per month, and its Omega ratios between 2.98 and 4.90, while its drawdowns remain contained below -2.6%. When both momentum and volatility are used, GFLFM dominates all competing models: its Sharpe ratios range from 1.50 to 1.86 and its Omega ratios are exceptionally high, (between 3.64 and 6.76), when its drawdowns are limited to about -1.5%.

One might wonder whether it is still necessary to consider latent factors given the availability of the classical Fama–French three-factor model. Table 3 reports the out-of-sample performance of long–short portfolios under the Fama–French specification. Compared with Table 2, which incorporates latent factors, the performance improvements are substantial. For the GFLFM approach, Sharpe ratios increase from roughly 0.68 under the Fama–French model to values above 1.5, while Omega ratios rise from about 1.75 to levels between 3 and 6. This comparison underscores the flexibility of the GFLFM framework and its ability to deliver sizable economic gains relative to standard benchmarks.

Fama-French 3-factors model	Sharpe Ratio	0.68
	Portfolio returns	0.032
	Standard deviation	0.047
	Maximum Drawdown (%)	-8.57
	Omega Ratio	1.75

Table 3: Out-of-sample Sharpe ratios, returns and standard deviations of long–short portfolios for the Fama-French three factors model.

6. Conclusion

In this paper, we developed a methodology for estimating panel factor models in which both observable and latent factors affect the dependent variable through smooth, nonlinear, covariate-dependent loadings. These loading maps may vary from one individual to another. In this new specification called (GFLFM), we encompass and extend many widely used factor models. Unlike standard approaches that rely exclusively on either observable or latent factors, our framework integrates both types simultaneously, allowing them to jointly shape the behavior of the outcome variable. Moreover, traditional factor models typically assume constant factor loadings across units and over time, an assumption that may obscure important heterogeneities driven by observable characteristics. Our approach addresses this limitation by allowing the influence of both latent and observable factors to vary flexibly with observed covariates.

We proposed a two-step estimation strategy. In the first step, the loading functions associated with the observable factors are estimated nonparametrically (e.g., via kernel smoothing). In the second step, we estimate the latent factors and their covariate-dependent loadings using localized PCA techniques within a "large N and large T" framework. Convergence rates in probability for our estimators have been established, but their asymptotic normality has only been briefly discussed. Formally deriving it would go beyond the scope of this already lengthy paper, given the high level of technicality required and the unavoidable complexity of the exposition. We illustrated the empirical relevance of our methodology through simulations and an empirical application to stock returns. We also introduced a data-driven procedure for selecting the optimal bandwidth in the kernel-based PCA. Our simulations show that our method yields accurate estimates of the factors and loadings and even more precise estimates of the common components. Furthermore, in a standard financial portfolio construction exercise, our model delivers superior empirical performance compared with the classical latent factor model of Bai (2003) and the IPCA model of Kelly et al. (2020).

Our general framework opens the way to the study of particular specifications that are useful in practice. For example, one could impose that the loading maps are identical within certain buckets and develop a clustering procedure based on our estimated loading functions. Moreover, our method can only accommodate low-dimensional covariate vectors of size q, due to the curse of dimensionality. A further challenge would be to identify a limited number of relevant covariates when many potential candidates are available. However, a direct treatment of high dimensional covariates ($q >> 1$) would require different tools and methodology.

References

Alessi, L., Barigozzi, M., Capasso, M., 2010. Improved penalization for determining the number of factors in approximate factor models. Statistics & Probability Letters 80, 1806–1813.

Bai, J., 2003. Inferential theory for factor models of large dimensions. Econometrica 71, 135–171.

Bai, J., 2009. Panel data models with interactive fixed effects. Econometrica 77, 1229–1279.

Bai, J., Li, K., 2014. Theory and methods of panel data models with interactive effects. Annals of Statistics 42, 142–170.

Bai, J., Ng, S., 2002. Determining the number of factors in approximate factor models. Econometrica 70, 191–221.

Bai, J., Ng, S., 2013. Principal components estimation and identification of static factors. Journal of Econometrics 176, 18–29.

Bai, J., Wang, P., 2016. Econometric analysis of large factor models. Annual Review of Economics 8, 53–80.

Barigozzi, M., Hallin, M., 2024. Dynamic factor models: A genealogy, in: Partial Identification in Econometrics and Related Topics. Springer, pp. 3–24.

Barillas, F., Shanken, J., 2017. Which alpha? The Review of Financial Studies 30, 1316–1338.

Bosq, R., 1998. Nonparametric Statistics for Stochastic Processes. Springer.

Bryzgalova, S., Pelger, M., Zhu, J., 2025. Forest through the trees: Building cross-sections of stock returns. The Journal of Finance 80, 2447–2506.

Cai, Z., Fan, J., Yao, Q., 2000. Functional-coefficient regression models for nonlinear time series. Journal of the American Statistical Association 95, 941–956.

Chen, M., Fernández-Val, I., Weidner, M., 2021. Nonlinear factor models for network and panel data. Journal of Econometrics 220, 296–324.

Chen, Q., Roussanov, N., Wang, X., 2023. Semiparametric conditional factor models: Estimation and inference. National Bureau of Economic Research .

Cheng, T., Dong, C., Gao, J., Linton, O., 2024. GMM estimation for high-dimensional panel data models. Journal of Econometrics 244, 105853.

Connor, G., Hagmann, M., Linton, O., 2007. Efficient semiparametric estimation of the fama-french model and extensions. Econometrica 75, 683–709.

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

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

Fama, E.F., French, K.R., 1993. Common risk factors in the returns on stocks and bonds. Journal of financial economics 33, 3–56.

Fama, E.F., French, K.R., 1995. Size and book-to-market factors in earnings and returns. The journal of finance 50, 131–155.

Fan, J., Li, K., Liao, Y., 2021. Recent developments in factor models and applications in econometric learning. Annual Review of Financial Economics 13, 401–430.

Fan, J., Liao, Y., Wang, W., 2016. Projected principal component analysis in factor models. Annals of Statistics 44, 219–254.

Feng, G., Giglio, S., Xiu, D., 2020. Taming the factor zoo: A test of new factors. The Journal of Finance 75, 1327–1370.

Feng, G., Peng, B., Su, L., Yang, T.T., 2019. Semi-parametric single-index panel data models with interactive fixed effects: Theory and practice. Journal of Econometrics 212, 607–622.

Fermanian, J.D., Thélot, L., 2025. Latent factor models with functional single-index loadings. Available at SSRN 5548704 .

Forni, M., Hallin, M., Lippi, M., Reichlin, L., 2000. The generalized dynamic-factor model: Identification and estimation. Review of Economics and statistics 82, 540–554.

Franklin, J., 2012. Matrix Theory. Dover Books.

Gagliardini, P., Ma, H., 2019. Extracting statistical factors when betas are time-varying. Swiss Finance Institute Research Paper .

Gagliardini, P., Ossola, E., Scaillet, O., 2019. A diagnostic criterion for approximate factor structure. Journal of Econometrics 212, 503–521.

Gagliardini, P., Ossola, E., Scaillet, O., 2020. Estimation of large dimensional conditional factor models in finance. Handbook of econometrics 7, 219–282.

Gu, S., Kelly, B., Xiu, D., 2021. Autoencoder asset pricing models. Journal of Econometrics 222, 429–450.

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

Hansen, B.E., 2008. Uniform convergence rates for kernel estimation with dependent data. Econometric Theory 24, 726–748.

Harvey, C.R., Liu, Y., Zhu, H., 2016. ... and the cross-section of expected returns. Review of Financial Studies 29, 5–68.

Kan, R., Zhang, C., 1999. Two-pass tests of asset pricing models with useless factors. Journal of Finance 54, 203–235.

Keilbar, G., Rodriguez-Poo, J.M., Soberon, A., Wang, W., 2025. A projection based approach for interactive fixed effects panel data models. arXiv , 2201.11482.

Kelly, B., Pruitt, S., Su, Y., 2020. Instrumented principal component analysis. SSRN 2983919.

Kelly, B.T., Pruitt, S., Su, Y., 2019. Characteristics are covariances: A unified model of risk and return. Journal of Financial Economics 134, 501–524.

Kozak, S., Nagel, S., Santosh, S., 2020. Shrinking the cross-section. Journal of Financial Economics 135, 271–292.

Lewellen, J., Nagel, S., Shanken, J., 2010. A skeptical appraisal of asset-pricing tests. Review of Financial Studies 23, 1417–1463.

Lütkepohl, H., 1997. Handbook of matrices. John Wiley & Sons.

Ma, S., Linton, O., Gao, J., 2021. Estimation and inference in semiparametric quantile factor models. Journal of Econometrics 222, 295–323.

Masry, E., 1996. Multivariate local polynomial regression for time series: uniform strong consistency and rates. Journal of Time Series Analysis 17, 571–599.

Merlevède, F., Peligrad, M., Rio, E., 2011. A bernstein type inequality and moderate deviations for weakly dependent sequences. Probability Theory and Related Fields 151, 435–474.

Moon, H.R., Weidner, M., 2015. Linear regression for panel with unknown number of factors as interactive fixed effects. Econometrica 83, 1543–1579.

Moon, H.R., Weidner, M., 2017. Dynamic linear panel regression models with interactive fixed effects. Econometric Theory 33, 158–195.

Motta, G., Hafner, C.M., von Sachs, R., 2011. Locally stationary factor models: Identification and nonparametric estimation. Econometric Theory 27, 1279–1319.

Pelger, M., Xiong, R., 2022. State-varying factor models of large dimensions. Journal of Business & Economic Statistics 40, 1315–1333.

Pelger, M., Xiong, R., 2023. Large-dimensional dynamic factor models with missing data. Journal of Econometrics .

Peligrad, M., 1996. On the asymptotic normality of sequences of weak dependent random variables. Journal of Theoretical Probability 9, 703–715.

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

Robinson, P.M., 1983. Nonparametric estimators for time series. Journal of Time Series Analysis 4, 185–207.

Shiryaev, A., 1996. *Probability.* Springer.

Stewart, G., Sun, J.g., 1990. Matrix Perturbation Theory. Academic Press.

Stock, J., Watson, M., 2011. Dynamic Factor Models. Oxford University Press, Oxford.

Stock, J.H., Watson, M.W., 2002. Forecasting using principal components from a large number of predictors. Journal of the American Statistical Association 97, 1167–1179.

Stock, J.H., Watson, M.W., 2016. Dynamic factor models, factor-augmented vector autoregressions, and structural vector autoregressions in macroeconomics, in: Handbook of macroeconomics. Elsevier. volume 2, pp. 415–525.

Su, L., Wang, X., 2017. On time-varying factor models: estimation and testing. Journal of Econometrics 198, 84–101.

Van Der Vaart, A.W., Wellner, J.A., 1996. Weak convergence, in: Weak convergence and empirical processes: with applications to statistics. Springer, pp. 16–28.

Welch, I., Goyal, A., 2008. A comprehensive look at the empirical performance of equity premium prediction. The Review of Financial Studies 21, 1455–1508.

Wendland, H., 2004. Scattered data approximation. volume 17. Cambridge university press.

Xiong, R., Pelger, M., 2023. Large dimensional latent factor modeling with missing observations and applications to causal inference. Journal of Econometrics 233, 271–301.

Zhang, L., Zhou, W., Wang, H., 2021. A semiparametric latent factor model for large scale temporal data with heteroscedasticity. Journal of Multivariate Analysis 186, 104786.

Appendix A. Proofs of consistency results

In our proofs, we denote C_0, C_1, etc., some generic constants that do not depend on (N, T, h). The dependence of some quantities as the weights w_{it}, N_{ts}, etc., w.r.t. z will be implicit, to lighten notations.

Proof of Theorem 1. Recall $\tilde{\Sigma} = [\tilde{\Sigma}_{ts}]_{1\leqslant s,t\leqslant T}$ and $\tilde{\Sigma}_{ts} = N_{ts}^{-1}\sum_{i=1}^N w_{it}w_{is}\widehat{u}_{it}\widehat{u}_{is}$. Due to the definition of $\widehat{\boldsymbol{F}}$ and (3.11), we have

$$\frac{1}{T}\sum_{s=1}^T\sum_{i=1}^N \frac{w_{it}w_{is}}{N_{ts}}\widehat{u}_{it}\widehat{u}_{is}\widehat{F}_s^\top = \widehat{F}_t^\top V_z, \tag{A.1}$$

for all $t \in \{1, \dots, T\}$. However, for all $s, t \in \{1, \dots, T\}$, we can expand $\widehat{u}_{it}\widehat{u}_{is}$ as

$$\begin{aligned}
\widehat{u}_{it}\widehat{u}_{is} &= (X_{it} - \widehat{C}_i(Z_{it})^\top G_t)(X_{is} - \widehat{C}_i(Z_{is})^\top G_s) \\
&= (u_{it} + \Delta\widehat{C}_i(Z_{it})^\top G_t)(u_{is} + \Delta\widehat{C}_i(Z_{is})^\top G_s) \\
&= u_{it}u_{is} + u_{it}\Delta\widehat{C}_i(Z_{is})^\top G_s + u_{is}\Delta\widehat{C}_i(Z_{it})^\top G_t + \Delta\widehat{C}_i(Z_{it})^\top G_t\Delta\widehat{C}_i(Z_{is})^\top G_s,
\end{aligned} \tag{A.2}$$

by recalling $u_{it} = B_i(Z_{it})^\top F_t + \varepsilon_{it}$ and denoting $\Delta\widehat{C}_i(Z_{it}) := C_i(Z_{it}) - \widehat{C}_i(Z_{it})$. By expanding the product $u_{it}u_{is}$, we get from (A.1)

$$\widehat{F}_t - \widehat{\pi}_t^\top F_t = \widehat{\zeta}_t + \widehat{\eta}_t + \widehat{\tau}_t + \widehat{\mu}_{1,t} + \widehat{\mu}_{2,t} + \widehat{\mu}_{3,t}, \text{ with} \tag{A.3}$$

$$\begin{aligned}
\widehat{\pi}_t &:= \frac{1}{T}\sum_{i=1}^N\sum_{s=1}^T \frac{w_{it}w_{is}}{N_{ts}} B_i(Z_{it})B_i(Z_{is})^\top F_s\widehat{F}_s^\top V_z^{-1}, \\
\widehat{\tau}_t^\top &:= \frac{1}{T}\sum_{i=1}^N\sum_{s=1}^T \frac{w_{it}w_{is}}{N_{ts}} \varepsilon_{it}\varepsilon_{is}\widehat{F}_s^\top V_z^{-1}, \\
\widehat{\zeta}_t^\top &:= \frac{1}{T}\sum_{i=1}^N\sum_{s=1}^T \frac{w_{it}w_{is}}{N_{ts}} B_i(Z_{it})^\top F_t\varepsilon_{is}\widehat{F}_s^\top V_z^{-1}, \\
\widehat{\eta}_t^\top &:= \frac{1}{T}\sum_{i=1}^N\sum_{s=1}^T \frac{w_{it}w_{is}}{N_{ts}} B_i(Z_{is})^\top F_s\varepsilon_{it}\widehat{F}_s^\top V_z^{-1}, \\
\widehat{\mu}_{1,t}^\top &:= \frac{1}{T}\sum_{i=1}^N\sum_{s=1}^T \frac{w_{it}w_{is}}{N_{ts}} u_{it}(\Delta\widehat{C}_i(Z_{is})^\top G_s)\widehat{F}_s^\top V_z^{-1}, \\
\widehat{\mu}_{2,t}^\top &:= \frac{1}{T}\sum_{i=1}^N\sum_{s=1}^T \frac{w_{it}w_{is}}{N_{ts}} u_{is}(\Delta\widehat{C}_i(Z_{it})^\top G_t)\widehat{F}_s^\top V_z^{-1}, \text{ and} \\
\widehat{\mu}_{3,t}^\top &:= \frac{1}{T}\sum_{i=1}^N\sum_{s=1}^T \frac{w_{it}w_{is}}{N_{ts}} (\Delta\widehat{C}_i(Z_{it})^\top G_t)(\Delta\widehat{C}_i(Z_{is})^\top G_s)\widehat{F}_s^\top V_z^{-1}.
\end{aligned}$$

Since we have assumed that the r largest eigenvalues of $\tilde{\Sigma}$ are bounded below a.s.

(Assumption 4), $\|V_z^{-1}\|$ is bounded above almost surely. Note that all the latter remainder terms are double sums over $i \in \{1,\dots,N\}$ and $s \in \{1,\dots,T\}$. The case $s = t$ will necessitate a particular treatment, but the corresponding terms will be finally negligible.

Study of $\widehat{\tau}_t$: Note that

$$\frac{1}{T}\sum_{t'=1}^{T}\|\widehat{F}_{t'}\|^2 = \frac{1}{T}\sum_{t'=1}^{T}\sum_{j=1}^{r}\widehat{F}_{t'j}^2 = \mathrm{trace}(\widehat{\boldsymbol{F}}^\top\widehat{\boldsymbol{F}}/T) = \mathrm{trace}(I_r) = r.$$

The Cauchy-Schwarz inequality provides

$$\begin{aligned}
\|\widehat{\tau}_t\|^2 &\leqslant \|V_z^{-1}\frac{1}{T}\sum_{i=1}^{N}\sum_{s=1}^{T}\frac{w_{it}w_{is}}{N_{ts}}\varepsilon_{it}\varepsilon_{is}\widehat{F}_s\|^2 \\
&\leqslant \|V_z^{-1}\|^2\Big(\frac{1}{T}\sum_{t'=1}^{T}\|\widehat{F}_{t'}\|^2\Big)\Big(\frac{1}{T}\sum_{s=1}^{T}\sum_{i,k=1}^{N}\frac{w_{it}w_{is}w_{kt}w_{ks}}{N_{ts}^2}\varepsilon_{it}\varepsilon_{is}\varepsilon_{kt}\varepsilon_{ks}\Big) \\
&\leqslant C_0\Big(\sup_{t,s\leqslant T}\frac{N}{N_{ts}}\Big)^2\Big(\frac{1}{TN^2}\sum_{s=1}^{T}\sum_{i,k=1}^{N}w_{it}w_{is}w_{kt}w_{ks}\varepsilon_{it}\varepsilon_{is}\varepsilon_{kt}\varepsilon_{ks}\Big). \qquad \text{(A.4)}
\end{aligned}$$

Due to Lemma 3, we have

$$\sup_{1\leqslant t,s\leqslant T, s\neq t}\frac{N}{N_{ts}} = O_P(1), \text{ and } \sup_{1\leqslant t\leqslant T}\frac{N}{N_{tt}} = O_P(h^q). \qquad \text{(A.5)}$$

From Assumption 5, the process $(W_t)_{t\leqslant T}$ is α-mixing, where $W_t = (\varepsilon_{it}, Z_{it})_{i\geqslant 1}$. Thus, any continuous map of this process is also α-mixing with a smaller mixing coefficient than α_t. Set $g_{ik}(W_t) := w_{it}w_{kt}\varepsilon_{it}\varepsilon_{kt}$. Thus, for any couple (i,k), the process $\big(g_{ik}(W_t)\big)_{t\geqslant 1}$ is α-mixing.

When $i \neq k$, $\mathbb{E}[w_{is}w_{ks}\varepsilon_{is}\varepsilon_{ks}] = 0$ due to (4.1), and we can apply the Davydov's inequality with the mixing coefficients (α_t): for any $\nu > 2$, $i \neq k$ and $s \neq t$, we have

$$\begin{aligned}
&|\mathbb{E}[w_{it}w_{is}w_{kt}w_{ks}\varepsilon_{it}\varepsilon_{is}\varepsilon_{kt}\varepsilon_{ks}]| = |\mathrm{Cov}\big(g_{ik}(W_t), g_{ik}(W_s)\big)| \\
&\quad\leqslant 2\alpha_{|t-s|}^{1-2/\nu}\|g_{ik}(W_t)\|_\nu\|g_{ik}(W_s)\|_\nu \\
&\quad\leqslant 2\exp\big(-(1-\frac{2}{\nu})C_\alpha|t-s|\big)\mathbb{E}\big[|g_{ik}(W_t)|^\nu\big]^{2/\nu} \\
&\quad\leqslant \frac{2}{h^{4q}}\exp\big(-(1-\frac{2}{\nu})C_\alpha|t-s|\big)\Big\{\int|K|^\nu\big(\frac{x-z}{h}\big)|K|^\nu\big(\frac{y-z}{h}\big)\,\big(f_i(x)f_k(y)h_{ik,\nu}\big)(x,y)\,dx\,dy\Big\}^{2/\nu},
\end{aligned}$$

denoting $h_{ik,\nu}(x,y) := \mathbb{E}\big[|\varepsilon_{it}\varepsilon_{kt}|^\nu|Z_{it} = x, Z_{kt} = y\big]$ and using the independence between Z_{it} and Z_{kt}. By stationarity, $h_{ik,\nu}$ does not depend on t and it is uniformly bounded above from (4.7). Thus, by Jensen inequality, we obtain

$$\begin{aligned}
&|\mathbb{E}[\frac{1}{TN^2}\sum_{s=1,s\neq t}^{T}\sum_{i,k=1,i\neq k}^{N} w_{it}w_{is}w_{kt}w_{ks}\varepsilon_{it}\varepsilon_{is}\varepsilon_{kt}\varepsilon_{ks}]| \\
&\leqslant \frac{C_0}{Th^{4q(1-1/\nu)}}\sum_{s=1,s\neq t}^{T}\exp\big(-(1-\frac{2}{\nu})C_\alpha|t-s|\big)\Big(\frac{1}{N}\sum_{i=1}^{N}\sup_{\{u;\|u\|<h\}} f_i^{2/\nu}(z+u)\Big)^2 \\
&\leqslant \frac{C_0}{Th^{4q(1-1/\nu)}}\sum_{s=1,s\neq t}^{T}\exp\big(-(1-\frac{2}{\nu})C_\alpha|t-s|\big)\Big(\frac{1}{N}\sum_{i=1}^{N}\sup_{\{u;\|u\|<h\}} f_i(z+u)\Big)^{4/\nu} \\
&= O(\frac{1}{Th^{4q(1-1/\nu)}}).
\end{aligned}$$

If $i=k$ and $s\neq t$, the expectation of the corresponding term in (A.4) is

$$\mathbb{E}\big[\frac{1}{TN^2}\sum_{s=1,s\neq t}^{T}\sum_{i=1}^{N} w_{it}^2w_{is}^2\varepsilon_{it}^2\varepsilon_{is}^2\big] = \mathbb{E}\big[\frac{1}{TN^2}\sum_{s=1,s\neq t}^{T}\sum_{i=1}^{N} w_{it}^2w_{is}^2\mathbb{E}[\varepsilon_{it}^2\varepsilon_{is}^2|Z_{it},Z_{is}]\big] = O(\frac{1}{Nh^{2q}}),$$

by the same change of variables as above, recalling that $(x,y)\mapsto\mathbb{E}\big[\varepsilon_{it}^2\varepsilon_{is}^2|Z_{it}=x,Z_{is}=y\big]$ is index-regular of order zero (Assumption 8).

When $i\neq k$ and $s=t$, we have similarly

$$\mathbb{E}\big[\frac{1}{TN^2}\sum_{i\neq k} w_{it}^2w_{kt}^2\varepsilon_{it}^2\varepsilon_{kt}^2\big] = \mathbb{E}\big[\frac{1}{TN^2}\sum_{i\neq k} w_{it}^2w_{kt}^2\mathbb{E}[\varepsilon_{it}^2\varepsilon_{kt}^2|Z_{it},Z_{kt}]\big] = O(\frac{1}{Th^{2q}}),$$

because $(x,y)\mapsto\mathbb{E}\big[\varepsilon_{it}^2\varepsilon_{kt}^2|Z_{it}=x,Z_{kt}=y\big]$ is uniformly bounded in a neighborhood of (z,z) because of (4.7).

When $i=k$ and $s=t$, since $\sup_N N^{-1}\sum_{i=1}^N\mathbb{E}[\varepsilon_{it}^4]<\infty$, we get

$$\mathbb{E}\big[\frac{1}{TN^2}\sum_{i=1}^{N} w_{it}^4\varepsilon_{it}^4\big] = O(\frac{1}{NTh^{4q}}).$$

Recalling (A.4) and noting that $1-1/\nu>1/2$, we have obtained

$$\|\widehat{\tau}_t\|^2 = O_P(\max(\frac{1}{Nh^{2q}},\frac{1}{Th^{4q(1-1/\nu)}})),$$

by Markov inequality. Moreover, noting that $T^{-1}\sum_{s,t=1}^T\exp(-C_\alpha|t-s|)<\infty$, we get the same upper bound for $T^{-1}\sum_{t=1}^T\|\tau_t\|^2$ by the same reasoning.

Remark 11. *If the innovations ε_{it} are uniformly bounded a.s., i.e. if $\sup_{i,t}|\varepsilon_{it}|<M_\varepsilon$ a.s. for some constant M_ε, then we can apply the version of the Davydov's inequality with sup-norms. In such a case, it is possible to set $\nu=2$ and the latter upper bounds become $O_P\big(1/(Th^{2q})+1/(Nh^{2q})\big)$.*

Study of $\widehat{\zeta}_t$: Now, let us deal with $\widehat{\zeta}_t$. The Cauchy-Schwarz inequality provides

$$\begin{aligned}
\|\widehat{\zeta}_t\|^2 &= \|V_z^{-1}\frac{1}{T}\sum_{i=1}^N\sum_{s=1}^T \frac{w_{it}w_{is}}{N_{ts}} B_i(Z_{it})^\top F_t \varepsilon_{is}\widehat{F}_s\|^2 \\
&\leqslant \|V_z^{-1}\|^2\Big(\frac{1}{T}\sum_{t'=1}^T \|\widehat{F}_{t'}\|^2\Big)\Big(\frac{1}{T}\sum_{s=1}^T\big(\sum_{i=1}^N \frac{w_{it}w_{is}}{N_{ts}}\varepsilon_{is}B_i(Z_{it})^\top F_t\big)^2\Big) \\
&\leqslant C_0\big(\sup_{t,s\leqslant T}\frac{N}{N_{ts}}\big)^2 \times \frac{1}{TN^2}\sum_{s=1}^T\sum_{i,k=1}^N w_{it}w_{is}w_{kt}w_{ks}\varepsilon_{is}\varepsilon_{ks}(B_i(Z_{it})^\top F_t)(B_k(Z_{kt})^\top F_t).
\end{aligned}$$

Due to (A.5), the first term $(\sup_{t,s\leqslant T} N/N_{ts})^2 = O_P(1)$ and it is sufficient to evaluate the expectation of the latter triple sum above (that is nonnegative). Recall the orthogonality condition (4.1) in Assumption 5. When $s \neq t$, deduce

$$\begin{aligned}
&|\mathbb{E}\Big[\frac{1}{TN^2}\sum_{s=1,s\neq t}^T\sum_{i,k=1}^N w_{it}w_{is}w_{kt}w_{ks}\varepsilon_{is}\varepsilon_{ks}(B_i(Z_{it})^\top F_t)(B_k(Z_{kt})^\top F_t)\Big]| \\
&\leqslant \mathbb{E}\Big[\frac{M_B^2}{TN^2}\sum_{s;s\neq t}\sum_{i=1}^N w_{it}^2 w_{is}^2 \mathbb{E}[\|F_t\|^2\varepsilon_{is}^2|Z_{is},Z_{it}]\Big] \\
&\leqslant \frac{C_0}{TN^2h^{2q}}\sum_{s;s\neq t}\sum_{i=1}^N \sup_{\{(u,v);\|u\|\leqslant h,\|v\|\leqslant h\}}\big(h_i^{(t,s)}f_{i,t-s}\big)(z+u,z+v) = O\Big(\frac{1}{Nh^{2q}}\Big),
\end{aligned}$$

by noting that the map $h_i^{(t,s)} : (x,y) \mapsto \mathbb{E}[\|F_t\|^2\varepsilon_{is}^2|Z_{is}=x, Z_{it}=y]$ is index-regular of order zero from Assumption 8 and using the boundedness of the loadings. If $s=t$, the same arguments as in the case $s\neq t$ can be applied:

$$\begin{aligned}
&|\mathbb{E}\Big[\frac{1}{TN^2}\sum_{i,k=1}^N w_{it}^2w_{kt}^2\varepsilon_{it}\varepsilon_{kt}(B_i(Z_{it})^\top F_t)(B_k(Z_{kt})^\top F_t)\Big]| \leqslant \mathbb{E}\Big[\frac{M_B^2}{TN^2}\sum_{i=1}^N w_{it}^4\mathbb{E}[\|F_t\|^2\varepsilon_{it}^2|Z_{it}]\Big] \\
&\leqslant \frac{C_1}{TN^2h^{3q}}\sum_{i=1}^N\sup_{\{u;\|u\|\leqslant h\}}\big(h_if_i\big)(z+u) = O\Big(\frac{1}{TNh^{3q}}\Big),
\end{aligned}$$

using that the map $h_i : (x,y) \mapsto \mathbb{E}[\|F_t\|^2\varepsilon_{it}^2|Z_{it}=x]$ is index-regular of order zero from Assumption 8. Globally, we have obtained $\|\widehat{\zeta}_t\|^2 = O_P\big(1/(Nh^{2q})\big)$. By exactly similar reasoning, we get the same rate for $T^{-1}\sum_{t=1}^T\|\widehat{\zeta}_t\|^2$.

Study of $\widehat{\eta}_t$: using exactly the same arguments and the same technical assumptions as for $\widehat{\zeta}_t$, we obtain $\mathbb{E}[\|\widehat{\eta}_t\|^2] = O_P(N^{-1}h^{-2q})$ and $T^{-1}\sum_{t=1}^T\|\widehat{\eta}_t\|^2 = O_P(N^{-1}h^{-2q})$.

Study of $\widehat{\mu}_{1,t}$: The Cauchy-Schwarz inequality and (3.5) lead to

$$\begin{aligned}
\|\widehat{\mu}_{1,t}\|^2 &= \|V_z^{-1}\frac{1}{T}\sum_{i=1}^N\sum_{s=1}^T \frac{w_{it}w_{is}}{N_{ts}}u_{it}\Delta\widehat{C}(Z_{is})^\top G_s\widehat{F}_s\|^2 \\
&\leqslant 2\|V_z^{-1}\frac{1}{T}\sum_{i=1}^N\sum_{s=1}^T \frac{w_{it}w_{is}}{N_{ts}}(B_i(Z_{it})^\top F_t)\Delta\widehat{C}(Z_{is})^\top G_s\widehat{F}_s\|^2 \\
&+ 2\|V_z^{-1}\frac{1}{T}\sum_{i=1}^N\sum_{s=1}^T \frac{w_{it}w_{is}}{N_{ts}}\varepsilon_{it}\Delta\widehat{C}(Z_{is})^\top G_s\widehat{F}_s\|^2 \\
&\leqslant 2\|V_z^{-1}\|^2\Big(\frac{1}{T}\sum_{t'=1}^T\|\widehat{F}_{t'}\|^2\Big)\Big(\frac{1}{T}\sum_{s=1}^T\|\sum_{i=1}^N\frac{w_{it}w_{is}}{N_{ts}}(B_i(Z_{it})^\top F_t)\Delta\widehat{C}(Z_{is})^\top G_s\|^2 \\
&+ \frac{1}{T}\sum_{s=1}^T\|\sum_{i=1}^N\frac{w_{it}w_{is}}{N_{ts}}\varepsilon_{it}\Delta\widehat{C}(Z_{is})^\top G_s\|^2\Big) \\
&\leqslant \frac{C_0}{T}\sum_{s=1}^T\Big(\sum_{i=1}^N\frac{w_{it}w_{is}}{N_{ts}}\|\Delta\widehat{C}(Z_{is})\|^2\|G_s\|^2\Big)\Big(\sum_{i=1}^N\frac{w_{it}w_{is}}{N_{ts}}\{(B_i(Z_{it})^\top F_t)^2+\varepsilon_{it}^2\}\Big) \\
&\leqslant \frac{O_P(\theta_{N,T}^2)}{T}\sum_{s=1}^T\|G_s\|^2\{\|F_t\|^2+\sum_{i=1}^N\frac{w_{it}w_{is}}{N_{ts}}\varepsilon_{it}^2\}.
\end{aligned}$$

Note that we have invoked the uniform boundedness of the the maps $B_i(\cdot)$ in a neighborhood of z. Since the map $H_i^{(t,s)}:(x,y)\mapsto\mathbb{E}[\varepsilon_{it}^2\|G_s\|^2|Z_{is}=x,Z_{it}=y]$ is index-regular of order zero from Assumption 8 with $s\neq t$, deduce

$$\frac{1}{T}\sum_{s=1,s\neq t}^T\sum_{i=1}^N\frac{w_{it}w_{is}}{N_{ts}}\varepsilon_{it}^2\|G_s\|^2 = O_P(1).$$

The term corresponding to $s=t$ is

$$\frac{O_P(\theta_{N,T}^2)\|G_t\|^2}{T}\sum_{i=1}^T\frac{w_{it}^2}{N_{tt}}\varepsilon_{it}^2 = \frac{O_P(\theta_{N,T}^2)\|G_t\|^2}{T}\times\frac{N}{N_{tt}h^{2q}}\times\frac{1}{N}\sum_{i=1}^N\varepsilon_{it}^2 = \frac{O_P(\theta_{N,T}^2)\|G_t\|^2}{Th^q}$$

and is then $o_P(\theta_{N,T}^2)$. Moreover, $T^{-1}\sum_{t=1}^T\|G_t\|^2=O_P(1)$ (Assumption 7). Globally, we have obtained $\|\widehat{\mu}_{1,t}\|^2=\|F_t\|^2O_P(\theta_{N,T}^2)$. We get $T^{-1}\sum_{t=1}^T\|\widehat{\mu}_{1,t}\|^2=O_P(\theta_{N,T}^2)$ similarly, using the index-regularity of the maps $x\mapsto H_i^{(t,t)}(x,x)$.

Study of $\widehat{\mu}_{2,t}$: The treatment is similar to that of $\widehat{\mu}_{1,t}$.

Study of $\widehat{\mu}_{3,t}$: Similarly, the Cauchy-Schwarz inequality, (3.5) and Assumption 7 yield

$$\begin{aligned}
\|\widehat{\mu}_{3,t}\|^2 &= \|V_z^{-1}\frac{1}{T}\sum_{i=1}^N\sum_{s=1}^T\frac{w_{it}w_{is}}{N_{ts}}\Delta\widehat{C}_i(Z_{is})^\top G_s\Delta\widehat{C}(Z_{it})^\top G_t\widehat{F}_s\|^2 \\
&\leqslant \|V_z^{-1}\|^2\Big(\frac{1}{T}\sum_{t'=1}^T\|\widehat{F}_{t'}\|^2\Big)\Big(\frac{1}{T}\sum_{s=1}^T\{\sum_{i=1}^N\frac{w_{it}w_{is}}{N_{ts}}\Delta\widehat{C}_i(Z_{is})^\top G_s\Delta\widehat{C}(Z_{it})^\top G_t\}^2\Big) \\
&\leqslant \frac{C_0}{T}\sum_{s=1}^T\big(\sum_{i=1}^N\frac{w_{it}w_{is}}{N_{ts}}(\Delta\widehat{C}(Z_{it})^\top G_t)^2\big)\big(\sum_{i=1}^N\frac{w_{it}w_{is}}{N_{ts}}(\Delta\widehat{C}_i(Z_{is})^\top G_s)^2\big) \\
&= O_P(\theta_{N,T}^4)\|G_t\|^2.
\end{aligned}$$

It follows also that $T^{-1}\sum_{t=1}^T\|\widehat{\mu}_{3,t}\|^2 = O_P(\theta_{N,T}^4)$.

Finally, we have shown that

$$\|\widehat{F}_t - \widehat{\pi}_t^\top F_t\|^2 = O_P\Big(\frac{1}{Nh^{2q}} + \frac{1}{Th^{4q(1-1/\nu)}} + \theta_{N,T}^2\Big), \tag{A.6}$$

and similarly for $T^{-1}\sum_{t=1}^T\|\widehat{F}_t - \widehat{\pi}_t^\top F_t\|^2$. Moreover, from Lemma 4, we know that

$$\sup_{1\leqslant t\leqslant T}\|\widehat{\pi}_t - \widehat{\pi}\| = O_P\big(\frac{\ln N}{\sqrt{N}h^q} + h^2\big), \tag{A.7}$$

with $\widehat{\pi} = (\boldsymbol{B}(z)^\top\boldsymbol{B}(z)/N)(\boldsymbol{F}^\top\widehat{\boldsymbol{F}}/T)V_z$. Conclude by noting that

$$\|\widehat{F}_t - \widehat{\pi}^\top F_t\|^2 \leqslant 2\|\widehat{F}_t - \widehat{\pi}_t^\top F_t\|^2 + 2\|(\widehat{\pi} - \widehat{\pi}_t)F_t\|^2 = O_P\Big(\frac{\ln^2 N}{Nh^{2q}} + \frac{1}{Th^{4q(1-1/\nu)}} + h^4 + \theta_{N,T}^2\Big).$$

For the average of the latter quantities over $t \in \{1,\dots,T\}$, invoke Corollary 6:

$$\begin{aligned}
\frac{1}{T}\sum_{t=1}^T\|\widehat{F}_t - \widehat{\pi}^\top F_t\|^2 &\leqslant \frac{2}{T}\sum_{t=1}^T\|\widehat{F}_t - \widehat{\pi}_t^\top F_t\|^2 + \frac{2}{T}\sum_{t=1}^T\|\widehat{\pi} - \widehat{\pi}_t\|^2\|F_t\|^2 \\
&\leqslant O_P(v_{N,T}) + O_P\big(\frac{\ln^2(N)}{Nh^{2q}} + h^4\big) + \frac{O_P(1)}{T^2}\sum_{t=1}^T\|F_t\|^3\,\|\widehat{F}_t\| \\
&= O_P(v_{N,T}) + \frac{O_P(1)}{T}\Big(\frac{1}{T}\sum_{t=1}^T\|F_t\|^6 \times \frac{1}{T}\sum_{t=1}^T\|\widehat{F}_t\|^2\Big)^{1/2} = O_P(v_{N,T}),
\end{aligned}$$

when $\mathbb{E}[\|F_t\|^6] < \infty$ by Assumption 7. This concludes the proof. □

Proof of Theorem 2. To lighten notations, set $\widehat{\Omega}_i := T^{-1}\sum_{t=1}^T w_{it}\widehat{F}_t\widehat{F}_t^\top$ and $\Omega_i := T^{-1}\sum_{t=1}^T w_{it}F_tF_t^\top$.

Set $\Delta B_{it} := B_i(Z_{it}) - B_i(z)$. Simple algebraic manipulations yield

$$
\begin{aligned}
\widehat{B}_i(z) &= \widehat{\Omega}_i^{-1}\big(\frac{1}{T}\sum_{t=1}^T w_{it}\widehat{F}_t\widehat{u}_{it}\big) \\
&= \widehat{\Omega}_i^{-1}\Big(\frac{1}{T}\sum_{t=1}^T w_{it}\widehat{F}_t\big(F_t - (\widehat{\pi}^{-1})^\top\widehat{F}_t + (\widehat{\pi}^{-1})^\top\widehat{F}_t\big)^\top B_i(z) + \frac{1}{T}\sum_{t=1}^T w_{it}\varepsilon_{it}\widehat{F}_t \\
&+ \frac{1}{T}\sum_{t=1}^T w_{it}\Delta B_{it}^\top F_t\widehat{F}_t + \frac{1}{T}\sum_{t=1}^T w_{it}\Delta\widehat{C}_i(Z_{it})^\top G_t\widehat{F}_t\Big) \\
&= \widehat{\pi}^{-1}B_i(z) + \widehat{\Omega}_i^{-1}\Big(\frac{1}{T}\sum_{t=1}^T w_{it}\widehat{F}_t\big(F_t - (\widehat{\pi}^{-1})^\top\widehat{F}_t\big)^\top B_i(z) + \frac{1}{T}\sum_{t=1}^T w_{it}\varepsilon_{it}\widehat{F}_t \\
&+ \frac{1}{T}\sum_{t=1}^T w_{it}\Delta B_{it}^\top F_t\widehat{F}_t + \frac{1}{T}\sum_{t=1}^T w_{it}\Delta\widehat{C}_i(Z_{it})^\top G_t\widehat{F}_t\Big) \\
&=: \widehat{\pi}^{-1}B_i(z) + \widehat{\Omega}_i^{-1}\big(r_{i1} + r_{i2} + r_{i3} + r_{i4}\big). \qquad \text{(A.8)}
\end{aligned}
$$

As a preliminary stage, let us show that $\widehat{\Omega}_i^{-1}$ is bounded above in probability uniformly w.r.t. i. By simple algebra, we have

$$
\begin{aligned}
\widehat{\Omega}_i &= \frac{1}{T}\sum_{t=1}^T w_{it}(\widehat{F}_t - \widehat{\pi}^\top F_t)(\widehat{F}_t - \widehat{\pi}^\top F_t)^\top + \frac{1}{T}\sum_{t=1}^T w_{it}(\widehat{F}_t - \widehat{\pi}^\top F_t)F_t^\top\widehat{\pi} \\
&+ \frac{1}{T}\sum_{t=1}^T w_{it}\widehat{\pi}^\top F_t(\widehat{F}_t - \widehat{\pi}^\top F_t)^\top + \widehat{\pi}^\top\Omega_i\widehat{\pi} = O_P\big(\frac{v_{N,T}}{h^q}\big) + \widehat{\pi}^\top\Omega_i\widehat{\pi}, \qquad \text{(A.9)}
\end{aligned}
$$

invoking Theorem 1 and the Cauchy-Schwarz inequality. Note that the factor h^q comes from the inequality $w_{it} \leqslant \|K\|_\infty/h^q$ and that $\widehat{\pi} = O_P(1)$ due to Lemma 5. Thus, let us focus on Ω_i. For any i and any couple $(l,l') \in \{1,\dots,r\}^2$, deduce from Assumption 10 that we have

$$
\begin{aligned}
&\mathbb{P}(|w_{it}(F_tF_t^\top)_{ll'} - \mathbb{E}\big[w_{it}(F_tF_t^\top)_{ll'}\big]| > x) \leqslant \mathbb{P}(|w_{it}(F_tF_t^\top)_{ll'}| > x/2) \\
&+ \mathbb{P}(|\mathbb{E}[w_{it}\|F_t\|^2]| > x/2) \\
&\leqslant \mathbb{P}(\|K\|_\infty|(F_tF_t^\top)_{ll'}| > xh^q/2) + \mathbf{1}\big\{\Big|\int K(u)(f_iG_{F,i})(z+hu)\,du\Big| > x/2\big\} \\
&\leqslant \mathbb{P}(\|F_t\|_\infty^2 > xh^q/(2\|K\|_\infty)) + \mathbf{1}\big\{\sup_i \sup_{\{u;\|u\|<h\}} (f_iG_{F,i})(z+u) > x/2\big\} \\
&\leqslant \exp\big(1 - (xh^q)^{\gamma_0/2}/(2b^2\|K\|_\infty)^{\gamma_0/2}\big) + 0.
\end{aligned}
$$

when $x > M_0$, for some constant $M_0 > 0$ that does not depend on i (Assumption 11). Set $1/\gamma = 1 + 2/\gamma_0$. Recall that the sequence $\big(w_{it}(F_tF_t^\top)_{ll'}\big)_t$ is strongly mixing, and its mixing coefficients are not larger than (α_t). For any $\epsilon > 0$, deduce from Merlevède et al.

(2011, Equation (1.8))

$$\mathbb{P}\Big(|\frac{1}{T}\sum_{t=1}^T \{w_{it}(F_tF_t^\top)_{ll'} - \mathbb{E}\big[w_{it}(F_tF_t^\top)_{ll'}\big]\}| > \epsilon\Big) \leqslant C_0T^\eta \exp\big(-C_1 h^{q\gamma}(T\epsilon)^{\gamma/(\gamma+1)}\big),$$

for some $\eta \geqslant 0$, and some constants C_0 and C_1 that do not depend on (i, T, h, l, l'). For any $\epsilon > 0$, this yields

$$\begin{aligned}\mathbb{P}\Big(\sup_{i\in\{1,\dots,N\}} |\frac{1}{T}\sum_{t=1}^T \{w_{it}(F_tF_t^\top)_{ll'} - \mathbb{E}\big[w_{it}(F_tF_t^\top)_{ll'}\big]\}| > \epsilon\Big) \\ \leqslant \quad C_0NT^\eta \exp\Big(-C_1\big\{h^q(T\epsilon)^{1/(\gamma+1)}\big\}^\gamma\Big),\end{aligned} \tag{A.10}$$

that tends to zero when N and T tend to the infinity. Indeed, $1/(\gamma+1) > 1/2$ and $Th^{2q}/\ln T$ tends to the infinity by assumption. Moreover, due to (4.8), we have

$$\begin{aligned}\frac{1}{T}\sum_{t=1}^T \mathbb{E}\big[w_{it}(F_tF_t^\top)_{ll'}\big] &= \mathbb{E}\big[w_{it}(F_tF_t^\top)_{ll'}\big] \\ &= \int K(u)(G_{ll',i}f_i)(z+hu)\,du = (G_{ll',i}f_i)(z) + O(h^2),\end{aligned} \tag{A.11}$$

and the remainder term is uniform w.r.t. i because of (4.8). Since the latter analysis can be led for any couple (l, l'), we deduce from (A.10) and (A.11) that

$$\sup_i \|\Omega_i - \mathbb{E}[F_tF_t^\top | Z_{it} = z] f_i(z)\| = o_P(1). \tag{A.12}$$

Since $\inf_i f_i(z) > 0$ and $\inf_i \min\big(\text{Spectrum of } \mathbb{E}[F_tF_t^\top|Z_{it}=z]\big) > 0$, deduce $\sup_i \|\Omega_i^{-1}\| = O_P(1)$ by the continuity of eigenvalues w.r.t. matrix coefficients. Since $\widehat{\pi}$ and $\widehat{\pi}^{-1}$ are $O_P(1)$ (Lemma 5) and $v_{N,T}/h^q = o(1)$, deduce from (A.9) that $\sup_i \|\widehat{\Omega}_i^{-1}\| = O_P(1)$.

Therefore, the result follows if we show that the remainder terms $r_{i,k}$ (resp. $N^{-1}\sum_{i=1}^N \|r_{ik}\|^2$), $k \in \{1,\dots,4\}$, tend in probability to zero at the rate $v_{N,T}/h^q$ (resp. $v_{N,T}$).

First, simple algebra provides

$$\begin{aligned} r_{i1} &= \frac{1}{T}\sum_{t=1}^T w_{it}(\widehat{F}_t - \widehat{\pi}^\top F_t)\big(\widehat{\pi}^\top F_t - \widehat{F}_t\big)^\top \widehat{\pi}^{-1}B_i(z) + \frac{\widehat{\pi}^\top}{T}\sum_{t=1}^T w_{it}F_t\big(\widehat{\pi}^\top F_t - \widehat{F}_t\big)^\top\widehat{\pi}^{-1}B_i(z) \quad \text{(A.13)} \\ &= O_P\Big(\frac{1}{T}\sum_{t=1}^T \|\widehat{F}_t - \widehat{\pi}^\top F_t\|^2 \frac{M_B}{h^q}\Big) + \widehat{\pi}^\top \bar{r}_{i1}\widehat{\pi}^{-1}B_i(z), \text{ and}\end{aligned}$$

$$\|\bar{r}_{i1}\|^2 \leqslant \frac{1}{T}\sum_{t=1}^T \|\widehat{F}_t - \widehat{\pi}^\top F_t\|^2 \times \frac{1}{T}\sum_{t=1}^T w_{it}^2 \|F_t\|^2$$
$$\leqslant \frac{\|K\|_\infty^2}{Th^q} \sup_{\{z'; \|z-z'\|<a\}} f_i(z')\mathbb{E}[\|F_t\|^2 | Z_{it} = z'] \sum_{t=1}^T \|\widehat{F}_t - \widehat{\pi}^\top F_t\|^2 \big(1 + o_P(1)\big).$$

Deduce, for any i,

$$\|r_{i1}\|^2 = O_P\big(\frac{v_{N,T}^2}{h^{2q}} + \frac{v_{N,T}}{h^q}\big) = O_P\big(\frac{v_{N,T}}{h^q}\big).$$

The order of magnitude of $N^{-1}\sum_{i=1}^N \|r_{i1}\|^2$ is smaller due to the averaging w.r.t. i. Indeed, deduce from (A.13) that

$$\frac{1}{N}\sum_{i=1}^N \|r_{i1}\|^2 \leqslant \frac{2\|\widehat{\pi}^{-1}\|^2 M_B^2}{T^2} \sum_{s,t=1}^T \|\widehat{F}_t - \widehat{\pi}^\top F_t\|^2 \, \|\widehat{F}_s - \widehat{\pi}^\top F_s\|^2 \, \frac{1}{N}\sum_{i=1}^N w_{it}w_{is}$$
$$+ \frac{2\|\widehat{\pi}\|^2 \, \|\widehat{\pi}^{-1}\|^2 M_B^2}{T^2} \sum_{s,t=1}^T \|\widehat{F}_t - \widehat{\pi}^\top F_t\| \, \|\widehat{F}_s - \widehat{\pi}^\top F_s\| \, \|F_t\| \, \|F_s\| \, \frac{1}{N}\sum_{i=1}^N w_{it}w_{is}.$$

Due to Hoeffding's inequality and the mutual independence between the Z_{it}, $i \in \{1, \dots, n\}$, we have

$$\mathbb{P}\Big(\sup_{(s,t)} \frac{1}{N}\big|\sum_{i=1}^N \big\{w_{it}w_{is} - \mathbb{E}[w_{it}w_{is}]\big\}\big| > \epsilon\Big) \leqslant 2T^2 \exp\big(-2Nh^{2q}\epsilon^2/\|K\|_\infty^2\big),$$

for any $\epsilon > 0$. Since $Nh^{2q}/\ln N$ tends to the infinity and invoking (4.4), we easily get

$$\sup_{(s,t);s\neq t} \frac{1}{N}\sum_{i=1}^N w_{it}w_{is} \leqslant \sup_{(s,t);s\neq t} \frac{1}{N}\sum_{i=1}^N \mathbb{E}[w_{it}w_{is}] + o_P(1) = O_P(1). \tag{A.14}$$

Similarly, invoking (4.6), we prove

$$\sup_{1\leqslant t\leqslant T} \frac{1}{N}\sum_{i=1}^N w_{it}^2 \leqslant \sup_{t\leqslant T} \frac{1}{N}\sum_{i=1}^N \mathbb{E}[w_{it}^2] + o_P(1) = O_P(h^{-q}). \tag{A.15}$$

By distinguishing between the cases $s \neq t$ and $s = t$, we obtain

$$\frac{1}{N}\sum_{i=1}^N \|r_{i1}\|^2 \leqslant O_P\Big(\big(\frac{1}{T}\sum_{t=1}^T \|\widehat{F}_t - \widehat{\pi}^\top F_t\|^2\big)^2 h^{-q}\Big)$$
$$+ O_P\Big(\big(\frac{1}{T}\sum_{t=1}^T \|\widehat{F}_t - \widehat{\pi}^\top F_t\| \, \|F_t\|\big)^2\Big) + O_P\Big(\frac{1}{T^2h^q}\sum_{t=1}^T \|\widehat{F}_t - \widehat{\pi}^\top F_t\|^2 \, \|F_t\|^2\Big). \tag{A.16}$$

To manage the last term on the r.h.s. of (A.16), note that Assumption 10 implies

$$\mathbb{P}\big(\sup_{1\leqslant t\leqslant T}\|F_t\|^2 > Th^q\big) \leqslant T\exp\big(1-(Th^q/b^2)^{\gamma_0/2}\big) = o(1),$$

since $Th^{2q}\to\infty$. Thus, $\sup_{1\leqslant t\leqslant T}\|F_t\|^2/(Th^q) = o_P(1)$. As a consequence,

$$\frac{1}{N}\sum_{i=1}^N \|r_{i1}\|^2 = O_P\big(v_{N,T}^2 h^{-q} + v_{N,T} + v_{N,T}\sup_{1\leqslant t\leqslant T}\frac{\|F_t\|^2}{Th^q}\big) = O_P(v_{N,T}).$$

Second, for some arbitrarily chosen index $l\in\{1,\dots,r\}$, we have

$$\frac{1}{T}\sum_{t=1}^T w_{it}\varepsilon_{it}\widehat{F}_{tl} = \widehat{\pi}^\top \frac{1}{T}\sum_{t=1}^T w_{it}\varepsilon_{it}F_{tl} + \frac{1}{T}\sum_{t=1}^T w_{it}\varepsilon_{it}(\widehat{F}_t - \widehat{\pi}^\top F_t)_l =: \widehat{\pi}^\top r_{i2}^{(1)} + r_{i2}^{(2)}.$$

By the same reasoning as for $\bar{r}_{i1}$ above (replace F_t with ε_{it}), $\big(r_{i2}^{(2)}\big)^2 = O_P\big(v_{N,T}h^{-q}\big)$. Similarly, it can be proven that $N^{-1}\sum_{i=1}^N \big(r_{i2}^{(2)}\big)^2 = O_P(v_{N,T})$ when $\sup_{i,t}|\varepsilon_{it}| < \infty$ a.s. because

$$\sup_{(s,t),s\neq t}\frac{1}{N}\Big|\sum_{i=1}^N w_{it}w_{is}\varepsilon_{it}\varepsilon_{is}\Big| = O_P(1), \tag{A.17}$$

is then a direct consequence of (A.14): deduce from (A.17)

$$\begin{aligned}\frac{1}{N}\sum_{i=1}^N \big(r_{i2}^{(2)}\big)^2 &\leqslant \frac{1}{T^2}\sum_{s,t=1}^T \|\widehat{F}_t - \widehat{\pi}^\top F_t\|\,\|\widehat{F}_s - \widehat{\pi}^\top F_s\|\,\Big|\frac{1}{N}\sum_{i=1}^N w_{it}w_{is}\varepsilon_{it}\varepsilon_{is}\Big| \\ &\leqslant \frac{O_P(1)}{T^2}\sum_{s\neq t}\|\|\widehat{F}_t - \widehat{\pi}^\top F_t\|\,\|\widehat{F}_s - \widehat{\pi}^\top F_s\| + \frac{C_0}{h^{2q}T^2}\sum_{t=1}^T \|\widehat{F}_t - \widehat{\pi}^\top F_t\|^2 \\ &\leqslant O_P(1)\Big(\frac{1}{T}\sum_{s=1}^T \|\widehat{F}_t - \widehat{\pi}^\top F_t\|\Big)^2 + \frac{O_P(v_{N,T})}{h^{2q}T} = O_P(v_{N,T}).\end{aligned}$$

Concerning $r_{i2}^{(1)}$, note that $\mathbb{E}\big[w_{it}\varepsilon_{it}F_t\big] = 0$ and

$$\begin{aligned}\mathbb{E}\Big[\big(r_{i2}^{(1)}\big)^2\Big] &= \mathrm{Var}(r_{i2}^{(1)}) = \frac{1}{T^2}\sum_{s,t=1}^T \mathbb{E}\big[w_{it}w_{is}\varepsilon_{it}\varepsilon_{is}F_{tl}F_{sl}\big] \\ &= \frac{1}{T^2}\sum_{s\neq t}\mathrm{Cov}\big(w_{it}\varepsilon_{it}F_{tl}, w_{is}\varepsilon_{is}F_{sl}\big) + \frac{1}{T^2}\sum_{t=1}^T \mathbb{E}\big[w_{it}^2\varepsilon_{it}^2F_{tl}^2\big].\end{aligned} \tag{A.18}$$

Recall that the process $\big((\varepsilon_{it}, Z_{it}, F_t), i\geqslant 1\big)_{t\geqslant 1}$ is α-mixing (Assumption 5). Thus, any continuous map of this process is also α-mixing with a smaller mixing coefficient than α_t. In particular, the process $\big(w_{it}\varepsilon_{it}F_{tl}\big)_{t\leqslant T}$ is α-mixing. When $s\neq t$, apply the Davydov's inequality with the mixing coefficients (α_t): for some $\bar\nu > 2$ and $s \neq t$, we have

$$
\begin{aligned}
|\mathrm{Cov}\big(w_{it}\varepsilon_{it}F_{tl}, w_{is}\varepsilon_{is}F_{sl}\big)| &\leqslant 2\alpha_{|t-s|}^{1-2/\bar\nu}\|w_{it}\varepsilon_{it}F_{tl}\|_{\bar\nu}^2 \\
&\leqslant \frac{2}{h^{2q(1-1/\bar\nu)}}\exp\big(-(1-\frac{2}{\bar\nu})C_\alpha|t-s|\big)\Big\{\int |K|^{\bar\nu}(u)\,(f_iH_{i,\bar\nu,l})(z+hu)\,du\Big\}^{2/\bar\nu}
\end{aligned}
$$

Thus, for some $a > 0$, we obtain

$$
\begin{aligned}
\frac{1}{T^2}\sum_{s\neq t}|\mathrm{Cov}\big(w_{it}\varepsilon_{it}F_{tl}, w_{is}\varepsilon_{is}F_{sl}\big)| &\leqslant \frac{C_0\sup_{\{z';\|z-z'\|<a\}}(f_iH_{i,\bar\nu,l})(z')}{Th^{2q(1-1/\bar\nu)}} \\
\times \sum_{s\neq t}\exp\big(-(1-\frac{2}{\bar\nu})C_\alpha|t-s|\big) &= O\big(\frac{1}{Th^{2q(1-1/\bar\nu)}}\big).
\end{aligned}
$$

Due to stationarity, the corresponding diagonal term in (A.18) is

$$
\frac{1}{T^2}\sum_{t=1}^T\mathbb{E}\big[(w_{it}\varepsilon_{it}F_{tl})^2\big] = \frac{\mathbb{E}\big[(w_{it}\varepsilon_{it}F_{tl})^2\big]}{T} = O\big(\frac{1}{Th^q}\big).
$$

We have obtained

$$
\big(r_{i2}^{(1)}\big)^2 = O_P\big(\frac{1}{Th^{2q(1-1/\bar\nu)}} + \frac{1}{Th^q}\big).
$$

Since the maps $H_{i,\bar\nu,l}$ and $H_{i,2,l}$ are index-regular of order zero, the same rate applies to $N^{-1}\sum_{i=1}^N\big(r_{i2}^{(1)}\big)^2$ and the latter reasoning can be done for any l.

Third, we have

$$
r_{i3} = \frac{\widehat\pi^\top}{T}\sum_{t=1}^T w_{it}(\Delta B_{it}^\top F_t)F_t + \frac{1}{T}\sum_{t=1}^T w_{it}(\Delta B_{it}^\top F_t)\big\{\widehat F_t - \widehat\pi^\top F_t\big\} =: \widehat\pi^\top r_{i3}^{(1)} + r_{i3}^{(2)}.
$$

Since $B_i(\cdot)$ is Lipschitz continuous in a neighborhood of z, Markov's inequality implies

$$
\begin{aligned}
\|r_{i3}^{(2)}\|^2 &\leqslant \frac{1}{T}\sum_{t=1}^T w_{it}^2(\Delta B_{it}^\top F_t)^2 \times \frac{1}{T}\sum_{t=1}^T\|\widehat F_t - \widehat\pi^\top F_t\|^2 \\
&\leqslant \frac{C_0h^2\|K\|_\infty}{Th^q}\sum_{t=1}^T w_{it}\|F_t\|^2 \times O_P(v_{N,T}) = O_P\big(\frac{h^2}{h^q}v_{N,T}\big),
\end{aligned}
$$

since $\sup_{\{u;\|u\|<a\}}(f_iG_{F,i})(z+u) < \infty$ for some $a > 0$.

Concerning $N^{-1}\sum_{i=1}^N \|r_{i3}^{(2)}\|^2$, note that

$$\begin{aligned}
\frac{1}{N}\sum_{i=1}^N \|r_{i3}^{(2)}\|^2 &\leqslant \frac{C_0 h^2}{T^2}\sum_{s,t=1}^T \|\widehat{F}_t - \widehat{\pi}^\top F_t\|\,\|\widehat{F}_s - \widehat{\pi}^\top F_s\|\,\|F_t\|\,\|F_s\|\Big(\frac{1}{N}\sum_{i=1}^N w_{it}w_{is}\Big) \\
&\leqslant \frac{O_P(h^2)}{T^2}\sum_{s,t=1,s\neq t}^T \|\widehat{F}_t - \widehat{\pi}^\top F_t\|\,\|\widehat{F}_s - \widehat{\pi}^\top F_s\|\,\|F_t\|\,\|F_s\| \\
&+ \frac{O_P(h^2)}{T^2 h^q}\sum_{t=1}^T \|\widehat{F}_t - \widehat{\pi}^\top F_t\|^2\,\|F_t\|^2 =: T_1 + O_P(h^2)\times T_2.
\end{aligned}$$

by recalling (A.14) and (A.15). We obviously have

$$T_1 \leqslant \frac{O_P(h^2)}{T}\sum_{t=1}^T \|\widehat{F}_t - \widehat{\pi}^\top F_t\|^2 \times \frac{1}{T}\sum_{t=1}^T \|F_t\|^2 = O_P(h^2 v_{N,T}).$$

Moreover, as in (A.16), we have $T_2 = O_P(v_{N,T})$. We have obtained

$$\frac{1}{N}\sum_{i=1}^N \|r_{i3}^{(2)}\|^2 = O_P(h^2 v_{N,T}).$$

Moreover, to manage $r_{i3}^{(1)}$, fix again an index $l\in\{1,\dots,r\}$ arbitrarily. We have

$$\begin{aligned}
\mathbb{E}\Big[\big(r_{i3l}^{(1)}\big)^2\Big] &= \frac{1}{T^2}\sum_{s\neq t}\mathbb{E}\Big[w_{it}w_{is}(\Delta B_{it}^\top F_t)(\Delta B_{is}^\top F_s)F_{tl}F_{sl}\Big] + \frac{1}{T^2}\sum_{t=1}^T \mathbb{E}\Big[w_{it}^2(\Delta B_{it}^\top F_t)^2 F_{tl}^2\Big] \\
&= \frac{1}{T^2 h^{2q}}\sum_{s\neq t}\sum_{k,k'=1}^r \int K_h(x-z)K_h(y-z)\big(B_{ik}(x)-B_{ik}(z)\big)\big(B_{ik'}(y)-B_{ik'}(z)\big) \\
&\times \big(\bar{G}_{kk'l,i}^{(s,t)} f_{i,t-s}\big)(x,y)\,dx\,dy + O\Big(\frac{h^2}{Th^{2q}}\mathbb{E}[\|F_t\|^4]\Big).
\end{aligned}$$

Note that Assumption 10 implies that all the moments of $\|F_t\|$ are finite, in particular the fourth. By second order limited expansions of $B_i(\cdot)$ (c.f. Assumption 3) and the maps $\bar{G}_{kk'l,i}^{(s,t)} f_{i,t-s}$, note that

$$\begin{aligned}
&\int K_h(x-z)K_h(y-z)\big(B_{ik}(x)-B_{ik}(z)\big)\big(B_{ik'}(y)-B_{ik'}(z)\big)\big(\bar{G}_{kk'l,i}^{(s,t)} f_{i,t-s}\big)(x,y)\,dx\,dy \\
&= h^2\int K(u)K(v)\big(\nabla B_{ik}(z).u + \frac{h}{2}u^\top\nabla^2 B_{ik}(\tilde{z})u\big)\big(\nabla B_{ik'}(z).v + \frac{h}{2}v^\top\nabla^2 B_{ik'}(\bar{z})v\big) \\
&\times \Big\{\big(\bar{G}_{kk'l,i}^{(s,t)} f_{i,t-s}\big)(z,z) + h\nabla_1\big(\bar{G}_{kk'l,i}^{(s,t)} f_{i,t-s}\big)(\check{z},\check{z}).u + h\nabla_2\big(\bar{G}_{kk'l,i}^{(s,t)} f_{i,t-s}\big)(\check{z},\check{z}).v\Big\}\,du\,dv,
\end{aligned}$$

for some vectors $\tilde{z}$, $\bar{z}$ and $\check{z}$ in a ball with center z and radius h in $\mathbb{R}^q$, with obvious notations. Since K is even and the maps $\nabla^j B_i$, $j\in\{1,2\}$, are uniformly bounded w.r.t.

i in a neighborhood of z (Assumption 3), we obtain

$$\int K_h(x-z)K_h(y-z)\big(B_{ik}(x)-B_{ik}(z)\big)\big(B_{ik'}(y)-B_{ik'}(z)\big)\big(\bar{G}^{(s,t)}_{kk'l,i}f_{i,t-s}\big)(x,y)\,dx\,dy = O(h^4).$$

This yields $\mathbb{E}\Big[\big(r^{(1)}_{i3l}\big)^2\Big] = O\big(h^4 + h^2/(Th^{2q})\big)$. and then

$$\|r^{(1)}_{i3l}\|^2 = O_P\big(h^4 + \frac{h^2}{Th^{2q}}\big). \tag{A.19}$$

Invoking the index-regularity of the maps $\bar{G}^{(s,t)}_{kk'l,i}$, we similarly obtain the same rate for $N^{-1}\sum_{i=1}^N \|r^{(1)}_{i3}\|^2$.

Fourth, we have

$$r_{i4} = \frac{\widehat{\pi}^\top}{T}\sum_{t=1}^T w_{it}\Delta\widehat{C}_i(Z_{it})^\top G_t F_t + \frac{1}{T}\sum_{t=1}^T w_{it}\Delta\widehat{C}_i(Z_{it})^\top G_t\big\{\widehat{F}_t - \widehat{\pi}^\top F_t\big\} =: \widehat{\pi}^\top r^{(1)}_{i4} + r^{(2)}_{i4}.$$

Noting that $2\mathbb{E}[\|G_t\|\,\|F_t\||Z_{it}=z'] \leqslant \mathbb{E}[\|G_t\|^2|Z_{it}=z'] + \mathbb{E}[\|F_t\|^2|Z_{it}=z']$, we have

$$\begin{aligned}\|r^{(1)}_{i4}\| &\leqslant \frac{\sup_{\{z';\|z-z'\|<h\}}\|\Delta\widehat{C}_i(z')\|}{T}\sum_{t=1}^T w_{it}\|G_t\|\,\|F_t\| \\ &= O_P\big(\theta_{N,T}\big)\sup_t \sup_{\{z';\|z-z'\|<h\}} f_i(z')\mathbb{E}[\|G_t\|\,\|F_t\||Z_{it}=z'] = O_P\big(\theta_{N,T}\big).\end{aligned}$$

Concerning the averaging w.r.t. i, we have

$$\begin{aligned}\frac{1}{N}\sum_{i=1}^N\|r^{(1)}_{i4}\|^2 &\leqslant \frac{\sup_i\sup_{\{z';\|z-z'\|<a\}}\|\Delta\widehat{C}_i(z')\|^2}{NT^2}\sum_{i=1}^N\Big(\sum_{t=1}^T w_{it}\|G_t\|\,\|F_t\|\Big)^2 \\ &= O_P\big(\theta^2_{N,T}\big)\times\frac{1}{NT^2}\sum_{s,t=1}^T\Big(\sum_{i=1}^N w_{it}w_{is}\Big)\|G_t\|\,\|F_t\|\,\|G_s\|\,\|F_s\|.\end{aligned}$$

Since $\sup_{(s,t);s\neq t}\Big(\sum_{i=1}^N w_{it}w_{is}\Big) = O_P(1)$, we get

$$\begin{aligned}\frac{1}{NT^2}\sum_{s,t=1}^T\Big(\sum_{i=1}^N w_{it}w_{is}\Big)\|G_t\|\,\|F_t\|\,\|G_s\|\,\|F_s\| &= O_P(1)\Big(\frac{1}{T}\sum_{t=1}^T\|G_t\|\,\|F_t\|\Big)^2 \\ &+ \frac{1}{NT^2}\sum_{t=1}^T\Big(\sum_{i=1}^N w^2_{it}\Big)\|G_t\|^2\,\|F_t\|^2 \\ &= O_P(1)\Big(\frac{1}{T}\sum_{t=1}^T\|G_t\|^2\Big)\Big(\frac{1}{T}\sum_{t=1}^T\|F_t\|^2\Big) + \frac{O_P(1)}{T^2h^q}\sum_{t=1}^T\|G_t\|^2\,\|F_t\|^2,\end{aligned}$$

that is bounded in probability with our assumptions. We have obtained $N^{-1}\sum_{i=1}^N \|r_{i4}^{(1)}\|^2 = O_P(\theta_{N,T}^2)$. Moreover,

$$\begin{aligned}\|r_{i4}^{(2)}\|^2 &\leqslant \frac{1}{T}\sum_{t=1}^T w_{it}^2\big(\Delta\widehat{C}_i(Z_{it})^\top G_t\big)^2 \times \frac{1}{T}\sum_{t=1}^T \big\|\widehat{F}_t - \widehat{\pi}^\top F_t\big\|^2 \\ &= O_P\big(\frac{\upsilon_{N,T}}{h^q}\theta_{N,T}^2 \sup_t \sup_{\{z';\|z-z'\|<h\}} f_i(z')\mathbb{E}[\|G_t\|^2|Z_{it}=z']\big).\end{aligned}$$

The orders of magnitude of $\|r_{i4}^{(2)}\|^2$ and $N^{-1}\sum_{i=1}^N \|r_{i4}^{(2)}\|^2$ are $O_P(\upsilon_{N,T})$ since the maps $z' \mapsto \mathbb{E}[\|G_t\|^2|Z_{it}=z']$ are index-regular of order zero and $\theta_{N,T}^2 = O(h^q)$ by assumption.

By gathering the orders of magnitude we obtained, the convergence rate of $\widehat{B}_i(z)$ for a given i is as follows:

$$\|\widehat{B}_i(z) - \widehat{\pi}^{-1}B_i(z)\|^2 = O_P\big(\frac{\upsilon_{N,T}}{h^q} + \frac{1}{Th^{2q(1-1/\bar{\nu})}} + \frac{1}{Th^q} + h^4 + \frac{h^2}{Th^{2q}} + \theta_{N,T}^2\big) = O_P\big(\frac{\upsilon_{N,T}}{h^q}\big),$$

since $2(1-1/\bar{\nu}) < 4(1-1/\nu)$ and $1 < 4(1-1/\nu)$ because $\nu > 2$. Moreover, we have

$$\frac{1}{N}\sum_{i=1}^N \|\widehat{B}_i(z) - \widehat{\pi}^{-1}B_i(z)\|^2 = O_P\big(\upsilon_{N,T} + \frac{1}{Th^{2q(1-1/\bar{\nu})}} + \frac{1}{Th^q} + h^4 + \frac{h^2}{Th^{2q}} + \theta_{N,T}^2\big) = O_P\big(\upsilon_{N,T}\big).$$

□

Proof of Corollary 3. By the definition of the common components $Com_{it}(z)$, we have

$$\begin{aligned}\widehat{Com}_{it}(z) - Com_{it}(z) &= \widehat{B}_i(z)^\top \widehat{F}_t - B_i(z)^\top F_t \\ &= \widehat{B}_i(z)^\top(\widehat{F}_t - \widehat{\pi}^\top F_t) + (\widehat{B}_i(z) - \widehat{\pi}^{-1}B_i(z))^\top \widehat{\pi}^\top F_t \\ &= \big(\widehat{B}_i(z) - \widehat{\pi}^{-1}B_i(z)\big)^\top(\widehat{F}_t - \widehat{\pi}^\top F_t) + B_i(z)^\top(\widehat{\pi}^{-1})^\top(\widehat{F}_t - \widehat{\pi}^\top F_t) + \big(\widehat{B}_i(z) - \widehat{\pi}^{-1}B_i(z)\big)^\top \widehat{\pi}^\top F_t.\end{aligned}$$

Due to Lemma 5, it is known that $\|\widehat{\pi}\|$ and $\|\widehat{\pi}^{-1}\|$ are $O_P(1)$. Moreover, $T^{-1}\sum_{t=1}^T \|F_t\|^2 = O_P(1)$ (Assumption 7) and the loadings $B_i(z)$ are bounded uniformly w.r.t. i. Thus, deduce the statements from Theorem 1, Theorem 2 and the Cauchy-Schwarz inequality.

□

Lemma 3 (Lower bound for N_{ts})**.** Under Assumption 2 and Assumption 6, for any couple of positive numbers (c,ε) s.t. $c+\varepsilon < \underline{\ell}_\infty(z)$, there exists some positive constant M_0 that does not depend on (t,s,N,T) s.t.

$$\mathbb{P}\big(\min_{s,t\leqslant T, s\neq t} N_{ts} \leqslant cN\big) \leqslant 2T^2\exp\big(-M_0 N h^{2q}\varepsilon^2\big),$$

when N is sufficiently large and h sufficiently small. Moreover, for any couple of positive numbers $(\bar{c},\bar{\varepsilon})$ s.t. $\bar{c}+\bar{\varepsilon} < \|K^2\|_\infty \underline{\tilde{\ell}}_\infty(z)$, there exists some positive constant M_1 that does not depend on (t, N, T) s.t.

$$\mathbb{P}\big(\min_{t\leqslant T} h^q N_{tt} \leqslant \bar{c}N\big) \leqslant 2T\exp\big(-M_1 N h^q \bar{\varepsilon}^2\big),$$

when N is sufficiently large and h sufficiently small.

As a consequence, when T is smaller than a power of N and $Nh^{2q}/\ln N \to \infty$, $\mathbb{P}\big(\min_{s,t\leqslant T, s\neq t} N_{ts} \leqslant cN\big) = o(1)$. Moreover, when $Nh^q/\ln N \to \infty$, $\mathbb{P}\big(\min_{t\leqslant T} h^q N_{tt} \leqslant \bar{c}N\big) = o(1)$. In particular, Lemma 3 implies $\sup_{1\leqslant t,s\leqslant T} N/N_{ts} = O_P(1)$.

Proof. By direct calculation, for any $e > 0$ and when $s \neq t$, we have

$$\begin{aligned}
\mathbb{E}\big[\frac{N_{ts}}{N}\big] &= \frac{1}{Nh^{2q}}\sum_{i=1}^N \mathbb{E}\Big[K\big(\frac{Z_{it}-z}{h}\big)K\big(\frac{Z_{is}-z}{h}\big)\Big] \\
&= \frac{1}{Nh^{2q}}\sum_{i=1}^N \int K\big(\frac{x-z}{h}\big)K\big(\frac{y-z}{h}\big) f_{i,t-s}(x,y)dxdy \\
&= \frac{1}{N}\sum_{i=1}^N \int K(u)K(v) f_{i,t-s}(z+hu, z+hv)\, du\, dv \\
&\geqslant \liminf_N \inf_{s,t} \frac{1}{N}\sum_{i=1}^N \inf_{\{(u,v);\|u\|<h,\|v\|<h\}} f_{i,t-s}(z+u, z+v) - e = \underline{\ell}_\infty(z) - e,
\end{aligned}$$

when $N > N_0$, for some N_0 that does not depend on (t,s). Since the vectors (Z_{it}, Z_{is}), $i \in \{1,\dots,N\}$, are mutually independent, invoke Bernstein inequality (Van Der Vaart and Wellner, 1996, Lemma 2.2.9): For any positive constant ε, we have

$$\begin{aligned}
\mathbb{P}\Big(\big|\frac{N_{ts}(z)}{N} - \frac{\mathbb{E}[N_{ts}]}{N}\big| > \varepsilon\Big) &= \mathbb{P}\Big(\big|\sum_{i=1}^N \{K\big(\frac{Z_{it}-z}{h}\big)K\big(\frac{Z_{is}-z}{h}\big) \\
&- \mathbb{E}\big[K\big(\frac{Z_{it}-z}{h}\big)K\big(\frac{Z_{is}-z}{h}\big)\big]\}\big| > Nh^{2q}\varepsilon\Big) \leqslant 2\exp\big(-M_0 N h^{2q}\varepsilon^2\big) =: \Delta_{N,T}(\varepsilon),
\end{aligned} \tag{A.20}$$

where the constant M_0 does not depend on (t,s,N,T) because of (4.4). Thus, we obtain

$$\begin{aligned}
\mathbb{P}\big(N_{ts} \leqslant cN\big) &\leqslant \mathbb{P}\big(\frac{N_{ts}}{N} \leqslant c, |N_{ts} - \mathbb{E}[N_{ts}]| \leqslant N\varepsilon\big) + \mathbb{P}\big(|N_{ts} - \mathbb{E}[N_{ts}]| > N\varepsilon\big) \\
&\leqslant \mathbf{1}\big(\frac{\mathbb{E}[N_{ts}]}{N} \leqslant c+\varepsilon\big) + \mathbb{P}\big(|N_{ts} - \mathbb{E}[N_{ts}]| > N\varepsilon\big) \leqslant 0 + \Delta_{N,T}(\varepsilon),
\end{aligned}$$

when N is sufficiently large, h and e are sufficiently small, recalling $c + \varepsilon < \underline{\ell}_\infty(z)$. This yields $\mathbb{P}\big(\min_{1\leqslant s,t\leqslant T; s\neq t} N_{ts} \leqslant cN\big) \leqslant T^2\Delta_{N,T}(\varepsilon)$.

In the case $s = t$, we can lead a similar reasoning: for any $\bar{e} > 0$, we have

$$\mathbb{E}\big[\frac{h^q N_{tt}}{N}\big] = \frac{1}{Nh^q}\sum_{i=1}^N \mathbb{E}\Big[K^2\big(\frac{Z_{it}-z}{h}\big)\Big] = \frac{1}{N}\sum_{i=1}^N \int K^2(u) f_i(z+hu)\,du$$
$$\geqslant \quad \|K^2\|_\infty \liminf_N \frac{1}{N}\sum_{i=1}^N \inf_{\{u;\|u\|<h\}} f_i(z+u) - \bar{e} = \|K^2\|_\infty \tilde{\underline{\ell}}_\infty(z) - \bar{e},$$

when N is sufficiently large. When $\bar{e}$ is sufficiently small, the lower bound is positive due to (4.3). Again, invoke Bernstein inequality to get

$$\mathbb{P}\Big(\big|\frac{h^q N_{tt}}{N} - \frac{\mathbb{E}\big[h^q N_{tt}\big]}{N}\big| > \bar{\varepsilon}\Big) = \mathbb{P}\Big(\big|\sum_{i=1}^N\big\{K^2\big(\frac{Z_{it}-z}{h}\big) - \mathbb{E}\big[K^2\big(\frac{Z_{it}-z}{h}\big)\big]\big\}\big| > Nh^q\bar{\varepsilon}\Big)$$
$$\leqslant \quad 2\exp\big(-M_1 N h^q \bar{\varepsilon}^2\big) =: \bar{\Delta}_{N,T}(\bar{\varepsilon}),$$

where the constant M_1 does not depend on (t, N, T) by stationarity and (4.6). Thus, we obtain

$$\mathbb{P}\big(h^q N_{tt} \leqslant \bar{c}N\big) \leqslant \mathbb{P}\big(\frac{h^q N_{tt}}{N} \leqslant \bar{c}, h^q|N_{tt} - \mathbb{E}[N_{tt}]| \leqslant N\bar{\varepsilon}\big) + \mathbb{P}\big(h^q|N_{tt} - \mathbb{E}[N_{tt}]| > N\bar{\varepsilon}\big)$$
$$\leqslant \quad \mathbf{1}\big(\frac{\mathbb{E}\big[h^q N_{ts}\big]}{N} \leqslant \bar{c} + \bar{\varepsilon}\big) + \mathbb{P}\big(h^q|N_{ts} - \mathbb{E}[N_{ts}]| > N\bar{\varepsilon}\big) \leqslant 0 + \bar{\Delta}_{N,T}(\bar{\varepsilon}),$$

when N is sufficiently large, h and $\bar{e}$ are sufficiently small, keeping in mind that $\bar{c} + \bar{\varepsilon} < \|K^2\|_\infty \tilde{\underline{\ell}}_\infty(z)$. This yields $\mathbb{P}\big(\min_{1\leqslant t\leqslant T} h^q N_{tt} \leqslant \bar{c}N\big) \leqslant T\bar{\Delta}_{N,T}(\bar{\varepsilon})$. □

Lemma 4. Under Assumption 2, Assumption 3, Assumption 6 and Assumption 9, if $Nh^{2q}/(\ln N)^2 \to \infty$ then

$$\delta_{NT} := \sup_{1\leqslant t,s\leqslant T, t\neq s}\Big\|\frac{1}{N_{ts}}\sum_{i=1}^N w_{it}w_{is}B_i(Z_{it})B_i(Z_{is})^\top - \frac{1}{N}\sum_{i=1}^N B_i(z)B_i(z)^\top\Big\| = O_P\big(\frac{\ln(N)}{\sqrt{Nh^q}} + h^2\big).$$

Proof. Using the maximum norm for matrices, we have

$$\mathbb{P}(\delta_{NT} > \epsilon) \leqslant T^2 \sup_{1\leqslant t,s\leqslant T, s\neq t} \sup_{1\leqslant k,l\leqslant r} q_{ts,kl}(\epsilon), \text{ where}$$

$$q_{ts,kl}(\epsilon) := \mathbb{P}\Big(\big|\frac{1}{N_{ts}}\sum_{i=1}^N w_{it}w_{is}\big(B_i(z)B_i(z)^\top\big)_{kl} - \frac{1}{N}\sum_{i=1}^N\big(B_i(z)B_i(z)^\top\big)_{kl}\big| > \epsilon\Big).$$

By a standard reasoning, we have for any $\alpha > 0$

$$
\begin{aligned}
q_{ts,kl}(\epsilon) &\leqslant \mathbb{P}(|N_{ts} - \mathbb{E}[N_{ts}]| > \alpha) \\
&+ \mathbb{P}\Big(|\{\frac{1}{N_{ts}} - \frac{1}{\mathbb{E}[N_{ts}]}\}\sum_{i=1}^N w_{it}w_{is}\big(B_i(Z_{it})B_i(Z_{is})^\top\big)_{kl}| > \frac{\epsilon}{3}; |N_{ts} - \mathbb{E}[N_{ts}]| \leqslant \alpha\Big) \\
&+ \mathbb{P}\Big(\frac{1}{\mathbb{E}[N_{ts}]}|\sum_{i=1}^N w_{it}w_{is}\big(B_i(Z_{it})B_i(Z_{is})^\top - B_i(z)B_i(z)^\top\big)_{kl}| > \frac{\epsilon}{3}\Big) \\
&+ \mathbb{P}\Big(|\frac{1}{\mathbb{E}[N_{ts}]}\sum_{i=1}^N w_{it}w_{is}\big(B_i(z)B_i(z)^\top\big)_{kl} - \frac{1}{N}\sum_{i=1}^N \big(B_i(z)B_i(z)^\top\big)_{kl}| > \frac{\epsilon}{3}\Big) \\
&=: q_1 + q_2 + q_3 + q_4.
\end{aligned}
$$

Note that the quantities q_k, $k \in \{1,2,3,4\}$ implicitly depend on (s,t,k,l). First, let us evaluate $q_1 = \mathbb{P}\big(|\sum_{i=1}^N \{w_{it}w_{is} - \mathbb{E}[w_{it}w_{is}]\}| > \alpha\big)$. Note that $|w_{it}w_{is}| \leqslant \|K\|_\infty^2/h^{2q}$ and

$$
\begin{aligned}
\mathbb{E}[(w_{it}w_{is})^2] &= \int K_h^2(x-z)K_h^2(y-z)f_{i,t-s}(x,y)\,dx\,dy \\
&= \frac{1}{h^{2q}}\int K^2(u)K^2(v)f_{i,t-s}(z+hu,z+hv)\,du\,dv \leqslant \frac{\|K\|_\infty^4}{h^{2q}}\sup_{\{(u,v);\|u\|<h,\|v\|<h\}} f_{i,t-s}(z+u,z+v).
\end{aligned}
$$

Due to (4.4) and since the vectors (Z_{it}, Z_{is}), $i \in \{1,\dots,N\}$, are mutually independent from Assumption 6, we have

$$
\sup_{(t,s),t\neq s}\mathrm{Var}\big(\sum_{i=1}^N w_{it}w_{is}\big) = \sup_{(t,s),t\neq s}\sum_{i=1}^N \mathrm{Var}(w_{it}w_{is}) = O\big(Nh^{-2q}\big).
$$

Thus, we can invoke Bernstein inequality (Van Der Vaart and Wellner, 1996, Lemma 2.2.9) and we obtain

$$
q_1 \leqslant 2\exp\big(-\frac{\alpha^2}{C_1Nh^{-2q} + C_2h^{-2q}\alpha}\big), \tag{A.21}
$$

where C_1 and C_2 are two constants that do not depend on (t,s,k,l,T,N).

For the treatment of q_2, note that

$$
\begin{aligned}
q_2 &\leqslant \mathbb{P}\big(|\frac{1}{N_{ts}} - \frac{1}{\mathbb{E}[N_{ts}]}| \times |\sum_{i=1}^N w_{it}w_{is}(B_i(z)B_i(z)^\top)_{kl}| > \frac{\epsilon}{3}; |N_{ts} - \mathbb{E}(N_{ts})| \leqslant \alpha\big) \\
&\leqslant \mathbb{P}\big(\frac{r\alpha M_B^2}{N_{ts}\mathbb{E}[N_{ts}]}\sum_{i=1}^N w_{it}w_{is} > \frac{\epsilon}{3}; |N_{ts} - \mathbb{E}(N_{ts})| \leqslant \alpha\big),
\end{aligned}
$$

since we supposed that $\sup_i \|B_i(z)\| \leqslant M_B$. Recalling (4.2), $\liminf_N \inf_{t,s}\mathbb{E}[N_{ts}]/N > C_e$

for some positive constant C_e. Since the kernel K is nonnegative, we obtain

$$q_2 \leqslant \mathbb{P}\big(\frac{r\alpha M_B^2}{\mathbb{E}[N_{ts}]} > \frac{\epsilon}{3}\big) \leqslant \mathbb{P}\big(\frac{r\alpha M_B^2}{N} > \frac{C_e\epsilon}{4}\big) = 0,$$

when N is sufficiently large, if $\alpha/(N\epsilon) \to 0$.

To manage q_3, denote

$$\delta B_{ist} := \big(B_i(Z_{it})B_i(Z_{is})^\top - B_i(z)B_i(z)^\top\big)_{kl},$$

forgetting (k,l) for notational convenience. Note that $|\delta B_{ist}| \leqslant 2rM_B^2$. Thus, for N sufficiently large, this provides

$$\begin{aligned} q_3 &\leqslant \mathbb{P}\Big(\frac{1}{N}\big|\sum_{i=1}^N \{w_{it}w_{is}\delta B_{ist} - \mathbb{E}[w_{it}w_{is}\delta B_{ist}]\}\big| > \frac{C_e\epsilon}{8}\Big) \\ &+ \mathbb{P}\Big(\frac{1}{N}\big|\sum_{i=1}^N \mathbb{E}[w_{it}w_{is}\delta B_{ist}]\big| > \frac{C_e\epsilon}{8}\Big) =: q_3^{(1)} + q_3^{(2)}. \end{aligned}$$

Since $t \neq s$, some changes of variables and limited expansions yield

$$\begin{aligned} \mathbb{E}[w_{it}w_{is}\delta B_{ist}] &= \int K(u)K(v)\big(B_i(z+hu)B_i(z+hv)^\top - B_i(z)B_i(z)^\top\big)_{kl} \\ \times\;& f_{i,t-s}(z+hu,z+hv)\,du\,dv = h\int K(u)K(v)\Big\{B_i(z)\big(\nabla B_i(z).v\big)^\top + \nabla B_i(z).u\,B_i(z)^\top\Big)_{kl} \\ +\;& h\,O\big(\sup_{\{z'\mid\|z-z'\|<h\}}\|\nabla^2 B_i(z')\| + \sup_{\{z'\mid\|z-z'\|<h\}}\|\nabla B_i(z')\|^2\big)\Big\} \\ \times\;& \big\{f_{i,t-s}(z,z) + h\nabla_1 f_{i,t-s}(z+\tilde u, z+\tilde v).u + h\nabla_2 f_{i,t-s}(z+\tilde u,z+\tilde v).v\big\}\,du\,dv \end{aligned}$$

where $\tilde u$ (resp. $\tilde v$) satisfies $\|\tilde u\| < h$ (resp. $\|\tilde v\| < h$), with obvious notations. Due to the uniform boundedness of the maps $B_i(\cdot)$ and their first derivatives (Assumption 3), this yields

$$\mathbb{E}[w_{it}w_{is}\delta B_{ist}] \leqslant h^2 C_3\Big\{f_{i,t-s}(z,z) + \sup_{\{(\tilde u,\tilde v)\mid\|\tilde u\|<h,\|\tilde v\|<h\}}\|\nabla f_{i,t-s}(z+\tilde u,z+\tilde v)\|\Big\},$$

since $\int uK(u)\,du = 0$ (K is even). Under (4.5), this provides

$$\sup_{\{(s,t);s\neq t\}}\big|\frac{1}{N}\sum_{i=1}^N \mathbb{E}[w_{it}w_{is}\delta B_{ist}]\big| = O(h^2).$$

Thus, if $\epsilon \geqslant C_\epsilon h^2$ with a sufficiently large constant C_ϵ, then $q_3^{(2)} = 0$ for every (t,s), $t \neq s$, when N is sufficiently large.

To deal with $q_3^{(1)}$, apply Bernstein's inequality. By invoking the boundedness of $B_i(\cdot)$, the mutual independence of the covariates and by a usual change of variables, we get

$$\sup_{(s,t);s\neq t} \mathrm{Var}\Big(\sum_{i=1}^N w_{it}w_{is}\delta B_{ist}\Big) = \sup_{(s,t);s\neq t}\sum_{i=1}^N \mathrm{Var}(w_{it}w_{is}\delta B_{ist}) = O(\frac{N}{h^{2q}}).$$

This yields $q_3^{(1)} \leqslant 2\exp(-C_6 N h^{2q}\epsilon^2)$, for some positive constant C_6 that does not depend on (s,t,k,l,N,T).

To deal with q_4, rewrite

$$q_4 \leqslant \mathbb{P}(|\frac{1}{\mathbb{E}[N_{ts}]}\sum_{i=1}^N \{w_{it}w_{is} - \mathbb{E}[w_{it}w_{is}]\}\big(B_i(z)B_i(z)^\top\big)_{kl}| > \frac{\epsilon}{6})$$
$$+ \quad \mathbb{P}(|\frac{1}{\mathbb{E}[N_{ts}]}\sum_{i=1}^N \mathbb{E}[w_{it}w_{is}]\big(B_i(z)B_i(z)^\top\big)_{kl} - \frac{1}{N}\sum_{i=1}^N \big(B_i(z)B_i(z)^\top\big)_{kl}| > \frac{\epsilon}{6}) =: q_4^{(1)} + q_4^{(2)}.$$

Note that $\sup_{t\neq s} |(w_{it}w_{is})/\mathbb{E}[N_{ts}]| \leqslant 2||K||^2/(Nh^{2q}C_e)$ for N sufficiently large, and

$$\sup_{(s,t);s\neq t} \mathrm{Var}\big(\sum_{i=1}^N w_{it}w_{is}\big(B_i(z)B_i(z)^\top\big)_{kl}\big) = \sup_{(s,t);s\neq t}\sum_{i=1}^N \mathrm{Var}\big(w_{it}w_{is}\big(B_i(z)B_i(z)^\top\big)_{kl}\big) = O(\frac{N}{h^{2q}}),$$

by (4.4). Thus, applying Bernstein inequality again provides

$$q_4^{(1)} \leqslant 2\mathrm{exp}\Big(-\frac{N\epsilon^2 h^{2q}}{C_1' + C_2'\epsilon}\Big),$$

for two constants C_1' and C_2'. Denote $D_{N,t-s} := \sum_{i=1}^N f_{i,t-s}(z,z)$. Thanks to second-order limited expansions, it can be easily proved that

$$\sup_{s\neq t, s\leqslant T} |\frac{1}{\mathbb{E}[N_{ts}]}\sum_{i=1}^N \mathbb{E}[w_{it}w_{is}]\big(B_i(z)B_i(z)^\top\big)_{kl} - \frac{1}{D_{N,t-s}}\sum_{i=1}^N f_{i,t-s}\big(B_i(z)B_i(z)^\top\big)_{kl}| = O(h^2),$$

using the fact that $\sup_{s\neq t} N^{-1}\sum_{i=1}^N |\mathbb{E}[w_{it}w_{is}] - f_{i,t-s}(z,z)| = O(h^2)$ ((4.5) with $k=2$). Moreover, by Assumption 9 and (4.2), we have

$$\sup_{1\leqslant s\leqslant T} |\frac{1}{D_{N,t-s}}\sum_{i=1}^N f_{i,t-s}\big(B_i(z)B_i(z)^\top\big)_{kl} - \frac{1}{N}\sum_{i=1}^N \big(B_i(z)B_i(z)^\top\big)_{kl}| = O(h^2).$$

Now, set $\epsilon = \bar{C}_\epsilon\big(h^2 + \sqrt{N}^{-1} h^{-q}\ln(N)\big)$ that tends to zero. This choice is sufficient to manage q_3 and q_4. In particular, $q_4^{(2)} = 0$ when N is sufficiently large and if $\bar{C}_\epsilon$ is sufficiently large.

Setting $\alpha = \sqrt{N\ln N}/h^q$, check that $\liminf_N \alpha^2 h^{2q}/\big\{(N+\alpha)\ln N\big\} > 0$ (recall (A.21)) and $\alpha/(N\epsilon) \to 0$. With these choices and when $\bar{C}_\epsilon$ is sufficiently large, $T^\gamma \sup_{s\neq t} q_k = o(1)$

for any $\gamma > 0$ and every $k \in \{1, 2, 3, 4\}$. Therefore, we obtained

$$T^2 \sup_{1 \leqslant t,s \leqslant T} q_{ts,kl}(\epsilon) \longrightarrow 0,$$

Then, we have shown that $\delta_{NT} = O_P(\epsilon)$, i.e. the announced result. $\square$

Corollary 6. *Under the assumptions of Lemma 3 and Lemma 4, for any $t \leqslant T$, $\widehat{\pi}_t - \widehat{\pi} = r_{1t} + r_{2t}\|F_t\|\,\|\widehat{F}_t\|$, where*

$$\sup_{1 \leqslant t \leqslant T} \|r_{1t}\| = O_P\Big(\frac{\ln(N)}{\sqrt{N h^q}} + h^2\Big), \text{ and } \sup_{1 \leqslant t \leqslant T} \|r_{2t}\| = O_P\Big(\frac{1}{T}\Big).$$

Proof. Since $\sup_t N_{tt}^{-1} = O_P(h^q/N)$ by Lemma 3, we have

$$\begin{aligned}
\widehat{\pi}_t - \widehat{\pi} &= \frac{1}{T}\sum_{s=1}^{T}\Big(\frac{1}{N_{ts}}\sum_{i=1}^{N} w_{it}w_{is}B_i(Z_{it})B_i(Z_{is})^\top - \frac{1}{N}\sum_{i=1}^{N} B_i(z)B_i(z)^\top\Big)F_s\widehat{F}_s^\top V_z^{-1} \\
&= O_P\Big(\frac{\ln(N)}{\sqrt{N h^q}} + h^2\Big) \times \frac{1}{T}\sum_{s=1}^{T}\|F_s\widehat{F}_s^\top V_z^{-1}\| + \frac{1}{TN_{tt}}\sum_{i=1}^{N} w_{it}^2 B_i(Z_{it})B_i(Z_{it})^\top F_t\widehat{F}_t^\top V_z^{-1} \\
&- \frac{1}{NT}\sum_{i=1}^{N} B_i(z)B_i(z)^\top F_t\widehat{F}_t^\top V_z^{-1} \\
&= O_P\Big(\frac{\ln(N)}{\sqrt{N h^q}} + h^2\Big)\Big(\frac{1}{T}\sum_{s=1}^{T}\|F_s\|^2 \times \frac{1}{T}\sum_{s=1}^{T}\|\widehat{F}_s\|^2\Big)^{1/2} + \frac{O_P(h^q)}{NT}\sum_{i=1}^{N} w_{it}^2\|F_t\widehat{F}_t^\top\| \\
&+ O\Big(\frac{1}{T}\Big)\|F_t\widehat{F}_t^\top\| = O_P\Big(\frac{\ln(N)}{\sqrt{N h^q}} + h^2\Big) + \|F_t\|\,\|\widehat{F}_t\| O_P\Big(\frac{1}{T}\Big). \qquad \text{(A.22)}
\end{aligned}$$

Note that $\sup_{t \leqslant T} N^{-1}\sum_{i=1}^{N} w_{it}^2 = O_P(h^{-q})$ because $x \mapsto f_i(x)$ is index-regular of order zero. Moreover, the two terms $O_P(\cdot)$ in (A.22) do not depend on $t \in \{1, \ldots, T\}$ by recalling Lemma 4 and Lemma 3 respectively. $\square$

Lemma 5. Under the assumptions of Theorem 1, we have

$$\frac{\widehat{\mathbf{F}}^\top \mathbf{F}}{T} \xrightarrow{P} Q_z := V_{z,\infty}^{1/2}\Upsilon_z^\top \Sigma_B^{-1/2}$$

for some deterministic (r, r) matrices $V_{z,\infty}$ and Υ_z such that $\Upsilon_z^\top\Upsilon_y = I_r$ and $V_z = V_{z,\infty} + o_P(1)$. Moreover, the matrix $\hat{\pi}$ converges to Q_z^{-1} in probability, i.e. $\hat{\pi} = Q_z^{-1} + o_P(1)$.

Proof. The proof is similar to the proof of Proposition 1 in Bai (2003). Left multiplying $T^{-1}\tilde{\Sigma}\widehat{\mathbf{F}} = \widehat{\mathbf{F}}V_z$ by $\frac{1}{T}(\frac{\boldsymbol{B}(z)^\top\boldsymbol{B}(z)}{N})^{1/2}\mathbf{F}^\top$, we obtain

$$\frac{1}{T}\Big(\frac{\boldsymbol{B}(z)^\top\boldsymbol{B}(z)}{N}\Big)^{1/2}\frac{\mathbf{F}^\top\tilde{\Sigma}\widehat{\mathbf{F}}}{T} = \Big(\frac{\boldsymbol{B}(z)^\top\boldsymbol{B}(z)}{N}\Big)^{1/2}\frac{\mathbf{F}^\top\widehat{\mathbf{F}}}{T}V_z. \qquad \text{(A.23)}$$

Moreover, note that we can decompose any (t,s)-component of the $T\times T$ matrix $\tilde{\Sigma}$ as

$$\begin{aligned}\tilde{\Sigma}_{ts} &= \frac{1}{N}\sum_{i=1}^{N} F_t^\top B_i(z)B_i(z)^\top F_s + F_t^\top\Big(\frac{1}{N_{ts}}\sum_{i=1}^{N} w_{it}w_{is}B_i(Z_{it})B_i(Z_{is})^\top - \frac{1}{N}\sum_{i=1}^{N} B_i(z)B_i(z)^\top\Big)F_s\\ &+ \frac{1}{N_{ts}}\sum_{i=1}^{N} w_{it}w_{is}\Big\{B_i(Z_{it})^\top F_t\varepsilon_{is} + B_i(Z_{is})^\top F_s\varepsilon_{it} + \varepsilon_{it}\varepsilon_{is} + u_{it}\Delta\widehat{C}_i(Z_{is})^\top G_s\\ &+ u_{is}\Delta\widehat{C}_i(Z_{it})^\top G_t + \Delta\widehat{C}_i(Z_{it})^\top G_t\Delta\widehat{C}_i(Z_{is})^\top G_s\Big\}\\ &= F_t^\top\frac{\boldsymbol{B}(z)^\top\boldsymbol{B}(z)}{N}F_s + [r_{NT}]_{ts}.\end{aligned}$$

where $[r_{NT}]_{ts}$ is the (s,t)-term of the matrix r_{NT} that is composed of the remaining terms in the decomposition of $\tilde{\Sigma}_{ts}$ above. Note their similarities with the remaining terms in (A.2), at the beginning of the proof of Theorem 1. Thus, this provides $\tilde{\Sigma} = \boldsymbol{F}\boldsymbol{B}(z)^\top\boldsymbol{B}(z)\boldsymbol{F}^\top/N + r_{N,T}$ and

$$(\frac{\boldsymbol{B}(z)^\top\boldsymbol{B}(z)}{N})^{1/2}\frac{\boldsymbol{F}^\top\boldsymbol{F}}{T}(\frac{\boldsymbol{B}(z)^\top\boldsymbol{B}(z)}{N})\frac{\boldsymbol{F}^\top\widehat{\boldsymbol{F}}}{T} + d_{NT} = (\frac{\boldsymbol{B}(z)^\top\boldsymbol{B}(z)}{N})^{1/2}\frac{\boldsymbol{F}^\top\widehat{\boldsymbol{F}}}{T}V_z, \quad \text{(A.24)}$$

where $d_{NT} = \frac{1}{T}(\frac{\boldsymbol{B}(z)^\top\boldsymbol{B}(z)}{N})^{1/2}\frac{\boldsymbol{F}^\top r_{NT}\widehat{\boldsymbol{F}}}{T}$ is a $r\times r$ matrix. Moreover, rewrite the $r\times r$ matrix

$$\begin{aligned}\frac{\boldsymbol{F}^\top r_{NT}\widehat{\boldsymbol{F}}}{T^2} &= \frac{1}{T}\sum_{t=1}^{T} F_t(\widehat{\zeta}_t + \widehat{\eta}_t + \widehat{\tau}_t + \sum_{k=1}^{3}\widehat{\mu}_{k,t})^\top V_z\\ &+ \frac{1}{T^2}\sum_{s,t=1}^{T} F_tF_t^\top\Big(\frac{1}{N_{ts}}\sum_{i=1}^{N} w_{it}w_{is}B_i(Z_{it})B_i(Z_{is})^\top - \frac{1}{N}\sum_{i=1}^{N} B_i(z)B_i(z)\Big)F_s\widehat{F}_s^\top. \qquad \text{(A.25)}\end{aligned}$$

We will show below that $d_{N,T} = o_P(1)$. Assume this is the case for the moment. Deduce from (A.24) the identity

$$[S_{N,T} + d_{N,T}R_{N,T}^{-1}]R_{N,T} = R_{N,T}V_z, \text{ with}$$

$$R_{N,T} := (\frac{\boldsymbol{B}(z)^\top\boldsymbol{B}(z)}{N})^{1/2}\frac{\boldsymbol{F}^\top\widehat{\boldsymbol{F}}}{T}, \text{ and } S_{N,T} := (\frac{\boldsymbol{B}(z)^\top\boldsymbol{B}(z)}{N})^{1/2}\frac{\boldsymbol{F}^\top\boldsymbol{F}}{T}(\frac{\boldsymbol{B}(z)^\top\boldsymbol{B}(z)}{N})^{1/2}.$$

Note that the diagonal matrix V_z stacks the eigenvalues of $S_{N,T} + d_{N,T}R_{N,T}^{-1}$. Moreover, every column vector of $R_{N,T}$ is an eigenvector of the latter matrix. Rescale the latter eigenvectors in the matrix $\Upsilon_{N,T}^z := R_{N,T}V_z^{-1/2}$. This yields

$$\{S_{N,T} + d_{N,T}R_{N,T}^{-1}\}\Upsilon_{N,T}^z = \Upsilon_{N,T}^z V_z. \qquad \text{(A.26)}$$

Note that

$$R_{N,T}^{\top} R_{N,T} = \widehat{\boldsymbol{F}}^{\top} \boldsymbol{F} (\frac{\boldsymbol{B}(z)^{\top} \boldsymbol{B}(z)}{N}) \boldsymbol{F}^{\top} \widehat{\boldsymbol{F}} / T^2 = \widehat{\boldsymbol{F}}^{\top} \tilde{\Sigma} \widehat{\boldsymbol{F}} / T - \widehat{\boldsymbol{F}}^{\top} r_{N,T} \widehat{\boldsymbol{F}} / T = V_z + o_P(1).$$

Indeed, it can be proved that $\widehat{\boldsymbol{F}}^{\top} r_{N,T} \widehat{\boldsymbol{F}}/T = o_P(1)$ in exactly the same way as $d_{N,T} = o_P(1)$. The minimum eigenvalue of $R_{N,T}$ is $\min_x x^{\top} R_{N,T}^{\top} R_{N,T} x/\|x\|^2 = \min_x x^{\top} V_z x/\|x\|^2 + o_P(1)$ (Lütkepohl, 1997, Section 5.5.1). Thus, for N and T sufficiently large, the probability that the minimum eigenvalue of $R_{N,T}$ is larger than $c_v/2 > 0$ (c_v being defined in Assumption 4) is arbitrarily close to one and $R_{N,T}^{-1} = O_P(1)$. As a consequence, $d_{N,T} R_{N,T}^{-1} = o_P(1)$. Therefore, because of Assumption 4 and Assumption 3, we have

$$S_{N,T} + d_{N,T} R_{N,T}^{-1} = \Sigma_B^{1/2} \Sigma_F \Sigma_B^{1/2} + o_P(1) =: S + o_P(1).$$

By Assumption 4, the eigenvalues of the symmetrical matrix S are distinct and, by continuity, the matrices $S_{N,T}$ have distinct eigenvalues for sufficiently large N and large T. By (A.26) and the perturbation theory for eigenvalues of Hermitian matrices (Stewart and Sun, 1990, Theorem 4.11), keeping in mind that $S_{N,T} + d_{N,T} R_{N,T}^{-1}$ is a perturbation of S, the eigenvalues of the former matrix tend to the eigenvalues of the latter. In other words, there exists a deterministic matrix $V_{z,\infty}$ s.t. $V_z = V_{z,\infty} + o_P(1)$, where $V_{z,\infty}$ is obviously a diagonal matrix with distinct entries.

Moreover, since $R_{N,T}^{\top} R_{N,T}$ tends to V_z in probability, the matrix of eigenvectors $\Upsilon_{N,T}^z$ satisfies $(\Upsilon_{N,T}^z)^{\top} \Upsilon_{N,T}^z \to I_r$ in probability. Asymptotically, the column vectors of $\Upsilon_{N,T}^z$ build an orthonormal basis. Remind they are eigenvectors of $S_{N,T} + o_P(1)$ and then of the positive definite matrix S asymptotically. Since S has distinct positive eigenvalues, the column vectors of $\Upsilon_{N,T}^z$ are essentially unique, apart some multiplicative factors (-1), by the continuity of eigenvectors w.r.t. matrix coefficients (Franklin, 2012, Section 6.12). Thus, there exists a matrix of orthonormal eigenvectors Υ_z such that $\|\Upsilon_{N,T}^z - \Upsilon_z\| = o_P(1)$. Deduce from the definitions of $R_{N,T}$ and $\Upsilon_{N,T}^z$ that

$$\frac{\boldsymbol{F}^{\top} \widehat{\boldsymbol{F}}}{T} = \big(\frac{\boldsymbol{B}(z)^{\top} \boldsymbol{B}(z)}{N}\big)^{-1/2} \Upsilon_{N,T}^z V_z^{1/2} = \Sigma_B^{-1/2} \Upsilon_z V_{z,\infty}^{1/2} + o_P(1),$$

yielding the announced result.

Now, let us show that $d_{N,T}$ is actually $o_P(1)$. From (A.25), we have

$$
\begin{aligned}
\|d_{NT}\|^2 &\leqslant \|\frac{\boldsymbol{B}(z)^\top\boldsymbol{B}(z)}{N}\|\|\frac{\boldsymbol{F}^\top r_{NT}\widehat{\boldsymbol{F}}}{T^2}\|^2 \\
&\leqslant \frac{O_P(1)\|V_z\|^2}{T}\sum_{t=1}^T\|F_t\|^2 \times \frac{1}{T}\sum_{t=1}^T\|\widehat{\zeta}_t+\widehat{\eta}_t+\widehat{\tau}_t+\sum_{k=1}^3\widehat{\mu}_{k,t}\|^2 \\
&+ \sup_{s,t}\Big\|\frac{1}{N_{ts}}\sum_{i=1}^N w_{it}w_{is}B_i(Z_{it})B_i(Z_{is})^\top - \frac{1}{N}\sum_{i=1}^N B_i(z)B_i(z)^\top\Big\|^2 \times \Big(\frac{1}{T^2}\sum_{s,t=1}^T\|F_t\|^2\,\|F_s\|\,\|\widehat{F}_s\|\Big)^2 \\
&\leqslant \frac{O_P(1)}{T}\sum_{t=1}^T\{\|\widehat{\zeta}_t\|^2+\|\widehat{\eta}_t\|^2+\|\widehat{\tau}_t\|^2+\sum_{k=1}^3\|\widehat{\mu}_{k,t}\|^2\} \\
&+ O_P\big(\frac{\ln^2(N)}{Nh^{2q}}+h^4\big)\Big(\frac{1}{T}\sum_{s=1}^T\|F_s\|^2\times\frac{1}{T}\sum_{s=1}^T\|\widehat{F}_s\|^2\Big) \\
&= O_P\big(v_{N,T}+\frac{\ln^2(N)}{Nh^{2q}}+h^4\big) = o_P(1).
\end{aligned}
$$

The last equality is obtained thanks to the proof of Theorem 1 and Lemma 4. Finally, we easily obtain

$$
\begin{aligned}
\hat{\pi}^\top &= V_z^{-1}\Big(\frac{\widehat{\boldsymbol{F}}^\top\boldsymbol{F}}{T}\Big)\frac{\boldsymbol{B}(z)^\top\boldsymbol{B}(z)}{N} = V_{z,\infty}^{-1}Q_z\Sigma_B + o_P(1) \\
&= V_{z,\infty}^{-1/2}\Upsilon_z^\top\Sigma_B^{1/2} + o_P(1) = \big(Q_z^{-1}\big)^\top + o_P(1),
\end{aligned}
$$

concluding the proof of the lemma. □

Proof of Corollary 4. The first assertion is stated in Lemma 5 above. Moreover, since we have $\hat{\pi}^\top := V_z^{-1}\frac{\widehat{\boldsymbol{F}}^\top\boldsymbol{F}}{T}\frac{\boldsymbol{B}(z)^\top\boldsymbol{B}(z)}{N}$, note first that

$$
\frac{\widehat{\boldsymbol{F}}^\top\boldsymbol{F}}{T} = \frac{(\widehat{\boldsymbol{F}}-\boldsymbol{F}\hat{\pi})^\top\boldsymbol{F}}{T} + \frac{\hat{\pi}^\top\boldsymbol{F}^\top\boldsymbol{F}}{T} = \frac{(\widehat{\boldsymbol{F}}-\boldsymbol{F}\hat{\pi})^\top\boldsymbol{F}}{T} + \hat{\pi}^\top\big(1+o_P(1)\big),
$$

because, under (3.14), $\boldsymbol{F}^\top\boldsymbol{F}/T \xrightarrow{P} I_r$. Moreover, using the same notations as in the proof of Theorem 1 and applying the Cauchy-Schwarz inequality, we have

$$
\|\frac{(\widehat{\boldsymbol{F}}-\boldsymbol{F}\hat{\pi})^\top\boldsymbol{F}}{T}\|^2 = \|\frac{1}{T}\sum_{t=1}^T(\widehat{F}_t-\hat{\pi}^\top F_t)F_t^\top\|^2 \leqslant \Big(\frac{1}{T}\sum_{t=1}^T\|F_t\|^2\Big)\Big(\frac{1}{T}\sum_{t=1}^T\|\widehat{F}_t-\hat{\pi}^\top F_t\|^2\Big),
$$

that tends to zero in probability with N and T (Theorem 1). Then, $\widehat{\boldsymbol{F}}^\top\boldsymbol{F}\hat{\pi}/T = \hat{\pi}^\top\hat{\pi}\big(1+o_P(1)\big)$.

Second, we have

$$\frac{\widehat{\boldsymbol{F}}^\top \boldsymbol{F}\hat{\pi}}{T} = \frac{\widehat{\boldsymbol{F}}^\top}{T}(\boldsymbol{F}\hat{\pi} - \widehat{\boldsymbol{F}} + \widehat{\boldsymbol{F}}) = \frac{1}{T}\widehat{\boldsymbol{F}}^\top(\boldsymbol{F}\hat{\pi} - \widehat{\boldsymbol{F}}) + I_r = I_r + o_P(1),$$

yielding $\hat{\pi}^\top\hat{\pi} = I_r + o_P(1)$. This means $\hat{\pi}$ is asymptotically an orthogonal matrix so that its eigenvalues tend to 1 or -1 when T and N tend to the infinity.

Third, recalling the definition of $\hat{\pi}$, we have got

$$\hat{\pi}^\top = V_z^{-1}\big(\frac{\widehat{\boldsymbol{F}}^\top \boldsymbol{F}}{T}\big)\frac{\boldsymbol{B}(z)^\top\boldsymbol{B}(z)}{N} = V_z^{-1}\hat{\pi}^\top\frac{\boldsymbol{B}(z)^\top\boldsymbol{B}(z)}{N} + o_P(1),$$

that can be rewritten $\hat{\pi}V_z + o_P(1) = (\boldsymbol{B}(z)^\top\boldsymbol{B}(z)/N)\hat{\pi}$. The latter relationship implies that $\hat{\pi}$ tends in probability to a matrix consisting of eigenvectors of $\boldsymbol{B}(z)^\top\boldsymbol{B}(z)/N$. By (3.13), this matrix is diagonal with distinct eigenvalues. Then each of its eigenvalues is associated with a unique eigenvector, and each of its eigenvectors has a single non-zero element in the canonical basis. This implies that $\hat{\pi}$ is asymptotically a diagonal matrix, but we already know that its eigenvalues tend to 1 or -1. Without loss of generality, we can assume all its elements are 1 (otherwise multiply the corresponding columns of $\widehat{\boldsymbol{F}}$ and $\hat{\boldsymbol{B}}$ by -1). Finally, this yields $\hat{\pi} = I_r + o_P(1)$.

□

Appendix B. Sketch of proofs for some asymptotic normality results

To simplify our arguments, we assume that the innovations are independent white noises.

Assumption 13. *For any i, the sequence $\varepsilon_{i.} := (\varepsilon_{it})_{t\geqslant 1}$ is a weak white noise, that is independent of the processes $(F_t)_{t\geqslant 1}$ and $(Z_{kt})_{k\geqslant 1, t\geqslant 1}$. Moreover, the random elements $(\varepsilon_{i.})_{i\geqslant 1}$ are mutually independent.*

The latter independence assumption could be replaced by conditional orthogonality conditions. Moreover, the process of innovations could be assumed to be strongly mixing, at the price of additional complexities.

Asymptotic normality of $\sqrt{Nh^q}\big(\widehat{F}_t - \hat{\pi}_t^\top F_t\big)$. Due to (A.3), the result follows if we prove that $\sqrt{Nh^q}\hat{\eta}_t$ is asymptotically normal, and

$$\hat{\zeta}_t + \hat{\tau}_t + \hat{\mu}_{1,t} + \hat{\mu}_{2,t} + \hat{\mu}_{3,t} = o_P\big(\frac{1}{\sqrt{Nh^q}}\big).$$

First, $V_z\hat{\eta}_t = \tilde{\eta}_t + \Delta\eta_t + \overline{\Delta\eta_t}$, where

$$\tilde{\eta}_t := \frac{\hat{\pi}_t^\top}{T}\sum_{i=1}^N\sum_{s=1}^T \frac{w_{it}w_{is}}{\mathbb{E}[N_{ts}]}\varepsilon_{it}F_sF_s^\top B_i(Z_{is}), \quad \Delta\eta_t := \frac{1}{T}\sum_{i=1}^N\sum_{s=1}^T \frac{w_{it}w_{is}}{N_{ts}}\varepsilon_{it}(\widehat{F}_s - \hat{\pi}_t^\top F_s)F_s^\top B_i(Z_{is}),$$

$$\overline{\Delta\eta_t} := \frac{\widehat{\pi}_t^\top}{T}\sum_{i=1}^N\sum_{s=1}^T w_{it}w_{is}\Big(\frac{1}{N_{ts}} - \frac{1}{\mathbb{E}[N_{ts}]}\Big)\varepsilon_{it}F_sF_s^\top B_i(Z_{is}).$$

Assume that the sequences $N^{-1}\sum_{i=1}^N f_{i,t-s}(z,z)$ tend to some limit $\ell_{\infty,t-s}(z)$ uniformly w.r.t. (s,t), when $s\neq t$. Due to (4.5), this implies $\sup_{s\neq t}|\mathbb{E}[N_{t,s}]/N - \ell_{\infty,t-s}(z)| = o(h^2)$. In other words, we can rewrite $\mathbb{E}[N_{t,s}] = Nc_{T,t-s}$ where $c_{T,t-s}\to\ell_{\infty,t-s}(z)$ uniformly w.r.t. $t-s$, when $s\neq t$. Thus,

$$\tilde{\eta}_t \simeq \frac{\widehat{\pi}_t^\top}{N}\sum_{i=1}^N w_{it}\varepsilon_{it}\Big(\frac{1}{T}\sum_{s=1,s\neq t}^T\frac{w_{is}}{c_{T,t-s}}F_sF_s^\top B_i(Z_{is})\Big) \simeq \frac{\widehat{\pi}_t^\top}{N}\sum_{i=1}^N w_{it}\varepsilon_{it}\check{\eta}_{i,t},$$

introducing the deterministic quantities

$$\check{\eta}_{i,t} := \lim_{T\to\infty}\frac{1}{T}\sum_{s=1,s\neq t}^T\frac{1}{\ell_{\infty,t-s}(z)}\mathbb{E}[F_sF_s^\top|Z_{is}=z]B_i(z)f_i(z).$$

Since the vectors $(Z_{it},\varepsilon_{it})_{i\geqslant 1}$ are independent by assumption, the CLT for triangular arrays in Shiryaev (1996, p. 334) can be invoked, to state that $\sqrt{Nh^q}\sum_{i=1}^N w_{it}\varepsilon_{it}\check{\eta}_{i,t}/N$ is asymptotically normal. Note that $\widehat{\pi}_t = \widehat{\pi} + o - p(1)$ (A.7) and $\widehat{\pi}$ is weakly convergent (Corollary 4). Thus, $\widehat{\pi}_t$ and deduce $\sqrt{Nh^q}\tilde{\eta}_t$ is asymptotically normal.

Concerning $\Delta\eta_t$, recall (A.6):

$$\|\widehat{F}_t - \widehat{\pi}_t^\top F_t\|^2 = O_P(\bar{\upsilon}_{N,T}),\ \text{with}\ \bar{\upsilon}_{N,T} := \frac{1}{Nh^{2q}} + \frac{1}{Th^{4q(1-1/\nu)}} + \theta_{N,T}^2 = O(\upsilon_{N,T}).$$

The Cauchy-Schwarz inequality and the uniform boundedness of the maps $B_i(\cdot)$ yield

$$\|\Delta\eta_t\|^2 \leqslant \frac{1}{T}\sum_{s=1}^T\|\widehat{F}_s - \widehat{\pi}_t^\top F_s\|^2 \times \frac{O_P(1)}{T}\sum_{s=1}^T\Big(\frac{1}{N}\sum_{i=1}^N w_{it}w_{is}\varepsilon_{it}\|F_s\|\Big)^2 = O_P(\frac{\bar{\upsilon}_{N,T}}{Nh^{2q}}),\quad \text{(B.1)}$$

by calculating the expectation of the second term on the r.h.s. of (B.1). Thus, $\sqrt{Nh^q}\Delta\eta_t = o_P(1)$ when $\bar{\upsilon}_{N,T} = o(h^q)$ (particularly when $\upsilon_{N,T} = o(h^q)$).

From Lemma 3 (or (A.20)), it is known that $\sup_{s\neq t}|N_{ts} - \mathbb{E}[N_{ts}]| = O_P(\ln N/\sqrt{Nh^{2q}})$, that is not $o_P(1/\sqrt{Nh^q})$ unfortunately. Thus, a rough upper bound is insufficient to manage $\overline{\Delta\eta_t}$. Nonetheless, by writing $N_{ts} - \mathbb{E}[N_{ts}]$ as a sum of independent random quantities, the order of magnitude of $\text{Var}(\overline{\Delta\eta_t})$ is the same as

$$\frac{1}{T^2N^4}\sum_{i,i'=1}^N\sum_{s,s'=1}^T\mathbb{E}\big[w_{is}w_{it}w_{i's'}w_{i't}\varepsilon_{it}\varepsilon_{i't}\sum_{k,k'=1}^N\big(w_{kt}w_{ks} - \mathbb{E}[w_{kt}w_{ks}]\big)\big(w_{k't}w_{k's'} - \mathbb{E}[w_{k't}w_{k's'}]\big)F_{sj}^2F_{s'j'}^2\big],$$

for some $(j,j')\in\{1,\ldots,r\}^2$. The (main) non zero terms of the latter expectation are related to $i=i'$ and $k=k'$. By usual changes of variables and reasoning, it can be easily checked that $\text{Var}(\overline{\Delta\eta_t}) = O(N^{-2}h^{-3q})$, yielding $\sqrt{Nh^q}\overline{\Delta\eta_t} = o_P(1)$ since Nh^{2q} tends to zero by assumption. Since V_z is weakly convergent, we have obtained that $\sqrt{Nh^q}\widehat{\eta}_t$ is asymptotically Gaussian.

Hereafter and due to (A.20), the denominators N_{ts} will not be a source of worry, except for the calculations of the exact limiting variance-covariance matrices (what we will not do here!). Thus, they will simply be replaced with N to analyze the remaining terms that will be negligible compared to $1/\sqrt{Nh^q}$.

Second, it can be proved as above that the order of magnitude of $\widehat{\zeta}_t$ is the same as

$$\tilde{\zeta}_t := \frac{1}{TN}\sum_{i=1}^N\sum_{s=1}^T w_{is}w_{it}\varepsilon_{is}F_sB_i(Z_{it})^\top F_t.$$

The latter term is centered. Its variance depends on conditional expectations of $\varepsilon_{is}\varepsilon_{ks'}$ given some covariates and some factors. Using Assumption 13, such expectations are zero only when $i=k$ and $s=s'$. By usual changes of variables and reasoning, it can be checked that $\mathrm{Var}(\tilde{\zeta}_t)$ is of order $(NTh^{2q})^{-1}$. Therefore, $\sqrt{Nh^q}\tilde{\zeta}_t = o_P(1)$ and similarly for $\sqrt{Nh^q}\widehat{\zeta}_t$.

Third, by a similar reasoning, the study of $\widehat{\tau}_t$ boils down to that of

$$\tilde{\tau}_t := \frac{1}{NT}\sum_{i=1}^N\sum_{s=1}^T w_{it}w_{is}\varepsilon_{it}\varepsilon_{is}F_s.$$

The calculation of the variance of $\tilde{\tau}_t$ is written as a fourfold sum involving the expectations of terms as $\varepsilon_{it}\varepsilon_{is}\varepsilon_{kt}\varepsilon_{ks'}$. The latter ones are non zero only when there are at least two identities among the indices i,k,s,s' and t. To be specific, it is necessary to impose $i=k$ and $s=s'$, or $s=s'=t$, yielding $\mathrm{Var}(\tilde{\tau}_t)\sim 1/(T^2h^{2q})+1/(TNh^{2q})$ that is $o\big(1/(Nh^q)\big)$. A remainder term is

$$\Delta\tau_t := \frac{1}{NT}\sum_{i=1}^N\sum_{s=1}^T w_{it}w_{is}\varepsilon_{it}\varepsilon_{is}(\widehat{F}_s-\widehat{\pi}_t^\top F_s).$$

Note that

$$\|\Delta\tau_t\|^2 \leqslant \frac{1}{T}\sum_{s=1}^T\|\widehat{F}_s-\widehat{\pi}_t^\top F_s\|^2\times\frac{1}{N^2T}\sum_{s=1}^T\sum_{i,k=1}^N w_{it}w_{is}w_{kt}\varepsilon_{it}\varepsilon_{is}\varepsilon_{kt}\varepsilon_{ks}.$$

The expectation of the later product may be evaluated by imposing $i=k$ or $s=t$, yielding $\|\Delta\tau_t\|^2\sim\bar{v}_{N,T}h^{-2q}(N^{-1}+T^{-1})$. Thus, $\Delta\tau_t = o_P(1/\sqrt{Nh^q})$ when $\bar{v}_{N,T} = o(Th^q/N)$, that we can be imposed.

Fourth, the other terms $\widehat{\mu}_{j,t}$, $j\in\{1,2,3\}$ are $O_P(\theta_{N,T})$. Once we impose $\sqrt{Nh^q}\theta_{N,T} = o(1)$, they are all negligible. □

Remark 12. *Note that Theorem 1 and its proof have been invoked to evaluate some residual terms in the study of $\sqrt{Nh^q}(\widehat{F}_t-\widehat{\pi}_t^\top F_t)$. In other words, the orders of magnitude for consistency purpose obtained in Section 4.2 are essential steps to prove our asymptotic* *normality results.*

Asymptotic normality of $\sqrt{Th^q}\big(\widehat{B}_i(z)-\widehat{\pi}^{-1}B_i(z)\big)$. Recalling (A.8), the result follows if we prove that

(1) $\sqrt{Th^q}\widehat{\pi}^\top\sum_{t=1}^T w_{it}\varepsilon_{it}F_t/T$ is asymptotically normal;

(2) $\widehat{\Omega}_i^{-1}$ is weakly convergent;

(3) $\sum_{t=1}^T w_{it}\varepsilon_{it}(\widehat{\pi}-\widehat{\pi}_t)^\top F_t/T = o_P(1/\sqrt{Th^q})$;

(4) $\sum_{t=1}^T w_{it}\varepsilon_{it}\nu_t F_t/T = o_P(1/\sqrt{Th^q})$ for all terms coming from the expansion (A.3), i.e. $\nu_t \in \{\widehat{\zeta}_t, \widehat{\eta}_t, \widehat{\tau}_t, \widehat{\mu}_{1,t}, \widehat{\mu}_{2,t}, \widehat{\mu}_{3,t}\}$;

(5) the remaining terms r_{ij}, $j\in\{1,3,4\}$, are $o_P(1/\sqrt{Th^q})$.

First, to prove (1), note that $\widehat{\pi}$ is weakly convergent (Corollary 4). Moreover, simple calculations show that the variance of the centered random vector $\sum_{t=1}^T w_{it}\varepsilon_{it}F_t/T$ is of order $1/(Th^q)$. A CLT for strongly mixing arrays (Peligrad, 1996, Theorem 2.2) yields the convenient tool to state (1).

Second, it has been proved that $\widehat{\Omega}_i$ is weakly convergent (A.12). When the limit is non-zero, this is still the case for its inverse.

Third,

$$\frac{1}{T}\Big\|\sum_{t=1}^T w_{it}\varepsilon_{it}(\widehat{\pi}-\widehat{\pi}_t)^\top F_t\Big\| \leqslant \sup_{1\leqslant t\leqslant T}\|\widehat{\pi}_t-\widehat{\pi}\| \times \frac{1}{T}\sum_{t=1}^T w_{it}\|\varepsilon_{it}F_t\| = O_P\Big(\frac{\ln N}{\sqrt{Nh^q}}+h^2\Big),$$

due to (A.7). The latter rate is sufficient to get (3) if $Nh^{q+4}=o(1)$ and $T/(Nh^q)=o(1)$. Note that the former condition can be difficult to reconcile with $v_{N,T}=o(h^q)$ in Theorem 2. At this stage, this is the case only when $q=1$. Otherwise, (3) has to be proved with other techniques to evaluate the discrepancies $\widehat{\pi}_t-\widehat{\pi}$ more precisely, typically with some terms that appear in the proof of Lemma 4. See Remark 8 too. Such refined developments are left aside for further work.

Point (4) can be checked relatively easily, by evaluating the variance of all the corresponding terms. Therefore, the presence of products as $\varepsilon_{it}\varepsilon_{it'}$ multiplied by innovations associated with other names necessitates some identities between the sets of indices and the obtained orders of magnitude are sufficient for our purpose. When $\sqrt{Th^q}\theta_{N,T}=o(1)$, the terms related to $\widehat{\mu}_{j,t}$, $j\in\{1,2,3\}$ are negligible too.

Fifth, the order of magnitude obtained for r_{i1} in the proof of Theorem 2 is not sufficient. Again, it is necessary to invoke all terms in the expansion of $\widehat{F}_t-\widehat{\pi}_t^\top F_t$ in (A.3), and to evaluate their orders of magnitude. Typically, this leads us to evaluate terms such as $T^{-1}\sum_{t=1}^T w_{it}F_t\zeta_t^\top$, and then

$$\tilde{r}_{i1} := \frac{1}{NT}\sum_{t=1}^T\sum_{k=1}^N\sum_{s=1}^T w_{it}w_{kt}w_{ks}B_k(Z_{kt})^\top F_t\varepsilon_{ks}F_tF_s^\top,$$

for instance. The variance of the latter triple sum is written as a sum of expectations induced by (t,t',k,k',s,s'), with obvious notations. To get non zero expectations, it is necessary to impose at least two identities across these six indices, say $k=k'$ and $s=s'$. Such a variance is then $O\big(1/(NTh^q)\big)=o\big(1/(Th^q)\big)$, that is $\tilde r_{i1}=o_P(1/\sqrt{Th^q})$. This reasoning can be applied to the other corresponding terms, to yield $r_{i1}=o_P(1/\sqrt{Th^q})$ and we omit the details. Concerning r_{i3}, note that $\sqrt{Th^q}r_{i3}^{(1)}=o_P(1)$ when $Th^{q+4}=o(1)$ and $q=1$, recalling (A.19). Revisiting $r_{i3}^{(2)}$ is more painful and recalling the expansion of (A.3) is again necessary. We skip the details, but the corresponding terms will be negligible. Finally, by the same technique, it can be proved that $r_{i4}=o_P(1/\sqrt{Th^q})$ when $Th^q\theta^2_{N,T}=o(1)$. □

Appendix C. Definition and uniform convergence of a particular $\widehat{C}_i(\cdot)$

Let us first state a technical lemma that is of interest per se and will constitute the key tool for proving (3.5). For any vector $z\in\mathbb{R}^q$, define

$$S_{iT}(z):=\frac{1}{T}\sum_{t=1}^{T}K_{h_0}(Z_{it}-z)Y_{it},$$

for arbitrary random variables Y_{it}, a kernel K that satisfies Assumption 2 and the previously defined covariates Z_{it}. Obviously, the bandwidth $h_0=h_0(T)$ is a deterministic sequence of positive numbers that tends to zero with T. For the sake of simplicity, it does not depend on i and it is the same for all components of Z_{it}, even if the latter features can be easily weakened. This bandwidth h_0 may be different (or not) from the bandwidth h used for the localized PCA (recall (3.10)). We now prove the weak consistency of $S_{iT}(z)$ uniformly with respect to $z\in\mathcal{C}$, a compact subset in $\mathbb{R}^q$, and w.r.t. $i\in\{1,\dots,N\}$. To this aim, we need a few generic assumptions.

Assumption 14. *For any $i\geqslant 1$, the sequence $(Y_{it},Z_{it})_{t\geqslant 1}$ is strictly stationary and strongly mixing, with mixing coefficients $(\alpha_{Y,m})_{m\geqslant 1}$ that satisfy $\alpha_{Y,m}\leqslant\exp(-C_Y m)$, for some constant $C_Y>0$.*

Note the assumed uniformity of the latter mixing coefficients w.r.t. i. Let $\tilde{\mathcal{C}}$ be an open neighborhood of $\mathcal{C}$.

Assumption 15. *For any i, the density f_i of Z_{it} exists, $\sup_i\|f_i\|_\infty<\infty$, $\inf_i\inf_{z\in\tilde{\mathcal{C}}}f_i(z)>0$, and $\sup_i\sup_{z\in\tilde{\mathcal{C}}}\mathbb{E}[|Y_{it}|^s\,|\,Z_{it}=z]f_i(z)<\infty$ for some $s>2$. Moreover, denoting $f_{i,t}$ the joint density of $(Z_{i,1},Z_{i,t+1})$, there is some index t^* such that, for all $t\geqslant t^*$,*

$$\sup_i\sup_{(z,z')\in\tilde{\mathcal{C}}^2}\mathbb{E}[|Y_{i,1}Y_{i,t+1}|\,|\,Z_{i,1}=z,Z_{i,t+1}=z']f_{i,t}(z,z')<\infty.$$

Note that lagged values of Y_{it} may be included in Z_{it}. Define $r_i(z):=\mathbb{E}[Y_{it}|Z_{it}=z]$.

Assumption 16. *For any $i \geqslant 1$, the map $z' \mapsto (f_i r_i)(z')$ is twice continuously differentiable on $\tilde{\mathcal{C}}$. Moreover,* $\sup_i \sup_{z' \in \tilde{\mathcal{C}}} \|\nabla^2 (f_i r_i)(z')\| < \infty$.

Lemma 6. Suppose that Assumption 2 and 14-16 hold. If $N = O(T^\xi)$ for some $\xi > 0$ and $\ln T / T^\theta = o(h_0^d)$ for some $\theta \in (0,1)$, then

$$\sup_{1 \leqslant i \leqslant N} \sup_{z \in \mathcal{C}} |S_{iT}(z) - (f_i r_i)(z)| = O_P(a_T + h_0^2), \text{ with } a_T := \left(\frac{\ln T}{T h_0^q}\right)^{1/2}. \tag{C.1}$$

Note that the uniform rate of convergence does not depend on N. Thus is due to the fact N is assumed to be smaller than a positive power of T. When the process $(Y_{it})_{i \geqslant 1, t \geqslant 1}$ satisfies Assumption 14-16, then it is said that $(Y_{it})_{i \geqslant 1, t \geqslant 1}$ is $Z_{it}-$adapted [13].

Proof. The proof of Lemma 6 is an adaptation of the proof of Hansen (2008, Theorem 2). For the sake of simplicifity and contrary to the latter paper, we have considered that the compact subset $\mathcal{C}$ does not depend on T, meaning that its sequence c_n is replaced by a constant. In the same vein, we have assumed exponentially decreasing mixing coefficients instead of polynomially decreasing ones. By carefully inspecting the proof of Hansen (2008, Theorem 2), its term T_{1n} is in our case $NT^{-\gamma}$, for an arbitrarily large constant $\gamma > 0$, and then tends to zero. Moreover, its term T_{2n} is here replaced by a constant times $NT^\rho \alpha_{Y,m}/m$ with $m = \lfloor a_T^{(2-s)/(s-1)} \rfloor$ and for some constant ρ. Since $s > 2$ and h_0 is larger than a power of $1/T$, m is larger than a power of $T^{\bar{\gamma}}$ for some $\bar{\gamma} > 0$. Deduce froml Assumption 14 that the corresponding term T_{2n} is here $O\big(T^{\xi+\rho} \exp(-C_Y T^{\bar{\gamma}})\big) = o(1)$. This yields

$$\sup_{1 \leqslant i \leqslant N} \sup_{z \in \mathcal{C}} |S_{iT}(z) - \mathbb{E}[S_{it}(z)]| = O_P(a_T). \tag{C.2}$$

Moreover, Assumption 16 and a limited Taylor expansions of $(f_i r_i)(\cdot)$ provides

$$\begin{aligned}\mathbb{E}[S_{iT}(z)] &= \mathbb{E}[K_{h_0}(Z_{it} - z) r_i(Z_{it})] = \int K_{h_0}(x - z)(r_i f_i)(x)\, dx \\ &= \int K(u)(r_i f_i)(z + h_0 u)\, du = (f_i r_i)(z) + \frac{h_0^2}{2} \int K(u) u^\top \nabla^2 (r_i f_i)(z^*) u\, du,\end{aligned}$$

for some vectors z^* s.t. $\|z - z^*\|_\infty \leqslant C_0 h_0$ for some constant C_0. Note that we have invoked that K is even and compactly supported. This yields

$$\sup_{1 \leqslant i \leqslant N} \sup_{z \in \mathcal{C}} |\mathbb{E}[S_{iT}(z)] - (f_i r_i)(z)| = O(h_0^2). \tag{C.3}$$

The result is deduced from (C.2) and (C.3). ☐

Now, we focus on the uniform consistency of $\widehat{C}_i(z)$, $i \in \{1, \ldots, N\}$ when $z \in \mathcal{C}$. By

[13] We keep this terminology even if Y_{it} does not depend on i.

kernel smoothing, $\widehat{C}_i(z)$ is defined here as follows:

$$\widehat{C}_i(z) := \widehat{\mathrm{Var}}(G_t|Z_{it}=z)^{-1}\widehat{\mathrm{Cov}}(X_{it},G_t|Z_{it}=z), \tag{C.4}$$

$$\widehat{\mathrm{Cov}}(X_{it},G_t|Z_{it}=z) = \widehat{\mathbb{E}}[X_{it}G_t|Z_{it}=z] - \widehat{\mathbb{E}}[X_{it}|Z_{it}=z]\widehat{\mathbb{E}}[G_t|Z_{it}=z],$$
$$\widehat{\mathrm{Var}}(G_t|Z_{it}=z) = \widehat{\mathbb{E}}[G_tG_t^\top|Z_{it}=z] - \widehat{\mathbb{E}}[G_t|Z_{it}=z]\widehat{\mathbb{E}}[G_t^\top|Z_{it}=z],$$

and all the latter quantities are particular cases of the formula

$$\widehat{\mathbb{E}}[\xi_{it}|Z_{it}=z] := \frac{\sum_{t=1}^T K_{h_0}(Z_{it}-z)\xi_{it}}{\sum_{t=1}^T K_{h_0}(Z_{it}-z)},$$

where ξ_{it} is chosen as X_{it}, G_t, $X_{it}G_t$, $G_tG_t^\top$ or even $\xi_{it}=1$ everywhere.

We apply Lemma 6 several times, when Y_{it} is X_{it}, or is an arbitrarily chosen component of G_t, $X_{it}G_t$, $G_tG_t^\top$, or even when $Y_{it}=1$ everywhere. This will straightforwardly yield the uniform consistency of $\widehat{\mathbb{E}}[X_{it}|Z_{it}=z]$, $\widehat{\mathbb{E}}[G_t|Z_{it}=z]$, $\widehat{\mathbb{E}}[X_{it}G_t|Z_{it}=z]$ and $\widehat{\mathbb{E}}[G_tG_t^\top|Z_{it}=z]$. As a consequence, we obtain

$$\sup_{1\leqslant i\leqslant N}\sup_{z\in\mathcal{C}}\|\widehat{\mathrm{Cov}}(X_{it},G_t|Z_{it}=z)-\mathrm{Cov}(X_{it},G_t|Z_{it}=z)\| = O_P(a_T+h_0^2), \text{ and}$$

$$\sup_{1\leqslant i\leqslant N}\sup_{z\in\mathcal{C}}\|\widehat{\mathrm{Var}}(X_{it},G_t|Z_{it}=z)-\mathrm{Var}(X_{it},G_t|Z_{it}=z)\| = O_P(a_T+h_0^2). \tag{C.5}$$

For any matrix norm (Lütkepohl, 1997, Section 8.4.1, Equation (11)), we have

$$\begin{aligned}\|\widehat{\mathrm{Var}}(G_t|Z_{it}=z)^{-1}-\mathrm{Var}(G_t|Z_{it}=z)^{-1}\| &\leqslant \|\widehat{\mathrm{Var}}(G_t|Z_{it}=z)\|^{-1}\,\|\mathrm{Var}(G_t|Z_{it}=z)\|^{-1}\\ \times\quad &\|\widehat{\mathrm{Var}}(G_t|Z_{it}=z)-\mathrm{Var}(G_t|Z_{it}=z)\|.\end{aligned} \tag{C.6}$$

Let us provide a sufficient condition so that the denominators in (C.6) do not matter.

Assumption 17. *Let $\lambda_{i,G}(z)$ be the smallest eigenvalue of Var$(G_t|Z_{it}=z)$. We assume* $\inf_i \inf_{z\in\mathcal{C}}\lambda_{i,G}(z)>0$.

Under Assumption 17 and recalling (C.5), this provides

$$\sup_{1\leqslant i\leqslant N}\sup_{z\in\mathcal{C}}\|\widehat{\mathrm{Var}}(X_{it},G_t|Z_{it}=z)^{-1}-\mathrm{Var}(X_{it},G_t|Z_{it}=z)^{-1}\| = O_P(a_T+h_0^2).$$

Finally, we get the announced result:

Lemma 7. Suppose that Assumption 2 and 17 hold, and that the processes $(X_{it})_{i,t}$, $(G_t)_t$, $(X_{it}G_t^\top)_{i,t}$ and $(G_tG_t^\top)_t$ are $Z_{it}-$adapted. If $N=O(T^\xi)$ for some $\xi>0$ and $\ln T/T^\theta = o(h_0^d)$ for some $\theta\in(0,1)$, then

$$\sup_{1\leqslant i\leqslant N}\sup_{z\in\mathcal{C}}\|\widehat{C}_i(z)-C_i(z)\| = O_P\Big(\Big(\frac{\ln T}{Th_0^q}\Big)^{1/2}+h_0^2\Big).$$

Apart from a logarithmic factor, the previous convergence rate is standard in kernel regression. Usual bandwidth selectors can then be safely invoked to choose a convenient h_0.

Appendix D. Figures and Tables

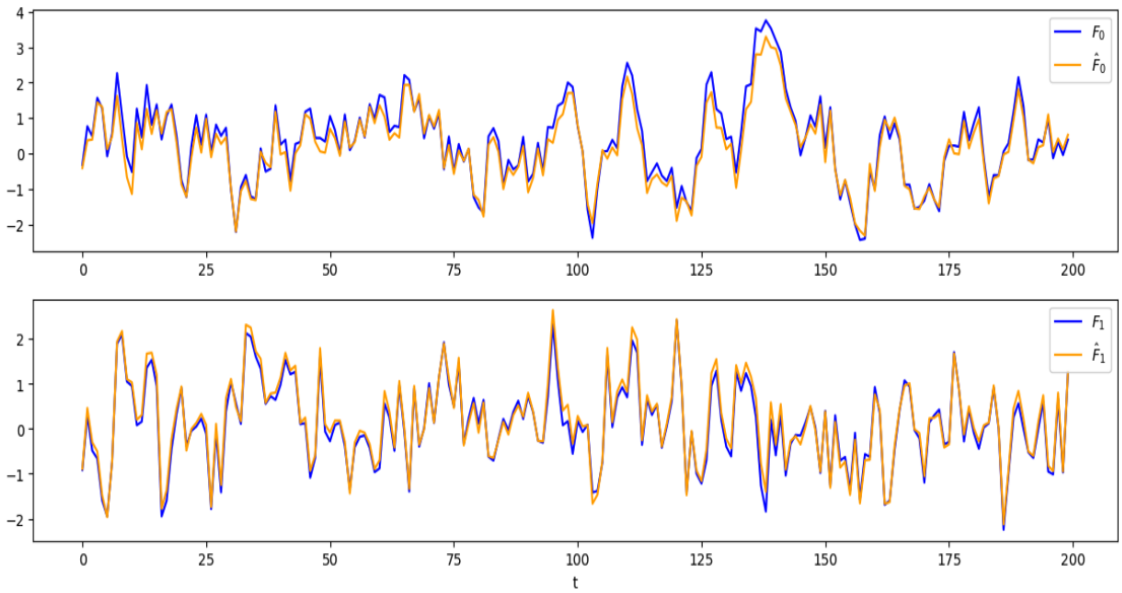

Figure D.2: A path of actual/estimated factors at 200 dates, in the simulated two factor model of Section 5.1.

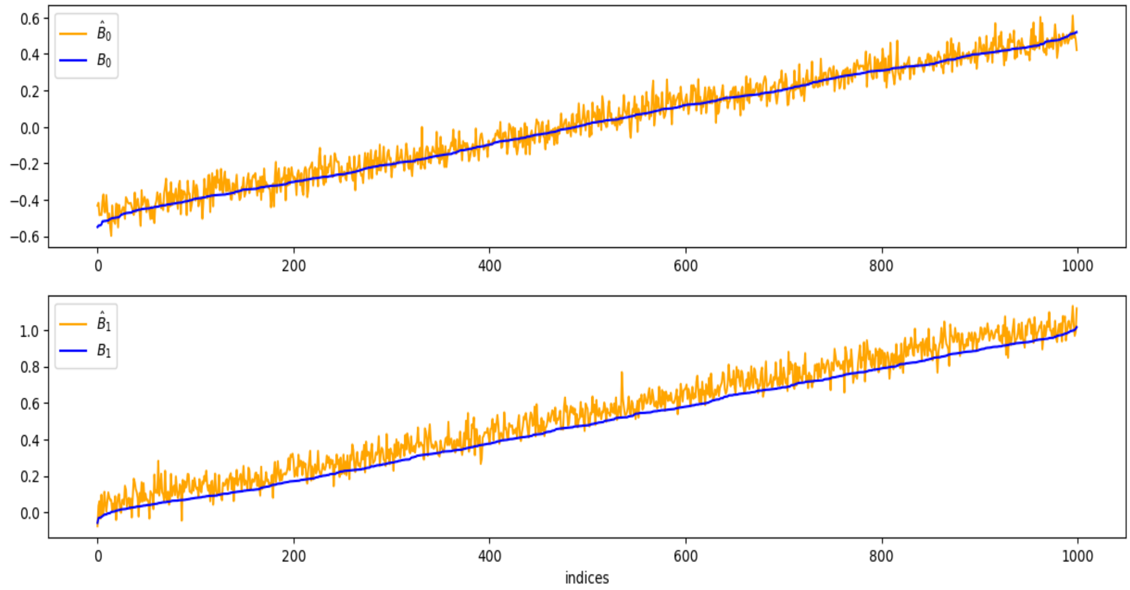

Figure D.3: Sorted values of the two loadings functions evaluated at z, compared to their corresponding estimated values

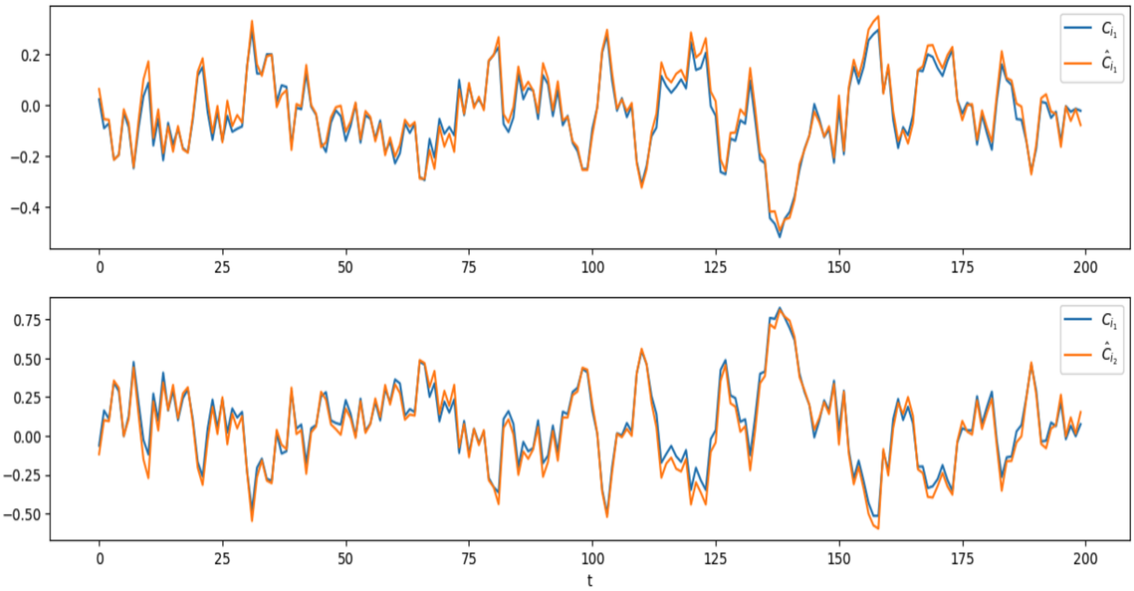

Figure D.4: An actual/estimated path of the common components, in the simulated two factor model of Section 5.1.

	$V_{F_1}^2$	$V_{F_2}^2$	$V_{F_3}^2$	M_{F_1}	M_{F_2}	M_{F_3}
r=1	1.263	-	-	0.082	-	-
r=2	1.131	1.003	-	0.144	0.092	-
r=3	0.999	1.042	0.975	0.135	0.150	0.121
	$V_{B_1}^2$	$V_{B_2}^2$	$V_{B_3}^2$	M_{B_1}	M_{B_2}	M_{B_3}
r=1	0.286	-	-	0.056	-	-
r=2	0.384	0.338	-	0.045	0.053	-
r=3	0.388	0.298	0.339	0.043	0.063	0.039
		V_C^2			M_{C_1}	
r=1		0.298			0.070	
r=2		0.337			0.081	
r=3		0.306			0.071	
		V_X^2			Com	
r=1		1.382			0.006	
r=2		1.014			0.010	
r=3		0.832			0.013	

Table D.4: Empirical MSE and variances of factors, loadings and common components. Results of Monte-Carlo experiment for $r \in \{1, 2, 3\}, p = 1$ and 100 simulated paths.

	$V_{F_1}^2$	$V_{F_2}^2$	$V_{F_3}^2$	M_{F_1}	M_{F_2}	M_{F_3}
r=1	0.976	-	-	0.124	-	-
r=2	1.112	0.994	-	0.175	0.122	-
r=3	1.017	0.971	0.997	0.152	0.169	0.213
	$V_{B_1}^2$	$V_{B_2}^2$	$V_{B_3}^2$	M_{B_1}	M_{B_2}	M_{B_3}
r=1	0.332	-	-	0.065	-	-
r=2	0.315	0.341	-	0.049	0.055	-
r=3	0.386	0.341	0.384	0.126	0.129	0.077
	$V_{C_1}^2$	$V_{C_2}^2$	$V_{C_3}^2$	M_{C_1}	M_{C_2}	M_{C_3}
r=1	0.335	0.332	0.336	0.074	0.072	0.079
r=2	0.341	0.342	0.340	0.081	0.082	0.085
r=3	0.304	0.305	0.313	0.076	0.066	0.083
		V_X^2			Com	
r=1		2.612			0.012	
r=2		2.261			0.017	
r=3		2.063			0.016	

Table D.5: Empirical MSE and variances of factors, loadings and common components. Results of Monte-Carlo experiment for $r \in \{1, 2, 3\}, p = 3$ and 100 simulated paths.

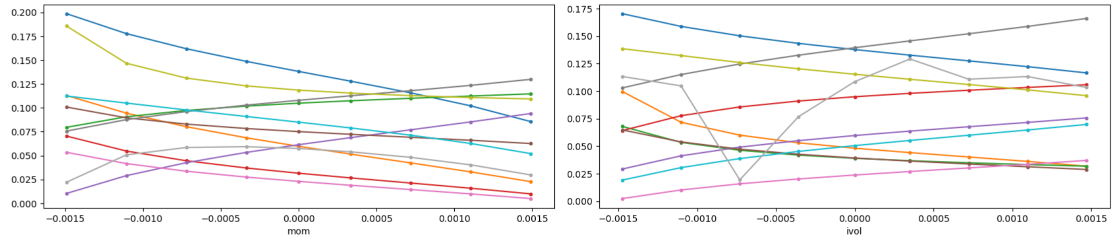

Figure D.5: Estimated loadings related to the second latent factor, on a grid of *mom* and *ivol* values, for some indexes i ($q = 1$ and $r = 3$)

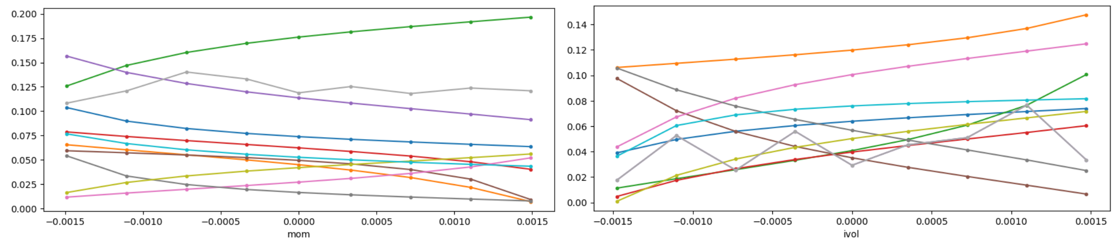

Figure D.6: Estimated loadings related to the third latent factor, on a grid of *mom* and *ivol* values, for some indexes i ($q = 1$ and $r = 3$)

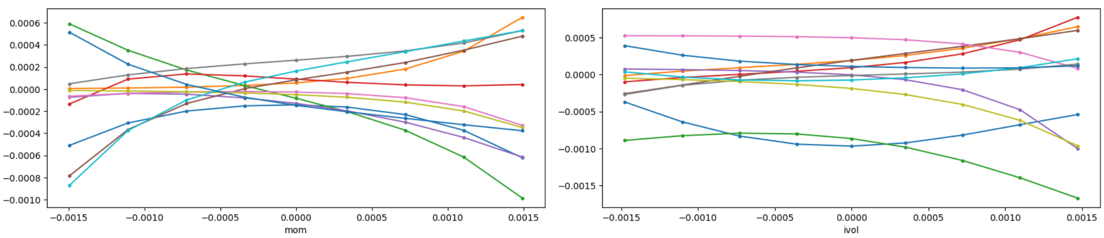

Figure D.7: Estimated loadings related to the first observed factor (*Market excess return*), on a grid of *mom* and *ivol* values, for some indexes i ($q = 1$ and $p = 3$)

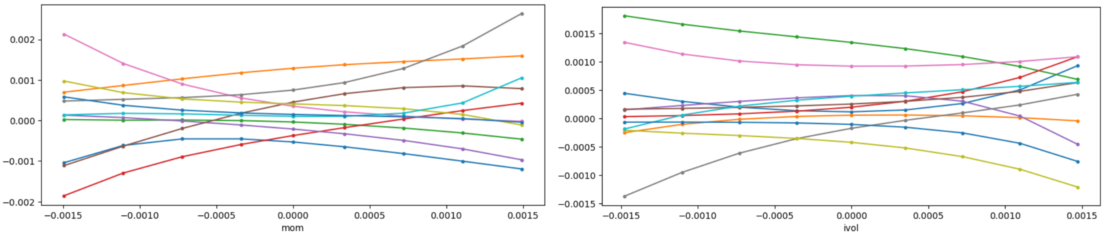

Figure D.8: Estimated loadings related to the second observed factor (SMB), on a grid of mom and $ivol$ values, for some indexes i ($q = 1$ and $p = 3$)

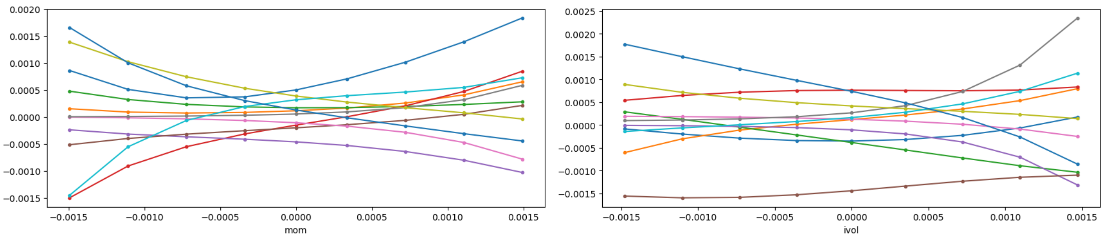

Figure D.9: Estimated loadings related to the third observed factor (HML), on a grid of mom and $ivol$ values, for some indexes i ($q = 1$ and $p = 3$)