Estimating a large dimensional factor model with noisy and asynchronous high frequency data*

Kun Lu[†]
Department of Statistics
University of Chicago

Dacheng Xiu[‡]
Booth School of Business
University of Chicago

This is a preliminary version: January 10, 2016

Abstract

In this paper, we estimate large dimensional factor models using high frequency data, which are asynchronous, and potentially contaminated by the market microstructure noise. Our strategy relies on the pre-averaging method to solve the microstructure problems, and on regressions to estimate models with known factors, or on the principal component analysis if factors are unknown. We provide the convergence rates of our estimators using the joint in-fill and increasing dimensionality asymptotics. On the empirical side, we provide an alternative measure of firms' network connectivity based upon correlations among their latent and idiosyncratic components.

Keywords: High-dimensional data, high-frequency latent factor model, pre-averaging estimator, regression, PCA, network connectivity.

1 Introduction

The arbitrage pricing theory by Ross (1976) and the inter-temporal capital asset pricing model by Merton (1973) suggest that factor models are sufficient to characterize the dynamics of asset prices. The literature has seen many successful factor models, e.g., the prominent three-factor (market, value, and size) model by Fama and French (1993), which demonstrate the empirical relevance of the theoretical prediction.

The large cross-section of transaction prices at high frequency presents a unique opportunity to explore the factor structure of asset returns empirically. On the one hand, the massive amount of

*Please find the most recent version using http://faculty.chicagobooth.edu/dacheng.xiu/research/FMN.pdf.

[†]Address: 5734 S. University Avenue, Chicago, IL 60637, USA. E-mail address: kunlu@galton.uchicago.edu.

[‡]Address: 5807 S Woodlawn Avenue, Chicago, IL 60637, USA. E-mail address: dacheng.xiu@chicagobooth.edu.

intraday data can turn the comovement among asset prices into "observable," in a simple and model-free way, e.g., Aït-Sahalia and Jacod (2014), so that several issues associated with low-frequency time series covering a long timespan become irrelevant, including structural breaks, latent variables, and time-varying parameters. On the other hand, statements about factor models can be made simpler and more general in very high-dimensional settings, e.g., Bai (2003) and Bai and Li (2012), which would otherwise be too complicated for moderate dimensions – the so called blessings of dimensionality, see Donoho et al. (2000).

The first step towards estimating a factor model is to estimate the covariance matrix as well as its eigenvalues and eigenvectors. However, the microstructure noise and the asynchronous arrival of trades, which come together with intraday data, result in biases of the sample covariance estimator with data sampled at a frequency higher than, say, every 5-minute. Most liquid stocks nowadays are traded every few seconds. Using 5-minute returns amounts to leaving a good deal of information aside, and more importantly, exacerbating the curse of dimensionality. With respect to illiquid stocks, microstructure noise plagues their volatility estimation even at the 5-minute sampling frequency. Moreover, there is the usual bias-variance trade-off in the sample covariance matrix estimator with intraday noisy data – the bias due to noise is larger when sampling at a higher frequency, and the variance becomes larger when the number of observations decreases with a lower frequency. This trade-off is magnified by the curse of dimensionality, which requires an even larger number of observations.

By adapting the pre-averaging estimator in the low-dimensional setting, e.g., Jacod, Li, Mykland, Podolskij, and Vetter (2009), we construct a noise-robust estimator of a large factor model, including the number of factors, the idiosyncratic covariance matrix, the total covariance matrix, and its inverse. We also estimate the latent factors along with their loadings up to certain rotations. If factors are known in priori, e.g., the Fama-French three-factor model, we adopt a regression approach using pre-averaged returns and pre-averaged factor returns. With respect to models with latent and unknown factors, we propose an estimator using the principal component analysis (PCA) of the pre-averaged returns. Using a large-deviation type of theory for martingales, we establish the desired consistency of our estimators under a variety of matrix norms. The simulation results agree with our theory.

Empirically, we estimate a network of the S&P 500 index components using factor models. The network is characterized by the correlations between the idiosyncratic shocks, similar to the network built by Bianchi, Billio, and Casarin (2014) using a Bayesian parametric approach. We also provide a network connectivity measure in the same spirit of Diebold and Yılmaz (2014). The estimated network suggests that firms within the same industry groups tend to form stronger links. Moreover, the connectedness tend to increase in financial turmoils.

Our paper is closely related to a growing literature on high frequency factor models. Fan, Furger, and Xiu (2015) and Aït-Sahalia and Xiu (2015b) develop the asymptotic theory for large dimensional models with known and unknown factors, respectively, based on synchronous and noiseless data. They provide the convergence rates of the covariance and precision matrices under various matrix norms. Pelger (2015a) and Pelger (2015b) also develop estimators of such models using noise-free data. Their asymptotic results are element-wise. In contrast, we analyze matrix-wise consistency results with noisy and asynchronous data. Wang and Zou (2010) propose the first noise-robust covariance matrix estimator in the high dimensional setting, by imposing the sparsity assumption on the covariance matrix itself, see also Tao, Wang, and Zhou (2013), Tao, Wang, Yao, and Zou (2011), and Tao, Wang, and Chen (2013) for related results. By comparison, we impose the sparsity condition on the covariance of the idiosyncratic components of a factor model, which is motivated from the economic theory and which fits the empirical data better.

Our paper is also related to the recent literature on the estimation of low-dimensional covariance matrix using noisy high frequency data. The noise-robust estimators include, among others, the multivariate realized kernels by Barndorff-Nielsen, Hansen, Lunde, and Shephard (2011), the quasi-maximum likelihood estimator by Aït-Sahalia, Fan, and Xiu (2010) and Shephard and Xiu (2014), the pre-averaging estimator by Christensen, Kinnebrock, and Podolskij (2010), the local method of moments by Bibinger, Hautsch, Malec, and Reiss (2014), and the two-scale and multi-scale estimators by Zhang (2011) and Bibinger (2012). We build our estimator based on the pre-averaging method because of its simplicity in deriving the large sample and high dimensional asymptotic results. Aït-Sahalia and Xiu (2015a) develop a related theory of principal component analysis for high frequency data.

In the high dimensional setting, Bai and Ng (2002) and Onatski (2010) propose estimators to determine the number of factors with low frequency data. Bai (2003) develops the element-wise inferential theory for factors and their loadings. They all consider approximate factor models introduced in Chamberlain and Rothschild (1983). Fan, Fan, and Lv (2008) propose a large covariance matrix estimator using a strict factor model with observable factors. Fan, Liao, and Mincheva (2011) extend this result to approximate factor models. Fan, Liao, and Mincheva (2013) develop the POET method for models with unobservable factors. Alternative covariance matrix estimators include the shrinkage method by Ledoit and Wolf (2004) and Ledoit and Wolf (2012), and the thresholding method proposed by Bickel and Levina (2008a), Bickel and Levina (2008b), Cai and Liu (2011) and Rothman, Levina, and Zhu (2009).

The rest of the paper is structured as follows. In section 2, we set up the model and discuss model assumptions. Section 3 proposes the estimation procedure for models with both observable and unobservable factors. Section 4 establishes the asymptotic properties of our estimators. Section 5 discusses the choice of tuning parameters. Section 6 provides simulation results. Section 7 gives an example that demonstrates the empirical relevance of our method. The appendix contains all mathematical proofs.

2 Model set-up and assumptions

Let $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}, \mathbb{P})$ be a filtered probability space. Throughout this paper, we use $\lambda_j(A)$, $\lambda_{\min}(A)$, and $\lambda_{\max}(A)$ to denote the jth, the minimum, and the maximum eigenvalues of a matrix A. In addition, we use $\|A\|_1$, $\|A\|$, and $\|A\|_{\mathrm{F}}$ to denote the $\mathbb{L}_1$ norm, the operator norm (or $\mathbb{L}_2$ norm), and the Frobenius norm of a matrix A, that is, $\max_{,j} \sum_i |A_{ij}|$, $\sqrt{\lambda_{\max}(A^\intercal A)}$, and $\sqrt{\mathrm{Tr}(A^\intercal A)}$, respectively. When A is a vector, both $\|A\|$ and $\|A\|_{\mathrm{F}}$ are equal to its Euclidean norm. We also use $\|A\|_{\mathrm{MAX}} = \max_{i,j} |A_{ij}|$ to denote the $\mathbb{L}_\infty$ norm of A on the vector space. We use e_i to denote a d-dimensional column vector whose ith entry is 1 and 0 elsewhere. We write $A_n \asymp B_n$ if $|A_n|/|B_n| = O(1)$. We use C to denote a generic constant that may change from line to line.

Let Y be a d-dimensional log-price process, X be a r-dimensional factor process, Z be the idiosyncratic component, and β be a constant factor loading matrix of size $d \times r$. We suppose that they satisfy a continuous-time factor model:

$$Y_t = \beta X_t + Z_t. \tag{1}$$

We consider two scenarios, depending on whether the factor X is known or not. In the case with known factors, we observe a noisy version of Y and a potentially noisy version of X, denoted as $Y^\star$ and $X^\star$, which satisfy:

$$Y^\star_{t_j} = Y_{t_j} + \varepsilon^y_j, \quad X^\star_{t_j} = X_{t_j} + \varepsilon^x_j,$$

where ε^y and ε^x are the additive noises. We can thereby rewrite the factor model as:

$$Y^\star_t = \beta X^\star_t + Z^\star_{ot},$$

where $Z^\star_{ot} = Z_t + \varepsilon^y_t - \beta \varepsilon^x_t$.

As to the case with unknown factors, we only observe a noisy version of Y, i.e., $Y^\star$, so the model can be written as:

$$Y^\star_t = \beta X_t + Z^\star_{ut},$$

where $Z^\star_{ut} = Z_t + \varepsilon^y_t$.

To fully specify the model, we make the following assumptions on the dynamics of factors and the idiosyncratic components.

Assumption 1. *Suppose the vector of the logarithm of asset prices Y_t follows a factor model given by (1), in which X_t is a continuous Itô semimartingale, that is,*

$$X_t = \int_0^t h_s \, ds + \int_0^t \eta_s dW_s,$$

and in addition, Z_t is another continuous Itô semimartingale satisfying

$$Z_t = \int_0^t f_s ds + \int_0^t \gamma_s dB_s.$$

We denote the spot covariance of X_t as $e_t = \eta_t \eta_t^\intercal$, and that of Z_t as $g_t = \gamma_t \gamma_t^\intercal$. W_t and B_t are independent Brownian motions. In addition, h_t and f_t are progressively measurable. Moreover, the processes η_t and γ_t are càdlàg, and e_t, e_{t-}, g_t, and g_{t-} are positive-definite. Finally, for all $1 \leq i, j \leq r$, $1 \leq k, l \leq d$, $|\beta_{kj}| \leq C$, for some $C > 0$, and there exists a locally bounded process H_t, such that $|h_{i,s}|$, $|\eta_{ij,s}|$, $|\gamma_{kl,s}|$, $|e_{ij,s}|$, $|f_{kl,s}|$, and $|g_{kl,s}|$ are all bounded by H_s for all $0 \leq s \leq t$.

As is common to the factor model literature, e.g., Bai and Ng (2002), Bai (2003), we impose a uniform bound for all processes. We also impose the usual exogeneity assumption:

Assumption 2. *For any $1 \leq j \leq r$, and $1 \leq k \leq d$, we have $[Z_{k,s}, X_{j,s}] = 0$, for any $0 \leq s \leq t$, where $[\cdot, \cdot]$ denotes the quadratic covariation.*

As a result, we can decompose the quadratic covariation of Y within $[0, t]$ as

$$\Sigma = \beta \mathrm{E} \beta^\intercal + \Gamma,$$

where for notational simplicity we omit the dependence of Σ, E, and Γ on the fixed t,

$$\Sigma = \frac{1}{t}\int_0^t c_s ds, \quad \Gamma = \frac{1}{t}\int_0^t g_s ds, \quad \text{and} \quad \mathrm{E} = \frac{1}{t}\int_0^t e_s ds. \tag{2}$$

We then impose that factors are pervasive, in the sense that they influence a large number of assets, see, e.g., Chamberlain and Rothschild (1983). In the case of known factors, we assume:

Assumption 3. *$\left\| d^{-1}\beta^\intercal \beta - B \right\| = o(1)$, as $d \to \infty$, for some positive-definite matrix B, with $\lambda_{\min}(B)$ bounded away from 0.*

For unknown factors, we can further impose restrictions on β without loss of generality:

Assumption 4. E *is a positive-definite covariance matrix, with distinct eigenvalues bounded away from 0. Moreover, $\left\| d^{-1}\beta^\intercal \beta - \mathbb{I}_r \right\| = o(1)$, as $d \to \infty$.*

With respect to the microstructure noise, we assume

Assumption 5. *Both $\{\varepsilon_j^x\}$ and $\{\varepsilon_j^y\}$ are sequence of i.i.d. variables with mean 0 and sub-Gaussian tails. Their variance is assumed to satisfy:*

$$\mathbb{E}\left(\varepsilon^x \varepsilon^{x\intercal}\right) = \Psi_1, \quad \textit{and} \quad \mathbb{E}\left(\varepsilon^y \varepsilon^{y\intercal}\right) = \Psi_2.$$

Moreover, they are independent of X, Y and Z.

Our results rely on the joint in-fill ($\Delta_n \to 0$) and increasing dimensionality ($d \to \infty$) asymptotics, with a fixed number of factors r. More specifically, we have

Assumption 6. *For some* $0 < \delta < 1/2$, $n^{-1/4+\delta/2}\sqrt{\log d} = o(1)$.

Finally, we impose a restriction on the degree of sparsity of the covariance matrix of idiosyncratic components.

Assumption 7. $n^{-1/4+\delta/2} m_d \sqrt{\log d} = o(1)$, *where,* m_d, *controls the degree of sparsity of* Γ.

$$m_d = \max_{i \leq d} \sum_{j \leq d} |\Gamma_{ij}|^q, \quad \text{for some } q \in [0, 1).$$

We assume that Γ belongs to a class of approximately sparse matrix as in Rothman, Levina, and Zhu (2009) and Bickel and Levina (2008b). Fan, Furger, and Xiu (2015) and Aït-Sahalia and Xiu (2015b) consider the case with $q = 0$, along with a block diagonal Γ.

Under this generalized measure of sparsity, we have

$$\|\Gamma\| \leq \|\Gamma\|_1 \leq \max_i \sum_{j=1}^{d} |\Gamma_{ij}|^q (\Gamma_{ii}\Gamma_{jj})^{(1-q)/2} = O(m_d).$$

3 Estimation

Since the observed transaction prices are usually asynchronous, we adopt the refresh time approach proposed by Barndorff-Nielsen, Hansen, Lunde, and Shephard (2011) to synchronize the data. We define the first refresh time as $t_1 = \max(t_1^1, t_1^2, \ldots, t_1^d)$, where t_j^i is the arrival time of the jth transaction price of asset i. The subsequent refresh times can be defined recursively as

$$t_{j+1} = \max\left(t^1_{N^1_{t_j}+1}, \ldots, t^d_{N^d_{t_j}+1}\right),$$

where $N^i_{t_j}$ is the number of arrivals for asset i prior to time t_j. We denote the resulting sample size after synchronization as n.

To deal with the bias due to the microstructure noise, we adopt the pre-averaging method proposed by Jacod, Li, Mykland, Podolskij, and Vetter (2009). In particular, we use a positive semi-definite version of it discussed by Christensen, Kinnebrock, and Podolskij (2010). This estimator is constructed as the sample covariance matrix of the pre-averaged returns over a sequence of overlapping blocks, so that the effect of the noise is dominated by a strengthened return signal of each block. Specifically, we choose the size of a local block to be k_n, such that

$$\frac{k_n}{n^{1/2+\delta}} = \theta + o(n^{-1/4+\delta/2}),$$

where θ is a positive constant and δ is given by Assumption 6.

The returns in each block are weighed by a continuously differentiable function g on $[0,1]$, with a piecewise Lipschitz derivative g'. Moreover, $g(0) = g(1) = 0$, and $\int_0^1 g^2(s)ds > 0$. We define $\psi_1 = \phi_1(0)$, $\psi_2 = \phi_2(0)$, where

$$\phi_1(s) = \int_s^t g'(u)g'(u-s)du, \quad \phi_2(s) = \int_s^t g(u)g(u-s)du.$$

For any sequence of synchronized vectors $\{V_j\}$, we define

$$\bar{V}_i = \sum_{j=1}^{k_n-1} g\left(\frac{j}{k_n}\right)\Delta_{i+j}^n V, \quad \text{for } i = 0, \ldots, n-k_n+1.$$

and $\Delta_i^n V = V_i - V_{i-1}$, for $i = 1, 2, \ldots, n$. Here V_i can denote $X_{t_i}, Y_{t_i}, Z_{t_i}$, $X_{t_i}^\star, Y_{t_i}^\star, Z_{t_i}^\star$, ε_i^x or ε_i^y.

3.1 Known factors

If the factor X is known but contaminated, we propose a regression-based approach using pre-averaged returns, i.e., $\bar{Y}^\star$ and $\bar{X}^\star$. We stack the $d \times 1$ and $r \times 1$ variables Y and X, $\bar{Y}^\star$ and $\bar{X}^\star$:

$$U := (Y^\intercal, X^\intercal)^\intercal, \quad \bar{U}^\star := (\bar{Y}^{\star\intercal}, \bar{X}^{\star\intercal})^\intercal.$$

The covariance matrix of U is given by

$$\Pi := \frac{1}{t}\int_0^t [dU_s, dU_s]ds = \frac{1}{t}\int_0^t \begin{pmatrix} \beta e_s \beta^\intercal + g_s & \beta e_s \\ e_s \beta^\intercal & e_s \end{pmatrix} ds =: \begin{pmatrix} \Pi_{11} & \Pi_{12} \\ \Pi_{21} & \Pi_{22} \end{pmatrix}.$$

which can be estimated by

$$\widehat{\Pi} = \frac{n}{n-k_n+2}\frac{1}{\psi_2 k_n t}\sum_{i=0}^{n-k_n+1} \bar{U}_i^\star \bar{U}_i^{\star\intercal}$$

where $\bar{U}_i^\star$ is the ith column of $\bar{U}^\star$. Therefore, we have

$$\widehat{\beta} = \widehat{\Pi}_{12}(\widehat{\Pi}_{22})^{-1}, \quad \widehat{\mathrm{E}} = \widehat{\Pi}_{22}, \quad \text{and} \quad \widehat{\Gamma} = \widehat{\Pi}_{11} - \widehat{\Pi}_{12}(\widehat{\Pi}_{22})^{-1}\widehat{\Pi}_{21}. \tag{3}$$

We then impose the sparsity of Γ, by applying hard-thresholding to the covariance matrix estimate $\widehat{\Gamma}$ and obtain:

$$\widehat{\Gamma}^S = \left(\widehat{\Gamma}_{ij}\mathbf{1}_{\{|\widehat{\Gamma}_{ij}| \geq S\}}\right),$$

for some threshold S. A plug-in covariance matrix estimator is therefore given by:

$$\widehat{\Sigma}_{\mathrm{REG}} = \widehat{\beta}\widehat{\mathrm{E}}\widehat{\beta}^\intercal + \widehat{\Gamma}^S.$$

We postpone the discussion of threshold selection to Section 5. Alternative thresholding methods, such as soft- or adaptive-thresholding, are also applicable, and our analysis below would be similar. We thereby only present the result of hard-thresholding. More discussions about thresholding techniques can be found in Rothman, Levina, and Zhu (2009) and Cai and Liu (2011).

3.2 Unknown factors

When factors are unknown, we only observe $Y^\star$. To estimate this model, we apply PCA to the pre-averaging estimates of the covariance matrix:

$$\widetilde{\Sigma} = \frac{n}{n-k_n+2}\frac{1}{\psi_2 k_n t}\sum_{i=0}^{n-k_n+1} \bar{Y}_i^\star \bar{Y}_i^{\star\intercal}.$$

Suppose that $\widehat{\lambda}_1 > \widehat{\lambda}_2 > \ldots > \widehat{\lambda}_d$ are the simple eigenvalues of $\widetilde{\Sigma}$, and that $\widehat{\xi}_1, \widehat{\xi}_2, \ldots, \widehat{\xi}_d$ are the corresponding eigenvectors. Then $\widetilde{\Sigma}$ can be decomposed as:

$$\widetilde{\Sigma} = \sum_{j=1}^{\widehat{r}} \widehat{\lambda}_j \widehat{\xi}_j \widehat{\xi}_j^\intercal + \widetilde{\Gamma}, \tag{4}$$

where $\widehat{r}$ is an estimator of r to be introduced below. Then we apply hard-thresholding on $\widetilde{\Gamma}$ and get:

$$\widetilde{\Gamma}^S = \left(\widetilde{\Gamma}_{ij}\mathbf{1}_{\{|\widetilde{\Gamma}_{ij}|\geq S)\}}\right). \tag{5}$$

Therefore we can obtain another plug-in estimator of the covariance matrix as:

$$\widehat{\Sigma}_{\text{PCA}} = \sum_{j=1}^{\widehat{r}} \widehat{\lambda}_j \widehat{\xi}_j \widehat{\xi}_j^\intercal + \widetilde{\Gamma}^S. \tag{6}$$

Our estimator can also be constructed from a least square point of view. We seek F and G such that

$$(F, G) = \arg \min_{F\in\mathcal{M}_{d\times\widehat{r}}, G\in\mathcal{M}_{\widehat{r}\times n}} \left\|\bar{Y}^\star - FG\right\|_{\text{F}}^2,$$

subject to the normalization

$$d^{-1}F^\intercal F = \mathbb{I}_{\widehat{r}}, \quad GG^\intercal \text{ is an } \widehat{r}\times\widehat{r} \text{ diagonal matrix.}$$

We then obtain the same $\widetilde{\Gamma}$ and $\widetilde{\Gamma}^S$ as above by:

$$\widetilde{\Gamma} = \frac{n}{n-k_n+2}\frac{1}{\psi_2 k_n t}\left(\bar{Y}^\star - FG\right)\left(\bar{Y}^\star - FG\right)^\intercal, \quad \widetilde{\Gamma}^S = \left(\widetilde{\Gamma}_{ij}\mathbf{1}_{\{|\widetilde{\Gamma}_{ij}|\geq S\}}\right),$$

with which we can construct

$$\widehat{\Sigma}_{\text{PCA}} = \frac{1}{t}FGG^\intercal F^\intercal + \widetilde{\Gamma}^S. \tag{7}$$

This representation is equivalent to (6). It is very useful in the proof.

To determine the number of factors, we propose the following estimator using a penalty function:

$$\widehat{r} = \arg \min_{1\leq j\leq r_{\max}} \left(d^{-1}\lambda_j(\widetilde{\Sigma}) + j\times g(n,d)\right) - 1,$$

where $r_{\max}$ is some upper bound of $r+1$. Our estimator is similar to that of Aït-Sahalia and Xiu (2015b), which shares the same spirit with Bai and Ng (2002).

The penalty function $g(n,d)$ satisfies two criterions. On the one hand, the penalty cannot dominate the signal, i.e., the value of $d^{-1}\lambda_j(\Sigma)$, when $1 \leq j \leq r$. Since $d^{-1}\lambda_r(\Sigma)$ is $O_p(1)$ as d increases, the penalty should shrink to 0. On the other hand, the penalty should dominate the estimation error as well as $d^{-1}\lambda_{r+1}(\Sigma)$ when $r+1 \leq j \leq d$ to avoid overshooting. The choice of $r_{\max}$ does not play any role in theory, yet it warrants an economically meaningful estimate of $\widehat{r}$ in finite sample or in practice.

4 Asymptotic properties

In this section, we provide the convergence rates of our estimators under a variety of matrix norms.

4.1 Known factors

Theorem 1. *Suppose that Assumptions 1 - 3 and 5 - 7 hold with $S \asymp n^{-1/4+\delta/2}\sqrt{\log d}$, then we have*

$$
\begin{aligned}
\left\|\widehat{\Sigma}_{\mathrm{REG}} - \Sigma\right\|_{\mathrm{MAX}} &= O_p\left(n^{-1/4+\delta/2}\sqrt{\log d}\right), \\
\left\|\widehat{\Sigma}_{\mathrm{REG}} - \Sigma\right\|_{\Sigma} &= O_p\left(d^{1/2}n^{-1/2+\delta}\log d + n^{-(1/4-\delta/2)(1-q)}m_d(\log d)^{(1-q)/2}\right), \\
\left\|\hat{\beta} - \beta\right\|_{\mathrm{MAX}} &= O_p\left(n^{-1/4+\delta/2}\sqrt{\log d}\right), \quad \textit{and} \\
\left\|\widehat{\Gamma}^S - \Gamma\right\|_{\mathrm{MAX}} &= O_p\left(n^{-1/4+\delta/2}\sqrt{\log d}\right),
\end{aligned}
$$

where $\|\cdot\|_{\Sigma}$ is as defined in Fan, Fan, and Lv (2008):

$$
\|\widehat{\Sigma}_{\mathrm{REG}} - \Sigma\|_{\Sigma} = \left(d^{-1}tr(\widehat{\Sigma}_{\mathrm{REG}}\Sigma^{-1} - \mathbb{I}_d)^2\right)^{1/2}.
$$

Theorem 1 establishes the consistency and convergence rate under both $\|\cdot\|_{\mathrm{MAX}}$ and $\|\cdot\|_{\Sigma}$ norms. The convergence rate depends on the sparsity level m_d, q, and δ. Under the $\|\cdot\|_{\mathrm{MAX}}$ norm, estimators with/without a factor model deliver the same rate; whereas under the $\|\cdot\|_{\Sigma}$ norm, without a factor model, the convergence rate is lower, i.e., $d^{1/2}n^{-1/4+\delta/2}$ under the $\|\cdot\|_{\Sigma}$ norm.

Theorem 2. *Suppose that Assumptions 1 - 3 and 5 - 7 hold and that $S \asymp n^{-1/4+\delta/2}\sqrt{\log d}$, then we have*

$$
\begin{aligned}
\left\|\left(\widehat{\Gamma}^S\right)^{-1} - \Gamma^{-1}\right\| &= O_p\left(n^{-(1/4-\delta/2)(1-q)}m_d(\log d)^{(1-q)/2}\right), \quad \textit{and} \\
\left\|\left(\widehat{\Sigma}_{\mathrm{REG}}\right)^{-1} - \Sigma^{-1}\right\| &= O_p\left(n^{-(1/4-\delta/2)(1-q)}m_d^3(\log d)^{(1-q)/2}\right).
\end{aligned}
$$

Moreover, $\lambda_{\min}(\widehat{\Sigma}_{\mathrm{REG}}) \geq \frac{1}{2}\lambda_{\min}(\Gamma)$, with probability approaching 1.

Theorem 2 establishes that our factor-based covariance matrix estimator is invertible with high probability, and the inverse is a consistent estimator of the precision matrix. More importantly, the minimum eigenvalue of the covariance matrix is bounded away from 0 with probability approaching 1, suggesting the matrix is well-conditioned. This property is essential to warrant an economically feasible optimal portfolio using our covariance matrix estimates as the input.

4.2 Unknown factors

In classical factor models, i.e., when dimension is small, the identification of factor models often require strong assumptions, such as a diagonal matrix Γ. In high dimensional setting, Chamberlain and Rothschild (1983) propose general approximate factor models and their identification requires the eigenvalues of Γ to be bounded. In contrast, Aït-Sahalia and Xiu (2015b) provide alternative identification conditions, allowing diverging eigenvalues of a block-diagonal Γ. Below, we provide an alterantive identification result:

Proposition 1. *Suppose that Assumptions 1 - 2 and 4 - 7 hold. Also, assume that* $\|\mathrm{E}\|_{\mathrm{MAX}} \leq K$, $\|\Gamma\|_{\mathrm{MAX}} \leq K$ *almost surely, and that* $d^{-1/2}m_d = o(1)$. *Then* r, $\beta \mathrm{E} \beta^\intercal$, *and* Γ *can be identified as* $d \to \infty$. *That is,* $\bar{r} = r$, *if* d *is sufficiently large. Moreover, we have*

$$\left\| \sum_{j=1}^{\bar{r}} \lambda_j \xi_j \xi_j^\intercal - \beta \mathrm{E} \beta^\intercal \right\|_{\mathrm{MAX}} \leq C d^{-1/2} m_d, \quad \text{and}$$
$$\left\| \sum_{j=\bar{r}+1}^{d} \lambda_j \xi_j \xi_j^\intercal - \Gamma \right\|_{\mathrm{MAX}} \leq C d^{-1/2} m_d,$$

where $\{\lambda_j, 1 \leq j \leq d\}$ *and* $\{\xi_j, 1 \leq j \leq d\}$ *are the eigenvalues and their corresponding eigenvectors of* Σ, *and* $\bar{r} = \arg\min_{1 \leq j \leq d}(d^{-1}\lambda_j + j d^{-1/2} m_d) - 1$.

Next, we show that the number of factors can be estimated consistently:

Theorem 3. *Suppose that Assumptions 1 - 2 and 4 - 7 hold, and suppose that* $d^{-1}m_d = o(1)$, $g(n,d) \to 0$, *and* $g(n,d)\left(n^{-1/4+\delta/2}\sqrt{\log d} + d^{-1}m_d\right)^{-1} \to \infty$, *we have* $\mathbb{P}(\widehat{r} = r) \to 1$.

Due to the fundamental indeterminacy in a factor model, we only identify the latent factors and their loadings up to some orthogonal transformations. That said, we can separate the idiosyncratic part of the covariance matrix from that explained by factors, which is the key step towards the construction of $\widehat{\Sigma}_{\mathrm{PCA}}$.

Theorem 4. *Suppose that Assumptions 1 - 2 and 4 - 7 hold with* $S \asymp n^{-1/4+\delta/2}\sqrt{\log d} + d^{-1/2}m_d$. *Also,* $d^{-1/2}m_d = o(1)$. *Suppose* $\widehat{r} \to r$ *with probability approaching 1, then there exists a* $r \times r$ *matrix*

H, such that with probability approaching 1, H is invertible, $\|HH^\intercal - \mathbb{I}_r\| = \|HH^\intercal - \mathbb{I}_r\| = o_p(1)$, then we have

$$\begin{aligned}
\left\|\widehat{\Sigma}_{\text{PCA}} - \Sigma\right\|_{\text{MAX}} &= O_p\left(n^{-1/4+\delta/2}\sqrt{\log d} + d^{-1/2}m_d\right), \\
\left\|\widehat{\Sigma}_{\text{PCA}} - \Sigma\right\|_{\Sigma} &= O_p\Big(d^{1/2}n^{-1/2+\delta}\log d + d^{-1/2}m_d^2 \\
&\quad + m_d\left(n^{-1/4+\delta/2}\sqrt{\log d} + d^{-1/2}m_d\right)^{1-q}\Big), \\
\|F - \beta H\|_{\text{MAX}} &= O_p\left(n^{-1/4+\delta/2}\sqrt{\log d} + d^{-1/2}m_d\right), \quad \textit{and} \\
\left\|G - H^{-1}X\right\| &= O_p\left(n^{-1/4+\delta/2}\sqrt{\log d} + d^{-1/2}m_d\right).
\end{aligned}$$

Theorem 5. *Suppose that Assumptions 1 - 2 and 4 - 7 hold with $S \asymp n^{-1/4+\delta/2}\sqrt{\log d} + d^{-1/2}m_d$. Suppose $d^{-1/2}m_d^2 = o(1)$ and $n^{-1/4+\delta/2}m_d\sqrt{\log d} = o(1)$, then $\lambda_{\min}(\widehat{\Sigma}_{\text{PCA}})$ is bounded away from 0 with probability approaching 1, and*

$$\begin{aligned}
\left\|\left(\widetilde{\Gamma}^S\right)^{-1} - \Gamma^{-1}\right\| &= O_p\left(m_d\left(n^{-1/4+\delta/2}\sqrt{\log d} + d^{-1/2}m_d\right)^{1-q}\right), \quad \textit{and} \\
\left\|\widehat{\Sigma}_{\text{PCA}}^{-1} - \Sigma^{-1}\right\| &= O_p\left(m_d^3\left(n^{-1/4+\delta/2}\sqrt{\log d} + d^{-1/2}m_d\right)^{1-q}\right).
\end{aligned}$$

There are two major differences of our theoretical results compared to those developed inFan, Liao, and Mincheva (2013). First, we do not assume a bounded eigenvalue or a bounded $\|\cdot\|_1$ of Γ as they do, which explains the m_d^3 in our rate for $\widehat{\Sigma}_{\text{PCA}}^{-1}$. Second, the because of the pre-averaging approach required for high-frequency data, the effective number of observations decreases, leading to the lower rate $n^{-1/4+\delta/2}$ instead of the usual $n^{-1/2}$ rate.

5 Practical considerations

5.1 Choice of the block-size k_n

The pre-averaging estimator involves a tuning parameter k_n, or equivalently, θ and δ. As discussed in a Christensen, Kinnebrock, and Podolskij (2010), our estimator is consistent if $0.1 < \delta < 0.5$. We fix δ at 0.2. With a large number of observations, the estimates are not sensitive to the choice of k_n. We follow the literature and use $\theta = 1$. In simulations, a range of k_ns are considered, all of which lead to similar estimates.

5.2 Choice of the penalty $g(n, d)$

A sensible choice of the penalty function could be

$$g(n, d) = \mu\left(n^{-1/4+\delta/2}\sqrt{\log d} + d^{-1}m_d\right)^{\kappa}, \tag{8}$$

for some μ and κ, which satisfies Assumption 3. As argued by Aït-Sahalia and Xiu (2015b), we can regard r as a tuning parameter. In our empirical studies, we find that as soon as $\widehat{r}$ is greater than 2 or 3, the interpretation of the results remain unchanged.

5.3 Choice of the threshold S

In general, the thresholding on Γ does not warrant a positive definite estimate in finite sample. We can achieve it by imposing this condition on the choice of S. Take the unknown factor case as an example. Recall that $S = C(n^{-1/4+\delta/2}\sqrt{\log d} + d^{-1/2}m_d)$. We need a C such that the smallest eigenvalue of $\widetilde{\Gamma}^{S(C)}$ is larger than 0. We define

$$\widetilde{C}_{\min} = \inf\left\{C > 0 : \lambda_{\min}\{\widetilde{\Gamma}^{S(C)}\} > 0\right\}.$$

$\widetilde{C}_{\min}$ is well-defined, because when C is sufficiently large, $\widetilde{\Gamma}^{S(C)}$ is diagonal. Therefore as long as $C > \widetilde{C}_{\min}$, we achieve a positive-definite estimate of $\widetilde{\Gamma}^{S}$. We can make a final choose of C by cross-validation, as discussed in Fan, Liao, and Mincheva (2013).

6 Monte Carlo Simulations

To verify the performance of our estimator, we sample 100 paths from a continuous-time r-factor model of d assets specified as:

$$dY_{i,t} = \sum_{j=1}^{r} \beta_{i,j} dX_{j,t} + dZ_{i,t}, \quad dX_{j,t} = b_j dt + \sigma_{j,t} dW_{j,t}, \quad dZ_{i,t} = \gamma_i^{\intercal} dB_{i,t},$$

where W_j is a standard Brownian motion and B_i is a d-dimensional Brownian motion, for $i = 1, 2, \ldots, d$, and $j = 1, 2, \ldots, r$. They are mutually independent. The (i, j)th entry of the spot covariance matrix of Z satisfies $g_{ij} = \chi^{|i-j|}$. We also allow for time-varying $\sigma_{j,t}$ which evolves according to the following system of equations:

$$d\sigma_{j,t}^2 = \kappa_j(\theta_j - \sigma_{j,t}^2)dt + \eta_j \sigma_{j,t} d\widetilde{W}_{j,t}, \quad j = 1, 2, \ldots, r,$$

where $\widetilde{W}_j$ is a standard Brownian motion with $\mathbb{E}[dW_{j,t} d\widetilde{W}_{j,t}] = \rho_j dt$. We choose $d = 500$ and $r = 3$. We fix t at 1 month, i.e., $t = 1/12$. In addition, $\kappa = (3, 4, 5)$, $\theta = (0.09, 0.04, 0.06)$, $\eta = (0.3, 0.4, 0.3)$, $\chi = 0.7$, $\rho = (-0.6, -0.4, -0.250)$, and $b = (0.05, 0.03, 0.02)$. $\beta_1 \sim \mathcal{U}[0.25, 2.25]$, and $\beta_2, \beta_3 \sim \mathcal{U}[-0.5, 0.5]$.

We simulate the noise in log prices from a normal distribution with mean 0 and variance 0.005^2, and the noise in factors (when used) from a normal distribution with mean 0 and variance 0.001^2. To mimic the asynchronicity, we censor the data using poisson sampling, where the number of observations for each stock is sampled from a truncated log normal distribution, as in Aït-Sahalia and Xiu (2015b).

	$\|\widehat{\Sigma}^S - \Sigma\|_{\text{MAX}}$		$\|(\widehat{\Sigma}^S)^{-1} - \Sigma^{-1}\|$			$\|\widehat{\Sigma}^S - \Sigma\|_{\text{MAX}}$		$\|(\widehat{\Sigma}^S)^{-1} - \Sigma^{-1}\|$	
k_n	REG	PCA	REG	PCA	k_n	REG	PCA	REG	PCA
	Δ_n=5					Δ_n=15			
36	1.971	1.971	4.960	4.960	16	1.378	1.378	4.726	4.728
96	0.561	0.561	3.627	3.627	36	0.483	0.483	3.540	3.541
156	0.365	0.365	10.266	10.265	56	0.309	0.309	11.289	11.289
	Δ_n=30					Δ_n=60			
10	1.109	1.109	4.683	4.682	7	0.774	0.774	4.512	4.512
20	0.412	0.412	3.595	3.596	12	0.367	0.367	3.748	3.747
30	0.262	0.262	10.163	10.162	17	0.260	0.260	11.226	11.225
	Δ_n=300					Δ_n=900			
2	1.255	1.255	4.931	4.908	2	0.623	0.623	4.268	4.244
3	0.515	0.515	4.029	4.048	3	0.479	0.479	33.477	13.357
4	0.381	0.381	65.987	51.709	4	0.441	0.440	436.734	272.753

Table 1: Choice of k_n in the pre-averaging estimator

Note: In this table, we report the values of $\|\widehat{\Sigma}^S - \Sigma\|_{\text{MAX}}$ and $\|(\widehat{\Sigma}^S)^{-1} - \Sigma^{-1}\|$ for different frequencies from 5 seconds to 15 mins and different case: observable factors and unobservable factors. The first column displays the sampling frequencies in seconds. Columns REG and PCA report the results of regression and the PCA methods respectively, using contaminated data. k_n stands for the averaging window size. We choose three different values for each scenario.

	$\|\widehat{\Sigma}^S - \Sigma\|_{\text{MAX}}$				$\|(\widehat{\Sigma}^S)^{-1} - \Sigma^{-1}\|$			
Freq	REG	PCA	REG	PCA	REG	PCA	REG	PCA
	Sample		Preaveraging		Sample		Preaveraging	
1	242.527	242.527	1.243	1.243	5.636	5.636	4.494	4.494
5	59.147	59.147	0.561	0.561	5.628	5.628	3.627	3.627
15	20.209	20.209	0.483	0.483	5.603	5.603	3.540	3.541
30	10.215	10.215	0.412	0.412	5.560	5.560	3.595	3.596
60	5.309	5.309	0.367	0.367	5.475	5.475	3.748	3.747
300	1.256	1.256	0.515	0.515	4.953	4.953	4.029	4.048
900	0.623	0.623	0.623	0.623	4.194	4.214	4.268	4.244

Table 2: Sample covariance matrix estimator versus the pre-averaging estimator

Note: In this table, we report the values of $\|\widehat{\Sigma} - \Sigma\|_{\text{MAX}}$ and $\|(\widehat{\Sigma}^S)^{-1} - \Sigma^{-1}\|$ for each subsampling frequency ranging from one observation every 1 seconds to 15 mins for both simple sample covariance matrix estimation and pre-averaging method. The first column displays the sampling frequencies in seconds. Columns REG and PCA report the results of regression and the PCA methods respectively, using contaminated data.

Table 1 shows that our pre-averaging estimator is robust to the choice of k_n. We also notice that the covariance matrix estimator performs relevantly well when k_n is large, whereas for the inverse estimator, the performance is better when k_n is smaller. Moreover, we also notice that the lower the sampling frequency, the smaller the window size we need. Especially, when the frequency is 900 seconds, i.e., 15-minute data, the optimal choice of k_n is around 2, so it is perhaps inappropriate and unnecessary to use pre-averaging method.

Table 2 compares our pre-averaging estimator with the simple covariance matrix estimators for under various frequencies, which do not use pre-averaging. We find that our pre-averaging estimator is robust to the microstructure noise and the asynchronity of the data. Whereas for the covariance matrix, when the frequency is high, the estimator performs badly. With a lower frequency, the performance of these two estimators improve.

7 Network connectivity measure of S&P 500 index components

We collect the data of intraday returns of all assets included in the S&P 500 index from Jan. 2004 to Dec. 2012 from the NYSE TAQ database. We also use the same factors as in Fan, Furger, and Xiu (2015) and Aït-Sahalia and Xiu (2015b): including the market portfolio, the small-minus-big market capitalization (SMB) portfolio, and high-minus-low price-earnings ratio (HML) portfolio in the Fama-French 3 factor model, as well as the daily-rebalanced momentum portfolio (MOM) formed by sorting stock returns between the past 250 days and 21 days. In addition, we collect 9 industry SDPR ETFs (Energy (XLE), Materials (XLB), Industrials (XLI), Consumer Discretionary (XLY), Consumer Staples (XLP), Health Care (XLV), Financial (XLF), Information Technology (XLK), and Utilities (XLU)).

We propose to build a network of the index components based on the correlations of the idiosyncratic parts with systematic factors. If there is a non-zero correlation between the idiosyncratic components of two firms, we draw a link in between the two corresponding nodes. This is an alternative to the variance decomposition approach by Diebold and Yılmaz (2014), the Granger-causality approach by Billio, Getmansky, Lo, and Pelizzon (2012), and the Bayesian approach by Bianchi, Billio, and Casarin (2014), etc.

Figure 1 provides the estimated network. For clarity of the presentation, we only show it for the financial sector. The length of the links reflects the magnitude of the residual correlations. We also color the notes according to their industry groups, given by the 3-4 digits of the Global Industry Classification Standard (GICS) codes. We find that within the finance sector, firms within the same industry group have a stronger correlation with each other.

In addition, we define a network connectivity measure as the total number of links in the estimated

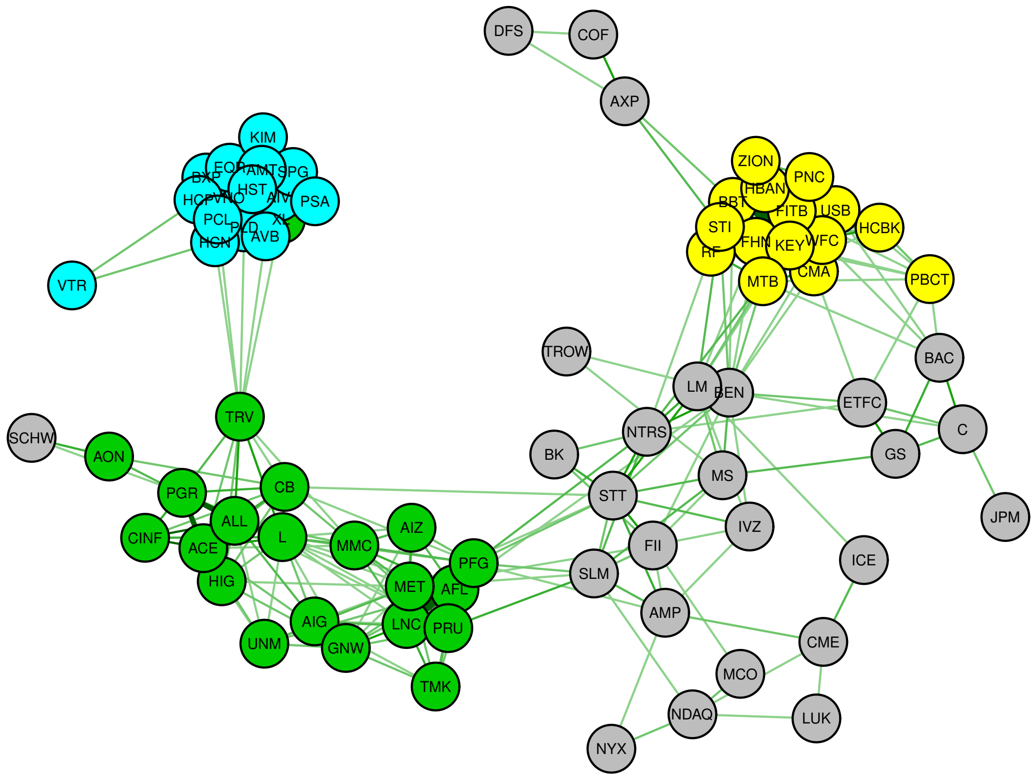

Figure 1: Network of the financial sector in 2012

Note: This figure provides the network of 80 financial sector firms which are components of the S&P 500 index in 2012. Each node stands for a firm marked by its ticker. We color the four industry groups defined by their GICS codes.

network. Figure 2 provides two monthly time series of the measure, based on a model with 4 known factors (market, momentum, value and size) and a model with 4 unknown ones. We make two observations. First, the model with unknown factors on average produces a smaller number of links, because the residual covariance matrix is more sparse than that of a 4-factor model with known factors. Second, the number of links increases sharply during financial turmoils.

References

Aït-Sahalia, Y., J. Fan, and D. Xiu (2010): "High-frequency covariance estimates with noisy and asynchronous financial data," *J. Amer. Statist. Assoc.*, 105(492), 1504–1517.

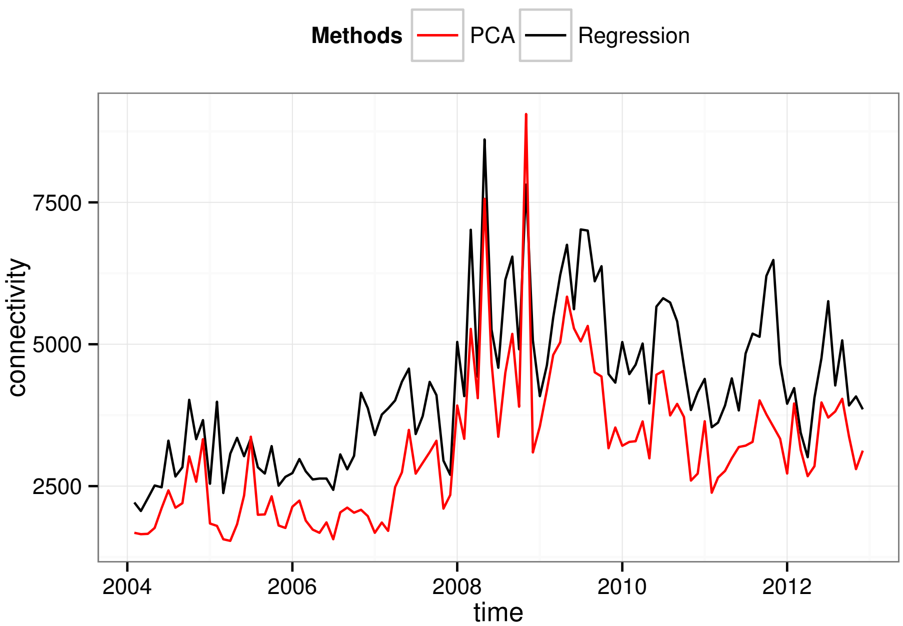

Figure 2: Time series plot of the network connectivity index

Note: In this figure, we plot the monthly time series of the number of total links for networks based on 4-factor models with known or unknown factors.

Aït-Sahalia, Y., and J. Jacod (2014): *High Frequency Financial Econometrics.* Princeton University Press.

Aït-Sahalia, Y., and D. Xiu (2015a): "Principal Component Analysis of High Frequency Data," Discussion paper, Princeton University and the University of Chicago.

——— (2015b): "Using Principal Component Analysis to Estimate a High Dimensional Factor Model with High-Frequency Data," *Chicago Booth Research Paper.*

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

Bai, J., and K. Li (2012): "Statistical analysis of factor models of high dimension," *Ann. Statist.*, 40(1), 436–465.

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

Barndorff-Nielsen, O. E., P. R. Hansen, A. Lunde, and N. Shephard (2011): "Multivariate realised kernels: consistent positive semi-definite estimators of the covariation of equity prices with noise and non-synchronous trading," *J. Econometrics*, 162(2), 149–169.

Bianchi, D., M. Billio, and R. Casarin (2014): "Modeling Contagion and Systemic Risk," *Available at SSRN 2537986.*

Bibinger, M. (2012): "An estimator for the quadratic covariation of asynchronously observed Itô processes with noise: Asymptotic distribution theory," *Stochastic Processes and their Applications*, 122, 2411–2453.

Bibinger, M., N. Hautsch, P. Malec, and M. Reiss (2014): "Estimating the quadratic covariation matrix from noisy observations: local method of moments and efficiency," *Ann. Statist.*, 42(4), 80–114.

Bickel, P. J., and E. Levina (2008a): "Regularized estimation of large covariance matrices," *Ann. Statist.*, 36(1), 199–227.

——— (2008b): "Covariance regularization by thresholding," *Ann. Statist.*, 36(6), 2577–2604.

Billio, M., M. Getmansky, A. W. Lo, and L. Pelizzon (2012): "Econometric measures of connectedness and systemic risk in the finance and insurance sectors," *Journal of Financial Economics*, 104(3), 535–559.

Cai, T., and W. Liu (2011): "Adaptive thresholding for sparse covariance matrix estimation," *J. Amer. Statist. Assoc.*, 106(494), 672–684.

Chamberlain, G., and M. Rothschild (1983): "Arbitrage, factor structure, and mean-variance analysis on large asset markets," *Econometrica*, 51(5), 1281–1304.

Christensen, K., S. Kinnebrock, and M. Podolskij (2010): "Pre-averaging estimators of the ex-post covariance matrix in noisy diffusion models with non-synchronous data," *J. Econometrics*, 159(1), 116–133.

Christensen, K., R. Oomen, and M. Podolskij (2010): "Realised quantile-based estimation of the integrated variance," *J. Econometrics*, 159(1), 74–98.

Diebold, F. X., and K. Yilmaz (2014): "On the network topology of variance decompositions: measuring the connectedness of financial firms," *J. Econometrics*, 182(1), 119–134.

Donoho, D. L., et al. (2000): "High-dimensional data analysis: The curses and blessings of dimensionality," *AMS Math Challenges Lecture*, pp. 1–32.

Fama, E. F., and K. R. French (1993): "Common risk factors in the returns on stocks and bonds," *Journal of financial economics*, 33(1), 3–56.

Fan, J., Y. Fan, and J. Lv (2008): "High dimensional covariance matrix estimation using a factor model," *J. Econometrics*, 147(1), 186–197.

Fan, J., A. Furger, and D. Xiu (2015): "Incorporating Global Industrial Classification Standard into Portfolio Allocation: A Simple Factor-Based Large Covariance Matrix Estimator with High Frequency Data," *Journal of Business and Economic Statistics, forthcoming.*

Fan, J., Y. Liao, and M. Mincheva (2011): "High-dimensional covariance matrix estimation in approximate factor models," *Ann. Statist.*, 39(6), 3320–3356.

——— (2013): "Large covariance estimation by thresholding principal orthogonal complements," *J. R. Stat. Soc. Ser. B. Stat. Methodol.*, 75(4), 603–680, With 33 discussions by 57 authors and a reply by Fan, Liao and Mincheva.

Jacod, J., Y. Li, P. A. Mykland, M. Podolskij, and M. Vetter (2009): "Microstructure noise in the continuous case: the pre-averaging approach," *Stochastic Process. Appl.*, 119(7), 2249–2276.

Ledoit, O., and M. Wolf (2004): "Honey, I shrunk the sample covariance matrix," *Journal of Portfolio Management*, (691).

——— (2012): "Nonlinear shrinkage estimation of large-dimensional covariance matrices," *Ann. Statist.*, 40(2), 1024–1060.

Merton, R. C. (1973): "An intertemporal capital asset pricing model," *Econometrica*, 41, 867–887.

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

Pelger, M. (2015a): "Large-Dimensional Factor Modeling Based on High-Frequency Observations," *Available at SSRN 2584172*.

——— (2015b): "Understanding systematic risk: A high-frequency approach," *Available at SSRN*.

Rivasplata, O. (2012): "Subgaussian random variables: An expository note," .

Ross, S. A. (1976): "The arbitrage theory of capital asset pricing," *J. Econom. Theory*, 13(3), 341–360.

Rothman, A. J., E. Levina, and J. Zhu (2009): "Generalized thresholding of large covariance matrices," *J. Amer. Statist. Assoc.*, 104(485), 177–186.

Saulis, L., and V. A. Statulevičius (1991): *Limit theorems for large deviations*, vol. 73 of *Mathematics and its Applications (Soviet Series)*. Kluwer Academic Publishers Group, Dordrecht, Translated and revised from the 1989 Russian original.

Shephard, N., and D. Xiu (2014): "Econometric analysis of multivariate realised QML: estimation of the covariation of equity prices under asynchronous trading," *Chicago Booth Research Paper*, (12-14).

Tao, M., Y. Wang, and X. Chen (2013): "Fast convergence rates in estimating large volatility matrices using high-frequency financial data," *Econometric Theory*, 29(4), 838–856.

Tao, M., Y. Wang, Q. Yao, and J. Zou (2011): "Large volatility matrix inference via combining low-frequency and high-frequency approaches," *J. Amer. Statist. Assoc.*, 106(495), 1025–1040.

Tao, M., Y. Wang, and H. H. Zhou (2013): "Optimal sparse volatility matrix estimation for high-dimensional Itô processes with measurement errors," *Ann. Statist.*, 41(4), 1816–1864.

Wang, Y., and J. Zou (2010): "Vast volatility matrix estimation for high-frequency financial data," *Ann. Statist.*, 38(2), 943–978.

Zhang, L. (2011): "Estimating covariation: Epps effect, microstructure noise," *J. Econometrics*, 160(1), 33–47.

Appendix A Proof of Theorem 1

To prove Theorem 1, we need to prove the following lemma 6 first.

Lemma 1. *Suppose that the entry-wise max norms of all the processes are bounded uniformly in [0,t]. Under assumptions 1, 2 and 3, for some C_0 large enough, we have*

$$\mathbb{P}\left(\left\|\widehat{\mathrm{E}}-\mathrm{E}\right\|_{\mathrm{MAX}} \geq C_0 n^{-1/4+\delta/2}\sqrt{\log d}\right) = O(r^2 C_1 d^{-C_0^2 C_2}), \tag{A.1}$$

$$\mathbb{P}\left(\left\|\widehat{\mathrm{E}}-\mathrm{E}\right\|_{F} \geq C_0 r n^{-1/4+\delta/2}\sqrt{\log d}\right) = O(r^2 C_1 d^{-C_0^2 C_2}), \tag{A.2}$$

$$\mathbb{P}\left(\max_{\substack{1\leq k\leq s\\1\leq l\leq d}}\left|\frac{n}{n-k_n+2}\frac{1}{\psi_2 k_n}\sum_{i=0}^{n-k_n+1}\bar{X}^{\star}_{k,i}\bar{Z}^{\star}_{o,li}\right| \geq C_0 n^{-1/4+\delta/2}\sqrt{\log d}\right) \tag{A.3}$$
$$= O(rC_1 d^{-C_0^2 C_2+1}),$$

$$\mathbb{P}\left(\max_{\substack{1\leq k\leq s\\1\leq l\leq d}}\left|\frac{n}{n-k_n+2}\frac{1}{\psi_2 k_n}\sum_{i=0}^{n-k_n+1}\bar{Z}^{\star}_{o,ki}\bar{Z}^{\star}_{o,li} - \int_0^t g_{s,lk}ds\right|\right. \tag{A.4}$$
$$\left.\geq C_0 n^{-1/4+\delta/2}\sqrt{\log d}\right) = O(C_1 d^{-C_0^2 C_2+2}),$$

$$\mathbb{P}\left(\max_{1\leq j\leq d}\left\|\hat{\beta}_j-\beta_j\right\| \geq C_0 r^{1/2} n^{-1/4+\delta/2}\sqrt{\log d}\right) = O(rC_1 d^{-C_0^2 C_2+1}), \tag{A.5}$$

$$\mathbb{P}\left(\left\|\hat{\beta}-\beta\right\|_{F} \geq C_0 r^{1/2} d^{1/2} n^{-1/4+\delta/2}\sqrt{\log d}\right) = O(rC_1 d^{-C_0^2 C_2+1}), \tag{A.6}$$

$$\mathbb{P}\left(\max_{1\leq k\leq d}\frac{n}{n-k_n+2}\frac{1}{\psi_2 k_n}\sum_{i=0}^{n-k_n+1}\left(\left(\beta-\hat{\beta}\right)^{\intercal}\bar{X}^{\star}_i\right)_k\right. \tag{A.7}$$
$$\left.\geq C_0 r^2 n^{-1/2+\delta}\log d\right) = O(r^2 C_1 d^{-C_0^2 C_2+1}),$$

$$\mathbb{P}\left(\max_{1\leq k,l\leq d}\left|\widehat{\Gamma}_{kl}-\Gamma_{kl}\right| \geq C_0 r n^{-1/4+\delta/2}\sqrt{\log d}\right) = O(C_1 d^{-C_0^2 C_2+2}), \tag{A.8}$$

$$\mathbb{P}\left(\max_{1\leq k,l\leq d}\left|\widehat{\Gamma}^{S}_{kl}-\Gamma_{kl}\right| \geq C_0 r n^{-1/4+\delta/2}\sqrt{\log d}\right) = O(C_1 d^{-C_0^2 C_2+2}). \tag{A.9}$$

Proof.

$$\begin{aligned}\widehat{\mathrm{E}}_{kl} =& \frac{n}{n-k_n+2}\frac{1}{\psi_2 k_n}\left(\sum_{i=0}^{n-k_n+1}\bar{X}_{k,i}\bar{X}_{l,i} + \sum_{i=0}^{n-k_n+1}\bar{X}_{k,i}\bar{\varepsilon}_{l,i} + \sum_{i=0}^{n-k_n+1}\bar{X}_{l,i}\bar{\varepsilon}_{k,i} + \sum_{i=0}^{n-k_n+1}\bar{\varepsilon}_{k,i}\bar{\varepsilon}_{l,i}\right)\\ =& T_1+T_2+T_3+T_4.\end{aligned}$$

Firstly, we decompose by

$$\begin{aligned}\mathbb{P}(|\widehat{\mathrm{E}}_{kl}-\mathrm{E}_{kl}|\geq u) \leq& \mathbb{P}(|T_1-\mathrm{E}_{kl}|\geq u/4) + \mathbb{P}(|T_2|\geq u/4)\\ &+ \mathbb{P}(|T_3|\geq u/4) + \mathbb{P}(|T_4|\geq u/4).\end{aligned}$$

For T_1, this expression can be furthered decomposed as

$$T_1 = \frac{n}{n-k_n+2}\frac{1}{\psi_2 k_n}\{\sum_{i=1}^{n-k_n+1} a_{1,i}(k,l)[X_{k,t_i^k}-X_{k,t_{i-1}^k}][X_{l,t_i^l}-X_{l,t_{i-1}^l}]$$

$$+\sum_{(i,j)\in F_{k,l}} b_{1,ij}(k,l)[X_{k,t_i^k}-X_{k,t_{i-1}^k}][X_{l,t_i^l}-X_{l,t_{i-1}^l}]\},$$

for certain numbers $a_{1,i}(k,l)$ and $b_{1,ij}(k,l)$. The set $F_{k,l}$ is given by

$$F_{k,l}=\{(i,j)|1\le i\le n-k_n+1, 1\le j\le n-k_n+1, |i-j|\le k_n-1, i\ne j\}.$$

and

$$|a_{1,i}(k,l)|+|b_{1,ij}(k,l)|\le Ck_n.$$

Let

$$A_{k,l}^i:=\frac{n}{n-k_n+2}\frac{a_{1,i}(k,l)}{\psi_2 k_n}=O_p(1).$$

We insert synchronized true price X_{k,t_i} and $X_{k,t_{i-1}}$ in between X_{k,t_i^k} and X_{k,t_{i-1}^k} and write

$$X_{k,t_i^k}-X_{k,t_{i-1}^k}=X_{k,t_i^k}-X_{k,t_i}+X_{k,t_i}-X_{k,t_{i-1}}+X_{k,t_{i-1}}-X_{k,t_{i-1}^k}.$$

Then using the above expression to expand $[X_{k,t_i^k}-X_{k,t_{i-1}^k}][X_{l,t_i^l}-X_{l,t_{i-1}^l}]$, we obtain the following decomposition

$$\begin{aligned}
&\sum_{i=1}^{n-k_n+1} A_{k,l}^i[X_{k,t_i^k}-X_{k,t_{i-1}^k}][X_{l,t_i^l}-X_{l,t_{i-1}^l}]\\
=&\sum_{i=1}^{n-k_n+1} A_{k,l}^i\Big\{[X_{k,t_i}-X_{k,t_{i-1}}][X_{l,t_i}-X_{l,t_{i-1}}]+[X_{k,t_i^k}-X_{k,t_i}][X_{l,t_i^l}-X_{l,t_i}]\\
&+[X_{k,t_i^k}-X_{k,t_i}][X_{l,t_i}-X_{l,t_{i-1}}]+[X_{k,t_i^k}-X_{k,t_i}][X_{l,t_{i-1}}-X_{l,t_{i-1}^l}]\\
&+[X_{k,t_i}-X_{k,t_{i-1}}][X_{l,t_i^l}-X_{l,t_i}]+[X_{k,t_i}-X_{k,t_{i-1}}][X_{l,t_{i-1}}-X_{l,t_{i-1}^l}]\\
&+[X_{k,t_{i-1}}-X_{k,t_{i-1}^k}][X_{l,t_i^l}-X_{l,t_i}]+[X_{k,t_{i-1}}-X_{k,t_{i-1}^k}][X_{l,t_i}-X_{l,t_{i-1}}]\\
&+[X_{k,t_{i-1}}-X_{k,t_{i-1}^k}][X_{l,t_{i-1}}-X_{l,t_{i-1}^l}]\Big\}.\\
\equiv&\sum_{i=1}^{n-k_n+1} A_{k,l}^i\Delta_i^n X_k\Delta_i^n X_l+H_{kl}^1(1)+\cdots+H_{kl}^1(8).
\end{aligned}$$

For $\sum_{i=1}^{n-k_n+1} A_{k,l}^i\Delta_i^n X_k\Delta_i^n X_l$, denote $X_t^*=\int_0^{\mathsf{T}}\sigma_s dW_s$, and denote for $1\le i\le n$, $1\le k,l\le p$,

$$\zeta_{i,kl}=(\Delta_i^n X_k^*)(\Delta_i^n X_l^*),\quad \zeta'_{i,kl}=\mathbb{E}((\Delta_i^n X_k^*)(\Delta_i^n X_l^*)|\mathcal{F}_{(i-1)\Delta_n}),$$
$$\text{and}\quad \zeta''_{i,kl}=\zeta_{i,kl}-\zeta'_{i,kl},$$

then $M_t=\sum_{i=1}^{n-k_n+1} A_{k,l}^i\zeta''_{i,kl}$ is a continuous-time martingale. By Itô's lemma, we have

$$\begin{aligned}
&\left(X_{t,k}^\star-X_{s,k}^*\right)\left(X_{t,l}^*-X_{s,l}^*\right)-\int_s^{\mathsf{T}}(\sigma\sigma^\mathsf{T})_{v,kl}dv\\
=&\int_s^{\mathsf{T}}\left(X_{v,k}^*-X_{s,k}^*\right)(\sigma_v dW_v)_l+\int_s^{\mathsf{T}}\left(X_{v,l}^*-X_{s,l}^*\right)(\sigma_v dW_v)_k.
\end{aligned}$$

Therefore

$$\zeta''_{i,kl} = \int_{(i-1)\Delta_n}^{i\Delta_n} \left(X^*_{s,k} - X^*_{(i-1)\Delta_n,k} \right) (\sigma_s dW_s)_l + \int_{(i-1)\Delta_n}^{i\Delta_n} \left(X^*_{s,l} - X^*_{(i-1)\Delta_n,l} \right) (\sigma_s dW_s)_k.$$

We can now write the target as M_t plus some remainder terms related to the drift:

$$\begin{aligned}
&\left| \sum_{i=1}^{n-k_n+1} A^i_{k,l} \left[X_{k,t^k_i} - X_{k,t^k_{i-1}} \right] \left[X_{l,t^l_i} - X_{l,t^l_{i-1}} \right] - \sum_{i=1}^{n-k_n+1} \int_{(i-1)\Delta_n}^{i\Delta_n} A^i_{k,l} (\sigma\sigma^\intercal)_{s,kl} ds \right| \\
=&\left| \sum_{i=1}^{n-k_n+1} A^i_{k,l} \Delta^n_i X^*_k \int_{(i-1)\Delta_n}^{i\Delta_n} b_{s,l} ds + \sum_{i=1}^{n-k_n+1} A^i_{k,l} \Delta^n_i X^*_l \int_{(i-1)\Delta_n}^{i\Delta_n} b_{s,k} ds \right. \\
&\left. + \sum_{i=1}^{n-k_n+1} A^i_{k,l} \int_{(i-1)\Delta_n}^{i\Delta_n} b_{s,l} ds \int_{(i-1)\Delta_n}^{i\Delta_n} b_{s,k} ds + M_t \right|.
\end{aligned}$$

We proceed with each of the four terms, starting with M_t.

The quadratic variation of M_t is given by

$$\begin{aligned}
\langle M, M \rangle_t =& \Delta_n \sum_{i=1}^{n-k_n+1} \int_{(i-1)\Delta_n}^{i\Delta_n} (A^i_{k,l})^2 \left(\left(X^*_{s,k} - X^*_{(i-1)\Delta_n,k} \right)^2 \sum_{r=1}^{q} \sigma^2_{s,lr} \right. \\
&+ \left(X^*_{s,l} - X^*_{(i-1)\Delta_n,l} \right)^2 \sum_{r=1}^{q} \sigma^2_{s,kr}. \\
&\left. +2 \left(X^*_{s,k} - X^*_{(i-1)\Delta_n,k} \right) \left(X^*_{s,l} - X^*_{(i-1)\Delta_n,l} \right) \sum_{r=1}^{q} \sigma_{s,lr}\sigma_{s,kr} \right) ds.
\end{aligned}$$

According to Assumption 1, here we assume that $\|X_t\|_\infty \le K$, $\|h_t\| \le K$, and $\|\sigma_t\sigma_t^\intercal\|_{\text{MAX}} \le K$, for some constant $K > 0$.Therefore by Cauchy-Schwarz inequality, we have

$$\langle M, M \rangle_t \le 16K^3 t\Delta_n.$$

Then by the exponential inequality for continuous martingale, we have

$$\mathbb{P}\left(\sup_{0\le s\le t} \left| \sum_{i=1}^{n-k_n+1} A^i_{k,l} \zeta''_{i,kl} \right| > u \right) \le \exp\left(-\frac{nu^2}{32K^3 t} \right). \tag{A.10}$$

In addition, by Cauchy-Schwarz inequality:

$$\begin{aligned}
&\mathbb{P}\left(\left| \sum_{i=1}^{n-k_n+1} A^i_{k,k} \Delta^n_i X^*_k \int_{(i-1)\Delta_n}^{i\Delta_n} b_{s,l} ds \right| > u \right) \\
\le&\mathbb{P}\left(\sum_{i=1}^{n-k_n+1} A^i_{k,k} (\Delta^n_i X^*_k)^2 - \sum_{i=1}^{n-k_n+1} \int_{(i-1)\Delta_n}^{i\Delta_n} A^i_{k,k} (\sigma\sigma^\intercal)_{s,kk} ds \right. \\
&\left. > \frac{x^2}{tK\Delta_n} - \sum_{i=1}^{n-k_n+1} \int_{(i-1)\Delta_n}^{i\Delta_n} A^i_{k,k} (\sigma\sigma^\intercal)_{s,kk} ds \right)
\end{aligned}$$

$$
\begin{aligned}
&\leq \mathbb{P}\left(\left|\sum_{i=1}^{n-k_n+1} A^i_{k,k}(\Delta^n_i X^*_k)^2 - \sum_{i=1}^{n-k_n+1}\int_{(i-1)\Delta_n}^{i\Delta_n} A^i_{k,k}(\sigma\sigma^\intercal)_{s,kk}ds\right| > \frac{u^2}{tK\Delta_n} - tK\right)\\
&\leq \exp\left(-\frac{\left(\frac{u^2}{tK\Delta_n} - tK\right)^2}{32K^3t\Delta_n}\right),
\end{aligned}
$$

where the last inequality follows from (A.10).

Finally, notice that

$$
\left|\sum_{i=1}^{n-k_n+1} A^i_{k,l}\int_{(i-1)\Delta_n}^{i\Delta_n} b_{s,l}ds\int_{(i-1)\Delta_n}^{i\Delta_n} b_{s,k}ds\right| \leq tK^2\Delta_n \max_i |A^i_{k,l}|,
$$

we can derive

$$
\begin{aligned}
&\mathbb{P}\left(\left|\sum_{i=1}^{n-k_n+1} A^i_{k,l}[X_{k,t_i} - X_{k,t_{i-1}}][X_{l,t_i} - X_{l,t_{i-1}}]\right.\right.\\
&\quad\left.\left. - \sum_{i=1}^{n-k_n+1}\int_{(i-1)\Delta_n}^{i\Delta_n} A^i_{k,l}(\sigma\sigma^\intercal)_{s,kl}ds\right| > 4u\right)\\
&\leq \mathbb{P}\left(|M_t| > u\right) + 2\mathbb{P}\left(\left|\sum_{i=1}^{n-k_n+1} A^i_{k,l}\Delta^n_i X^\star_k\int_{(i-1)\Delta_n}^{i\Delta_n} b_{s,l}ds\right| > u\right)\\
&\quad + \mathbf{1}_{\{tK^2\Delta_n \max_i |A^i_{k,l}| > u\}}\\
&\leq \exp\left(-\frac{u^2}{32K^3t\Delta_n}\right) + 2\exp\left(-\frac{\left(\frac{u^2}{tK\Delta_n} - tK\right)^2}{32K^3t\Delta_n}\right)\\
&\leq C_1\exp\left(-\frac{16C_2u^2}{\Delta_n}\right),
\end{aligned}
$$

where the above inequality holds if $x > (tK^2\Delta_n \max_i |A^i_{k,l}|) \vee (tK\sqrt{\Delta_n}) \vee (\frac{tK\Delta_n}{2}\sqrt{1 + \frac{4}{\Delta_n}})$, and $C_1 \geq 3$, $C_2 \leq (512K^3t)^{-1}$. On the other hand, if x violates this bound, i.e. $x \leq C'\sqrt{\Delta_n}$, we can choose C_1 such that $C_1\exp(-16C_2C'^2) \geq 1$, so that the inequality follows trivially. For $H^1_{kl}(1), \ldots H^1_{kl}(8)$, we can use exactly the same technique for proving $\sum_{i=1}^{n-k_n+1} A^i_{k,l}\Delta^n_i X_k \Delta^n_i X_l$, and have that

$$
\mathbb{P}\left(\left|H^1_{kl}(1) + \cdots + H^1_{kl}(8)\right| \geq u\right) \leq C_p e^{-C_p n u^2}.
$$

Since it is easy to check that

$$
\sum_{i=1}^{n-k_n+1}\int_{(i-1)\Delta_n}^{i\Delta_n} A^i_{k,l}(\sigma\sigma^\intercal)_{s,kl}ds = \int_0^\intercal (\sigma\sigma^\intercal)_{s,kl}ds + o(n^{-1/4+\delta/2}).
$$

Therefore, we have

$$
\mathbb{P}\left(\left|\frac{n}{n-k_n+2}\frac{1}{\psi_2 k_n}\sum_{i=1}^{n-k_n+1} a_{1,i}(k,l)A^i_{k,l}[X_{k,t^k_i} - X_{k,t^k_{i-1}}][X_{l,t^l_i} - X_{l,t^l_{i-1}}]\right.\right.
$$

$$
\begin{aligned}
&- \int_0^{\mathsf{T}} (\sigma\sigma^{\mathsf{T}})_{s,kl} ds \Bigg| \geq u \Bigg) \\
\leq & \mathbb{P}\left(\left| \sum_{i=1}^{n-k_n+1} A_{k,l}^i [X_{k,t_i} - X_{k,t_{i-1}}][X_{l,t_i} - X_{l,t_{i-1}}] \right.\right. \\
&\left.\left. - \sum_{i=1}^{n-k_n+1} \int_{(i-1)\Delta_n}^{i\Delta_n} A_{k,l}^i (\sigma\sigma^{\mathsf{T}})_{s,kl} ds \right| > u/2 \right) \\
&+ \mathbb{P}\left(\left| H_{kl}^1(1) + \cdots + H_{kl}^1(8) \right| \geq u/3 \right) \\
&+ \mathbb{P}\left(\left| \sum_{i=1}^{n-k_n+1} \int_{(i-1)\Delta_n}^{i\Delta_n} A_{k,l}^i (\sigma\sigma^{\mathsf{T}})_{s,kl} ds - \int_0^{\mathsf{T}} (\sigma\sigma^{\mathsf{T}})_{s,kl} ds \right| \geq u/3 \right) \\
\leq & C_p e^{-C_p n u^2} + I_{\{u < o(n^{-1/4+\delta/2})\}} \\
\leq & C_p e^{-C_p n u^2}.
\end{aligned}
$$

On the other hand, since we have

$$|a_{1,i}(k,l)| + |b_{1,ij}(k,l)| \leq C k_n,$$

and

$$
\begin{aligned}
&[X_{k,t_i^k} - X_{k,t_{i-1}^k}][X_{l,t_i^l} - X_{l,t_{i-1}^l}] = O(n^{-1}), \\
&[X_{k,t_i^k} - X_{k,t_{i-1}^k}][X_{l,t_i^l} - X_{l,t_{i-1}^l}] = O(n^{-1}).
\end{aligned}
$$

Let

$$X_{1,ij} = \frac{n}{n-k_n+2} \frac{n}{\psi_2 k_n} b_{1,ij}(k,l)[X_{k,t_i^k} - X_{k,t_{i-1}^k}][X_{l,t_i^l} - X_{l,t_{i-1}^l}],$$

then for $(i,j) \in F_{k,l}$,

$$X_{1,ij} = O_p(n^{-1}).$$

Using the same decomposition method, we can decompose

$$\sum_{(i,j)\in F_{k,l}} X_{1,ij} = \sum_{(i,j)\in F_{k,l}} X_{1,ij}^* + H_{kl}^2(1) + \cdots + H_{kl}^2(8).$$

Since $|X_t|$ is bounded by K, then $X_{1,ij} - \mathrm{E}(X_{1,ij})$ is also bounded by some constants. Then according to the Hoeffding's lemma, we know $X_{1,ij}^\star - \mathrm{E}(X_{1,ij}^\star)$ is a sub-Gaussian. Similar arguments can be extended to $H_{kl}^2(1), \ldots, H_{kl}^2(8)$. Then according to Hoeffding inequality, and $\sharp F_{kl} \leq C n k_n$, we have

$$
\begin{aligned}
& \mathbb{P}\left(\left| \frac{1}{n} \sum_{(i,j)\in F_{k,l}} X_{1,ij} \right| \geq u/8 \right) \\
\leq & \mathbb{P}\left(\left| \frac{1}{n} \sum_{(i,j)\in F_{k,l}} X_{1,ij}^* - \mathbb{E}\left(\frac{1}{n} \sum_{(i,j)\in F_{k,l}} X_{1,ij}^* \right) \right| \geq u/24 \right)
\end{aligned}
$$

$$
\begin{aligned}
&+\mathbb{P}\left(\left|H_{kl}^2(1)+\cdots+H_{kl}^2(8)\right|\geq u/24\right)\\
&+I\left(\left|\mathbb{E}\left(\frac{1}{n}\sum_{(i,j)\in F_{k,l}}X_{1,ij}\right)\right|\geq u/24\right)\\
&\leq C_p e^{-\frac{C_p(nu)^2}{nk_n}}+\mathbf{1}_{\left\{\left|\mathbb{E}\left(\frac{1}{n}\sum_{(i,j)\in F_{k,l}}X_{1,ij}\right)\right|\geq u/24\right\}}\\
&\leq C_p e^{-C_p n^{1/2-\delta}u^2}.
\end{aligned}
$$

This inequality holds when $u \geq Ck_n/n$ for some constant C.

As for T_4, it can be decomposed as

$$
T_4=\frac{n}{n-k_n+2}\frac{1}{\psi_2 k_n}\left(\sum_{i=0}^{n-k_n+1}a_{4,i}(k,l)\varepsilon_{k,t_i^k}\varepsilon_{l,t_i^l}+\sum_{(i,j)\in F_{k,l}}b_{4,ij}(k,l)\varepsilon_{k,t_i^k}\varepsilon_{l,t_j^l}\right).
$$

Since $\bar{\varepsilon}_k(t_i^k)=\sum_{h=1}^{k_n}\left\{g\left(\frac{i}{k_n}\right)-g\left(\frac{i-1}{k_n}\right)\right\}\varepsilon_k(t_{i+h}^k)$, and g is continuous, piecewise continuously differentiable with a piecewise Lipschitz derivative g'. Then

$$
|a_{4,i}(k,l)|+|b_{4,ij}(k,l)|\leq\frac{C}{k_n}.
$$

Let

$$
\varepsilon_{4,i}=\frac{n}{n-k_n+2}\frac{k_n}{\psi_2}a_{4,i}(k,l)\varepsilon_{k,t_i^k}\varepsilon_{l,t_i^l},
$$

and

$$
\varepsilon_{4,ij}=\frac{n}{n-k_n+2}\frac{k_n}{\psi_2}a_{4,ij}(k,l)\varepsilon_{k,t_i^k}\varepsilon_{l,t_j^l}.
$$

A straightforward calculation shows that

$$
\mathbb{E}\left(\frac{1}{k_n^2}\sum_{i=0}^{n-k_n+1}\varepsilon_{4,i}\right)=\frac{\psi_1}{\theta^2\psi_2 n^{2\delta}}\Psi_{1,kl}+o(n^{-1/4+\delta/2}),\quad\mathbb{E}(\varepsilon_{4,ij})=0.
$$

Firstly, we assume ε_t to be a sub-gaussian, then according to Proposition 3.2 in Rivasplata (2012), we know

$$
\mathbb{E}\left[\varepsilon_{k,t_i^k}\varepsilon_{l,t_i^l}\right]^k\leq\sqrt{\mathbb{E}(|\varepsilon_{k,t_i^k}|^{2k})\mathbb{E}(|\varepsilon_{l,t_i^l}|^{2k})}\leq C(k!)^2k^2 2^{k+1}.
$$

Then applying Theorem 4.16 in Saulis and Statulevičius (1991), we can derive a bound on the kth cumulant of $\sum_{i=0}^{n-k_n+1}\varepsilon_{4,i}$, then using Lemma 2.3 and Lemma 2.4 in Saulis and Statulevičius (1991), we can establish a exponential tail probability for $1/k_n^2\sum_{i=0}^{n-k_n+1}\varepsilon_{4,i}$:

$$
\begin{aligned}
&\mathbb{P}\left(\left|\frac{1}{k_n^2}\sum_{i=0}^{n-k_n+1}\varepsilon_{4,i}\right|\geq u/8\right)\\
\leq&\mathbb{P}\left(\left|\frac{1}{k_n^2}\sum_{i=0}^{n-k_n+1}\varepsilon_{4,i}-\mathbb{E}\left(\frac{1}{k_n^2}\sum_{i=0}^{n-k_n+1}\varepsilon_{4,i}\right)\right|\geq u/16\right)
\end{aligned}
$$

$$
\begin{aligned}
&+ \mathbf{1}_{\{(u/16 \leq \frac{\psi_1}{\theta^2 \psi_2 n^{2\delta}} \Psi_{1,kl} + o(n^{-1/4+\delta/2})\}} \\
\leq & C_p e^{-C_p n^{1+4\delta} u^2}.
\end{aligned}
$$

Correspondingly, we have

$$
\mathbb{P}\left(\left| \frac{1}{k_n^2} \sum_{(i,j) \in F_{k,l}} \varepsilon_{4,ij} \right| \geq u/8 \right) \leq C_p e^{-C_p n^{\frac{1}{2}+3\delta} u^2},
$$

For T_2 and T_3, since they are similar, we only give inference details for T_2,

$$
T_2 = \frac{n}{n - k_n + 2} \frac{1}{\psi_2 k_n} \sum_{(i,j) \in G_{k,l}} a_{2,ij}(k,l)[X_{k,t_i^k} - X_{k,t_{i-1}^k}] \varepsilon_{l,t_j^l}.
$$

Here

$$
G_{k,l} = \{(i,j) | 0 \leq i \leq n - k_n + 1, 0 \leq j \leq n - k_n + 1, |i - j| \leq k_n - 1\},
$$

and

$$
|a_{2,ij}(k,l)| \leq C.
$$

Since $\sharp G_{k,l} \leq C n k_n$, and $[X_{k,t_i^k} - X_{k,t_{i-1}^k}] = O(\frac{1}{\sqrt{n}})$, we define

$$
\varepsilon_{2,ij} = \frac{n}{n - k_n + 2} \frac{\sqrt{n}}{\psi_2} a_{2,ij}(k,l)[X_{k,t_i^k} - X_{k,t_{i-1}^k}] \varepsilon_{l,t_j^l}.
$$

If ε_t has sub-gaussian tail, then using the same technique for proving T_4, we have

$$
\mathbb{P}\left(\left| \frac{1}{\sqrt{n} k_n} \sum_{(i,j) \in G_{k,l}} \varepsilon_{2,ij} \right| \geq u/4 \right) \leq C_p e^{-C_p n^{\frac{1}{2}+\delta} u^2}.
$$

Similarly,

$$
\mathbb{P}(|T_3| \geq u/4) \leq C_p e^{-C_p n^{\frac{1}{2}+\delta} u^2}.
$$

respectively. Above all, we prove that

$$
\mathbb{P}(|\widehat{\mathrm{E}}_{kl} - \mathrm{E}_{k,l}| \geq u) \leq C_1 e^{-C_2 n^{1/2-\delta} u^2}.
$$

Thus

$$
\begin{aligned}
&\mathbb{P}\left(\left\| \widehat{\mathrm{E}} - \mathrm{E} \right\|_{\mathrm{MAX}} \geq C_0 n^{-1/4+\delta/2} \sqrt{\log d} \right) \\
\leq & C_1 r^2 e^{-C_2 n^{1/2-\delta} (C_0 n^{-1/4+\delta/2} \sqrt{\log d})^2}. \\
= & C_1 d^{-C_0^2 C_2}.
\end{aligned}
$$

Since

$$
\left\| \widehat{\mathrm{E}} - \mathrm{E} \right\|_F \leq r \left\| \widehat{\mathrm{E}} - \mathrm{E} \right\|_{\mathrm{MAX}},
$$

then

$$\mathbb{P}\left(\left\|\widehat{\mathrm{E}}-\mathrm{E}\right\|_F \geq C_0 r n^{-1/4+\delta/2}\sqrt{\log d}\right) \leq C_1 d^{-C_0^2 C_2}.$$

Because $\bar{X}^\star_{k,i}\bar{Z}^\star_{o,li}$ is similar to $\bar{X}^\star_{k,i}\bar{X}^\star_{l,i}$, therefore we can directly get

$$\mathbb{P}\left(\max_{1\leq k\leq s, 1\leq l\leq d}\left|\frac{n}{n-k_n+2}\frac{1}{\psi_2 k_n}\sum_{i=0}^{n-k_n+1}\bar{X}^\star_{k,i}\bar{Z}^\star_{o,li}\right| \geq C_0 n^{-1/4+\delta/2}\sqrt{\log d}\right) \leq C_1 d^{-C_0^2 C_2+1}.$$

In addition, we have

$$\begin{aligned}&\mathbb{P}\left(\max_{1\leq k\leq s, 1\leq l\leq d}\left|\frac{n}{n-k_n+2}\frac{1}{\psi_2 k_n}\sum_{i=0}^{n-k_n+1}\bar{Z}^\star_{o,ki}\bar{Z}^\star_{o,li} - \int_0^t g_{s,kl}ds\right| \geq C_0 n^{-1/4+\delta/2}\sqrt{\log d}\right)\\ \leq& C_1 d^{-C_0^2 C_2+2}.\end{aligned}$$

Moreover, note that

$$\hat{\beta}_j - \beta_j = \left(\hat{\Pi}^{22}\right)^{-1}\hat{\Pi}^{12}_j,$$

therefore, under the event that

$$\begin{aligned}A = &\left\{\max_{1\leq i\leq s, 1\leq j\leq d}\left|\frac{n}{n-k_n+2}\frac{1}{\psi_2 k_n}\sum_{i=0}^{n-k_n+1}\bar{X}^\star_{k,i}\bar{Z}^\star_{o,li}\right| \leq C_0 n^{-1/4+\delta/2}\sqrt{\log d}\right\}\\ &\cap\left\{\lambda_{\min}\left(\widehat{\mathrm{E}}\right) \geq \frac{1}{2}\lambda_{\min}\left(\int_0^t e_s ds\right)\right\},\end{aligned}$$

We have

$$\|\hat{\beta}_j - \beta_j\|^2 \leq \frac{4}{\lambda^2_{\min}\left(\int_0^t e_s ds\right)}\sum_{i=1}^r\left(\hat{\Pi}^{12}_{ij}\right)^2 \leq \frac{4rC_0^2 n^{-1/2+\delta}\log d}{\lambda^2_{\min}(\int_0^t e_s ds)}.$$

and

$$\|\hat{\beta} - \beta\|_F^2 \leq \frac{4rC_0^2 n^{-1/2+\delta} d\log d}{\lambda^2_{\min}(\int_0^t e_s ds)}.$$

Therefore, it suffices to show that $\mathbb{P}(A) \geq 1 - O(rn^{3/2+\delta-3\nu/4-\nu\delta/2}d^{-1})$.

We assume $\lambda_{\min}\left(\int_0^t e_s ds\right)$ is bounded away from 0 and that $r = o(n^{-1/4+\delta/2}\sqrt{\log d})$, then it follows that

$$\begin{aligned}&\mathbb{P}\left(\left\|\widehat{\mathrm{E}} - \int_0^t e_s ds\right\| \leq \frac{1}{2}\lambda_{\min}\left(\int_0^t e_s ds\right)\right)\\ \geq&\mathbb{P}\left(r\max_{1\leq i,j\leq s}\left|\widehat{\mathrm{E}}_{ij} - \int_0^t e_{ij,s}ds\right| \leq \frac{1}{2}\lambda_{\min}\left(\int_0^t e_s ds\right)\right)\\ \geq&1 - O(C_1 d^{-C_0^2 C_2}),\end{aligned}$$

then by Lemma A.1 of Fan, Liao, and Mincheva (2011), we have

$$\mathbb{P}\left(\lambda_{\min}(\widehat{\mathrm{E}}) \geq \frac{1}{2}\lambda_{\min}\left(\int_0^t e_s ds\right)\right)$$

$$\geq 1 - O(C_1 d^{-C_0^2 C_2}). \tag{A.11}$$

Combining this with (A.3), we have $\mathbb{P}(A) \geq 1 - O(C_1 d^{-C_0^2 C_2 + 1})$.

(vii) To prove (A.7), we note that

$$\max_{1\leq k\leq d} \frac{n}{n-k_n+2}\frac{1}{\psi_2 k_n}\sum_{i=0}^{n-k_n+1}\left(\sum_{l=1}^{r}(\beta_{k,l}-\hat{\beta}_{k,l})\bar{X}_{l,i}\right)^2$$
$$\leq \max_{1\leq k\leq d}\|\hat{\beta}_k-\beta_k\|^2\frac{n}{n-k_n+2}\frac{1}{\psi_2 k_n}\sum_{i=0}^{n-k_n+1}\|\bar{X}_i^\star\|^2,$$

Then by (A.1) with $C > \max_{1\leq k\leq s}\int_0^t e_{s,kk}ds$,

$$\mathbb{P}\left(\frac{n}{n-k_n+2}\frac{1}{\psi_2 k_n}\sum_{i=0}^{n-k_n+1}\|\bar{X}_i^\star\|^2\leq sC\right)$$
$$\geq \mathbb{P}\left(s\max_{1\leq k\leq s}\left|\frac{n}{n-k_n+2}\frac{1}{\psi_2 k_n}\sum_{i=0}^{n-k_n+1}\|\bar{X}_i^\star\|^2-\int_0^t e_{s,kk}ds\right|+r\max_{1\leq k\leq s}\int_0^t e_{s,kk}ds\leq sC\right)$$
$$\geq 1-O(C_1 d^{-C_0^2 C_2}),$$

and by (A.5), we get

$$\mathbb{P}\left(\max_{1\leq k\leq d}\frac{n}{n-k_n+2}\frac{1}{\psi_2 k_n}\sum_{i=0}^{n-k_n+1}\left(\sum_{l=1}^{r}(\beta_{k,l}-\hat{\beta}_{k,l})\bar{X}_{l,i}^\star\right)^2 > Cn^{-1/2+\delta}r^2\log d\right)$$
$$\leq O(C_1 d^{-C_0^2 C_2+1}).$$

(viii) Finally, under the event of

$$\left\{\max_{1\leq l\leq d}\left|\frac{n}{n-k_n+2}\frac{1}{\psi_2 k_n}\sum_{i=0}^{n-k_n+1}\bar{Z}_{o,ki}^\star\bar{Z}_{o,li}^\star-\int_0^t g_{s,ll}ds\right|\leq\frac{1}{4}\max_{1\leq l\leq d}\int_0^t g_{s,ll}ds\right\}$$
$$\cap\left\{\max_{1\leq k\leq d}\frac{n}{n-k_n+2}\frac{1}{\psi_2 k_n}\sum_{i=0}^{n-k_n+1}\left(\sum_{l=1}^{r}(\beta_{k,l}-\hat{\beta}_{k,l})\bar{X}_{l,i}^\star\right)^2\leq Cn^{-1/2+\delta}r^2\log d\right\},$$

Then according to Cauchy-Schwarz inequality that

$$\max_{1\leq k,l\leq d}\left|\frac{n}{n-k_n+2}\frac{1}{\psi_2 k_n}\sum_{i=0}^{n-k_n+1}\left[(\bar{Y}_i^{\star k}-(\hat{\beta}\bar{X}_i^\star)_k)(\bar{Y}_i^{\star l}-(\hat{\beta}\bar{X}_i^\star)_l)-\bar{Z}_{o,ki}^\star\bar{Z}_{o,li}^\star\right]\right|$$
$$\leq\max_{1\leq k,l\leq d}\left|\frac{n}{n-k_n+2}\frac{1}{\psi_2 k_n}\sum_{i=0}^{n-k_n+1}((\hat{\beta}-\beta)\bar{X}_i^\star)_k(\hat{\beta}-\beta)\bar{X}_i^\star)_l\right|$$
$$+2\max_{1\leq l,k\leq d}\left|\frac{n}{n-k_n+2}\frac{1}{\psi_2 k_n}\sum_{i=0}^{n-k_n+1}\bar{Z}_{o,ki}^\star((\hat{\beta}-\beta)\bar{X}_i^\star)_l\right|$$

$$
\begin{aligned}
&\leq \max_{1\leq k\leq d} \frac{n}{n-k_n+2}\frac{1}{\psi_2 k_n}\sum_{i=0}^{n-k_n+1}((\hat\beta-\beta)\bar X_i^\star)_k^2 \\
&\quad + 2\sqrt{\max_{1\leq k\leq d} \frac{n}{n-k_n+2}\frac{1}{\psi_2 k_n}\sum_{i=0}^{n-k_n+1}(\bar Z^\star_{o,ki})^2 \max_{1\leq k\leq d} \frac{n}{n-k_n+2}\frac{1}{\psi_2 k_n}\sum_{i=0}^{n-k_n+1}((\hat\beta-\beta)\bar X_i^\star)_k^2} \\
&\leq C_0 n^{-1/2+\delta} r^2 \log d + 2\sqrt{\left(\frac{5}{4}Ct\right)\left(C_0 n^{-1/2+\delta} r^2 \log d\right)} \\
&\leq C_0' n^{-1/4+\delta} s\sqrt{\log d}.
\end{aligned}
$$

Consequently, we have

$$
\begin{aligned}
&\max_{1\leq k,l\leq d}\left|\frac{n}{n-k_n+2}\frac{1}{\psi_2 k_n}\sum_{i=0}^{n-k_n+1}\left[\left(\bar Y_i^{\star k}-(\hat\beta\bar X_i^\star)_k\right)\left(\bar Y_i^{\star l}-(\hat\beta\bar X_i^\star)_l\right)-\bar Z^\star_{o,ki}\bar Z^\star_{o,li}\right]\right| \\
&\leq C_0' n^{-1/4+\delta} s\sqrt{\log d}.
\end{aligned}
$$

with probability $1-O\left(n^{3/2+\delta-3\nu/4-\nu\delta/2}\right)$ by (A.4) and (A.6). Finally, by triangle inequality, we obtain

$$
\begin{aligned}
&\max_{1\leq l,k\leq d}\left|\widehat\Gamma_{kl}-\Gamma_{kl}\right| \\
&\leq \max_{1\leq k\leq s, 1\leq l\leq d}\left|\frac{n}{n-k_n+2}\frac{1}{\psi_2 k_n}\sum_{i=0}^{n-k_n+1}\bar Z^\star_{k,i}\bar Z^\star_{l,i}\right| \\
&\quad + \max_{1\leq k,l\leq d}\left|\frac{n}{n-k_n+2}\frac{1}{\psi_2 k_n}\sum_{i=0}^{n-k_n+1}\left[\left(\bar Y_i^{\star k}-(\hat\beta\bar X_i^\star)_k\right)\left(\bar Y_i^{\star l}-(\hat\beta\bar X_i^\star)_l\right)-\bar Z^\star_{o,ki}\bar Z^\star_{o,li}\right]\right|,
\end{aligned}
$$

which leads to the result by (A.4).

Under the event of

$$
B=\left\{\max_{k,l}|\widehat\Gamma_{ij}-\Gamma|\leq Csn^{-1/4+\delta/2}\sqrt{\log d}\right\}.
$$

Firstly, we bound

$$
\left\|\widehat\Gamma_o^S-\Gamma\right\|_{\mathrm{MAX}}\leq\left\|\Gamma^S-\Gamma\right\|_{\mathrm{MAX}}+\left\|\widehat\Gamma_o^S-\Gamma^S\right\|_{\mathrm{MAX}}.
$$

The first term can be bounded by

$$
\max_{k,l}|\Gamma_{kl}|\mathbf{1}_{\{|\Gamma_{kl}|\leq S\}}\leq S.
$$

On the other hand, for the second term,

$$
\begin{aligned}
\left\|\widehat\Gamma_o^S-\Gamma^S\right\|_{\mathrm{MAX}} &\leq \max_{k,l}\left|\widehat\Gamma_{kl}\right|\mathbf{1}_{\left\{|\widehat\Gamma_{kl}|\geq S, |\Gamma_{kl}|<S\right\}} \\
&\quad+\max_{k,l}|\Gamma_{kl}|\mathbf{1}_{\left\{|\widehat\Gamma_{kl}|<S, |\Gamma_{kl}|\geq S\right\}} \\
&\quad+\max_{k,l}\left|\widehat\Gamma_{kl}-\Gamma_{kl}\right|\mathbf{1}_{\left\{|\widehat\Gamma_{kl}|\geq S, |\Gamma_{kl}|\geq S\right\}}
\end{aligned}
$$

$$\equiv T_1 + T_2 + T_3.$$

For T_1, we have

$$T_1 \leq \max_{k,l} \left|\widehat{\Gamma}_{kl} - \Gamma_{kl}\right| \mathbf{1}_{\left\{\left|\widehat{\Gamma}_{kl}\right| \geq S, \left|\Gamma_{kl}\right| < S\right\}} + \max_{k,l} \left|\Gamma_{kl}\right| \mathbf{1}_{\{|\Gamma_{kl}| < S\}}$$
$$\equiv T_4 + T_5.$$

We have already proved $T_5 \leq S$. Now take $\gamma \in (0,1)$. Then

$$\begin{aligned} T_4 \leq & \max_{k,l} \left|\widehat{\Gamma}_{kl} - \Gamma_{kl}\right| \mathbf{1}_{\left\{\left|\widehat{\Gamma}_{kl}\right| \geq S, \left|\Gamma_{kl}\right| \leq \gamma\lambda\right\}} \\ & + \max_{k,l} \left|\widehat{\Gamma}_{kl} - \Gamma_{kl}\right| \mathbf{1}_{\left\{\left|\widehat{\Gamma}_{kl}\right| \geq S, \gamma t \leq \left|\Gamma_{kl}\right| \leq S\right\}} \\ \leq & \max_{k,l} \left|\widehat{\Gamma}_{kl} - \Gamma_{kl}\right| \max_{k,l} N_{kl}(1-\gamma) + \max_{k,l} \left|\widehat{\Gamma}_{kl} - \Gamma_{kl}\right|, \end{aligned}$$

where $N_{kl}(t) \equiv I\left(\left|\widehat{\Gamma}_{kl} - \Gamma_{kl}\right| > t\lambda\right)$. Note that for according to Lemma 1, we can choose some constant such that

$$\mathbb{P}\left(\max_{k,l} N_k(1-\gamma) > 0\right) = \mathbb{P}\left(\max_{k,l} \left|\widehat{\Gamma}_{kl} - \Gamma_{kl}\right| > (1-\gamma)\lambda\right) \to 0.$$

Therefore

$$T_4 \leq Csn^{-1/4+\delta/2}\sqrt{\log d},$$

which means

$$T_1 \leq C\lambda + sn^{-1/4+\delta/2}\sqrt{\log d}.$$

To bound T_2,

$$\begin{aligned} T_2 \leq & \max_{k,l} \left|\widehat{\Gamma}_{kl} - \Gamma_{kl}\right| \mathbf{1}_{\{|\widehat{\Gamma}_{kl}| < S, |\Gamma_{kl}| \geq S)\}} \\ \leq & \max_{k,l} \left|\widehat{\Gamma}_{kl} - \Gamma_{kl}\right| \mathbf{1}_{\{|\Gamma_{kl}| \geq S\}} + \lambda \max_{k,l} \mathbf{1}_{\{|\Gamma_{kl}| \geq S\}} \\ \leq & C(\lambda + sn^{-1/4+\delta/2}\sqrt{\log d}). \end{aligned}$$

As for term T_3, we have

$$T_3 \leq \max_{kl} \left|\widehat{\Gamma}_{kl} - \Gamma_{kl}\right| I_{\{|\widehat{\Gamma}_{kl}| \geq S, |\Gamma_{kl}| \geq S\}} \leq Csn^{-1/4+\delta/2}\sqrt{\log d}.$$

Above all, we get

$$\left\|\widehat{\Gamma}_o^S - \Gamma\right\|_{\text{MAX}} \leq C(\lambda + sn^{-1/4+\delta/2}\sqrt{\log d}).$$

Here we choose $\lambda \leq Csn^{-1/4+\delta/2}\sqrt{\log d}$. Then

$$\left\|\widehat{\Gamma}_o^S - \Gamma\right\|_{\text{MAX}} \leq Csn^{-1/4+\delta/2}\sqrt{\log d},$$

with probability at least $1 - O(n^{3/2+\delta-3\nu/4-\nu\delta/2d^{-1}})$. Based on Lemma 1, and following the same proving steps as proof of theorem 1 in Fan, Furger, and Xiu (2015), we can get

$$\left\|\widehat{\Sigma}_o^S - \Sigma\right\| = O\left(n^{-1/4+\delta/2}\sqrt{\log d}\right).$$

∎

On the other hand, we have

$$\begin{aligned}\left\|\hat{\Sigma}_o^S - \Sigma\right\|_\Sigma^2 \leq& 4\left\|\beta(\widehat{\mathrm{E}} - \mathrm{E})\beta^\intercal\right\|_\Sigma^2 + 24\left\|\beta\widehat{\mathrm{E}}(\hat{\beta} - \beta)\right\|_\Sigma^2 \\ &+ 16\left\|(\hat{\beta} - \beta)\widehat{\mathrm{E}}(\hat{\beta} - \beta)^\intercal\right\|_\Sigma^2 + 2\left\|\hat{\Gamma}_o^S - \Gamma\right\|_\Sigma^2.\end{aligned} \tag{A.12}$$

Lemma 2. *Under assumptions and conditions, we have*

$$\mathbb{P}\left(\left\|\beta(\widehat{\mathrm{E}} - \mathrm{E})\beta^\intercal\right\|_\Sigma^2 + \left\|\beta\widehat{\mathrm{E}}(\hat{\beta} - \beta)^\intercal\right\|_\Sigma^2 \geq C_0 d^{-1} n^{-1/2+\delta}\log d\right) = O(C_1 d^{-C_0^2 C_2 + 1}),$$

and

$$\mathbb{P}\left(\left\|(\hat{\beta} - \beta)\widehat{\mathrm{E}}(\hat{\beta} - \beta)^\intercal\right\|_\Sigma^2 \geq C_0 d n^{-1+2\delta}\log^2 d\right) = O(C_1 d^{-C_0^2 C_2 + 1}).$$

Proof. (i) Using the same argument in proof of theorem 2 in Fan, Fan, and Lv (2008), we have

$$\left\|\beta^\intercal\Sigma^{-1}\beta\right\| \leq 2\left\|cov^{-1}(X)\right\|.$$

Therefore

$$\begin{aligned}\left\|\beta(\widehat{\mathrm{E}} - \mathrm{E})\beta^\intercal\right\|_\Sigma^2 =& d^{-1}tr\left(\Sigma^{-1/2}\beta(\widehat{\mathrm{E}} - \mathrm{E})\beta^\intercal\Sigma^{-1}\beta(\widehat{\mathrm{E}} - \mathrm{E})\beta^\intercal\Sigma^{-1/2}\right) \\ =& d^{-1}tr\left((\widehat{\mathrm{E}} - \mathrm{E})\beta^\intercal\Sigma^{-1}\beta(\widehat{\mathrm{E}} - \mathrm{E})\beta^\intercal\Sigma^{-1}\beta\right) \\ \leq& d^{-1}\left\|(\widehat{\mathrm{E}} - \mathrm{E})\beta^\intercal\Sigma^{-1}\beta\right\|_F^2 \\ \leq& O(d^{-1})\left\|\widehat{\mathrm{E}} - \mathrm{E}\right\|_F^2.\end{aligned}$$

On the other hand, we also have

$$\begin{aligned}\left\|\beta\widehat{\mathrm{E}}(\hat{\beta} - \beta)^\intercal\right\|_\Sigma^2 \leq& d^{-1}\left\|\beta^\intercal\Sigma^{-1}\beta\widehat{\mathrm{E}}(\hat{\beta} - \beta)\right\|_F\left\|\widehat{\mathrm{E}}\Sigma^{-1}(\hat{\beta} - \beta)^\intercal\right\|_F \\ \leq& d^{-1}\left\|\beta^\intercal\Sigma^{-1}\beta\right\|_F\left\|\widehat{\mathrm{E}}\right\|_F^2\left\|\hat{\beta} - \beta\right\|_F^2.\end{aligned}$$

Then by Lemma 1 (A.2) and (A.6), and $\mathbb{P}(\|\widehat{\mathrm{E}}\|_F^2 > C) = O(C_1 r^2 d^{-C_0^2 C_2})$. We can get the final results.

(ii) Moreover, we have

$$\left\|(\hat{\beta} - \beta)\widehat{\mathrm{E}}(\hat{\beta} - \beta)^\intercal\right\|_\Sigma^2 = d^{-1}tr\left((\hat{\beta} - \beta)\widehat{\mathrm{E}}(\hat{\beta} - \beta)^\intercal\Sigma^{-1}(\hat{\beta} - \beta)\widehat{\mathrm{E}}(\hat{\beta} - \beta)^\intercal\Sigma^{-1}\right)$$

$$
\begin{aligned}
&\leq d^{-1}\left\|(\hat{\beta}-\beta)\widehat{\mathrm{E}}(\hat{\beta}-\beta)^{\intercal}\Sigma^{-1}\right\|_F^2\\
&\leq d^{-1}\lambda_{\mathrm{MAX}}^2(\Sigma^{-1})\lambda_{\mathrm{MAX}}^2(\widehat{\mathrm{E}})\left\|\hat{\beta}-\beta\right\|_F^4.
\end{aligned}
$$

Since $\lambda_{\mathrm{MAX}}^2(\Sigma^{-1})$ and $\lambda_{\mathrm{MAX}}^2(\widehat{\mathrm{E}})$ are both bounded. Then the result follows from (A.6). ∎

Finally, we have

$$
\begin{aligned}
\left\|\hat{\Gamma}_o^S-\Gamma\right\|_\Sigma =& d^{-1/2}\left\|\Sigma^{-1/2}(\hat{\Gamma}_o^S-\Gamma)\Sigma^{-1/2}\right\|_F \\
\leq& \left\|\Sigma^{-1/2}(\hat{\Gamma}_o^S-\Gamma)\Sigma^{-1/2}\right\|\\
\leq& \left\|\hat{\Gamma}_o^S-\Gamma\right\|\lambda_{\mathrm{MAX}}(\Sigma^{-1}).
\end{aligned}
\tag{A.13}
$$

Then based on (A.12), Lemma 2 and Lemma 3 (B.15), and the fact that

$$
\begin{aligned}
&d^{-1}n^{-1/2+\delta}\log d+dn^{-1+2\delta}\log^2 d+m_d^2n^{-(1/2-\delta)(1-q)}(\log d)^{(1-q)}\\
=&O\left(dn^{-1+2\delta}\log^2 d+m_d^2n^{-(1/2-\delta)(1-q)}(\log d)^{(1-q)}\right).
\end{aligned}
$$

On the other hand, if we do not assume the factor structure, using a direct pre-averaging estimator $\hat{\Sigma}^\star$. Then we will get

$$
\left\|\hat{\Sigma}^\star-\Sigma\right\|_\Sigma^2\leq C\left\|\beta(\widehat{\mathrm{E}}-\mathrm{E})\beta^\intercal\right\|_\Sigma^2+C\left\|\beta\bar{X}\bar{Z}^\intercal\right\|_\Sigma^2+C\left\|\bar{Z}\bar{Z}^\intercal-\Gamma\right\|_\Sigma^2. \tag{A.14}
$$

According to the proof of Lemma 2, we know $\left\|\beta(\widehat{\mathrm{E}}-\mathrm{E})\beta^\intercal\right\|_\Sigma^2=O_p(d^{-1}n^{-1/2+\delta}\log d)$; it is also easy to get that $\left\|\beta\bar{X}\bar{Z}^\intercal\right\|_\Sigma^2=O_p(d^{-1}n^{-1/2+\delta}\log d)$. We can also get that $\left\|\bar{Z}\bar{Z}^\intercal-\Gamma\right\|_\Sigma^2=O_p(dn^{-1/2+\delta}\log d)$.

Appendix B Proof of Theorem 2

Lemma 3. *Suppose that the entry-wise max norms of all the processes are bounded uniformly in* $[0,t]$. *Under Assumptions 1 - 7, we have*

$$
\mathbb{P}\left(\left\|\widehat{\Gamma}_o^S-\Gamma\right\|>C_0m_dn^{-(1/4-\delta/2)(1-q)}(\log d)^{(1-q)/2}\right)=O(n^{3/2+\delta-3\nu/4-\nu\delta/2}), \tag{B.15}
$$

$$
\mathbb{P}\left(\lambda_{\min}\left(\widehat{\Gamma}_o^S\right)\geq\frac{1}{2}\lambda_{\min}\left(\Gamma\right)\right)>1-O(n^{3/2+\delta-3\nu/4-\nu\delta/2}), \tag{B.16}
$$

$$
\mathbb{P}\left(\left\|(\widehat{\Gamma}_o^S)^{-1}-\Gamma^{-1}\right\|>C_0m_dn^{-(1/4-\delta/2)(1-q)}(\log d)^{(1-q)/2}\right)=O(n^{3/2+\delta-3\nu/4-\nu\delta/2}), \tag{B.17}
$$

$$
\begin{aligned}
&\mathbb{P}\left(\left\|\widehat{\beta}^\intercal(\widehat{\Gamma}_o^S)^{-1}\widehat{\beta}-\beta^\intercal\Gamma^{-1}\beta\right\|>C_0m_dn^{-(1/4-\delta/2)(1-q)}\cdot d(\log d)^{(1-q)/2}\right)\\
&=O(n^{3/2+\delta-3\nu/4-\nu\delta/2}),
\end{aligned}
\tag{B.18}
$$

$$
\begin{aligned}
&\mathbb{P}\left(\left\|\left(\widehat{\mathrm{E}}^{-1}+\widehat{\beta}^\intercal(\widehat{\Gamma}_o^S)^{-1}\widehat{\beta}\right)-\left(\mathrm{E}^{-1}+\beta^\intercal(\Gamma^S)^{-1}\beta\right)\right\|>C_0m_dn^{-(1/4-\delta/2)(1-q)}d(\log d)^{(1-q)/2}\right)\\
&=O(n^{3/2+\delta-3\nu/4-\nu\delta/2}),
\end{aligned}
\tag{B.19}
$$

$$\mathbb{P}\left(\left\|\left(\widehat{\mathrm{E}}^{-1}+\widehat{\beta}^{\intercal}(\widehat{\Gamma}_o^S)^{-1}\widehat{\beta}\right)^{-1}\right\|>C_0 m_d d^{-1}\right)=O(n^{3/2+\delta-3\nu/4-\nu\delta/2}), \tag{B.20}$$

$$\mathbb{P}\left(\left\|\widehat{\beta}\left(\widehat{\mathrm{E}}^{-1}+\widehat{\beta}^{\intercal}(\widehat{\Gamma}_o^S)^{-1}\widehat{\beta}\right)^{-1}\widehat{\beta}^{\intercal}(\widehat{\Gamma}_o^S)^{-1}\right\|>C_0 m_d\right)=O(n^{3/2+\delta-3\nu/4-\nu\delta/2}), \tag{B.21}$$

$$\mathbb{P}\left(\left\|\widehat{\beta}\left(\widehat{\mathrm{E}}^{-1}+\widehat{\beta}^{\intercal}(\widehat{\Gamma}_o^S)^{-1}\widehat{\beta}\right)^{-1}\widehat{\beta}^{\intercal}\Gamma^{-1}\right\|>C_0 m_d\right)=O(n^{3/2+\delta-3\nu/4-\nu\delta/2}). \tag{B.22}$$

Proof. (i) Because $\widehat{\Gamma}^S-\Gamma$ is symmetric, its operator norm is bounded by the ∞-norm:

$$\left\|\widehat{\Gamma}_o^S-\Gamma\right\|\le\max_{1\le l\le d}\sum_{k=1}^{d}\left|\widehat{\Gamma}_{lk}^S-\Gamma_{lk}\right|.$$

Then using the same technique for proving (A.9), we can prove that , with probability no less than $O(n^{3/2+\delta-3\nu/4-\nu\delta/2})$, we have

$$\left\|\widehat{\Gamma}_o^S-\Gamma\right\|\le C_0 m_d n^{-(1/4-\delta/2)(1-q)}(\log d)^{(1-q)/2}.$$

Proof of (B.16)-(B.22) is the same as Lemma 5 in Fan, Furger, and Xiu (2015), therefore we omit the details. ∎

By the localization argument, we only need to prove the result under a stronger assumption that the entry-wise norms of all the processes are bounded uniformly in $[0,t]$. By the Sherman - Morrison -Woodbury formula, we have

$$\begin{aligned}
&\left\|(\widehat{\Sigma}_o^S)^{-1}-\Sigma^{-1}\right\|\\
\le&\left\|(\widehat{\Gamma}_o^S)^{-1}-\Gamma^{-1}\right\|+\left\|\left((\widehat{\Gamma}_o^S)^{-1}-\Gamma^{-1}\right)\widehat{\beta}\left(\widehat{\mathrm{E}}^{-1}+\widehat{\beta}^{\intercal}(\widehat{\Gamma}_o^S)^{-1}\widehat{\beta}\right)^{-1}\widehat{\beta}^{\intercal}(\widehat{\Gamma}_o^S)^{-1}\right\|\\
&+\left\|\left((\widehat{\Gamma}_o^S)^{-1}-\Gamma^{-1}\right)\widehat{\beta}\left(\widehat{\mathrm{E}}^{-1}+\widehat{\beta}^{\intercal}(\widehat{\Gamma}_o^S)^{-1}\widehat{\beta}\right)^{-1}\widehat{\beta}^{\intercal}\Gamma^{-1}\right\|\\
&+\left\|\Gamma^{-1}(\widehat{\beta}-\beta)\left(\widehat{\mathrm{E}}^{-1}+\widehat{\beta}^{\intercal}(\widehat{\Gamma}_o^S)^{-1}\widehat{\beta}\right)^{-1}\widehat{\beta}^{\intercal}\Gamma^{-1}\right\|\\
&+\left\|\Gamma^{-1}(\widehat{\beta}-\beta)\left(\widehat{\mathrm{E}}^{-1}+\widehat{\beta}^{\intercal}(\widehat{\Gamma}_o^S)^{-1}\widehat{\beta}\right)^{-1}\beta^{\intercal}\Gamma^{-1}\right\|\\
&+\left\|\Gamma^{-1}\beta\left(\left(\widehat{\mathrm{E}}^{-1}+\widehat{\beta}^{\intercal}(\widehat{\Gamma}_o^S)^{-1}\widehat{\beta}\right)^{-1}-\left(\mathrm{E}^{-1}+\beta^{\intercal}(\Gamma^S)^{-1}\beta\right)^{-1}\right)\beta^{\intercal}\Gamma^{-1}\right\|\\
:=&L_1+L_2+L_3+L_4+L_5+L_6.
\end{aligned}$$

We now bound each term above with probability no less than $1-O(n^{3/2+\delta-3\nu/4-\nu\delta/2})$. First of all, by (B.15)

$$L_1\le Cm_d n^{-(1/4-\delta/2)(1-q)}(\log d)^{(1-q)/2}.$$

To bound L_2, we have, by (B.17) and (B.21),

$$L_2\le\left\|(\widehat{\Gamma}_o^S)^{-1}-\Gamma^{-1}\right\|\cdot\left\|\widehat{\beta}\left(\widehat{\mathrm{E}}^{-1}+\widehat{\beta}^{\intercal}(\widehat{\Gamma}_o^S)^{-1}\widehat{\beta}\right)^{-1}\widehat{\beta}^{\intercal}(\widehat{\Gamma}_o^S)^{-1}\right\|$$

$$\leq Cm_d^2 n^{-(1/4-\delta/2)(1-q)}(\log d)^{(1-q)/2}.$$

Similarly, L_3 can be bounded using (B.17) and (B.22).

Next, for L_4, we use (A.6), and (B.20), $\|\cdot\| \leq \|\cdot\|_{\mathrm{F}}$, and $\lambda_{\min}(\Gamma) \geq C'$,

$$\begin{aligned} L_4 \leq & \left\|\Gamma^{-1}\right\|^2 \cdot \left\|\widehat{\beta} - \beta\right\| \cdot \left\|\widehat{\beta}\right\| \cdot \left\|\left(\widehat{\mathrm{E}}^{-1} + \widehat{\beta}^{\intercal}(\widehat{\Gamma}_o^S)^{-1}\widehat{\beta}\right)^{-1}\right\| \\ \leq & Cm_d^2 n^{-(1/4-\delta/2)(1-q)}(\log d)^{(1-q)/2}. \end{aligned}$$

Similarly, using the fact that $\|\beta\| \leq \|\beta\|_{\mathrm{F}} = O(\sqrt{d})$, we can establish the same bound for L_5.

Finally, we have

$$\begin{aligned} L_6 \leq & \left\|\Gamma^{-1}\right\|^2 \cdot \|\beta\|^2 \cdot \left\|\left(\widehat{\mathrm{E}}^{-1} + \widehat{\beta}^{\intercal}(\widehat{\Gamma}_o^S)^{-1}\widehat{\beta}\right)^{-1} - \left(\mathrm{E}^{-1} + \beta^{\intercal}(\Gamma^S)^{-1}\beta\right)^{-1}\right\| \\ \leq & \left\|\Gamma^{-1}\right\|^2 \cdot \|\beta\|^2 \cdot \left\|\left(\widehat{\mathrm{E}}^{-1} + \widehat{\beta}^{\intercal}(\widehat{\Gamma}_o^S)^{-1}\widehat{\beta}\right) - \left(\mathrm{E}^{-1} + \beta^{\intercal}(\Gamma^S)^{-1}\beta\right)\right\| \\ & \cdot \left\|\left(\widehat{\mathrm{E}}^{-1} + \widehat{\beta}^{\intercal}(\widehat{\Gamma}_o^S)^{-1}\widehat{\beta}\right)^{-1}\right\| \cdot \left\|\left(\mathrm{E}^{-1} + \beta^{\intercal}\Gamma^{-1}\beta\right)^{-1}\right\|. \end{aligned}$$

Note that for any vector v such that $\|v\| = 1$, by the definition of operator norm, we have

$$v^{\intercal}\beta^{\intercal}\Gamma^{-1}\beta v \geq \lambda_{\min}(\Gamma^{-1})v^{\intercal}\beta^{\intercal}\beta v \geq \lambda_{\min}(\Gamma^{-1})\lambda_{\min}(\beta^{\intercal}\beta).$$

It then follows that

$$\lambda_{\min}(\beta^{\intercal}\Gamma^{-1}\beta) \geq \lambda_{\min}(\Gamma^{-1})\lambda_{\min}(\beta^{\intercal}\beta).$$

On the other hand, by Assumption 3, we have

$$d^{-1}v^{\intercal}\beta^{\intercal}\beta v = v^{\intercal}Bv - v^{\intercal}(B - d^{-1}\beta^{\intercal}\beta)v \geq \lambda_{\min}(B) - \left\|d^{-1}\beta^{\intercal}\beta - B\right\| > C,$$

where C is some constant. Thus, $\lambda_{\min}(\beta^{\intercal}\beta) > Cd$. Therefore $\lambda_{\min}(\beta^{\intercal}\Gamma^{-1}\beta) > C'd$, following from the fact that $\lambda_{\max}(\Gamma) \leq Km_d$. It then implies that $\lambda_{\min}(\mathrm{E}^{-1} + \beta^{\intercal}\Gamma^{-1}\beta) \geq \lambda_{\min}(\beta^{\intercal}\Gamma^{-1}\beta) > C'm_d^{-1}d$.

$$\left\|\left(\mathrm{E}^{-1} + \beta^{\intercal}\Gamma^{-1}\beta\right)^{-1}\right\| = O_p(m_d d^{-1}).$$

Using (B.19) and (B.20) , we have

$$L_6 \leq Cm_d^3 n^{-(1/4-\delta/2)(1-q)}(\log d)^{(1-q)/2}.$$

Finally, combining these estimates, we can obtain, for some $C > 0$,

$$\begin{aligned} \left\|(\widehat{\Sigma}_o^S)^{-1} - \Sigma^{-1}\right\| \leq & C\left(m_d^3 n^{-(1/4-\delta/2)(1-q)}(\log d)^{(1-q)/2}\right. \\ & \left. + m_d^2 n^{-(1/4-\delta/2)(1-q)}(\log d)^{(1-q)/2}\right), \end{aligned}$$

where the second term on the right is dominated by the first one, which yields the desired result.

To prove the second statement, note that for any vector v such that $\|v\| = 1$, we have

$$v^\intercal \widehat{\Sigma}^S v = v^\intercal \widehat{\beta} \widehat{\mathrm{E}} \widehat{\beta}^\intercal v + v^\intercal \widehat{\Gamma}^S v \geq \lambda_{\min}\left(\widehat{\Gamma}^S\right),$$

which implies that

$$\lambda_{\min}\left(\widehat{\Sigma}^S\right) \geq \lambda_{\min}\left(\widehat{\Gamma}^S\right).$$

This inequality, combining with (ii) of Lemma 3, concludes the proof.

Appendix B.1 Proof of Theorem 4

Proof.

Lemma 4. *Suppose Assumption 1 and 2 hold, then we have*

$$\max_{1\leq k\leq s, 1\leq l\leq d} \left| \frac{n}{n-k_n+2} \frac{1}{\psi_2 k_n} \sum_{i=0}^{n-k_n+1} \bar{X}_i^k \bar{X}_i^l - \int_0^t e_{s,kl} ds \right| = O_p(n^{-1/4+\delta/2}\sqrt{\log d}), \tag{B.23}$$

$$\max_{1\leq k\leq s, 1\leq l\leq d} \left| \frac{n}{n-k_n+2} \frac{1}{\psi_2 k_n} \sum_{i=0}^{n-k_n+1} \bar{X}_i^k \bar{Z}_{ui}^{\star l} \right| = O_p(n^{-1/4+\delta/2}\sqrt{\log d}), \tag{B.24}$$

$$\max_{1\leq k\leq s, 1\leq l\leq d} \left| \frac{n}{n-k_n+2} \frac{1}{\psi_2 k_n} \sum_{i=0}^{n-k_n+1} \bar{Z}_{ui}^{\star k} \bar{Z}_{ui}^{\star l} - \int_0^t g_{s,kl} ds \right| = O_p(n^{-1/4+\delta/2}\sqrt{\log d}). \tag{B.25}$$

These results can be derived directly from Lemma 1.

∎

Lemma 5. *Suppose Assumptions 1 - 7, and 4 hold with $S = O_p(n^{-1/4+\delta/2}\sqrt{\log d})$. Suppose $d^{-1/2} m_d = o(1)$, $n^{-1/4+\delta/2}\sqrt{\log d} = o(1)$, and $\widehat{r} \to r$ with probability approaching 1, then there exists a $r \times r$ matrix H, such that with probability approaching 1, H is invertible, $\|HH^\intercal - \mathbb{I}_r\| = \|H^\intercal H - \mathbb{I}_r\| = o_p(1)$, and more importantly,*

$$\|F - \beta H\|_{\mathrm{MAX}} = O_p\left(n^{-1/4+\delta/2}\sqrt{\log d} + d^{-1/2} m_d\right),$$
$$\left\|G - H^{-1}\bar{\mathcal{X}}\right\| = O_p\left(n^{-1/4+\delta/2}\sqrt{\log d} + d^{-1/2} m_d\right).$$

This lemma is very important during the proof. First, we can assume $\widehat{r} = r$. Since it holds with probability approaching 1 as established by Lemma 3, a simple conditioning argument, see, e.g., footnote 5 of, is sufficient to show this is without loss of rigor. Recall that

$$\Lambda = \mathrm{Diag}\left(\widehat{\lambda}_1, \widehat{\lambda}_2, \ldots, \widehat{\lambda}_r\right), \quad F = d^{1/2}\left(\widehat{\xi}_1, \widehat{\xi}_2, \ldots, \widehat{\xi}_r\right), \quad \text{and} \quad G = d^{-1} F^\intercal \mathcal{Y}.$$

We write

$$H = t^{-1} \frac{n}{n-k_n+2} \frac{1}{\psi_2 k_n} \bar{\mathcal{X}}^{\star} \bar{\mathcal{X}}^{\star\intercal} \beta^{\intercal} F \Lambda^{-1}.$$

It is easy to verify that

$$\widehat{\Sigma} F = F\Lambda, \quad GG^{\intercal} = td^{-1} \times \Lambda, \quad F^{\intercal}F = d \times \mathbb{I}_r, \quad \text{and}$$
$$\widehat{\Gamma} = t^{-1} \left(\mathcal{Y} - FG\right) \left(\mathcal{Y} - FG\right)^{\intercal} = t^{-1} \mathcal{Y}\mathcal{Y}^{\intercal} - d^{-1} F\Lambda F^{\intercal}.$$

We now need a few more lemmas.

Lemma 6. *Under Assumptions 1 - 7, and 4, $d^{-1/2} m_d = o(1)$, and $n^{-1/2+\delta} \log d = o(1)$, we have*

$$\|F - \beta H\|_{\mathrm{MAX}} = O_p\left(n^{-1/4+\delta/2} \sqrt{\log d} + d^{-1/2} m_d\right). \tag{B.26}$$

$$\left\|H^{-1}\right\| = O_p(1). \tag{B.27}$$

$$\left\|G - H^{-1} \bar{\mathcal{X}}\right\| = O_p\left(n^{-1/4+\delta/2} \sqrt{\log d} + d^{-1/2} m_d\right). \tag{B.28}$$

Lemma 7. *Under Assumptions 1 - 7, and 4, $d^{-1/2} m_d = o(1)$, and $n^{-1/2+\delta} \log d = o(1)$, we have*

$$\left\|\widehat{\Gamma}_u^S - \Gamma\right\|_{\mathrm{MAX}} \leq \left\|\widehat{\Gamma} - \Gamma\right\|_{\mathrm{MAX}} = O_p\left(n^{-1/4+\delta/2} \sqrt{\log d} + d^{-1/2} m_d\right). \tag{B.29}$$

Lemma 8. *Under Assumptions 1 - 7, and 4, $d^{-1/2} m_d = o(1)$, and $n^{-1/2+\delta} \log d = o(1)$, we have*

$$\left\|t^{-1} FGG^{\intercal}F^{\intercal} - \beta \mathrm{E} \beta^{\intercal}\right\|_{\mathrm{MAX}} = O_p\left(n^{-1/4+\delta/2} \sqrt{\log d} + d^{-1/2} m_d\right).$$

Proof of Lemma 6 - 8 follows directly from Aït-Sahalia and Xiu (2015).

Proof. Note that

$$\widehat{\Sigma}_u^S = d^{-1} F\Lambda F^{\intercal} + \widehat{\Gamma}_u^S = t^{-1} FGG^{\intercal}F^{\intercal} + \widehat{\Gamma}_u^S.$$

By Lemma 7, we have

$$\left\|\widehat{\Gamma}_u^S - \Gamma\right\|_{\mathrm{MAX}} = O_p\left(n^{-1/4+\delta/2} \sqrt{\log d} + d^{-1/2} m_d\right).$$

By the triangle inequality, we have

$$\left\|\widehat{\Sigma}_u^S - \Sigma\right\|_{\mathrm{MAX}} \leq \left\|d^{-1} F\Lambda F^{\intercal} - \beta \mathrm{E} \beta^{\intercal}\right\|_{\mathrm{MAX}} + \left\|\widehat{\Gamma}_u^S - \Gamma\right\|_{\mathrm{MAX}}$$

Therefore, the desired result follows from Lemmas 7 and 8. ∎

Lemma 9. *Under Assumptions 1 - 7, and 4, $d^{-1/2} m_d = o(1)$, and $n^{-1/2+\delta} \log d = o(1)$, we have*

$$\|F - \beta H\|_F^2 = O_p\left(dn^{-1/2+\delta} \log d + m_d^2\right). \tag{B.30}$$

$$\|(F - \beta H)(F - \beta H)^{\intercal}\|_F^2 = O_p\left(dn^{-1+2\delta} \log^2 d + d^{-1} m_d^4\right). \tag{B.31}$$

$$\|\beta H (F - \beta H)^{\intercal}\|_\Sigma^2 = O_p\left(n^{-1/2+\delta} \log d + d^{-1} m_d^2\right). \tag{B.32}$$

$$\|\beta(H^{\intercal}H - I_r)\beta^{\intercal}\|_\Sigma^2 = o_p(1). \tag{B.33}$$

Proof. (i) We have $\|F-\beta H\|_F^2 \leq \|F-\beta H\|_{\mathrm{MAX}}^2\, d$.
(ii) According to the definition of $\|\cdot\|_\Sigma$, we have

$$\begin{aligned}&\|(F-\beta H)(F-\beta H)^\intercal\|_\Sigma^2\\=&O_p(d^{-1}\|F-\beta H\|_F^4)\\=&O_p(d\,\|F-\beta H\|_{\mathrm{MAX}}^4).\end{aligned}$$

(iii) By $\left\|\beta^\intercal\Sigma^{-1}\beta\right\| = O(1)$, We have

$$\begin{aligned}&\|\beta H(F-\beta H)^\intercal\|_\Sigma^2\\=&d^{-1}tr(H(F-\beta H)^\intercal\Sigma^{-1}(F-\beta H)H\beta^\intercal\Sigma^{-1}\beta)\\\leq&d^{-1}\,\|H\|^2\left\|\beta^\intercal\Sigma^{-1}\beta\right\|\left\|\Sigma^{-1}\right\|\|F-\beta H\|_F^2\\=&O_p(\|F-\beta H\|_{\mathrm{MAX}}^2).\end{aligned}$$

(iv) by $\left\|\beta^\intercal\Sigma^{-1}\beta\right\| = O(1)$, We have

$$\begin{aligned}&\|\beta(H^\intercal H-I_r)\beta^\intercal\|_\Sigma^2\\\leq&d^{-1}tr\{(H^\intercal H-I_r)\beta^\intercal\Sigma^{-1}\beta(H^\intercal H-I_r)\beta^\intercal\Sigma^{-1}\beta\}\\\leq&d^{-1}\left\|\beta^\intercal\Sigma^{-1}\beta\right\|^2\|H^\intercal H-I_r\|_F^2\\=&o_p(1).\end{aligned}$$

∎

Thus, for some constant C, we have

$$\begin{aligned}\|\widehat{\Sigma}_o^S-\Sigma\|_\Sigma^2\leq&C\Big[\|\beta(H^\intercal H-I_r)\beta^\intercal\|_\Sigma^2+\|\beta H(F-\beta H)^\intercal\|_\Sigma^2\\&+\|(F-\beta H)(F-\beta H)^\intercal\|_\Sigma^2+\left\|\widehat{\Gamma}_u^S-\Gamma\right\|\Big]\\=&O_p(dn^{-1+2\delta}\log^2 d+d^{-1}m_d^4\\&+m_d^2(n^{-1/4+\delta/2}\sqrt{\log d}+d^{-1/2}m_d)^{2(1-q)}).\end{aligned}$$

Appendix B.2 Proof of Theorem 5

Lemma 10. *Under Assumptions 1 - 7, and 4, $d^{-1/2}m_d = o(1)$, and $n^{-1/4+\delta/2}\sqrt{\log d} = o(1)$, we have*

$$\left\|\widehat{\Gamma}_u^S-\Gamma\right\| = O_p\left(m_d(n^{-1/4+\delta/2}\sqrt{\log d}+d^{-1/2}m_d)^{1-q}\right). \tag{B.34}$$

Moreover, if in addition, $d^{-1/2}m_d^2 = o(1)$ and $m_d n^{-1/4+\delta/2}\sqrt{\log d} = o(1)$ hold, then $\lambda_{\min}\left(\widehat{\Gamma}^S\right)$ is bounded away from 0 with probability approaching 1, and

$$\left\|\left(\widehat{\Gamma}_u^S\right)^{-1}-\Gamma^{-1}\right\| = O_p\left(m_d(n^{-1/4+\delta/2}\sqrt{\log d}+d^{-1/2}m_d)^{1-q}\right). \tag{B.35}$$

Proof of Lemma 10. Note that since $\widehat{\Gamma}^S - \Gamma$ is symmetric,

$$\left\|\widehat{\Gamma}_u^S - \Gamma\right\| \le \left\|\widehat{\Gamma}_u^S - \Gamma\right\|_\infty = \max_{1\le l\le d}\sum_{k=1}^d \left|\widehat{\Gamma}_{lk}^S - \Gamma_{lk}\right|.$$

By Lemma 7, and using the same technique as proving (A.9), we have

$$\left\|\widehat{\Gamma}_u^S - \Gamma\right\| = O_p\left(m_d S^{-q}(n^{-1/4+\delta/2}\sqrt{\log d} + d^{-1/2}m_d) + m_d S^{1-q}\right).$$

Choosing $S = M'(n^{-1/4+\delta/2}\sqrt{\log d} + d^{-1/2}m_d)$, we have

$$\left\|\widehat{\Gamma}_u^S - \Gamma\right\| = O_p\left(m_d(n^{-1/4+\delta/2}\sqrt{\log d} + d^{-1/2}m_d)^{1-q}\right).$$

Moreover, since $\lambda_{\min}(\Gamma) > K$ for some constant K and by Weyl's inequality, we have $\lambda_{\min}(\widehat{\Gamma}_u^S) > K - o_p(1)$. As a result, we have

$$\begin{aligned}\left\|\left(\widehat{\Gamma}_u^S\right)^{-1} - \Gamma^{-1}\right\| =& \left\|\left(\widehat{\Gamma}_u^S\right)^{-1}\left(\Gamma - \left(\widehat{\Gamma}_u^S\right)\right)\Gamma^{-1}\right\| \le \lambda_{\min}(\widehat{\Gamma}_u^S)^{-1}\lambda_{\min}(\Gamma)^{-1}\left\|\Gamma - \widehat{\Gamma}_u^S\right\| \\ \le& O_p\left(m_d(n^{-1/4+\delta/2}\sqrt{\log d} + d^{-1/2}m_d)^{1-q}\right).\end{aligned}$$

∎

Proof of Theorem 5. First, by Lemma 10 and the fact that $\lambda_{\min}(\widehat{\Sigma}_u^S) \ge \lambda_{\min}(\widehat{\Gamma}_u^S)$, we can establish the first two statements.

To bound $\left\|(\widehat{\Sigma}_u^S)^{-1} - \Sigma^{-1}\right\|$, by the Sherman - Morrison - Woodbury formula, we have

$$\begin{aligned}&\left(\widehat{\Sigma}_u^S\right)^{-1} - \left(\widetilde{\Sigma}\right)^{-1} \\ =&\left(t^{-1}FGG^\intercal F^\intercal + \widehat{\Gamma}_u^S\right)^{-1} - \left(t^{-1}\beta HH^{-1}\bar{\mathcal{X}}^\star\bar{\mathcal{X}}^{\star\intercal}(H^{-1})^\intercal H^\intercal\beta^\intercal + \Gamma\right)^{-1} \\ =&\left((\widehat{\Gamma}_u^S)^{-1} - \Gamma^{-1}\right) - \left((\widehat{\Gamma}_u^S)^{-1} - \Gamma^{-1}\right)F\left(d\Lambda^{-1} + F^\intercal(\widehat{\Gamma}_u^S)^{-1}F\right)^{-1}F^\intercal(\widehat{\Gamma}_u^S)^{-1} \\ &- \Gamma^{-1}F\left(d\Lambda^{-1} + F^\intercal(\widehat{\Gamma}_u^S)^{-1}F\right)^{-1}F^\intercal\left((\widehat{\Gamma}_u^S)^{-1} - \Gamma^{-1}\right) \\ &+ \Gamma^{-1}(\beta H - F)\left(tH^\intercal\left(\bar{\mathcal{X}}\bar{\mathcal{X}}^\intercal\right)^{-1}H + H^\intercal\beta^\intercal\Gamma^{-1}\beta H\right)^{-1}H^\intercal\beta^\intercal\Gamma^{-1} \\ &- \Gamma^{-1}F\left(tH^\intercal\left(\bar{\mathcal{X}}\bar{\mathcal{X}}^\intercal\right)^{-1}H + H^\intercal\beta^\intercal\Gamma^{-1}\beta H\right)^{-1}(F^\intercal - H^\intercal\beta^\intercal)\Gamma^{-1} \\ &+ \Gamma^{-1}F\left(\left(tH^\intercal\left(\bar{\mathcal{X}}\bar{\mathcal{X}}^\intercal\right)^{-1}H + H^\intercal\beta^\intercal\Gamma^{-1}\beta H\right)^{-1} - \left(d\Lambda^{-1} + F^\intercal(\widehat{\Gamma}_u^S)^{-1}F\right)^{-1}\right)F^\intercal\Gamma^{-1} \\ =&L_1 + L_2 + L_3 + L_4 + L_5 + L_6.\end{aligned}$$

By Lemma 10, we have

$$\|L_1\| = O_p\left(m_d(n^{-1/4+\delta/2}\sqrt{\log d} + d^{-1/2}m_d)^{1-q}\right).$$

For L_2, because $\|F\| = O_p(d^{1/2})$, $\lambda_{\max}\left((\widehat{\Gamma}_u^S)^{-1}\right) \leq \left(\lambda_{\min}(\widehat{\Gamma}_u^S)\right)^{-1} \leq K + o_p(1)$,

$$\begin{aligned}\lambda_{\min}\left(d\Lambda^{-1} + F^\intercal(\widehat{\Gamma}_u^S)^{-1}F\right) &\geq \lambda_{\min}\left(F^\intercal(\widehat{\Gamma}_u^S)^{-1}F\right)\\ &\geq \lambda_{\min}\left(F^\intercal F\right)\lambda_{\min}\left((\widehat{\Gamma}_u^S)^{-1}\right)\\ &\geq m_d^{-1}d,\end{aligned}$$

and by Lemma 10, we have

$$\begin{aligned}\|L_2\| &\leq \left\|\left((\widehat{\Gamma}_u^S)^{-1} - \Gamma^{-1}\right)\right\| \|F\| \left\|\left(d\Lambda^{-1} + F^\intercal(\widehat{\Gamma}_u^S)^{-1}F\right)^{-1}\right\| \left\|F^\intercal(\widehat{\Gamma}_u^S)^{-1}\right\|\\ &= O_p\left(m_d(n^{-1/4+\delta/2}\sqrt{\log d} + d^{-1/2}m_d)^{1-q}\right).\end{aligned}$$

The same bound holds for $\|L_3\|$. As for L_4, note that $\|\beta\| = O_p(d^{1/2})$, $\|H\| = O_p(1)$, $\left\|\Gamma^{-1}\right\| \leq (\lambda_{\min}(\Gamma))^{-1} \leq K$, and $\|\beta H - F\| \leq \sqrt{rd}\,\|\beta H - F\|_{\text{MAX}} = O_p(n^{-1/4+\delta/2}d^{2/\nu+1/2} + m_d)$, and that

$$\begin{aligned}\lambda_{\min}\left(tH^\intercal\left(\bar{\mathcal{X}}\bar{\mathcal{X}}^\intercal\right)^{-1}H + H^\intercal\beta^\intercal\Gamma^{-1}\beta H\right) &\geq \lambda_{\min}\left(H^\intercal\beta^\intercal\Gamma^{-1}\beta H\right)\\ &\geq \lambda_{\min}(\Gamma^{-1})\lambda_{\min}(\beta^\intercal\beta)\lambda_{\min}(H^\intercal H)\\ &> Km_d^{-1}d,\end{aligned}$$

hence we have

$$\begin{aligned}\|L_4\| &\leq \left\|\Gamma^{-1}\right\| \|(\beta H - F)\| \left\|\left(tH^\intercal\left(\bar{\mathcal{X}}^\star\bar{\mathcal{X}}^{\star\intercal}\right)^{-1}H + H^\intercal\beta^\intercal\Gamma^{-1}\beta H\right)^{-1}\right\| \|H^\intercal\beta^\intercal\| \left\|\Gamma^{-1}\right\|\\ &= O_p(m_d n^{-1/4+\delta/2}\sqrt{\log d} + d^{-1/2}m_d^2).\end{aligned}$$

The same bound holds for L_5. Finally, with respect to L_6, we have

$$\begin{aligned}&\left\|\left(tH^\intercal\left(\bar{\mathcal{X}}\bar{\mathcal{X}}^\intercal\right)^{-1}H + H^\intercal\beta^\intercal\Gamma^{-1}\beta H\right)^{-1} - \left(d\Lambda^{-1} + F^\intercal(\widehat{\Gamma}_u^S)^{-1}F\right)^{-1}\right\|\\ \leq &Kd^{-2}m_d^2\left\|\left(tH^\intercal\left(\bar{\mathcal{X}}\bar{\mathcal{X}}^\intercal\right)^{-1}H + H^\intercal\beta^\intercal\Gamma^{-1}\beta H\right) - \left(d\Lambda^{-1} + F^\intercal(\widehat{\Gamma}_u^S)^{-1}F\right)\right\|.\end{aligned}$$

Moreover, since we have

$$\left\|tH^\intercal(\bar{\mathcal{X}}\bar{\mathcal{X}}^\intercal)^{-1}H - d\Lambda^{-1}\right\| = \left\|\Lambda^{-1}F^\intercal(\beta H - F)\right\| = O_p\left(n^{-1/4+\delta/2}\sqrt{\log d} + d^{-1/2}m_d\right)$$

and

$$\begin{aligned}&\left\|H^\intercal\beta^\intercal\Gamma^{-1}\beta H - F^\intercal(\widehat{\Gamma}_u^S)^{-1}F\right\|\\ \leq &\left\|(H^\intercal\beta^\intercal - F^\intercal)\Gamma^{-1}\beta H\right\| + \left\|F^\intercal\Gamma^{-1}(\beta H - F)\right\| + \left\|F^\intercal\left(\Gamma^{-1} - (\widehat{\Gamma}_u^S)^{-1}\right)F\right\|\\ = &O_p\left(dm_d(n^{-1/4+\delta/2}\sqrt{\log d} + d^{-1/2}m_d)^{1-q}\right),\end{aligned}$$

combining these inequalities yields

$$\|L_6\| = O_p\left(m_d^3(n^{-1/4+\delta/2}\sqrt{\log d} + d^{-1/2}m_d)^{1-q}\right).$$

On the other hand, using the Sherman - Morrison - Woodbury formula again,

$$\begin{aligned}
&\left\|\widetilde{\Sigma}^{-1} - \Sigma^{-1}\right\| \\
=&\left\|\left(t^{-1}\beta\bar{\mathcal{X}}\bar{\mathcal{X}}^\intercal\beta^\intercal + \Gamma\right)^{-1} - (\beta\mathrm{E}\beta^\intercal + \Gamma)^{-1}\right\| \\
\leq& \left\|\Gamma^{-1}\right\|^2 \|\beta H\|^2 \left\|\left(\left(tH^\intercal\left(\bar{\mathcal{X}}\bar{\mathcal{X}}^\intercal\right)^{-1}H + H^\intercal\beta^\intercal\Gamma^{-1}\beta H\right)^{-1}\right.\right. \\
&\left.\left. - \left(H^\intercal\mathrm{E}^{-1}H + H^\intercal\beta^\intercal\Gamma^{-1}\beta H\right)^{-1}\right)\right\| \\
\leq& Kd\left\|tH^\intercal\left(\bar{\mathcal{X}}\bar{\mathcal{X}}^\intercal\right)^{-1}H + H^\intercal\beta^\intercal\Gamma^{-1}\beta H\right\|^{-1}\left\|H^\intercal\mathrm{E}^{-1}H + H^\intercal\beta^\intercal\Gamma^{-1}\beta H\right\|^{-1} \\
&\left\|t\left(\bar{\mathcal{X}}\bar{\mathcal{X}}^\intercal\right)^{-1} - \mathrm{E}^{-1}\right\| \\
=&O_p\left(m_d n^{-1/4+\delta/2}\sqrt{\log d}\right).
\end{aligned}$$

By the triangle inequality, we obtain

$$\begin{aligned}
\left\|(\widehat{\Sigma}_u^S)^{-1} - \Sigma^{-1}\right\| \leq& \left\|(\widehat{\Sigma}_u^S)^{-1} - \widetilde{\Sigma}^{-1}\right\| + \left\|\widetilde{\Sigma}^{-1} - \Sigma^{-1}\right\| \\
=&O_p\left(m_d^3(n^{-1/4+\delta/2}\sqrt{\log d} + d^{-1/2}m_d)^{1-q}\right).
\end{aligned}$$

∎