



Knowing factors or factor loadings, or neither? Evaluating estimators of large covariance matrices with noisy and asynchronous data

Chaoxing Dai^a, Kun Lu^b, Dacheng Xiu^{a,*}

^a Booth School of Business, University of Chicago, 5807 S Woodlawn Avenue, Chicago, IL 60637, USA

^b Department of ORFE, Princeton University, Sherrerd Hall, Charlton Street, Princeton, NJ 08544, USA

ARTICLE INFO

Article history:

Available online 11 October 2018

JEL classification:

C13
C14
C55
C58
G01

Keywords:

High-dimensional data
High-frequency data
Factor model
Pre-averaging estimator
Portfolio allocation
Low-rank plus sparse covariance matrix
Barra covariance matrix estimator

ABSTRACT

We investigate estimators of factor-model-based large covariance (and precision) matrices using high-frequency data, which are asynchronous and potentially contaminated by the market microstructure noise. Our estimation strategies rely on the pre-averaging method with refresh time to solve the microstructure problems, while using three different specifications of factor models with a variety of thresholding methods, respectively, to battle the curse of dimensionality. To estimate a factor model, we either adopt the time-series regression (TSR) to recover loadings if factors are known, or use the cross-sectional regression (CSR) to recover factors from known loadings, or use the principal component analysis (PCA) if neither factors nor their loadings are assumed known. We compare the convergence rates in these scenarios using the joint in-fill and increasing dimensionality asymptotics. To evaluate the empirical trade-off between robustness to model misspecification and statistical efficiency among all 30 combinations of estimation strategies, we run a horse race on the out-of-sample portfolio allocation with Dow Jones 30, S&P 100, and S&P 500 index constituents, respectively, and find the pre-averaging-based strategy using TSR or PCA with location thresholding dominates, especially over the subsampling-based alternatives.

© 2018 Elsevier B.V. All rights reserved.

1. Introduction

Factor models provide a parsimonious representation of the dynamics of asset returns, as motivated by Ross (1976)'s arbitrage pricing theory. Since this seminal work, researchers have devoted significant effort to the search for proxies of factors (e.g., Fama and French (1993, 2015)) or characteristics of stocks (e.g., Daniel and Titman (1997)) to explain the cross-sectional variation of expected returns. These factors and characteristics also serve as natural candidates for factors and loadings that drive the time-series dynamics of stock returns. In this paper, we make use of a factor model to assist the estimation of large covariance matrices among stock returns.

The factor-model specification leads to a low-rank plus sparse structure of the covariance matrix, which guarantees a well-conditioned estimator as well as a desirable performance of its inverse (precision matrix). To estimate the low-rank component, we consider three scenarios: known factors, known factor loadings, or unknown factors and factor loadings. In the first two scenarios, we employ a time-series regression (TSR) or a cross-sectional regression (CSR) to estimate the

* Corresponding author.

E-mail addresses: chaoxingdai@chicagobooth.edu (C. Dai), kunl@princeton.edu (K. Lu), dacheng.xiu@chicagobooth.edu (D. Xiu).

unknown loadings or factors, using either portfolios and ETFs as proxies for factors, or characteristics as proxies for factor loadings. In the third scenario, we employ the principal component analysis (PCA) to identify latent factors and their loadings. Combining the estimated factors and/or their loadings yields the low-rank component of the covariance matrix. With respect to the sparse component, we adopt a variety of thresholding methods that warrant positive semi-definite estimates of the covariance matrix.

In addition, we use the large cross section of transaction-level prices available at high frequencies. High-frequency data provide a unique opportunity to measure the variation and covariation among stock returns. The massive amount of data facilitates the use of simple nonparametric estimators within a short window, such as the sample covariance matrix estimator on a daily, weekly, or monthly basis, so that several issues associated with parametric estimation using low-frequency time series covering a long timespan become irrelevant, such as structural breaks and time-varying parameters; see, e.g., [Aït-Sahalia and Jacod \(2014\)](#) for a review. 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 15 min, exacerbating the curse of dimensionality due to data elimination.

By adapting the pre-averaging estimator designed for low-dimensional covariance matrix estimation, e.g., [Jacod et al. \(2009\)](#), we construct noise-robust estimators for large covariance matrices, making use of a factor model in each of the three aforementioned scenarios. Using the large deviation theory of martingales, we establish the desired consistency of our covariance matrix estimators under the infinity norm (on the vector space) and the weighted quadratic norm, as well as the precision matrix estimators under the operator norm. Moreover, we show TSR converges as the sample size increases, regardless of a fixed or an increasing dimension. By contrast, the convergence of CSR and PCA requires a joint increase of the dimensionality and the sample size – the so-called blessings of dimensionality; see [Donoho et al. \(2000\)](#).

Empirically, we analyze the out-of-sample risk of optimal portfolios in a horse race among various estimators of the covariance matrix as inputs. The portfolios comprise constituents of Dow Jones 30, S&P 100, and S&P 500 indices, respectively. We find covariance matrix estimators based on pre-averaged returns sampled at refresh times outperform those based on returns subsampled at a fixed 15-min frequency, for almost all combinations of estimation strategies and thresholding methods. Moreover, either TSR or PCA, plus the location thresholding that utilizes the Global Industry Classification Standards (GICS) codes, yield the best performance for constituents of the S&P 500 and S&P 100 indices, whereas TSR dominates in the case of Dow Jones 30.

Our paper is closely related to a growing literature on continuous-time factor models for high-frequency data. [Fan et al. \(2016\)](#) and [Aït-Sahalia and Xiu \(2017b\)](#) develop the asymptotic theory for large dimensional factor models with known and unknown factors, respectively, assuming a synchronous and noiseless dataset. Their simulations show a clear breakdown of either TSR or PCA when noise is present and the sampling frequency is more than every 15 min. [Pelger \(2015a\)](#) and [Pelger \(2015b\)](#) develop the central limit results of such models in the absence of the noise. Their asymptotic results are element-wise, whereas the theoretical results in this paper focus on matrix-wise properties. [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 et al. \(2013b\)](#), [Tao et al. \(2011, 2013a\)](#), and [Kim et al. \(2016\)](#) for related results. [Brownlees et al. \(2017\)](#) impose the sparsity condition on the inverse of the covariance matrix or the inverse of the idiosyncratic covariance matrix. By contrast, we impose the sparsity assumption on the covariance of the idiosyncratic components of a factor model, as motivated by the economic theory, which also fits the empirical data better.

Our paper is also related to the recent literature on the estimation of the low-dimensional covariance matrix using noisy high-frequency data. The noise-robust estimators include, among others, the multivariate realized kernels by [Barndorff-Nielsen et al. \(2011\)](#), the quasi-maximum likelihood estimator by [Aït-Sahalia et al. \(2010\)](#) and [Shephard and Xiu \(2017\)](#), the pre-averaging estimator by [Christensen et al. \(2010\)](#), the local method of moments by [Bibinger et al. \(2014\)](#), and the two-scale and multi-scale estimators by [Zhang \(2011\)](#) and [Bibinger \(2012\)](#). [Shephard and Xiu \(2017\)](#) document the advantage of using a factor model in their empirical study, even when the dimension of assets is as low as 13. We build our estimator based on the pre-averaging method because of its simplicity in deriving the in-fill and high-dimensional asymptotic results. Allowing for increasing dimensionality asymptotics sheds light on important statistical properties of the covariance matrix estimators, such as minimum and maximum eigenvalues, condition numbers, etc, which are critical for portfolio allocation exercises. [Aït-Sahalia and Xiu \(2017a\)](#) develop a related theory of PCA for low-dimensional high-frequency data.

Our paper is also related to the recent literature on large covariance matrix estimation with low-frequency data. [Fan et al. \(2008\)](#) propose a large covariance matrix estimator using a strict factor model with observable factors. [Fan et al. \(2011\)](#) extend this result to approximate factor models. [Fan et al. \(2013\)](#) 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 \(2008a,b\)](#); [Cai and Liu \(2011\)](#); [Rothman et al. \(2009\)](#).

Our paper shares the theoretical insight with the existing literature of factor models specified in discrete time. [Bai and Ng \(2002\)](#) and [Onatski \(2010\)](#) propose estimators to determine the number of factors. [Bai \(2003\)](#) develops the element-wise inferential theory for factors and their loadings. These papers, including ours, allow for more general models than the approximate factor models introduced in [Chamberlain and Rothschild \(1983\)](#). The above factor models are static, as opposed to the dynamic factor models in which the lagged values of the unobserved factors may also affect the observed dependent variables. Inference on dynamic factor models are developed in [Forni et al. \(2000\)](#); [Forni and Lippi \(2001\)](#), [Forni et al. \(2004\)](#), and [Doz et al. \(2011\)](#); see [Croux et al. \(2004\)](#) for a discussion.

Factor models based on stock characteristics date back to [Rosenberg \(1974\)](#), who suggests modeling factor betas of stocks as linear functions of observable security characteristics. [Connor and Linton \(2007\)](#) and [Connor et al. \(2012\)](#) further extend this model to allow for nonlinear or nonparametric functions. One of our covariance matrix estimators, namely, CSR, is designed to leverage the linear factor model with characteristics as loadings. Such an estimation strategy is widely used in the financial industry, but is largely ignored by the academic literature.¹ Our asymptotic analysis of this estimator fills in this gap.

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 each scenario of the factor model and establishes the asymptotic properties of these estimators. Section 4 discusses the choice of tuning parameters. Section 5 provides Monte Carlo simulation evidence. Section 6 evaluates these estimators in an out-of-sample optimal portfolio allocation race. The [Appendix](#) contains all mathematical proofs.

2. Model setup and assumptions

2.1. Notation

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 j th (descending order), the minimum, and the maximum eigenvalues of a square matrix A , respectively. In addition, we use $\|A\|_1$, $\|A\|$, $\|A\|_F$, and $\|A\|_\Sigma$ to denote the \mathbb{L}_1 norm, the operator norm (or \mathbb{L}_2 norm), the Frobenius norm, and the weighted quadratic norm of a matrix A , that is, $\max_j \sum_i |A_{ij}|$, $\sqrt{\lambda_{\max}(A^\top A)}$, $\sqrt{\text{Tr}(A^\top A)}$, and $d^{-1/2} \|\Sigma^{-1/2} A \Sigma^{-1/2}\|_F$, respectively. Note $\|A\|_\Sigma$ is only defined for a $d \times d$ square matrix. We use \mathbb{I}_d to denote a $d \times d$ identity matrix. All vectors are regarded as column vectors, unless otherwise specified. When A is a vector, both $\|A\|$ and $\|A\|_F$ are equal to its Euclidean norm. We also use $\|A\|_{\text{MAX}} = \max_{i,j} |A_{ij}|$ to denote the \mathbb{L}_∞ norm of A on the vector space. We use e_i to denote a d -dimensional column vector whose i th 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.

2.2. Factor dynamics

Let Y be a d -dimensional log-price process, X be an r -dimensional factor process, Z be the idiosyncratic component, and β be a constant factor loading matrix of size $d \times r$. We make the following assumption about their dynamic relationship:

Assumption 1. Suppose Y_t follows a continuous-time factor model:

$$Y_t = \beta X_t + Z_t, \quad (1)$$

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 Z_t is another continuous Itô semimartingale, satisfying

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

where W_s and B_s are standard Brownian motions. In addition, h_s and f_s are progressively measurable. Moreover, the processes η_s and γ_s are càdlàg, and, writing $e_s = \eta_s \eta_s^\top$, $g_s = \gamma_s \gamma_s^\top$, e_s , e_{s-} , g_s , and g_{s-} are positive-definite. Finally, for all $1 \leq i, j \leq r$, $1 \leq k, l \leq d$, $|\beta_{ij}| \leq C$, for some $C > 0$, and there exists a locally bounded process H_s , such that $|h_{i,s}|$, $|\eta_{ij,s}|$, $|\gamma_{kl,s}|$, $|e_{ij,s}|$, $|f_{kl,s}|$, and $|g_{kl,s}|$ are bounded by H_s uniformly for $0 \leq s \leq t$.

Assumption 1 is fairly general except for two important limitations: β is constant and jumps are excluded. The same constant β assumption has been adopted by e.g., [Todorov and Bollerslev \(2010\)](#) in a low-dimensional setting and [Fan et al. \(2016\)](#) and [Aït-Sahalia and Xiu \(2017b\)](#) in high-dimensional settings. In our empirical study, we impose a constant β within each month, because β is available from the MSCI Barra at a monthly frequency.

To emphasize and highlight the theoretical trade-offs in the estimation from a high-dimensional perspective, we exclude jumps from our theoretical analysis to avoid delving into unnecessary technicalities. As a result, our empirical covariance matrix estimates contain the quadratic covariation contributed by co-jumps. Although co-jumps may be an important component, we do not find it particularly important to separate them from the total quadratic covariation for the large portfolio allocation exercise in which we are interested. Moreover, separating jumps would substantially complicate the estimation procedure, e.g., with more tuning parameters. In this paper, we desire simpler estimators while leaving analysis of jumps for future work.

¹ Barra Inc., which was acquired by MSCI Inc., was a leading provider of this type of covariance matrix to practitioners; see, e.g., [Kahn et al. \(1998\)](#).

Next, we impose the usual exogeneity assumption:

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

Our main goal is to estimate the integrated covariance matrix of Y , denoted as $\Sigma = \frac{1}{t} \int_0^t c_s ds$, where c_s is the spot covariance of Y_s . [Assumptions 1](#) and [2](#) infer a factor structure on c_s :

$$c_s = \beta e_s \beta^\top + g_s, \quad 0 \leq s \leq t.$$

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

$$\Sigma = \beta E \beta^\top + \Gamma,$$

where

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

We omit the dependence of Σ , E , and Γ on t for brevity in notation.

Next we impose that factors are pervasive, in the sense that they influence a large number of assets; see, e.g., [Chamberlain and Rothschild \(1983\)](#).

Assumption 3. E has distinct eigenvalues, with $\lambda_{\min}(E)$ bounded away from 0. Moreover, there exists some positive-definite matrix B such that $\|d^{-1} \beta^\top \beta - B\| = o(1)$, as $d \rightarrow \infty$, and $\lambda_{\min}(B)$ is bounded away from 0.

As in [Aït-Sahalia and Xiu \(2017b\)](#), [Assumption 3](#) leads to the identification of the number of factors when factors and their loadings are latent. [Fan et al. \(2016\)](#) also use it in the case of known factors, when building the operator norm bound for the precision matrix. Such an assumption may also be restrictive in that it excludes the existence of weak factors; see, e.g., [Onatski \(2010\)](#). Dealing with weak factors requires a rather different setup, so we leave it for future work.

2.3. Sparsity

For high-dimensional covariance matrix estimation, a certain “sparsity” condition is necessary for dimension reduction, in addition to a factor model, because the idiosyncratic component of the covariance matrix, once the low-rank component is removed, is equally large. One cannot obtain a good estimate of it without additional assumptions. Sparsity seems a reasonable choice for both known-factor and unknown-factor models given the empirical findings of [Aït-Sahalia and Xiu \(2017b\)](#).

We define m_d as the degree of sparsity of Γ , where

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

The sparsity assumption imposes $m_d/d \rightarrow 0$. This notion of sparsity follows from [Rothman et al. \(2009\)](#) and [Bickel and Levina \(2008b\)](#). When $q = 0$, m_d is equal to the maximum number of non-zero elements in rows of Γ , the usual notion used by [Bickel and Levina \(2008a\)](#). In this case, sparsity simply means each row of Γ contains few non-zero elements. [Fan et al. \(2016\)](#) and [Aït-Sahalia and Xiu \(2017b\)](#) consider this special case, while requiring Γ to be block diagonal. Our assumption below is more general.

Under this notion of sparsity, we have

$$\|\Gamma\| \leq \|\Gamma\|_1 = \max_{i \leq d} \sum_{j \leq d} |\Gamma_{ij}| = O(m_d).$$

Therefore, imposing $m_d/d \rightarrow 0$ creates a gap between eigenvalues of the low rank ($\beta E \beta^\top$) and the sparse (Γ) components of the covariance matrix Σ , which is essential for identification in a latent factor model specification, and for estimation of the factor models in general.

To use sparsity for estimation, we define a class of thresholding functions $s_\lambda(z) : \mathbb{R} \rightarrow \mathbb{R}$, which satisfies

$$(i) |s_\lambda(z)| \leq |z|; \quad (ii) s_\lambda(z) = 0 \text{ for } |z| \leq \lambda; \quad (iii) |s_\lambda(z) - z| \leq \lambda.$$

As discussed in [Rothman et al. \(2009\)](#), Condition (i) imposes shrinkage, condition (ii) enforces thresholding, and condition (iii) restricts the amount of shrinkage to be no more than λ . The exact three requirements of $s_\lambda(\cdot)$ ensure desirable statistical properties of the estimated covariance matrix. Examples of such thresholding functions we use include hard thresholding, soft thresholding, smoothly clipped absolute deviation (SCAD) ([Fan and Li, 2001](#)), and adaptive lasso (AL) ([Zou, 2006](#)):

$$s_\lambda^{\text{Hard}}(z) = z \mathbf{1}(|z| > \lambda), \quad s_\lambda^{\text{Soft}}(z) = \text{sign}(z)(|z| - \lambda)_+, \quad s_\lambda^{\text{AL}}(z) = \text{sign}(z)(|z| - \lambda^{\eta+1} |z|^{-\eta})_+,$$

$$s_{\lambda}^{\text{SCAD}}(z) = \begin{cases} \text{sign}(z)(|z| - \lambda)_+, & \text{when } |z| \leq 2\lambda; \\ \frac{(a-1)z - \text{sign}(z)a\lambda}{a-2}, & \text{when } 2\lambda < |z| \leq a\lambda; \\ z, & \text{when } a\lambda < |z|. \end{cases}$$

where $a = 3.7$ and $\eta = 1$, as suggested by Rothman et al. (2009). We adopt these functions in the construction of the estimators in Section 3. Although these choices lead to the same convergence rate from our analysis, the resulting finite sample performance of the covariance matrices differ quite a bit, which we investigate in simulations and the empirical study.

2.4. Microstructure noise

We analyze three scenarios of factor models, depending on whether the factor X or its loading β are known. We use the term “known” instead of “observable”, because even if we assume factor X can be proxied by certain portfolios in the literature, for instance, the Fama–French three factors by Fama and French (1993), we allow for potential microstructure noise so that the true factors are always latent in our setup.

The first scenario assumes X is known, in which case, we denote the observed factor as X^* :

$$Y_{t_j^i}^* = Y_{t_j^i} + \varepsilon_{t_j^i}^y, \quad X_{t_j^i}^* = X_{t_j^i} + \varepsilon_{t_j^i}^x,$$

for $1 \leq i \leq d$ and $1 \leq j \leq N_t^i$, where ε^y and ε^x are some additive noises associated with the observations at sampling times t_j^i s, t_j^i denotes the arrival time of the j th transaction of asset i , and N_t^i is the number of transactions for asset i . We can thereby rewrite the factor model (1) as

$$Y_{t_j^i}^* = \beta X_{t_j^i}^* + Z_{o,t_j^i}^*, \quad (2)$$

where $Z_{o,t_j^i}^* = Z_{t_j^i} + \varepsilon_{t_j^i}^y - \beta \varepsilon_{t_j^i}^x$. Barring from the noise, this model is a standard linear regression. In the empirical study, we regard those portfolios that are useful to explain the cross section of expected asset returns as factors, including the five Fama–French factors Fama and French (2015) and the momentum factor Carhart (1997).² We also add industry portfolios as suggested by King (1966).

The second scenario assumes β is known and perfectly observed, yet Y is again contaminated, so we can write the model as

$$Y_{t_j^i}^* = \beta X_{t_j^i} + Z_{u,t_j^i}^*, \quad (3)$$

where $Z_{u,t_j^i}^* = Z_{t_j^i} + \varepsilon_{t_j^i}^y$. This model dates back to Rosenberg (1974), who developed a factor model of stock returns in which the factor loadings of stocks are linear functions of observable security characteristics. This model is equivalent to a model with characteristics as β s associated with some linear latent factors. In the empirical study, we use 13 characteristics obtained from the MSCI Barra, a leading company that provides factors and covariance matrices using this method.

In the third scenario, we only observe a noisy Y , so the model can be written into the same form as (3). This model is a “noisy” version of the approximate factor model by Chamberlain and Rothschild (1983); Bai (2003), which can only be identified as the dimension of Y increases to ∞ , thanks to the “blessing” of the dimensionality.

With respect to the microstructure noises, following (Kim et al., 2016), we assume

Assumption 4. Both $\{\varepsilon_i^x\}$ and $\{\varepsilon_i^y\}$ have the following structure:

$$\varepsilon_{i,t_j^i} = u_{i,t_j^i} + v_{i,t_j^i},$$

where, for each $1 \leq i \leq d$, $1 \leq j \leq N_t^i$, writing $\Delta_{t_j^i} = t_j^i - t_{j-1}^i$,

$$u_{i,t_j^i} = \sum_{l=0}^{\infty} \rho_{i,t_j^i-l} \xi_{i,t_j^i-l}, \quad v_{i,t_j^i} = \sum_{l=0}^{\infty} b_{i,t_j^i-l} \Delta_{t_j^i-l}^{-1/4} [\tilde{B}_{i,t_j^i-l} - \tilde{B}_{i,t_j^i-l-1}].$$

We assume ξ_{i,t_j^i} and ξ_{j,t_j^j} are random variables with mean 0, and independent when $l \neq l'$, but potentially dependent for $l = l'$. Also, ρ_{i,t_j^i} is bounded in probability with $\sum_{l=0}^{\infty} |\rho_{i,t_j^i-l}| < \infty$ uniformly in i , and ξ_{i,t_j^i} is independent of the filtration $\{\mathcal{F}_t\}$ generated by X and Z . Moreover, \tilde{B}_i is a Brownian motion independent of ξ_i but potentially correlated with W and B of Assumption 1, and b_{i,t_j^i} is adapted to the filtration $\{\mathcal{F}_{t_j^i}\}$, and bounded in probability with $\sum_{l=0}^{\infty} |b_{i,t_j^i-l}| < \infty$ uniformly in i . Additionally, we assume there exists $\varsigma > 2$, such that $\max_{i,j} \mathbb{E}|\varepsilon_{i,t_j^i}|^{2\varsigma} < \infty$.

² Although these factors explain the cross section of expected returns, they also account for significant variations in the time series of realizations.

The microstructure noise has two independent components: u and v , where u allows for serial dependence, and v allows for correlation with returns. Here v_{i,t_j^i} has scaled Brownian increments, where the scale factor is of order $\Delta_{t_{j-1}^i}^{-1/4}$.³ This assumption is motivated from the microstructure theory and existing empirical findings that order flows tend to cluster in the same direction and to be correlated with returns in the short run; see, e.g., Hasbrouck (2007), Brogaard et al. (2014). This assumption is therefore more realistic in particular for data sampled at ultra-high frequencies. Kalnina and Linton (2008) and Barndorff-Nielsen et al. (2011) adopt a similar assumption.

2.5. Asynchronicity

Because the transactions arrive asynchronously, adopting the refresh time sampling scheme proposed by Martens (2004) prior to estimation is common. The first refresh time is defined as $t_1 = \max \{t_1^1, t_1^2, \dots, t_1^d\}$. The subsequent refresh times are defined recursively as

$$t_{j+1} = \max \left\{ t_{N_{t_j}^1+1}^1, \dots, t_{N_{t_j}^d+1}^d \right\},$$

where $N_{t_j}^i$ is the number of transactions for asset i prior to time t_j . We denote by n the resulting sample size after refresh time sampling.

Effectively, this sampling scheme selects the most recent transactions at refresh times, which avoids adding zero returns artificially, because by design, all assets have at least one update between two refresh times. By comparison, the alternative previous-tick subsampling scheme by Zhang (2011) discards more data in order to avoid artificial zero returns. That said, the refresh time scheme is notoriously influenced by the most illiquid asset, which largely determines the number of observations after sampling. Pair-wise refresh time, as suggested by Aït-Sahalia et al. (2010), is more efficient for entry-wise consistency of the covariance matrix. In the same spirit, Hautsch et al. (2012) adopt a more general strategy that conducts refresh-time sampling on blocks of assets formed by sorting on liquidity. The resulting covariance matrix has desirable entry-wise consistency, but its matrix-wise properties are rather involved to analyze. Since our focus is the consistency under matrix-wise norms, we adopt the refresh time sampling throughout. We also find empirically that the refresh time approach delivers satisfactory performance. We make the following assumption on the observation times, following Kim et al. (2016):

Assumption 5. The observation time t_j^i s, $1 \leq j \leq N_t^i$, $1 \leq i \leq d$, are independent of the price process X_t and Z_t , and the noise ε^x and ε^y . We assume the intervals between two adjacent observations are independent, and that there exist constants \bar{n} , C , and $\varsigma > 2$, such that

$$\max_{1 \leq i \leq d} \mathbb{E} |t_j^i - t_{j-1}^i|^a \leq C \bar{n}^{-a}, \text{ for any } 1 \leq a \leq 2\varsigma,$$

and that $c_1 \bar{n} \leq n \leq c_2 \bar{n}$ holds with probability approaching 1, where c_1 and c_2 are some positive constants.

A large literature is devoted to the modeling of durations, i.e., time intervals between adjacent transactions, since the seminal paper by Engle and Russell (1998), which proposes an autoregressive conditional duration (ACD) model and shows this parametric model can successfully describe the evolution of time durations for (heavily traded) stocks. Our focus here is the covariance matrix of returns, so we are agnostic about the dynamics of durations. The independence between durations and prices is a strong assumption, yet is commonly used in the literature, with the exception of Li et al. (2014). This assumption means we can make our inference regarding the times of trades, and therefore refresh times, fixed.

To simplify the notation, we treat n as if it were deterministic in what follows. Also, we use $Y_{t_i}^*$ to denote the most recent observation prior to or at the i th refresh time, and relabel the associated noise as ε_i^x and ε_i^y . Note $Y_{t_i}^* - \varepsilon_i^y$ is not necessarily equal to Y_{t_i} . Instead, it equals the value of Y at actual transaction times. The same convention applies to X^* and Z^* .

3. Three estimators and their asymptotic properties

We now proceed to the estimators and their asymptotic properties. Our results rely on the joint large sample ($n \rightarrow \infty$) and increasing dimensionality ($d \rightarrow \infty$) asymptotics, with a fixed number of factors r and a fixed time window $[0, t]$.

To deal with the bias due to the microstructure noise, we adopt the pre-averaging method proposed by Jacod et al. (2009) to pre-weight the returns. Specifically, we divide the whole sample into a sequence of blocks, with the size of each block being k_n , which satisfies:

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

³ The scaled factor is slightly more restrictive than that of the low dimensional setting considered in Ikeda (2016) and Varneskov (2016), where they use $\Delta_{t_{j-1}^i}^{-1/2}$. This rate only allows for moderate endogeneity which our estimator is robust to.

where $\theta > 0$ and $\delta > 0$. We set $n_\delta^{-1/2} := n^{-1/4+\delta/2} + n^{-2\delta}$, in which $n^{-2\delta}$ reflects the bias due to the microstructure noise, whereas $n^{-1/4+\delta/2}$ is the convergence rate of the pre-averaging estimator given by Christensen et al. (2010) in the low dimensional setting. $n_\delta^{-1/2}$ is the effective sample size as we will see below.

Our choice of $\delta > 0$ is not optimal. Nonetheless, it results in a simpler estimator without need of bias-correction, which also relies on fewer tuning parameters. More importantly, it guarantees a semi-definite covariance matrix estimate in any finite sample, a desirable property on which our following-up thresholding procedure relies.

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

$$\phi_1(s) = \int_s^1 g'(u)g'(u-s)du, \quad \phi_2(s) = \int_s^1 g(u)g(u-s)du.$$

For a sequence of vectors V , we define its weighted average return as \bar{V} , given by

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

and $\Delta_i^n V = V_i - V_{i-1}$, for $i = 1, 2, \dots, n$, and $V_i \in \{X_{t_i}, Y_{t_i}, Z_{t_i}, X_{t_i}^*, Y_{t_i}^*, Z_{t_i}^*, \varepsilon_i^x, \varepsilon_i^y\}$.

In what follows, we propose three estimators corresponding to different scenarios of the factor model. Each estimator uses the sample covariance matrix of these pre-averaged returns as an input, which leads to robustness to the microstructure noise. Intuitively, the effect of the noise is dominated by a strengthened return signal of each block.

3.1. Time-Series Regression (TSR)

When factors are known, we adopt a time-series regression-based approach using \bar{Y}^* and \bar{X}^* . We stack the d - and r -dimensional processes Y and X into U , and their pre-average returns \bar{Y}^*, \bar{X}^* into \bar{U}^* , respectively:

$$U := (Y^\top, X^\top)^\top, \quad \bar{U}^* := (\bar{Y}^{*\top}, \bar{X}^{*\top})^\top,$$

where U is a $(d+r)$ -dimensional process and \bar{U}^* is a $d \times (n - k_n + 2)$ dimensional matrix. The quadratic covariation of U is given by

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

which can be estimated by the sample covariance matrix of \bar{U}^* :

$$\hat{\Pi} = \frac{n}{n - k_n + 2} \frac{1}{\psi_2 k_n t} \sum_{i=0}^{n-k_n+1} \bar{U}_i^* \bar{U}_i^{*\top},$$

where \bar{U}_i^* is the $(i+1)$ th column of \bar{U}^* . We then construct estimators of each component of the covariance matrix as:

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

We apply thresholding to the covariance matrix estimates and obtain:

$$\hat{\Gamma}^S = (\hat{\Gamma}_{ij}^S), \quad \hat{\Gamma}_{ij}^S = \begin{cases} \hat{\Gamma}_{ij} & i = j, \\ s_{\lambda_{ij}}(\hat{\Gamma}_{ij}) & i \neq j, \end{cases} \quad (5)$$

A plug-in covariance matrix estimator is therefore given by:

$$\hat{\Sigma}_{\text{TSR}} = \hat{\beta} \hat{E} \hat{\beta}^\top + \hat{\Gamma}^S.$$

We postpone the choice of the thresholding method $s(\cdot)$ and tuning parameters λ_{ij} to Section 4, where we provide a procedure to guarantee the positive semi-definiteness of $\hat{\Sigma}_{\text{TSR}}$.

We provide the convergence rates under several matrix norms for both the covariance matrix $\hat{\Sigma}_{\text{TSR}}$ and its inverse:

Theorem 1. Suppose Assumptions 1–5 hold, and $n_\delta^{-1/2} \sqrt{\log d} = o(1)$. Then we have

$$\begin{aligned} \|\hat{\Sigma}_{\text{TSR}} - \Sigma\|_{\text{MAX}} &= O_p\left(n_\delta^{-1/2} \sqrt{\log d}\right), \\ \|\hat{\Sigma}_{\text{TSR}} - \Sigma\|_{\Sigma} &= O_p\left(d^{1/2} n_\delta^{-1} \log d + (n_\delta^{-1} \log d)^{(1-q)/2} m_d\right), \\ \|\hat{\beta} - \beta\| &= O_p\left(n_\delta^{-1/2} \sqrt{\log d}\right). \end{aligned}$$

Moreover, if $(n_\delta^{-1} \log d)^{(1-q)/2} m_d = o(1)$, we have

$$\left\| (\widehat{\Sigma}_{\text{TSR}})^{-1} - \Sigma^{-1} \right\| = O_p \left((n_\delta^{-1} \log d)^{(1-q)/2} m_d \right),$$

and $\lambda_{\min}(\widehat{\Sigma}_{\text{TSR}}) \geq \frac{1}{2} \lambda_{\min}(\Gamma)$, with probability approaching 1.

Theorem 1 establishes the convergence rate, which depends on the degree of sparsity m_d , q , the dimension d , and the local window length parameter δ . Under $\|\cdot\|_{\text{MAX}}$ norm, covariance matrix estimators with/without a factor model deliver the same rate, because even in a factor model, too many parameters for estimation remain from the low-rank component, which determines the low convergence rate under $\|\cdot\|_{\text{MAX}}$ norm.

In terms of the inverse, when $d > n$, estimating Σ^{-1} becomes infeasible without a factor model, whereas the factor-based covariance matrix is invertible with high probability, and the inverse converges to the target under the operator norm. Because $\|\cdot\|_{\Sigma}$ norm depends on both Σ and Σ^{-1} , and the latter is more accurately estimated using a factor model, under $\|\cdot\|_{\Sigma}$ norm, using a factor model gives a better rate than the rate without it, which would be $d^{1/2} n_\delta^{-1/2} \sqrt{\log d}$.

More importantly, the minimum eigenvalue of the resulting covariance matrix is bounded away from 0 with high probability, so that the covariance matrix estimate is well-conditioned. This property is essential to warrant an economically feasible optimal portfolio using the estimated covariance matrix as the input.

3.2. Cross-Sectional Regression (CSR)

If we observe the factor loading matrix β , we propose a cross-sectional regression approach that recovers X at first. We start with a scenario in which the data are synchronous and noise-free, because the asymptotic property of such an estimator is not available in the literature, to the best of our knowledge. The estimator can be constructed as

$$\widehat{\Sigma}_{\text{CSR}}^* = \beta \check{E}^* \beta^\top + \check{I}^{*S},$$

where

$$\check{X} = (\beta^\top \beta)^{-1} \beta^\top Y, \quad \check{E}^* = \frac{1}{t} \check{X} \check{X}^\top, \quad \check{I}^{*S} = \frac{1}{t} (Y - \beta \check{X}) (Y - \beta \check{X})^\top,$$

and

$$\check{I}^{*S} = (\check{I}_{ij}^{*S}), \quad \check{I}_{ij}^{*S} = \begin{cases} \check{I}_{ij} & i = j, \\ s_{\lambda_{ij}}(\check{I}_{ij}) & i \neq j, \end{cases}.$$

This estimator is similar in spirit to the covariance matrix estimator provided by the MSCI Barra; see [Kahn et al. \(1998\)](#). We analyze its properties as follows:

Theorem 2. Suppose [Assumptions 1–3, 5](#) hold, $n^{-1/2} \sqrt{\log d} = o(1)$. Then we have

$$\begin{aligned} \left\| \widehat{\Sigma}_{\text{CSR}}^* - \Sigma \right\|_{\text{MAX}} &= O_p \left(n^{-1/2} \sqrt{\log d} + d^{-1/2} m_d^{1/2} \right), \\ \left\| \widehat{\Sigma}_{\text{CSR}}^* - \Sigma \right\|_{\Sigma} &= O_p \left(\left[n^{-1/2} d^{1/4} \sqrt{\log d} + d^{-1/4} m_d^{1/2} \right]^2 + \left[n^{-1/2} \sqrt{\log d} + d^{-1/2} m_d^{1/2} \right]^{1-q} m_d \right), \\ \left\| \check{X} - X \right\| &= O_p \left(d^{-1/2} m_d^{1/2} \right). \end{aligned}$$

Moreover, if $\left[n^{-1/2} \sqrt{\log d} + d^{-1/2} m_d^{1/2} \right]^{1-q} m_d = o(1)$, we have

$$\left\| (\widehat{\Sigma}_{\text{CSR}}^*)^{-1} - \Sigma^{-1} \right\| = O_p \left(\left[n^{-1/2} \sqrt{\log d} + d^{-1/2} m_d^{1/2} \right]^{1-q} m_d \right),$$

and $\lambda_{\min}(\widehat{\Sigma}_{\text{CSR}}^*) \geq \frac{1}{2} \lambda_{\min}(\Gamma)$, with probability approaching 1.

Theorem 2 shows the CSR estimator does not converge under $\|\cdot\|_{\text{MAX}}$ when d is fixed, due to the second term $d^{-1/2} m_d^{1/2}$, unlike the TSR estimator. This finding is not surprising, because the cross-sectional regression exploits an increasing dimensionality to estimate X . The first term $n^{-1/2} \sqrt{\log d}$ is the same as that in the convergence rate of TSR in the absence of noise (see [Fan et al. \(2016\)](#)), because both approaches estimate Γ based on a thresholded sample covariance matrix estimator. Comparing this convergence rate with that of the PCA estimator given by [Aït-Sahalia and Xiu \(2017b\)](#), which is $n^{-1/2} \sqrt{\log d} + d^{-1/2} m_d$, is also interesting. The rate improvement in the CSR estimator comes from the second term and is due to the knowledge of β . Overall, the convergence rate of CSR depends on a striking trade-off between n and d .

In the general scenario where the noise plagues the data, we construct a pre-averaging-based covariance matrix estimator:

$$\widehat{\Sigma}_{\text{CSR}} = \beta \check{E} \beta^\top + \check{I}^S,$$

where

$$\tilde{X}^* = (\beta^\top \beta)^{-1} \beta^\top \tilde{Y}^*, \quad \tilde{E} = \frac{n}{n - k_n + 2} \frac{1}{\psi_2 k_n t} \tilde{X}^* \tilde{X}^{*\top}, \quad \tilde{\Gamma} = \frac{n}{n - k_n + 2} \frac{1}{\psi_2 k_n t} (\tilde{Y}^* - \beta \tilde{X}^*) (\tilde{Y}^* - \beta \tilde{X}^*)^\top,$$

and

$$\tilde{r}^S = (\tilde{\Gamma}_{ij}^S), \quad \tilde{\Gamma}_{ij}^S = \begin{cases} \tilde{\Gamma}_{ij} & i = j, \\ s_{\lambda_{ij}}(\tilde{\Gamma}_{ij}) & i \neq j, \end{cases}.$$

The next theorem gives the convergence rates

Theorem 3. Suppose Assumptions 1–5 hold, and $n_\delta^{-1/2} \sqrt{\log d} + d^{-1/2} m_d^{1/2} = o(1)$. Then we have

$$\begin{aligned} \|\hat{\Sigma}_{\text{CSR}} - \Sigma\|_{\text{MAX}} &= O_p \left(n_\delta^{-1/2} \sqrt{\log d} + d^{-1/2} m_d^{1/2} \right), \\ \|\hat{\Sigma}_{\text{CSR}} - \Sigma\|_{\Sigma} &= O_p \left(\left[n_\delta^{-1/2} \sqrt{\log d} d^{1/4} + d^{-1/2} m_d^{1/2} \right]^2 + \left[n_\delta^{-1/2} \sqrt{\log d} + d^{-1/2} m_d^{1/2} \right]^{1-q} m_d \right), \\ \|\tilde{X}^* - \bar{X}^*\| &= O_p \left(d^{-1/2} m_d^{1/2} \right). \end{aligned}$$

Moreover, if $\left[n_\delta^{-1/2} \sqrt{\log d} + d^{-1/2} m_d^{1/2} \right]^{1-q} m_d = o(1)$, we have

$$\left\| (\hat{\Sigma}_{\text{CSR}})^{-1} - \Sigma^{-1} \right\| = O_p \left(\left[n_\delta^{-1/2} \sqrt{\log d} + d^{-1/2} m_d^{1/2} \right]^{1-q} m_d \right),$$

and $\lambda_{\min}(\hat{\Sigma}_{\text{CSR}}) \geq \frac{1}{2} \lambda_{\min}(\Gamma)$, with probability approaching 1.

Compared to the results of Theorem 2, the rates in Theorem 3 remain the same except that n is replaced by n_δ , which is the effective sample size when noise is present. The same intuition as in the no-noise case holds as well. Moreover, as discussed previously, the choice of $\delta > 0$ leads to a simpler estimator despite it being less efficient.

3.3. Principal component analysis (PCA)

Without prior knowledge of factors or their loadings, we apply PCA to the pre-averaged covariance matrix estimate based on \tilde{Y}^* :

$$\tilde{\Sigma} = \frac{n}{n - k_n + 2} \frac{1}{\psi_2 k_n t} \sum_{i=0}^{n-k_n+1} \tilde{Y}_i^* \tilde{Y}_i^{*\top}.$$

Suppose $\hat{\lambda}_1 > \hat{\lambda}_2 > \dots > \hat{\lambda}_d$ are the simple eigenvalues of $\tilde{\Sigma}$, and $\hat{\xi}_1, \hat{\xi}_2, \dots, \hat{\xi}_d$ are the corresponding eigenvectors. Then $\tilde{\Sigma}$ can be decomposed as

$$\tilde{\Sigma} = \sum_{j=1}^{\hat{r}} \hat{\lambda}_j \hat{\xi}_j \hat{\xi}_j^\top + \tilde{\Gamma}, \quad (6)$$

where \hat{r} is an estimator of r to be introduced below. Similar to TSR, we apply thresholding on $\tilde{\Gamma}$ and obtain

$$\tilde{r}^S = (\tilde{\Gamma}_{ij}^S), \quad \tilde{\Gamma}_{ij}^S = \begin{cases} \tilde{\Gamma}_{ij} & i = j, \\ s_{\lambda_{ij}}(\tilde{\Gamma}_{ij}) & i \neq j, \end{cases}$$

and the resulting estimator of Σ is

$$\hat{\Sigma}_{\text{PCA}} = \sum_{j=1}^{\hat{r}} \hat{\lambda}_j \hat{\xi}_j \hat{\xi}_j^\top + \tilde{r}^S. \quad (7)$$

This estimator is motivated by the POET strategy by Fan et al. (2013) for low-frequency data, and adapted from the PCA approach by Ait-Sahalia and Xiu (2017b) for high-frequency data. The idea behind this strategy is that the eigenvectors corresponding to the first r eigenvalues of the sample covariance matrix can be used to construct proxies of β , so that the low-rank component of Σ can be approximated by the first term on the right-hand side of (6). The second term then approximates the sparse component of Σ , which leads to the construction of the estimator given by (7).

This estimator can also be constructed from a least-squares point of view, which seeks F and G such that

$$(F, G) = \arg \min_{F \in \mathcal{M}_{d \times \hat{r}}, G \in \mathcal{M}_{\hat{r} \times n}} \|\tilde{Y}^* - FG\|_F^2,$$

subject to the normalization:

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

(Bai and Ng, 2002) and Bai (2003) propose this estimator to estimate factors and their loadings in a factor model for low-frequency data.

Once we have estimates of factors and loadings, we can obtain the same \tilde{r} and \tilde{r}^S as above by:

$$\tilde{r} = \frac{n}{n - k_n + 2} \frac{1}{\psi_2 k_n t} (\bar{Y}^* - FG) (\bar{Y}^* - FG)^{\top}, \quad \tilde{r}^S = (\tilde{r}_{ij}^S), \quad \tilde{r}_{ij}^S = \begin{cases} \tilde{r}_{ij} & i = j, \\ s_{\lambda_{ij}}(\tilde{r}_{ij}) & i \neq j, \end{cases}$$

with which we can construct

$$\hat{\Sigma}_{\text{PCA}} = t^{-1} F G G^{\top} F^{\top} + \tilde{r}^S. \quad (8)$$

Although (7) is easier to implement, this equivalent form of $\hat{\Sigma}_{\text{PCA}}$ is useful in the proof and provides estimates of factors and their loadings (up to some rotation).

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

$$\hat{r} = \arg \min_{1 \leq j \leq r_{\max}} (d^{-1} \lambda_j(\tilde{\Sigma}) + j \times f(n, d)) - 1,$$

where r_{\max} is some upper bound. This estimator is similar to that of Aït-Sahalia and Xiu (2017b), which shares the spirit with Bai and Ng (2002). The penalty function $f(n, d)$ satisfies two criteria. On the one hand, the penalty is dominated by the signal, i.e., the value of $d^{-1} \lambda_j(\Sigma)$, for $1 \leq j \leq r$. Because $d^{-1} \lambda_r(\Sigma)$ is $O_p(1)$ as d increases, we select a penalty that shrinks to 0. On the other hand, we require the penalty to dominate the estimation error as well as $d^{-1} \lambda_j(\Sigma)$ for $r + 1 \leq j \leq d$ to avoid overshooting. The choice of r_{\max} does not play any role in theory, yet it warrants an economically meaningful estimate of \hat{r} in a finite sample or in practice.

Theorem 4. Suppose Assumptions 1–5 hold. Also, $n_{\delta}^{-1/2} \sqrt{\log d} + d^{-1/2} m_d = o(1)$, $f(n, d) \rightarrow 0$, and $f(n, d) (n_{\delta}^{-1/2} \sqrt{\log d} + d^{-1/2} m_d)^{-1} \rightarrow \infty$. Then we have

$$\begin{aligned} \|\hat{\Sigma}_{\text{PCA}} - \Sigma\|_{\text{MAX}} &= O_p \left(n_{\delta}^{-1/2} \sqrt{\log d} + d^{-1/2} m_d \right), \\ \|\hat{\Sigma}_{\text{PCA}} - \Sigma\|_{\Sigma} &= O_p \left(d^{1/2} n_{\delta}^{-1} \log d + d^{-1/2} m_d^2 + m_d \left(n_{\delta}^{-1/2} \sqrt{\log d} + d^{-1/2} m_d \right)^{1-q} \right). \end{aligned}$$

Also, there exists a $r \times r$ matrix H , such that with probability approaching 1, H is invertible, $\|HH^{\top} - \mathbb{I}_r\| = o_p(1)$, and

$$\begin{aligned} \|F - \beta H\|_{\text{MAX}} &= O_p \left(n_{\delta}^{-1/2} \sqrt{\log d} + d^{-1/2} m_d \right), \\ \|G - H^{-1}X\| &= O_p \left(n_{\delta}^{-1/2} \sqrt{\log d} + d^{-1/2} m_d \right). \end{aligned}$$

If, in addition, $n_{\delta}^{-1/2} \sqrt{\log d} = o(1)$, then $\lambda_{\min}(\hat{\Sigma}_{\text{PCA}})$ is bounded away from 0 with probability approaching 1, and

$$\|\hat{\Sigma}_{\text{PCA}}^{-1} - \Sigma^{-1}\| = O_p \left(m_d^3 \left(n_{\delta}^{-1/2} \sqrt{\log d} + d^{-1/2} m_d \right)^{1-q} \right).$$

Due to the fundamental indeterminacy of a factor model, we only identify the latent factors and their loadings up to some invertible matrix H . That said, the covariance matrix estimator itself is invariant to H .

3.4. A comparison of the three estimators

So far, we have obtained the convergence rates of all scenarios of factor models for the covariance matrix and its inverse, under a variety of norms.

We observe the usual tradeoff between efficiency and robustness in all scenarios. In terms of efficiency, TSR dominates CSR, which in turn dominates PCA. Nonetheless, PCA is more robust to model misspecification in that its construction utilizes the least amount of prior information. Interestingly, the loss of efficiency diminishes as the dimension of assets increases, thanks to the blessings of dimensionality. In light of this tradeoff, we resort to simulations in Section 5 for further comparison of the finite sample performance of these estimators, and to empirical data in Section 6 for evaluation of their relevance in practice.

4. Practical considerations

4.1. Choice of δ and k_n

As discussed in Christensen et al. (2010), the pre-averaging estimator we adopt is consistent in the low-dimensional case if $0 < \delta < 0.5$, but its CLT requires $0.1 < \delta < 0.5$ so that the asymptotic bias due to noise is negligible compared to the asymptotic variance.

The two terms in $n_\delta^{-1/2}$ exactly characterizes the bias–variance trade-off in our setting. Similar to Kim et al. (2016), the $n^{-2\delta}$ term is due to the bias of the microstructure noise, whereas $n^{-1/4+\delta/2}$ is due to the variance of the estimator. We thereby select $\delta = 0.1$ to balance the bias and variance.

The tuning parameter k_n is determined by θ , once δ is given. With a large number of observations, the estimates are not sensitive to the choice of k_n as long as d is moderately large. In simulations and empirical studies, we adopt a range of θ s, and thus k_n s, all of which lead to similar estimates that do not change our interpretations.

4.2. Choice of r and $f(n, d)$

A sensible choice of the penalty function could be

$$f(n, d) = \mu \left(n_\delta^{-1/2} \sqrt{\log d} + d^{-1} m_d \right)^\kappa \cdot \text{median}(\{\hat{\lambda}_1, \dots, \hat{\lambda}_d\}), \quad (9)$$

for some tuning parameters μ and κ . One might also use the perturbed eigenvalue ratio estimator in Pelger (2015a) to determine r , which gives almost the same result in simulations. The latter requires one less tuning parameter, but its proof is more involved. Alternatively, as argued by Aït-Sahalia and Xiu (2017b), we can simply regard r as a tuning parameter from the practical point of view. In fact, the performance of the estimator is not sensitive to \hat{r} , as long as \hat{r} is greater than or equal to r , yet is not too large, as shown from our simulation results. We conjecture that the same convergence rate holds for our estimators as long as $\hat{r} \geq r$ and \hat{r} is finite, which is indeed the case for parameter estimation in the interactive effect models; see, e.g., Moon and Weidner (2015). The proof likely involves the random matrix theory that is not available for continuous martingales. We leave this investigation for future work. In our empirical studies, we find that as soon as r is greater than 3 but not as large as 20, the comparison results remain the same qualitatively and the interpretations are identical.

4.3. Choice of the thresholding methods

We compare two types of thresholding methods on the residual covariance matrix, e.g., $\hat{\Gamma}$ constructed in TSR.⁴ The same applies to $\tilde{\Gamma}$ (CSR) and $\tilde{\Gamma}$ (PCA).

The first one is the location-based thresholding utilizing domain knowledge (denoted as location thresholding), as in Fan et al. (2016). This approach preserves positive semi-definiteness in a finite sample and is computationally efficient, because neither tuning nor optimization is involved. Specifically, we first sort the residual covariance matrix $\hat{\Gamma}$ into blocks by assets' industrial classifications (sector, industry group, industry, or sub-industry), and then apply a block-diagonal mask to this residual covariance matrix. The thresholding function can be written explicitly as:

$$s_{\lambda_{ij}}^{\text{loc}}(z) = z \mathbf{1}(\lambda_{ij} = 1), \quad \text{where } \lambda_{ij} = 1, \text{ if and only if } (i, j) \in \text{the same block.}$$

For each block, we have a positive semi-definite sub-matrix because $\hat{\Gamma}$ is positive semi-definite by construction, so that stacking these blocks on the diagonal produces a positive semi-definite $\hat{\Gamma}^s$.

The second class of methods we consider employ a threshold based on the sample correlation matrix. Specifically, we set

$$\lambda_{ij} = \tau \sqrt{\hat{\Gamma}_{ii} \hat{\Gamma}_{jj}},$$

where τ is some constant to be determined. With this threshold, we then apply Hard, Soft, SCAD, and AL thresholding methods with $s_{\lambda_{ij}}(\cdot)$ s given by Section 2.3, respectively, in the construction of (5). These methods do not always guarantee a positive semi-definite $\hat{\Gamma}^s$ in a finite sample. Also, when τ is small, $\hat{\Gamma}^s$ might not be sufficiently sparse.

To fix these issues, we find an appropriate τ via a grid search algorithm, following Fan et al. (2013). We start from a small value of τ , and gradually increase it until a positive semi-definite $\hat{\Gamma}^s$ is obtained and the degree of sparsity is below a certain threshold. As τ increases, the degree of sparsity decreases, and the solution shrinks toward the diagonal of $\hat{\Gamma}$, which is positive semi-definite. Thus, our grid search is guaranteed to produce a solution. In other words, this algorithm yields a positive semi-definite estimate in a finite sample. Note the grid search for τ is easier here than that of Fan et al. (2013), because τ is bounded between 0 and 1. In practice, a natural choice of the desired degree of sparsity can be obtained using that of the location-based thresholding, which is computationally less expensive than the cross-validation method.

5. Monte Carlo simulations

In this section, we investigate the finite sample performance of the pre-averaging estimator and compare it with the subsampling method discussed in Fan et al. (2016) and Aït-Sahalia and Xiu (2017b). The latter estimators are built upon the

⁴ Alternatively, one can adopt the adaptive thresholding method by Cai and Liu (2011). In our simulations, this approach does not perform as well as the location-based and correlation-based thresholding methods we consider.

realized covariance estimators using subsampled returns. We simulate 1000 paths from a continuous-time r -factor model of d assets specified as

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

where W_j is a standard Brownian motion and B_i is a d -dimensional Brownian motion, for $i = 1, 2, \dots, d$, and $j = 1, 2, \dots, r$. They are mutually independent. X_j is the j th factor, and we set X_1 as the market factor, with the associated loadings being positive. The covariance matrix of Z , Γ , is a block-diagonal matrix with $\Gamma_{i,l} = \gamma_i^\top \gamma_l$. We also allow for time-varying $\sigma_{j,t}^2$ which evolves as

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

where \tilde{B}_j is a standard Brownian motion with $\mathbb{E}[dB_{j,t}d\tilde{B}_{j,t}] = \rho_j dt$. We choose $d = 500$ and $r = 3$. We fix t at 21 trading days, i.e., $t = 1/12$. In addition, $\kappa = (3, 4, 5)$, $\theta = (0.09, 0.04, 0.06)$, $\eta = (0.3, 0.4, 0.3)$, $\rho = (-0.6, -0.4, -0.250)$, and $b = (0.05, 0.03, 0.02)$. As for the factor loadings, we sample $\beta_1 \sim \mathcal{U}[0.25, 1.75]$, and $\beta_2, \beta_3 \sim \mathcal{N}(0, 0.5^2)$. The diagonal elements of Γ are sampled independently from $\mathcal{U}[0.1, 0.2]$, with constant within-block correlations sampled from $\mathcal{U}[0.10, 0.40]$ for each block. To generate blocks with random sizes, we fix the largest block size at 35, and randomly generate the sizes of the blocks from a uniform distribution between 10 and 35, such that the total size of all blocks is d . β s and block sizes are randomly generated but fixed across Monte Carlo repetitions.

We simulate the noise in log prices for each asset as an MA(1) process, i.e., $\epsilon_{i,t}^x = \xi_{i,t} - 0.5\xi_{i,t-1}$, where $\xi_{i,t}$ is an i.i.d. normal noise 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. The log-normal distribution $\log \mathcal{N}(\mu, \sigma^2)$ has parameters $\mu = 2500$ and $\sigma = 0.8$, and the lower and upper truncation boundaries are 1000 and 23,400, respectively, which matches the empirical data on S&P 500 index constituents.

For pre-averaging estimators, we compare a range of local window lengths by varying θ in (4). For subsampling estimators, we compare a range of subsampling frequencies from every 5 min to every 65 min, denoted as Δ_n in seconds. We also add the benchmark estimates using noiseless returns sampled at a 5-min frequency without asynchronicity for comparison. In addition, we consider a “mixed” approach by applying the pre-averaging method to the subsampled data, in order to check the marginal effect of the refresh-time versus subsampling. Because the simulated Γ is a block-diagonal matrix, we regard the location-based thresholding method as the benchmark, and compare its performance with other thresholding methods in Section 4.3. We present the simulation results in Tables 1 and 2.

First, we find pre-averaging estimators achieve smaller errors compared to subsampling estimators, and are closer to the noiseless benchmark, for almost all model specifications and almost all criteria. The pre-averaging method achieves a slightly better estimate of the low-rank part $\beta E \beta^\top$, and a better residual part Γ , especially under the operator norm for Γ^{-1} , which in turn gives a better estimate of the precision matrix Σ^{-1} . Throughout, the pre-averaging estimators are robust to a wide range of tuning parameters θ . The sweet spot appears to be $\theta = 0.08$, whereas the optimal frequency for the subsampling method seems to be achieved near $\Delta_n = 900$.

Second, for comparison across various thresholding methods, we find in almost all scenarios that the Location thresholding achieves the best result, followed by Soft, AL, and SCAD thresholding, whereas Hard thresholding appears to be the worst. The differences are smaller in terms of $\|\hat{\Sigma} - \Sigma\|_{\text{MAX}}$ and $\|\hat{\Gamma} - \Gamma\|_{\text{MAX}}$, because the largest entry-wise errors are either achieved along the diagonal or on entries off the diagonal with large magnitudes, which are least affected by various thresholding methods. Nevertheless, the differences are most salient in terms of $\|(\hat{\Sigma})^{-1} - \Sigma^{-1}\|$ and $\|(\hat{\Gamma})^{-1} - \Gamma^{-1}\|$. The former is arguably the most important metric in this table, because it dictates the performance of the portfolio allocation in the empirical exercise.

Third, the comparison between the “mixed” approach versus subsampling depends on a bias–variance tradeoff. If the bias due to noise is large, the mixed approach outperforms; if the noise effect becomes negligible for a sufficiently low sampling frequency, the subsampling approach can outperform because its convergence rate is faster. By contrast, the “mixed” approach is always dominated by the pre-averaging based on the refresh time scheme, because the latter reserves more data and the two estimators are equally efficient given the same amount of data.

Finally, to demonstrate the impact of the selected number of factors on the PCA, we present in Table 3 the errors with a range of \hat{r} pre-set instead of being estimated. Due to space constraints, we only report the results with $\theta = 0.08$ for the pre-averaging, $\Delta_n = 900$ for the subsampling method, and the benchmark no-noise and synchronous case with $\Delta_n^* = 300$. Location thresholding is used in all cases. For the estimation of Σ and Σ^{-1} , we find when $\hat{r} < r$, the performance is much worse in every metric than the case with $\hat{r} = r$, whereas in the case of $\hat{r} > r$, the performance is only slightly worse, in particular when \hat{r} is within a reasonable range (smaller than 20). For the purpose of covariance matrix estimation, this result justifies treating r as a tuning parameter without estimating it. With respect to estimating the low-rank and sparse components of Σ , using an incorrect number of factors is harmful.

Table 1

Pre-averaging vs. subsampling using location and hard thresholding.

Estimator		Pre-Averaging				Subsampling				Noiseless	Mixed
Tuning Parameters		θ	θ	θ	θ	Δ_n	Δ_n	Δ_n	Δ_n	Δ_n^*	θ, Δ_n
		0.04	0.06	0.08	0.10	300	900	1,800	3,900	300	0.08, 900
Location Thresholding											
$\ \widehat{\Sigma} - \Sigma\ _{\text{MAX}}$	TSR	0.06	0.06	0.06	0.06	0.12	0.08	0.09	0.13	0.03	0.09
	CSR	0.05	0.04	0.04	0.04	0.11	0.07	0.07	0.09	0.03	0.06
	PCA	0.06	0.06	0.06	0.06	0.12	0.08	0.09	0.17	0.04	0.09
$\ \widehat{\beta}\widehat{E}\beta^\top - \beta E\beta^\top\ _{\text{MAX}}$	TSR	0.03	0.04	0.05	0.05	0.04	0.06	0.08	0.11	0.03	0.07
	CSR	0.02	0.02	0.03	0.03	0.02	0.03	0.04	0.06	0.02	0.04
	PCA	0.04	0.04	0.05	0.06	0.04	0.06	0.08	0.19	0.04	0.08
$\ \widehat{\Gamma} - \Gamma\ _{\text{MAX}}$	TSR	0.05	0.04	0.04	0.04	0.12	0.06	0.06	0.07	0.02	0.05
	CSR	0.05	0.04	0.03	0.03	0.11	0.06	0.06	0.07	0.02	0.05
	PCA	0.05	0.04	0.03	0.03	0.11	0.06	0.06	0.17	0.02	0.05
$\ \widehat{\Sigma} - \Sigma\ _{\Sigma}$	TSR	0.31	0.24	0.24	0.26	0.80	0.40	0.43	0.62	0.14	0.39
	CSR	0.30	0.23	0.22	0.23	0.79	0.37	0.37	0.48	0.13	0.33
	PCA	0.31	0.25	0.24	0.26	0.79	0.40	0.43	0.93	0.15	0.40
$\ (\widehat{\Sigma})^{-1} - \Sigma^{-1}\ $	TSR	5.70	4.62	4.42	5.92	9.43	6.50	10.45	30.16	3.57	9.82
	CSR	5.71	4.66	4.42	5.84	9.42	6.52	10.26	28.90	3.54	9.62
	PCA	5.70	4.63	4.44	5.93	9.41	6.50	10.49	28.68	3.56	9.84
$\ (\widehat{\Gamma})^{-1} - \Gamma^{-1}\ $	TSR	5.72	4.64	4.45	5.96	9.45	6.52	10.54	30.59	3.59	9.90
	CSR	5.74	4.68	4.45	5.89	9.44	6.54	10.38	29.33	3.57	9.71
	PCA	5.72	4.65	4.48	5.98	9.44	6.52	10.59	28.89	3.58	9.95
$\widehat{\Gamma}$	PCA	3.00	3.00	3.00	3.00	3.00	3.00	3.00	1.71	3.00	3.04
Hard Thresholding											
$\ \widehat{\Sigma} - \Sigma\ _{\text{MAX}}$	TSR	0.08	0.08	0.09	0.09	0.12	0.09	0.11	0.13	0.08	0.10
	CSR	0.08	0.08	0.08	0.08	0.11	0.08	0.08	0.09	0.07	0.08
	PCA	0.08	0.08	0.08	0.08	0.12	0.09	0.10	0.17	0.08	0.10
$\ \widehat{\Gamma} - \Gamma\ _{\text{MAX}}$	TSR	0.08	0.08	0.08	0.08	0.12	0.08	0.08	0.08	0.08	0.08
	CSR	0.08	0.08	0.08	0.08	0.11	0.08	0.08	0.08	0.08	0.08
	PCA	0.08	0.08	0.08	0.08	0.11	0.08	0.08	0.17	0.08	0.08
$\ \widehat{\Sigma} - \Sigma\ _{\Sigma}$	TSR	0.60	0.50	0.44	0.43	1.09	0.61	0.53	0.58	0.37	0.49
	CSR	0.57	0.46	0.41	0.39	1.04	0.56	0.46	0.41	0.35	0.42
	PCA	0.57	0.47	0.42	0.41	1.04	0.58	0.51	0.67	0.35	0.47
$\ (\widehat{\Sigma})^{-1} - \Sigma^{-1}\ $	TSR	8.40	7.65	7.20	7.02	10.65	8.52	7.79	7.56	6.58	7.44
	CSR	8.25	7.50	7.05	6.90	10.53	8.38	7.67	7.52	6.67	7.33
	PCA	8.23	7.46	7.13	7.22	10.52	8.35	7.61	8.45	7.43	7.38
$\ (\widehat{\Gamma})^{-1} - \Gamma^{-1}\ $	TSR	8.42	7.67	7.22	7.04	10.67	8.54	7.81	7.59	6.60	7.46
	CSR	8.27	7.52	7.07	6.92	10.55	8.41	7.70	7.55	6.69	7.36
	PCA	8.25	7.49	7.03	6.88	10.54	8.37	7.62	8.40	6.70	7.30

Note: In this table, we compare the performance of pre-averaging estimators with subsampling estimators using different thresholding methods under a variety of matrix norms. The tuning parameter θ is for the pre-averaging local window length ($k_n \approx \theta n^{1/(2+\delta)}$), whereas the Δ_n is the subsampling frequency in seconds. The “Noiseless” column provides the estimates using clean and synchronous data with a sampling frequency at $\Delta_n^* = 300$ (5-min), and the “Mixed” column provides the estimates using the pre-averaging approach on the subsampled data ($\theta = 0.08, \Delta_n = 900$). Because the low-rank part $\widehat{\beta}\widehat{E}\beta^\top$ is identical across thresholding methods, we only report it in the upper panel. We also report the average estimated number of factors for the PCA approach. The number of Monte Carlo repetitions is 1000.

6. Empirical applications

6.1. Data

We collect data from the TAQ database intraday transaction prices of the constituents of Dow Jones 30 index, S&P 100 index, and S&P 500 index from January 2004 to December 2013. The indices have 42, 152, and 735 stocks, respectively, during this sampling period.

We select stocks that are members of these indices on the last day of each month, and exclude those among them that have no trading activities on one or more trading days of this month, as well as the bottom 5% stocks in terms of the number of observations for liquidity concerns. To clean the data, we adopt the procedure detailed in [Da and Xiu \(2017\)](#), which only relies on the condition codes from the exchanges and the range of NBBO quotes to identify outliers. We exclude overnight returns to avoid dividend issuances and stock splits. Days with half trading hours are also excluded. We do not, however, remove jumps from these intraday returns as they do not seem to matter for our purpose. We sample the stocks using the refresh time approach, as well as the previous tick method at a 15-min frequency. We select 15 min because the Hausman

Table 2

Pre-averaging vs. subsampling using Soft, SCAD, and AL thresholding.

Estimator		Pre-Averaging				Subsampling				Noiseless	Mixed
Tuning Parameters		θ	θ	θ	θ	Δ_n	Δ_n	Δ_n	Δ_n	Δ_n^*	θ, Δ_n
		0.04	0.06	0.08	0.10	300	900	1,800	3,900	300	0.08, 900
Soft Thresholding											
$\ \hat{\Sigma} - \Sigma\ _{\text{MAX}}$	TSR	0.06	0.06	0.07	0.07	0.12	0.09	0.10	0.13	0.05	0.09
	CSR	0.06	0.05	0.05	0.06	0.11	0.07	0.07	0.09	0.04	0.07
	PCA	0.06	0.06	0.06	0.07	0.12	0.09	0.10	0.15	0.05	0.09
$\ \hat{\Gamma} - \Gamma\ _{\text{MAX}}$	TSR	0.05	0.05	0.05	0.05	0.12	0.06	0.06	0.08	0.04	0.06
	CSR	0.05	0.05	0.05	0.05	0.11	0.06	0.06	0.08	0.05	0.06
	PCA	0.05	0.05	0.05	0.05	0.11	0.06	0.06	0.17	0.05	0.06
$\ \hat{\Sigma} - \Sigma\ _{\Sigma}$	TSR	0.51	0.40	0.36	0.35	1.05	0.55	0.50	0.60	0.25	0.45
	CSR	0.49	0.38	0.34	0.32	1.01	0.51	0.44	0.45	0.25	0.40
	PCA	0.50	0.40	0.35	0.35	1.02	0.53	0.49	0.74	0.25	0.45
$\ (\hat{\Sigma})^{-1} - \Sigma^{-1}\ $	TSR	7.28	6.26	5.74	5.61	10.39	7.76	7.14	7.79	5.43	6.62
	CSR	7.27	6.26	5.76	5.64	10.36	7.76	7.18	7.99	5.54	6.67
	PCA	7.25	7.20	7.76	8.11	10.35	7.73	8.12	12.54	8.24	8.42
$\ (\hat{\Gamma})^{-1} - \Gamma^{-1}\ $	TSR	7.31	6.28	5.76	5.63	10.42	7.78	7.16	7.82	5.68	6.64
	CSR	7.29	6.28	5.78	5.67	10.39	7.78	7.20	8.03	5.76	6.69
	PCA	7.28	6.26	5.75	5.63	10.38	7.75	7.17	12.05	5.77	6.68
SCAD Thresholding											
$\ \hat{\Sigma} - \Sigma\ _{\text{MAX}}$	TSR	0.06	0.06	0.07	0.07	0.12	0.09	0.10	0.13	0.05	0.09
	CSR	0.06	0.05	0.05	0.06	0.11	0.07	0.07	0.09	0.04	0.07
	PCA	0.06	0.06	0.06	0.07	0.12	0.09	0.10	0.15	0.05	0.09
$\ \hat{\Gamma} - \Gamma\ _{\text{MAX}}$	TSR	0.05	0.05	0.05	0.05	0.12	0.06	0.06	0.08	0.04	0.06
	CSR	0.05	0.05	0.05	0.05	0.11	0.06	0.06	0.08	0.05	0.06
	PCA	0.05	0.05	0.05	0.05	0.11	0.06	0.06	0.17	0.05	0.06
$\ \hat{\Sigma} - \Sigma\ _{\Sigma}$	TSR	0.51	0.40	0.36	0.36	1.05	0.55	0.52	0.61	0.26	0.47
	CSR	0.49	0.39	0.34	0.33	1.01	0.52	0.45	0.45	0.25	0.41
	PCA	0.50	0.40	0.36	0.35	1.02	0.54	0.51	0.74	0.26	0.46
$\ (\hat{\Sigma})^{-1} - \Sigma^{-1}\ $	TSR	7.42	6.51	6.10	6.35	10.39	7.99	12.21	9.69	5.44	11.96
	CSR	7.37	6.53	6.14	6.22	10.36	7.94	10.78	9.96	5.55	10.75
	PCA	7.35	7.26	7.92	8.43	10.35	7.91	11.04	12.96	8.39	11.40
$\ (\hat{\Gamma})^{-1} - \Gamma^{-1}\ $	TSR	7.44	6.53	6.12	6.38	10.42	8.01	12.36	9.78	5.68	12.10
	CSR	7.39	6.55	6.17	6.26	10.39	7.97	10.91	10.05	5.76	10.88
	PCA	7.37	6.53	6.14	6.25	10.38	7.94	10.68	12.60	5.77	10.97
AL Thresholding											
$\ \hat{\Sigma} - \Sigma\ _{\text{MAX}}$	TSR	0.06	0.06	0.06	0.07	0.12	0.09	0.10	0.13	0.05	0.09
	CSR	0.05	0.05	0.05	0.05	0.11	0.07	0.07	0.09	0.04	0.07
	PCA	0.06	0.06	0.06	0.07	0.12	0.09	0.10	0.16	0.05	0.09
$\ \hat{\Gamma} - \Gamma\ _{\text{MAX}}$	TSR	0.05	0.04	0.04	0.05	0.12	0.06	0.06	0.08	0.04	0.06
	CSR	0.05	0.05	0.05	0.05	0.11	0.06	0.06	0.08	0.04	0.06
	PCA	0.05	0.05	0.05	0.05	0.11	0.06	0.06	0.17	0.04	0.06
$\ \hat{\Sigma} - \Sigma\ _{\Sigma}$	TSR	0.48	0.38	0.35	0.36	1.03	0.55	0.55	0.67	0.23	0.50
	CSR	0.47	0.37	0.34	0.34	1.00	0.52	0.50	0.53	0.23	0.45
	PCA	0.48	0.38	0.35	0.36	1.00	0.54	0.55	0.75	0.24	0.50
$\ (\hat{\Sigma})^{-1} - \Sigma^{-1}\ $	TSR	6.70	5.65	5.29	6.73	10.26	7.46	16.46	41.68	5.34	14.24
	CSR	6.76	5.72	5.37	6.67	10.30	7.53	17.40	40.75	5.46	15.06
	PCA	6.94	7.94	8.82	9.55	10.29	7.71	18.10	25.70	9.12	16.06
$\ (\hat{\Gamma})^{-1} - \Gamma^{-1}\ $	TSR	6.72	5.66	5.32	6.79	10.29	7.48	16.70	42.81	5.59	14.42
	CSR	6.78	5.75	5.43	6.78	10.32	7.56	17.72	41.92	5.71	15.32
	PCA	6.77	5.73	5.44	6.89	10.32	7.54	17.95	25.58	5.71	15.87

Note: This table is a continuation of Table 1, where we report the simulation results using different thresholding methods. All other settings remain the same.

tests proposed in Aït-Sahalia and Xiu (2016) suggest it is a safe frequency at which to use realized covariance estimators for this pool of stocks.

Fig. 1 plots the daily sample sizes after refresh time sampling for S&P 500, S&P 100, and Dow Jones 30 index constituents. In addition, it presents the quartiles of the daily sample sizes for S&P 500 index constituents by different shading. As we increase the number of assets, the daily refresh time observations decrease rapidly. Still, we are able to obtain on average 284 observations per day for S&P 500, which is approximately equivalent to sampling every 90 s. The average number of observations for S&P 100 and Dow Jones 30 constituents are 905 and 2105, respectively.

Table 3

Impact of the selected number of factors on the PCA method.

		Location Thresholding											
\hat{r}		1	2	3	4	5	6	8	10	15	20	30	50
$\ \hat{\Sigma} - \Sigma\ _{\text{MAX}}$	Pre-Averaging	0.20	0.14	0.06	0.06	0.06	0.06	0.06	0.06	0.06	0.06	0.06	0.06
	Subsampling	0.20	0.15	0.08	0.08	0.08	0.08	0.08	0.08	0.08	0.08	0.08	0.08
	Noiseless	0.20	0.14	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04
$\ \beta\hat{E}\beta^T - \beta E\beta^T\ _{\text{MAX}}$	Pre-Averaging	0.24	0.16	0.05	0.08	0.09	0.10	0.10	0.10	0.11	0.11	0.12	0.14
	Subsampling	0.24	0.16	0.06	0.09	0.10	0.10	0.11	0.11	0.11	0.12	0.13	0.16
	Noiseless	0.24	0.16	0.04	0.08	0.08	0.09	0.09	0.09	0.09	0.10	0.10	0.12
$\ \hat{\Gamma} - \Gamma\ _{\text{MAX}}$	Pre-Averaging	0.25	0.17	0.03	0.07	0.07	0.08	0.09	0.09	0.09	0.09	0.09	0.11
	Subsampling	0.27	0.19	0.06	0.07	0.08	0.08	0.09	0.09	0.09	0.09	0.09	0.10
	Noiseless	0.24	0.16	0.02	0.07	0.07	0.08	0.09	0.09	0.09	0.09	0.09	0.11
$\ \hat{\Sigma} - \Sigma\ _{\Sigma}$	Pre-Averaging	1.11	0.66	0.24	0.25	0.26	0.27	0.29	0.31	0.36	0.39	0.55	0.75
	Subsampling	1.19	0.76	0.40	0.41	0.43	0.44	0.46	0.48	0.54	0.59	0.78	1.03
	Noiseless	1.07	0.62	0.15	0.15	0.16	0.16	0.18	0.19	0.21	0.23	0.34	0.49
$\ (\hat{\Sigma})^{-1} - \Sigma^{-1}\ \times 10^{-2}$	Pre-Averaging	0.10	0.09	0.04	0.04	0.04	0.05	0.05	0.05	0.05	0.07	0.07	0.09
	Subsampling	0.10	0.09	0.07	0.06	0.06	0.06	0.06	0.06	0.06	0.07	0.08	0.09
	Noiseless	0.10	0.09	0.04	0.04	0.04	0.04	0.04	0.04	0.05	0.06	0.06	0.07
$\ (\hat{\Gamma})^{-1} - \Gamma^{-1}\ \times 10^{-2}$	Pre-Averaging	0.10	0.09	0.04	0.08	0.13	0.21	0.36	0.50	0.87	1.35	1.69	2.06
	Subsampling	0.10	0.09	0.07	0.08	0.11	0.17	0.26	0.35	0.59	0.84	1.03	1.26
	Noiseless	0.10	0.09	0.04	0.14	0.23	0.42	0.79	1.09	2.03	3.23	4.36	5.01

Note: We present the errors of the PCA approach for a variety of norms, using different number of factors \hat{r} , with the true value of r being equal to 3. For tuning parameters, we select $\theta = 0.08$ for pre-averaging estimators, $\Delta_n = 900$ for subsampling estimators, and the benchmark no-noise and synchronous case with $\Delta_n^* = 300$. Location thresholding is used in all cases. The number of Monte Carlo repetitions is 1000.

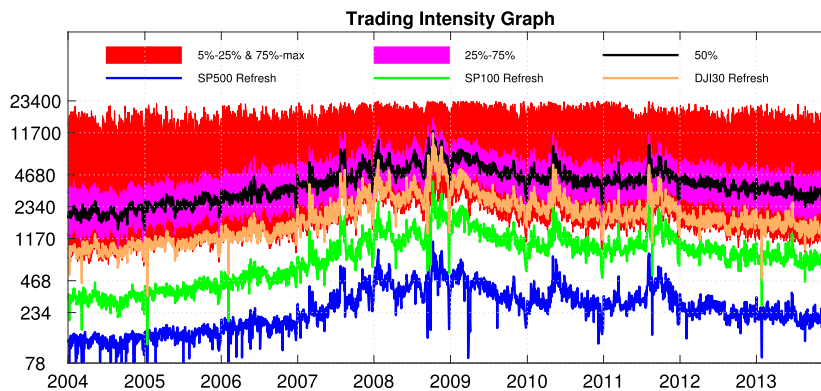


Fig. 1. Trading Intensity of Assets. Note: This figure plots the max, 5%, 25%, 50%, and 75% quantiles of the daily number of observations for the S&P 500 index constituents after cleaning. It also plots the sample size after refresh time sampling for the Dow Jones 30, S&P 100, and S&P 500 index constituents, respectively.

We collect the GICS codes from the Compustat database for the Location thresholding method. The codes have 8 digits. Digits 1–2 indicate the company's sector; digits 3–4 describe the company's industry group; digits 5–6 describe the industry; digits 7–8 describe the sub-industry. Throughout 120 months and among the assets we consider, the time-series median of the largest block size is 80 for sector-based classification, 39 for industry group, 27 for industry, and 15 for sub-industry categories, for S&P 500 index constituents.

We construct observable factors from high-frequency transaction prices at a 15-min frequency. The factors include the market portfolio, the small-minus-big (SMB) portfolio, the high-minus-low (HML) portfolio, the robust-minus-weak (RMW) portfolio, and the conservative-minus-aggressive (CMA) portfolio in the Fama–French 5-factor model. We also include the momentum (MOM) portfolio formed by sorting stock returns between the past 250 days and 21 days. We also collect the 9 industry SPDR ETFs from the TAQ database (Energy (XLE), Materials (XLB), Industrials (XLI), Consumer Discretionary (XLY), Consumer Staples (XLP), Health Care (XLV), Financial (XLF), Information Technology (XLK), and Utilities (XLU)). The time series of cumulative returns of all factors are plotted in Fig. 2.

We obtain monthly factor loadings (exposures) from the MSCI Barra USE3 by Kahn et al. (1998). The loadings we utilize include Earnings Yield, Momentum, Trading Activity, Currency Sensitivity, Earnings Variation, Growth, Volatility, Dividend Yield, Size, Size Nonlinearity, Leverage, and Value. In addition, we construct and add the market exposure for each stock, using the slope coefficient in a time-series regression of its weighted daily returns on the weighted S&P 500 index returns over the trailing 252 trading days. The weights are chosen to have a half life of 63 days, so as to match the method documented

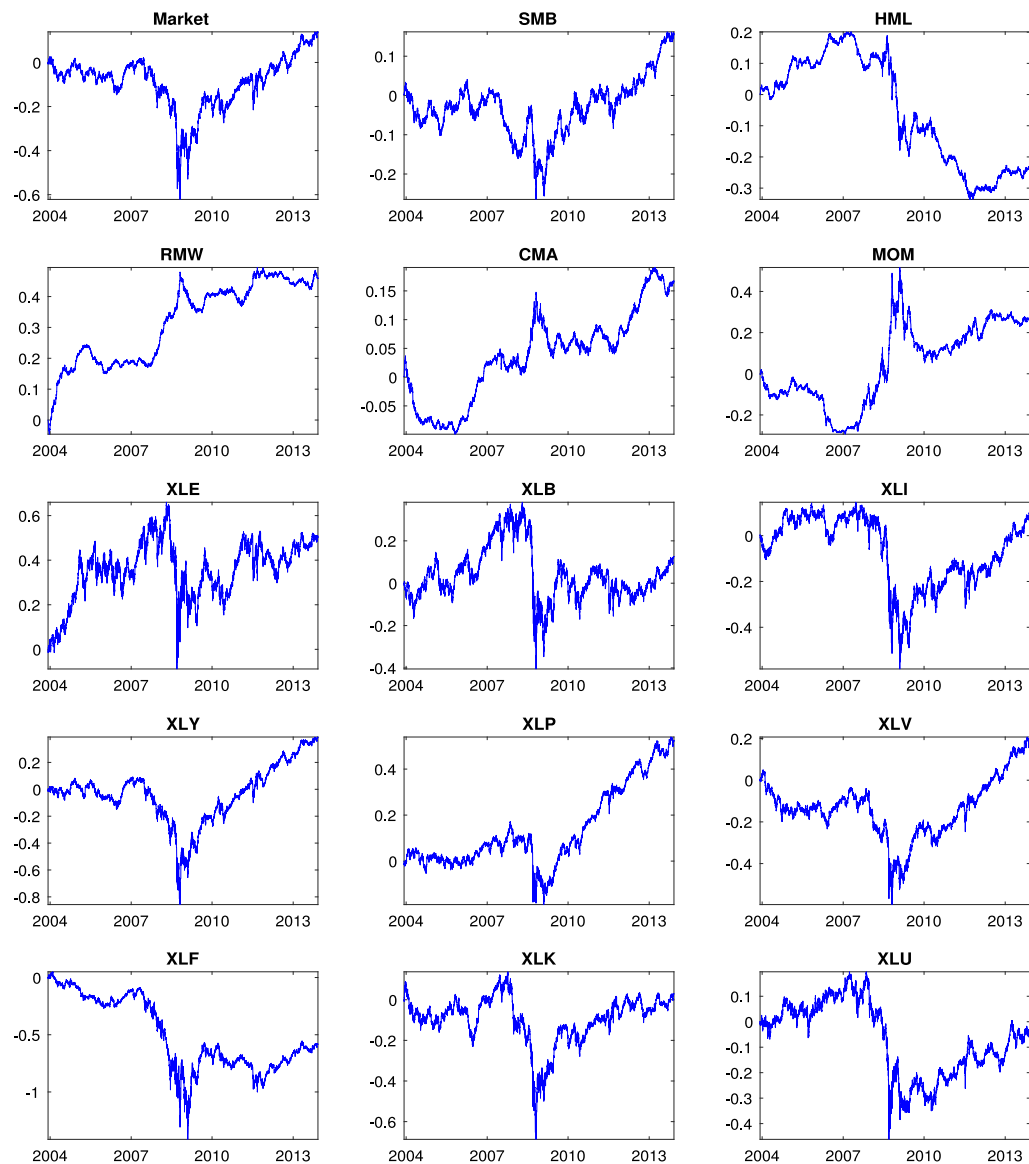


Fig. 2. Time Series of Factors Used in TSR. Note: This figure plots the cumulative returns of the factors we have used in TSR, including the market portfolio, the small-minus-big market capitalization (SMB) portfolio, the high-minus-low price-earning ratio (HML) portfolio, the robust-minus-weak (RMW) portfolio, the conservative-minus-aggressive (CMA) portfolio, the momentum (MOM) portfolio, as well as 9 industry SPDR ETFs (Energy (XLE), Materials (XLB), Industrials (XLI), Consumer Discretionary (XLY), Consumer Staples (XLP), Health Care (XLV), Financial (XLF), Information Technology (XLK), and Utilities (XLU)). The overnight returns are excluded, same for the half trading days.

by USE3. In total, we have 14 observable loadings including the intercept term. We normalize the factor exposures such that their cross-sectional means are 0 and variances are 1 for each month. Although the covariance matrix estimation is invariant under such transformations, the estimated factors now have similar scales. In case of missing exposures, we use their latest available values, or set them to 0 if they are missing throughout the entire sample period for certain stocks.⁵ The cumulative returns of the estimated factors based on S&P 500 constituents are shown in Fig. 3.

We plot the cumulative leading principal components of S&P 500 constituents using our PCA method in Fig. 4. Recognizing a one-to-one correspondence among the factors in Figs. 2–4 is difficult, because the list of characteristics available from the MSCI Barra does not match the observed factors we obtain. Instead, we plot their generalized correlations using 15-min returns in Fig. 5, which measure how correlated two sets of factors are, as suggested by Bai and Ng (2006) and recently

⁵ The empirical findings remain the same if we exclude stocks whose loadings are missing from the USE3 dataset.

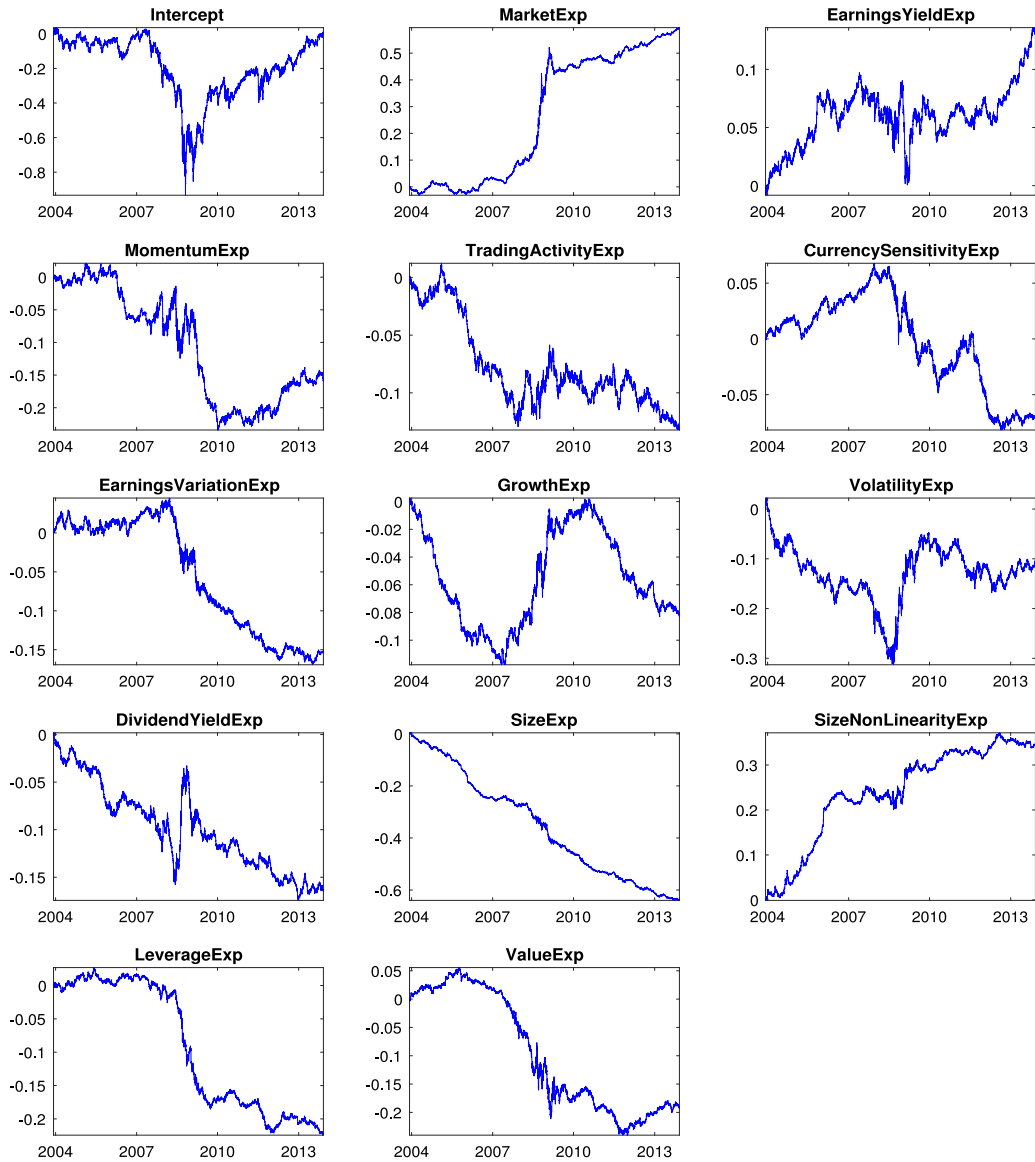


Fig. 3. Time Series of Estimated Factors by CSR. Note: This figure plots the cumulative returns of factors we estimate using CSR, based on S&P 500 constituents. The corresponding factor exposures include the intercept, the market beta, and 12 other variables from MSCI Barra USE3 (Earnings Yield, Momentum, Trading Activity, Currency Sensitivity, Earnings Variation, Growth, Volatility, Dividend Yield, Size, Size Nonlinearity, Leverage, and Value).

employed in [Pelger \(2015b\)](#). Indeed, strong coherence exists among the observed and inferred factors using different approaches, in particular among the PCA and the CSR factors.

6.2. Out-of-Sample portfolio allocation

We then examine the performance of the covariance matrix estimators in a constrained minimum variance portfolio allocation exercise because it requires only the estimated covariance matrix as an input. This approach to evaluating estimators of large covariance matrices is common; see, e.g., [Fan et al. \(2012\)](#). Specifically, we consider the following optimization problem:

$$\min_{\omega} \omega^T \hat{\Sigma} \omega, \text{ subject to } \omega^T \mathbf{1} = 1, \|\omega\|_1 \leq \gamma, \quad (10)$$

where $\|\omega\|_1 \leq \gamma$ imposes an exposure constraint. As explained in [Fan et al. \(2016\)](#), when $\gamma = 1$, all portfolio weights must be non-negative, i.e., no short selling occurs. When γ is small, $\|\hat{\Sigma} - \Sigma\|_{\text{MAX}}$ dictates the performance of the portfolio risk because the optimal portfolio comprises a relatively small number of stocks. By contrast, when γ is large, the portfolio is close

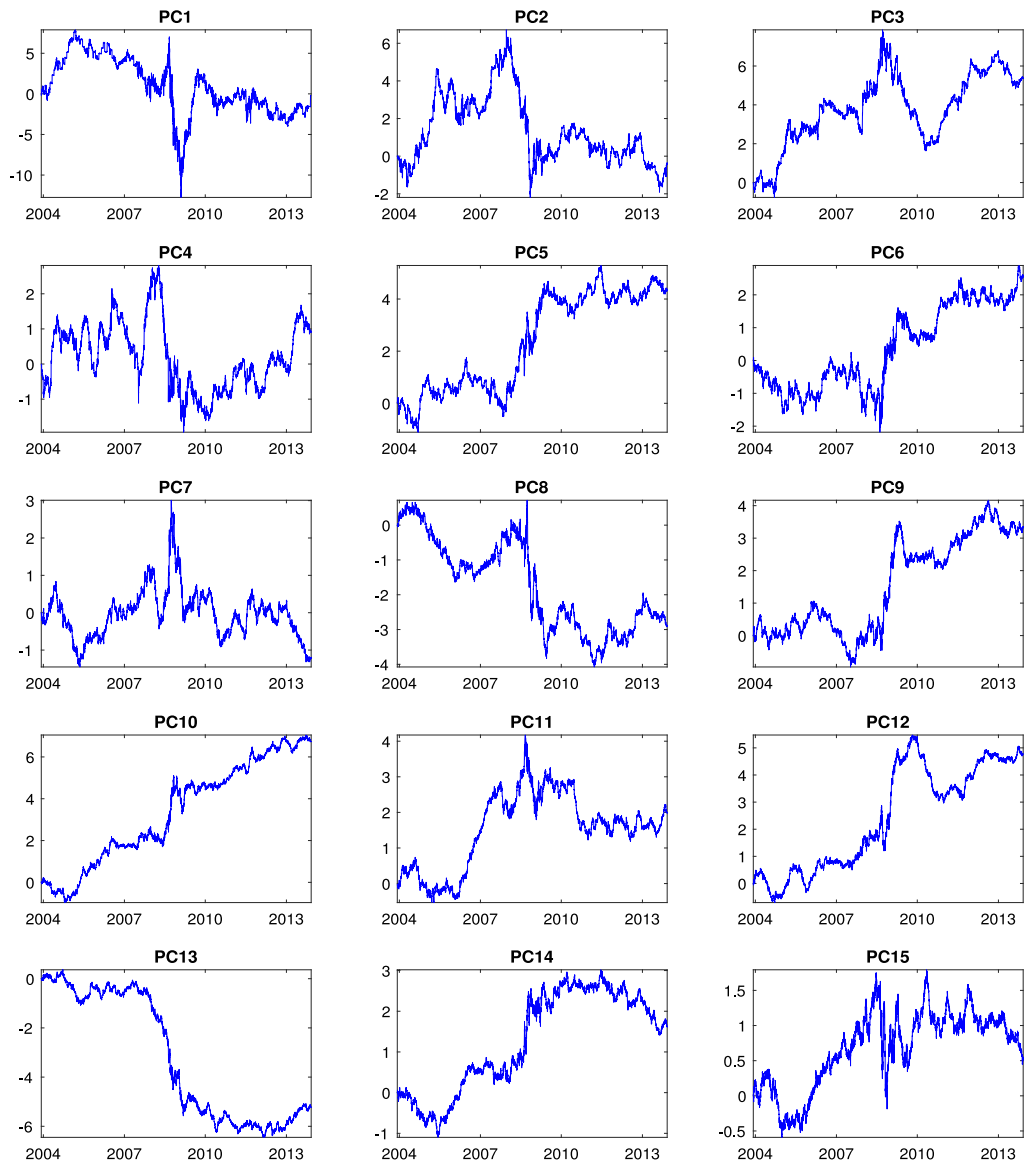


Fig. 4. Time Series of Estimated Factors by PCA. Note: This figure plots the cumulative returns of the leading principal components of S&P 500 constituents we estimate using PCA.

to the global minimum variance portfolio, for which $\|\hat{\Sigma}^{-1} - \Sigma^{-1}\|$ drives the performance of the portfolio risk. Therefore, investigating the out-of-sample risk of the portfolios in (10) for a variety of exposure constraints is informative about the quality of the covariance matrix estimation.

To focus on the evaluation of covariance matrix estimators, we intentionally adopt the simplest random walk forecasting model, i.e., $\hat{\Sigma}_t \approx \mathbb{E}_t(\Sigma_{t+1})$, so that the estimated realized covariance matrices using data of the previous month are directly used for the portfolio construction the next month. For a range of exposure constraints, we measure the out-of-sample portfolio risk using 15-min realized volatility. We compare the covariance matrices based on pre-averaging and subsampling methods, with many choices of θ s and subsampling frequencies Δ_n s. Figs. 6–8 provide the results for the best choice of $\theta = 0.08$ and $\Delta_n = 900$ in simulations and the five thresholding methods we consider.

Fig. 6 shows that (i) for S&P 500 constituents, the Location thresholding (black) performs the best among all thresholding methods, followed by Soft (blue), AL (red), and SCAD (green) approaches, with Hard thresholding (yellow) being the worst. (ii) The TSR approach appears to be the best, with the lowest out-of-sample risk and most stable performance across different thresholding methods. PCA is almost the same as TSR when Location thresholding is applied, but its performance deteriorates if we apply other thresholding techniques. CSR is dominated by the other two by a very large margin. This differs from our simulation results, indicating CSR suffers from more serious model misspecification. (iii) When the exposure constraint γ is

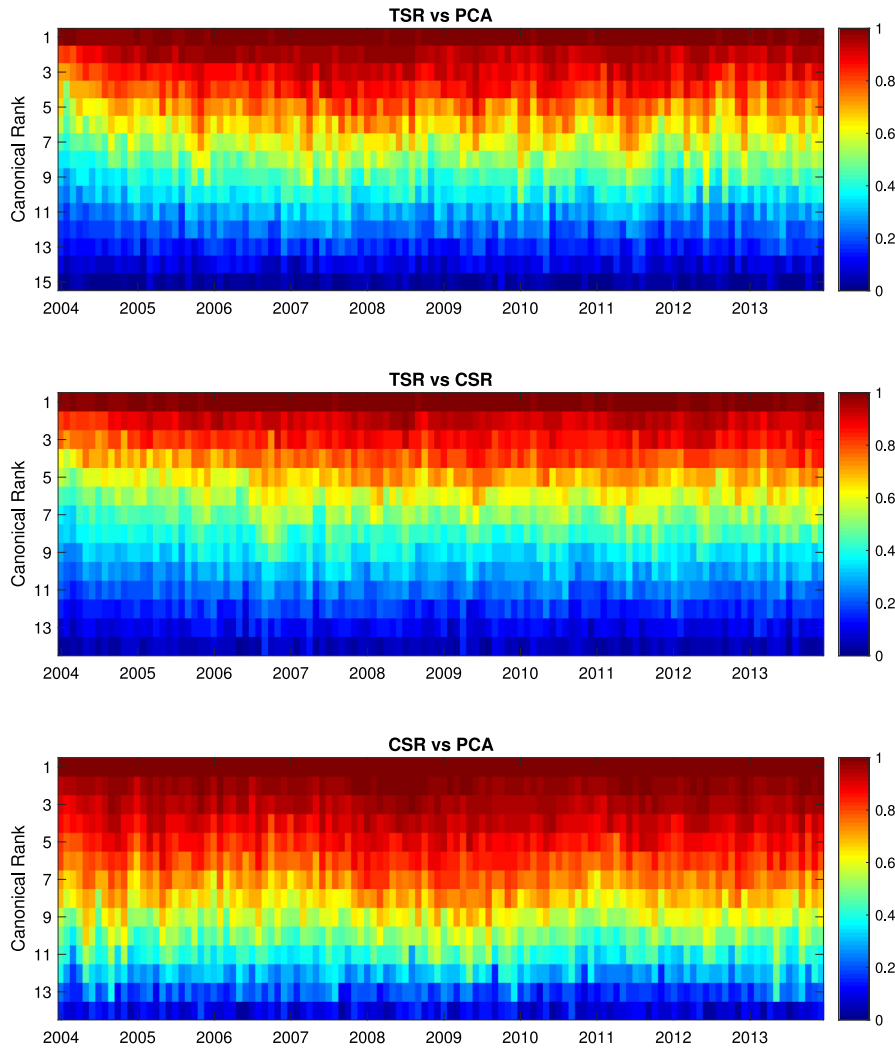


Fig. 5. Generalized Correlation Plot. Note: This figure provides the heatmap of the monthly generalized (canonical) correlations among the time series of factors used in TSR and estimated from CSR and PCA methods. The y-axis is the rank of the generalized correlation. For two sets of factors X_a and X_b , the generalized correlation of rank k is calculated as the square-root of the k th-largest eigenvalue of the matrix $[X_a, X_a]^{-1}[X_a, X_b][X_b, X_b]^{-1}[X_b, X_a]$, where $[\cdot, \cdot]$ denotes the quadratic covariation.

small, the performance gap among different thresholding methods is small. This result is expected because the $\|\cdot\|_{\text{MAX}}$ norm differences across all methods are similar, and in this case, the portfolios are so heavily constrained that they are effectively low-dimensional. (iv) The pre-averaging estimators (solid lines) dominate the subsampling estimators (dash-dotted lines) across almost all cases, which agrees with our simulation results.

For S&P 100, we observe similar patterns from Fig. 7, namely, that pre-averaging estimators outperform the subsampling estimators. PCA performs slightly worse than TSR. With respect to the Dow Jones 30, Fig. 8 shows the CSR and PCA perform considerably worse than TSR. This finding is not surprising, given that our theory suggests consistency of these two estimators requires a large dimension, whereas for TSR, a smaller cross section works better.

Finally, we report in Table 4 the Diebold–Mariano (Diebold and Mariano, 2002) tests for comparison of the out-of-the-sample risk of portfolios based on the pre-averaging estimators against the subsampling estimators. Negative test statistics favor the pre-averaging approach. Similar to Figs. 6–8, when γ is large, pre-averaging estimators deliver significantly smaller out-of-the-sample risk using TSR and PCA across all thresholding methods for S&P 500 index constituents. For S&P 100 and Dow-Jones index constituents, the difference between pre-averaging and subsampling is significant only if using TSR.

7. Conclusion

Leveraging a variety of factor models, we construct pre-averaging-based large covariance matrix estimators using high-frequency transaction prices, which are robust to the asynchronous arrival of trades and the market microstructure

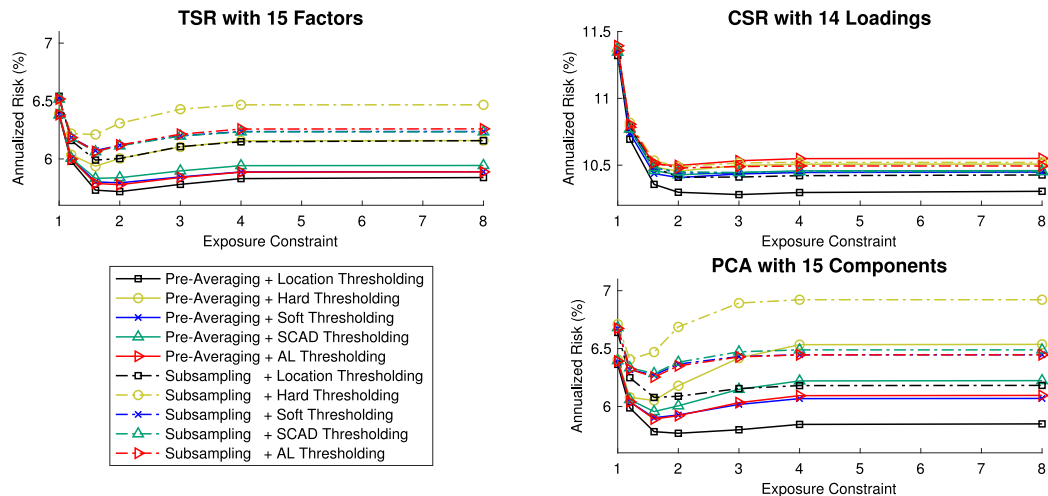


Fig. 6. Out-of-Sample Risk of the S&P 500 Portfolio. Note: This figure compares the time-series average of the out-of-sample monthly volatility from 2004 to 2013 using S&P 500 Index constituents. The x-axis is the exposure constraint γ in the optimization problem (10). The results are based on a combination of methods: (pre-averaging, subsampling) \times (TSR, CSR, PCA) \times (Location, Hard, Soft, SCAD, AL thresholding). We use the GICS codes at the industry group level for the location thresholding method. The pre-averaging estimator uses $\theta = 0.08$, whereas the subsampling estimator uses 15-min returns. The out-of-sample volatility is calculated using 15-min subsampled returns each month. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

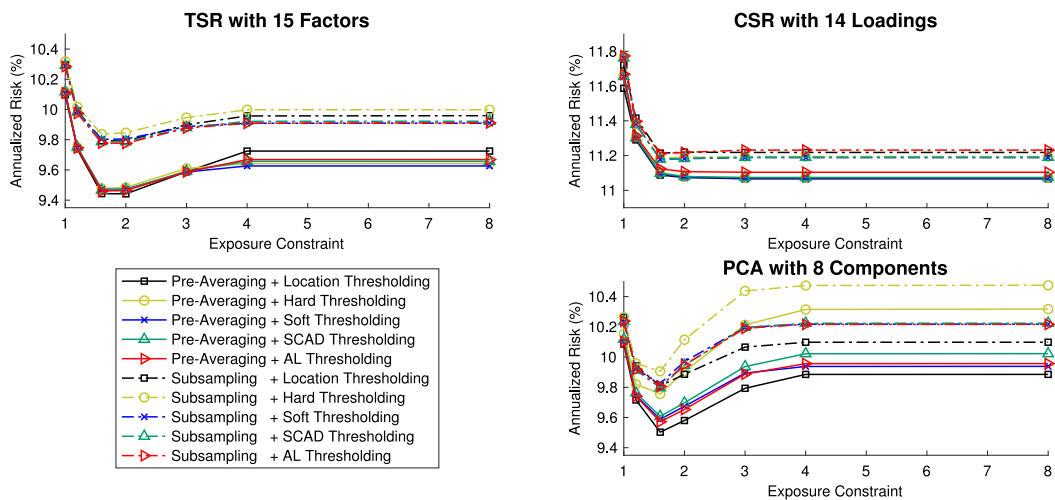


Fig. 7. Out-of-Sample Risk of the S&P 100 Portfolio. Note: This figure compares the time-series average of the out-of-sample monthly volatility from 2004 to 2013 using S&P 100 Index constituents. All other settings remain the same as those in Fig. 6.

noise. We compare various estimators based on different combinations of factor model specifications and thresholding methods, in terms of their convergence rates, their finite sample behavior, and their empirical performance in a portfolio allocation horse race. Throughout, we find that pre-averaging plus TSR or PCA with Location thresholding dominates the other combinations, in particular the subsampling method. Also, CSR, the method MSCI Barra adopts for low-frequency data, performs considerably worse in almost all scenarios we study. This bad performance is perhaps driven by model misspecification, which can be alleviated with a potentially better set of factor exposures.

Acknowledgments

We thank the co-editor Oliver Linton, the associate editor, and two anonymous referees for their useful comments. We also benefit from useful discussions with Rui Da, Markus Pelger, and seminar participants at the 2017 Joint Statistical Meetings. Dacheng Xiu acknowledges support from the IBM Faculty Scholar Fund at the University of Chicago Booth School of Business.

Table 4

Diebold–Mariano tests for the out-of-sample risk comparison.

Exposure Constraint γ		1	1.2	1.6	2	3	4	8
S&P 500								
Location	TSR	−1.51	−1.84*	−2.25**	−2.74***	−3.04***	−3.01***	−3.01***
	CSR	0.01	−0.39	−1.36	−1.43	−1.74*	−1.60	−1.56
	PCA	−2.20**	−2.44**	−2.53**	−2.69***	−2.85***	−2.69***	−2.66***
Hard	TSR	−1.43	−2.01**	−2.46**	−2.64***	−2.60***	−2.46**	−2.44**
	CSR	0.59	0.56	0.35	0.76	1.23	1.31	1.33
	PCA	−2.29**	−2.81***	−3.18***	−3.37***	−3.44***	−3.23***	−3.22***
Soft	TSR	−1.52	−2.09**	−2.79***	−3.28***	−3.44***	−3.42***	−3.40***
	CSR	0.49	0.35	0.08	0.43	0.78	0.85	0.89
	PCA	−2.21**	−2.65***	−3.20***	−3.42***	−3.59***	−3.50***	−3.48***
SCAD	TSR	−1.53	−2.04**	−2.57**	−2.98***	−3.08***	−2.98***	−2.97***
	CSR	0.56	0.50	0.31	0.51	0.99	1.09	1.11
	PCA	−2.21**	−2.65***	−3.09***	−3.32***	−3.23***	−2.91***	−2.88***
AL	TSR	−1.54	−2.15**	−2.84***	−3.33***	−3.53***	−3.46***	−3.44***
	CSR	0.61	0.45	0.47	0.80	1.30	1.43	1.44
	PCA	−2.18**	−2.60***	−3.10***	−3.27***	−3.32***	−3.10***	−3.09***
S&P 100								
Location	TSR	−1.57	−2.09**	−2.34**	−2.46**	−1.86*	−1.51	−1.53
	CSR	−2.43**	−2.79***	−3.23***	−3.19***	−3.30***	−3.30***	−3.30***
	PCA	−1.55	−1.93*	−1.58	−1.40	−0.83	−0.64	−0.64
Hard	TSR	−1.69*	−2.25**	−2.47**	−2.43**	−2.03**	−2.05**	−2.05**
	CSR	−1.87*	−1.89*	−1.67*	−1.87*	−1.93*	−1.93*	−1.93*
	PCA	−1.38	−1.64	−0.93	−1.06	−0.57	−0.31	−0.30
Soft	TSR	−1.67*	−2.31**	−2.63***	−2.59***	−2.12**	−2.05**	−2.05**
	CSR	−2.17**	−2.50**	−2.21*	−2.32*	−2.40**	−2.40**	−2.40**
	PCA	−1.70*	−2.16**	−1.53	−1.62	−1.22	−1.16	−1.16
SCAD	TSR	−1.69*	−2.34**	−2.63***	−2.56**	−2.02**	−1.93*	−1.93*
	CSR	−2.15**	−2.41**	−2.14**	−2.24**	−2.34**	−2.34**	−2.34**
	PCA	−1.60	−2.03**	−1.22	−1.32	−0.89	−0.72	−0.72
AL	TSR	−1.68*	−2.30**	−2.61***	−2.61***	−2.28**	−2.00**	−2.01**
	CSR	−2.17**	−2.39**	−2.11**	−2.25**	−2.35**	−2.35**	−2.35**
	PCA	−1.70*	−2.11**	−1.48	−1.48	−1.02	−0.91	−0.91
Dow Jones 30								
Location	TSR	−1.83*	−2.22**	−2.70***	−2.64***	−2.55**	−2.55**	−2.55**
	CSR	−0.04	0.00	0.15	0.14	0.20	0.20	0.20
	PCA	0.71	0.38	−0.07	−0.69	−0.75	−0.75	−0.75
Hard	TSR	−1.74*	−1.91*	−2.65***	−2.50**	−2.44**	−2.44**	−2.44**
	CSR	−0.69	−0.76	−0.68	−0.60	−0.49	−0.49	−0.49
	PCA	0.63	0.23	−0.50	−1.06	−1.00	−0.97	−0.97
Soft	TSR	−0.61	−1.09	−2.39**	−2.28**	−2.21**	−2.21**	−2.21**
	CSR	−0.58	−0.76	−1.02	−0.97	−0.84	−0.84	−0.84
	PCA	0.62	0.23	−0.44	−0.97	−1.10	−1.10	−1.10
SCAD	TSR	−0.58	−1.07	−2.39**	−2.28**	−2.21**	−2.21**	−2.21**
	CSR	−0.58	−0.76	−1.02	−0.98	−0.85	−0.85	−0.85
	PCA	0.62	0.22	−0.46	−0.96	−1.10	−1.10	−1.10
AL	TSR	−0.06	−0.62	−2.08**	−2.01**	−1.98**	−1.98**	−1.98**
	CSR	−0.51	−0.65	−1.06	−1.01	−0.88	−0.88	−0.88
	PCA	0.64	0.25	−0.43	−0.85	−1.02	−1.02	−1.02

Note: This table reports the Diebold–Mariano test statistics against the null that the out-of-sample risk of the portfolios based on the pre-averaging method is equal to that based on the subsampling approach. Negative numbers favor the pre-averaging approach.

*to reveal the significance at the 10% level.

**to reveal the significance at the 5% level.

***to reveal the significance at the 1% level.

Appendix. Mathematical proofs

A.1. Proof of Theorem 1

We need a few lemmas.

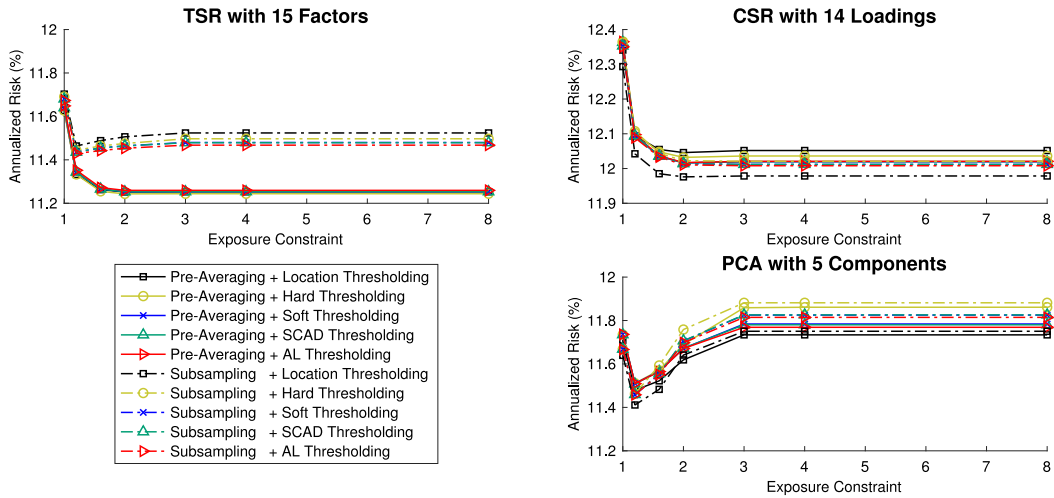


Fig. 8. Out-of-Sample Risk of the Dow Jones 30 Portfolio. Note: This figure compares the time-series average of the out-of-sample monthly volatility from 2004 to 2013 using Dow Jones 30 Index constituents. All other settings remain the same as those in Fig. 6.

Lemma 1. Suppose that $n_\delta^{-1/2} \sqrt{\log d} = o(1)$. Under Assumptions 1–5, and for some constants C_0, C_1 and C_2 , we have

$$(i) \quad \mathbb{P} \left(\left\| \hat{\mathbf{E}} - \mathbf{E} \right\|_{\text{MAX}} \geq C_0 n_\delta^{-1/2} \sqrt{\log d} \right) = O(r^2 C_1 d^{-C_0^2 C_2}), \quad (\text{A.1})$$

$$(ii) \quad \mathbb{P} \left(\left\| \hat{\mathbf{E}} - \mathbf{E} \right\|_{\text{F}} \geq C_0 m_\delta^{-1/2} \sqrt{\log d} \right) = O(r^2 C_1 d^{-C_0^2 C_2}), \quad (\text{A.2})$$

$$(iii) \quad \mathbb{P} \left(\max_{\substack{1 \leq k \leq r \\ 1 \leq l \leq d}} \left| \frac{n}{n - k_n + 2} \frac{1}{\psi_2 k_n} \sum_{i=0}^{n-k_n+1} \bar{X}_{k,i}^* \bar{Z}_{o,li}^* \right| \geq C_0 n_\delta^{-1/2} \sqrt{\log d} \right) = O(r C_1 d^{-C_0^2 C_2 + 1}), \quad (\text{A.3})$$

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

$$(v) \quad \mathbb{P} \left(\max_{1 \leq j \leq d} \left\| \hat{\beta}_j - \beta_j \right\| \geq C_0 n_\delta^{-1/2} \sqrt{\log d} \right) = O(C_1 d^{-C_0^2 C_2 + 1}), \quad (\text{A.5})$$

$$(vi) \quad \mathbb{P} \left(\left\| \hat{\beta} - \beta \right\|_{\text{F}} \geq C_0 d^{1/2} n_\delta^{-1/2} \sqrt{\log d} \right) = O(C_1 d^{-C_0^2 C_2 + 1}), \quad (\text{A.6})$$

$$(vii) \quad \mathbb{P} \left(\left\| \hat{\beta} - \beta \right\|_{\text{MAX}} \geq C_0 n_\delta^{-1/2} \sqrt{\log d} \right) = O(C_1 d^{-C_0^2 C_2 + 1}), \quad (\text{A.7})$$

$$(viii) \quad \mathbb{P} \left(\max_{1 \leq k \leq d} \frac{n}{n - k_n + 2} \frac{1}{\psi_2 k_n} \sum_{i=0}^{n-k_n+1} ((\beta - \hat{\beta})^\top \bar{X}_i^*)_k \geq C_0 r^2 n_\delta^{-1} \log d \right) = O(C_1 d^{-C_0^2 C_2 + 1}), \quad (\text{A.8})$$

$$(ix) \quad \mathbb{P} \left(\max_{1 \leq k, l \leq d} |\hat{\Gamma}_{kl} - \Gamma_{kl}| \geq C_0 n_\delta^{-1/2} \sqrt{\log d} \right) = O(C_1 d^{-C_0^2 C_2 + 2}), \quad (\text{A.9})$$

$$(x) \quad \mathbb{P} \left(\max_{1 \leq k, l \leq d} |\hat{\Gamma}_{kl}^S - \Gamma_{kl}| \geq C_0 n_\delta^{-1/2} \sqrt{\log d} \right) = O(C_1 d^{-C_0^2 C_2 + 2}). \quad (\text{A.10})$$

Proof of Lemma 1. (i) Note that we have

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

therefore,

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

For T_1 , this expression can be furthered decomposed as:

$$T_1 = \frac{n}{n - k_n + 2} \frac{1}{\psi_2 k_n} \left\{ \sum_{i=1}^{n-k_n+1} a_{1,i}(k, l)(X_{k,t_i^k} - X_{k,t_{i-1}^k})(X_{l,t_i^l} - X_{l,t_{i-1}^l}) \right. \\ \left. + \sum_{(i,j) \in \mathcal{F}} b_{1,ij}(k, l)(X_{k,t_i^k} - X_{k,t_{i-1}^k})(X_{l,t_i^l} - X_{l,t_{i-1}^l}) \right\},$$

for certain numbers $a_{1,i}(k, l)$ and $b_{1,ij}(k, l)$ such that

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

The set \mathcal{F} is given by

$$\mathcal{F} = \{(i, j) | 1 \leq i \leq n - k_n + 1, 1 \leq j \leq n - k_n + 1, |i - j| \leq k_n - 1, i \neq j\}.$$

Let

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

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

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

Now 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

$$\sum_{i=1}^{n-k_n+1} A_{k,l}^i (X_{k,t_i^k} - X_{k,t_{i-1}^k})(X_{l,t_i^l} - X_{l,t_{i-1}^l}) \\ = \sum_{i=1}^{n-k_n+1} A_{k,l}^i \left\{ (X_{k,t_i} - X_{k,t_{i-1}})(X_{l,t_i} - X_{l,t_{i-1}}) + (X_{k,t_i^k} - X_{k,t_i})(X_{l,t_i^l} - X_{l,t_i}) \right. \\ + (X_{k,t_i} - X_{k,t_{i-1}})(X_{l,t_i} - X_{l,t_{i-1}}) + (X_{k,t_i^k} - X_{k,t_i})(X_{l,t_{i-1}} - X_{l,t_{i-1}^l}) \\ + (X_{k,t_i} - X_{k,t_{i-1}})(X_{l,t_i^l} - X_{l,t_i}) + (X_{k,t_i} - X_{k,t_{i-1}})(X_{l,t_{i-1}} - X_{l,t_{i-1}^l}) \\ + (X_{k,t_{i-1}} - X_{k,t_{i-1}^k})(X_{l,t_i^l} - X_{l,t_i}) + (X_{k,t_{i-1}} - X_{k,t_{i-1}^k})(X_{l,t_i} - X_{l,t_{i-1}}) \\ \left. + (X_{k,t_{i-1}} - X_{k,t_{i-1}^k})(X_{l,t_{i-1}} - X_{l,t_{i-1}^l}) \right\} \\ \equiv \sum_{i=1}^{n-k_n+1} A_{k,l}^i \Delta_i^n X_k \Delta_i^n X_l + H_{kl}^1(1) + \cdots + H_{kl}^1(8).$$

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

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

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

$$(X_{t,k}^* - X_{s,k}^*)(X_{t,l}^* - X_{s,l}^*) - \int_s^t (\sigma \sigma^\top)_{v,kl} dv = \int_s^t (X_{v,k}^* - X_{s,k}^*)(\sigma_v dW_v)_l + \int_s^t (X_{v,l}^* - X_{s,l}^*)(\sigma_v dW_v)_k.$$

Therefore

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

Now we have

$$\left| \sum_{i=1}^{n-k_n+1} A_{k,l}^i (X_{k,t_i^k} - X_{k,t_{i-1}^k})(X_{l,t_i^l} - X_{l,t_{i-1}^l}) - \sum_{i=1}^{n-k_n+1} \int_{(i-1)\Delta_n}^{i\Delta_n} A_{k,l}^i (\sigma \sigma^\top)_{s,kl} ds \right|$$

$$= \left| \sum_{i=1}^{n-k_n+1} A_{k,l}^i \Delta_i^n X_k^* \int_{(i-1)\Delta_n}^{i\Delta_n} b_{s,l} ds + \sum_{i=1}^{n-k_n+1} A_{k,l}^i \Delta_i^n X_l^* \int_{(i-1)\Delta_n}^{i\Delta_n} b_{s,k} ds \right. \\ \left. + \sum_{i=1}^{n-k_n+1} A_{k,l}^i \int_{(i-1)\Delta_n}^{i\Delta_n} b_{s,l} ds \int_{(i-1)\Delta_n}^{i\Delta_n} b_{s,k} ds + M_t \right|.$$

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

The quadratic variation of M_t is given by

$$[M, M]_t = \Delta_n \sum_{i=1}^{n-k_n+1} \int_{(i-1)\Delta_n}^{i\Delta_n} (A_{k,l}^i)^2 \left((X_{s,k}^* - X_{(i-1)\Delta_n,k}^*)^2 \sum_{r=1}^q \sigma_{s,lr}^2 + (X_{s,l}^* - X_{(i-1)\Delta_n,l}^*)^2 \sum_{r=1}^q \sigma_{s,kr}^2 \right. \\ \left. + 2(X_{s,k}^* - X_{(i-1)\Delta_n,k}^*)(X_{s,l}^* - X_{(i-1)\Delta_n,l}^*) \sum_{r=1}^q \sigma_{s,lr} \sigma_{s,kr} \right) ds.$$

According to [Assumption 1](#), here we assume that $\|X_t\|_\infty \leq K$, $\|h_t\|_\infty \leq K$, and $\|\sigma_t \sigma_t^\top\|_{\text{MAX}} \leq K$, for some constant $K > 0$. Therefore by Cauchy–Schwarz inequality, we have

$$[M, M]_t \leq 16K^3 t \Delta_n.$$

Then by the exponential inequality for a continuous martingale, we have

$$\mathbb{P} \left(\sup_{0 \leq s \leq t} \left| \sum_{i=1}^{n-k_n+1} A_{k,l}^i \zeta_{i,kl}'' \right| > u \right) \leq \exp \left(-\frac{nu^2}{32K^3 t} \right). \quad (\text{A.11})$$

In addition, by Cauchy–Schwarz inequality:

$$\mathbb{P} \left(\left| \sum_{i=1}^{n-k_n+1} A_{k,k}^i \Delta_i^n X_k^* \int_{(i-1)\Delta_n}^{i\Delta_n} b_{s,l} ds \right| > u \right) \\ \leq \mathbb{P} \left(\left| \sum_{i=1}^{n-k_n+1} A_{k,k}^i (\Delta_i^n X_k^*)^2 - \sum_{i=1}^{n-k_n+1} \int_{(i-1)\Delta_n}^{i\Delta_n} A_{k,k}^i (\sigma \sigma^\top)_{s,kk} ds \right| > \frac{u^2}{tK \Delta_n} - \sum_{i=1}^{n-k_n+1} \int_{(i-1)\Delta_n}^{i\Delta_n} A_{k,k}^i (\sigma \sigma^\top)_{s,kk} ds \right) \\ \leq \mathbb{P} \left(\left| \sum_{i=1}^{n-k_n+1} A_{k,k}^i (\Delta_i^n X_k^*)^2 - \sum_{i=1}^{n-k_n+1} \int_{(i-1)\Delta_n}^{i\Delta_n} A_{k,k}^i (\sigma \sigma^\top)_{s,kk} ds \right| > \frac{u^2}{tK \Delta_n} - tK \right) \\ \leq \exp \left(-\frac{\left(\frac{u^2}{tK \Delta_n} - tK \right)^2}{32K^3 t \Delta_n} \right),$$

where the last inequality follows from [\(A.11\)](#). Finally, notice that

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

we can derive

$$\mathbb{P} \left(\left| \sum_{i=1}^{n-k_n+1} A_{k,l}^i (X_{k,t_i} - X_{k,t_{i-1}})(X_{l,t_i} - X_{l,t_{i-1}}) - \sum_{i=1}^{n-k_n+1} \int_{(i-1)\Delta_n}^{i\Delta_n} A_{k,l}^i (\sigma \sigma^\top)_{s,kl} ds \right| > 4u \right) \\ \leq \mathbb{P}(|M_t| > u) + 2\mathbb{P} \left(\left| \sum_{i=1}^{n-k_n+1} A_{k,l}^i \Delta_i^n X_k^* \int_{(i-1)\Delta_n}^{i\Delta_n} b_{s,l} ds \right| > u \right) + \mathbf{1}_{\{tK^2 \Delta_n \max_i |A_{k,l}^i| > u\}} \\ \leq \exp \left(-\frac{u^2}{32K^3 t \Delta_n} \right) + 2 \exp \left(-\frac{\left(\frac{u^2}{tK \Delta_n} - tK \right)^2}{32K^3 t \Delta_n} \right) \\ \leq C_1 \exp \left(-\frac{16C_2 u^2}{\Delta_n} \right),$$

where the above inequality holds if $x > (tK^2 \Delta_n \max_i |A_{k,l}^i|) \vee (tK \sqrt{\Delta_n}) \vee (tK \Delta_n / 2\sqrt{1+4/\Delta_n})$, and $C_1 \geq 3$, $C_2 \leq (512K^3 t)^{-1}$. On the other hand, if x violates this bound, i.e., $x \leq C' \sqrt{\Delta_n}$, we can choose C_1 such that $C_1 \exp(-16C_2 C'^2) \geq 1$, so that the

inequality follows trivially. For $H_{kl}^1(1), \dots, H_{kl}^1(8)$, we can use exactly the same technique for proving $\sum_{i=1}^{n-k_n+1} A_{k,l}^i \Delta_i^n X_k \Delta_i^n X_l$, and have that

$$\mathbb{P}(|H_{kl}^1(1) + \dots + H_{kl}^1(8)| \geq u) \leq C_p e^{-C_p n u^2}.$$

Since it is easy to show that

$$\sum_{i=1}^{n-k_n+1} \int_{(i-1)\Delta_n}^{i\Delta_n} A_{k,l}^i(\sigma \sigma^\top)_{s,kl} ds = \int_0^\tau (\sigma \sigma^\top)_{s,kl} ds + o(n_\delta),$$

we have

$$\begin{aligned} & \mathbb{P}\left(\left|\frac{n}{n-k_n+2} \frac{1}{\psi_2 k_n} \sum_{i=1}^{n-k_n+1} a_{1,i}(k, l) A_{k,l}^i(X_{k,t_i^k} - X_{k,t_{i-1}^k})(X_{l,t_i^l} - X_{l,t_{i-1}^l}) - \int_0^\tau (\sigma \sigma^\top)_{s,kl} ds\right| \geq u\right) \\ & \leq \mathbb{P}\left(\left|\sum_{i=1}^{n-k_n+1} A_{k,l}^i(X_{k,t_i} - X_{k,t_{i-1}})(X_{l,t_i} - X_{l,t_{i-1}}) - \sum_{i=1}^{n-k_n+1} \int_{(i-1)\Delta_n}^{i\Delta_n} A_{k,l}^i(\sigma \sigma^\top)_{s,kl} ds\right| > u/2\right) \\ & \quad + \mathbb{P}(|H_{kl}^1(1) + \dots + H_{kl}^1(8)| \geq u/3) + \mathbb{P}\left(\left|\sum_{i=1}^{n-k_n+1} \int_{(i-1)\Delta_n}^{i\Delta_n} A_{k,l}^i(\sigma \sigma^\top)_{s,kl} ds - \int_0^\tau (\sigma \sigma^\top)_{s,kl} ds\right| \geq u/3\right) \\ & \leq C_p e^{-C_p n u^2} + \mathbf{1}_{\{u < o(n_\delta)\}} \\ & \leq C_p e^{-C_p n u^2}. \end{aligned}$$

On the other hand, since we have

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

and

$$(X_{k,t_i^k} - X_{k,t_{i-1}^k})(X_{l,t_i^l} - X_{l,t_{i-1}^l}) = O_p(n^{-1}),$$

$$(X_{k,t_i^k} - X_{k,t_{i-1}^k})(X_{l,t_i^l} - X_{l,t_{i-1}^l}) = O_p(n^{-1}),$$

writing

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

then for $(i, j) \in \mathcal{F}$,

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

Using a similar decomposition, we can obtain

$$\sum_{(i,j) \in \mathcal{F}} X_{1,ij} = \sum_{(i,j) \in \mathcal{F}} B_{k,l}^i \Delta_i^n X_k \Delta_i^n X_l + H_{kl}^2(1) + \dots + H_{kl}^2(8).$$

Since $\|X_t\|_\infty$ is bounded, then $B_{k,l}^i \Delta_i^n X_k \Delta_i^n X_l - \mathbb{E}(B_{k,l}^i \Delta_i^n X_k \Delta_i^n X_l)$ is also bounded. Then according to the Hoeffding's lemma, we obtain $B_{k,l}^i \Delta_i^n X_k \Delta_i^n X_l - \mathbb{E}(B_{k,l}^i \Delta_i^n X_k \Delta_i^n X_l)$ is a sub-Gaussian random variable. Similar arguments can be extended to $H_{kl}^2(1), \dots, H_{kl}^2(8)$. Then according to Hoeffding inequality, and $\sharp F_{kl} \leq C n k_n$, where $\sharp F_{kl}$ denotes the number of elements in the set F_{kl} , then we have

$$\begin{aligned} & \mathbb{P}\left(\left|\frac{1}{n} \sum_{(i,j) \in \mathcal{F}} X_{1,ij}\right| \geq u/8\right) \\ & \leq \mathbb{P}\left(\left|\frac{1}{n} \sum_{(i,j) \in \mathcal{F}} B_{k,l}^i \Delta_i^n X_k \Delta_i^n X_l - \mathbb{E}\left(\frac{1}{n} \sum_{(i,j) \in \mathcal{F}} B_{k,l}^i \Delta_i^n X_k \Delta_i^n X_l\right)\right| \geq u/24\right) \\ & \quad + \mathbb{P}(|H_{kl}^2(1) + \dots + H_{kl}^2(8) - \mathbb{E}(H_{kl}^2(1) + \dots + H_{kl}^2(8))| \geq u/24) + \mathbf{1}_{\left\{\left|\mathbb{E}\left(\frac{1}{n} \sum_{(i,j) \in \mathcal{F}} X_{1,ij}\right)\right| \geq u/24\right\}} \\ & \leq C_p e^{-\frac{C_p (n u)^2}{n k_n}} + \mathbf{1}_{\left\{\left|\mathbb{E}\left(\frac{1}{n} \sum_{(i,j) \in \mathcal{F}} X_{1,ij}\right)\right| \geq u/24\right\}} \\ & \leq C_p e^{-C_p n^{1/2-\delta} u^2}. \end{aligned}$$

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

As for T_4 , it can be decomposed as

$$T_4 = \frac{n}{n - k_n + 2} \frac{1}{\psi_2 k_n} \sum_{i=0}^{n-k_n+1} \bar{u}_{k,t_i^k} \bar{u}_{l,t_i^l} + \bar{v}_{k,t_i^k} \bar{v}_{l,t_i^l} + \bar{u}_{k,t_i^k} \bar{v}_{l,t_i^l} + \bar{v}_{k,t_i^k} \bar{u}_{l,t_i^l} \\ = T_4^1 + T_4^2 + T_4^3 + T_4^4.$$

Using similar techniques of proving Proposition 10 in Kim et al. (2016), and under Assumption 4, we can prove that

$$\max_{1 \leq k, l \leq d} \mathbb{E} |T_4| \leq C_p (n^{-2\delta} + n^{-1/4-2\delta} + n^{-1/2-2\delta}).$$

With respect to T_2 , note that

$$T_2 = \frac{n}{n - k_n + 2} \frac{1}{\psi_2 k_n} \sum_{i=0}^{n-k_n+1} \bar{X}_{k,t_i^k} \bar{u}_{l,t_i^l} + \bar{X}_{k,t_i^k} \bar{v}_{l,t_i^l} = T_2^1 + T_2^2.$$

For T_2^1 , by Jensen's inequality, we have

$$\max_{1 \leq k, l \leq d} \mathbb{E} |T_2^1| \leq C_p \frac{1}{k_n} \sum_{i=0}^{n-k_n+1} \mathbb{E} |\bar{X}_{k,t_i^k}| \mathbb{E} |\bar{u}_{l,t_i^l}| \\ \leq C_p \frac{1}{k_n} \sum_{i=0}^{n-k_n+1} \mathbb{E} |\bar{X}_{k,t_i^k}| k_n^{-1/2} \leq C_p \frac{1}{k_n} n n^{-1} k_n^{-1/2} = k_n^{-3/2}.$$

For T_2^2 , by Jensen's inequality and Holder's inequality, we have

$$\max_{1 \leq k, l \leq d} \mathbb{E} |T_2^2| \leq C_p \frac{1}{k_n} \sum_{i=0}^{n-k_n+1} [\mathbb{E} |\bar{X}_{k,t_i^k}|^2]^{1/2} [\mathbb{E} |\bar{v}_{l,t_i^l}|^2]^{1/2} \leq C_p k_n^{-1} n^{1/4\varsigma}.$$

Therefore,

$$\max_{1 \leq k, l \leq d} \mathbb{E} |T_2| \leq C_p (k_n^{-3/2} + k_n^{-1} n^{1/4\varsigma}).$$

Similarly, we can show that

$$\max_{1 \leq k, l \leq d} \mathbb{E} |T_3| \leq C_p (k_n^{-3/2} + k_n^{-1} n^{1/4\varsigma}).$$

Using Talagrand's concentration inequality and some simple calculation, we have

$$\mathbb{P}(\max_{1 \leq k, l \leq d} |T_2| \geq u) \leq d^2 e^{-C_p(k_n^{-3} + k_n^{-2} n^{1/2\varsigma}) - 2u^2},$$

$$\mathbb{P}(\max_{1 \leq k, l \leq d} |T_3| \geq u) \leq d^2 e^{-C_p(k_n^{-3} + k_n^{-2} n^{1/2\varsigma}) - 2u^2},$$

and

$$\mathbb{P}(\max_{1 \leq k, l \leq d} |T_4| \geq u) \leq d^2 e^{-C_p(n^{-2\delta} + n^{-1/4-2\delta} + n^{-1/2-2\delta}) - 2u^2}.$$

Above all, we prove that

$$\mathbb{P}(|\hat{E}_{kl} - E_{kl}| \geq u) \leq C_1 e^{-C_2[n^{1/2-\delta} + n^{4\delta}]u^2}.$$

Thus

$$\mathbb{P}\left(\|\hat{E} - E\|_{\text{MAX}} \geq C_0 n_\delta^{-1/2} \sqrt{\log d}\right) \\ \leq C_1 r^2 e^{-C_2[n^{1/2-\delta} + n^{4\delta}](C_0 n_\delta^{-1/2} \sqrt{\log d})^2} = C_1 d^{-C_0^2 C_2}.$$

(ii) Since

$$\|\hat{E} - E\|_F \leq r \|\hat{E} - E\|_{\text{MAX}},$$

then

$$\mathbb{P}\left(\|\hat{E} - E\|_F \geq C_0 m_\delta^{-1/2} \sqrt{\log d}\right) \leq C_1 d^{-C_0^2 C_2}.$$

(iii) The derivation of $\bar{X}_{k,i}^* \bar{Z}_{o,li}^*$ is similar to that of $\bar{X}_{k,i}^* \bar{X}_{l,i}^*$ given by (i), and under Assumption 3, we obtain

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

(iv) By the similar argument as in (i), we have

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

(v)-(vii) Moreover, note that

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

therefore, under the event that

$$A = \left\{ \max_{1 \leq i \leq s, 1 \leq j \leq d} \left| \frac{n}{n - k_n + 2} \frac{1}{\psi_2 k_n} \sum_{i=0}^{n-k_n+1} \bar{X}_{k,i}^* \bar{Z}_{o,li}^* \right| \leq C_0 n_\delta^{-1/2} \sqrt{\log d} \right\} \cap \left\{ \lambda_{\min}(\hat{\mathbf{E}}) \geq \frac{1}{2} \lambda_{\min} \left(\int_0^t e_s ds \right) \right\},$$

we have

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

and

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

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

We assume $\lambda_{\min} \left(\int_0^t e_s ds \right)$ is bounded away from 0 and r is finite, so it follows that

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

By Lemma A.1 of Fan et al. (2011), we have

$$\mathbb{P} \left(\lambda_{\min}(\hat{\mathbf{E}}) \geq \frac{1}{2} \lambda_{\min} \left(\int_0^t e_s ds \right) \right) \geq 1 - O(C_1 d^{-C_0^2 C_2}).$$

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

(viii) To prove (A.8), we note that

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

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

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

By (A.5), we obtain

$$\mathbb{P} \left(\max_{1 \leq k \leq d} \frac{n}{n - k_n + 2} \frac{1}{\psi_2 k_n} \sum_{i=0}^{n-k_n+1} \left(\sum_{l=1}^r (\beta_{k,l} - \hat{\beta}_{k,l}) \bar{X}_{l,i}^* \right)^2 > C[n^{-1/2+\delta} + n^{-4\delta}]r^2 \log d \right) \leq O(C_1 d^{-C_0^2 C_2 + 1}).$$

(ix) Finally, under the event of

$$\left\{ \max_{1 \leq l \leq d} \left| \frac{n}{n - k_n + 2} \frac{1}{\psi_2 k_n} \sum_{i=0}^{n-k_n+1} (\bar{Z}_{o,li}^*)^2 - \int_0^t g_{s,ll} ds \right| \leq \frac{1}{4} \max_{1 \leq l \leq d} \int_0^t g_{s,ll} ds \right\}$$

$$\cap \left\{ \max_{1 \leq k \leq d} \frac{n}{n - k_n + 2} \frac{1}{\psi_2 k_n} \sum_{i=0}^{n-k_n+1} \left(\sum_{l=1}^r (\beta_{k,l} - \hat{\beta}_{k,l}) \bar{X}_{l,i}^* \right)^2 \leq C[n^{-1/2+\delta} + n^{-4\delta}]r^2 \log d \right\},$$

according to Cauchy–Schwarz inequality, we have

$$\begin{aligned} & \max_{1 \leq k, l \leq d} \left| \frac{n}{n - k_n + 2} \frac{1}{\psi_2 k_n} \sum_{i=0}^{n-k_n+1} [(\bar{Y}_{k,i}^* - (\hat{\beta} \bar{X}_i^*)_k)(\bar{Y}_{l,i}^* - (\hat{\beta} \bar{X}_i^*)_l) - \bar{Z}_{o,ki}^* \bar{Z}_{o,li}^*] \right| \\ & \leq \max_{1 \leq k, l \leq d} \left| \frac{n}{n - k_n + 2} \frac{1}{\psi_2 k_n} \sum_{i=0}^{n-k_n+1} ((\hat{\beta} - \beta) \bar{X}_i^*)_k ((\hat{\beta} - \beta) \bar{X}_i^*)_l \right| \\ & \quad + 2 \max_{1 \leq l, k \leq d} \left| \frac{n}{n - k_n + 2} \frac{1}{\psi_2 k_n} \sum_{i=0}^{n-k_n+1} \bar{Z}_{o,ki}^* ((\hat{\beta} - \beta) \bar{X}_i^*)_l \right| \\ & \leq \max_{1 \leq k \leq d} \frac{n}{n - k_n + 2} \frac{1}{\psi_2 k_n} \sum_{i=0}^{n-k_n+1} ((\hat{\beta} - \beta) \bar{X}_i^*)_k^2 \\ & \quad + 2 \sqrt{\max_{1 \leq k \leq d} \frac{n}{n - k_n + 2} \frac{1}{\psi_2 k_n} \sum_{i=0}^{n-k_n+1} (\bar{Z}_{o,ki}^*)^2 \max_{1 \leq k \leq d} \frac{n}{n - k_n + 2} \frac{1}{\psi_2 k_n} \sum_{i=0}^{n-k_n+1} ((\hat{\beta} - \beta) \bar{X}_i^*)_k^2} \\ & \leq C_0[n^{-1/2+\delta} + n^{-4\delta}]r^2 \log d + 2\sqrt{\left(\frac{5}{4}Ct\right)}(C_0[n^{-1/2+\delta} + n^{-4\delta}]r^2 \log d) \\ & \leq C'_0 n_\delta^{-1/2} r \sqrt{\log d}. \end{aligned}$$

Consequently, we have

$$\begin{aligned} & \max_{1 \leq k, l \leq d} \left| \frac{n}{n - k_n + 2} \frac{1}{\psi_2 k_n} \sum_{i=0}^{n-k_n+1} [(\bar{Y}_{k,i}^* - (\hat{\beta} \bar{X}_i^*)_k)(\bar{Y}_{l,i}^* - (\hat{\beta} \bar{X}_i^*)_l) - \bar{Z}_{o,ki}^* \bar{Z}_{o,li}^*] \right| \\ & \geq C'_0 n_\delta^{-1/2} r \sqrt{\log d}, \end{aligned}$$

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

$$\begin{aligned} & \max_{1 \geq l, k \geq d} |\hat{\Gamma}_{kl} - \Gamma_{kl}| \\ & \geq \max_{1 \geq k \geq r, 1 \geq l \geq d} \left| \frac{n}{n - k_n + 2} \frac{1}{\psi_2 k_n} \sum_{i=0}^{n-k_n+1} \bar{Z}_{k,i}^* \bar{Z}_{l,i}^* - \int_0^t g_{s,lk} ds \right| \\ & \quad + \max_{1 \geq k, l \geq d} \left| \frac{n}{n - k_n + 2} \frac{1}{\psi_2 k_n} \sum_{i=0}^{n-k_n+1} [(\bar{Y}_{k,i}^* - (\hat{\beta} \bar{X}_i^*)_k)(\bar{Y}_{l,i}^* - (\hat{\beta} \bar{X}_i^*)_l) - \bar{Z}_{o,ki}^* \bar{Z}_{o,li}^*] \right|, \end{aligned}$$

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

(x) Under the event of

$$B_1 = \left\{ \max_{k,l} |\hat{\Gamma}_{kl} - \Gamma| \geq C n_\delta^{-1/2} \sqrt{\log d} \right\},$$

and

$$B_2 = \left\{ C_1 < \sqrt{\hat{\Gamma}_{kk} \hat{\Gamma}_{ll}} < C_2, \quad \text{for all } k \geq d, l \geq d \right\},$$

here C_1 and C_2 are some constant, and $\omega_n = n_\delta^{-1/2} \sqrt{\log d}$. Because of $s_{kl}(z) = s_{kl}(z \mathbf{1}_{|z| > \tau \omega_n \sqrt{\hat{\Gamma}_{kk} \hat{\Gamma}_{ll}}})$, we have

$$\begin{aligned} \|\hat{\Gamma}^S - \Gamma\|_{\text{MAX}} & \leq \max_{1 \leq k, l \leq d} \left| \hat{\Gamma}_{kl} \mathbf{1}_{|\hat{\Gamma}_{kl}| \geq \tau \omega_n \sqrt{\hat{\Gamma}_{kk} \hat{\Gamma}_{ll}}} + \Gamma_{kl} \mathbf{1}_{|\hat{\Gamma}_{kl}| \geq \tau \omega_n \sqrt{\hat{\Gamma}_{kk} \hat{\Gamma}_{ll}}} - \Gamma_{kl} \mathbf{1}_{|\hat{\Gamma}_{kl}| \leq \tau \omega_n \sqrt{\hat{\Gamma}_{kk} \hat{\Gamma}_{ll}}} \right| \\ & \leq \max_{1 \leq k, l \leq d} \left| s_{kl}(\hat{\Gamma}_{kl}) - \hat{\Gamma}_{kl} \mathbf{1}_{|\hat{\Gamma}_{kl}| \geq \tau \omega_n \sqrt{\hat{\Gamma}_{kk} \hat{\Gamma}_{ll}}} + |\hat{\Gamma}_{kl} - \Gamma_{kl}| \mathbf{1}_{|\hat{\Gamma}_{kl}| \leq \tau \omega_n \sqrt{\hat{\Gamma}_{kk} \hat{\Gamma}_{ll}}} + \Gamma_{kl} \mathbf{1}_{|\hat{\Gamma}_{kl}| \leq \tau \omega_n \sqrt{\hat{\Gamma}_{kk} \hat{\Gamma}_{ll}}} \right| \\ & \leq C \omega_n \sqrt{\hat{\Gamma}_{kk} \hat{\Gamma}_{ll}} \mathbf{1}_{|\hat{\Gamma}_{kl}| \geq C \omega_n \theta_1} + C \omega_n \mathbf{1}_{|\hat{\Gamma}_{kl}| \geq C \omega_n \theta_1} + C \omega_n \sqrt{\hat{\Gamma}_{kk} \hat{\Gamma}_{ll}} \\ & \leq C \omega_n. \end{aligned}$$

Then we obtain

$$\|\widehat{\Gamma}^S - \Gamma\|_{\text{MAX}} \leq C m_\delta^{-1/2} \sqrt{\log d},$$

with probability at least $1 - O(C_1 d^{-C_0^2 C_2 + 2})$. ■

Lemma 2. Under [Assumptions 1–4](#), and $n_\delta^{-1/2} \sqrt{\log d} = o(1)$. We have

$$(i) \mathbb{P} \left(\|\beta(\widehat{E} - E)\beta^\top\|_\Sigma^2 + \|\beta\widehat{E}(\widehat{\beta} - \beta)^\top\|_\Sigma^2 \geq C_0 d^{-1} n_\delta^{-1} \log d \right) = O(C_1 d^{-C_0^2 C_2 + 1}),$$

and

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

Proof of Lemma 2. (i) For the first part, using the same argument in proof of theorem 2 in [Fan et al. \(2008\)](#), we have

$$\|\beta^\top \Sigma^{-1} \beta\| \leq 2 \|\text{cov}^{-1}(X)\|.$$

Therefore

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

On the other hand, we also have

$$\begin{aligned} \|\beta\widehat{E}(\widehat{\beta} - \beta)^\top\|_\Sigma^2 &\leq \frac{1}{d} \|\beta^\top \Sigma^{-1} \beta\widehat{E}(\widehat{\beta} - \beta)^\top\|_F \|\widehat{E} \Sigma^{-1} (\widehat{\beta} - \beta)^\top\|_F \\ &\leq \frac{1}{d} \|\beta^\top \Sigma^{-1} \beta\|_F \|\widehat{E}\|_F^2 \|\widehat{\beta} - \beta\|_F^2. \end{aligned}$$

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

(ii) For the second part, we have

$$\begin{aligned} \|(\widehat{\beta} - \beta)\widehat{E}(\widehat{\beta} - \beta)^\top\|_\Sigma^2 &= \frac{1}{d} \text{tr} \left((\widehat{\beta} - \beta)\widehat{E}(\widehat{\beta} - \beta)^\top \Sigma^{-1} (\widehat{\beta} - \beta)\widehat{E}(\widehat{\beta} - \beta)^\top \Sigma^{-1} \right) \\ &\leq \frac{1}{d} \|(\widehat{\beta} - \beta)\widehat{E}(\widehat{\beta} - \beta)^\top \Sigma^{-1}\|_F^2 \\ &\leq \frac{1}{d} \lambda_{\text{MAX}}^2(\Sigma^{-1}) \lambda_{\text{MAX}}^2(\widehat{E}) \|\widehat{\beta} - \beta\|_F^4. \end{aligned}$$

Since $\lambda_{\text{MAX}}^2(\Sigma^{-1})$ and $\lambda_{\text{MAX}}^2(\widehat{E})$ are both bounded, then the result follows from [\(A.6\)](#). ■

Lemma 3. Under [Assumptions 1–4](#), and $n_\delta^{-1/2} \sqrt{\log d} = o(1)$, we have

$$(i) \mathbb{P} \left(\|\widehat{\Gamma}^S - \Gamma\| > C_0 m_d n_\delta^{-(1-q)/2} (\log d)^{(1-q)/2} \right) = O(n^{3/2+\delta-3\nu/4-\nu\delta/2}), \quad (\text{A.12})$$

$$(ii) \mathbb{P} \left(\lambda_{\min}(\widehat{\Gamma}^S) \geq \frac{1}{2} \lambda_{\min}(\Gamma) \right) > 1 - O(n^{3/2+\delta-3\nu/4-\nu\delta/2}), \quad (\text{A.13})$$

$$\begin{aligned} (iii) \mathbb{P} \left(\|(\widehat{\Gamma}^S)^{-1} - \Gamma^{-1}\| > C_0 m_d n_\delta^{-(1-q)/2} (\log d)^{(1-q)/2} \right) \\ = O(n^{3/2+\delta-3\nu/4-\nu\delta/2}), \end{aligned} \quad (\text{A.14})$$

$$\begin{aligned} (iv) \mathbb{P} \left(\|\widehat{\beta}^\top (\widehat{\Gamma}^S)^{-1} \widehat{\beta} - \beta^\top \Gamma^{-1} \beta\| > C_0 m_d n_\delta^{-(1-q)/2} (\log d)^{(1-q)/2} \right) \\ = O(n^{3/2+\delta-3\nu/4-\nu\delta/2}), \end{aligned} \quad (\text{A.15})$$

$$\begin{aligned} (v) \mathbb{P} \left(\|(\widehat{E}^{-1} + \widehat{\beta}^\top (\widehat{\Gamma}^S)^{-1} \widehat{\beta}) - (E^{-1} + \beta^\top (\Gamma^S)^{-1} \beta)\| \\ > C_0 m_d n_\delta^{-(1-q)/2} (\log d)^{(1-q)/2} \right) = O(n^{3/2+\delta-3\nu/4-\nu\delta/2}), \end{aligned} \quad (\text{A.16})$$

$$(vi) \mathbb{P} \left(\|(\widehat{E}^{-1} + \widehat{\beta}^\top (\widehat{\Gamma}^S)^{-1} \widehat{\beta})^{-1}\| > C_0 m_d d^{-1} \right) = O(n^{3/2+\delta-3\nu/4-\nu\delta/2}), \quad (\text{A.17})$$

$$(vii) \quad \mathbb{P} \left(\left\| \hat{\beta} (\hat{E}^{-1} + \hat{\beta}^\top (\hat{\Gamma}^S)^{-1} \hat{\beta})^{-1} \hat{\beta}^\top (\hat{\Gamma}^S)^{-1} \right\| > C_0 m_d \right) = O(n^{3/2+\delta-3\nu/4-\nu\delta/2}), \quad (\text{A.18})$$

$$(viii) \quad \mathbb{P} \left(\left\| \hat{\beta} (\hat{E}^{-1} + \hat{\beta}^\top (\hat{\Gamma}^S)^{-1} \hat{\beta})^{-1} \hat{\beta}^\top \Gamma^{-1} \right\| > C_0 m_d \right) = O(n^{3/2+\delta-3\nu/4-\nu\delta/2}). \quad (\text{A.19})$$

Proof of Lemma 3. (i) Since $\hat{\Gamma}^S - \Gamma$ is symmetric, its operator norm is bounded by the ∞ -norm:

$$\|\hat{\Gamma}^S - \Gamma\| \leq \max_{1 \leq l \leq d} \sum_{k=1}^d |\hat{\Gamma}_{lk}^S - \Gamma_{lk}|.$$

Then using the same technique for proving (A.10), we can prove that, with probability no less than $O(n^{3/2+\delta-3\nu/4-\nu\delta/2})$, we have

$$\|\hat{\Gamma}^S - \Gamma\| \leq C_0 m_d n_\delta^{-(1-q)/2} (\log d)^{(1-q)/2}.$$

Proofs of (A.13)–(A.19) are similar to that of Lemma 5 in Fan et al. (2016), therefore we omit the details. ■

Proof of Theorem 1. Based on Lemma 1, and following the same steps as that of theorem 1 in Fan et al. (2016), we can obtain

$$\|\hat{\Sigma}_{\text{TSR}} - \Sigma\|_{\text{MAX}} = O\left(n_\delta^{-1/2} \sqrt{\log d}\right).$$

For the next part, we will prove the convergence results based on the Σ norm:

$$\begin{aligned} \|\hat{\Sigma}_{\text{TSR}} - \Sigma\|_\Sigma^2 &\leq 4 \|\beta(\hat{E} - E)\beta^\top\|_\Sigma^2 + 24 \|\beta\hat{E}(\hat{\beta} - \beta)\|_\Sigma^2 \\ &\quad + 16 \|(\hat{\beta} - \beta)\hat{E}(\hat{\beta} - \beta)^\top\|_\Sigma^2 + 2 \|\hat{\Gamma}^S - \Gamma\|_\Sigma^2. \end{aligned} \quad (\text{A.20})$$

Finally, we have

$$\begin{aligned} \|\hat{\Gamma}^S - \Gamma\|_\Sigma &= d^{-1/2} \|\Sigma^{-1/2}(\hat{\Gamma}^S - \Gamma)\Sigma^{-1/2}\|_F \leq \|\Sigma^{-1/2}(\hat{\Gamma}^S - \Gamma)\Sigma^{-1/2}\| \\ &\leq \|\hat{\Gamma}^S - \Gamma\|_{\lambda_{\max}(\Sigma^{-1})}. \end{aligned}$$

Then based on (A.20), Lemma 2 and Lemma 3 (A.12), and the fact that

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

we prove that

$$\|\hat{\Sigma}_{\text{TSR}} - \Sigma\|_\Sigma = O_p\left(d^{1/2} n_\delta^{-1} \log d + m_d^2 m_d n_\delta^{-(1-q)/2} (\log d)^{(1-q)/2}\right).$$

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

$$\|\hat{\Sigma}^* - \Sigma\|_\Sigma^2 \leq C \|\beta(\hat{E} - E)\beta^\top\|_\Sigma^2 + C \|\beta\bar{X}\bar{Z}^\top\|_\Sigma^2 + C \|\bar{Z}\bar{Z}^\top - \Gamma\|_\Sigma^2.$$

According to the proof of Lemma 2, we obtain $\|\beta(\hat{E} - E)\beta^\top\|_\Sigma^2 = O_p(d^{-1} n_\delta^{-1} \log d)$ and $\|\beta\bar{X}\bar{Z}^\top\|_\Sigma^2 = O_p(d^{-1} n_\delta^{-1} \log d)$. We can also get $\|\bar{Z}\bar{Z}^\top - \Gamma\|_\Sigma^2 = O_p(d n_\delta^{-1} \log d)$. Therefore $\|\hat{\Sigma}^* - \Sigma\|_\Sigma = O_p(d^{1/2} n_\delta^{-1/2} \sqrt{\log d})$, which has slower convergence rate than our estimator.

For the inverse part, 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

$$\begin{aligned} &\|(\hat{\Sigma}_{\text{TSR}})^{-1} - \Sigma^{-1}\| \\ &\leq \|(\hat{\Gamma}^S)^{-1} - \Gamma^{-1}\| + \|((\hat{\Gamma}^S)^{-1} - \Gamma^{-1}) \hat{\beta} (\hat{E}^{-1} + \hat{\beta}^\top (\hat{\Gamma}^S)^{-1} \hat{\beta})^{-1} \hat{\beta}^\top (\hat{\Gamma}^S)^{-1}\| \\ &\quad + \|((\hat{\Gamma}^S)^{-1} - \Gamma^{-1}) \hat{\beta} (\hat{E}^{-1} + \hat{\beta}^\top (\hat{\Gamma}^S)^{-1} \hat{\beta})^{-1} \hat{\beta}^\top \Gamma^{-1}\| \\ &\quad + \|\Gamma^{-1}(\hat{\beta} - \beta) (\hat{E}^{-1} + \hat{\beta}^\top (\hat{\Gamma}^S)^{-1} \hat{\beta})^{-1} \hat{\beta}^\top \Gamma^{-1}\| + \|\Gamma^{-1}(\hat{\beta} - \beta) (\hat{E}^{-1} + \hat{\beta}^\top (\hat{\Gamma}^S)^{-1} \hat{\beta})^{-1} \beta^\top \Gamma^{-1}\| \\ &\quad + \|\Gamma^{-1} \beta ((\hat{E}^{-1} + \hat{\beta}^\top (\hat{\Gamma}^S)^{-1} \hat{\beta})^{-1} - (\hat{E}^{-1} + \beta^\top (\Gamma^S)^{-1} \beta)^{-1}) \beta^\top \Gamma^{-1}\| \\ &:= L_1 + L_2 + L_3 + L_4 + L_5 + L_6. \end{aligned}$$

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 (A.12)

$$L_1 \leq Cm_d n_\delta^{-(1-q)/2} (\log d)^{(1-q)/2}.$$

To bound L_2 , by (A.14) and (A.18), we have

$$L_2 \leq \|(\widehat{\Gamma}^S)^{-1} - \Gamma^{-1}\| \cdot \|\widehat{\beta}(\widehat{E}^{-1} + \widehat{\beta}^\top(\widehat{\Gamma}^S)^{-1}\widehat{\beta})^{-1}\widehat{\beta}^\top(\widehat{\Gamma}^S)^{-1}\| \leq Cm_d^2 n_\delta^{-(1-q)/2} (\log d)^{(1-q)/2}.$$

Similarly, L_3 can be bounded using (A.14) and (A.19).

Next, for L_4 , we use (A.6), and (A.17), $\|\cdot\| \leq \|\cdot\|_F$, and $\lambda_{\min}(\Gamma)$ is bounded below by some constant,

$$L_4 \leq \|\Gamma^{-1}\|^2 \cdot \|\widehat{\beta} - \beta\| \cdot \|\widehat{\beta}\| \cdot \|\widehat{E}^{-1} + \widehat{\beta}^\top(\widehat{\Gamma}^S)^{-1}\widehat{\beta})^{-1}\| \leq Cm_d^2 n_\delta^{-(1-q)/2} (\log d)^{(1-q)/2}.$$

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

Finally, we have

$$\begin{aligned} L_6 &\leq \|\Gamma^{-1}\|^2 \cdot \|\beta\|^2 \cdot \|\widehat{E}^{-1} + \widehat{\beta}^\top(\widehat{\Gamma}^S)^{-1}\widehat{\beta})^{-1} - (E^{-1} + \beta^\top(\Gamma^S)^{-1}\beta)^{-1}\| \\ &\leq \|\Gamma^{-1}\|^2 \cdot \|\beta\|^2 \cdot \|\widehat{E}^{-1} + \widehat{\beta}^\top(\widehat{\Gamma}^S)^{-1}\widehat{\beta}) - (E^{-1} + \beta^\top(\Gamma^S)^{-1}\beta)\| \\ &\quad \cdot \|\widehat{E}^{-1} + \widehat{\beta}^\top(\widehat{\Gamma}^S)^{-1}\widehat{\beta})^{-1}\| \cdot \|(E^{-1} + \beta^\top\Gamma^{-1}\beta)^{-1}\|. \end{aligned}$$

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

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

It then follows that

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

On the other hand, by Assumption 3, we have

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

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

$$\|(E^{-1} + \beta^\top \Gamma^{-1} \beta)^{-1}\| = O_p(m_d d^{-1}).$$

Using (A.16) and (A.17), we have

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

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

$$\|(\widehat{\Sigma}_{\text{TSR}})^{-1} - \Sigma^{-1}\| \leq C \left(m_d^3 n_\delta^{-(1-q)/2} (\log d)^{(1-q)/2} + m_d^2 n_\delta^{-(1-q)/2} (\log d)^{(1-q)/2} \right).$$

We find that the second term on the right is dominated by the first one, then replace the whole above equation by the first term, which yields the desired result.

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

$$v^\top \widehat{\Sigma}_{\text{TSR}} v = v^\top \widehat{\beta} \widehat{E} \widehat{\beta}^\top v + v^\top \widehat{\Gamma}^S v \geq \lambda_{\min}(\widehat{\Gamma}^S),$$

which implies that

$$\lambda_{\min}(\widehat{\Sigma}_{\text{TSR}}) \geq \lambda_{\min}(\widehat{\Gamma}^S).$$

This inequality, combining with (A.13) of Lemma 3, concludes the proof. ■

A.2. Proof of Theorem 2

Proof of Theorem 2 follows the same arguments as that of Theorem 3.

A.3. Proof of Theorem 3

We note that

$$\check{X}^* - \bar{X}^* = (\beta^\top \beta)^{-1} \beta^\top \bar{Z}^*.$$

Similar to the proof of (A.1)–(A.3), and by Hoeffding inequality, we have

$$\begin{aligned}\|\beta^\top \bar{Z}^*\| &= \sqrt{\|\beta^\top \bar{Z}^* \bar{Z}^{*\top} \beta\|} \leq \sqrt{\|\beta^\top (\bar{Z}^* \bar{Z}^{*\top} - \Gamma) \beta\|} + \|\beta^\top \Gamma \beta\| \\ &\leq \sqrt{\|\beta^\top\| \|\bar{Z}^* \bar{Z}^{*\top} - \Gamma\| \|\beta\|} + \|\beta^\top\| \|\Gamma\| \|\beta\| \\ &\leq C \sqrt{d m_d n_\delta^{-(1-q)/2} (\log d)^{(1-q)/2} + d m_d} \\ &\leq C d^{1/2} m_d^{1/2} n_\delta^{-(1-q)/4} (\log d)^{(1-q)/4} + d^{1/2} m_d^{1/2},\end{aligned}$$

where we use $\|\Gamma\| \leq m_d$, and $\|\bar{Z}^* \bar{Z}^{*\top} - \Gamma\| \stackrel{p}{=} O_p(n_\delta \sqrt{\log d})$.

Then we have

$$\begin{aligned}\|\check{X}^* - \bar{X}^*\| &\leq \|(\beta^\top \beta)^{-1}\| \|\beta^\top \bar{Z}^*\| \leq \frac{\|\beta^\top \bar{Z}^*\|}{\lambda_{\min}(\beta^\top \beta)} \\ &\leq \frac{\|\beta^\top \bar{Z}^*\| d^{-1}}{\lambda_{\min}(d^{-1} \beta^\top \beta)} \leq C d^{-1/2} m_d^{1/2} n_\delta^{-(1-q)/4} (\log d)^{(1-q)/4} + d^{-1/2} m_d^{1/2}.\end{aligned}$$

Moreover, we note that

$$\begin{aligned}\max_{1 \leq k \leq d} \sum_{i=0}^{n-k_n+1} \left(\sum_{l=1}^r \beta_{k,l} (\hat{X}_{l,i}^* - \bar{X}_{l,i}^*) \right)^2 &\leq \max_{1 \leq k \leq d, 1 \leq i \leq n} \|\beta_k\|^2 \|\check{X}^* - \bar{X}^*\|_F^2 \\ &\leq \max_{1 \leq k \leq d, 1 \leq i \leq n} \|\beta_k\|^2 r \|\check{X}^* - \bar{X}^*\|^2, \\ &\leq C d^{-1} m_d n_\delta^{-(1-q)/2} (\log d)^{(1-q)/2} + d^{-1} m_d.\end{aligned}$$

Using these estimates, we obtain

$$\begin{aligned}&\max_{1 \leq k, l \leq d} \left| \sum_{i=0}^{n-k_n+1} [(\bar{Y}_{k,i}^* - (\beta \hat{X}_i^*)_k) (\bar{Y}_{l,i}^* - (\beta \hat{X}_i^*)_l) - \bar{Z}_{o,ki}^* \bar{Z}_{o,li}^*] \right| \\ &\leq \max_{1 \leq k, l \leq d} \left| \sum_{i=0}^{n-k_n+1} (\beta (\check{X}^* - \bar{X}_i^*))_k (\beta (\check{X}^* - \bar{X}_i^*))_l \right| + 2 \max_{1 \leq l, k \leq d} \left| \sum_{i=0}^{n-k_n+1} \bar{Z}_{o,ki}^* (\beta (\check{X}^* - \bar{X}_i^*))_l \right| \\ &\leq \max_{1 \leq k \leq d} \sum_{i=0}^{n-k_n+1} (\beta (\check{X}^* - \bar{X}_i^*))_k^2 + 2 \sqrt{\max_{1 \leq k \leq d} \sum_{i=0}^{n-k_n+1} (\bar{Z}_{o,ki}^*)^2 \max_{1 \leq k \leq d} \sum_{i=0}^{n-k_n+1} (\beta (\check{X}^* - \bar{X}_i^*))_k^2} \\ &\leq C d^{-1/2} m_d^{1/2} n_\delta^{-(1-q)/4} (\log d)^{(1-q)/4} + d^{-1/2} m_d^{1/2}.\end{aligned}\tag{A.21}$$

Therefore, according to (A.4) and (A.21), and using triangle inequality, we have

$$\begin{aligned}&\max_{1 \leq l, k \leq d} |\hat{\Gamma}'_{kl} - \Gamma_{kl}| \\ &\leq \max_{1 \leq k \leq r, 1 \leq l \leq d} \left| \frac{n}{n - k_n + 2} \frac{1}{\psi_2 k_n} \sum_{i=0}^{n-k_n+1} \bar{Z}_{k,i}^* \bar{Z}_{l,i}^* - \int_0^t g_{s,lk} ds \right| \\ &\quad + \max_{1 \leq k, l \leq d} \left| \frac{n}{n - k_n + 2} \frac{1}{\psi_2 k_n} \sum_{i=0}^{n-k_n+1} [(\bar{Y}_{k,i}^* - (\beta \hat{X}_i^*)_k) (\bar{Y}_{l,i}^* - (\beta \hat{X}_i^*)_l) - \bar{Z}_{o,ki}^* \bar{Z}_{o,li}^*] \right| \\ &\leq C \left(n_\delta^{-1/2} \sqrt{\log d} + C d^{-1/2} m_d^{1/2} n_\delta^{-(1-q)/4} (\log d)^{(1-q)/4} + d^{-1/2} m_d^{1/2} \right) \\ &\leq C \left(n_\delta^{-1/2} \sqrt{\log d} + d^{-1/2} m_d^{1/2} \right).\end{aligned}$$

The rest steps are similar to the TSR case. This concludes the proof.

A.4. Proof of Theorem 4

Proposition 1. Suppose that Assumptions 1–4 hold. Also, assume that $\|E\|_{\text{MAX}} \leq K$, $\|\Gamma\|_{\text{MAX}} \leq K$ almost surely for some constant K , $n_\delta^{-1/2} \sqrt{\log d} = o(1)$, and $d^{-1/2} m_d = o(1)$. Then r , $\beta E \beta^\top$, and Γ can be identified as $d \rightarrow \infty$. That is, $\bar{r} = r$, if d is sufficiently large. Moreover, we have

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

$$\left\| \sum_{j=\bar{r}+1}^d \lambda_j \xi_j \xi_j^\top - \Gamma \right\|_{\text{MAX}} \leq Cd^{-1/2} m_d,$$

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 $\bar{r} = \arg \min_{1 \leq j \leq d} (\frac{\lambda_j}{d} + jd^{-1/2} m_d) - 1$.

Lemma 4. Suppose [Assumptions 1–4](#) hold, and $n_\delta^{-1/2} \sqrt{\log d} = o(1)$, then we have

$$\max_{1 \leq k \leq r, 1 \leq l \leq d} \left| \frac{n}{n - k_n + 2} \frac{1}{\psi_2 k_n} \sum_{i=0}^{n-k_n+1} \bar{X}_i^k \bar{X}_i^l - \int_0^t e_{s,kl} ds \right| = O_p(n_\delta^{-1/2} \sqrt{\log d}), \quad (\text{A.22})$$

$$\max_{1 \leq k \leq r, 1 \leq l \leq d} \left| \frac{n}{n - k_n + 2} \frac{1}{\psi_2 k_n} \sum_{i=0}^{n-k_n+1} \bar{X}_i^k \bar{Z}_{u,li}^* \right| = O_p(n_\delta^{-1/2} \sqrt{\log d}), \quad (\text{A.23})$$

$$\max_{1 \leq k \leq r, 1 \leq l \leq d} \left| \frac{n}{n - k_n + 2} \frac{1}{\psi_2 k_n} \sum_{i=0}^{n-k_n+1} \bar{Z}_{u,ki}^* \bar{Z}_{u,li}^* - \int_0^t g_{s,kl} ds \right| = O_p(n_\delta^{-1/2} \sqrt{\log d}). \quad (\text{A.24})$$

Recall that

$$\Lambda = \text{Diag}(\hat{\lambda}_1, \hat{\lambda}_2, \dots, \hat{\lambda}_r), \quad F = d^{1/2} (\hat{\xi}_1, \hat{\xi}_2, \dots, \hat{\xi}_r), \quad \text{and} \quad G = d^{-1} F^\top \mathcal{Y}.$$

We write

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

It is easy to verify that

$$\begin{aligned} \widehat{\Sigma} F &= F \Lambda, \quad G G^\top = t d^{-1} \times \Lambda, \quad F^\top F = d \times \mathbb{I}_r, \quad \text{and} \\ \widehat{\Gamma} &= \frac{1}{t} (\mathcal{Y} - F G) (\mathcal{Y} - F G)^\top = \frac{1}{t} \mathcal{Y} \mathcal{Y}^\top - \frac{1}{d} F \Lambda F^\top. \end{aligned}$$

Lemma 5. Suppose [Assumptions 1–4](#) hold with $\lambda_{ij} = O(n_\delta^{-1/2} \sqrt{\log d})$. Suppose $d^{-1/2} m_d = o(1)$, $n_\delta^{-1/2} \sqrt{\log d} = o(1)$, and $\hat{r} \rightarrow r$ with probability approaching 1, then there exists a $r \times r$ matrix H , such that with probability approaching 1, H is invertible, $\|H H^\top - \mathbb{I}_r\| = \|H^\top H - \mathbb{I}_r\| = o_p(1)$, and more importantly,

$$\begin{aligned} \|F - \beta H\|_{\text{MAX}} &= O_p(n_\delta^{-1/2} \sqrt{\log d} + d^{-1/2} m_d), \\ \|G - H^{-1} \bar{\mathcal{X}}\| &= O_p(n_\delta^{-1/2} \sqrt{\log d} + d^{-1/2} m_d). \end{aligned}$$

Lemma 6. Under [Assumptions 1–4](#), $d^{-1/2} m_d = o(1)$, and $n_\delta^{-1/2} \sqrt{\log d} = o(1)$, we have

$$\|F - \beta H\|_{\text{MAX}} = O_p(n_\delta^{-1/2} \sqrt{\log d} + d^{-1/2} m_d). \quad (\text{A.25})$$

$$\|H^{-1}\| = O_p(1). \quad (\text{A.26})$$

$$\|G - H^{-1} \bar{\mathcal{X}}\| = O_p(n_\delta^{-1/2} \sqrt{\log d} + d^{-1/2} m_d). \quad (\text{A.27})$$

Lemma 7. Under [Assumptions 1–4](#), $d^{-1/2} m_d = o(1)$, and $n_\delta^{-1/2} \sqrt{\log d} = o(1)$, we have

$$\|\widetilde{F}^S - \Gamma\|_{\text{MAX}} \leq \|\widehat{\Gamma} - \Gamma\|_{\text{MAX}} = O_p(n_\delta^{-1/2} \sqrt{\log d} + d^{-1/2} m_d). \quad (\text{A.28})$$

Lemma 8. Under [Assumptions 1–4](#), $d^{-1/2} m_d = o(1)$, and $n_\delta^{-1} \log d = o(1)$, we have

$$\left\| \frac{1}{t} F G G^\top F^\top - \beta E \beta^\top \right\|_{\text{MAX}} = O_p(n_\delta^{-1/2} \sqrt{\log d} + d^{-1/2} m_d).$$

Proof of Proposition 1, Lemmas 4–8. The proofs follow the same arguments as in [Aït-Sahalia and Xiu \(2017b\)](#), thus we omit the details. ■

Lemma 9. Under [Assumptions 1–4](#), $d(n_\delta^{-1/2} \sqrt{\log d})^2 = o(1)$, $n_\delta^{-1} \log d = o(1)$, and $d^{-1/2} m_d = o(1)$, we have

$$(i) \quad \|F - \beta H\|_F^2 = O_p(d(n_\delta^{-1/2} \sqrt{\log d})^2 + m_d^2). \quad (\text{A.29})$$

$$(ii) \| (F - \beta H)(F - \beta H)^\top \|_F^2 = O_p(d n^{2\delta-1} \log^2 d + d^{-1} m_d^4). \quad (A.30)$$

$$(iii) \|\beta H(F - \beta H)^\top\|_\Sigma^2 = O_p(n^{\delta-1/2} \log d + d^{-1} m_d^2). \quad (A.31)$$

$$(iv) \|\beta(H^\top H - I_r)\beta^\top\|_\Sigma^2 = o_p(1). \quad (A.32)$$

Proof of Lemma 9. (i) We have $\|F - \beta H\|_F^2 \geq d \|F - \beta H\|_{\text{MAX}}^2$.

(ii) According to the definition of $\|\cdot\|_\Sigma$, we have

$$\|(F - \beta H)(F - \beta H)^\top\|_\Sigma^2 = O_p\left(\frac{1}{d} \|F - \beta H\|_F^4\right) = O_p(d \|F - \beta H\|_{\text{MAX}}^4).$$

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

$$\begin{aligned} \|\beta H(F - \beta H)^\top\|_\Sigma^2 &= \frac{1}{d} \text{tr}(H(F - \beta H)^\top \Sigma^{-1} (F - \beta H) H \beta^\top \Sigma^{-1} \beta) \\ &\leq \frac{1}{d} \|H\|^2 \|\beta^\top \Sigma^{-1} \beta\| \|\Sigma^{-1}\| \|F - \beta H\|_F^2 \\ &= O_p(\|F - \beta H\|_{\text{MAX}}^2). \end{aligned}$$

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

$$\begin{aligned} \|\beta(H^\top H - I_r)\beta^\top\|_\Sigma^2 &\geq \frac{1}{d} \text{tr}\{(H^\top H - I_r)\beta^\top \Sigma^{-1} \beta (H^\top H - I_r)\beta^\top \Sigma^{-1} \beta\} \\ &\geq \frac{1}{d} \|\beta^\top \Sigma^{-1} \beta\|^2 \|H^\top H - I_r\|_F^2 \\ &= o_p(1). \quad \blacksquare \end{aligned}$$

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

$$\|\tilde{\Gamma}^S - \Gamma\| = O_p\left(m_d(n_\delta^{-1/2} \sqrt{\log d} + d^{-1/2} m_d)^{1-q}\right). \quad (A.33)$$

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

$$\|(\tilde{\Gamma}^S)^{-1} - \Gamma^{-1}\| = O_p\left(m_d(n_\delta^{-1/2} \sqrt{\log d} + d^{-1/2} m_d)^{1-q}\right). \quad (A.34)$$

Proof of Lemma 10.

Note that since $\tilde{\Gamma}^S - \Gamma$ is symmetric,

$$\|\tilde{\Gamma}^S - \Gamma\| \leq \|\tilde{\Gamma}^S - \Gamma\|_\infty = \max_{1 \leq l \leq d} \sum_{k=1}^d |\hat{\Gamma}_{lk}^S - \Gamma_{lk}|.$$

By Lemma 7, and using the same technique as proving (A.10), we have

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

Choosing $\lambda_{ij} = M'(n_\delta^{-1/2} \sqrt{\log d} + d^{-1/2} m_d)$, M' is some positive constant, we have

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

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

$$\begin{aligned} \|(\tilde{\Gamma}^S)^{-1} - \Gamma^{-1}\| &= \|(\tilde{\Gamma}^S)^{-1} (\Gamma - \tilde{\Gamma}^S) \Gamma^{-1}\| \leq \lambda_{\min}(\tilde{\Gamma}^S)^{-1} \lambda_{\min}(\Gamma)^{-1} \|\Gamma - \tilde{\Gamma}^S\| \\ &\leq O_p\left(m_d(n_\delta^{-1/2} \sqrt{\log d} + d^{-1/2} m_d)^{1-q}\right). \quad \blacksquare \end{aligned}$$

Proof of Theorem 4.

Note that

$$\hat{\Sigma}_{\text{PCA}} = \frac{1}{d} F \Lambda F^\top + \tilde{\Gamma}^S = \frac{1}{t} F G G^\top F^\top + \tilde{\Gamma}^S.$$

By Lemma 7, we have

$$\|\tilde{\Gamma}^S - \Gamma\|_{\text{MAX}} = O_p\left(n_\delta^{-1/2} \sqrt{\log d} + d^{-1/2} m_d\right).$$

By the triangle inequality, we have

$$\|\widehat{\Sigma}_{\text{PCA}} - \Sigma\|_{\text{MAX}} \leq \left\| \frac{1}{d} F \Lambda F^\top - \beta E \beta^\top \right\|_{\text{MAX}} + \|\widetilde{\Gamma}^S - \Gamma\|_{\text{MAX}}$$

Therefore, the desired result follows from [Lemmas 7](#) and [8](#).

Using [Lemma 9](#), for some constant C , we have

$$\begin{aligned} \|\widehat{\Sigma}_{\text{PCA}} - \Sigma\|_{\Sigma}^2 &\geq C \left[\|\beta(H^\top H - I_r)\beta^\top\|_{\Sigma}^2 + \|\beta H(F - \beta H)^\top\|_{\Sigma}^2 + \|(F - \beta H)(F - \beta H)^\top\|_{\Sigma}^2 + \|\widetilde{\Gamma}^S - \Gamma\| \right] \\ &= O_p(d(n_\delta^{-1/2} \sqrt{\log d})^4 + \frac{1}{d} m_d^4 + m_d^2(n_\delta^{-1/2} \sqrt{\log d} + d^{-1/2} m_d)^{2(1-q)}). \end{aligned}$$

For the inverse, firstly, by [Lemma 10](#) and the fact that $\lambda_{\min}(\widehat{\Sigma}_{\text{PCA}}) \geq \lambda_{\min}(\widetilde{\Gamma}^S)$, we can establish the first two statements.

To bound $\|(\widehat{\Sigma}_{\text{PCA}})^{-1} - \Sigma^{-1}\|$, by the Sherman–Morrison–Woodbury formula, we have

$$\begin{aligned} &(\widehat{\Sigma}_{\text{PCA}})^{-1} - (\widetilde{\Sigma})^{-1} \\ &= (t^{-1} F G G^\top F^\top + \widetilde{\Gamma}^S)^{-1} - (t^{-1} \beta H H^{-1} \bar{\chi}^* \bar{\chi}^{*\top} (H^{-1})^\top H^\top \beta^\top + \Gamma)^{-1} \\ &= ((\widetilde{\Gamma}^S)^{-1} - \Gamma^{-1}) - ((\widetilde{\Gamma}^S)^{-1} - \Gamma^{-1}) F (d\Lambda^{-1} + F^\top (\widetilde{\Gamma}^S)^{-1} F)^{-1} F^\top (\widetilde{\Gamma}^S)^{-1} \\ &\quad - \Gamma^{-1} F (d\Lambda^{-1} + F^\top (\widetilde{\Gamma}^S)^{-1} F)^{-1} F^\top ((\widetilde{\Gamma}^S)^{-1} - \Gamma^{-1}) \\ &\quad + \Gamma^{-1} (\beta H - F) \left(t H^\top (\bar{\chi} \bar{\chi}^\top)^{-1} H + H^\top \beta^\top \Gamma^{-1} \beta H \right)^{-1} H^\top \beta^\top \Gamma^{-1} \\ &\quad - \Gamma^{-1} F \left(t H^\top (\bar{\chi} \bar{\chi}^\top)^{-1} H + H^\top \beta^\top \Gamma^{-1} \beta H \right)^{-1} (F^\top - H^\top \beta^\top) \Gamma^{-1} \\ &\quad + \Gamma^{-1} F \left(\left(t H^\top (\bar{\chi} \bar{\chi}^\top)^{-1} H + H^\top \beta^\top \Gamma^{-1} \beta H \right)^{-1} - (d\Lambda^{-1} + F^\top (\widetilde{\Gamma}^S)^{-1} F)^{-1} \right) F^\top \Gamma^{-1} \\ &= L_1 + L_2 + L_3 + L_4 + L_5 + L_6. \end{aligned}$$

By [Lemma 10](#), we have

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

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

$$\lambda_{\min}(d\Lambda^{-1} + F^\top (\widetilde{\Gamma}^S)^{-1} F) \geq \lambda_{\min}(F^\top (\widetilde{\Gamma}^S)^{-1} F) \geq \lambda_{\min}(F^\top F) \lambda_{\min}((\widetilde{\Gamma}^S)^{-1}) \geq m_d^{-1} d,$$

and by [Lemma 10](#), we have

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

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

$$\begin{aligned} \lambda_{\min} \left(t H^\top (\bar{\chi} \bar{\chi}^\top)^{-1} H + H^\top \beta^\top \Gamma^{-1} \beta H \right) &\geq \lambda_{\min}(H^\top \beta^\top \Gamma^{-1} \beta H) \\ &\geq \lambda_{\min}(\Gamma^{-1}) \lambda_{\min}(\beta^\top \beta) \lambda_{\min}(H^\top H) \\ &> K m_d^{-1} d, \end{aligned}$$

hence we have

$$\begin{aligned} \|L_4\| &\leq \|\Gamma^{-1}\| \|(\beta H - F)\| \left\| \left(t H^\top (\bar{\chi} \bar{\chi}^\top)^{-1} H + H^\top \beta^\top \Gamma^{-1} \beta H \right)^{-1} \right\| \|H^\top \beta^\top\| \|\Gamma^{-1}\| \\ &= O_p(m_d n_\delta^{-1/2} \sqrt{\log d} + d^{-1/2} m_d^2). \end{aligned}$$

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

$$\begin{aligned} &\left\| \left(t H^\top (\bar{\chi} \bar{\chi}^\top)^{-1} H + H^\top \beta^\top \Gamma^{-1} \beta H \right)^{-1} - (d\Lambda^{-1} + F^\top (\widetilde{\Gamma}^S)^{-1} F)^{-1} \right\| \\ &\leq K d^{-2} m_d^2 \left\| \left(t H^\top (\bar{\chi} \bar{\chi}^\top)^{-1} H + H^\top \beta^\top \Gamma^{-1} \beta H \right) - (d\Lambda^{-1} + F^\top (\widetilde{\Gamma}^S)^{-1} F) \right\|. \end{aligned}$$

Moreover, since we have

$$\begin{aligned}\|tH^{\top}(\bar{\chi}\bar{\chi}^{\top})^{-1}H - d\Lambda^{-1}\| &= \|\Lambda^{-1}F^{\top}(\beta H - F)\| \\ &= O_p\left(n_{\delta}^{-1/2}\sqrt{\log d} + d^{-1/2}m_d\right),\end{aligned}$$

and

$$\begin{aligned}&\|H^{\top}\beta^{\top}\Gamma^{-1}\beta H - F^{\top}(\tilde{\Gamma}^S)^{-1}F\| \\ &\leq \|(H^{\top}\beta^{\top} - F^{\top})\Gamma^{-1}\beta H\| + \|F^{\top}\Gamma^{-1}(\beta H - F)\| + \|F^{\top}(\Gamma^{-1} - (\tilde{\Gamma}^S)^{-1})F\| \\ &= O_p\left(dm_d(n_{\delta}^{-1/2}\sqrt{\log d} + d^{-1/2}m_d)^{1-q}\right).\end{aligned}$$

Combining these inequalities yields

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

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

$$\begin{aligned}&\|\tilde{\Sigma}^{-1} - \Sigma^{-1}\| \\ &= \|(t^{-1}\beta\bar{\chi}\bar{\chi}^{\top}\beta^{\top} + \Gamma)^{-1} - (\beta E\beta^{\top} + \Gamma)^{-1}\| \\ &\leq \|\Gamma^{-1}\|^2\|\beta H\|^2\left\|\left((tH^{\top}(\bar{\chi}\bar{\chi}^{\top})^{-1}H + H^{\top}\beta^{\top}\Gamma^{-1}\beta H)^{-1} - (H^{\top}E^{-1}H + H^{\top}\beta^{\top}\Gamma^{-1}\beta H)^{-1}\right)\right\| \\ &\leq Kd\|tH^{\top}(\bar{\chi}\bar{\chi}^{\top})^{-1}H + H^{\top}\beta^{\top}\Gamma^{-1}\beta H\|^{-1}\|H^{\top}E^{-1}H + H^{\top}\beta^{\top}\Gamma^{-1}\beta H\|^{-1}\|t(\bar{\chi}\bar{\chi}^{\top})^{-1} - E^{-1}\| \\ &= O_p\left(m_d n_{\delta}^{-1/2}\sqrt{\log d}\right).\end{aligned}$$

By the triangle inequality, we obtain

$$\begin{aligned}\|(\hat{\Sigma}_{PCA})^{-1} - \Sigma^{-1}\| &\leq \|(\hat{\Sigma}_{PCA})^{-1} - \tilde{\Sigma}^{-1}\| + \|\tilde{\Sigma}^{-1} - \Sigma^{-1}\| \\ &= O_p\left(m_d^3(n_{\delta}^{-1/2}\sqrt{\log d} + d^{-1/2}m_d)^{1-q}\right).\end{aligned}$$

This concludes the proof. ■

References

- Aït-Sahalia, Y., Fan, J., Xiu, D., 2010. High-frequency covariance estimates with noisy and asynchronous financial data. *J. Amer. Statist. Assoc.* 105 (492), 1504–1517.
- Aït-Sahalia, Y., Jacod, J., 2014. *High Frequency Financial Econometrics*. Princeton University Press.
- Aït-Sahalia, Y., Xiu, D., 2016. A Hausman test for the presence of market microstructure noise in high frequency data. *J. Econometrics* (forthcoming).
- Aït-Sahalia, Y., Xiu, D., 2017a. Principal component analysis of high frequency data. *J. Amer. Statist. Assoc.* (forthcoming).
- Aït-Sahalia, Y., Xiu, D., 2017b. Using principal component analysis to estimate a high dimensional factor model with high-frequency data. *J. Econometrics* 201 (2), 384–399.
- Bai, J., 2003. Inferential theory for factor models of large dimensions. *Econometrica* 71 (1), 135–171.
- Bai, J., Ng, S., 2002. Determining the number of factors in approximate factor models. *Econometrica* 70 (1), 191–221.
- Bai, J., Ng, S., 2006. Evaluating latent and observed factors in macroeconomics and finance. *J. Econometrics* 131 (1), 507–537.
- Barndorff-Nielsen, O.E., Hansen, P.R., Lunde, A., Shephard, N., 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.
- Bibinger, M., 2012. An estimator for the quadratic covariation of asynchronously observed Itô processes with noise: Asymptotic distribution theory. *Stochastic Process. Appl.* 122, 2411–2453.
- Bibinger, M., Hautsch, N., Malec, P., Reiss, M., 2014. Estimating the quadratic covariation matrix from noisy observations: Local method of moments and efficiency. *Ann. Statist.* 42 (4), 80–114.
- Bickel, P.J., Levina, E., 2008a. Regularized estimation of large covariance matrices. *Ann. Statist.* 36 (1), 199–227.
- Bickel, P.J., Levina, E., 2008b. Covariance regularization by thresholding. *Ann. Statist.* 36 (6), 2577–2604.
- Brogaard, J.A., Hendershott, T., Riordan, R., 2014. High frequency trading and price discovery. *Rev. Financ. Stud.* 27, 2267–2306.
- Brownlees, C.T., Nualart, E., Sun, Y., 2017. Realized networks. Discussion Paper, Universitat Pompeu Fabra - Department of Economics and Business; Barcelona Graduate School of Economics (Barcelona GSE).
- Cai, T., Liu, W., 2011. Adaptive thresholding for sparse covariance matrix estimation. *J. Amer. Statist. Assoc.* 106 (494), 672–684.
- Carhart, M.M., 1997. On persistence in mutual fund performance. *J. Finance* 52 (1), 57–82.
- Chamberlain, G., Rothschild, M., 1983. Arbitrage, factor structure, and mean-variance analysis on large asset markets. *Econometrica* 51 (5), 1281–1304.
- Christensen, K., Kinnebrock, S., Podolskij, M., 2010. Pre-averaging estimators of the ex-post covariance matrix in noisy diffusion models with non-synchronous data. *J. Econometrics* 159 (1), 116–133.
- Connor, G., Hagmann, M., Linton, O., 2012. Efficient semiparametric estimation of the fama–franch model and extensions. *Econometrica* 80 (2), 713–754.
- Connor, G., Linton, O., 2007. Semiparametric estimation of a characteristic-based factor model of common stock returns. *J. Empir. Finance* 14 (5), 694–717.
- Croux, C., Renault, E., Werker, B., 2004. Dynamic factor models. *J. Econometrics* 119, 223–230.
- Da, R., Xiu, D., 2017. When Moving-Average Models Meet High-Frequency Data: Uniform Inference on Volatility. Discussion Paper, University of Chicago.
- Daniel, K., Titman, S., 1997. Evidence on the characteristics of cross sectional variation in stock returns. *J. Finance* 52 (1), 1–33.
- Diebold, F.X., Mariano, R.S., 2002. Comparing predictive accuracy. *J. Bus. Econom. Statist.* 20 (1), 134–144.

- Donoho, D.L., et al., 2000. High-dimensional data analysis: the curses and blessings of dimensionality. In: AMS Math Challenges Lecture, pp. 1–32.
- Doz, C., Giannone, D., Reichlin, L., 2011. A two-step estimator for large approximate dynamic factor models based on Kalman Filtering. *J. Econometrics* 164, 188–205.
- Engle, R.F., Russell, J.R., 1998. Autoregressive conditional duration: a new model for irregularly spaced transaction data. *Econometrica* 1127–1162.
- Fama, E.F., French, K.R., 1993. Common risk factors in the returns on stocks and bonds. *J. Financ. Econ.* 33 (1), 3–56.
- Fama, E.F., French, K.R., 2015. A five-factor asset pricing model. *J. Financ. Econ.* 116 (1), 1–22.
- Fan, J., Fan, Y., Lv, J., 2008. High dimensional covariance matrix estimation using a factor model. *J. Econometrics* 147 (1), 186–197.
- Fan, J., Furger, A., Xiu, D., 2016. Incorporating global industrial classification standard into portfolio allocation: a simple factor-based large covariance matrix estimator with high-frequency data. *J. Bus. Econom. Statist.* 34 (4), 489–503.
- Fan, J., Li, R., 2001. Variable selection via nonconcave penalized likelihood and its oracle properties. *J. Amer. Statist. Assoc.* 96 (456), 1348–1360.
- Fan, J., Liao, Y., Mincheva, M., 2011. High-dimensional covariance matrix estimation in approximate factor models. *Ann. Statist.* 39 (6), 3320–3356.
- Fan, J., Liao, Y., Mincheva, M., 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.
- Fan, J., Zhang, J., Yu, K., 2012. Vast portfolio selection with gross-exposure constraints. *J. Amer. Statist. Assoc.* 107 (498), 592–606.
- Forni, M., Hallin, M., Lippi, M., Reichlin, L., 2000. The generalized dynamic-factor model: Identification and estimation. *Rev. Econ. Stat.* 82, 540–554.
- Forni, M., Hallin, M., Lippi, M., Reichlin, L., 2004. The generalized dynamic factor model: Consistency and rates. *J. Econometrics* 119 (2), 231–255.
- Forni, M., Lippi, M., 2001. The generalized dynamic factor model: Representation theory. *Econometric Theory* 17, 1113–1141.
- Hasbrouck, J., 2007. *Empirical Market Microstructure*. Oxford University Press, New York, NY.
- Hautsch, N., Kyj, L.M., Oomen, R.C., 2012. A blocking and regularization approach to high-dimensional realized covariance estimation. *J. Appl. Econometrics* 27 (4), 625–645.
- Ikedu, S.S., 2016. A bias-corrected estimator of the covariation matrix of multiple security prices when both microstructure effects and sampling durations are persistent and endogenous. *J. Econometrics* 193 (1), 203–214.
- Jacod, J., Li, Y., Mykland, P.A., Podolskij, M., Vetter, M., 2009. Microstructure noise in the continuous case: the pre-averaging approach. *Stochastic Process. Appl.* 119 (7), 2249–2276.
- Kahn, R.N., Brougham, P., Green, P., 1998. *United States Equity Model (USE3) - Risk Model Handbook*. MSCI INC.
- Kalnina, I., Linton, O., 2008. Estimating quadratic variation consistently in the presence of endogenous and diurnal measurement error. *J. Econometrics* 147, 47–59.
- Kim, D., Wang, Y., Zou, J., 2016. Asymptotic theory for large volatility matrix estimation based on high-frequency financial data. *Stochastic Process. Appl.* 126 (11), 3527–3577.
- King, B.F., 1966. Market and industry factors in stock price behavior. *J. Bus.* 39 (1), 139–190.
- Ledoit, O., Wolf, M., 2004. Honey, i shrunk the sample covariance matrix. *J. Portfolio Manage.* 30 (4), 110–119.
- Ledoit, O., Wolf, M., 2012. Nonlinear shrinkage estimation of large-dimensional covariance matrices. *Ann. Statist.* 40 (2), 1024–1060.
- Li, Y., Mykland, P., Renault, E., Zhang, L., Zheng, X., 2014. Realized volatility when sampling times are possibly endogenous. *Econometric Theory* 30, 580–605.
- Martens, M., 2004. *Estimating Unbiased and Precise Realized Covariances*. Discussion Paper, Erasmus University Rotterdam (EUR); Robeco Asset Management.
- Moon, H.R., Weidner, M., 2015. Linear regression for panel with unknown number of factors as interactive fixed effects. *Econometrica* 83 (4), 1543–1579.
- Onatski, A., 2010. Determining the number of factors from empirical distribution of eigenvalues. *Rev. Econ. Stat.* 92, 1004–1016.
- Pelger, M., 2015a. Large-dimensional factor modeling based on high-frequency observations. Available at SSRN 2584172.
- Pelger, M., 2015b. Understanding systematic risk: a high-frequency approach. Available at SSRN 2647040.
- Rosenberg, B., 1974. Extra-market components of covariance in security returns. *J. Financ. Quant. Anal.* 9 (02), 263–274.
- Ross, S.A., 1976. The arbitrage theory of capital asset pricing. *J. Econom. Theory* 13 (3), 341–360.
- Rothman, A.J., Levina, E., Zhu, J., 2009. Generalized thresholding of large covariance matrices. *J. Amer. Statist. Assoc.* 104 (485), 177–186.
- Shephard, N., Xiu, D., 2017. Econometric analysis of multivariate realised QML: Estimation of the covariation of equity prices under asynchronous trading. *J. Econometrics* 201 (1), 19–42.
- Tao, M., Wang, Y., Chen, X., 2013a. Fast convergence rates in estimating large volatility matrices using high-frequency financial data. *Econometric Theory* 29 (4), 838–856.
- Tao, M., Wang, Y., Yao, Q., Zou, J., 2011. Large volatility matrix inference via combining low-frequency and high-frequency approaches. *J. Amer. Statist. Assoc.* 106 (495), 1025–1040.
- Tao, M., Wang, Y., Zhou, H.H., 2013b. Optimal sparse volatility matrix estimation for high-dimensional Itô processes with measurement errors. *Ann. Statist.* 41 (4), 1816–1864.
- Todorov, V., Bollerslev, T., 2010. Jumps and betas: A new framework for disentangling and estimating systematic risks. *J. Econometrics* 157, 220–235.
- Varneskov, R.T., 2016. Flat-top realized kernel estimation of quadratic covariation with nonsynchronous and noisy asset prices. *J. Bus. Econom. Statist.* 34 (1), 1–22.
- Wang, Y., Zou, J., 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.
- Zou, H., 2006. The adaptive lasso and its oracle properties. *J. Amer. Statist. Assoc.* 101 (476), 1418–1429.